To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Search this book:

Go!

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Table of Contents

Introduction

You need not be reminded how fast and radically IT is changing in today's organizations. What was essentially a centralized backbone function has become an organization's lifeblood, running throughout it and supplying each department with the information needed for survival. Accordingly, your job has changed from ensuring the efficient operation of the central mainframe to supporting IT functions to users throughout an entire organization, which may range across several locations. And accordingly, we have made some major changes to this handbook.

Auerbach publishes this handbook as a guide to help you get your job done efficiently and effectively. Over the past several years, we editors have noted how your jobs have been changing to include new and different responsibilities. As we watched these changes, we wondered where they were heading and where we should take the handbook. We realized these changes were transforming the traditional data center into something different, but what exactly was that something? And what were we going to call it?

To answer these questions, we went to the best source: you, our readers. What we learned from talking to you, we used as our guide in making changes to the handbook so we can provide the information you need. The first change was the title, which went from *Handbook of Data Center Management to Handbook of Enterprise Operations Management*. Then, we organized a set of sections and chapters that address the issues facing you on the job everyday.

Client/server and then the Internet have propelled IT to users throughout organizations, and organizations increasingly rely on these technologies to get their business done. The term enterprise is a reflection of these realities. It means a business organization and the business it conducts, which is what you and your staff support by ensuring IT operations.

ENTERPRISE TECHNOLOGY

We have provided overviews and how-to info on a broad range of information technology that run today's enterprises. This technology includes:

- Mainframe-based systems
- Local area networks (LANs)
- Internet-based systems
- Telecommunications

Managing these types of systems are included in the following sections of this handbook:

- Data center management
- Enterprise systems planning and management
- Networking

We included this mix of technology and topics to reflect the streams of technology running through today's organizations. Mainframes and batch systems are still the workhorses of the enterprise, and operations of these systems still require management. LANs and Internet-based systems are the technology driving many businesses and giving a competitive edge. Of course, these systems are interconnected, and the handbook also addresses these interfacing issues.

Supporting enterprise operations does not involve only hardware and software. It also involves the interaction between technology and people, and this edition of the handbook includes a group of chapters on help desk and end-user support issues. Coverage of these topics ranges from management and planning of end-user support to tips on how best to run a help desk.

Help desks are designed to help people use computer technology correctly. Another important concern is ensuring that an organization's systems are being used properly by those who are authorized to use them. The handbook includes an entire section on security issues. Its chapters cover basic security technologies, the issues involved in designing an enterprisewide security architecture, and methods for catching intruders.

Enterprise Business Issues

IT has become more integrated into business operations. This integration was a driving force for the distribution of information technology and the popularity of the Internet and client/server. This integration also signaled a change in the job description of IT managers, who now had to be as well-versed in business management as they have been in technology management.

The handbook provides several sections of business management information written by IT professionals. Chapters include basic overviews of planning and business strategy to help you effectively manage personnel and departments. An emphasis has been placed on management techniques that achieve quality service from data centers and help desks as well as personnel. These quality techniques are well-suited to current business conditions, where strong competition and tight labor markets require organizations to do more with less staffing.

PERSONAL RESOURCE

A particular strength of the handbook has been the attention it pays to career issues. Not only is it a guide for ensuring smooth operations and ways to plan for more-effective operations support, the handbook is also a personal resource for managers. Changing technology and changing business conditions require changing skill sets from an IT manager. By taking charge of their careers and managing them, IT managers can adapt more readily to change. This proactive approach also empowers IT managers, who can no longer trust a single organization to provide a career track.

BEYOND ENTERPRISE COMPUTING

As long as organizations continue to rely on IT to run business operations, they will need managers to ensure the operation of computer systems. Although more and more IT operations are being automated, information technology is being used more often and by more types of business people. No matter how much they use and depend on IT, business people will always be more concerned with getting business done than with IT operations. They will always need support and someone to manage IT operations, despite whatever advances in technology may come. Who will they turn to? The one to whom they've always turned — you. And this handbook, no matter what title it will have, will be there to make sure you have the answer they (and you) are looking for.

John Wyzalek

Weehawken, NJ

April 1998

Table of Contents

HOME ACCOUNT INFO SUBSCRIBE LOGIN SEARCH MY ITKNOWLEDGE FAQ SITEMAP CONTACT US

SEARCH ITKNOWLEDGE

Go!

Keyword

Brief Full

- Advanced Search
- Search Tips

BROWSE BY TOPIC

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Bookmark It

Search this book:

Go!

Table of Contents

About the Editor

John Wyzalek has been a writer and editor specializing in IT topics for more than 10 years. In that time he has managed magazines and journals, edited handbooks, and was involved in the publication of more than 100 books for IT professionals. He holds a master's degree in computer science from the New Jersey Institute of Technology, Newark NJ, and a bachelor's degree in German literature from McGill University, Montrèal, Quèbec.

Table of Contents

Products | **Contact Us** | **About Us** | **Privacy** | **Ad Info** | **Home**

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Section I
Strategic Planning

To keep focus on their jobs and their data centers, data center managers should periodically step back and take a look at what is happening on the job and in the data center. They also need to take a look forward, and this where this section will help data center managers. The look forward is not so much a good guess about where the data center is going but where it should be in the future. To realize today's look forward as tomorrow's reality, data center managers must rely on strategy. However, strategy can often be as immaterial as any forecast, and in the day-to-day buzz of getting the job done, data center managers may forget about strategic plans as well as all else that is not right at hand.

Chapter I–1, "Managing Enterprise Systems without a Strategy: A Case Study, " shows the consequences of not having strategy to coordinate operations and the growth of technology. As the chapter relates, a lot of time and resources were spent on inadequate systems. If a strategy had been in place, the organization profiled in the chapter would not have wasted so much and would have implemented systems that satisfied the organization's needs. However, in the hectic atmosphere created to keep systems running in that organization, strategy was left behind.

To break the vicious circle that can spin out of control when no strategy is in place, Chapter I–2, "Developing a Strategic Information Plan," shows data center managers how to design a framework and develop a strategic plan. The chapter begins with taking a look at current conditions and how to take stock of what is missing and needed. From this baseline data, a strategic plan can be drawn up.

Strategic planning can be used for achieving high-levels goals such as improving quality or gaining a competitive edge, and it can be used for more-concrete goals such as how to update systems with new technology. This section covers these two sides of using strategic planning.

Chapter I–3, “New IS Strategies for Competitive Leadership,” presents an overall organization of IS designed to provide competitive advantage and shows how systems operations fits into this organization. It explains the new technologies that IS and the data center must support. Most important, it shows how strategy affects systems operations, IS, and the entire organization.

For the data center to be a leader in the IS, it must provide quality service to the entire organization. Delivering quality service to the user must be a primary concern of any data center manager. Chapter I–4, “Delivering IS Service Quality,” discusses how to implement a successful quality service strategy.

The data center has had many of its functions outsourced to third-party companies. How does a data center manager guarantee quality service in areas that are run by an outside organization? As Chapter I–5, “Achieving Quality Outsourcing” explains, quality outsourcing is looking for a business partner who shares the same values and principles and can add capabilities. The chapter provides a framework for data center managers to find just such a partner.

Technology is changing every six months. Planning to implement new technology on a long-term basis is very difficult. Chapter I–6, “Acquiring and Deploying Advanced Information Technology,” provides information to help managers bring some order to the changes in technology and how to deploy it throughout the enterprise.

The previous chapters have discussed developing strategic plans. The section ends with how a unique method for implementing strategies. Chapter I–7, “Using JAD for Strategic Initiatives,” borrows the joint application development (JAD) technique used in systems development for use in implementing strategies. Through a series of facilitated sessions, JAD provides an efficient process for collecting information and generating novel solutions to problems. The result is generally higher quality and smoother project implementation.

Chapter I-1
Managing Enterprise Systems Without a Strategy: A Case Study

Michael J. Masterson

Enterprise systems began and developed in the Clover Corporation (not the real name of the company examined in this case study) in much the same way they have in many organizations. A few stand-alone mainframe-based systems were initiated in the 1970s. By the 1980s, the number of incompatible stand-alone systems in Clover had mushroomed to more than 100 systems that were linked by a network, but systems interoperability was not approached until there was a centralized LAN strategy implemented. Today, the LAN and telecommunications systems have become the backbone of the organization, but unless newer IT strategies are developed, these systems will no longer be able to keep up with the changing needs of Clover and its employees.

AN OVERVIEW OF THE ORGANIZATION AND ITS SYSTEMS

More than 2,300 personnel work in the numerous structures housing this organization. Information technology (IT) provides the backbone for this task-diverse, operationally dispersed organization, while the local area network (LAN) has evolved into its electronic central nervous system. Clover's communications and information technology policy, despite its importance to the company, was left to staff members best described as junior technical specialists. Senior management abdicated providing the IT vision for the Clover technological infrastructure.

MIXED AND MISMATCHED DEVELOPMENT

In about 1971, two of the first automated data program offices created and managed at Clover were for aircraft fuel systems and a system for tracking/tasking parts for aircraft maintenance. These offices functioned as stand-alone entities with host-mainframe-based data and hardware to support the field-site users of these automated systems. Concurrent with the initiation of these stand-alone applications, a Field Assistance Branch (FAB) was created to support customer requests for assistance from the users of the automated systems in the field. The FAB remains dedicated to supporting end-users, from basic trouble-shooting to program code correction. A host-mainframe system linked to the LAN and a wide-area network serves as the repository of all problem reports and effective solutions arrived at by the FAB.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

More and more automated systems were placed under development at Clover, frequently as immediate solutions to immediate problems. A tremendous mismatch of systems evolved under this ad hoc, by-default strategy for information technology implementation by Clover management. Interoperability was never a consideration in systems development, consequently the accessibility and information sharing capability (reach and range) of these automated systems — and their supporting offices — was no broader than each unique system. The result was a large number of “stovepipe” systems and offices — two automated systems created by people working literally 15 feet apart might be completely incompatible.

BOSS, I HAVE A BETTER IDEA

In the early 1980s, a trio of energetic junior electrical engineers recognized the problems caused by these stovepipe systems and proposed a solution — a dual broadband network (i.e., a simplex broadband circuit, every node having separate input/output broadband feeds), for data communications. This broadband backbone was approved and built incorporating twisted-wire pair circuits, running at 9600 bps, in use by each separate automated system office. The data network structure allowed network movement of host mainframe information within Clover, as well as allowing users in the field to network into the appropriate host mainframes. The enthusiastic response to this data network led to a medium upgrade for more bandwidth. The data communications broadband was upgraded to coaxial cable, running at 10 Mbps, which allowed faster response and more user access.

Initial Failures with Implementing a LAN

After completion of the data communications network, the junior technicians understood that with a software upgrade to the existing applications, an information network could be piggybacked onto the data communications network backbone. The perceived value (by the technicians) of coordinating interoffice communications drove the creation of the information network. The LAN operated at 10 Mbps, the same speed as the data communications network. However, the diverse nature of stand-alone office-specific software regularly overloaded the network software. In addition, conflicts between network management software and the plethora of office computer hardware led to interoperability blow-ups, resulting in a complete network shutdown. Clover's director of computer operations described the technology platform as "one big kludge."

ENTERPRISE-WIDE POLICY CREATES A SUCCESS

The next step in the network-creation process was the designation of the LAN shop as the responsible office for network operations. Unfortunately, Clover senior management fought the Medusa of information technology within Clover for 23 years before authorizing this step, which brought with it the authority to correct the kludged-together systems at the company. Senior management's tasking did include support for the upgrade to a Fiber Distributed Data Interface (FDDI) LAN. Clover's FDDI LAN operates with carrier sense multiple access with collision detection (CSMA/CD) protocol, bus configuration, and runs at 100 Mbps as the high-speed LAN for information handling. Also, the LAN office was able to collect and standardize the numerous hardware and software systems still operating within the company. Formal office automation standards and programs were implemented to prevent a return of kludged systems. However, interoperability problems were discovered among the new versions of office software, the application programs, and the still-diverse Clover computer hardware.

Centralized LAN office authority for control of all Clover office automation, both software and hardware, has eliminated many of these problems. This control led to the installation of interoperable, 386-processor-based hardware and Microsoft Windows 3.1 software. The LAN office now had a common body of knowledge to focus on when troubleshooting problems.

It must be noted that this brief history of the Clover LAN comes from outside the organization, with a clear view of a mature, robust network. As previously noted, a 23-year timeframe shaped the evolution of the Clover LAN. This long trip clearly demonstrates what the lack of IT Vision from senior leadership created and sustained — segmented operations, sets of stand-alone systems, and internal barriers to adaptation (Keen) for the Clover information technology platform. However, the positive aspect is the use of technology and software to overcome hardware limitations.

CURRENT LAN CONFIGURATION

An overview of the fiber optic LAN configuration is covered first, followed by the data communications network. This is a logical order since the fiber optic LAN is the most common link into the host-mainframe-based data communications network.

Desktop. Every person employed in an office at Clover has at least a 386 IBM-compatible personal computer on his/her desk. Most offices are equipped with 486sx33 machines loaded with Microsoft Windows. This platform enables approximately 2,100 people to have individual links into the FDDI LAN. The information network uses Novell software operating with Ethernet 2 LAN implementation. As previously mentioned, CSMA/CD broadcast protocol is used with a bus configuration. Also, a transmission control procedure/Internet protocol (TCP/IP) for file transfers into the Clover Data Communications Network (DCN), a Corporate Data Network, or Internet is used. All network operations behind the desktop Window functions are transparent to users.

Monitors. Network monitors indicate the Clover FDDI LAN handles at least 126,000 e-mail data messages per month. No estimate is available for the number of file transfers or host-mainframe data packets transfers per month. In addition, some of the layers in the FDDI LAN meet open-system-interconnection (OSI) model standards.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Local Gateways. From every network user node, access into the LAN is via RG-58A, 10base2 Thinnet cable operating at 10 Mbps, connecting to the office local gateway — the multimedia access center (MMAC)(interview/briefing). MMACs isolate all traffic destined for the local office subnet, switching local subnet traffic to the correct node. Any data packets addressed outside the local subnet are managed and controlled by the MMAC. The heart of the MMAC is the intelligent repeater module (IRM). The IRM performs all repeater functions for each of the MMAC's 30 ports, regardless of media type. In addition, the IRM provides the network management and control functions for the MMAC, including error detection, retiming data packets, and regenerating data preamble. A media interface module (Mim) provides the ability to use a variety of media outbound from the MMAC, to include twisted wire pair, fiber optic cable, or thick/thin Ethernet coaxial cable. Of note, each MMAC can support an office LAN of up to 30 connections. However, it is LAN office policy to have only 20 individual IP addresses per MMAC. This prevents segment overload and allows flexibility for diagnostic troubleshooting.

In the Clover LAN, the MMAC Mim is set for RG-6 thicknet coaxial cable operating at 10 Mbps and linking to the office LAN file server, a 486dx66 machine running Novell Network software. The 21 file servers in the FDDI LAN provide the shared data/programs and disk drives among the office LAN users.

Routers. In addition, the thicknet cable connects to the LAN Routers. These routers operate at OSI model layer 3 and act as the traffic cops for LAN traffic (interview). Routers look at the layer 3 protocol and make decisions based on the type of media/data packet. This requires that every router in the Clover LAN be able to interpret any layer 3 protocol used in the system. The more protocols in use, the more processing time on each packet and the more of a bottleneck the router will become. The routers are multilegged, usually connecting up to four separate office LANs from one building. Routers are also called fiber hubs for each Clover LAN operating location. As the hub for the LAN, the router is where the switch from the information network into the broadband data communications host-mainframes computer occurs.

- **Bridges.** The routers connect via fiber optic cable operating at 100 Mbps to the fiber bridges — the backbone of the FDDI LAN. Each Clover operating location has at least one FX 8210-N-E / fiber bridge node. These Remote Bridges transfer data between two Ethernet/IEEE 802.3 LANs over the FDDI backbone (interview). In addition, the FX 8210-N-E2 operates as a Local Bridge that transfers data between any two Ethernet/IEEE 802.3 LANs connected to it by the router. Therefore, this fiber bridge provides data paths controlled by the fiber bridge CPU: *remote bridging*, for LAN to FDDI and FDDI to LAN connections, and *local bridging*, for LAN to LAN connections.

For data handling, all LAN data frames are received by the Local Area Network Controller in the fiber bridge and stored in the Bridge Buffer Memory (BBM). The destination addresses of the frames are checked against the addresses stored in the address table. Frames addressed to remote stations (i.e., to be transferred to FDDI) are forwarded by optical transmitters to the FDDI ring.

Gateways. One other critical level of the LAN is the Gateways into other network environments. Any LAN data packets addressed for an automated system host-mainframe, the Corporate Information Network, or Internet locations are routed to the appropriate gateway. These gateways operate at OSI layer 6 and are concerned with translating the media/data packets to allow transfer between two very different computer environments. The first is internal to Clover, e.g., linking the desktop environment to a data communications network host-mainframe. The second is external, translating out to (or inbound from) the Corporate Information Network to work with users at field locations, or into the Internet. All data transferred from the FDDI LAN to the Gateways moves at the broadband speed of 10 Mbps. There are no fiber optic gateways off the Clover LAN.

Backups. As the history of the networks shows, the Data Communications Network (DCN) was the first network created at Clover. In addition, the 10 Mbps broadband DCN functions as the backup to the FDDI LAN. If a fiber optics connection is broken, the entire Clover LAN reverts to the broadband DCN, and everything slows to 10 Mbps. Normally, the LAN office's first indication of a problem on the FDDI network originates with multiple complaints about the “slowdown” of the network.

Mainframe Access. DCN host-mainframes are the repositories of all the databases for the 147-plus automated data systems developed, deployed, and managed by Clover. Several requirements must be met to allow a user in a desktop environment to gain access to one of the DCN host-mainframes. The user must require access to the database, be an authorized operator on the Clover FDDI LAN, have the correct host-mainframe emulation package installed on his/her desktop computer, and must be fluent in COBOL. All these factors help control (or prohibit) access to the host-mainframe. This is a valid concern for Clover program offices, since once access to the mainframe is gained, permanent, undetected changes to the database can be accomplished. The LAN routers are the key to user access from the desktop, with the router sending the data packet into the 10 Mbps broadband loop to the appropriate gateway for the host-mainframe. The emulation software in the desktop environment masks all the network actions from the user. Once connected, the user operates the desktop hardware like a host-mainframe-connected dumb terminal.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

The Other Communication System

One other platform at Clover is the Plain Old Telephone Service (POTS). Of course, there is a telephone on every employee’s desk. However, Clover employees rely on the LAN almost to the exclusion of POTS. Common practice is to send an e-mail and go on to other business, and voice contact is eliminated. All too often, the result is complaints to the LAN office because e-mail didn’t go through. Occasionally, the LAN office finds a system problem. Usually, the problem is solved with the advice: “RY*!?%M” (Read Your *!?% Mail).

Future of IT at Clover — Time for New Strategies. LAN architecture at Clover is a mature, robust system. The company’s business would be critically damaged without the technology platform currently in use. However, many concerns and decision points create uncertainty about the future information technology platform at Clover. Problems of hardware and software already identified by computer operations personnel include:

- Hard-drive memory of the 21 file servers is already at maximum capacity. Six-gigabit hard-drive memory units are to be installed to correct this problem.
- Prior to designation of an authorized LAN office, new versions of office automation software were impossible to implement across Clover due to inability to secure site licenses. File server-based office automation software was purchased (21 copies), and site licenses of 100 users/copy obtained. The result is easier control, upgrade, and troubleshooting of office software.
- All office automation software versions are forward compatible, but

not backward compatible, e.g., a file built on Microsoft Word 3.1 and sent to a machine using Word for Office 4.0 would work, but could not be sent back to the 3.1 version without special action. In some cases, there is complete incompatibility from old to new software.

- Acquisition and installation of a new, Pentium-based computer system will start in the spring of 1995. The greater speed and power of these machines will increase the strain on the LAN's slowest link — thin/thicknet medium from the desktop to fiber optic cable. Cost estimates to upgrade to fiber optics to each desktop are estimated to be millions of dollars. No funding is available.
- Full open systems interconnection model standards will be implemented throughout Clover. This will clear up software-compatibility difficulties mentioned previously. No completion date is set.

A serious problem is the soon-to-be reached hardware limit of available nodes on the FDDI LAN. As presently configured, the LAN can support 2,250 individual users; 2,100 users are already on the net. However, corporate organization downsizing and restructuring means Clover will be increasing by at least 300 users in the near future. No solution is available to increase LAN capacity.

Another corporate reorganization action with serious future impact on the Clover LAN is the designation of Clover as one of four IT centers under corporate headquarters. Clover's information technology platform is not compatible or interoperable with either the other three centers or the corporate headquarters. Restructuring was finalized before the question of information technology reach and range was addressed.

Myopic Funding Strategy

A driving force contributing directly to virtually all of these current and future problems for the Clover LAN is the corporate funding and budget process. As the evolution of the network shows, funding for new systems is readily available. However, none of the 147+ automated data systems Clover has brought online included system maintenance, training, and upgrade funding. According to one IT staff member at Clover, the senior leadership IT vision is summed up as: "Here's the new system, make it work /last /succeed." The IT funding process severely limits the IT vision of the future for Clover.

CONCLUSION

This case study of the Clover Local Area Network shows clearly the importance of the system for current successful operations. However, senior leadership has consistently undermanaged the information technology platform at Clover. Every step of growth of the company's information systems improved business operations. Despite this importance, the Clover information technology platform will have reached its limits of accessibility and interoperability by the year 2000. Without total senior management commitment to a clear Information Technology vision for the future, Clover as a corporate entity will be unable to pursue future business options.

Recommended Reading

Carr, Houston H. *Managing End User Computing, 2nd ed.,* Auburn University Press, 1994.

Carr, Houston H., Snyder, Charles A. *The Management of Business Telecommunications*, Auburn University Press, April, 1994.

Keen, Peter G.W. *Shaping the Future: Business Design Through Information Technology*, Harvard Business School Press, 1991.

Stamper, David A. *Business Data Communications, 3rd ed.,* The Benjamin/Cummings Publishing Company, Inc., 1991.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter I-2
Developing a Strategic Information Plan

Raymond J. Posch

Business decisions may involve difficult choices affecting large-scale investments, strategic directions, and critical organizational goals (i.e., those in which the CEO or chief operations officer would be involved). Business decisions can also involve somewhat simpler choices regarding issues related to production, administration, or any other department (i.e., those decisions handled daily and hourly by line workers, clerks, supervisors, specialists, and managers).

The variety of these decisions demonstrates that businesses depend on information at all levels for both daily operations and longer-term strategies. When high-quality information is made easily accessible to the various decision makers at all levels throughout an organization, the decisions are more effective, and if these decisions are implemented properly in subsequent actions, the organization is better positioned to achieve its goals. The content and form of the information as well as the way it is communicated and managed are important factors to success. These are the basic, underlying premises of information strategy—that information is crucial to business success, that its importance must be recognized, and that the information resources must be managed as strategic assets.

Information does not exist entirely on its own. Most organizations have information systems comprising computer hardware, software, and networks to manage information and perform the intended processing as well as the facilities and people necessary to develop and support those systems.

Information resources can be grouped into three types:

- *Information services.* The computer applications for the processing, storage, retrieval, transmission, and presentation of information.
- *Information infrastructure.* The hardware, software, systems, staff, facilities, and other elements that support the delivery of information services.
- *The information itself.* The data that allows the organization to conduct business operations.

The more robust (in terms of functional capabilities) and flexible the infrastructure, the more easily applications can be developed and integrated. The more capable the infrastructure and the more rich the base of information, the more powerful the services that can be delivered.

Information resources are complex and require a great deal of attention, money, and work. The IS infrastructure in particular needs careful long-range planning, investment, development, and management. In addition, because business functions are increasingly becoming automated, information resources are becoming more vital to conducting business. They must be viewed globally and strategically not just operationally. An organization therefore must have a cohesive long-term strategy for information resource management at both the business and technical levels to effectively support its goals.

To develop such an information strategy, the following actions are recommended:

- Developing a strategic business plan.
- Conducting an information resource audit.
- Developing a strategic information systems plan.

The objective of these actions is to produce a written strategic information plan that can be clearly understood by senior management. The strategic information plan should consist of the following components:

- An information resource audit statement.
- An information strategy statement.
- An information systems architecture statement.
- A strategic IS organizational statement.
- An information strategy implementation plan.

This chapter discusses in detail the steps to be taken in each of the three phases of the plan and includes guidelines for writing the plan.

DEVELOPING A STRATEGIC BUSINESS PLAN

The purpose of strategic business planning is to specify a few important business goals and a strategy for attaining them. The purpose of strategic information planning is to supplement the business plan by defining a strategy that supports the important business goals through appropriate information management activities. Strategic business planning must precede strategic information planning.

Information management plans that are not linked to strategic business goals, no matter how well thought out, lead to problems that are often prevalent in operations and systems departments, Such systems plans typically specify information technology (IT) strategies and IS plans that support the tactical requirements of the various user departments but do not address the critical business goals. The resulting misalignment between IS goals and business goals results in conflicting IS priorities among internal groups and the misallocation of IS resources.

If an organization does not have a strategic plan with well-defined business goals, the business strategy must handle whatever needs arise at any given time. This results in ever-changing priorities and numerous disputes as to the exact nature of the goals. In such environments, any systems plan is imperiled from the start.

Strategic planning is often confused with long-range planning. A strategy is a plan for achieving an objective by focusing on critical goals. Therefore, the objective of strategic planning is to identify the most important goals and determine the most important actions for attaining these goals. An organization can develop a long-range plan that is not strategic or can develop a strategic plan that need not be long range.

In reality, strategic planning may need to address multiple time frames—short range (i.e., one year or less), mid range (i.e., two to three years), and long range (i.e., four years or longer)—with different goals for each. Strategic plans focus primarily on goals. In contrast, tactical plans focus on actions that satisfy immediate requirements, including those that implement strategy, and operational plans explain in detail steps needed to make tactics fully operational.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

The Partnership Between Corporate and IS Management

The ideal situation is one in which information strategy development is viewed as an extension of the strategic business planning process. This requires that business managers recognize that information is critical to their success and perceive information management as a strategic business issue. Senior IS managers must often take the lead in educating senior corporate managers, especially in helping them view information strategy as an extension of business strategy. The strategic planning process provides an excellent opportunity for IS and corporate managers to build a partnership, in which they work closely together to plan and implement a business strategy that acknowledges information's vital role in the functioning of the organization.

The Strategic Business Planning Process

Strategic business planning, when its focus is broadened to include information strategy development, typically involves the following six steps:

1. Identifying the planning time frames.

2. Identifying the main goals for the time frames.

3. Identifying the critical information required to support the goals.

4. Estimating the time and resource, including information resources, required to achieve the goals.

5. Evaluating the high-level trade-offs of expected benefits versus costs. If appropriate, goals should be redefined and steps 3, 4, and 5 repeated until satisfied.

6. Writing the strategic business plan to explain the main goals and the strategies for accomplishing those goals in clear and concise language. The plan must then be communicated, at last in summary form, to the staff.

Strategic business planning begins by identifying the planning time frame (usually one to three years) and the main goals that the organization wants to achieve in that time frame. Goals are commonly stated in relation to such areas as products, markets, sales, positioning relative to competition, and profitability. Goals, however, should also be included in terms of strengths and capabilities (to be obtained from the organization's infrastructure). The main goals should be strategic in nature and critical to the success of the organization in the planned time frame.

One method of determining the main goals is to identify critical success factors and then translate them into specific goals. The number of goals should be limited—three to six goals for each business unit and, overall, three to six most-critical goals for the organization. In general, the fewer the goals, the more likely they will be met. Once the goals are determined, they should be ranked according to priority so that it is clear which are most and least important. An important step in incorporating information strategy into the business planning process is identification of the critical information needs. This is discussed in the following section.

Identifying Critical Information Needs

For each main business goal, the business managers must identify the critical high-level information needed to support the goal. For example, the first-cut, high-level, critical information requirements for a marketing division in a highly competitive industry might be as follows:

- Detailed reporting on market penetration for the organization and prominent competitors worldwide.
- Graphic presentations of market shares and market trends.
- Quantitative and qualitative analyses of customer or buyer perceptions.

For the information needs to be well understood, however, these first-cut requirements must be expanded into more detailed descriptions. The first-cut requirements in the example have been expanded into more detail as follows:

- Monthly market penetration reports must be developed that provide estimated unit sales, percentage of total sales, and percentage of total estimated market, categorized by North America first, then Europe, Asia, and the rest of the world for the organization and its top 10 competitors.
- Charts must be developed showing the same information graphically with the addition of time-line trends.
- Online sales and performance figures must be upgraded at least weekly but preferably daily for the most recent day, week, month, year-to-date, or other specified time period. These figures must show unit sales, revenues, sales expenses, operating expenses, and net profit organized according to product line, sales region, and sales team, and a

comparison must be made to the corresponding figures of the previous year.

- Numeric and narrative analyses of surveys must be conducted at least annually, showing trends, satisfaction levels, and other important data regarding customers and, for comparative purposes, other buyers in the market who are not customers. The information should be organized according to product categories, regions, customer industries, and other pertinent factors.

This level of detail will be required for evaluating whether existing information resources can satisfy the marketing division's needs or, if they cannot, what changes will be necessary to help the existing resources meet those needs. In cases in which a significant degree of uncertainty exists about how well existing systems meet the strategic needs, an information resource audit (i.e., the next phase of the strategic information plan, which is described in detail in a subsequent section of this chapter) might be needed at this point to clarify the information management activities required.

Preparing the Strategic Plan

After the critical information needs have been identified, the necessary information resource requirements (e.g., computers, networks, and data bases), along with other resource requirements (e.g., people and dollars), must then be estimated for each goal. Operations and systems managers are responsible for estimating information resource requirements; other business managers must estimate the requirements for resources from their business units. Certain resources can be used to serve multiple corporate goals; therefore, the planning process must reflect the plan as a whole, not just the individual goals independently. Resources for building and maintaining the infrastructure are especially important and should not be diverted to short-term needs. External as well as internal resources can be used (e.g., at least part of the needs of the marketing division described in the previous section could be met by using external resources). Specific high-level tactics, activities, and projects (e.g., development of new systems or enhancement of existing systems) must be roughly defined at this stage to gain some appreciation of what is needed to achieve the goals.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Bookmark It

Search this book:

Go!

Previous | Table of Contents | Next

On the basis of those high-level tactics, time and resource requirements for the business goals must be estimated to evaluate the costs. Whenever possible, quantifiable benefits and returns on investment should also be estimated. Estimates do not need to be exact but should be reliable within an order of magnitude. Significant impacts and risks (which may or may not be quantified) should also be identified. In addition, because the plan may address multiple time periods, it must have a time line as well. Certain projects (e.g., those that develop the information infrastructure) may take multiple years to implement. Ultimately, it must be decided whether the expected benefits of the goals are worth the predicted costs and risks or whether a different set of goals should be specified.

Furthermore, strategic planning, like all types of planning, is an iterative process consisting of rounds of specifying and estimating, followed by evaluating the risks and trade-offs (i.e., costs versus benefits). When the iterations have finished it is helpful to rank the high-level tactics and projects for each business unit and for the organization as a whole within each planning time frame.

The final task of strategic business planning is to write a strategic business plan as a document (at least in summary form) for everyone in the organization, though the information must be treated as proprietary and confidential. It should explain, as clearly and concisely as possible, the most important business goals and the assumptions for both the overall organizational strategy and the high-level tactics of each business unit, including the IS division. Developed this way, the information strategy will be fully integrated into the business plan.

CONDUCTING AN INFORMATION RESOURCE AUDIT

An information resource audit is an assessment of the assets and liabilities of existing information resources. The audit should cover the information systems (i.e., applications), hardware, software, networks, tools, IS department and staff, data bases, and metadata (i.e., data definitions) of the organization. It is useful to differentiate among the three main types of resources: infrastructure, applications, and information. The audit can be applied at multiple levels and for multiple purposes; for the strategic information plan, however, it is most useful at a high level. In other words, the assessment should avoid too much detail and must clearly define the greatest strengths and weaknesses of the information resources.

Performing the Audit

The audit is best performed after the strategic business plan has been completed or at least after the critical information needs have been identified. The audit should focus on how well those critical information needs are presently being satisfied. Other critical high-level strengths and weaknesses should also be assessed. Any systems under development that are pertinent to major assets or liabilities must be included. The audit can be performed by IS managers (from a service provider's perspective) with a review conducted by corporate management, or the audit can be performed by non-IS business managers (from a client's perspective) with a review conducted by IS management.

Writing and Reviewing the Assessment

The product of the audit is a clearly written statement disclosing the findings of the audit. On the liability side, the audit statement should describe what is needed to correct deficiencies (e.g., subject data bases or a centralized data dictionary or repository) and why, but it should not explain in fine detail how the changes are to be executed. Similarly, the assets should be briefly described in terms of what is needed and why.

The audit statement should relate directly to the strategic business plan and should be written for senior management. A review process of one or more iterations will be needed to ensure that the assessment is strategic in nature and indeed correctly identifies the significant assets and liabilities among corporate information resources.

The Enterprise Information Atlas

The enterprise information atlas consists of information maps and is a useful but optional product of the audit. Exhibit 1 presents one type of information map, which shows the sources and uses for information in one particular information services company.

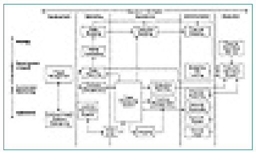


Exhibit 1 An information map for an information service company.

The purpose of the information map is to help managers as well as others in the organization comprehend the organization's information resources and how they relate to the goals and functioning of the organization. Information maps are not only educational tools but reference sources and, perhaps from the IS perspective, marketing tools as well. Information maps depict the relationships of information resources at various functional levels in the organization and can be used to help answer such questions as, What systems (i.e., applications) or data bases are sources of information or services to the various business units or corporate functions? The maps should be as graphic as possible, showing, for example, how information relates organizationally, geographically, and operationally as well as how it relates to the planning and management process.

If automated, information maps can be stored in a data base or repository with online access and can depict the organization's information resources at several levels: high level (i.e., strategic), mid level (i.e., tactical), and detail level (i.e., operational). To support the strategic plan, different maps may be required for different time frames (e.g., in the current state, with a projection at the end of a short-range plan, or with the projection at the end of a mid-range plan). The maps would not contain the details of a full data dictionary or repository but might well include such organizational models as data models or entity-relationship models. If the information maps are not automated, they can be included in a document serving as a supplement to the audit statement (though they can also stand alone). The maps should be updated with each subsequent cycle of strategic planning and information resource auditing, if not more frequently.

Information maps can be most easily implemented at a high level and in document form and as such would probably derive the most benefits, especially in educating senior managers about the organization's information resources and the importance of information. If they are automated, however, information maps would provide a useful resource for the organization as a whole and in particular for the data administration or data resource management group that would be responsible for their maintenance. CASE tools that are encyclopedia or repository based would probably offer the best technology for automating the information maps.

| Previous | Table of Contents | Next |

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

DEVELOPING A STRATEGIC INFORMATION SYSTEMS PLAN

The first phase—strategic business planning—identified key business goals, critical information needs, and the information management strategy for meeting the business goals. The second phase—the information resource audit—identified the significant assets and liabilities among the corporate information resources, especially as they relate to the information requirements of the strategic plan. The purpose of the third phase—strategic information systems planning—is to complete the strategic information planning process and to develop the written strategic information plan.

Five steps are involved in strategic information systems planning. They are:

1. Developing or refining the IS architecture.
2. Developing or refining the strategic IS organizational plan.
3. Completing the business-level IS plan.
4. Completing the information strategy statement.
5. Preparing the information strategy implementation plan.

Developing or Refining the IS Architecture

With the great number of new IS technologies and options and with the ever-increasing need for consistency and integration, it is important that the organization have an architecture that defines the preferred technologies, methods, and directions that the organization will use and follow for its business information systems. The purpose of an IS architecture is to clearly define how corporate systems are to be constructed and how they will appear or act. The purpose is to narrow the approach, set expectations for both users and IS staff, and specify standards and guidelines that support efficient, effective information systems.

Development of the IS architecture should begin with an architectural reference model. Most organizations are developing systems on the basis of a distributed model. Because of the critical issues related to systems integration, however, planners must not only evaluate information resources by their location but also clearly differentiate between infrastructure, applications (i.e., business and user service), and information. Therefore, a two-dimensional information resource model such as that illustrated in Exhibit 2 is a more effective starting point. The infrastructure is the foundation and the public utility; it directs the information flow and provides facilities for information management. Applications deliver services to users; they are built on the basis of the infrastructure and are integrated and share information through it. Information is the asset that is stored, processed, moved, and managed by means of the applications and infrastructure.

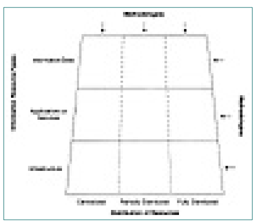


Exhibit 2 Information resource architectural reference model.

Each of these three main information resource types can be distributed centrally, regionally, or departmentally (i.e., partially distributed), or they can be distributed to the user (fully distributed). Those information resources that provide the means of interconnection and integration of other distributed information resources (e.g., wide area networks, local area networks, phone systems, and integrative software layers and services) are some of the most critical elements of the information infrastructure.

IS architectures are quite complex. The reference model (such as the one in Exhibit 2) decomposes the architecture into smaller architectural pieces, thereby making it easier to understand. On the basis of this model, the process of developing the architecture is probably best organized in the following sequence of architectural units:

- Information or data.
- Information applications, or high-level services.
- The information infrastructure, consisting of networking and communications, computing hardware, and software.
- Methodologies.

For each of these architectural units, the architecture should address the distribution of resources (i.e., whether resources are distributed centrally, departmentally, or to the user). Methodologies are also an important part of the architecture and can be thought of as applying to both dimensions (i.e., resource types as well as the distribution of the resources). Each architectural unit is described in further detail in the following sections.

Information or Data. The information should specify the following details:

- How information will support the business goals.
- What the objectives are for information access and information services for the users.
- What technologies will be used for information storage and presentation.
- What kinds of information are important.
- How the information will usually be managed.

This architectural unit must address such issues as data base technology and applications (including issues related to distinctions between nonrelational, relational, and object-oriented data bases and between distributed and centralized data bases, as well as determining which DBMSs are to be supported), data resource management (including issues of what repository should be used, distributed versus centralized ownership, and the balancing of priorities), and data administration (including issues of data dictionary maintenance, access control, and standards enforcement). In the future, it will become increasingly important that this architectural unit address information regarding all media types (e.g., data, voice, video, audio, graphics, or document images).

Information Applications. This architectural unit must define what the general approach will be to applications and user services. It should address such issues as which applications should run on which hardware and software platforms, development of applications versus purchase of packaged applications, levels of integration between applications, use of underlying services (i.e., hardware, software, and communications), and user interface standards.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Bookmark It

Search this book:

Go!

Previous | Table of Contents | Next

The Information Infrastructure. The infrastructure architectural unit can be broken down into three primary areas: networking, hardware, and software. The networking portion should identify the strategies or target technologies for interconnecting computing equipment, both locally and across wide areas, and how, more generally, users will connect with departmental, regional, and central services. The networking architectural unit must define the voice (i.e., telephone), video conferencing, and data networking strategies as well as the degrees of integration of these networks. For long-range time frames, goals (e.g., gigabit bandwidths supporting multiple protocols) can be identified without specifying which technologies will be used. For short-range time frames, specific technologies (e.g., token-ring LANs bridged across an SNA network with T1 trunks) might be selected as standards. The hardware architecture should specify which computer systems and user workstations will be used to perform the actual computing and information processing work. The software architecture should identify the preferred operating systems, data base management systems, programming languages, and software strategies to be used for consistency and integration (e.g., low-level software services).

Methodologies. Methodologies should be established governing stages of the information resource life cycles (e.g., planning, analysis, design, development, testing, deployment, maintenance, origination, update, archival, and management) and across both dimensions of the architectural model (see Exhibit 2). It may be appropriate to have life cycle methodologies that are distinctly different among the main information resource types (i.e., information, applications, and infrastructure). The methodologies will also likely have differences among central, departmental, and user-level resources. Such techniques as systems engineering, CASE, joint and rapid application design, information engineering, and applications development environments (e.g., AD/Cycle) may provide useful methodologies and tools, though they may not address full information resource management as defined by the architectural model.

In addition, because IS architecture is primarily future oriented, the architecture should address different architectural goals at different points in time as well as the issues concerning the current base of systems (i.e., the applications and infrastructure) and how they will conform over time to a new architecture. One of the objectives of the architecture is to provide reasonable consistency over time while adapting to new requirements and new technologies.

The Architecture Statement. IS management should write the information systems architecture statement describing the overall architecture clearly and succinctly for senior management. The architecture statement should begin with an executive summary followed by supporting levels of architectural strategy. Depending on the magnitude (i.e., the size and diversity) of the organization and the scope (i.e., number and size) of its systems, the statement may need to be supplemented by multiple documents (perhaps elaborating on each of the architectural units described in the preceding paragraphs). The statement must be reviewed annually as part of the strategic planning cycle and, if appropriate, modified to reflect new business conditions, strategies, and technologies.

Developing or Refining the Strategic IS Organizational Plan

Because the IS department is such a critical element in how the organization manages its information, the organization also must establish a plan for the IS department itself. Changing business structures, more distributed (i.e., lower-level) information technology deployment, and advances in computing and networking technologies are causing changes in the way operations and systems personnel are organized and managed. The IS department must effectively support the information strategy and the IS architecture. However, because of its importance, this issue should be addressed and documented in a separate strategic IS organizational statement.

IS management should develop the IS organizational strategy in close partnership with senior management. The current information processing environment is more highly distributed and therefore usually requires IS support that is likewise more distributed. For example, many businesses have hybrid-structure IS departments with decentralized IS development and support staff actually reporting to or being part of the business units and centralized IS staff for planning, maintenance, and enforcement of the architecture and for data administration. Whatever IS organizational structure is used, the strategy must be well defined in this statement, specifying a time line and developmental as well as service-level objectives. The IS organizational statement also must be reviewed and updated with the annual strategic planning cycle.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Completing the Business-Level IS Plan

Strategic IS planning at a business level (i.e., with an integrated view of the organization) should be completed during the strategic business planning phase, when the information strategy was originally developed. Before proceeding with implementation planning, however, IS management should review the information needs, high-level tactics, identified projects, expected benefits, and associated resource requirements for their continuing validity in relation to architectural and organizational decisions or other new developments that may have subsequently occurred.

Planning is a process of iteration and refinement. IS management must ensure that the process does not become too detailed and drawn out. During this step, changes to the information strategy statement (usually with regard to the tactics and projects or the estimated resource requirements) may result. If the time or cost estimates change significantly, the impact on the original business plan must be evaluated.

Completing the Information Strategy Statement

Most of the effort for the information strategy statement has already been completed in the first two phases. What remains at this point is to bring together and clarify the overall information strategy for the organization, explain the critical information needs and how they relate to the business goals, and describe the information management plans and resource assumptions in the strategic business plan. Architectural and organizational strategies should be reflected at a high level, with differentiation between the development of infrastructure and the development of applications (i.e., business services). Strategic directions, important goals, and the orders of magnitude of project time and cost are more important than specific breakdowns of the projects and their costs.

The information strategy statement should explain how the goals, strategies, and high-level tactics are arranged according to priorities across business units. It should be clear, concise, and written for senior management (i.e., the CEO, chief operations officer, and business unit managers). IS management should write the information strategy statement, and the document should be reviewed and approved by senior management.

Preparing the Information Strategy Implementation Plan

The final step in strategic information resource planning is to create a plan for implementing the information strategy. The implementation plan should be prepared by IS management and addressed at a high level (i.e., in less detail) to senior management, and it should consolidate the planned activities of the information strategy statement, the information systems architecture statement, and the strategic IS organizational statement into a single reference document. It should identify and explain critical assumptions, dependencies, resource issues, and events. The information strategy implementation plan, when completed, will become a primary reference for IS project planning, management, and control as well as status reviews.

SUMMARY

Information strategy requires that business executives recognize information resources as critical corporate assets (and future information requirements) that must be taken into account in the business planning process. Once this is clearly understood and accepted, information management becomes a business issue in which senior management must play an active role. Information strategy is an important part of information management and is best communicated through a strategic information plan.

The strategic information planning process, however, can become a hindrance or a barrier if it is allowed to get mired in too much analysis and debate. To avoid this problem, the strategic information plan must be developed at a high level, with active participation by senior management, and in a reasonably short time (three to six weeks beyond the strategic business planning process). It is important that the strategic information plan be seen as strategic, be developed quickly, and become the focal point for an ongoing business partnership between senior and IS management.

The plan does not have to address every issue but must establish the overall strategic direction. (A more thorough strategic information plan, especially in the area of IS architecture, which requires in-depth knowledge of information technologies and the foresight to judge the best architecture for a particular organization, will actually take years to evolve.) When it is initially prepared, the strategic information plan should be considered a basic plan that will undoubtedly change as the process becomes better understood and as the architecture matures with time. More effort can be channeled into refining the strategic information plan later, but having an effective strategic information plan that will implement the business vision should be high priority for every business executive, CIO, and IS manager.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter I-3
New IS Strategies for Competitive Leadership

Robert G. Simko

The greatest challenge facing business today is complexity. Markets and competitive structures have grown increasingly complicated and volatile. Freed of traditional hierarchies of control, corporate structures have become difficult to coordinate, and information technology, which could reduce complexity, has too often increased it by generating vast amounts of data that few can assimilate, let alone understand.

This problem is being solved by organizations throughout the world that have adopted the tactics, techniques, and tools to support an information strategy. A strategy allows a company to deal with complexity by identifying its underlying patterns, and by isolating and acting on critical performance variables. Business success is based on simple, rather than complex, parameters.

This principle should be applied, above all, to IS technology. The real issues are not whether mainframes or new types of large-scale servers should be employed to anchor infrastructures, the relative merits of parallel processing or symmetric multiprocessing, or whether relational or multidimensional databases are better. These are essentially tactical questions.

The critical decision is simpler — and it is the most important IS strategy decision a CIO will make this decade. It is whether to take the initiative.

INFORMATION INFRASTRUCTURE

Data is the starting point and the raw material. However, in most organizations, it is fragmented into incompatible databases and files and dispersed across a wide range of different systems and servers. Much of the data is inaccessible, and, even where accessible, its quality is possibly questionable. It may be collected and stored inconsistently, in ways that make it difficult to apply for other purposes.

The first challenge is to make data accessible. The second is to make it usable. To become genuine information, data must also be comprehensive, consistent, credible, timely, and available in practical and useful form. The third challenge is to meet the first two without endangering the integrity and security of core data resources on which organizations depend for their basic business operations. These challenges are met through the following initiatives:

A Database Infrastructure is Established

During this effort, critical categories of data are distilled from production systems. They are converted to common formats and reloaded into separate relational database servers to support new, query-intensive applications. This process focuses on data that is organizational in nature, because it supports business processes, operations and decisions to apply access to organizations or functions because multiple divisions, business units, and departments within the organization require access to it, or both. Databases may be located on single or multiple systems, and on various platforms.

A data warehouse is created. This high-level software framework refines incoming data for consistency and accuracy and reorganizes it by using logical business models that ensure its relevance to current and new users and applications. The database infrastructure ensures the basic accessibility of data; the data warehouse addresses its quality.

Additional functions, such as summarization (i.e., delivery of specific, condensed types of data to users based on predefined criteria), integration (that is, single logical views of data from multiple sources) and drill-down (or the ability to start with high levels of data generality and then move to progressively finer levels of detail), are also provided. However, this description refers to an organizational data warehouse. Data warehouses that support specific departments and specialized applications may differ from this one.

Data Management is Implemented. This includes the staff and tools that administer core databases and data warehouses and ensures the security of organizational data. By addressing these functions at the infrastructure level, an organization provides increased flexibility to individuals and departments in how they handle data locally.

Combined, these initiatives create the information infrastructure illustrated in Exhibit 1.

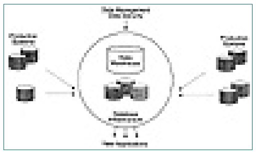


Exhibit 1 Information infrastructure.

APPLICATION DEPLOYMENT

If a concerted effort is made to identify them, dozens or even hundreds of application opportunities will emerge in most organizations. These will typically fall into four main categories:

- *Decision support systems* provide value wherever business decisions must be based on complex data and multiple variables. Properly used, these systems enable large volumes of data about markets, customers, competitors, partners, and the internal operations of the company itself to be rendered into a concise, actionable form.
- *Data marketing systems* enable mass customization; that is, they allow product and service offerings to be tailored to the individual preferences of any number of customers. Direct marketing yields can be increased by orders of magnitude, and customer loyalties can be reinforced and expanded over time by using new relationship-marketing techniques.
- *General-purpose information access applications* support the departmental managers, professional staff, sales personnel and front-line employees who account for more than 90% of the total workforce in many organizations. Separate initiatives will address the varied, tactically oriented informational needs of these individuals, many of whom are located at remote sites or use mobile computers.
- *Integrated information systems* are the next generation of informational computing. Designed to exploit new advances in database and workstation technology, they initially provide value in sales, service, and other applications involving interaction with customers. A few leading companies are also developing broader, more ambitious systems that use information to integrate processes throughout the enterprise that contribute to competitive performance and organizational efficiency.

As Exhibit 2 indicates, potential applications extend across a wide range of departments and functions in most industries and businesses, as well as in public-sector organizations.

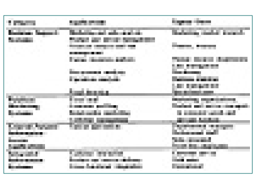


Exhibit 2 Representative informational applications.

Depending on application size, functional characteristics and user communities, various tools and technologies might be appropriate. Solutions range from large-scale, customized systems running on parallel processing servers to simple, single-user PC-based applications. Organizations should select whatever works best.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

ORGANIZATIONAL CHANGES

The technical realization of an information strategy is the easy part. Data can be extracted from online transaction processing (OLTP) and batch systems, regardless of their physical location, and be redeployed, without affecting their performance, functionality, and organizational ownership. Parallel database and server infrastructures can be created alongside existing production systems, and new applications can be deployed independently.

The main obstacles to deployment are organizational. The emergence of information as a distinct and valuable resource is a recent phenomenon. IS thinking and management practices are only now starting to catch up with this change. Three key changes may be required. First, IS structures may need to be revised. Most central IS organizations are subdivided into systems operations, networks, and applications functions. Where IS responsibilities have been decentralized, similar structures have typically been replicated at the division or business-unit level. Two new functions should be created. One would manage organizational data resources (including database infrastructures, data warehouses and data security mechanisms) for query exploitation. The other would oversee new informational applications, supporting user projects and assisting in selecting and implementing tools to support end-user information access. Exhibit 3 depicts this new structure.

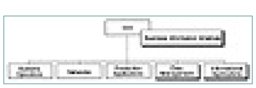


Exhibit 3 New IS structure.

Second, CIO-level initiatives should be launched. In many organizations, CIOs have begun to reinterpret their role as chief information officer rather than chief IS officer. They have begun to add new value to their roles by helping business functions identify new ways of using information. This process can start with compiling and publishing a directory of the production data collected within the organization. Using this, the CIO can act as critical intermediary among business functions, needs for information and the IS functions that generate and manage it.

Finally, it may be necessary to modify business practices. Where databases and data warehouses support multiple dispersed-user communities, it may be difficult to coordinate initiatives, define application-specific returns on investment (ROIs), and allocate costs. Information infrastructures should thus be treated in the same way as enterprisewide area networks, that is, as an inherently cross-functional business resource. It may be appropriate to define ROIs at the organizational level or simply accept that there are broader, nonquantifiable, competitive benefits that justify this type of investment.

Whichever approach is used, the investment must be made; the fastest and most cost-effective way to create an information infrastructure is to work from the center out. A strong, enterprise-level initiative can have multiplier effects, enabling divisions, business units, departments, and end users to implement applications with increased speed and lowered costs.

COMPETITIVE ALIGNMENT

The objective of an information strategy is to increase competitiveness. Infrastructure initiatives and application-deployment priorities should be targeted to have maximum competitive impact as rapidly as possible. Planners should begin with a model of how the organization competes, specifically by defining the organization’s model based on one of the following descriptions:

- The focused model applies to organizations (and divisions or business units in larger companies) that compete on the basis of a single or a few closely related products or services. Key data usually originates from a few major production systems. Infrastructures tend to be concentrated and easy to implement. Identifying application opportunities is usually clear cut. The objective is to improve the organization’s understanding of customers’ needs (i.e., decision support) and its marketing effectiveness (i.e., database marketing), and to provide new services that add value to relationships (i.e., customer information systems).
- The hybrid model applies to organizations that operate in multiple lines of business. These organizations serve the same customer base, or they have leverageable commonalities across customer bases in the products or services they offer as well as in marketing, operational, and administrative processes. This type of organization is typically large and is subdivided into divisions, business units, or equivalents. It tends to be cross-functional. Decision support systems analyze production data from multiple lines of business to identify integration opportunities, as well as to better manage cross-organizational risks and assets. General-purpose applications provide users with access to data from other operating units. Database marketing and customer information

systems are geared to realizing business synergies.

- The decentralized model applies to organizations in which most business operations and production systems are decentralized to individual operating units. Competitive strategies are specific to, and determined by, the management of these units. Infrastructures typically support the consolidation of data mainly to support decisions such as business analysis and planning, and to support corporate finance, human resources, procurement, and other staff functions.

By initially focusing an information strategy on critical variables of competitive performance, planners can make early gains and prove their concept. Other applications can then be deployed to exploit the existing infrastructure. Benefits are often immediate and powerful, as the following examples illustrate.

FOCUSED MODEL

The following sections describe the experiences of some organizations operating according to the focused model.

Data Mining by a Specialty Retailer

This retailer, specializing in apparel for style-conscious individuals, began as a new player in a crowded market, facing strong competition from established department stores and a wide range of specialty outlets. However, it has consistently expanded its sales and market share and is now a recognized leader in its field.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Early in the company's history, while it was still relatively small, management started investing heavily in decision-support and database marketing systems. The company began by mining production data from its point-of-sale (POS) systems. Using advanced data analysis tools, it sought key patterns in the sales of each outlet, department, and item. As it turned out, there were major differences in garment sizes among different sections of the country, as well as significant variations in preferences for colors, patterns, and materials. This information was used not only to realize major inventory and procurement savings, but also to select designs, plan product mixes, optimize store layouts and displays, structure promotions, and improve advertising and direct-marketing yields.

By collecting records on credit card usage, then assembling external data on their owners, management made another important discovery. Although the company considered young people to be its main market, it found that its customers included a significant number of mature individuals. They tended to have higher disposable incomes than did the young customers and, on average, to make more-frequent, larger purchases. New programs were introduced to target this group.

Analyzing POS data also enabled management to identify emerging trends at an early stage, which had an even more important competitive impact. Fashions change rapidly. By integrating different phases of its planning, procurement, marketing and delivery processes and using trend data to drive these processes, the company responded more rapidly and effectively to changes than its competitors did.

The company recently began to expand internationally. To understand local market dynamics in each new country, the company has initiated a data-mining pilot program.

Customer Service and a Package Delivery Company

This company faced an increasingly common challenge, how to compete in a business that offers few opportunities for competitive differentiation. It already offered guaranteed delivery, extensive geographic coverage, and aggressive pricing. However, it added a new edge — the ability to locate any package, anywhere within its network, at any time.

This was no small challenge because the company handles more than 11 million packages daily, employs more than 50,000 drivers, and has more than 1,800 operational centers. Packages might pass through the hands of more than 60 different people during the delivery process, and many packages are not bar-coded when they are collected.

The company designed and implemented an integrated delivery-information system built around some of the world's largest, most powerful, and most reliable database servers. Now, when drivers pick up a bar-coded package, it is immediately scanned and data is transmitted by cellular radio links directly to a central production database. Other packages are coded on delivery to local offices, and data is uploaded from these to the same database. Packages are scanned through all subsequent phases of the delivery process, including final delivery. Drivers record an electronic signature from the final recipient.

Data is transferred continuously from the production system to a separate relational database server that holds immediately accessible records for the previous 12 months of deliveries (more than 200,000 gigabytes of data). Earlier records can be retrieved from archives.

Customer telephone inquiries are routed to more than 60 delivery-information offices where service specialists can access the central query database to determine the exact status and location of a customer's package. If it has been delivered, representatives can cite the time and place of delivery, as well as the name of the person who signed for it. Electronic signatures can also be retrieved for inspection if necessary.

Using a specialized PC-based software package, customers can also access the relational database directly. They thus obtain the same information about any of their packages without even calling a service representative. The overall system is shown in Exhibit 4.

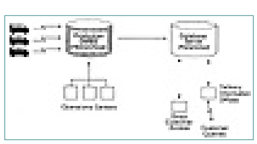


Exhibit 4 Database infrastructure of package delivery company.

HYBRID MODEL

The following sections describe the experiences of some organizations operating according to the hybrid model.

Data Warehousing by an Insurance Company

This company, a large, multiline insurance and health-services supplier, faced increasing competition. In its health-care business, it was also experiencing mounting pressures to control costs and improve service quality.

The company had trouble dealing with these challenges because its data was fragmented. Data generated by different production systems (e.g., application and eligibility verification, policy issue, claims-processing, provider payments, agent commissions) was stored in separate, difficult-to-access databases. It was thus difficult to retrieve data for analytical purposes, and it was even more difficult to obtain a logical view of the broader processes by which the company interacted with agents, policy holders, and service providers.

This fragmentation had contributed to major business inefficiencies. Key business segments were suffering from high service provider costs, premium delinquencies, and fraud. In addition, the company's ability to detect and quantify problems, let alone respond to them, was minimal.

Similarly, potential synergies between different product lines were not being exploited. In addition, the company operated in many different markets, each with its own regulatory environment that required extensive — and usually unavailable — regional information for effective policy planning and management.

As part of a transition to a strategy of cross-functional managed care, management approved the creation of a companywide data warehouse. Now in place, this system organizes data according to a custom-designed business model, and it integrates inputs from a wide range of production sources, as pictured in Exhibit 5.

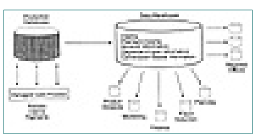


Exhibit 5 Insurance company data warehouse.

Previous | Table of Contents | Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Departments throughout the company can now access a wide range of data from all production sources, integrate and manipulate it in various ways, and perform highly granular drill-down analyses. Benefits include identification of new opportunities for rate increases; identification of a large volume of fraudulent claims, enhanced procedures for analyzing risk exposures and preventing losses improved benchmarking of service provider and agent performance; and changes in policies on pricing, content, and marketing that have increased the company's competitive position. New applications are being added on an ongoing basis. Many benefits have been tied to specific ROI gain. However, management has also recognized that warehousing has become a requirement for competitiveness in the insurance industry, as computerized claims processing was 30 years ago.

Database Marketing by a Regional Bank

Historically, this financial institution enjoyed a strong regional position through its hundreds of local branches, large ATM network, and established customer loyalties. The company was, however, facing new competitive pressures. The ability to use ATMs in a wide range of different outlets, telebanking trends, and changes in demographic mixes and buying patterns were eroding its customer base. New forms of competitive differentiation were required.

The bank chose a strategy of mass customization. This included broadening its range of financial services while packaging and marketing them more effectively to appeal to different groups of customers. Critical to the success of this strategy was the creation of a database infrastructure, illustrated in Exhibit 6, which allowed customer data to be pooled by different divisions and departments, analyzed to identify cross-sell opportunities, and linked to a powerful database marketing system.

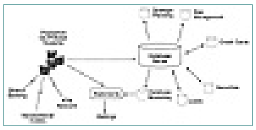


Exhibit 6 Bank database infrastructure.

Young households, among which the bank had suffered serious competitive inroads, were particularly targeted. Using highly granular data on demographics, individual and household profiles, and purchasing cycles, the bank customized offerings that would appeal to this group.

The ability to work with high levels of precision also enabled the company to leverage its regional position effectively. For example, marketing campaigns were targeted at individual rural communities and at specific zones in metropolitan areas. Relationship marketing programs were also tailored, and partners selected, for maximum local appeal.

The bank has now rebuilt its market share. It has also established a new competitive identity as a regionally focused supplier of a complete range of family financial services.

A Sales Information System for a Manufacturing Company

The business strategy of a major electronics manufacturer focused on supplying a range of products to meet a wide spectrum of customer needs. In principle, this should have been a significant competitive edge, but, all too often, it was the reverse.

One of the company's primary competitive weaknesses was information overload. The complexity of the product line, along with frequent changes in it, caused confusion among its salesforce. Each product division generated, and delivered separately, copious amounts of product documentation, marketing literature, pricing updates, and technical notices. Representatives faced stacks of new material virtually every day. Reading it, let alone digesting it, was a full-time job.

To make matters worse, sales personnel often covered large numbers of customers. With even basic data about customers difficult to obtain, putting together a detailed account profile for a sales call could take days of work.

Now, all sales-critical data is consolidated in common databases that are continuously updated and easily accessible from the mobile computers used by all representatives. Databases can be queried to obtain information on specific products, prices, availability schedules, and promotions, as well as on the status of individual customer orders. Representatives can also call up account histories, organizational details, and an online library of press stories and financial information about their customers.

Users can request that the system track specific customers, competitors, or types of products for them. Data will be duly collected and organized, and it can be accessed and downloaded for review when needed.

This system has brought about major gains. Sales have increased across the board, and customer satisfaction has improved. Many customers have reported that the company's representatives are the most knowledgeable they deal with. Reaction among the salesforce has also been highly positive, not least because the focus can be placed on selling, not reading.

DECENTRALIZED MODEL

The following section describes the experience of an organization operating according to the decentralized model.

Decision Support: Diversified Corporation

This diversified, multinational corporation, with interests in manufacturing and several service businesses, has more than 200 major subsidiaries and divisions grouped into major lines of business. Corporate management focuses on a single primary variable of performance: high earnings per share.

The company's business model is highly decentralized. Business unit executives possess a high degree of autonomy in most areas of decision making, subject only to meeting corporate financial targets. Because this model had proved successful in the past, management had no desire to change it. However, a weak economy was weakening demand for many of the company's products and services and was beginning to undercut profitability. The challenge was thus to improve organizationwide financial performance without increasing centralized intervention.

The company focused on financial analysis as a critical tool in achieving this goal. It extracted data from accounting, asset management, and certain other cross-functional systems used by all business units, and transferred this data to high-performance decision support systems. A complex range of financial parameters were then analyzed for individual business units, lines of business, and the company as a whole.

Important opportunities for new cost savings were identified. Certain units have been merged; others have been divested; and operational, support, and administrative functions have been consolidated in a number of areas. A corporate procurement group has also been established to identify common products and services required by multiple business units and to coordinate purchases.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Search this book:

Go!

Previous | Table of Contents | Next

Chapter I-4
Delivering IS Service Quality

Jeffery M. Ferguson
Robert A. Zawacki

Providing quality service to customers is now recognized as a key competitive strategy in many organizations. The growing interest in service is, to a large extent, a reaction to the demands of customers who have not only lost their tolerance for inadequate service but who, for the most part, enjoy the benefits of a buyer's market and can purchase similar products from any one of a number of competitors.

IS organizations within such customer-focused companies also realize the importance of quality in the services they provide. Just as the company's external customers are becoming more demanding, so are the IS department's internal customers. The trend toward out-sourcing IS services represents these internal customers' ability to find similar support elsewhere. Therefore, to retain the substantial impact they currently have on the business and to continue to provide a valuable contribution to the business units of the company, IS organizations must examine their own ability to provide quality services and to meet the IS needs of the company.

COMMON COMPETITIVE STRATEGIES

To gain a better understanding of a quality service strategy, IS managers must understand the basic strategies organizations use to differentiate themselves from the competition.

Technical Quality Strategy

One of the most widely used strategies, this approach is based on developing and maintaining superior technical quality in goods and services. Although a product will definitely not be successful if its technical capabilities are not up to par, the technical dimensions of a product do not represent the whole story for most customers. This strategy may therefore work well when technical excellence is of critical value to the customer; however, in the competitive environment facing many organizations, it may be difficult to demonstrate technical superiority. For many organizations, it may be too expensive or impossible to create a differential based solely on technical quality. The technical strategy therefore provides necessary but insufficient benefits to customers.

Price Strategy

Another common approach is the price strategy. This strategy has a strong intuitive appeal because low cost can be a powerful motivator. In addition, price is a quantitative dimension that can be easily communicated to the customer. The biggest drawbacks to this approach are the difficulty of maintaining an advantage over the long run and the subtle message it relays to customers to shop around for the best price. It is difficult to develop a lasting relationship with customers who have become preoccupied with finding the lowest price. The drive for the lowest operating costs may also suppress efforts to deliver a quality product or service. The organization that skimps on technical quality to lower prices may be unlucky enough to find that its customers are willing to pay more for a better-quality product from a different company.

Image Strategy

The image strategy focuses on creating an environment or aura around the service or product through marketing efforts. For example, clothing and cosmetics are traditionally marketed through image appeals. Services such as American Express and IBM computer leasing use a strong element of image marketing to make customers feel confident and even proud of the service vendor they are using. This approach can be successful at setting customer expectations and reinforcing value if the services that are promised are actually delivered. The tendency to concentrate more on talking about the service than on delivering it should be avoided at all cost.

Service Strategy

The service strategy creates a differential advantage by developing a range of services that add value for the customers and build strong relationships with them. It does not mean giving less attention to technical quality. The emphasis is on delivering a broader component of quality that includes both functional and technical quality. The goal of this strategy is to cement relationships with customers so that the organization is less susceptible to outside competition.

To be successful with its internal customers, an IS organization should combine various elements of each of these strategies. However, the service strategy represents the best approach for IS organizations, and they should consider it to be their driving philosophy.

THE CHALLENGES OF SERVICE QUALITY

Before attempting to improve the service quality the IS department offers, IS executives must understand the basic complications inherent in improving service or implementing a service strategy. Because of the overwhelming problems that can arise if the IS organization overpromises the level of services it offers, three principles must be taken into consideration as part of a program to improve the quality of IS services.

Service Depends on the Performance of Employees. Because quality service occurs during delivery, it is subject to the vagaries of the human disposition. In addition, services are performances rather than objects, which makes precise specifications difficult to establish and enforce.

Service Quality Is Defined by the Customers. Because service quality cannot be objectively measured by weight, size, or durability, organizations must rely on customers' perceptions of service quality. Service delivery personnel must view service from the perspective of their customers. The IS organization must therefore collect input about service quality from the customers and base its definition of quality on that input.

Services Are Evaluated Not Only by What Is Done but by How It Is Done. Technical quality relates to what is delivered; functional quality relates to how the what is delivered. For an IS organization, technical quality might pertain to the accuracy, timeliness, and relevance of a particular application. Functional quality, on the other hand, would be represented by courtesy, responsiveness to any special needs of the user, recovery from errors, and a smooth change from a previous application.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

The IS organization is not a collection of material assets; it depends on the quality of the contacts between IS systems and operations people and their customers in user departments. It is during these contacts that the IS department has the opportunity to prove to its customers that it is a quality provider of the services they need.

ADVANTAGES OF A SERVICE STRATEGY

The service strategy has two primary advantages for IS organizations. First, it views total service quality as a function of both technical quality and functional quality. This balanced view is more in keeping with customer (i.e., user) demands. When functional quality is neglected, total service quality suffers. Technical dimensions remain important in the service strategy but are not considered the main points of differentiation. They are a necessary but not a sufficient condition for delivering total service quality.

Organizations focus on technical quality when their highest priority is developing state-of-the-art technology. Organizations driven by research and development would probably view their priorities this way. Service customers, on the other hand, have a different set of needs representing more of a balance between technical and functional quality. They are not as concerned about the technology being used as they are about the benefits they receive. The strengths of the IS organization using the technical quality strategy do not match the needs of its clients, leaving the organization susceptible to competitors. The service strategy, on the other hand, provides the balance required by most customers.

The second strength of the service strategy is that the organization's strategic focus becomes customer relationships. Because competitors often can duplicate the technical aspects of services, the customer service strategy provides a greater opportunity for maintaining a differential advantage. The customers who experience only the outcome of a service may not even be aware of the technical aspects of service quality. Interactions with the service personnel (i.e., functional quality), on the other hand, are highly visible, and, as a result, customers evaluate these activities in detail. IS organizations should therefore focus more on these relationships with customers—relationships they may have been ignoring in favor of internal process improvement.

To encourage this new focus, IS management should look for such service skills as implementing effective change with the customer, negotiation and conflict resolution, and group dynamics when hiring new employees. In addition, IS management should be providing training in these skills for current employees.

IMPLEMENTING A SERVICE STRATEGY

One approach for implementing a service strategy in the IS organization has five basic steps to be considered:

- Identifying customers' definition of service quality.
- Developing a customer orientation.
- Training employees in both technical and functional quality skills.
- Providing excellent internal service.
- Making the most out of all contacts with customers.

Customer Assessment

Customer assessment is the cornerstone of the service strategy for two important reasons. First, without an understanding of how customers define quality service, any actions taken are merely guesses about what benefits customers truly are seeking. Second, quality will not be improved unless it is measured, and the only way to measure it is by having a definition to compare results to.

The assessment of customers' service expectations should include both functional and technical dimensions of service quality, including:

- Reliability.
- Responsiveness.
- Competence.
- Access.
- Courtesy.
- Communication.
- Credibility.
- Security.
- Understanding the customer.

Reliability and competence relate to technical quality, but all the rest are elements of functional quality.

Customer Orientation

For IS organizations, adopting a service strategy means shifting from a technical orientation to a customer orientation. Technically oriented organizations focus on internal procedures. Their objective is to develop systems that allow for efficient workflow. A production-dominated system is designed and evaluated in terms of readily measurable criteria, such as uptime, or the number of jobs completed, as opposed to the extent to which customers' needs are being met. The focus is usually on tasks rather than on people. Production-oriented organizations embrace the bureaucratic model, value efficiency, and specialize in the unwavering and timely performance of routine tasks.

Fundamentally and strategically, a customer-oriented organization focuses on ways of meeting customers' needs rather than the systems' needs. It takes a proactive approach to developing and delivering benefits for those it serves. This is a key difference from a production-oriented organization.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Because service quality is defined by the customers, a strong customer orientation is a requisite for the service strategy. It means giving priority to customers' needs and delivering to them functional as well as technical quality. In sum, the service strategy delivers a more complete package of quality to the customer.

The customer behavioral response to technologically driven change must be a primary concern of IS management. The probability of implementing effective change within and outside of IS can be increased by:

- The extent of information employees have.
- The extent of their participation in the planning and design process.
- The participants' trust in the IS executive and project officer.
- The participants' past experience with IS and how change was implemented in the past.

Training Employees

Technically oriented organizations with highly skilled professionals run the risk of defining quality too narrowly. Many of these organizations require that their employees receive substantial training in the technical aspects of the business but neglect customer interaction skills and procedures. Service delivery personnel must be versed in both technical and functional aspects of quality.

Internal Marketing

Excellent service to internal customers is a necessary prerequisite for delivering quality service to external customers. Internal marketing involves applying the philosophy and practices of marketing to employees. Its main objectives are to employ and keep the best people and to motivate them to high levels of performance.

In effect, internal marketing sells the employees of the organization on the organization and its products or services and encourages them to communicate their commitment to service to outside clients. This works within the IS organization as well. In essence, the way that IS employees are treated is reflected in the way they treat their internal clients. As a result, any service quality program must begin with service to employees before service for customers can be expected.

ACTION PLAN

Implementing an effective service quality program is a major cultural shift for an IS organization. These cultural changes usually take between two and three years to implement; they are evolutionary, not revolutionary. The following guidelines and the checklist shown in Exhibit 1 should assist the IS executive in successfully implementing the changes that will lead to improved IS service quality:

- *Implement a service quality improvement program only after a careful diagnosis of the organization’s readiness for change.* Do not start a change program if the IS organization is going through extreme stress (e.g., a movement of systems people to the end-user community or a reduction in workforce).
- *Plan for institutionalization of the change program.* Many change programs in IS do not last because more effort goes into planning and implementing the change program than into maintaining it. A steering committee should be established that reports directly to the senior IS person. In addition to managers, three or four individual contributors should serve on this committee. It will increase the quality of the decisions and provide links to the other individual contributors.
- *The senior IS leader must constantly reinforce the quality program.* For example, the senior IS manager must make certain that quality is mentioned at least three times in any presentation being made to other IS people.
- *Ensure that change efforts are compatible with organizational values.* For example, the senior IS manager may express the need for open communication; however, when a quality group makes a recommendation that challenges senior management, does senior management send a signal that it devalues openness?
- *Establish measures of customer service before the quality training begins and track progress.* Senior IS managers need to demonstrate value added to the CEO.
- *Review the reward structure.* Link rewards (e.g., bonuses) to desired new behavior. Consider giving the steering committee or key contributors a reward. The behaviors that are rewarded are repeated in IS organizations.

- *Stay the course and establish adequate links across functional areas and from top to bottom.* Prepare a simple one-page report that describes what happened at quality meetings and then have that report sent to other quality teams and IS managers. Some IS organizations have an online data base that anyone can have access to and review what other quality teams are doing.

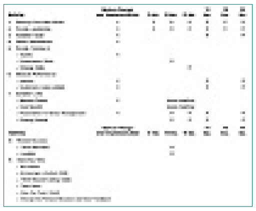


Exhibit 1 Checklist for implementing a service quality program.

In summary, high-performance IS organizations need leaders who recognize service quality as a critical success factor. Furthermore, service quality will be measured not only by the user or customer but by how IS leaders respond to the needs of their individual contributors within IS. The effective IS manager must value employees and develop a service quality orientation within the IS staff so they can provide exceptional service to the customers.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter I-5
Achieving Quality Outsourcing

William Perry
Steve Devinney

Quality outsourcing is the process of protecting an investment by managing and controlling the risks to the organization. Quality outsourcing is more than looking for outside competencies, it's looking for a business partner who practices the same values and business principles as one's own company and who augments the company's process capability so that the company can achieve and sustain customer satisfaction, both internally and externally.

QUALITY CHALLENGES

As a risky business and a process that must be planned, managed, and controlled, outsourcing can have dire consequences under the following circumstances:

- The organization fails to identify its business goals and objectives to prospective outsourcers.
- The vendor doesn't perform to the agreed-on levels of performance.
- The organization fails to adequately define requirements.
- The organization unknowingly signs away its rights to seek relief in the event the vendor fails to perform.
- The vendor does not have the competencies needed.

The Quality Assurance Institute's research and interaction with more than 1,000 member companies has helped distill these consequences into two major challenges in outsourcing preparation:

1. selecting a quality outsourcer;
2. controlling product quality and services received.

SELECTING A QUALITY OUTSOURCER

There are two important considerations when selecting an outsourcer:

1. the outsourcer's competency;
2. whether the outsourcer has implemented well-defined processes that meet organizational needs.

Obviously, no organization wants to engage the services of a vendor less competent than itself or one that hasn't already implemented disciplined processes to meet goals and objectives. Well-defined, disciplined processes are an effective way of managing performance risks.

In most companies, a process known as the request for proposal (RFP) is the driver for vendor selection. This is usually a rout process just waiting to happen. What organizations fail to realize is that the RFP process must be customized for each opportunity and that each opportunity should be founded on existing business needs, goals, and objectives.

Because the selection process is the first opportunity to display a professional awareness and understanding of business risks and business needs, an ad hoc or chaotic process is likely to cause loss of the opportunity to positively affect the direction or quality of a vendor's work. By exposing its vulnerability during the process, the organization places itself at further risk.

Defining the Selection Process

The selection process should be defined through a method based on team dynamics. The following prerequisite tasks need to be accomplished to ensure success:

- *Appointing an outsourcing manager*. One individual should be responsible for planning, facilitating, and coordinating all project activities. This individual becomes the prime spokesperson for the company during negotiations with prospective vendors.
- *Appointing an outsourcing committee.* The outsourcing committee should comprise business experts capable of defining the attributes of the process and product and technical experts who can communicate technology goals and constraints. As a whole, the committee must understand why a particular process/product is being considered for outsourcing and management's vision, goals, and objectives.

Once these tasks are completed, the committee defines the process for selecting an outsourcer, i.e., how bids will be solicited and evaluated and final selection made.

Evaluating Vendors

The competency of the outsourcer and the process discipline of prospective vendors are equally important and should not be taken lightly. Both reflect the vendor's ability to perform and meet immediate and future needs. Following are some of the most important criteria for vendor evaluation.

Certification. It is important to ascertain whether prospective vendors have achieved certification for an industry-accepted standards model (e.g., SEI CMM, ISO 9000, SPICE). If an assessment has been performed, it is valuable to request copies of the assessment findings and the improvement plan. Having an assessment performed for candidates who have not done so on their own helps determine their level of process maturity and competence.

Employee Interviews. It is becoming a common practice to interview employees of prospective vendors to ascertain whether they have the proper skills and experience to meet organizational needs. At a minimum, the individuals who will be responsible for project management should be interviewed. These interviews should be conducted in the same way that interviews of prospective in-house employees are conducted.

Client Interviews. Another method of evaluating vendors is to interview their current and past clients regarding performance received. Vendors should be asked to supply a list of all clients served during a specified time frame, usually of from three to five years.

Documentation. The request for proposal should ask each vendor to submit the documentation for one business process. This process could be requirements management, project planning, management and control, or configuration management. The documentation can be reviewed and assessed internally or by engaging the services of a specialist.

It is critical not to rush through the vendor-selection process. Because the time spent performing this procedure is the most challenging and at times the most frustrating, it is easy to rely on the industry reputation of prospective outsourcers. This is a mistake. IS managers who take nothing for granted, leave no stone unturned, and do their homework diligently will reap the most rewarding investment from their efforts.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

CONTROLLING QUALITY

Controls are a method of managing the risk that an outsourcing vendor will fail to meet organizational objectives, schedules, budgets, and deliverables. A 20-page statement of work, augmented by a 60-page legal contract, is not sufficient to manage this risk and protect what can be billions of dollars invested over an extended period of time. As retrospective documents that impose penalties after the fact, contracts are not a suitable tool for remedying customer dissatisfaction.

The outsourcing process is more effectively controlled through these methods:

- requirements definition
- formal change management systems
- linking outsourcer milestones with quality reviews

The Requirements Definition Process

Outsourcing puts certain obligations on both parties, whether the requirements are functional, constraint, or quality. It is incumbent on IS managers to ensure that needs are defined at a level of detail sufficient not only to enable a vendor to bid accurately on the project, but, more importantly, to deliver the goods or services.

Defining organizational needs and expectations through the requirements definition process is thus the driver for vendor selection and the baseline for assessing vendor performance. Documenting requirements is accomplished through the request for proposal; it serves as the tool for defining the vendor's work and products and internal organizational constraints (i.e., budget, schedule, and operating environment).

The following guidelines are important to ensure an effective requirements definition process.

- If something is important to the organization, document it as a requirement.
- If something is not documented, it will not be performed.
- Requirements that are omitted, not clearly stated, or are constantly changing lead to a negative relationship that could result in litigation.

A major component of the RFP and the vehicle for defining needs is the statement of work (SOW), which is prepared by the outsourcing committee. Following is a list of the major components of a comprehensive SOW.

- project goals, objectives, and measures of success (i.e., performance measures)
- constraint requirements
- product specification
- quality and control requirements
- roles and responsibilities
- security, interface, conversion, configuration, and communications requirements
- start and completion dates, or milestones
- description of project end
- warranty or continuing support needs
- instructions for prospective bidders and description of the selection process

To develop a statement of work that encompasses all the preceding elements, the outsourcing committee should answer the following questions for each process or product:

1. Why is this process/product critical to the success of the project?

2. What must we do to satisfy the process/product objective?

3. How must we perform to ensure process/product compliance?

4. How will we ensure that the products produced meet our expectations?

5. What measurement data is needed to ensure that business goals and objectives are being met?

Incorporating into the statement of work process standards and procedures that delineate what must be performed and how, and product standards that effectively predefine the attributes of each product and establish a basis for compliance checking help ensure the quality of the outsourcing effort.

Formal Change Management Systems

Organizational change inevitably occurs. During the outsourcing engagement, two types of change take place. Controlled changes involve new products and services, redesigned work processes, new technology, and customer demands. They are usually aligned to a change in strategic direction, initiated at the executive level, and are closely managed throughout the organization. Conversely, organizations encounter uncontrolled or subtle changes that are more elusive and usually result from informal interactions, memos, and daily problem solving. Uncontrolled changes are frequently the source of friction, confrontation, and dissatisfaction among all participants in the outsourcing project and are the prime contributor to runaway projects. This phenomenon is often referred to as “scope creep.”

All vendors will certainly entertain changes to the original statement of work, but the possibility of the need for rework or additional resources means that changes increase the project time line and costs. To effectively manage change and maintain a positive working relationship with the vendor, IS managers should ensure that a well-defined change-management process is implemented. It should stipulate who can initiate a request for change as well as the method of classifying and justifying the urgency of change. It should determine whether all affected parties have been involved in the change and delineate decision-making criteria. Finally, it must specify who has approval authority and how changes will be monitored and controlled.

Previous | Table of Contents | Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

It is important to review the change-management process with the vendor, because most outsourcing vendors have their own formal methods of change control that effectively protect their interests. Before a contract is signed, it is imperative that all organizational participants understand and agree to the vendor's process and that the organization's procedure can be integrated into the vendor's.

The following attributes of the vendor's change management process should be closely scrutinized:

- how the vendor will assess the effects of change on completed work or planned future work;
- how the vendor will reforecast costs, schedules, and resource allocations;
- how the vendor will manage and control the change(s).

Change will happen, so success lies in how well it is managed, justified, and controlled.

Linking Outsourcer Milestones with Quality Reviews

Every project undertaken has been completed on time and within budget. Of course, this is not to say that the projects were completed according to original schedules or budgets. Given unlimited time and resources, a project of any magnitude can be completed. Does this mean we can solve the national debt? Hardly.

Successful completion of a project depends on the customer's understanding the vendor's processes. The project plan is a combination of organizational requirements and the methods the vendor will use to achieve them. It is impossible for an IS manager or the outsourcing committee to understand the vendor's work breakdown structure if the entire process is not understood.

To gain this understanding, the outsourcing committee should participate with the vendor during the project-planning process. Because not all activities are equally important, the committee should devote greatest attention to principal activities, or milestones. Completion of requirements gathering, for example, is certainly a major event in the development of software.

Major milestones often gain added importance because they serve as interfaces to other project activities. It is at these points that a control method should be deployed to ensure that performance/product expectations have been met. Some of the most common forms of control are reviews, walkthroughs, and inspections. The organization, as the customer, should participate in the control method to ensure that performance meets expectations and that corrective actions are taken to remedy any nonconformance issue. Controls should be installed as close to the completion of a product as possible and before interfacing to the next activity. Thus, the overall goal is targeted control, not overcontrol.

One of the most important milestones actually occurs much earlier in the outsourcing procedure. It is the contracting process. This critical point should not be taken for granted. Organizations can literally sign away all of their rights to relief by affixing a signature to a contract. All goals and objectives, the statement of work, methods to manage and control the outsourcer, and contract verbiage should be reviewed with legal counsel before contract signing.

RECOMMENDED PLAN OF ACTION

The following steps will help data center managers and their organizations achieve quality outsourcing.

- Establish management goals and objectives that are clearly stated, understood by all, and measurable.
- Select an outsourcing manager and committee.
- Define processes to solicit bids, evaluate responses, select the vendor, and manage and control the vendor.
- Define a statement of work by completing the following activities:
    - predefine the attributes of processes and products through standards and procedures;
    - emphasize measurement that is consistent with management's goals and objectives;
    - predefine the control methods to be deployed, who will deploy them, and when they are to be deployed;
    - obtain legal counsel's approval of the statement of work and contracts. Submit these documents to management for review and approval.
- Define the internal change-management process.

- Gain understanding of the vendor's processes.
- Establish project milestones and methods of control.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Search this book:

Go!

Previous | Table of Contents | Next

Chapter I-6
Acquiring and Deploying Advanced Information Technology

Louis Fried

The drive to improve has caused a burgeoning interest in new technology and its management. Executives have awakened to the fact that technology management is as important for their information systems as it is for their manufacturing facilities and their products. As a result, IS departments are under pressure to meet the same goals of efficiency, cost reduction, and responsiveness as other departments within their companies. This article details the objectives of technology management and the link between the introduction of new information technology and business process redesign.

TECHNOLOGY MANAGEMENT

Two major issues, inexorably linked together, trouble IS directors of major corporations. These issues are:

- Supporting the redesign of company business processes.
- Replacing legacy systems that are impeding the IS function's ability to respond flexibly to business needs.

The link between these two issues is the ability of the IS function to plan for, acquire, and deploy new information technology for the development and operation of new applications.

The internal needs of IS—better price/performance ratios, faster software development paradigms, specialized application capabilities, smaller increments of capacity increases, and improved architectural flexibility—generally require a substantial short-term increase in new technology costs before the long-term benefits can be realized. Although business needs have spurred the adoption of new information technology, in many cases, IS departments are poorly equipped to deal with the entire scope of managing the introduction of new technology to the organization. Technology management is not only a problem for IS. In fast-paced industries such as biotechnology or electronics, technology management is critical to their continued survival.

The scope of technology management includes a broad range of activities such as:

- Strategic technology positioning.
- Tracking technology trends.
- Aligning technology needs with business needs.
- Identifying appropriate new technology.
- Identifying the technology rendezvous—that is, identifying the relative importance of technologies to the business compared with the time at which the technology should be adopted by the company.
- Justifying technology acquisition.
- Acquiring new technology.
- Introducing new technology.
- Adapting technology to the business needs.
- Deploying technology.

This article offers a strategic perspective on some of these issues.

STRATEGIC TECHNOLOGY POSITIONING

Although many aspects of technology management could be assumed under this activity, this article takes a narrower view. Strategic technology positioning consists of adopting policies and procedures that set forth the management position regarding technology. An organization may determine that its competitive position is best served by being an adopter, an adapter, or an inventor of technology. Adopters frequently use off-the-shelf products and thus trail others in technology acquisition. Adapters make technology an essential element of their value-based planning and use new technology in innovative ways. Inventors seek opportunities through creating new or innovative uses of technology to stay far ahead of competition.

Furthermore, producers must accommodate the feedback of technology opportunities to the business plans. Strategic business plans developed without regard to the competitive threats and opportunities supported by technology advances can be blindsided by more aggressive users of technology. The clearly stated positions embodied in policies and procedures reinforce a strategic approach to technology planning and acquisition. They also ensure that valuable employee time is spent in accordance with corporate policy.

TRACKING TECHNOLOGY TRENDS

Large IS groups frequently have specific positions created to track technology. Technology tracking activities are often part of a technology planning or systems architecture group within the IS division. Although some managers feel that technology tracking is a part of every system analyst's job, the more successful results are obtained when the effort is not so diffused. Because successful technology planning must be continuous, specific assignment of responsibility is necessary. In smaller companies that cannot dedicate full-time personnel to this task, the responsibility should be made explicit for one or two individuals as a part-time function.

Technology tracking can only work properly in the framework of strategic technology positioning. It is futile to track emerging technologies if the company's position is that of an adopter. However, for technology adapters and inventors, a vision and understanding of the future and of technology life cycles is imperative. This understanding of the technology life cycle serves as a means to determine areas where skills need to be developed, to identify new projects, to improve productivity and quality, and to anticipate potential competitive advantages and disadvantages.

Technology Life Cycle. A technology's life cycle consists of six stages:

1. *Breakthrough and basic research.* The technology is invented and advanced to a stage at which product development is feasible.

2. *Research and development.* Initial products are developed.

3. *Emergence.* Products are introduced and the market is educated to accept the products.

4. *Growth.* New products using the technology continue to be offered at a rapid pace.

5. *Maturity.* The market for the technology stabilizes and products become commonly applied.

6. *Decline.* The technology is superseded by newer technologies that have functional, cost, or performance advantages.

By monitoring the stages of information technologies, which develop at different paces, a company can observe the applications to which technologies are applied. Usually, businesses must consider technology applications rather than simply individual technologies. Technology applications are generally constructed by blending individual technologies; personal computing, for example, was made possible by technology trends in miniaturization, local area networking, and graphical user interfaces, among other technologies.

Previous	Table of Contents	Next

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

ALIGNING TECHNOLOGY STRATEGY WITH BUSINESS NEEDS

Tracking technology without regard to the needs of the business can waste a lot of time and money. It is absolutely critical that those tracking the technology be able to:

- Appraise those technologies that may be used immediately for competitive advantage.
- Appraise changes or new developments that may be used for competitive advantage over the next three to five years.
- Identify potential applications of current and future technology and how such applications may affect the competition.
- Identify potential changes in current applications driven by market demands or technology developments.

Aligning technology strategy with business needs is one of the most frequently identified problems facing corporate IS executives. This alignment requires knowledge of the business operations of each strategic business unit in the corporation, their competitive business strategies, and the best available information on the technology and business strategies of their competitors. In addition, it requires the active participation of both information technologists and users to develop an understanding of the potentials for use of information technology and a consistent vision of the future.

Importance of Teams. Research in technology management and product development has shown the success of triad management. Triad management techniques create teams of marketing, technology suppliers (e.g., engineering or R&D), and manufacturing representatives to rapidly introduce new technologies or bring new products to market. Similarly, most successful IS implementations are those that were required and driven by the users. To achieve alignment, IS must form alliances and occasional task forces or teams with user organizations. Technology planning is no exception if the alignment of technology and business strategy is the goal.

Alignment of technology strategy with business strategy requires two modes of operations:

- The technology strategy must be able to respond quickly to changing business needs.
- The technology potential and vision must be able to influence the development of business strategy.

Technology Planning Specialists

Although the profusion of personal computers has forced employees and managers to become more computer-literate, most noncomputing professionals do not have insights into the full potential of computing and communications technology. Technology planning must continue to be the responsibility of technology planners.

Technology planners should make a point of meeting with user managers, not only to educate them informally, but to learn about their business operations and needs. Most people are flattered to be asked about their job, and most line managers readily respond to requests from IS personnel to learn more about their business operations.

Technology planners also need to discover how competing companies are using information technology. This does not imply industrial espionage, but simply tracking the trade press, attending industry or information technology conferences, and talking with prospective vendors and suppliers. Innovative applications can arise outside the company’s industry and be applicable to company needs, so this intelligence effort should not be confined to the company’s industry.

Increasingly, close cooperation with both suppliers and customers is needed to be competitive and to respond to market conditions. It is now necessary to view the organization and its business processes as a part of an extended enterprise composed of the organization, its allies, suppliers, and major customers. Technology planners must either create relationships within the businesses that can provide the perspective of the extended enterprise or they must initiate relationships with key suppliers and customers to understand their uses of information technology and how the company’s processes must interface with its business partners.

Building the Business Case

As technology planners acquire a knowledge of the industry and the business, they must document this knowledge so it may be used to build a business case for new applications. They also need to find a way to translate the needs of strategic business units into a projection of when and how new or emerging technologies will influence the company's industry.

With a knowledge of the industry's business processes, technology planners can construct a value chain. For example, in a manufacturing company, the value chain may contain the major elements of R&D, engineering, logistics, operations and manufacturing, marketing, distribution, sales, and service. New technology can affect any aspect of this chain through such areas as product design tools, materials or components procurement, inventory management, manufacturing methods and controls, maintenance of plant and equipment, packaging processes, and sales and service tracking. The potential applications that can provide leverage and maximize the contribution to the corporation from technology investment can be recognized. In addition, the technologies capable of supporting those applications can be identified.

Technology Rendezvous. Working with user managers, technology planners can gain an understanding of both the leverage points and the perceived priorities of managers. These factors determine what is called *the technology rendezvous*, which is the relative importance of various technologies to the organization compared to the time at which the technology should be adopted by the company. Exhibit 1 illustrates how the technology rendezvous may be presented to management.

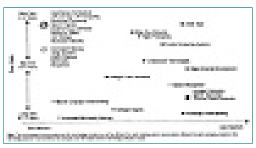


Exhibit 1 The technology rendevouz—charting the relative importance of information technologies.

From the example in Exhibit 1, it is possible to identify clusters of technology that must be introduced over time to keep a company competitive. In this example, the technologies in the upper-left corner of the chart will be most important in the shortest period of time. Technologies of immediate and short-term importance must be considered in the early stages of implementing a technology plan. The plan must consider other technologies in an evolutionary fashion.

The technology rendezvous is important for meeting strategic alignment objectives. It provides a senior-executive view of the needs for introducing new technology or extending the use of currently available technology to remain competitive, thus influencing corporate business strategy. Simultaneously, it provides IS with a view of when and how it should anticipate training, experimentation, and application development using these technologies.

Previous	Table of Contents	Next

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

INFORMATION TECHNOLOGY DEPLOYMENT

Technology is deployed through application. For example, a new technology for the manufacture of integrated circuits is deployed through the change of processes and equipment on the production line. Similarly, new information technology is deployed through its application for the benefit of the company and its processes and users.

Deploying information technology through applications means changes and requisite training for the application designers and developers as well as for the users. The processes that will change may include requirements definition, systems design specification, programming, user documentation development, application testing, system installation, system support, and end-user business processes. Substantive changes that allow companies to take full advantage of enabling information systems technology almost invariably require changes in the business processes and in the organizations that perform these processes.

The actual acquisition and introduction of new technology requires justification in terms that meet the approval of senior management. Even if the planning stages have been carried out with appropriate involvement and approval by senior management, the introduction of specific elements of the plan requires detailed planning and justification based on the benefits of applications to the business.

Reengineering

The goal of many organizations is to replace old systems with new ones that are easier to maintain and modify, support the types of user interfaces to which users of personal computers have become accustomed, and support new modes of business operation. For many applications, downsizing of equipment and the introduction of client-server architectures will provide significantly lower operating costs.

One reason that apparently successful systems development projects have not achieved expected benefits is that the applications were designed to support existing business processes. When existing business processes form the foundation for the requirements definition, even anticipated gains in efficiency may be unrealized. The result: companies have begun using a collection of methods and tools under the general name of business process redesign (BPR) or reengineering.

Many consulting firms have adopted variations of the BPR methods. SRI International defines BPR as a methodology for transforming the business process of an enterprise to achieve breakthroughs in the quality, responsiveness, flexibility, and cost of those processes in order to compete more effectively and efficiently in the enterprise's chosen market. BPR uses a combination of industrial engineering, operations research, management theory, performance measurement, quality management, and systems analysis techniques and tools simultaneously to redesign business processes and to harness the power of information technology to support these restructured business processes more effectively.

BPR projects are designed to take a fresh look at a major business process from a customer perspective. The customer of a process may be the external customer of its products or services or may be the internal recipient of the process output.

Redesigning business processes using new technology benefits the company by improving efficiency and making business processes more responsive to customer needs. In the end, it represents a clear manner in which technology influences the strategy of the company.

BPR Methodology. SRI's seven-step BPR methodology and some of the associated elements, tools, and techniques used are outlined in Exhibit 2. By using this or similar BPR or reengineering approaches, organizations can overcome resistance to change and the problems of maintaining the effectiveness of employees during the period of transition to the new systems and business processes that affect the way employees perform their jobs.

Exhibit 2 Seven-step BPR methodology.

GUIDELINES FOR SUCCESSFUL TECHNOLOGY DEPLOYMENT

Project Framework. Few companies succeed with projects that massively change the entire company's operating structure. New technology introduced within such a project framework very often adds to a company's problems rather than solves them. Generally, projects that address a single process (e.g., materials procurement, loan approval processing, or claims processing) have a greater likelihood of success. Success in the initial project that introduces new technology is crucial to expanding the use of the technology in the organization. In fact, it may be critical to the entire process of acceptance of new technology by management and employees.

Keep Employees Well-Informed. All affected or potentially affected company personnel should be provided with regular information about the BPR project's purpose, status, effect on existing processes and employees, training schedules, and interface needs. This information should be delivered within the context of the expected benefits to the company's competitive position so that everyone involved retains a focus on the value of the change.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Acknowledge Downsizing Effects. If staff reductions are anticipated, employees should receive a statement from senior management at the beginning of the project about how such reductions will be managed. For example, a statement that all staff reduction will be managed through normal attrition can make a major difference in employee cooperation. Management positions on how employees will be retrained, what options exist for transfer or early retirement, or what types of outplacement support will be provided should be publicized. Some companies have offered bonuses to employees facing displacement so that they will remain at their positions until new systems or processes are completely installed.

Continuity. The implementation time for redesigning a major business process and introducing new technology that enables the new process may be two or more years. It is vital to the success of the project that the senior management sponsors of the project be committed to this time frame and that they be kept informed of progress throughout the implementation.

Continuity and consistency of implementation must be ensured. Midstream changes in the project plan should be discouraged, unless they are critical to the end result of the project (for example, a midcourse decision to outsource the function supporting the process) or caused by changes in the external environment of the business process (as in a change in the organization’s investment capability). Expect minor changes and finetuning during the period right after implementation.

Coordinated Assistance. Advice or assistance should be sought and encouraged from both management and the employees involved in the project during implementation. For example, it may be appropriate to seek the assistance of the human resources or personnel functions to deal with changes that affect employees. Furthermore, coordination with the managers of functions that use the output of the redesigned process must be maintained to ensure that new interfaces operate smoothly. If the process involves external suppliers or distributors, it may be necessary to set up a help desk or hotline during the implementation phase or train field representatives to assist such external participants.

Evaluate and Monitor the New Process. Where possible, it is essential to set up and evaluate a pilot operation of the redesigned business process. Debugging is always easier when only a limited part of the company's activities have been committed.

The results of the process change should be monitored in terms of the measures and goals established during the analysis phase, and these results should be periodically reported for at least the first year or more of the implementation. The measurement devices should be built into the new process to provide continuous measurement and to form a basis for further evolution in the future.

SUMMARY

Acquiring appropriate technology that can advance an organization's competitive position requires a dedicated and continuing effort. Acquisition of appropriate technology first requires that the technology planners understand the needs of the business and the strategic technology position that senior management has adopted.

Second, technology planners must maintain an awareness of technology trends to ensure the organization's ability to support its business needs and technology position. Technology planners must position IS not only to support the organization's business directions but to influence company or business unit strategy. Next, IS and technology planners must justify the adoption of new technology in terms of the new or improved business processes that the new systems will support. The key to successful introduction and deployment of new information technology is business process redesign that combines the business strategy, the process improvement justification, and the technology into a coherent approach that can be readily conveyed to senior management.

Previous	Table of Contents	Next

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter I-7
Using JAD for Strategic Initiatives

Michael C. Kettelhut

Increasing competition in the business environment has generated new management trends. One suggests that successful organizations are those that create environments that encourage continuous learning, which in turn fosters continuous change.[1]

The impetus behind the trend toward change-focused learning is found in such developments as total quality management (TQM); the advent of high-performing, self-directed teams; participatory management; global competition; and changing information technology. Change-focused organizational learning has the potential to improve response times, broaden general knowledge, lower costs, and improve customer service. Because learning requires transfer of knowledge between organizational members, it also reduces organizational dependence on individuals and may add depth to organizational resources. Sharing knowledge empowers individuals to act.

Another management trend suggests that employee empowerment yields substantial benefits to organizations as a whole and to individual functions. For example, empowered groups at Corning Inc.'s IS division are credited with achieving the following results:

- The division functions with 48% fewer operations personnel than the industry average, and 29% fewer total data center personnel than the industry average.

- Help desk calls resolved immediately increased from 75% to more than 90%, and the customer satisfaction rating improved from 78% to 100%.
- Operations services reduced staff by 20% and errors by 70%.
- Employees deadlocked in the same positions for more than 20 years developed job skills allowing them to move into new positions.
- Resources were reallocated to other pressing IS needs.
- A direct contribution was made to Corning's bottom line through generated savings of more than $500,000 annually.[2]

Thus, in the last few years, organizations experimenting with leaderless groups, quality circles, and participative management have reported increases in productivity and quality. They have also reported increased employee satisfaction and lower turnover. Yet, despite the benefits of empowerment, implementation is frequently problematic.

There is a process that can help the IS department reap the benefits of empowerment. Joint application development (JAD) — a facilitated process for using employees in groups to specify requirements for information systems — offers the promise of:

- improving the quality of delivered information systems;
- enhancing employee participation and learning;
- yielding many of the same advantages of employee-empowerment programs.

Through a series of facilitated sessions, JAD provides an efficient process for collecting information and generating novel solutions to problems. The result is generally higher-quality applications and smoother project implementation.

This chapter briefly examines the JAD process and then suggests ways that JAD can be used to more tightly link the development of information systems to corporate strategy. It discusses the benefits of JAD for the IS department and the applicability of JAD techniques to broader organizational tasks.

THE PROCESS OF FACILITATION

Developed in 1977 by Chuck Morris at IBM Canada, JAD methodology applies facilitation techniques to the development of system specifications. These techniques involve the use of formal procedures to prepare for and manage group sessions.

Although JAD is generally used by IS groups to generate detailed specifications, the methodology is useful as a general approach to managing group work across a broad range of organizational decision-making or systems-development tasks. For example, team building is a facilitated process that uses a neutral facilitator or counselor to help organizational members define mutually supportive goals and processes. Facilitated sessions are useful when companies begin to initiate the open dialogue necessary to form organizational strategies, create mission statements, or build common mental models of organizational processes and problems.[3]

The JAD process is straightforward. An executive sponsor approves a project and selects a neutral facilitator with the assistance of IS management. Other participants are selected for their functional business knowledge or their detailed task-specific knowledge.

All participants are interviewed and informed of the upcoming session. Information is collected to identify potential conflicts, requirements, operating assumptions, work flows, and system processes. Details of current reports, screen interfaces, menu structures, or other elements of the current system or work process are also documented. Identification of open issues (i.e., potential conflicts) is essential for later resolution during the sessions. The facilitator and scribe prepare an agenda, ground rules, lists of issues and assumptions, and lists of the basic requirements and functions of the proposed system.

Following the interviews, functional representatives participate in a group review of application requirements and design specifications. Participants are asked to adhere to specific ground rules during the sessions. For instance, attendance is mandatory; participants must attend a three- to five-day group session.

Because consensus is a goal, decisions are based on shared agreement. Consensus is defined as agreement that all participants recognize as necessary and will support, even though some individuals may not agree with what has been decided. Building consensus requires resolution of open issues. If an issue is not resolved, the facilitator assigns someone responsibility for its resolution.

The facilitator uses an agenda, charts, and visual aids to keep the process on track. If conflict occurs, the facilitator can invoke a 10-minute rule stipulating that disagreements not resolved in 10 minutes are documented as open issues. The final documentation is similar to documentation developed in traditional systems analysis and design efforts.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Joint Requirements Planning

Facilitated sessions can be used to define strategic requirements, functional needs, user requirements, and implementation details. A project with distinct phases may involve a series of sessions with different participants in each phase. For example, the development process frequently begins with joint-requirements planning (JRP). The facilitator interviews individual senior executives to identify their key requirements. Then a joint-requirements review is held. This is a facilitated session, and the desired result is an approved set of clearly stated functional requirements. A signed document containing the requirements is forwarded to the executive sponsor for approval before initiation of the next phase.

Exhibit 1 depicts the potential series of requests and inputs used in successive sessions and the corresponding status reports provided to the executive sponsor. The executive sponsor and three groups are involved in the process — senior managers who define requirements, end users, and the development team. Numbers associated with information flows denote the possible sequence of events; each number represents one cycle, and appropriate subevents are denoted with letters. As indicated, decisions made by the executive sponsor trigger the process and are required for continuation from phase to phase.

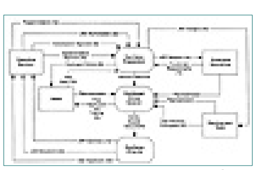


Exhibit 1 The facilitation cycle.

To begin, the facilitator solicits input for definition of key requirements. If there is disagreement over the functional requirements, a JRP session may be held. Later in the development cycle, the user group could review screen designs and prototypes. At each iteration, participants determine acceptability of the current product and the facilitator reports their decision to the executive sponsor. The groups are empowered to make key decisions that they agree are essential. On the other hand, the sponsor retains veto power and can stop the development project.

This iterative group approach has three benefits:

1. The exchange of information and discussion of new points of view increases quality;

2. The sharing and documentation of knowledge increases organizational learning and reduces organizational dependence on individual employees;

3. The executive sponsor exercises control over the project as it progresses toward completion and has the opportunity at the conclusion of each phase to assess the group's work.

LINKING APPLICATIONS TO ORGANIZATIONAL STRATEGY

Use of an iterative group approach can increase the effectiveness of IS groups involved in implementing organizational strategies that depend on information technologies. An organization should begin by defining its strategic goals, competitive position, and critical success factors (CSFs). These lead to specific functional requirements that are detailed through decomposition, which provides task-level definition for organizational procedures. Implementation of business applications requires detailed, task-specific knowledge.

Suppose processing time in physical distribution is a CSF. Physical distribution functions can be decomposed into, for example, order processing, inventory, warehousing, packaging, and shipping. Organizations that focus on total quality might benefit from having the project manager, the facilitator, or the JAD team work directly with customers to determine which physical distribution tasks are most important to the customer.

As the project moves from requirements definition to detailed design, the knowledge required of participants shifts dramatically (see Exhibit 2). At project initiation, strategic perspectives address critical factors, the external environment, competitors, and strategic alternatives. These perspectives must be transformed into functional requirements and then into application modules, a process requiring knowledge of the functional processes under consideration and their relationship to other organizational processes. Staff with detailed knowledge of procedures is required at this stage to ensure accurate mapping of functional tasks to application modules. As processing steps for each module are defined, programming knowledge becomes more important.

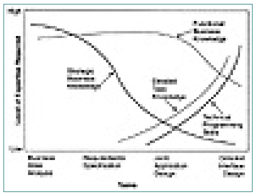


Exhibit 2 Knowledge/task changes in the development cycle.

Although Exhibit 2 depicts the level of expertise required through various phases, it does not clearly indicate the source of information used or the type of information required to implement new strategies based on changes in IT applications or architecture. Exhibit 3 presents a framework suggesting linkages between organizational hierarchy, key management responsibilities, and the functional requirements of organizational systems. Types of activities appropriate in facilitated sessions are also noted. As the exhibit illustrates:

- Strategy is defined at the apex of the organization and implemented by executive management;
- for requirements definition, executive management translates responsibilities into performance standards;
- requirements are translated into specific applications within functions and details of implementation are defined at the operational task level of the organization;
- task-specific user knowledge is required for specific application input, output, and processing requirements.

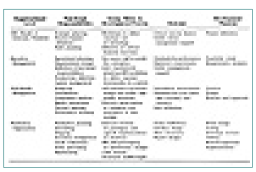


Exhibit 3 Linking applications to organizational strategy.

The participation of senior management ensures the transfer of strategic knowledge. This is particularly important when an organization's ability to compete depends on its ability to leverage information technology. The participation of middle management builds commitment and provides visible indication of support to end users. Finally, successful construction of an application requires end-user involvement to define system interfaces, tests of functionality, and prototyping of changes.

STRATEGIC BENEFITS

Studies report that use of JRP and JAD for system specifications reduces total project time by an average of 40%. Controlled experiments conducted by CNA Insurance Co. found that JAD methods increased productivity in the analysis and design phases of a project by 50%. The real difference between JAD and traditional approaches may lie in the compression of the one-on-one interviewing process into a three- to five-day meeting in which a group defines and approves the system specifications.[4]

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Although saving time is important, JAD also provides a formal structure supporting group interaction in a participative environment. The group approach provides direct benefits. It

- reduces the probability of developing the wrong application;
- reduces errors in specification that are one to two orders of magnitude more expensive to correct after implementation than during the analysis and design phase;
- increases user acceptance and improves implementation;
- broadens the base of participation and therefore furthers learning while reducing organizational dependencies on any single individual.

Learning is facilitated by a participatory, democratic environment, and workplace learning is maximized when people bring what they are learning into conscious awareness. Conscious awareness takes place during the questioning, reflection, and feedback that occur in the facilitated session. It permits deeper understanding to emerge from otherwise everyday activities.[5] Effective facilitators skillfully manage the interactive processes used in group sessions to help participants develop communication skills that focus on common meaning and encourage reflection about both the process and content of the discussion.[6]

Facilitated development should be viewed as a natural step in the evolution of applications-development methodology. Just as quality circles have proven effective in improving the quality and efficiency of manufacturing processes, JAD contributes to the IS development process.

CSFs OF STRATEGIC IMPLEMENTATION

Some organizations adopt participative practices only superficially. These organizations may publicly state their philosophy, but closer examination suggests that groups are not really empowered and that participation actually means more meetings but not more autonomy for participants.

On the other hand, several organizations have built their reputation and success around participative management approaches and the use of teams. These organizations are frequently industry leaders.[7]

The strategies for moving to empowered organizations with self-managed teams are not complex. One common thread underlying the implementation of quality circles, JAD, empowerment programs, and other participative management initiatives is the creation of an appropriate communications environment.

Organizations that capitalize on their human resources follow specific communication practices. One of these is full disclosure of operating information, such as specifics on the organization's cost structure and earnings.

Many organizations create unique cultures or value systems that support participation and empowerment.[8] R.M. Schwarz suggests that organizations' value systems must ensure:

- the availability of valid information;
- delegation of decision-making authority that lets employees make free and informed choices;
- a decision-making process that fosters internal commitment and choice.

The importance of these values to JAD is described in the following sections.

Valid Information. During JAD sessions, the facilitator reviews all the information collected from participants in the interviews that preceded the sessions. This public review lets participants challenge and validate the information. To maintain a focus on problem solving, the facilitator emphasizes the group's opportunity to start with a clean sheet of paper and discard the processes they have used in the past without respect to who designed them or how they were implemented.

Schwarz's values are reflected in these practices: information to ensure understanding is shared with the help of the facilitator. Furthermore, the independent presentation by the facilitator allows the group to validate information without regard to its source. Finally, the clean-sheet approach formally suggests the group's ability to change practices that result from previous decisions.

Free and Informed Choice. In well-orchestrated JAD projects, the support of senior management ensures that individuals who are selected as participants have relevant knowledge and interest. The participants are empowered to make decisions, and the facilitator helps prevent coercion or manipulation of any participant.

Internal Commitment to the Choice. Virtually all research examining commitment and participation suggests that involvement in the decision-making processes increases employee commitment. Furthermore, user participation in the development of information systems is one of three recurring themes in studies of successful development projects. The other two are clearly defined objectives and strong management support.

Organizations that adopt value systems such as those proposed by Schwarz empower their employees. One well-known organization that empowers its employees is Nordstroms, the Dallas-based high-service department store that outperforms most competitors. Nordstroms' value system is based on just one simple rule: "Use your good judgment in all situations. There will be no additional rules."[9]

The values prescribed by Schwarz suggest that organizations must avoid three dysfunctional behaviors.

- *Misuse of participation.* Such situations occur when participants recognize that there is a legitimate need to work as a team to solve a problem but are neither interested nor qualified.
- *False participation.* Here, a leader invites a group to a meeting but proceeds to sell his or her own ideas or agenda without allowing real participation.
- *Scapegoating.* This behavior occurs when the organization finds fault with individuals or groups.

The first two behaviors lead subordinates to recognize the organization's initiatives as phony and to react negatively to them. The third contradicts the need for an organization to use participative approaches, and specifically delegation, to be willing to allow individuals or groups to make an occasional mistake. When problems are identified, organizations must react proactively by focusing on finding solutions and learning how to prevent recurrence.

Proactive organizations that avoid these behaviors are better positioned to adapt facilitated approaches to systems development because participants recognize that they can make a difference during the process.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

JAD AND CHANGE MANAGEMENT

Today's competitive business environment presents several challenges to IS departments. Foremost is the pressure to reduce costs; many corporations are still downsizing, rightsizing, or outsourcing. Many organizations that have outsourced the operations portion of their IS department must still complete the analysis and design tasks, if not the actual programming of new applications. This environment highlights the importance of JRP and JAD for efficiently defining requirements and design specifications and producing what are often more-effective, higher-quality solutions.

Other business trends — service orientation, changing technology, TQM, teams, participatory management — are driving organizations to adopt a culture supportive of learning and change. These trends have also heightened interest in organizational reengineering.

Reengineering is a complex process. Reengineering projects that depend on leveraging information technology increase the pressure on the IS department. A great many reengineering projects have failed because of mismanagement of the change process. Participative approaches like JAD provide a formal framework for managing change by making affected groups part of the process and by building consensus.

The growing use of JAD reflects its potential in the area of change management. JAD sessions may make use of computer-aided software engineering (CASE) tools to capture information or group-decision support systems to automate the brainstorming sessions. IS organizations that market facilitation services can extend the benefits of facilitation to other organizational tasks such as strategic planning, creating an organizational vision, or reengineering projects.

Organizations whose IS departments use and market JAD techniques gain a competitive edge for the following reasons.

- They develop systems faster, with fewer problems in specification, greater user acceptance, and smoother implementation.
- They improve the quality of delivered applications and provide developmental opportunities for IS personnel.
- They provide opportunities for IS developers to gain sorely needed business expertise.
- They are better positioned to take advantage of end-user technologies such as client-server computing, distributed data bases, and graphical user interfaces.[10]

These benefits accrue to the organization as a whole. In the IS department, the process leads to more stable application designs, less maintenance, and lower staff turnover. As turnover declines and maintenance requirements are reduced, organizations improve their capability to respond to environmental changes within their current resource constraints.

CONCLUSION

Two steps are required to realize JAD's potential. First, the organization must understand that implementation of JAD is a step toward employee team building and empowerment. The organization must be committed to changing its culture as needed to reflect values that support participation and empowerment.

The process of empowerment requires that management allow workers to take responsibility for their day-to-day tasks, including planning, scheduling, human-resource decisions, quality control or quality assurance, and customer satisfaction. In essence, the process is one of building trust. The team at Corning asserts that there are eight key success factors for group empowerment.

- Start with a vision and clear goals.
- Ensure management commitment, visible support, and a willingness to take risks.
- Pay particular attention to middle managers and supervisors.
- Involve staff in all phases of the project.
- Communicate, communicate, and communicate some more.
- Keep your eye on the ball.
- Educate all those involved.
- Develop a reward system that promotes success.[11]

These factors also form the basis for successful use of JAD techniques in the IS organization. However, JAD is usually viewed as a technique used by a development group for a specific project rather than as a process of group development/empowerment that can be institutionalized. Many organizations that use JAD establish new teams for each new project, and team membership usually changes. The creation of functional JAD teams responsible for the ongoing development, implementation and maintenance of applications in a given organizational function is the first step toward the creation of empowered development groups.

Implementing JAD, a methodology with known benefits, supports the process of group empowerment and increases organizational learning. The structure used in JAD resembles the processes organizations have used for structured reviews of code.

IS managers will find useful information in several of the works discussed in this chapter, including information on where to receive training. Investing in formal training that prepares managers and analysts to act as facilitators in group sessions involves minimal risk and offers high returns that encompass more than the initial capital investment.

Notes

1. Watkins, K. and V. Marsick. *Sculpting the Learning Organization: Lessons in the Art and Science of Systematic Change*. Jossey-Bass, San Francisco, 1993.

2. Shrednick, H. R. Shutt, and M. Weiss. “Empowerment: Key to IS World-Class Quality,” *MIS Quarterly*, 12/92, 491–505.

3. Senge, P. *The Fifth Discipline: The Art and Practice of the Learning Organization*, Doubleday,New York, 1990.

4. Wood, J. and D. Silver. *Joint Application Design, 2nd ed*. John Wiley & Sons, New York, 1995.

5. Watkins and Marsick.

6. Senge.

7. Pfeffer, J. *Competitive Advantage Through People,* Harvard Business School Press, Boston,1994.

8. Schwarz, R.M. *The Skilled Facilitator — Practical Wisdom for Developing Effective Groups,* Jossey-Bass, San Francisco, 1994.

9. Pfeffer.

10. M. Kettelhut, “Group-Centered Systems Development: Improving the Development Process of High-Quality Systems,” *Strategic Systems* 4 6–8/92, 3–9.

11. Shrednick, Shutt, and Weiss.

HOME ACCOUNT INFO SUBSCRIBE LOGIN SEARCH MY ITKNOWLEDGE FAQ SITEMAP CONTACT US

SEARCH ITKNOWLEDGE

Go!

Keyword

Brief Full

Advanced Search

Search Tips

BROWSE BY TOPIC

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Bookmark It

Search this book:

Go!

Previous | Table of Contents | Next

Section II
Data Center Management

Section I of this handbook helped data center managers look at the "big picture" and see how the data center can fit into the IT organization as well as the entire corporation. This section homes in on the management of the data center and looks at shorter-term planning issues as well as day-to-day management of the center.

The workplace in the 1990s has become one where change and having to do more with less are both constants. Chapter II–1, "Centers of Excellence: Empowering People to Manage Change," explains a way that data center mangers can structure the data center so it can change with the organization and still deliver optimal service where and when it is needed most. It includes a case study from Texas Instruments, where the center-of-excellence model was used to restructure the entire IT department. This type of structuring is ideal for helping the data center make the transformation to a more distributed computing environment.

Most of each day in a data center is spent answering phone calls from users in the organization, who are not necessarily calling to chat, but to report a problem they are having with a system. How well a data center answers these calls and solves users' problems greatly influences how well the data center is perceived in an organization. Often, answering a user's call well involves not so much technological expertise as it does service expertise. Data center staff members are technically savvy, but they may not be customer service experts.

Chapter II–2, “Building a Data Center Service Culture,” shows data center managers how to enhance their staff’s service skills.

Another large concern in organizations today is the bottom line, which in some companies takes precedence over all else. Data center managers, especially those who are pressured by upper management to contain cost, will appreciate the tips offered in Chapter II–3, “Cost-Effective Management Practices for the Data Center.”

Controlling costs has motivated many managers to outsource certain IT functions. In particular, many center operations have been outsourced, and data center operations managers must deal with outsourcing on a daily basis. Chapter II–4, “IT Outsourcing: Current Trends, Benefits, and Risks” gives data center managers an overview of what is going on in the multibillion dollar industry of outsourcing. Chapter II–5, “How to Select an Outsourcing Vendor,” offers more practical advice by giving pointers in choosing outsourcing partners, drafting contracts, and negotiating the final terms.

Areas most often targeted for outsourcing are ones that do not contribute directly to an organization’s overall business. Although support services are critical in keeping an organization running, they usually do not play a direct part in its business activities. More and more data center managers are finding themselves responsible for hardware that is not actually in the data center. Desktop systems require a lot of support, which may be more than a data center can provide. Chapter II–6, “Outsourcing Desktop Support Services,” shows how to control costs of supporting end-user desktop systems.

As organizations acquire more and more computer technology, there are bound to be more and more calls to the help desk about it. In many cases, data center managers struggle to keep their staff on top of new technology so that they can answer trouble calls from end-users. Outsourcing some or all of help desk support is an attractive option in this situation, and Chapter II–7, “Outsourcing the Help Desk Function” provides a six-step approach for taking the outsourcing option.

As the data center has evolved from the glass box where the mainframe was kept, to today’s center, where mainframes, minicomputers, and servers reside, it has taken on an increasing support role. Managing the help desk has become a primary responsibility for many data center managers, though many of them come from a traditional, mainframe-oriented data center background. Chapter II–8, “The Help Desk in a Distributed Environment,” gives such data center managers needed background in managing help desks in today’s distributed environment, where systems and user calls have multiplied.

Chapter II–9, “Fine Tuning the Help Desk: Goals and Objectives” asks a series of questions that data center managers can use to make sure their help desks are meeting the needs of users. These questions are specifically designed for help desks that deal with questions about systems distributed throughout an organization.

Many data center operations have become more efficient and cost effective because they have been automated. Automation is providing needed relief to data centers overwhelmed by help desk calls. Chapter II–10, “Milestones in

Implementing Help Desk Automation," explains the various phases in managing the implementation of help desk automation. Of course, automation does not benefit only the running of the help desk. This section concludes with Chapter II–11, "Automating the Data Center," which covers planning and implementation issues involved in data center automation.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter II-1
Centers of Excellence: Empowering People to Manage Change

Steven W. Lyle
Robert A. Zawacki

Against a backdrop of endless organizational restructuring, today’s data center managers attempt to add value to the bottom line of their business units. While facing reductions in staff, they are challenged to maintain or even increase customer satisfaction and productivity. Positioning for the future becomes even more of a challenge when legacy systems must be integrated into an organizational vision.

Data center managers often look to structural change as the answer to the multiple challenges they face. The efforts of many managers to apply the latest quick fix or management tenets often fail, however, because they aim to implement changes designed for the organizations of the 1950s, not the 1990s and beyond.

In the past, change flowed along a reasonably predictable course. Individual contributors in the data center adjusted to change by working harder and smarter, and by introducing technology that helped them stay ahead of change. Today’s organizations, however, are facing high-speed random change that affects the direction, focus, strategy, and behaviors of the organization and its members.[1]

Several similar organizational structures have been designed to facilitate random change: the learning organization, the horizontal organization, the shamrock organization, the STAR (i.e., strategic goals in a constant state of transition and renewal) organization, and the high-velocity organization. One additional evolving design that has great potential as data centers make the transformation from legacy systems to client/server environments is the center of excellence (COE). This chapter describes the center of excellence and reports on Texas Instruments’ experience with it.

THE CENTER OF EXCELLENCE MODEL

Historically, organizations have bounced from centralized organizational structures to decentralized structures and back again. This cycle of redesign and restructuring has traditionally involved the transfer of control — the control of people. Yet organizational redesign is not about controlling people. It is about providing a strategy and structure that facilitates the growth of people by giving them the opportunity to do their best work. True organizational restructuring enables people to use their unique talents and abilities to the best interest of the business or institution — in other words, it empowers them.

The concept of a center of excellence is designed to help prevent history from repeating itself and stop the constant transfer of control over people. However, effective implementation of a center of excellence can be accomplished only by change agents who have abandoned the mindset of the 1950s.

The COE model concentrates on the acquisition and development of the skill sets required to foster the distinctive competencies organizations need to compete in today’s market. Two mutually dependent imperatives are key to success.

1. An organizational structure must be implemented that ensures fast mobilization and development of intellectual property (i.e., time-based competition).

2. The organization must engage and keep the customer’s attention (i.e., provide customer satisfaction).

Most organizations have focused more on the second imperative than on the first — that is, they focus on the execution of projects (i.e., engagement of customers) rather than on investing in the development of their people. This pattern does not result from either a lack of desire to invest in them or poor management per se but rather from a lack of focus. Organizational leaders and managers are human, and they can focus their attention on only a limited number of tasks.

Lack of focus on the development of people has caused many organizations to perform poorly in their ability to engage customers. The end result is that they not only lose customers, they lose their people too. Many companies never recover from this costly spiral.

CHARACTERISTICS OF A COE

The COE model provides a framework for creating an environment that allows organizations to address the imperatives for success in the 1990s and beyond and gain competitive advantage. This environment not only fosters superior customer support for legacy systems but also provides increased coaching and training to the people who maintain and work with these systems.

A COE is defined by the following characteristics.

- It is a logical grouping of related skills or disciplines.
- It is an administrative entity focused on the well-being and development of people.
- It is a place where individuals learn skills and share knowledge across functional boundaries.
- It is a physical organizational unit in which members are all together, or a virtual unit that is only a learning and communications vehicle.
- It matches resources to demand.

The mission of a COE is to place people resources where they are needed most by the business and to ensure that people are trained appropriately and have the necessary experiences and background to succeed on projects. Appropriate placement, training, and development of human resources necessitates that the COE be staffed with a forward-thinking coach who is able to stay ahead of the need curve. To ensure an equal emphasis on the two organizational imperatives to success, the responsibilities of the coach must be separate from those of the organization leader or project leader. Separating the coach's responsibilities not only helps achieve this equality of emphasis, it promotes the cultural change process by sending a clear signal to the organization that management is serious about valuing its people resources.

The COE coach thus provides clearer and more meaningful feedback to legacy systems maintenance and development people, who in turn achieve a better understanding of the future needs of the organization and their place in that future. When people feel valued, they add value to the customer. The COE is about valuing people.

IT TRANSFORMATION AT TEXAS INSTRUMENTS

Like many data centers around the world, the IT organization at Dallas-based Texas Instruments is faced with increased pressure to perform and deliver at greatly reduced cycle times. To meet the challenge, the IT organization commissioned a project to reengineer information technology at Texas Instruments. The project is known by the acronym RITTI.

The IT leadership team recognized that a transformation of the IT environment requires a concurrent engineering approach involving several elements:

- organization
- people
- business processes
- technology

The team realized that each of these elements alone would not guarantee successful business leadership for the IT organization or its customers.

Considering these elements together, however, could achieve major improvements.

Three major strategies address the elements:

1. A process strategy necessitates that the team map, understand, and address entire business processes versus piecemeal patchwork;

2. an architecture strategy based on Texas Instruments' component-based applications development methodology and object-modeling techniques separates the presentation level, the data level, and the logic (i.e., business rules) level. The methodology is facilitated by Composer and Arranger, two business products of Texas Instruments Software, as well as by the repository technology currently being jointly developed by Texas Instruments Software and Microsoft Corp.;

3. an organization and people strategy promotes the ability to develop, deploy, and retain the critical skills needed to compete.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Texas Instruments has successfully implemented the COE concept to achieve an equality of focus between the development of human resources and the engagement of customers. The COE model lets the IT team develop the talent necessary to support existing legacy systems while refining and executing the three transformation strategies. Unlike many organizational leadership bodies whose members assert that people are their most valued resource, the leadership team at Texas Instruments backs its words with actions.

TEXAS INSTRUMENTS' COE

Horizontal Skill-Centered Units

Texas Instruments' IT organization went from a vertical functional-department structure to a horizontal skills-centered COE structure that allocates people to vertical-project teams across the organization. The COE structure separates the traditional elements of control between two distinct roles: a COE coach and a project or organization leader.

To ensure that everyone understands the roles and the accountability within the COE structure, the IT organization rewrote organizational processes to reflect the new roles. The processes were published and then explained during open discussion meetings. Some of the major processes modified for the IT organization at Texas Instruments include:

- the performance-evaluation process
- the development planning process
- the compensation planning process
- the staffing and assignment process

- the knowledge-capture process
- the cost-management and labor-tracking process

The processes were originally published by an implementation team and later refined. Each process is currently owned by a COE coach who is responsible for leading any further refinement activity. Both the coach and project or organization leader work with COE members to ensure that each individual understands — from both a project standpoint and a self-development standpoint — the requirements of success within the organization and the market.

The Coach's Responsibilities

The COE coach has no responsibilities outside the center of excellence. The responsibilities of the coach include:

- training, developing, and assessing COE members in support of projects
- mentoring COE members in his or her areas of expertise or discipline
- facilitating and enabling the exchange and sharing of ideas and information
- recruiting and facilitating the assignment of individuals to projects
- managing the base salaries of COE members
- providing administrative support to COE members
- managing costs within the COE

The Project or Organization Leader's Responsibilities

The project or organization leader/manager has the following responsibilities.

- attaining performance objectives for all project milestones
- forecasting time-phased and skill-based resource requirements
- providing technical project direction and setting task-level priorities
- validating COE core competencies and future skills development
- providing coaches with feedback from team members' performance evaluations in support of the performance-evaluation feedback process and the promotion and base-salary adjustment process
- initiating and participating in the corrective-action process regarding performance issues
- managing and distributing variable compensation to project members

SUPPORTING HUMAN RESOURCES PROGRAMS AND SYSTEMS

The center of excellence at Texas Instruments is supported by three human-resources-related processes that are of great importance to the individual and the organization. These are:

- the staffing and assignment process

- the performance-evaluation process
- the compensation planning process

The following sections summarize the significant characteristics of these processes.

The Staffing and Assignment Process. Under the staffing and assignment process, administrative responsibility for individuals resides with only one COE. Individuals are encouraged to remain with their project assignment until completion of a major phase or milestone. They are also encouraged to participate in a wide range of assignments that provide exposure across the IT environment.

An open process for communication of assignment opportunities helps individuals achieve maximum exposure across the organization. Assignment changes are discussed with all involved parties (i.e., the individual, the project or organization leader, and the coach) before they occur.

The Performance-Evaluation Process. Performance-review sessions are held with the coach when individuals change assignments, or at least annually. They are initiated by individuals and based on feedback the coach obtains from various project or organization leaders. In addition, individuals are expected to collect 360-degree feedback throughout the year. As much as possible, performance-review sessions are separate from the compensation-review process.

The Compensation Planning Process. COE coaches manage the base pay (i.e., salary) of COE members. Base pay is determined according to the individual's competency level within the skill set/discipline, customer results, skills acquisition, teamwork, and knowledge sharing. Project and organization leaders manage variable compensation based on superb execution of tasks that map back to a key customer or organizational success criterion. All leaders are compensated based on attainment of performance objectives and staff development, not organization size.

IMPLEMENTATION OF CSFs

Many data centers are limited in their ability to enable the type of business change required by their customers. These organizations should take care not to repeat the historical change pattern of centralization and decentralization.

Markets characterized by random change warrant implementation of a learning organization that is based on the COE model, which helps legacy-systems people develop and improve their skills while organizations integrate legacy systems into their vision for the future.

Successful implementation of a center of excellence depends on several factors:

- Leaders who are visibly committed to the change in focus.
- A clearly articulated vision.
- Benchmarking activity that facilitates learning.
- Clear communication of the reason for and benefits of the change.

- Involvement of as many organizational members as possible.
- Thorough communications planning throughout the transition.

A center of excellence implemented along these lines provides the flexibility organizations need to meet today's two imperatives for success:

1. It engages the customer.

2. It provides continuity to an individual's career and development while adding value to the bottom line of the customer in a timely fashion.

Organizations should not waste time with repeated reorganizations. An organizational design that is flexible, responsive to random change, and customer-focused will outlive the next technology wave. The center of excellence model offers a solution, not just another management fad.

Note

1. R.A. Zawacki et al. *Transforming the Mature Information Technology Organization* (Colorado Springs, CO: EagleStar Publishing, 1995), 22–23.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Chapter II-2
Building A Data Center Service Culture

Richard D. Hays

Excessive cost and insensitive response (whether perceived or real) are the focal point of most complaints about data center services. Although some data center managers are able to deal effectively with these complaints, most are caught up in a continuing battle with service consumers to prove their departments' worth and contribution. Unfortunately, this effort often devolves into a downward spiral of accusation, defense, and conflict.

As the relationship deteriorates, internal customers often express their displeasure with the type, level, and cost of data center service, based on their definition of what that service should be. Any explanations and clarifications offered by the data center may be seen as defensive and insensitive. The situation, if unresolved, can breed negative stereotypes based on misunderstanding, not facts. The conflict can become personalized and entrenched. Eventually, any data center effort to improve service is met with doubt and condemnation. Most management time and effort must focus on handling complaints rather than on improving the data center contribution.

Relationships with internal customers, trust, and overall organizational productivity are the victims of this ongoing conflict. The turmoil created provides data center managers with strong incentive to improve service. Good intentions alone are far from sufficient, however, to bring about a significant change in service, or perceptions of service. Building a service-driven culture is essential to achieving satisfactory results.

THE NATURE OF SERVICE

Data center managers know what is needed to build a technically competent organization. It is second nature for data center managers to emphasize technical excellence — it is known territory that is consistent with their training and experience. The data center function is really a service organization, however, and it is ultimately judged in terms of customer satisfaction. The demands and objectives of a service-driven organization differ markedly from those required of a technology-driven one. Technology is important, but only as it contributes to service. However, few data center managers have minimal training or experience in the mechanisms and methods of building a service-driven organization.

Service is a difficult-to-measure and perishable commodity that immediately disappears upon execution. It usually cannot be examined retrospectively, and judgments of its quality are almost totally based on consumer perceptions. For most data center managers, these perceptions are soft, elusive variables that do not necessarily respond to concrete, often rigid, technology solutions. User perceptions are an unavoidable part of service delivery, however, and the effective manager of a service-driven organization must master the science and art of building a data center that can consistently deliver high quality and responsive service at an appropriate cost.

The brief time in which the customer directly receives service from the data center is a critical factor in defining perception of the data center and its services. Jan Carlzon, the CEO of Scandinavian Airlines System (SAS), has referred to these key interchanges as “moments of truth” for the customer. If the data center can identify and manage these interchanges, a key step has been taken to dramatically improve customers’ perceptions of service. In many cases, most data centers cannot predict when customers will ask for and use data-center-generated information. The interactions that tend to shape the overall view of service most heavily, however, are those face-to-face interactions between data center personnel and departmental users or senior managers. The effectiveness of these moments is largely determined by the culture of the data center.

THE IMPORTANCE OF CULTURE

Every work group — jury, military squad, sports team, or data center team — develops informal guidelines and values that shape the views and behavior of its members. This shared organizational culture helps individuals understand how the unit functions and their role and expected behavior within it. The culture is a pervasive and ever-present influence on the assumptions used by individuals and the way they see themselves and their customers. Because the service interaction is so transitory, its quality depends heavily on the skills and attitudes of the people delivering the service. This, in turn, is a function of the overall culture of the unit.

Culture can be evaluated on both strength and fit. A strong culture has many elements that are mutually reinforcing and can greatly influence individual behavior, whereas a weaker culture is less internally aligned and exerts less influence over behavior. Some cultures reinforce key goals and desired values, others do not support positive accomplishment or are not aligned well with each other.

IBM and Apple Computer are both organizations with strong, well-identified cultures that heavily shape individual behavior. In 1979, for example, IBM launched an effort to create a product to head off the market inroads of the highly successful Apple II personal computer. IBM realized that its strong culture had served well in prior competitive encounters, but it was likely to stifle the creative effort needed to quickly and successfully enter the PC market. Therefore, IBM created a special task force that was purposely separated and isolated from the broader corporate culture to foster the needed speed and creativity. Such norms as insistence on only IBM-manufactured components and a careful, plodding approach were purposely abandoned. Consequently, the IBM PC was introduced in 1981, the result of explicit recognition of how a unique culture was essential to a specific corporate goal. To provide a unique or improved service, a data center must carefully examine the elements that make up the existing culture as well as those that are needed to define the new service-driven culture.

CULTURAL REINFORCEMENTS

The specific culture of any group is built and reinforced by several elements that act to remind individuals of its nature and goals. They are discussed in the following sections.

Language. The language used within the work group serves as a continual reminder of a mindset and approach. A data center that provides service to users acts differently from one that deals with customers. Users simply use data center services — not a very demanding role for the supplier — but customers have needs that must be actively explored and met to their satisfaction, a much more challenging task. Service people who see themselves as employees tend to behave differently from associates or consultants, as do service agents who help clients solve problems rather than those who merely handle complaints.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Although simply changing language is insufficient to change behavior significantly, it is a critical element that may be most clearly understood using the Disney organization as an example. Disney believes that its theme parks are best served by building a show-business climate and culture and goes to great lengths to reinforce its cultural objectives through the use of appropriate language. Employees are cast members and customers are guests. Cast members don’t wear uniforms, they wear costumes. They are either onstage or backstage and stay in character accordingly. Every element of language provides a constant reminder to each cast member that he or she is there to ensure a quality experience for the guest. The terms and focus for a data center will differ considerably, but language that reinforces is no less important.

Heroes. Every organization has individuals who epitomize the values and norms of that culture. These individuals become local heroes, receive recognition and reward, and personify the values and behaviors that others in the organization are supposed to follow. If a data center culture is technology centered, its heroes are likely to be those with the greatest technical competence or expertise (i.e., a technical guru). If the culture is service centered, however, those who receive the highest recognition and reward will be those who exemplify responsive and creative service to customers. Heroes specifically define the desired values of a culture, and management largely determines who those heroes are.

Ceremony. The way in which meetings are conducted, greetings exchanged, organizational victories celebrated, and information transferred internally are all part of the ceremonies and rituals that reinforce culture. Ceremony is controlled primarily by management and can be used to define and support the desired service culture.

Reward system. Formal performance appraisals, promotions, and raises as well as informal gestures of praise and gratitude all need to be aligned with the basic desired service values of the culture. Too often, data center rewards vary from those promised upfront, creating a justified cynicism regarding the gap between what is said and what occurs. Data center professionals are unlikely to become more customer oriented if the major rewards and recognition continue to go to those representing values of technical excellence alone.

RESPECT AND SKILL

Two elements of a service culture are so crucial that they deserve special attention: respect and interpersonal skill. Few things shape perceptions of service so much as the respect that customers feel they receive in the moment-of-truth encounter. Almost everyone has experienced frustrating and unsatisfying service situations. And hostile and negative service encounters make customers feel they have been denied respect. Issues of power and control are important during any service encounter. If the server exercises dominance over the customer and treats him or her as a subordinate in the quest for solutions, perceptions of poor service inevitably follow. The server must have an attitude and relationship with the customer representative that acknowledges that they are respected partners in pursuing satisfactory results.

Likewise, several specific interpersonal skills are necessary from the server if the customer is to have a positive perception. Exhibit 1 illustrates these interpersonal skills. The ability to be respectful and interact effectively and consistently in service encounters is a function of both selection and training. The data center manager must carefully select data center agents who have the aptitude for these interpersonal skills and then thoroughly train their personnel in the practice of such abilities.

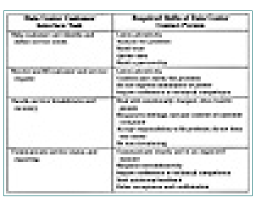


Exhibit 1 Required data center customer interpersonal skills.

The process of recovery from a service shortfall provides one of the most reliable tests of the level of respect and skill that actually exists in the data center. Everyone tries to avoid them, but service shortfalls inevitably occur. The skill used in dealing with recovery situations is a primary determinant of service-quality perceptions. In this situation, the customer is likely to be agitated and the feedback to the data center negative. If the customer's concerns are treated as points to be countered or debated, the stage is set for a confrontational exchange. If the negative feedback is seen as an opportunity for improvement, however, the impact is reversed. A data center that aggressively seeks feedback from customers — whether positive or negative — and uses that information to correct service shortfalls, exhibits the hallmarks of a service-driven organization.

SERVICE CULTURE VS. TECHNICAL

PROFESSIONALISM

IT units employ and are managed by individuals who are highly trained in their profession. Academic training and professional societies appropriately promote values of technical expertise, specialization, and knowledge as well as independent judgment — all characteristics that add competence and value to data center services. A strong data center promotes valuable outcomes and supports a belief that its professionals should function with considerable autonomy and act to maximize quality. In this type of environment, technical excellence itself can easily become the ultimate objective — that is, the data center becomes technology driven. A basic and inherent conflict exists, however, between professional training and orientation and service-culture requirements.

A vibrant service culture demands that customer needs be top priorities. Technology is useful only to the extent that it contributes to satisfying the customer. Professionalism is helpful if it aids in providing more-effective or less-expensive service. Formal professional training and experience, however, rarely emphasize the importance of service, nor do they provide the tools and perspective needed to skillfully identify customer needs or measure satisfaction. The successful data center service-driven culture must incorporate the positive aspects of professionalism with the dominant goal of satisfying customer needs effectively and economically.

ARTICULATING A DESIRED SERVICE CULTURE

Moving a data center toward a real service culture begins by carefully examining the dominant characteristics of the existing culture and explicitly articulating the desired culture. Exhibit 2 provides a sampling of factors that might be found in a current environment in relation to those that are desired.

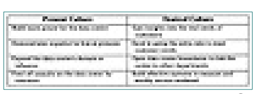


Exhibit 2 Identifying cultural values.

COMMIT TO SERVICE

With a strongly held commitment to move to a service-driven data center culture, and with a clear picture of exactly what that desired culture is, an organizational change plan can be constructed. This plan must identify specific indicators and measures of progress as well as key stakeholders who must be enrolled in the process. Developing a careful and comprehensive change plan will also help the data center avoid the lingering failure that a piecemeal approach engenders.

The new service-driven culture will dramatically increase data center productivity as time and attention formerly devoted to fruitless complaints and conflict can now be used to better identify and meet customer needs. In addition, it will improve internal morale and commitment as the data center moves from being seen as the problem to being valued as a respected partner in the pursuit of competitiveness for the overall organization. And finally, it will elevate quality ceilings for the entire organization as lack of data center service ceases to be an excuse for shortfalls in other areas.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter II-3
Cost-Effective Management Practices for the Data Center

Gilbert Held

Data center managers have the responsibility to maintain the highest possible level of economic efficiency in the data center. They must know the expenditures in the data center and its operations, and be aware of methods for reducing these costs. This chapter outlines the core set of activities and functions that represent the major expenses in operating a data center. For each activity and function, methods to control costs are presented. In addition, several examples illustrate the use of software productivity tools that can have a substantial effect on the ability to support internal and external customers, economize on the use of disk space, set up appointments and meetings in an orderly manner, and perform other efficiency-enhancing tasks.

MAJOR LINE-ITEM EXPENSES

The major line-item expenses associated with a typical data center are:

- Staff
    - Employees
    - Contractor personnel
- Hardware
    - Capital expenses

    - Noncapital and lease expenses
- Software
    - Purchases
    - Leases
- Maintenance
- Utilities
- Communications facilities
- Supplies

Although the expenditure for any line item and its percentage of total expenses varies among data centers, in aggregate these items represent between 90% and 95% of all data center expenditures. Data center managers should use the list of line items in planning methods to control their major expenses; these methods can mean large paybacks.

REDUCING STAFF COSTS

For most organizations, direct and indirect payroll expenses will significantly exceed all other costs associated with data center operations. Thus, methods for controlling staff costs can provide the manager with a significant level of success in regulating costs. Three of the more cost-effective methods managers should consider are a lights-out operation, downsizing, and the use of productivity tools.

Lights-Out Operation

The term *lights-out* describes an unattended data center, which of course implies an obvious reduction in staffing costs.

Most organizations using a lights-out data center restrict unattended operations to second and third shifts, when the need for manual intervention can be minimized through the use of such automation programs as job schedulers, task managers, and automated tape libraries. Even the few organizations that have implemented a third-shift lights-out data center must periodically run it with employees — drive heads need cleaning, printed output must be moved to the distribution center, and backup tapes must be transported to an off-site storage location. Nevertheless, utility programs can considerably reduce staffing needs, permitting a near lights-out operation to rapidly pay for itself through the avoidance of second- and third-shift premiums as well as a reduction in the number of primary-shift employees.

A wide selection of traditional mainframe automation software is available for purchase or lease. Such products include applications for job scheduling, disk management, and multiple-console support utility programs.

Job Schedulers. A job scheduler is a utility program that allows data center personnel to define criteria that the program then uses to queue jobs for execution. Some queues can be assigned a high priority to perform an immediate or near-immediate execution; placement in other queues may result in a deferred execution (e.g. in the evening). Job-scheduling software can perform a task previously performed by one or more data center employees.

Tape Management Software. This option provides a data center with volume-tracking capability, enabling tape librarians to more effectively manage hundreds, thousands, or tens of thousands of tapes and to locate and mount requested tapes much more rapidly.

Disk Management Software. This software enhances the data center's ability to allocate online storage to users and to back up data onto tapes or cartridges. In addition, some disk-management storage utilities include a feature that compresses data during a tape backup operation. For organizations with more than a few hundred tapes, disk-management software is used to compress onto one tape data that would usually require the storage capacity of three or more tapes. This saving can offset the cost of the program and possibly provide more payback.

Multiple-Console Support. A multiple-console support program provides large data centers with the ability to allow personnel to control computer operations from outside the center. Designated managers can be permitted to use their terminals or access the computer through a microcomputer, enabling them to assign priority to jobs and perform other functions.

LAN Administration. A lights-out operating philosophy is also applicable to the corporate local area network (LAN). Third-party programs are available to automate many key LAN functions, such as initiating file-backup operations of each server on the network at predefined times.

Downsizing

The four major expenses associated with staffing a data center are:

- the number of employees and pay rates;
- contractor support;
- overtime;
- shift operations and differential pay.

For most organizations, the number of employees directly supporting data center operations, and their salaries and benefits represent the largest staff expense. Some companies have outsourced all or a portion of data center operations either to reduce costs or to obtain an experienced pool of employees at locations where the hiring of data-processing personnel is difficult. As an alternative to outsourcing, downsizing data center operations is a practical method of controlling or reducing staffing expenses.

| Previous | Table of Contents | Next |

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

With a downsizing strategy, a company often encourages older employees to accept early retirement. A fraction of these employees are replaced by younger, lower-paid personnel. This method considerably reduces direct personnel expenses, but since many of those employees urged to retire early represent a pillar of knowledge and experience, the manager might not find it practical to lose this resource. Thus, managers should instead explore productivity tools, contractor support, overtime, and shift operations in the quest to control staffing expenses.

Productivity Tools. In addition to those automation tools previously mentioned to achieve a near lights-out operation, there are hundreds of programs for automating operations on platforms ranging from mainframes to LAN servers.

An automation tool replaces a repetitive process by automating its operation. As a result, employee productivity increases or fewer employees are needed to perform an activity. By comparison, a productivity tool helps an employee perform a job more efficiently but is not designed to actually replace a repetitive process.

The use of one or more productivity tools permits fewer organizational employees to work more efficiently and effectively. Although there are thousands of programs that can be classified as productivity tools, many, if not most, can be grouped into a core set of categories, including illustration or drawing programs, remote control programs, electronic mail and calendar scheduling programs, text retrieval programs, monitoring programs, and archiving and compression performing programs. The use of such programs can have a substantial effect on the ability of employees to perform their job assignments. The following section discusses the benefits to be obtained through the use of programs in each of these categories.

A remote control software program allows a LAN administrator to access and control networks at other corporate locations. One administrator can manage multiple networks at different geographical locations as if his or her personal computer were locally attached to each network. The productivity tool is used to extend the span of control of employees and provide a mechanism to more cost-effectively control distributed networks. Similarly, remote control software products, such as Hewlett-Packard's series of distributed systems (DS) control programs, allow an operator at a central site to control the operation of minicomputers at distributed data centers, thereby reducing the need to hire additional operational staff. In both these examples, the operations are random processes and the productivity tool extends the operational capability of an employee, but it does not replace the process.

Illustration or Drawing Program. One common activity performed in a data center involves the periodic creation of schematic diagrams for hardware, communications, and personnel. Through the use of an appropriate illustration or drawing program, the process can be significantly enhanced.

For example, Visio, a Windows-based program from Shapeware Corp., includes more than 20 built-in templates that facilitate the drawing of different objects and symbols. By selecting the network template, the symbols for a server and bus-based local area network can be moved into the program's drawing window. Using such a program can considerably reduce the time and effort involved in creating different types of diagrams. Thus, the use of Visio or a similar drawing program can substantially boost employee productivity in preparing different types of schematic diagrams.

Remote Control Programs. A variety of functions can be supported through the use of remote control software products, ranging in scope from LAN administration to enabling technical experts in a corporation at one location to test and troubleshoot different types of problems that might occur at another corporate location either around the block or thousands of miles away.

Electronic Mail and Calendar/Scheduling Programs. The use of a calendar/scheduling capability within an electronic mail program such as Novell GroupWise eases the establishment of meetings and conferences, as well as the dissemination of information. For example, if other users have access to your calendar, they can determine if you are available before attempting to schedule a meeting.

Text Location and Retrieval Programs. One common data center problem involves the location of information necessary to use different hardware and software products. In an IBM mainframe environment, it is often common for several offices to be used to store various programming and hardware-related manuals. Not only is this a waste of office space, but, in addition, it is difficult for more than one person at a time to effectively use the same manual.

Recognizing this problem, IBM and other hardware and software vendors now make their manuals available on CD-ROM. Although a single CD-ROM can store the equivalent of a stack of manuals 50 feet high, its utility is restricted unless it is placed on a LAN and users are provided with text retrieval software that facilitates its use.

Placing the CD-ROM on a LAN permits multiple users to simultaneously view different manuals contained on the disk, while appropriate text-retrieval software searches the contents of the disk. In this way, the contents of a 600-megabyte CD-ROM can be searched in seconds, permitting access to keyboard references that would be time-consuming or even impossible to retrieve when using paper manuals. Because many vendor programming and hardware-related manuals are conspicuous by their lack of an index, the use of text location and retrieval programs can enhance user productivity by more easily locating desired information.

Monitoring Programs. Data center managers can select from a wide range of monitoring programs to view the performance of computer hardware and network operations. Some programs, such as CPU monitors, provide information about the use of mainframe resources that can provide assistance in tuning an organization’s hardware to the work flow executed on it. For example, use of a monitoring program might help determine that delays in program execution are primarily the result of disk constraints. An upgrade of older 3380s to 3390 disk technology could therefore improve the flow of jobs through the mainframe.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

A second area where monitoring programs can be of assistance concerns the state of the corporate network. Through the use of the Simple Network Management Protocol (SNMP) and Remote Monitoring (RMON) probes, the activity occurring on remote network segments can be viewed from a central location. Consider the use of SimpleView, an SNMP RMON manager developed by Triticom Corp. for managing SNMP-compliant hardware and software devices. The addresses of an Ethernet and Token-Ring probe can be entered into the Network Map Window, with the Ethernet probe then being selected. Unlike some SNMP programs that require users to identify variables to be retrieved from a probe's management information base (MIB) by their numeric address in the global naming tree, SimpleView supports an MIB Walk feature in which the variable names associated with global naming tree entries are identified by their text description (e.g., sysObjectID). By simply selecting a variable by its name, information about the network management variable can be retrieved from the selected probe without having to know the numeric sequence identifier associated with the MIB entry.

Another valuable feature associated with SimpleView is its support of the SNMP GetNext command, which permits a user to retrieve the value of the next management variable after an initial location in the global naming tree is specified. Once the sysObject ID is selected from the MIB Walk window, use of subsequent GetNext commands displays the MIB Variable to be retrieved in terms of its position in the global naming tree.

Use of an SNMP management console thus enables the retrieval of information from different types of network probes. This in turn can allow trained personnel at one company location to view the operation of remote networks, possibly to examine the state of their health and initiate corrective action if a network-related problem should be noted.

Archiving and Compression-Performing Programs. Owing to the growth in client/server computing, a new function has been added to many corporate data centers — the development, testing, and distribution of programs and data files.

The distribution of programs and data files is facilitated through the use of an archiving and compression-performing program such as the popular PKZIP, ARJ, ARC, and similar programs. Any of these programs reduce the data storage required to distribute program and data files via diskette, magnetic tape or cartridge, removable disk, or any other type of physical media. If distribution occurs by telecommunications, the reduced storage achieved from archiving and compressing one or more programs and data files results in a reduction in communications time. Thus, using an archiving and compression-performing program can reduce both data storage and communications time associated with the distribution of programs and data files.

The user interface is important in selecting an appropriate archiving and compression-performing program, as it greatly affects the productivity of employees using it. Unfortunately, almost all archiving and compression-performing programs are command-based, requiring users to enter strings of codes to perform such operations as viewing the contents of an archive. One exception is WinZip, a Windows-based program developed by Nico Mak Computing.

Contractors. For most organizations, contractor support is used to add a level of expertise not available from in-house employees. Because most contractor employees are billed at a level 30% to 50% higher than employee salary levels, eliminating or minimizing the use of this resource can result in considerable savings. One way is to review the records of employees to determine whether they are capable of learning the skills performed by contractor personnel. If so, the manager could then plan to send employees to the appropriate vendor courses and attend public seminars to acquire the previously contractor-supplied skills. Once employees obtain adequate skill levels, they can work with contractor employees for a period of time to obtain on-the-job training, eventually phasing out contractor support completely.

Overtime. Controlling overtime and shift operations can also aid the data center manager in keeping data center operations within budget. Overtime, in many instances, gradually becomes an unnecessary expense. For example, overtime usually begins as a mechanism to perform a necessary operation, such as extending shift coverage to support the testing of new software or a similar function. Without appropriate controls, it becomes easy for employees to request and receive permission to use overtime to substitute for absent co-workers. All too often, overtime use expands significantly and can become the proverbial budget buster. Alternatives to overtime, such as compensatory time, should be considered.

Shift Operations. Shift operations not only require the use of additional employees but cause premium pay rates through shift differentials. Controlling shift operations can significantly control personnel costs. Possible methods of reducing shift operations include:

- obtaining additional processing and peripheral device capacity to finish jobs earlier;
- altering the job flow mixture to execute jobs that do not require operator intervention after the prime shift;
- requiring non-prime-shift jobs to be executed on specific days and at specific times.

HARDWARE COSTS

Over the life cycle of a computer system, its hardware cost typically represents the second or third largest expenditure of funds. Controlling the cost of hardware is a means for the data center manager to hold expenditures within budget.

Capital and Noncapital Expenses

Most organizations subdivide hardware expenditures into capital and noncapital expenses. Capital expenses must be depreciated over a period of years although paid for upon acquisition. Noncapital hardware expenses can be written off during the year the equipment is required. However, the current limit for this category of hardware is $17,500. Because most hardware expenditures exceed this limit, this chapter discusses hardware that is capitalized.

Several practical methods for controlling the cost of capitalized equipment include:

- obtaining plug-compatible peripherals to satisfy peripheral expansion requirements;
- using a leased equipment to minimize one-time expenditures of new additions;
- buying used equipment to satisfy some or all of the data center’s processing and peripheral expansion requirements;
- platform reduction.

The major risk in acquiring used equipment is the availability and cost of maintenance. If availability of maintenance is not a problem and cost is reasonable, used equipment can considerably reduce equipment costs.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Platform Reduction. Platform reduction refers to downsizing client/server technology as a replacement for conventional mainframe and minicomputer-based applications. A Pentium-based server is capable of operating at 50 millions of instructions per second (MIPS) for a fraction of the cost of an equivalent MIPS mainframe. Of course, previously developed applications must also be moved off the mainframe, which is not a trivial task. Organizations with extensive mainframe applications can require anywhere from 18 months to two or more years to move a series of mainframe applications to client/server technology. The additional expense of moving applications to a new platform includes a period in which many programmers are no longer available to enhance existing applications because their primary effort is focused on program conversion. The most successful platform reduction efforts are those that can identify an orderly migration of applications or the successful replacement of mainframe applications with client/server applications.

Example. An organization has an IBM 3090E supporting employees using Lotus 1-2-3/M, WordPerfect, a CICS application, and PROFS for electronic mail. The mainframe versions of 1-2-3 and WordPerfect can be replaced by equivalent and more-functional client /server products. PROFS can be replaced by 10 or more LAN e-mail programs. Thus, only the customized CICS application requires a conversion effect, and this conversion process is greatly simplified by the availability of CICS for Intel-based servers. An orderly migration to client/server technology can entail installing the network, training employees to use the LAN, loading the ready-made client/server programs, and providing any additional user training. The conversion of the customized CICS applications to run on the client/server platform would be planned to coincide with the end of this transition effort. Such planning minimizes the time required to move to a computing platform. In addition, such a plan would enable the organization to expediently cancel mainframe software licenses and maintenance service, which can range between $10,000 and $25,000 a month for a 3090E. An expeditious and orderly migration to a computing platform can have a considerable financial impact on the organization's bottom line.

SOFTWARE COSTS

Line-item expenses associated with software are generally subdivided into purchase and lease products. For example, the computer operating system and utilities may be available only on a lease basis, whereas application programs may be available for purchase only.

Leasing and Purchasing. In examining lease vs. purchasing software, the data center manager should be aware that many mainframe products are priced on a monthly lease basis. The expense of leasing for 30 months usually equals the cost of purchasing the product. Excluding the potential interest earned on the difference between the expense associated with an immediate purchase and the lease of software, the organization should generally purchase any software product that it anticipates using for more than 30 months, when it will begin to accrue savings. The savings cumulatively increase each month the product is used after this break-even point.

Multiple Processors. Multiple-processor data centers also can help control software costs. Because most computer manufacturers charge a monthly fee for each processor use site, based on the number and types of computers the processor operates on, reducing production installation to one or a few computers can provide savings.

For example, a data center with multiple IBM 3090 processors uses CICS and PROFS as well as other IBM software products. If multiple copies of each software product were licensed to run on each processor, the software licensing fee would significantly exceed the licensing fee charged for running CICS on one processor and PROFS on the second processor. In this case, it might be beneficial to establish a cross-domain network to enable users to access either processor from a common front-end processor channel attached to each 3090 computer system.

For client/server applications, the financial savings possible from obtaining LAN-compliant software rather than individual copies of a product can be considerable — a database program sold for $499 can be obtained as a network-compliant program capable of supporting 50 concurrent users for $1,995. The data center manager should examine the different types of PC software licenses and the cost of individual vs. LAN licenses.

MAINTENANCE

Although hardware is usually thought of as the main source of maintenance expenses, many software products have a maintenance cost to be considered. Expenses on hardware maintenance can be controlled through:

- third-party maintenance;

- conversion of on-site to on-call support;
- replacing old equipment with newer, more-sophisticated products.

Although savings from the use of third-party maintenance and the conversion of on-site to on-call maintenance support are relatively self-explanatory, the topic of reducing expenditures by replacing old equipment with more-sophisticated products requires some elaboration.

Most computer manufacturers guarantee the availability of spare parts for a specified time, such as five years from product introduction. After that time, spare parts may or may not be available from the original equipment manufacturer. If not, the organization may be forced to pay inflated prices for spare parts from third-party equipment vendors to prolong the life of aging hardware. In addition, monthly maintenance fees are usually based on the age of hardware, with maintenance costs rising as equipment life increases. At a certain point, it becomes more economical to purchase or lease newer equipment than to pay rising maintenance costs that can exceed the monthly lease and maintenance cost of the newer equipment.

UTILITIES AND CONSERVATION PRACTICES

Conservation can serve as a method for controlling the cost of data center operations. Two of the most practical utility costs over which an data center manager can have significant control are water and electricity.

If the data center processors use chilled water for cooling, the manager should stipulate that the water be recirculated. Not only does it reduce the organization’s water bill, it will also reduce the electric bill and possibly even the sewage bill. After chilled water is used for processor cooling, the water temperature is usually below that of natural water. Therefore, a smaller amount of energy is required to re-cool the used water than to chill new water. Because many cities and counties bill sewage charges by the amount of water consumed, recirculating chilled water also reduces water consumption, and, in turn, reduces the organization’s sewage bill.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Equipment Replacement

The cost of operating old equipment usually exceeds that of more-sophisticated products designed to use very large-scale integrated (VLSI) circuitry. For some data center configurations, including front-end processors, disk drives, and mainframes that might run nonstop throughout the year, the cost of electricity can be significant. By examining equipment for components that can consume an inordinate amount of energy, performing an operational analysis, and replacing equipment when justified, the data center manager can realize considerable operational cost savings.

Examples. A data center is located in a large metropolitan area such as New York City or Los Angeles, where the cost of electricity can easily exceed 10 cents per kilowatt hour (KWH).

Some front-end processors manufactured in the mid-1980s consume 15,000 watts per hour, whereas more-modern equipment using VLSI circuitry may require only 3,500 watts per hour. Because most front-end processors are run continuously except during upgrading or periodic maintenance, they can be estimated to consume power some 8,400 hours a year. If the front-end processor consumes 15,000 watts of electricity per hour, it will use 126,000 kilowatts of power (15,000 watts × 8,400 hours/year). At a cost of 10 cents per KHW, the cost of providing power to the processor becomes $12,600 during the year.

For the newer front-end processor that consumes 3,500 watts of electricity per hour, the yearly cost of power would be reduced to $2,940, or a decrease of $9,660. Over the typical five-year life of most front-end processors, the extra cost of electrical consumption of the older processor would exceed $48,000. This cost by itself would probably not justify acquiring new equipment. However, considered in tandem with other equipment, this alternative could provide a sufficient level of savings to justify the acquisition of new equipment.

COMMUNICATIONS

If the data center provides computational support for remote locations, the data center manager should consider analyzing its expenditures for voice and data communications. Depending on the number of remote locations and traffic volumes between those sites and the data center, the manager may find several methods to control the organization's communications cost. Two of the more practical methods are the development of a T1 backbone network and negotiation of a Tariff 12 agreement.

T1 Backbone Network. A T1 backbone network permits voice and data to share transmission on high-speed digital circuits that provide the capacity of 24 voice-grade channels at a monthly cost of approximately eight voice channels. If the organization's voice and data traffic requirements between two locations exceed that number, T1 circuits can significantly reduce the cost of communications.

Tariff 12 Negotiations. This tariff enables organizations with large communications requirements to negotiate special rates with communications carriers. According to vendors' information, discounts in excess of 50% have been obtained for multiyear contracts with a value of $5 million to $10 million per year. For organizations with communications requirements that do not justify Tariff 12 negotiations, savings may still be achieved by having each carrier provide service between the organizations' remote locations and its data center. In addition to the main carriers, the manager should also examine the cost of transmission facilities of alternative local access carriers, such as Metropolitan Fiber Systems of Chicago, which provides local access to long-distance carriers in more than 50 cities throughout the US. The choice of this alternative might permit the manager to reduce the cost of the organization's local and long-distance communications.

SUPPLIES

One of the most overlooked areas for controlling the cost of data center operations is consumable supplies. Most consumable supplies are purchased repeatedly throughout the year. The data center manager can reduce the cost of consumable supplies by grouping the requirements of several departments, consolidating expenditures on a monthly or quarterly basis to purchase larger quantities of needed supplies. This action typically results in a decrease in the unit cost per 100 disks, per cartons of 1,000 sheets of multipart paper and similar supplies.

Another technique is forecasting supply requirements for longer time periods. For example, purchasing multipart paper on a quarterly basis instead of each month may reduce both the per-carton cost of paper and the shipping charges.

CONCLUSION

This chapter has outlined several practical actions data center managers can consider to control data center costs. Although only some of these actions may be applicable to a particular data center, by considering each action as a separate entity, the data center manager is able to effectively consider the cost of many line items that cumulatively represent the major cost of operating a data center. This process is a helpful way to review current expenditures and serves as a guide to the various options for containing or reducing data center operations costs. In an era of budgetary constraints, such cost-control actions can only increase in importance to all managers.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter II-4
IT Outsourcing: Current Trends, Benefits, and Risks

Yvonne Lederer Antonucci
James J. Tucker, III

Initially, the search for cost-effective methods to acquire goods and services provided the impetus for outsourcing. However, the dimensions of outsourcing and the related benefits have since grown dramatically. This chapter presents the results of a survey of more than 1,200 firms, and offers insight into why managers outsource various functions to achieve both tactical and strategic benefits.

In recent years, there has been a total rethinking of organizational structures and business processes by management. This rethinking has resulted in a number of major management movements such as Process Reengineering, Just-in-Time, and Total Quality Management. These movements are responsible for nothing less than a revolution in business and have been the driving force behind firms' efforts to reconfigure their organizational structures (e.g., restructuring, downsizing, rightsizing, and organizational "flattening"), and to rethink how material, labor, and support services needed to feed business processes should be acquired. This latter rethinking is responsible for the dramatic increase in outsourcing.

The phenomenal growth of IT outsourcing has fundamentally reconfigured the IT landscape. Information systems and technology outsourcing can be defined as contracting with outside vendors for the provision of various IT functions, including data entry, data center operations, application maintenance and development, disaster recovery, and network management and operations. Vendors might be individual IT professionals, consulting firms, employee leasing companies, or full-service providers. Today, there is extreme variation in both size and type of computing services provided by outside vendors to the outsourcing of all facets of the IT function, both strategic and nonstrategic. These relationships are often referred to strategic alliances and partnering relationships.

TRENDS IN OUTSOURCING

Outsourcing of all types of services has experienced strong growth in recent years. The total U.S. market for outsourcing services will increase from the 1996 level of $100 billion to over $300 billion in the year 2001 — a 218% increase over this five-year period (see Exhibit 1).[1] A recent survey found that 51% of managers plan to increase expenditures for outsourcing, 49% plan no change, and only 2% plan to decrease expenditures.[2] Another survey of North American CIOs indicates that all IT areas are experiencing significant outsourcing-related activity for both large and small firms, ranging from 77% for disaster recovery to 23% for data entry (see Exhibit 2).[3] The U.S. market for IT outsourcing is expected to continue experiencing strong growth (see Exhibit 3).[1]

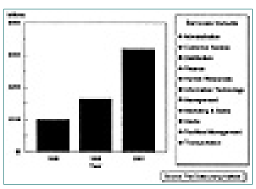


Exhibit 1 Total U.S. market for outsourcing services.

Exhibit 2 Percentage of firms planning to evaluate, planning to outsource, or currently outsourcing IT functions.

Exhibit 3 The U.S. market for IT outsourcing service.

Of the total U.S. market for outsourcing services, the growth rate of IT outsourcing, with projected revenues of $184 billion by the year 2001 (see Exhibit 4),[1] far outpaces other service areas. By contrast, the next-largest outsourcing service area, administration, has projected revenues of only $40 billion by the year 2001. The top three growth areas of individual IT functions will be applications development, applications maintenance, and desktop systems (see Exhibit 5).[1]

Exhibit 4 Total U.S. market for outsourcing services.

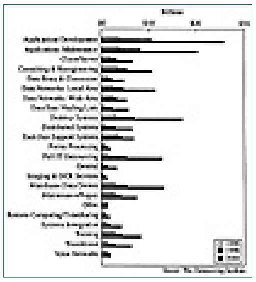


Exhibit 5 The U.S. market for outsourcing services.

Outsourcing in one form or another has existed almost from the dawn of data processing. In the mid-1960s, computer-services bureaus ran a variety of programs, including general ledger, payroll, inventory control, etc. The customers of these service bureaus were mostly small and medium-size firms, although some large firms used them for specialized needs or highly confidential items such as executive payroll.

In mid-size and large organizations, IT outsourcing prior to 1990 was significant but largely peripheral to the main IT activities that took place internally. During this era, the major drivers for IT outsourcing were cost-effective access to specialized or occasionally needed computing power or systems-development skills, avoidance of building in-house IT skills and skill sets (primarily an issue for small and very low-technology organizations), and access to special functional capabilities.[5]

Caveat Emptor

Before the advantages and disadvantages of IT outsourcing are examined, a word of caution is in order. Both proponents and critics of outsourcing offer numerous reasons to support their positions. Not surprisingly, information provided by individuals and organizations that benefit by outsourcing often focuses only on the benefits of this practice. Likewise, information provided by individuals or organizations that are negatively affected by outsourcing tends to reflect negatively on this practice. Therefore, in any analysis of outsourcing or insourcing, it is most important that the reader carefully consider the incentives of those providing the information.

WHY MANAGERS OUTSOURCE

Since the late 1980s, the general use of outsourcing has become a megatrend for a wide variety of goods and services. Initially, the search for more cost-effective methods to acquire goods and services provided the impetus for outsourcing. However, the dimensions of outsourcing and the related perceived benefits have grown dramatically. A survey of more than 1,200 firms provides insight into why managers outsource various functions to achieve both short-term tactical benefits and long-term strategic benefits. These reasons for outsourcing are presented below.[1]

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

The top five short-term tactical reasons for outsourcing are:

1. **Reduce or control operating costs —** The single most important tactical reason for outsourcing is to reduce or control operating costs. Access to the outside provider's lower cost structure is one of the most compelling short-term benefits of outsourcing. In a recent Outsourcing Institute survey, companies reported that on average they saw a 9% reduction in costs through outsourcing.
2. **Make capital funds available —** Outsourcing reduces the need to invest capital in non-core business functions. This makes capital funds more available for core areas. Outsourcing can also improve certain financial measurements of the firm by eliminating the need to show return on equity from capital investments in non-core areas.
3. **Cash infusion —** Outsourcing can involve the transfer of assets from the customer to the provider. Equipment, facilities, vehicles, and licenses used in the current operations all have a value and are, in effect, sold to the provider as if the transaction were resulting in a cash payment.
4. **Resources not available internally —** Companies outsource because they do not have access to the required resources within. For example, if an organization is expanding its operations, especially into a new geographic area, outsourcing is a viable and important alternative to building the needed capability from the ground up.
5. **Function difficult to manage or out of control —** Outsourcing is certainly one option for addressing these types of problems. Outsourcing does not, however, mean abdication of management responsibility, nor does it work well as a knee-jerk reaction by companies in trouble.

The top five long-term strategic reasons for outsourcing are:

1. **Improve business focus —** Outsourcing allows the company to focus on broader business issues while having operational details assumed by an outside expert. For many companies, the single most compelling reason for outsourcing is that several of the "how" type of issues are siphoning off huge amounts of management's resources and attention.
2. **Access to world-class capabilities —** By the very nature of their specialization, outsourcing providers bring extensive worldwide, world-class resources to meeting the needs of their customers. According to Norris Overton, vice president of reengineering, AMTRAK, partnering with an organization with world-class capabilities can offer access to new technology, tolls, and techniques that the organization may not currently possess: better career opportunities for personnel who transition to the outsourcing provider, more structured methodologies, procedures, and documentation; competitive advantage through expanded skills.
3. **Accelerated reengineering benefits —** Outsourcing is often a byproduct of another powerful management tool — business process reengineering. It allows an organization to immediately realize the anticipated benefits of reengineering by having an outside organization — one that is already reengineered to world-class standards — take over the process.
4. **Shared risks —** There are tremendous risks associated with the investments an organization makes. When companies outsource, they become more flexible, more dynamic, and better able to adapt to changing opportunities.
5. **Free resources for other purposes —** Every organization has limits on the resources available to it. Outsourcing permits an organization to redirect its resources from non-core activities toward activities that have a greater return in serving the customer.

BENEFITS

Focus on Core Competencies

As the environment becomes increasingly competitive and uncertain, many firms have focused on their core competencies as a way to gain competitive advantage and add unique value to their customers. Proponents argue that management should not squander valuable time worrying about support operations when it could be devoting its top talents to nurturing, building, and leveraging core competencies. Here, IT is viewed as important but not of core importance to the business and is more efficiently managed by outsourcing vendors who are better equipped than internal staff to keep pace with new technologies, skills, and processes. As one CIO comments: "As a consumer products company, we're in very competitive markets. We needed to unlock a lot of information that was trapped in our mainframes …. If we tried to do the same thing with our internal staff, we'd be shackled with a lot of training and not have any time to focus on the job."[6,7]

Access to State-of-the-Art Technology

The volatility of information technology has the effect of rapidly making IT skills obsolete and creating IT skills shortages. Because IT evolves so rapidly, by the time a firm invests in and trains its IT staff on a certain technology, that technology may be obsolete. Under these circumstances, retaining a permanent workforce may be cost prohibitive. (For training costs as a percentage of the IT budget, see Exhibit 6). As one CIO noted: "When we need special technical skills, or skills in a different part of the world, it's easier to turn to [an outsourcing] firm than to do your own hiring. We don't have the responsibility or the headache of training in new technological environments and new tool sets."[4,7]

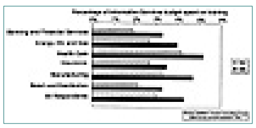


Exhibit 6 Training as a portion of information services budget.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Cost Savings

Fierce competition has led many firms to embark on massive cost-reduction efforts, including the reduction of staffing levels. Today, even thriving organizations strive for continuous reduction of costs and staffing. The plausibility of outsourcing IT functions was first considered by most firms as a result of corporate cost-cutting campaigns. Proponents cite a number of reasons why outsourcers may be able to provide comparable IT services at a significant cost savings:[5,8]

1. Outsourcing introduces market competition and increases economies of scale.

2. Outsourcers usually have much tighter overhead control of fringe benefits and run much leaner overhead structures than their clients.

3. Outsourcers are more aggressive in using low-cost labor pools and are more willing to move data centers to low-cost areas.

4. Outsourcers can better manage excess hardware capacity by selling under-used hardware or combining many clients’ work in the same operations center.

5. Outsourcers have the ability to operate a leaner management structure because of increased competence and critical mass volumes of work.

6. Outsourcers are very creative and realistic in structuring leases.

Flexibility

Organizations face an increased need for flexibility to adapt to a rapidly changing environment. This calls for an IT function that can respond quickly to changing demands. Proponents contend that this flexibility can be achieved by outsourcing, which allows firms to tap a wide range of resources, skills, and capacities provided by the vendor. In comparison, the capabilities of an internal IT function are limited by the capabilities of the current IS staff and technology.[8]

Protect the Firm's Image and Increase Job Security for Regular Employees

If outsourcing is used, workers are often hired with the explicit understanding that their employment will be for a limited duration. Therefore, they can more easily be dropped or added to the work force without jeopardizing the reputation of the firm as an unstable employer. More importantly, the use of outsourced workers buffers regular employees from fluctuations in demand and enables the firm to establish a stronger relationship with its regular work force than would otherwise be possible.[4]

VULNERABILITY OF INTERNAL IT: IMAGE PROBLEMS

Ever-increasing IT budgets (see Exhibit 7) and doubts about the resulting benefits have led to questions regarding the legitimacy of investments in information technology. A recent study revealed that a majority of senior managers viewed their IT functions as cost burdens rather than as strategic resources. Some internal IT departments are perceived to be outdated, inflexible, expensive, unmanageable, and lacking a customer orientation.[8,9]

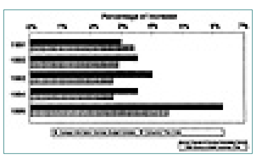


Exhibit 7 Annual increase in information services budget vs. the inflation rate.

This negative image, whether warranted or not, often acts to initiate discussions concerning outsourcing. Here, outsourcing is viewed as a way to solve problems related to IT management and control. When severe conflict exists between the CIO and the CEO or other top management officials, IT outsourcing may be employed to break the power of the central internal IT department. Outsourcing, accompanied by downsizing of hardware and increased end-user computing, may be a symptom of an IT legitimacy crisis and the resulting struggle for power over the IT function.

Given the negative image of IT often held by top management, it is not surprising that IT outsourcing has experience such growth. Nevertheless, there currently exists no conclusive theoretical or empirical support for the assumption that outsourcing will always lead to more-focused organizations, higher flexibility, lower costs and staffing levels, economies of scale, or to the solution of all problems with internal IT departments.[8]

RISKS

With all the media attention focused on the projected benefits of major IT outsourcing deals, a number of questions emerge. “Is IT outsourcing really as effective as proponents say it is?” “What are the risks, disadvantages, hidden costs?” To address these questions, an examination of a number of surveys, case studies, and other research provides insight into these issues.

All that Glitters ...”— Media Hype and Outsourcing Benefits

One study concluded that the overall lesson learned from scrutinizing the business-media coverage of IT outsourcing is that media coverage often portrays an overly optimistic view of outsourcing. First, reports are often made during the honeymoon period when clients first sign an outsourcing contract. At this point, the client and vendor possess high outsourcing expectations. Second, the media only reports projected savings made at the first press conference rather than actual savings. Public sources neglect to report that some outsourcing clients are charged exorbitant excess fees for modifications to the contract. Third, media coverage under-represents outsourcing failures because few companies wish to advertise a mistake. By focusing only on success stories, the media misrepresents the spectrum of outsourcing experiences.[9] A recent survey of North American CIOs reveals that, in contrast to the usual media coverage, the actual benefits of IT outsourcing fell short of expected benefits (see Exhibit 8).[3]

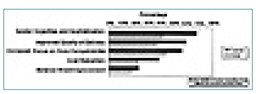


Exhibit 8 Expected vs. actual benefits of IT outsourcing.

Previous | Table of Contents | Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

IT: a "Commodity" that Can Be Easily Outsourced?

Some argue that IT is not like other organizational resources that have been successfully outsourced in the past. In particular, the following key characteristics associated with IT make it different from other outsourcing candidates:[9]

1. **IT permeates the entire organization** — Senior executives often question why the outsourcing of IT is any different from the outsourcing of other non-core business functions such as legal, administrative, and custodial services. Some CIOs argue that the IT function is fundamentally different from other functions frequently outsourced, since IT is not a process at the beginning or the end of an organization's production chain, but an integrated part of it.
   Unlike many other products and services, IT cannot be easily isolated from other organizational functions. IT permeates every business function in the value chain, as well as all support activities. IT is not a homogeneous function, but comprises a wide variety of IT applications that are integrated parts of almost all processes within an organization, and influence directly how these processes are performed. This is especially the case for information systems that support primary processes, and even more so if these primary processes are essentially information processing. As a result, IT outsourcing cannot be compared to outsourcing of catering, security, logistics, legal services, advertising, or the procurement of raw materials and components.
   When firms in the study tried to isolate IT activities for outsourcing, they often discovered that changes to the outsourced function affect other areas of the business. The vendors typically lacked the specialized

knowledge to cope with organizational interfaces outside the boundaries of the IT outsourcing contract.

2. **Information technology evolves rapidly —** Because IT evolves so rapidly, predicting beyond a three-year horizon is highly speculative at best. Hence, IT outsourcing decisions — particularly decisions that result in long-term contracts — are always accompanied by a high degree of uncertainty. Although participants initially perceived that outsourcing vendors would reduce the uncertainty associated with rapid IT evolution, many discovered that outsourcing actually locks them into older technologies and may act as an inhibitor to the adoption of new technology. Indeed, it is precisely because of the dizzying pace of IT change that outsourcing contracts typically only make nonspecific or vague reference to future technologies.

3. **The underlying economics of IT changes rapidly —** Although price/performance improvements occur in every industry, there can be few industries where the underlying economics shift as fast as in IT. A mainframe MIP that cost $1 million in 1965 costs less that $30,000 today. Today's computing resources will cost 20% to 30% less next year. The rapid change in the underlying economics causes extreme uncertainty and thus makes it difficult for decision makers to evaluate costs of outsourcing bids. While a 10% discount on IT for the next 10 years may be appealing to a senior executive today, by the next year he may be paying the vendor above-market prices for computer resources. Fundamentally, the problem appears to be a lack of a simple basis for gauging the economics of IT activity.

4. **The costs of switching to alternative ITs and IS suppliers is high —** As a result of increased competition in the IS market, there has been a shakeout among IT vendors with mergers and takeovers becoming commonplace. Mergers have led to the outsourcing market's being concentrated in the hands of a relatively small number of IT suppliers that are interrelated in many ways. In the U.S., 56% of the outsourcing market is held by seven companies.[8,9]

Loss of Control

Critics of IT outsourcing argue that no outside vendor can match the responsiveness and service levels offered by an in-house function, largely because the outsider is not subject to the same management direction and control as employees. In addition, there are often concerns about the confidentiality of data and strategic applications, and about provisions for disaster recovery with an outside vendor. These anxieties about control might be summed up under the rubric "nobody else cares about our business the way we do."[10]

Less Flexibility

Some CIOs argue that outsourcing is likely to remove, or at least significantly constrain, any flexibility in the delivery of IT. The outsourcing vendor will provide the level of IT services specified in the contract using the technological platform it deems appropriate. In such an environment, vendors typically run mainframes and mainframe software applications "into the ground," as they are hidden costs. Unless specifically spelled out in the contract, the flexibility of moving to new computing platforms may be lost. Or if not lost, the vendor is likely to be the sole benefactor of any cost advantage of moving to new, more cost-effective technology, such as client/server computing. Of more concern, however, is that outsourcing could constrain how an organization reacts or adapts to a changing business environment. Without the flexibility of having an internal IT capability, it might prove difficult (and/or expensive) to have the vendor provide a changed set of IT services.[9]

Hostage

IT professionals argue that outsourcing often results in the user's becoming a hostage of the vendor. The organization often will have given up certain technical skills, transferred hardware and software to the vendor, and potentially become locked into the vendor's proprietary software and hardware. Moreover, the longer the contract and the larger its scope, the more leverage the customer has in negotiations, but the more leverage the vendor has after the outsourcing is under way.[10]

Once IT functions have been turned over to an outsider, it becomes extremely difficult and costly to bring them back inside again. For example, a recent newsletter reported General Motors is in the "worst possible situation [of] bringing outsourced functions back in-house." General Motors is in the process of attempting to replace its IT outsourcing vendor. CIO Ralph Szygenda is trying to hire dozens of subordinate CIOs and also terminate the outsourcing contract early.[11]

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Cost Savings?

Many managers assume that outsourcing vendors are inherently more efficient due to economies of scale. The theory of economies of scale states that large-sized companies achieve lower average costs than small companies through mass production and labor specialization efficiencies. In the outsourcing arena, however, the applicability of the economies-of-scale model can be questioned. First, small shops may have lower costs per MIP than large shops by employing older technology, offering below-market wages, and maintaining tight controls and procedures. Second, the difference between hardware discounts granted to vendors vs. individual companies is often negligible. Third, changes in software licensing agreements diminish a vendor's advantage. Finally, labor expertise is largely a myth, since clients are usually supported by the same staff that were transferred to the vendor. Purchasing new expertise from the vendor may be costly.

When vendors submit bids that indicate savings, companies may question whether they can achieve similar results without vendor assistance. If the vendor is not inherently more efficient, perhaps the company can reduce its own IS expenses through data center consolidation, resource optimization, and chargeback implementation, which can also dramatically reduce IS costs.[9]

Subcontractors: Surprises and Problems

Companies that outsource are often unpleasantly surprised to find their vendor is not working on their projects — someone else is. Outsourcing vendors in search of hard-to-find technical skills often subcontract portions of their computer-system work to small, often unknown companies — and they sometimes do it without the knowledge of the IT managers who are their clients. In fact, a recent survey found an average of 36% of an outsourcing contract and 25% of systems integration contracts involve subcontractors. Along with these contracts can come problems, including viruses brought in by outsiders, poor communication, high costs, and low-quality service.[12]

The survey also reveals that subcontracting is indeed common practice — and the trouble it can cause is just as common. Subcontractors are employed by prime IT services vendors at more than 50% of responding organizations. While 65% reserve some management rights with regard to subcontractors, more than half of the survey respondents report that subcontractors caused problems. Service quality, at 67%, headed the list of woes cited, followed by costs (30%), viruses (17%), and security (10%). In addition, 40% of the respondents cited other problems with outsourcing contractors, including substandard technical abilities, poor communication, lack of documentation, sluggish response times, and late delivery.[12]

CONCLUSION

The evidence presented early in the study suggests that the outsourcing of major IT functions will continue to grow at a rapid pace into the foreseeable future. The continued strong growth in both the depth and breadth of IT outsourcing suggests that this management practice is more than just a passing fad, and under the right circumstances, IT outsourcing will provide at least some of the advantages noted by proponents. These benefits include increased ability to focus on core competencies, access to state-of-the-art technology, increased flexibility, and cost savings. For firms that have successfully outsourced various IT functions, the question is not, “Should we outsource?” but rather, “How much should we outsource?”

The study also suggests that the media often presents an overly optimistic portrayal of IT outsourcing, especially when outsourcing deals are initially announced. In addition, IT outsourcing is not a panacea for all IT problems, and in some situations, it may create as many problems as it is intended to resolve. Problems noted by critics include loss of control, less flexibility, negligible or questionable cost savings, and being “held hostage” by the vendor and subcontractors.

Given the numerous benefits and risks of IT outsourcing, IT managers would be well advised to carefully and cogently examine these issues when evaluating proposals to outsource various IT functions and when conducting periodic assessments regarding the success or failure of existing outsourcing relationships.

Notes

1. Information provided by The Outsourcing Institute, New York, 1997.

2. *The Outsourcing Index*, The Outsourcing Institute and Dunn & Bradstreet, December 1996, New York.

3. *Leading Trends in Information Services*, Deloitte & Touche LLP, 1996.

4. "Information Systems Outsourcing," H. Rao, K. Nam, & A. Chaudhury, *Communications of the ACM*, July 1996.

5. "How to Manage an IT Outsourcing Alliance," F. McFarlan and R. Nolan, *Sloan Management Review*, Winter 1995.

6. "Technology Outsourcing," V. Gurbaxani, *Communications of the ACM*, July 1996.

7. *Outsourcing For Impact — Partnering for Excellence*, Andersen Consulting, 1996.

8. *Information Systems Outsourcing Decision Making: A Managerial Approach*, L.A. de Looff, Idea Group Publishing (Hershey, PA), 1997.

9. *Beyond The Information Systems Outsourcing Bandwagon*, Mary Lacity and Rudy Hirscheim, John Wiley & Sons (New York), 1995.

10. "IT Outsourcing: Operational Improvement or Strategic Imperative?", Management Support Technology Corporation, 1997.

11. "Rethinking the Outsourcing Decision," *IT Cost Management Strategies* — Letter, March 1997.

12. "Hidden Partners, Hidden Dangers," *Informationweek*, January 20, 1997.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter II-5
How to Select an Outsourcing Vendor and Negotiate Terms

Douglas B. Hoyt

The outsourcing of computer operations work has proven to be an important means for many organizations to help them become or stay competitive today. With new IT technologies, global competition, growing availability of competent services vendors, and re-engineering efforts to redefine and improve traditional business processes, outsourcing has become an increasingly prevalent way to fulfill business improvement opportunities.

THE CHALLENGE

It has been estimated that 73% of the leading companies contract out part of their IS functions, that $6 billion of client/server work will have been outsourced in 1996, and that 80% of information-systems activities will possibly be done by contractors by 2000. Because organizations are finding that outsourcing can cut costs, sharpen management's focus, get access to expertise, and foster global expansion, computer operations managers should help lead the way to finding improvement opportunities via IT outsourcing. When outsourcing has been decided on by a company's administration, a manager must help by securing the most appropriate vendors and laying the groundwork for a mutually beneficial relationship to implement those decisions so they work to the organization's best advantage.

One or More Vendors?

Most organizations that outsource some computer work use several vendors. One reason is to select the best vendor for each function specialty. A second reason is to not "put all your eggs in one basket;" sometimes it is better to have more than one vendor for a single function, for comparison purposes and to protect against total disaster if one fails. Having several vendors can prove to be an advantage when it is decided to change vendors or return the work in-house, an eventuality that must be considered in every outsourcing action.

Having one vendor do everything that is outsourced simplifies the management of the work outsourced, and is fine so long as the vendor proves competent and reliable. But many outsourcing arrangements have their rocky periods, and their termination must be anticipated in all cases.

SOURCES OF OUTSOURCING VENDORS

Many of the larger and better-known outsourcing vendors are known to most computer operations managers from discussions with peers and reports in the press and magazines. Some vendors may have approached the managers' organizations in efforts to sell their services. Word of mouth and general-knowledge sources can be valuable, especially when they reflect opinions of results the vendors have achieved from people or media in which the prospective client has trust.

However, even though a worthy source might be familiar from prior knowledge, it is usually wise to look at other alternatives to make sure the vendor selected is the one best suited for the job to be done. Some sources to consider in seeking the best service vendors are The Outsourcing Institute, research services, consultants, and computer resellers.

The Outsourcing Institute

The Outsourcing Institute Inc. (45 Rockefeller Plaza, Suite 2000, New York, NY 10111, 800-421-6767) is a for-profit membership group to which both outsourcing users and vendors belong. Membership is free. One of its services to members is its Buyer's Guide, a list of outsourcing vendors detailing the services they provide and industries and geographical areas they cover.

Research Services

Dataquest Professional Service Trends and The Gartner Group publish reviews and reports on potential vendors. It has been said that such information should be sufficient for projects of $100,000 or less, and that consultants are suggested for advice on more sizable work to be contracted out.

Consultants

Consultants, especially consulting arms of large public accounting firms, have been engaged to advise and counsel on whether and what to contract out, vendors to select, and how to negotiate and manage outsourced work. Firms like Andersen, Price Waterhouse, and Coopers & Lybrand provide such advisory assistance, though their fees range from $100 per hour to much higher.

The leading consulting associations can also refer inquirers to members that provide outsourcing vendor advisory services. These include:

- Institute of Management Consultants (IMC), 521 Fifth Avenue, 35th floor, New York, New York 10175. 212-697-8262
- The Association of Management Consulting Firms (ACME), 521 Fifth Avenue, 35th floor, New York, New York 10175. 212-697-9693
- Independent Computer Consultants Association (ICCA), 11131 South Towne Square, Suite F, St. Louis, Missouri 63123. 800-774-4222

Consultants' capabilities can also be researched through the Internet via Management Consultant Network International, Inc. (MCNI) be reached via http://www.mcninet.com/ or at 413-755-0825.

Resellers

Many resellers of hardware and software also provide outsourcing services for a variety of computer operations, usually related to the products they sell. When a reseller does offer to perform outsourcing work, it will be extra motivated to please its customer, since its bread-and-butter profits depend on the sale of computer products to the same customer. Therefore, resellers are an important outsourcing option to consider.

HOW TO EVALUATE VENDORS

The final decision in selecting an outsourcing vendor may well be an intuition judgment, such as a belief that these are reliable, honest people who would be pleasant to work with. A feeling of mutual respect and trust based on such opinions can be important to a successful, continuing business relationship.

Nevertheless, the decision should be made only after a thorough analysis of the many characteristics that are essential to the success of contracted work. This analysis must include feedback from the vendor's other customers, and verification of the vendor's expertise and experience, financial stability, and reliability. Checks should be made into any potential conflicts of interest with the firm being considered. It is good to get at least two proposals and compare them, based on a carefully written Request for Proposal (RFP), or in a statement of the work that is to be required.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Checking References

The best reference is a referral from someone who is known personally and trusted, and who has experience with the vendor. Lacking an unbiased recommendation, it is well to discuss the vendor with some of its present or past clients, meet with them, and talk over in some detail the work done, how it met expectations, the vendor's technical expertise, resourcefulness, reliability, and related matters. It is important to make sure the clients met with are representative and random, not selected favorites of the vendor.

Appraising Expertise and Experience

One major reason for wanting to contract out some computer work is usually that vendors have greater expertise and experience than the client in the function involved. Therefore, it is essential to evaluate and verify the nature and degree of these factors in the vendors being considered. The sources of research into this area are many — vendors' literature, published articles, discussions with potential vendor representatives, consultants, and, most important, information from the vendors' past and current clients.

When appropriate, it may be wise to ask the potential vendors for résumés of the persons who would do the work or be in charge, and evaluate their qualifications in the specializations required. The presence of certifications can be significant, such as the ISO 9000 series certification, demonstrating quality-control discipline, and CCP, Certified Computer Professional.

Financial Stability

The client must be assured that any outsourcing vendor selected will be around long enough to complete the contract, usually a matter of from one to 10 years, and also be available to continue if it is desired to renew the contract. The vendor should be comfortably profitable, so as not to be influenced to cut corners to survive.

If the potential vendor is a public company, review its financial statements. Secure credit reports from credit agencies like Dun & Bradstreet and Moody's. Look back some years, as enduring years of success are an indication of likely future success at surviving and performing. Review published reports in periodicals like *The Wall Street Journal* and *Business Week* for any significant events such as legal or labor problems in the vendor's history.

Size

Values related to a vendor's size should be evaluated. Smaller firms might be more flexible, cheaper, more eager to please, but are less able to assure long-term survival. A one-man firm could be risky if the function is a critical one. Large firms indicate strength by having achieved that size, have greater resources, but may be more rigid and demanding in their relationships with clients. These are all matters to consider along with the nature of the job to be done.

Conflicts of Interest

An outsourcing firm serving competing clients can be inappropriate where the revelation of proprietary information is at stake. For example, Electronic Data Services (EDS), a subsidiary of General Motors until recently and now a key outsourcing vendor to GM, would not be asked to be a vendor for Ford or Chrysler. However, the serving of competitors can be acceptable and be of benefit to all, especially when the cost of developing a specialized expertise can be spread among many clients and the protection of proprietary information is assured; examples are Automated Data Processing (ADP), which provides payroll service for thousands of clients, and Hyatt, which provides reservation systems know-how to its competitors.

Fair Prices

The relative prices of competitive vendors is often a key factor. However, it sometimes is difficult to compare prices because vendors have different price structures, Also, there are many approaches to incorporating incentive and flexible pricing systems, and that nature of these systems is also one part of the price-evaluation analysis. The price-structure alternatives are reviewed below.

THE REQUEST FOR PROPOSAL

Major outsourcing contracts are often based on a Request for Proposal (RFP) sent to potential vendors. If an RFP is not used, it is essential for the client to prepare a statement of the work to be contracted out, including requirements such as functions to be performed, turnaround time, location, and equipment. All those who have worked with outsourcing arrangements agree that a principal ingredient for success is to have in advance a well-thought-out, complete statement of the work to be done and related requirements. The statement should specify results expected, in measurable terms where possible, e.g., cycle time, volumes, or costs. This statement ensures that the client has fully analyzed what is wanted and that the vendor has a complete and correct understanding of what is expected.

Whether an RFP or work description is used, the prospective vendor's competence should be evaluated by analyzing the qualities of the potential vendor's proposal, its promptness, clarity, and completeness, qualities that will be desired in the work to be done under the contract.

TERMS TO BUILD INTO CONTRACT

Books have been written on contract terms and negotiation methodologies. The main features of contracts are reviewed here with comments about options that are available. The important thing about the contract is to look at it as a means for the client and vendor to discuss and come to a clear and common agreement about what is to be done, how the process will work, and the pricing. The contract is the foundation of the relationship, which must be mutually beneficial if it is to be successful. Nothing can damage a relationship more than differences in the understanding of what is expected.

Requirements Statement

This section covers the nature and details of the work to be done, cycle times, where it will be done, programs and equipment to be used, and related matters. This is normally a repeat of the requirements statement given to the prospective vendor as part of the exploratory process, possibly revised or refined based on discussions with the vendor. One feature that is important to include is the specific measures the vendor will take to ensure prompt recovery in the event of a disaster, including backup for data and facilities. Another feature to include is the methods the client plans to use to monitor the performance of the vendor during the contract, and the standards and criteria that will be applied in determining the acceptability of the results.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Term or Duration

For a contract to maintain an organization's computer hardware, a term of one or two years may be appropriate. In such contracts, it would make for minimum disruption to change from one vendor to another. But for a contract in which an outsourcing vendor is going to buy much or all of the client's computer hardware, hire maybe 100 of the client's employees, and perform some of the client's basic computer operations, many such contracts are made for terms of seven to 10 years or more. These major contracts would involve significant problems for the client as well as for the vendor if they are terminated or changed in a few years.

Some recommend keeping the periods short to allow for flexibility and changes. The client's needs may change with new products, new markets, new competitors. Provisions must be made in the contract to facilitate these types of new conditions. Since hardware or software may develop in a year or two that could enable the vendor to cut its costs significantly, provision must be made to permit the client to receive at least part of that benefit. So, if the period of tenure is lengthy, terms must anticipate the need for flexibility to adapt to changes properly and fairly for the client and vendor. Finally, short contracts give the vendors more incentive to perform well in order to secure renewals.

Pricing and Incentives

If a major part of the client's purpose in outsourcing is to reduce or control costs, make sure that the contract terms will provide the means to measure and fulfill the accomplishment of that objective.

A simple price per unit of work basis is fine for short term arrangements where there is a uniformity of the tasks to be performed. However, for longer term or more complex agreements, it is generally desirable to incorporate incentives based on performance, as well as provisions for renegotiating when conditions may change. Incentives can be made to reward or penalize the vendor for variations in measurable results, such as speed of a cycle time and percent of defects. With such terms, the vendor can share the risks and rewards of poor or exemplary performance.

An example of revenue sharing is a 10-year agreement with Perot Systems Corp. to provide client/server outsourcing for Europcar International. After the second year, when a distributed system is established, the contract provides that Europcar will pay Perot a share of its revenue.

The Flexibility to Make Amendments

In long-term contracts, to deal with possible changes in the assumed conditions on which the original terms were based, it is a good idea to incorporate provisions for renegotiation of the price and other terms. Changes in technology, products introduced, entering global markets, disaster crises, and unanticipated conditions could alter the client's needs as to what is outsourced and how it is done. Therefore, such unforeseen events must be provided for in the contract terms so that the client can make changes as new conditions require. Vendors might initially balk at giving contract terms with this flexibility, but will usually consent to them to maintain a good relationship and remain competitive.

Non-Compete Terms

In situations where the vendor is in a business similar or possibly the same as the client, the client should insert a contract provision that the vendor will not do business with the client's clients for a specified period of time such as two or five years, or possibly in certain territories. Through their work, outsourcing vendors often have access to the clients' customer and potential customer lists, as well as to information about the clients' product features or services methods that give the client a competitive advantage. The noncompete agreement terms prevent the outsourcing firm from using such information to compete with its client for a specified period. It should also state that the vendor will not recruit any of the client's employees without prior written consent.

Nondisclosure Terms

When the vendor is in a position to learn from the client information about systems and methods, products, or services that give the client a competitive edge, the contract should contain terms that prevent the vendor from disclosing that proprietary and technical information to third parties. The provision can state examples of restricted information, such as lists of customers and potential customers, marketing plans, computer programs, documentation, and manuals. It might also specify that the vendor cannot reverse-engineer or decompile any of the client's software.

Ownership of Innovations

When the work to be outsourced includes the design and development of new systems, the contract must clearly specify who owns the new system and programs, and what restrictions apply to their sale or use. The new programs logically belong to the client because it paid for them. Equally as logical is that the vendor owns them because it did the work. These arrangements must be negotiated, as the programs often have considerable value, and could have an important influence on the client's competitive position if they were to be obtained by a competitor.

Use of Licensed or Purchased Software

Purchased and licensed software terms normally specify that the purchaser or licensee may not make the software available to other organizations without the vendor's permission. Therefore, if any such software is to be used by the outsourcing vendor, permission must be obtained from the software provider. If the permission is granted, the software vendor will insist that similar restrictions apply to the outsourcing vendor. Those restrictions must be built into the contract between the client and outsourcing service.

Subcontractors

If it is possible for the outsourcing vendor to use subcontractors, the contract must specify that any subcontractors are bound by nondisclosure, noncompete, and other provisions of the prime contract. If the client wishes to review and approve any subcontractors, the contract should so state.

Acceptance of Results

Where the work contracted out is fairly routine, such as processing the payroll or invoices, the results should be defined in the contract in terms of turnaround time, percent of errors, and similar yardsticks. But if the work is creative, such as designing a new system, the contract should specify who is responsible for making the decision of acceptability and should outline at least some general criteria on which that decision will be made.

Previous | Table of Contents | Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Asset and Staff Transfers

When hardware and software are to be purchased by the outsourcing contractor from the client, those items, as well as the financial terms, must be spelled out in detail in the contract. When employees are to be transferred to the vendor's payroll, those staff members must be specified or, if such transfers are to depend on interviews and reviews by the vendor, those arrangements and understandings should be made clear in the contract's provisions.

Documentation

Where the vendor will be upgrading or establishing new systems and programs, the contract should spell out the responsibilities and requirements for the documentation. In the case of instructions for the client's user personnel, the vendor, probably with the assistance of specified client personnel, should keep an up-to-date set of complete and clear users' instructions. The programming documentation and the instructions for the computer operating personnel must also be maintained completely and clearly in such a manner that they will be useful to those who will take over the system if and when it is decided to return the operations to the client or to another vendor. Who maintains and who owns the source code must be explicitly spelled out.

Arbitration of Disagreements

Hughes Aircraft Corp. outsourced a major part of its computer operations to

Computer Sciences Corp. for $1.5 billion. Hughes' contract provided for a committee representing several people from the client's and vendor's staff to resolve differences that might arise. Hughes found that the arrangement was very helpful in keeping the process moving ahead smoothly. It is suggested that a similar provision be considered in outsourcing contracts.

It is also recommended that the American Arbitration Association be specified in the contract as the agency to resolve any serious conflicts, instead of taking them to court. Arbitration is cheaper, faster, equally fair, and a bit less unfriendly. It can even be done without attorneys, although they are usually present.

Termination

Every contract has an end, and that event must be carefully provided for in the contract. Naturally, when preparing a contract, everyone's prime attention goes to how to get the transition started, and it is instinctive to think that the termination arrangements can be worked out later, or when the need arises.

Because every contract does end, the various possibilities must be thought out and planned for. The end may come when the client wishes to transfer the work to another vendor, decides to bring back the operation in-house, or the nature of the job is so modified that other arrangements must be made. The termination may be initiated by the client or the vendor, and the timing of the advance notice must be specified. Hopefully not, but the termination may be caused by the client or the vendor being dissatisfied with how the contract is working. It could be that either the client or vendor goes out of business. All of the myriad possibilities must be taken into account and provisions written to make the change as undisrupting as possible however it happens.

As the start of an outsourcing transition does not usually occur instantaneously, neither does the termination; it may be reasonable or necessary for the vendor to assist in the final conversion, and the terms for doing so should be planned and reflected in the contract terms.

HOW TO NEGOTIATE — STRATEGIES AND TACTICS

General Concepts

The objective of any outsourcing arrangement is to establish a relationship and a process that are beneficial to both client and vendor. Anything less is an unsatisfactory result. No outsourcing relationship is without its frictions and problems. However, if on balance either the vendor or client is seriously unhappy, then the other party cannot be pleased either.

Nevertheless, in the negotiation game, both client and vendor use strategies and tactics to gain favorable terms for themselves. The client should have prepared an RFP or other detailed statement of what it wants, and the vendor often will start off with a standard contract with terms that are favorable to it. Then both parties sit around a table and discuss all the features they want, and give and take to come to an accommodation that each considers is to its individual benefit.

Many articles and books have been written on the art and tactics of negotiation. Only the basic principles are covered here. The fundamental needs are to be forthright and fair, to consider as many contingencies as possible, and to try to listen to and understand the concerns of the other party. The contract provision areas outlined above can be used as a checklist. Attorneys are skilled at thinking of other issues, such as which state's rules apply, and most are proficient in negotiating methodology. Bringing in purchasing, audit, and human relations, in addition to information-systems specialists, can further help to make sure all relevant issues are considered.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Roles of Attorneys and Others

The computer operations manager, the CIO, or other IT executive should normally take the lead in determining what work to outsource, in reviewing possible vendors, in selecting the most likely one (or more) to confer with, and in negotiating the final contracts. The attorneys involved should be helpful from the beginning in seeing that the RFP, work statement, and contract are clear and unambiguous to all concerned, and cover all the legal issues that can reasonably be expected to arise.

In addition to the computer and legal officials, many organizations have found that other staff members are valuable in the planning and negotiation processes. These include human relations staff, who advise about the handling of employees to be transferred or terminated, and the morale of those who remain; auditors to help assure adequate controls of the work to be outsourced; and purchasing, which has the experience in selecting and negotiation with vendors.

DIRECTV® - Case Example

As reported in the August 1996 Managing Office Technology, DIRECTV® is a top direct-satellite company based in El Segundo, CA, with a goal or achieving 10 million subscribers by 2000. With an 18-inch dish, DIRECTV delivers more than 175 channels, both on a subscription and pay-per-view basis. Anxious to concentrate on its demanding marketing aspirations, DIRECTV elected to outsource its information systems and the preparing of invoices.

DIRECTV selected Output Technologies in Kansas City, MO, to prepare and mail over 3,400 invoices each day. The data about the purchases to be billed are collected for each customer by a smart card attached to the TV, and are relayed to DIRECTV and then to Output Technologies over the regular phone lines. Output Technologies also inserts information and marketing material in with the invoices, selecting the inserts by considering their weight and the needs of different demographic areas.

Output Technologies checks for input errors and reports quality problems to DIRECTV. It also provides security by using protected transmission lines, and by having alternate facilities that could continue in the event of a disaster.

This dynamic outsourcing relationship has given the broadcast satellite service a competitive advantage in the market by allowing DIRECTV to focus on its core competencies.

After deciding to outsource its bill processing and related functions, DIRECTV compared the capabilities of several potential vendors, and selected Output Technologies because its research showed that the vendor had a strong computer facility and volume capacity, its client references were favorable, it demonstrated a record accuracy in bill processing, its operating methods were sound, and it had the capability and flexibility for customized projects. After some years of working together, client and vendor agree this selection has proven to be sound.

CONCLUSION

The following points are highlights of the action a computer operations manager should take to secure the right outsourcing vendors and negotiate favorable terms with them:

- After researching where and how outsourcing can be beneficial to the organization, and supporting such research initiated by others, the computer operations manager should take the initiative in seeking the best vendors to do the work and in negotiating favorable terms with them.
- Select more than one vendor when the best expertise and experience desired cannot be found in any one vendor.
- Research thoroughly the reputation, capabilities, soundness, and compatibility of vendors selected, since the liaison is like a marriage,

which requires working together harmoniously to make for a successful alliance.

• Provide flexibility needed to adapt to changing needs and conditions, by enabling renegotiations as circumstances warrant.

• Take advantage of the skills and participation of attorneys for their negotiation and legal talents, of human resources for guiding employee changes and morale, and auditors and purchasing personnel for their advice and assistance in controls and vendor relations.

• Provide for arbitration of problems to avoid the hassles of court proceedings.

• Make sure the work requirements are clearly documented and fully understood by vendor and client personnel to avoid misunderstandings, a most common source of outsourcing problems.

• Include provisions for the vendor and client to share improvement benefits, so that the vendor is motivated in that direction.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter II-6
Outsourcing Desktop Support Services

William J. Beaumont

Many IS organizations struggle to control the skyrocketing costs associated with desktop assets. Companies that have traditionally relied on large, distributed systems as the backbone of their information systems are increasingly dependent on desktop systems. Although the use of desktop computers has grown dramatically, only recently have PCs and LANs been entrusted with mission-critical tasks. This new dependence on PCs has created certain problems because PCs and LANs do not have the centralized support infrastructure that data centers had.

GROWING UP WILD

Desktop systems have evolved in an undisciplined fashion, with few central controls and often with little central planning. This is not surprising, since most PCs started out as personal productivity tools — calculators were replaced by spreadsheets and typewriters by word processors. Computers became the tools of office task automation by virtue of the explosion of computing power, made possible through chip development and aided by sophisticated software.

Business managers purchased desktop computers without much coordination with legacy systems and often with only ad hoc guidance from the IS professionals running the data center. There was no standard PC platform, no standard for support, and usually little or no centralized management of desktop systems. Thus, the support underlying desktop technology has been a multilayered jumble in some organizations.

For years, everyone from technologists to business managers was preoccupied with buying equipment, installing it, integrating it with existing systems, and linking it with the necessary cabling and adapters. Often, PCs are not adequately administered, and this lack of coordination results in duplicated services, resources, and expertise, as well as difficulty in implementing standards and applying systems management tools.

THE TWO-HEADED MONSTER: COST AND CONTROL

As a business tool, a computer helps users increase productivity by performing tasks cheaper, better, and faster than alternate methods; if it does not, it is not worth buying. After all costs are counted — purchase, installation, operator training, and support — the computer should save the organization money. The purchase of desktop units must pass the same test that every business expenditure must pass: it should generate a return on investment. *Fortune* recently reported that only 20% of all CIOs believe they are obtaining a return on their IT investment.

Total Cost of Ownership. Whether a piece of equipment is a sound investment depends partly on what the cost is and how it is measured. For the past decade, the Gartner Group has compiled data on a measuring concept, now widely adopted by the IT industry, known as Total Cost of Ownership (TCO). TCO is defined as the total cost of the computer asset, including the hard costs of purchasing hardware plus the soft costs of labor associated with supporting and maintaining the asset.

As much as 85% of a PC's total cost of ownership is associated with soft costs, which can be as much as $10,000 per machine, per year. The TCO of a typical desktop PC is now approximately $44,000 over the PC lifecycle (i.e., a period of five years).

Gartner has been tracking costs using its TCO model since 1987. In that time, five-year costs have jumped more than 135%. A fully networked PC is even more expensive at approximately $11,900 per node, per year for hardware, software, support, administrative services, and end-user operations.

Although processors, costing a great deal less than before, are now extremely fast, increases in labor costs have been massive. Administrative expenditures have quadrupled and those for end-users have doubled, prompting the urgent need to bring distributed computing costs under control.

INTEGRATION OF DESKTOP SUPPORT SERVICES

The only way to control the escalating costs of desktop computing is through greater coordination of desktop-support services, since that is where the majority of corporate IT money is spent. Any company hoping to gain control of its IT environment must be able to measure and justify all of the expenses associated with end-user computing.

This justification can best be achieved with a fully integrated desktop-management solution for identifying, controlling, and reducing TCO, which should include ways to analyze improvement plans, implement optimization plans, and audit results to ensure that reductions are being achieved.

Savings of several thousand dollars per PC can be achieved by organizations committed to a serious, companywide cost-reduction effort. Even greater savings are possible — from 25% to as much as 50% — when the effort is managed at the enterprise level.

Outsourcing Support Services

The need for third-party support services continues to grow at a rapid pace as companies struggle to keep up with growing user demands within multivendor, multiplatform environments. Soft costs are escalating beyond affordable rates, making outsourcing of support services a more efficient and practical way of supporting employees. In addition, outsourcing these services allows companies to focus on business issues at hand instead of addressing individual problems that could be handled more effectively by another party.

The increase in outsourcing support services can largely be attributed to complex processing environments. IT managers are challenged by the sophistication of today’s networks and application software, the shortage of skills needed to support their users, and increased pressure from day-to-day business issues. It has become more difficult for even the largest in-house support center to stay abreast of new applications and technologies to support its own base of users and keep them satisfied.

Mixed PC/workstation environments spend more than $400 per user, per year on problem resolution. Multiplied throughout an organization with hundreds, or even thousands, of users, the cost of providing effective support becomes prohibitive.

Third-party organizations that focus on support services can leverage their investments in technology and staffing across a broad base of clients to obtain economies of scale that allow them to provide cost-effective solutions. A single point of contact to handle multiple support issues, ranging from software applications through operating systems through hardware problems, has become a more cost-effective way to resolve problems on a per-user basis.

The most practical solution helps customers optimize their investments in IT by taking on some or all support tasks facing overwhelmed, underbudgeted IS staffs. This allows IS to focus on strategic functions such as developing new systems and introducing new technology that benefits their businesses.

Previous | Table of Contents | Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Bookmark It
Search this book:
Go!

Previous | Table of Contents | Next

The third-party service provider should be:

- *A single source to consult for product management*. The vendor should singlehandedly coordinate workgroup management across the enterprise.
- *Vendor neutral*. The third party should support all brands and harbor no product bias.
- *Simple to deal with*. The third-party service should be a virtual extension of in-house support.
- *Low-risk*. The vendor selected should have an established track record.
- *Able to produce concrete, measurable benefits*. Third-party providers should be able to reduce costs, enhance control, maximize systems availability, and increase end-user satisfaction.

A STRATEGIC, COMPREHENSIVE APPROACH TO CONTROLLING COSTS

Most customers need a strategy for controlling costs at the desktop, including the following integrated services.

Planning and Consulting Services. Planning-support services help customers define their environments, set targets for improvement, and monitor progress against plans. These services should be offered either as a one-time service or used over the course of a project.

Consultants should be used on ad hoc projects. Sometimes the greatest value of outside consultants is having an independent, unbiased party review internal operational reports or organizational approaches to problems.

Hardware Services. Both remedial and preventive maintenance, with optional on-site staff support, should be provided for a variety of products. Almost every desktop unit undergoes at least one or two changes per year. Moves, adds, and changes to any system should be supported, providing a comprehensive solution.

Asset Management Services. Asset management contributes to service effectiveness in large organizations. Accurate information is vital to cost control. Companies that engage in IT asset management stand to reap significant savings and boost end-user productivity.

An effective asset-management service inventories a customer’s IT environment and creates an asset database through a combination of automated and manual data-collection activities. Once completed, the database should contain information detailing the customer’s IT assets from several perspectives including hardware, software, end users, location, and cost centers.

To ensure that all facts are up to date, it is a good idea for a client-support team to apply specialized software to probe the network for new, moved, or changed devices and see that all changes are captured in the database. Periodic audits of the IT environment should be performed.

End-User Support Services (Help Desk). End-user support services, more commonly known as the help desk, aid users with the navigation and resolution of computer technology questions, as well as supporting hardware, software, and system-related activities. Services should meet the growing needs of multivendor processing environments, offering new solutions that are flexible to a customer’s needs.

End-user support services should include:

- *Call management*. Support should include a problem triage service.
- *Basic end-user support*. The service should support popular shrink-wrapped applications.
- *Network end-user support*. Network operating systems should be supported.
- *Customized support*. The service should be able to support unique software packages often found in particular industries.
- *Advanced product support*. Complex software, such as Lotus Notes, should be supported.

Network Support Services. IS organizations spend approximately 40% of their time on networking issues and 60% of their tools budget on network management.

A customer’s environment should be remotely monitored continuously to improve costs, boost user productivity, and increase systems availability. Round-the-clock monitoring allows for the identification of potential trouble spots and enhances network performance by helping to highlight areas for improvement.

Program Management. Some customers need support for individual projects or complete management of their desktop environment. In either case, a tailored program should be designed, allowing companies to focus their resources on their core competencies.

Support Partner Services. When a customer requires additional services such as leasing, procurement, or disaster recovery, it is ideal if the vendor can respond with an integrated solution through support partners. These preferred vendor relationships offer quick, comprehensive responses to a customer's business requirements.

THE SOLUTION IS AT HAND

Technology is changing at a rapid pace, and companies are struggling to come to grips with managing their desktop assets. They are seeking innovative solutions to lowering the total cost of ownership of their IT environment.

To improve operational performance, service and support must be integrated. A tightly integrated suite of services to support desktop systems is the best solution for controlling the exploding costs of desktop computing.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Chapter II-7
Outsourcing the Help Desk Function

Fritz H. Grupe

Despite the debate over the productivity paradox, corporate managers appear confident that their IT investments yield a significant return. They continue to acquire and use computers and their attendant peripherals and networks at an accelerating pace. Not surprisingly, investments in help-desk-related goods and services are estimated to be increasing from 25% to 50% nationwide.

A major component of the IT outsourcing industry centers in the help-desk arena. The following several considerations drive this development:

- Business size
- Performance
- Cost accounting
- Service demands
- Business needs
- Staff development

Each of these drivers is discussed in the sections that follow.

Business Size. Some businesses are too small to provide adequate staff support. Although such businesses deploy as wide a variety of systems, software packages, and networks as their larger counterparts and encounter an equal number of related problems, they have fewer resources with which to resolve problems. Outsourcing offers these businesses a range of service options delivered in a professional manner. For example, an outsourcer specializing in help-desk operations can acquire and implement expensive support tools that facilitate end-user support but are not cost-effective when implemented by small companies.

Performance. There is little doubt that outsourcing is sometimes viewed as an alternative to what are perceived as ineffective, unresponsive IS departments. Help desks are challenged by a host of difficulties including high turnover, poorly trained and improperly motivated staff, staff-recruitment problems, and unsatisfactory problem tracking. Understaffed help desks often service end users reactively and fail to offer proactive solutions through training and end-user documentation.

Cost Accounting. In-house help-desk operations are frequently budgeted as general expense overhead. This form of budgeting does not encourage judicious use of help-desk resources. An outsourcing arrangement developed around a chargeback system focuses attention on the cost of unnecessary services. Services must be priced realistically so the so-called underground network does not shift the workload away from the help desk to unskilled and inefficient co-workers whose time does not appear on an expense sheet.

Service Demands. Some companies use outsourcing vendors to provide specific services for the in-house help desk, such as coverage for peak periods, nonbusiness hours, special constituencies, or specific problem areas.

Business Needs. Some companies need to downsize while focusing on their core specialties. Outsourcing the help-desk function offers a means of transferring the funds committed to positions that are not highly valued to the acquisition of a higher-quality service. The strategy of converting fixed costs into variable costs holds promise of reduced overall expenditures.

In addition, IS managers in companies experiencing a rapid growth curve may not have the time to properly staff, train, and equip a help desk. In spite of the help desk’s importance in enabling end users to fully utilize mission-critical applications, the help desk is not a mission-critical function. Many companies that outsource the help-desk function do so simply to avoid the necessity of staffing and managing such a unit while expanding their services. Considerable savings of management time are realized when scheduling, training, hiring, and coordination are off-loaded to an outsourcing vendor.

Staff Development. Help-desk positions are not considered a highly desirable career goal. An outsourcer can provide employees with more-extensive training opportunities and establish career paths within its management structure. Outsourcers may also have more opportunities to move promising staff into positions unrelated to help-desk operations.

OUTSOURCING OPTIONS

Managers reviewing their help-desk needs and the best way to address them should consider four options in respect to outsourcing. Selection of the most appropriate model depends, of course, on individual company circumstances.

- *Insourced help desk*. An organization with adequate resources or a high inventory of in-house developed systems may find it most effective to staff the help desk entirely with its own employees.
- *Outsourced help desk*. In this model, a firm specializing in help-desk operations provides all help-desk services.
- *Hybrid help desk operations*. In the hybrid model, the organization provides some of the help-desk employees, and the outsourcing firm provides others. This model offers several flexible service options. Companies can outsource support for shrink-wrap applications only and use the internal help desk for legacy and proprietary systems. Or, they can contract the first level of support and retain support for more-substantial problems. The reverse strategy is also possible.
- *Provision of supplementary support tools*. In this model, outsourcing vendors provide prebuilt expert systems, case-based reasoning tools, and automated text-retrieval systems (e.g., hypertext and indexing systems) to enhance the help-desk staff's access to up-to-date information.

Substantial numbers of help-desk outsourcing arrangements have been implemented. Following are some of the more illustrative examples.

Taco Bell. The Store Operations Support Project (SOS Taco) is a partnership between Taco Bell and Coopers & Lybrand. SOS provides assistance on inventory, staff scheduling, point-of-sale data gathering, and other back-office operations for more than 4,100 sites in the U.S. and abroad. Coopers & Lybrand provides nearly 20 part-time and full-time supplementary staff for the help desk during high-demand periods and night shifts. This group augments a much smaller Taco Bell support group, creating an in-house core of people who can work with other Taco Bell employees on development activities. Coopers & Lybrand provides each employee with four weeks of training. The help desk handles in excess of 20,000 calls each month.

Novell, Inc. This well-known vendor of networking and other software outsourced support services to Dallas-based CompuCom Systems, Inc. The agreement merged three help-desk operations into one and serves some 6,000 employees at four main sites.

Microsoft Corp. When Microsoft needed help in supporting new users of Windows 95, it contracted with Boston-based Keane, Inc., to set up a help desk with 350 support personnel. The Keane help-desk group augmented Microsoft's existing telephone hot-line staff.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

SELECTING A HELP-DESK OUTSOURCER

Given the options for operating a help-desk organization, the process of selecting an outsourcing vendor seems daunting. With forethought, proper information gathering, and careful planning, however, the process of identifying the right outsourcer becomes straightforward. The exercise of gathering requirements and current customer-support information is both informative and the key to ensuring that the outsourcer provides the correct type and level of service. These requirements are specified in a request for proposal (RFP) and become the basis for evaluating outsourcing vendors and negotiating the service-level agreement.

It is important to be honest with potential vendors about support requirements and company-specific idiosyncrasies. Working as a team with the vendor results in a win-win outsourcing arrangement. The following six steps provide guidance on how to achieve this arrangement.[1]

Step One: Developing Help Desk Requirements

The critical first step of any help-desk outsourcing venture is gathering the information needed to develop accurate requirements for the new or improved help-desk facility. Clear objectives and accurate requirements greatly enhance the ability of the outsourcer to develop an offer that matches company expectations. Conversely, a lack of accurate information results either in an overpriced bid designed to ensure that the vendor’s risks are covered or in an underpriced bid that inhibits the vendor from meeting service expectations. Experienced help-desk consultants, who can quickly assess the current environment and facilitate goal development, can provide valuable assistance during this first step.

The process of developing help-desk requirements involves determining outsourcing objectives, project scope, and, in the case of an existing help desk, current performance.

Determining Goals and Objectives. An organization must determine the goals of its help desk and why outsourcing is under consideration. Is the goal to implement a new help desk, increase the efficiency of an existing operation, or follow the company's strategic plan to outsource non-mission-critical functions? After general goals are established, the following more specific objectives are set:

- Level of coverage or hours of operation
- Level of responsiveness
- Customer-satisfaction criteria

Determining Project Scope. If the goal is to outsource an existing help-desk operation, information on the scope and size of the outsourcing project is gathered by examining current resource use. The existing operation, particularly the procedures and tools used, should be thoroughly documented and statistical information to determine call volumes and peak periods should be captured. If the goal is to implement a new help desk, outside assistance is particularly helpful during this phase. Specific scoping tasks include:

- Describing current processes, software tools, and equipment
- Calculating service capacity
- Determining call volumes
- Determining average length per call
- Identifying peak-period support requirements
- Describing the range of applications and products to be supported
- Calculating current costs
- Calculating support staff costs
- Calculating facilities and equipment costs
- Documenting skill requirements

Determining Current Performance. An accurate picture of the current level of performance is important for determining the service-level agreement, especially if service-level improvement is a goal of the outsourcing effort. Performance statistics should be captured as part of standard help-desk operations. Key questions to ask include:

- What is the average wait in queue?
- How many calls are resolved on first contact?
- What are the average response times by category of calls?
- What is the call-abandon rate?
- Is there a backlog of open problems?

Step Two: Determining the Appropriate Outsourcing Model

Once the goals, scope, and performance objectives are determined, the appropriate outsourcing model is selected. One fundamental question is whether to outsource the function or the solution. When an outsourcing vendor assumes the support function, the systems and processes currently in place are replicated. This approach has the advantage of appearing seamless to help-desk customers. Outsourcing the complete solution allows the outsourcer to re-engineer the operation as required to provide the optimum level of service. Other issues that must be decided include whether to outsource the entire help desk or only portions of the operation, to use onsite or offsite support, or to purchase support as part of a shared help-desk function.

Step Three: Formulating a Request for Proposal

Once the groundwork has been completed, the next task is to create a request for proposal. This document is the prime communication vehicle used to solicit bids from outsourcing vendors. Although each situation is unique, experience has shown that an RFP should contain several basic sections:

- *Overview*. The overview section describes the objectives for the outsourcing project and lists the reasons for creating the RFP.
- *Background*. This section describes the company's business, mission, locations, and organizational structure.
- *General requirements*. The general-requirements section, often the largest section of the RFP, describes guidelines for proposals, a time schedule for the evaluation, contract terms, and the logistical requirements for the project.
- *Required service levels and volume forecast*. This section is used to describe the expected quality results and projected volumes. It delineates current service levels and minimum acceptable service levels.
- *Selection method*. This section describes the criteria and method for choosing the outsourcing vendor. A partial list of possible selection criteria includes company stability, management organization, flexibility, facilities, service capabilities, training, price, implementation plan, and references. The selection criteria should be ranked by priority.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Step Four: Evaluating Vendors

The evaluation process begins as vendors respond to the RFP with their own proposals. If numerous vendors respond, a cursory evaluation of the responses may be sufficient to eliminate marginal vendors. A more thorough evaluation of the remaining proposals is needed to determine the vendor’s comprehension of company requirements and the quality of the proposed solution. This evaluation include several steps.

Reviewing Proposals. Proposals should be evaluated with a matrix containing the evaluation attributes, their weighting, and a column for each vendor. Each vendor’s solution is ranked for each attribute on a scale of 1 to 10. These scores are then adjusted by the appropriate weighting factor and added to produce an overall company score. Exhibit 1 depicts a sample evaluation matrix with recommended evaluation attributes. Key questions to ask include:

- Does the proposed solution meet the objectives described in the RFP?
- Does the vendor have the experience and resources to implement the solution?
- Does the vendor have a formal methodology for help-desk support?
- What is the vendor’s reputation for meeting service commitments?
- Does the vendor have a reasonable implementation or transition plan for phasing in the outsourcing service?
- Is the vendor’s pricing in line with the level of services provided?

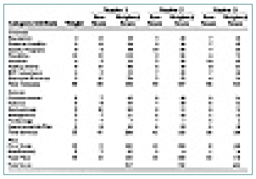


Exhibit 1 Sample vendor evaluation matrix.

Visiting the Vendor and Checking References. It is prudent to evaluate the veracity of vendor claims by checking references and visiting one of the vendor's help-desk facilities. Although references are helpful, a visit to an operational help desk is the most effective way of gaining insight into the vendor's practices. IS managers should be sure to examine the support procedures in action and talk to help-desk clients about their support satisfaction. Examining customer survey results is another good source of this information.

Step Five: Selecting a Vendor

Carefully following the selection process just described simplifies final vendor selection, because it is likely that one vendor more clearly meets requirements than others. IS managers should take care, however, not to let price alone determine the outcome of the selection process. Although price is an important consideration, lower price alone may not provide the highest value. Managers must carefully weigh cost with other considerations, such as service levels, to identify the vendor that provides the highest value per support dollar.

Although precise criteria for evaluating outsourcing vendors depends on the nature of the individual contract, some general areas of evaluation common to all help-desk outsourcing vendors are not covered by the contract. Weightings for each of these criteria should also be developed according to individual organizational needs. A list of such criteria follows:

- How much experience has the vendor had with help-desk operations? How many other contracts has the vendor implemented? Does the vendor have positive references? How well has the vendor adhered to the letter and the spirit of previous ventures? Does the vendor use effective project management skills? When contractual problems have arisen, have they been handled smoothly and to each party's mutual benefit?
- Is the vendor's staff capable of providing the required support? Can the vendor's staff handle all of the application programs, programming environments, and networking systems present and implement sophisticated knowledge bases of previous problems solved, interactive voice-response systems, call-center operations, automated support tools, voice and e-mail, and statistical programs? Will the vendor be able to recruit, train, retrain, and retain qualified staff who have the depth of specialization to meet difficult, company-specific problems?
- Do you know who the outsourcer's project manager will be? Do you have résumés of key players and do you know what the degree of their involvement is?
- Is the vendor stable financially and managerially? Does it have the

funding and the managerial competence to survive? Does the vendor have some knowledge of your business and can its people adapt to the types of employees in your company? Does the vendor have the resources to remain competitive in the outsourcing arena?

- Will your relationship with the vendor be synergistic? In other words, will the vendor realize anything other than financial gain from the partnership? Will the vendor make the commitment to form a productive relationship that offers both client and vendor a competitive advantage?
- Can you trust the outsourcer's personnel to exercise confidentiality regarding sensitive issues such as business objectives, technology plans, and business data? Is the vendor open to changing modes of operation to fit your corporate culture? Can the vendor's staff be seamlessly transitioned into your operations?

Step Six: Defining the Contract and Negotiating Terms

The final step in the outsourcing selection process is the definition and negotiation of the terms by which both parties in the outsourcing agreement function. It is crucial to ensure that all terms used to describe the help-desk operation are clearly defined, because assumptions about the meanings of terms such as calls, problems, and incidents can lead to contractual nightmares.

A clearly defined contract helps control the costs of outsourcing by providing an accurate description of which cost items are included in the arrangement, as well as the items that need to be assessed for demand measures. A distinctly outlined contract also provides vendors with a sharper definition of what services they have to provide and less opportunity for hedging cost estimates.

Despite the need for a clear contract, it is important that both parties be ready to modify their means and ends as the outsourcing relationship emerges and develops. Not all aspects of the contract are foreseeable. Following are some of the issues the contract should define.

Coverage Areas. The contract should specify which hardware and software is to be supported and whether support pertains to industrywide products and services or to company-specific legacy systems that require greater support.

Service-Level Agreements. The evaluation criteria specified in the service-level agreement include the types of response times and successful completion rates to be used to rate service as being above, at, or below average. Hours of operation should also be delineated, as should the expected volume of calls and any actions to be taken if the number exceeds or fails to meet expectations.

Other issues that need to be determined include which types of calls are considered top priority and which can be delayed, acceptable response times, and whether the calls/requests for assistance can be categorized by length of time needed for a response and by complexity. Finally, the predictability of the call distribution throughout the day must be considered.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Reporting. It is essential that explicit quality-control measures, such as periodic planned review sessions, be in place. The group or individual overseeing the contract must address several questions, such as:

- What statistical and other types of reports are needed and what types of data will gathered for use in these reports?
- What logs and tracking systems must be in place? Must they track abandoned calls and the length of time spent waiting for pickup and response?
- Will calls be monitored on a selective basis?
- Who is able to conduct call monitoring?

Determination of Service Adequacy. It is important to delineate which measures will be used to determine whether the help desk is succeeding. Available methods include end-user satisfaction surveys and help-desk statistics.

Staffing and Scheduling. The contract should address the issues of how many and what types of support personnel are expected to be available and the hours during which they will be available. Procedures for escalating problems to more-experienced personnel when initial contacts are unable to provide resolution should also be specified.

The contract must obviously deal with issues relating to hybrid work situations, such as how personnel problems will be managed and whether the necessity for co-workers to function under two different salary schedules and personnel policies will prove problematic. Finally, the contract should stipulate whether and for how long the outsourcing vendor is expected to employ organizational staff.

Implementation. Service implementation should specify how the service will be initiated and how a transition to full operation should be accomplished. One vendor conducted a preliminary test of the outsourced help-desk system by preparing several hundred calls that tested whether adequate escalation procedures, automated systems, consulting skills, and other processes were in place. Access to the help desk by corporate users was initiated only after adequate test results were achieved.

Location. The contract must specify whether calls can be received and assistance provided from a distance or on site. If onsite support is provided, the number of sites needing such support must be specified. Finally, the availability and use of alternative means of communication such as fax or electronic bulletin boards should be included.

Equipment. The area of equipment involves numerous and broad issues relating to both hardware and software. For example, the contract must specify whether software for logging and tracking problem resolution is needed, on whose machine the software will reside, and who holds the license and owns the data. The availability of other software, such as e-mail, corporate groupware, proprietary software, and change-management software, should be addressed.

Other issues include stipulation of needed telephone systems, use of artificial intelligence packages and whether their contents (e.g., rules, cases, and problems) will be provided immediately or built up over time according to the company’s own experience, and use of interactive voice-response systems. The contract should also address whether help-desk staff will operate end-user equipment such as printers, training laboratories, or general-purpose workstation facilities.

Termination and Renegotiation. In addition to stating an expiration date, the contract should cover mediation or arbitration procedures. It is should also delineate the exit procedure following termination. Other items for consideration include where automated software used for accumulating knowledge will reside after the contract period.

Payment and Costs. The contract must delineate total costs as well as the payment method. If a chargeback system will be used, its costing structure should specify whether the chargeback is by call, size of the user base, problem category, or a blanket fee. Bonuses or penalties for the vendor’s success or failure to meet objectives such as increased first-call problem resolution or degree of customer satisfaction must also be specified.

Change Service Orders. Workloads are rarely fixed or entirely predictable. The contract must therefore state how additional costs for a changing workload or the addition of new functions will be levied.

OUTSOURCING PROBLEMS AND PITFALLS

Several potential problems can derail a help-desk outsourcing effort.

Losing Track of Long-Term Goals. Outsourcing should be a long-term relationship. Short-term goals such as immediate cost savings or correcting a staff-turnover problem may have a high initial profile but are misleading in the long run.

Selecting an Unqualified Outsourcer. As is the case with many technology-oriented firms, specialization exists in the help-desk arena. Not all companies are equally equipped to effectively handle outsourcing. Some firms have extensive experience in selecting support tools, specialists who can address complex issues, and regularly trained staff. They also have a track record of building mutually beneficial partnerships with their clients based on financial stability, technical skills inventory, and management competence.

Failing to Consider all Contractual Consequences. An organization outsourcing its help desk cannot simply accept a proposed standard contract. Managers must fully understand the implications of the outsourcing decision and why specific services are being outsourced, as well as the effects of altered conditions such as corporate growth or stagnation, addition of functions, and other organizational changes on the help desk.

Assuming that a Contract Will Administer Itself. Signing an outsourcing contract does not end management responsibilities. A qualified contract manager or group is essential for overseeing the implementation of the contract and ensuring that the vendor handles personnel issues appropriately, corrects problem areas, and provides appropriate reports.

Omitting Items from the Contract. When it comes to contracts, the sins of omission are at least as serious as the sins of commission. What is left out of a contract can be expensive to add on later.

Losing Staff Control. The offset of having the outsourcer do the hiring and firing is that the contractee has to work through the outsourcer to correct problems. Because the outsourcer may choose or be forced to reassign people in its organization, the contractee could suffer from a higher turnover of trained staff.

Encountering Resistance from IS Departments. Not only does the outsourcer replace IT staff, new staff members become part of the IT operation. As external observers become aware of internal problems, response rates, user attitudes, and backlogs, a great potential for conflict is created. When a hybrid arrangement is in effect, differences in pay scales, leave and other personnel policies may become points of contention.

Setting Unrealistic Goals. Contract goals, however measured, should be realistic. Overly ambitious expectations of improvement can prove illusory and negate chances for improvement. For example, one vendor reported that an emphasis on first-call resolution hampered upper-level executives and their aides from receiving the help they needed. These people did not want to wait on the telephone while an answer to a help-desk query was being developed. When the answer was ready, or when assistance could be provided onsite, the executives wanted to be contacted again.

CONCLUSION

As the number of end users has increased, so has the demand for help-desk services. Although keeping end users productive is essential, organizations are finding it difficult to provide the breadth of support needed. Outsourcing all or part of the help-desk function offers one alternative for providing services economically and effectively. Selecting an appropriate vendor and preparing a detailed contract are critical to the eventual success of such a partnership. Following the steps presented in this article will help IS managers accomplish both of these processes.

Note

1. *The Help Desk Opportunity: Opportunities to Optimize Productivity*, white paper, (Keane Corp., Boston 1995).

Previous | Table of Contents | Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter II-8
The Help Desk in a Distributed Environment

John L. Connor

Help desks were originally implemented because there were too many customer calls for the data center to handle. There were too many problems and too little problem-solving time. A help desk consolidated, recorded, and tracked problems. However, help desks were typically set up without goals and objectives and with a minimal investment in training and tools.

This chapter examines what can be done to improve help desk operations and how traditional technology management—the acquisition by vertical specialization of such tools and product knowledge as MVS, CICS, and DB2—fails to satisfy the requirements of a help desk operating in a client/server or distributed environment. The chapter also examines the requirements and tools for a help desk to function successfully in a rapidly changing environment.

THE EVOLUTION OF THE HELP DESK

Historically, data center customers have had their problems solved more by luck and perseverance. Exhibit 1 shows that as newer technologies have emerged the level of problem calls to the data center has increased but the time available for solving these problems has decreased. This situation more than any reason triggered the rise of help desks.

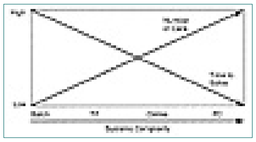


Exhibit 1 Progression of customer calls to the help desk.

Companies created help desks out of necessity. They found themselves having to support systems and products that were becoming more and more complex and increasingly critical to their survival.

The first step in creating help desks was to have programmers support systems, because they were the only ones who understood the systems. The programmers did not have, however, the right people skills and often annoyed customers. They were replaced by professional help desk personnel, who had terrific people skills but were usually clueless about systems. Support became friendly but shallow. A help desk was a place for a customer to call to get somebody to call somebody else who could help. This arrangement was not terribly efficient.

Reasons for forming a help desk included:

- Too many customers for the data center to handle.
- Too many problems to be solved.
- Many simple problems to be solved.
- Less time to solve problems.
- Vendors needed central problem management.
- Sharp increase in customer complaints.
- Help desks became a trend.

However, many help desks just happened to be formed because one person or section was delegated to answer calls from customers as the volume grew.

From these murky beginnings, the help desk began to emerge—complete with the poor reputation of many IS departments. As a result of their origins, many help desks were poorly planned and were not as successful as they could have been. Other reasons for lack of success included:

- No terms of reference (i.e., goals and objectives).
- No customer consultation.
- Inadequate staff training.
- Poor help desk launch.
- Lack of management support.
- Low investment.
- Poor staffing.
- Low profile factor.
- Lack of tools.

CHANGING USER REQUIREMENTS

The questions and expectations users have of a help desk have changed dramatically. Users once called a help desk to ask how to do a certain function with a product. Now, they want to know how to put a specific item into a specific spreadsheet in a certain document, manipulate it, insert it in a memo, distribute it to five people, get approval, print the document, and distribute it to their customers.

Questions to help desks are getting more complex. What do callers to the help desk ask about? According to a recent survey by the Help Desk Institute, callers to help desks ask about the following:

- Applications: 36%.
- Hardware: 24%.
- Data communications: 21%.
- Miscellaneous services: 16%.
- Voice communications: 3%.

The danger is that customers are gradually getting used to and accepting a lower level of support. They have already grown accustomed to having their phone calls answered by a recording. They are getting used to listening to recordings and pressing buttons to navigate through automatic call directors to get to real people who can answer their questions. Is this an efficient way to solve the help desk issues?

The answer is no. There must be a greater emphasis on customer service. Callers should be able to expect help desk staff to solve their problems. However, that means that the people who fill help desk slots must have in-depth knowledge of multiple platforms, software tools, and applications. It is not a matter of technology; it is whether technology is effectively used, and that requires a well-trained staff who can use automated tools to answer their customers' questions.

THE ROLE OF A HELP DESK

Often there are two areas in an IS organization that have responsibility for managing a company's computers. In this chapter they are called:

- *The Bridge*. The replacement for the old-style operations area.
- *The Help Desk or Service Center*. The new proactive, service-oriented replacement for the old-style help desk.

A pitfall in setting up a bridge function is that the expertise of a typical computer department is focused on a number of telephones. It is surprising how quickly users learn these strategic telephone numbers and call bridge operators for a considered opinion or advice. Many data centers have effectively implemented integrated, host-based automation, and a single bridge operator may control from one automated console 5 to 20 MVS systems and a multitude of subsystems. Because of this operator's scope of responsibility, these telephone calls can be very distracting. Distractions in the bridge area and their impact on productivity and efficiency are often seriously underestimated.

Because of these distractions, many data centers have set up their help desk facilities along with the bridge area. The actual location is not as important as a help desk's providing an effective level of service, thereby, relieving the bridge of user contact and first level problem diagnosis.

In a fully automated data center, a help desk assumes many functions of operations and operators have become service managers. Therefore, help desk staffers must know the working of the systems they manage. Detailed information on physical links, networks, and installed equipment is important to help desk workers; they need to be fully conversant with the complete operation, in most cases even more so than bridge personnel.

A help desk, like a bridge, relies on concise information being displayed accurately and as required. Without this capability, a help desk will lose credibility with users. But how many installations provide their help desks with this kind of support? Not enough, if the consensus of user opinion on the effectiveness of their IS department is to be believed.

Previous	Table of Contents	Next

Products | Contact Us | About Us | Privacy | Ad Info | Home

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

User Attitudes Toward Help Desks

The attitudes of help desk users are likely to differ from the attitudes of help desk staffers, and management should know the feelings of help desk customers. An organization should know the attitude of users to their help desk. A help desk's profile must be generally raised to a higher level if the help desk is to get the support it needs from management and its clients. In general, a low profile is a direct result of how help desks evolved, which was described in a previous section.

Failure to understand the current level of appreciation jeopardizes any attempts to improve a help desk's profile and image. The success of a help desk depends on its staff knowing what its customers want.

A Traditional Help Desk

To define how a help desk should now be, it is worthwhile to review how requirements for a help desk were defined a few years ago.

Senior management in a typical organization originally determined that its help desk staff should possess the following:

- Understanding of data processing terminology.
- Perseverance for tracking down responsibilities.
- Ability to troubleshoot problems.
- Understanding of a computer resources team.
- Ability to interact with people.

The basic job responsibilities were defined as following:

- Answering calls for help from users.
- Solving problems when possible.
- Referring calls for further action.
- Logging activities.
- Reporting activities.

On the basis of these job requirements, it seems that what IS organizations wanted as recently as 1989 was to staff their help desks with people who could take calls and look at problems. Problem solving was not high on the list. Companies have been building elaborate telephone systems, call management systems and problem management tools to accomplish an impossible task: enabling people to support technology they do not understand.

Evolution of problem Tracking Systems

Originally, help desk staffers would solve problems if they could. Otherwise, they wrote down the pertinent information, called the appropriate support person, and passed the information along. Once they passed the information on, there was no record of who had called, what the problem was, or who was handling it. Problems with this method include the following:

- No statistics.
- Lost information.
- Inability to handle multiple problems from one cause.
- Single problem solved multiple times.
- Poor service call management.
- Slow or ineffective call referral.

Thus, problem tracking systems came about. These systems allow a help desk staff person to enter on a screen all pertinent information about a call by just filling in the blanks. The reporting components of these packages can produce a wide variety of statistical information. However, entering information and monitoring results are still manual processes.

Help Desks in a Distributed Environment

For years there has been talk about the need for help desks to be proactive. However, this requirement to be proactive has been narrowly defined as reacting to problems in existing systems. To be proactive simply meant to solve problems with the existing technology infrastructure before users had to call the help desk. Is this being proactive? It is an improvement, but the focus is still on problem resolution rather than prevention.

To be truly proactive, a help desk must go beyond supporting existing systems and look at the needs of its customers. They must get involved at a much earlier stage and must:

- Assess customer needs.
- Select applications required to satisfy those needs.

- Supervise the installation of new applications.
- Train customers how to use new applications.
- Ensure that the necessary knowledge is delivered effectively to users.
- Resolve problems.

These are not the activities of a typical help desk, which focuses almost exclusively on reactively solving problems. A help desk should perform its traditional role of problem resolution, but its primary role should be that of a customer advocate. Not only should a help desk act as a customer advocate, it should also proactively work to avoid calls from customers. A help desk must solve problems through education, support, and effective knowledge bases to eliminate problems before they happen.

New Role of Help Desks

The roles of a new help desk or service center are still evolving, but will likely include the following:

- Assuring overall productivity of customers.
- Overseeing the quality of service provided to customers.
- Functioning as a customer advocate.
- Maintaining and distributing knowledge bases needed by customers.
- Providing support for existing applications.

A powerful reason for changing a help desk's mission are the changes happening in corporate computing today. The days of monolithic mainframe-based data processing departments are gone, and more important, so is the organizational structure that grew up around it. Today's information systems are evolving into a state of centralized decentralization.

Need for Central Control. Although networked systems are taking over, companies need a single organization to monitor and access the needs of customers and to act as customers' advocate for providing technology. A help desk staffer's job is more substantive and demanding than ever before, but an effective help desk analyst must understand mainframe hardware and software, operating systems, subsystems, personal computers, and applications, and company practices. Help desks continue to suffer from a poor image, which hampers recruitment efforts.

Also, as the responsibilities of a help desk change in an organization, so will its role. More and more, a help desk will report to the president or chief information officer (CIO) rather than to the IS organization.

Tools Required for a Distributed Environment

Today's help desk is the starting point for creating tomorrow's customer-oriented service center. An important question is "How are these new service centers supported through current tools?" Most tools available today are for a reactive help desk. They do not provide the capabilities necessary for proactive support. Many are little more than a place to keep notes about a problem; these products are often called *information* *management* or *problem entry systems*. They will not suffice for tomorrow's service center.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Most people immediately think of software or hardware when they think of tools for an IS environment. In this case, however, it also means effective people skills. Many individuals in a help desk do not have the basic skills or personal tools to make a service center concept work. However, these skills can be learned. Skills needed for a successful service center include:

- Effective and active listening.
- Knowing which types of questions to ask.
- Understanding problem solving by removing assumptions.
- Information gathering.
- Recognizing a problem.
- Knowing and applying problem-solving processes.
- Assessing problems and sorting facts from expectations.
- Decision making (i.e., judgmental versus rational decisions).
- Being able to apply and follow corrective, adaptive, interim, or procedural courses of action.

Centralization and Distribution

A centralized-distributed approach is needed for allocating the tools required for acquiring the skills needed for tomorrow’s help desk. Management tasks will be physically distributed but be under central management. Thus, an organization will be able to manage enterprise computing from one central location but still will have distributed agents on local platforms.

For example, a central help desk may be located in headquarters in Dallas and management sites may be located at offices in London, New York, Tokyo, and Chicago. Each management site manages its portion of the enterprise, but they would all be under the direction of the Dallas-based headquarters and follow the same standards.

A centralized-distributed help desk system is too complex for the tools of one vendor to handle. Previously, a multivendor environment meant one vendor for the tape system, and one vendor for the disk drives. Now more vendors than that are required for the inside of a single PC. Enterprise computing comes in a kit.

Exhibit 2 shows the enterprise kit implemented for a help desk. Various tools are integrated into a central help desk management tool. From one location and one display, all pieces of information needed to manage and solve customers' problems are available. When a call comes from a user, a help desk staffer can call up and display the user's configuration from a configuration management tool and continue talking to the user. The information collected from the user is entered in the information management system for future reference and problem management.

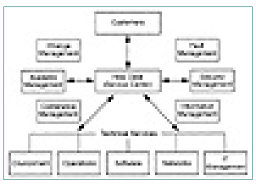


Exhibit 2 The help desk in a distributed environment.

Choosing Tools

The following steps can help an organization determine which tools are needed for a new help desk. This is an important step in the process of developing a new help desk or service center. It is inadvisable to rush into acquiring technology without proper planning.

1. *Taking a before picture*. Measure the performance of the current help desk operation. This is necessary to calculate the effect of the support technology and help justify expenditures.
2. *Becoming an internal salesperson*. Management must be convinced of the need and value of the support technology. The cost savings should be calculated and presented.
3. *Performing a needs analysis*. No assumptions should be made about how the current help desk functions. A flow chart must be made to explain every business process and how data is accessed, collected, and passed on at each step. Attention must be paid to steps and data that will not be required once new support technology is implemented.
4. *Having a systemwide strategy*. Such a strategy is a comprehensive blueprint and timetable for implementing new technology on different platforms and collecting information. This strategy must consider both the IS and the business plans for the entire company as well as customers' expectations.
5. *Asking vendors for help*. Vendors can be useful in conducting a needs

analysis and preparing a systems strategy. Late in the process, specific products can be measured against a request for proposal (RFP).

6. *Testing your RFP*. Many people just copy down product specifications or features from vendors' marketing literature. A better approach is to ask vendors how their product solves specific problems. Those who draft the RFP should try to answer the RFP before they distribute it. This helps ensure that the RFP actually asks for the information needed. This information is most often used to justify the expenditure of a significant sum of money.

Gathering Requirements for Tools

What are the characteristics of the tools needed to renovate a help desk? Key points to investigate are discussed in the following sections.

Designed for Automation. Computers can detect and resolve problems much faster than people. Local tools should filter out problems that can be corrected at the local level to reduce network traffic. The tools must be able to pass information back and forth. For example, if a minicomputer system is having a problem, the monitoring tool for that minicomputer passes only the required information to a central facility that help desk personnel monitor.

Designed for End-to-End Processes. The help desk staffers need to see what a customer sees and that means enabling them to view information from the customer's terminal or PC. Too often, tools are purchased by vertical specialty (e.g., for MVS, CICS, IMS, DB2, WAN, LAN, or minicomputer). Tools purchased in this manner do an excellent job of managing the specific vertical specialty, but they only work for a vertical tower. They cannot detect a hardware or minicomputer problem causing a service problem.

Customers' applications do not run in vertical trenches; they run across various platforms. Help desk tools must provide a view consistent with what a customer sees across all platforms. One example why vertical tools do not work any more is found in the operations of a circa-1989 help desk. For example, when a typical customer in 1989 phoned in a response problem, help desk personnel looked at the remote 3274, the modems, the fractional T1, the 3745, and the host application and quickly found the fault (e.g., a bad modem). This situation has changed. For example, when a customer today calls a help desk about a response problem, help desk personnel may not be able to find a problem and log the call as no trouble found. The problem could not be found because the customer is connected to a LAN, which is connected to a campus Ethernet backbone, which is hooked to a local ATM switch, which has a gateway on to a WAN system, which is mixed into a T3 circuit across to the 3745 and then to the host application.

This example illustrates two important facts. First, the level of expertise needed by help desk staff to solve problems has increased. Second, vertically specialized tools cannot solve such a problem. The solution requires tools that span vertical trenches.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Exploiting the Network. There is a wealth of information from the various network managers that are installed in an enterprise IS organization. This information needs to be integrated and correlated with the other service or platform information so that a true picture of the cause of a service problem can be identified.

Shared Data Bases. The various users of help desk tools must be able to share information. A specific local problem may need to be passed to the central help desk if it has not been serviced in the appropriate time frame. If enterprise problems should be located in one data base, why can they not be located in the help desk's management system as well? Technical services or support staff must be notified of problems and locations. Help desk staff needs up to the minute status on the state of a problem.

Capturing Data and Solving Problems as Close to the Source as Possible. An organization must be careful about the amount of service management information it places on a network. A network can be overloaded. By installing appropriate platform tools, local help desk services can manage their respective platforms and when there is an unsolvable problem, only pertinent information is forwarded to a centralized help facility.

Such an arrangement allows an organization to select the best tools for managing various platforms. As an example, if there is an MVS system, then an MVS automation tool would manage all except the most difficult problems. Perhaps an MVS problem is caused by a minicomputer node failure. The MVS automation tool can never know such information. However, by forwarding the problem from the MVS tool to a central help desk facility, it can be integrated with enterprise information from other tools and be output to a service display showing what the problem is and who is affected by the failure.

Multiple Levels of Control. With the capability to design a hierarchical structure of workstations for help desk staff, a manager or supervisor can monitor workstations, assign problems to each, and monitor elapsed time to completion. Complex problems can be shared among staff members and workstations.

The Help Desk's View

Exhibit 3 illustrates how automated help desk tools should look to help desk personnel. The exhibit also lists several specific characteristics of a central management tool that integrates all specific platform information from distributed tools, takes automatic action where appropriate, and correlates various sources of information into a coherent display for the help desk staff.

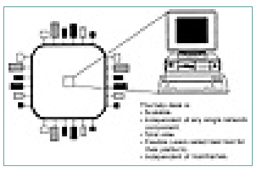


Exhibit 3 Help desk's view of itself in a distributed environment.

The tool proposed in Exhibit 3 is specifically designed for a multivendor environment. It allows help desk personnel to see the state of the entire enterprise at a single glance on a single workstation display screen. The entire enterprise can be managed from the data center to the networks and distributed systems and client/server applications that span technology platforms

The enterprise is an easy concept to understand. However, when it comes to implementing systems management tools within the enterprise, issues can be confusing. Management of enterprise systems requires that data center managers disassociate their thinking from traditional resource management. The tools' real targets are the services provided to customers (i.e., applications, individual sessions, or transactions).

Effective enterprise management systems must span the entire organization. A help desk staffer should be able to see instantly from a single workstation a component failure anywhere in the enterprise. Then, the help desk staffer should be able to select the next level of display for more detailed information and instantly select the appropriate tool to identify, diagnose, repair, reset, or dispatch the problem.

SUMMARY

A help desk is the interface to the data center that customers see. To a customer, a poor help desk usually means a poor data center. An organization's help desk can always improve. All help desk personnel must understand the strategic importance of their contribution to the success of not only the data center but the entire enterprise.

The success or failure of a help desk ultimately depends on its staff's skills. Therefore, help desk staffers must be experts. Also, an organization should anticipate the transition to enterprise management (i.e., a distributed environment). When building service management tools for a help desk, an organization must design end-to-end solutions. And finally, it must always choose tools that make use of automated processes.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Chapter II-9
Fine Tuning the Help Desk: Goals and Objectives

Christine B. Tayntor

When help desks were first introduced, their mission was simple: to help unsophisticated end users operate their personal computers and obtain a basic proficiency with the software installed on the personal computers. The challenges were straightforward. Not only were the personal computers standalone machines, but the software packages being used were normally limited to word processing and spreadsheets.

In retrospect, those were the halcyon days. Since then, the world of end-user computing has changed dramatically, with a new level of complexity being added in three key areas: hardware, software, and customer requirements.

In many cases, standalone computers have been replaced by personal computers linked into extensive local (LANs) and wide-area networks. This has increased both users’ dependence on their machines and the possibility that one person’s error could affect an entire community. It has also complicated the help desk’s responsibilities. To be successful, the help desk staff has had to expand its area of expertise to include at least the fundamentals of network operating systems.

Today’s software makes that of the early 1980s appear primitive, as vendors incorporate an increasing variety of features into their products. Word processing packages now include many of the tools that were previously found only in high-end desktop publishing products, and spreadsheets boast integrated spell checkers and query tools. In addition, many users have migrated from DOS to Windows, adding complexity as they run multiple products simultaneously. For the help desk staff, problem diagnosis has become substantially more difficult. Not only have they had to develop expertise in the new, more complex applications, but they now need to understand the relationships among the applications.

If the changes in hardware and software have been dramatic, they have been overshadowed by the evolution of the help desk's customer base. A decade ago, users were relatively unsophisticated and content to perform only basic functions on their computers, but today's customers have a far higher comfort level with personal computers. Many are highly computer literate and willing to experiment, stretching the envelope of packaged software. They are no longer content with simple directions on how to boot a computer or construct a basic spreadsheet. Instead, they want in-depth explanations of how to link a spreadsheet with a graphics package and tie them both into a word processing document, then send the results to a dozen people by electronic mail.

As customers' levels of expertise have grown, so too have their demands on the help desk staff. They expect nothing less than excellence and are frequently vocal if their expectations are not met. If the help desk is to be successful and continue to be an important part of the corporation, it must change, adjusting its services and skills to meet the new requirements.

THE ABCS OF FINE-TUNING

After a company has recognized the need for change in its help desk operation, the question becomes one of how and where to start. Three basic steps can facilitate an orderly transition to the help desk of the future:

- *Defining the ultimate help desk*. This is the vision of excellence for the future. It will become the benchmark against which performance can be measured.
- *Establishing the baseline*. This involves determining where the organization is now. Comparing current performance to the goals established in the first step can help identify the areas in which performance improvements are required.
- *Managing the evolution*. After the plan has been developed, basic management techniques are needed to ensure that the plan is implemented rather than becoming only a historical document.

The next section explores these steps and their tasks in more detail.

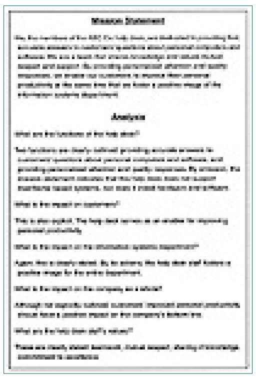
CHARTING THE COURSE

Before a company initiates any organizational changes, it should have clearly established goals and a plan to achieve them. Without such a formal vision, change is unmanageable and may have unpredictable results. Three fundamental tasks are required to define the ultimate help desk: establishing a vision, identifying the services to be provided, and defining optimum service levels.

Establishing a Vision

One of the key elements in defining the future is to clearly outline the functions that the help desk can and cannot perform. Before this level of detail can be reached, however, the organization needs an overall framework — a statement of who and what it is. In short, the help desk needs a mission statement.

In the simplest terms, a mission statement is a short written summary of a group's purpose and values. It explains what the group does and how those functions affect its customers, its own staff, and the company as a whole. An analysis of a sample mission statement for a help desk is included as Exhibit 1.


Mission Statement

Analysis

Exhibit 1 Mission statement and analysis.

Identifying the Services to Be Provided

Although the mission statement addresses services in general terms, management should define them very specifically, because what the team is expected to do directly affects both the staffing level and the skills that are required. Specific questions should be answered at this point.

Is the help desk responsible for resolving the majority of all problems, or is its primary function to route questions to the appropriate expert? The answer to this question determines how large the staff must be, because solving a problem takes substantially longer than simply routing it to another person. Similarly, the more problems the help desk is expected to resolve, the greater the depth of product knowledge and training required.

Although additional staff and more-extensive training may appear to have negative implications, positive effects can be realized if the help desk is made responsible for problem resolution. When responsibility is centralized, synergies can occur. The help desk staff is normally located together, allowing the members to share information and assist each other in solving problems. This is not always true of the individual experts to whom problems are forwarded. When knowledge is pooled, many problems can be resolved more quickly, thereby reducing costs and improving customer service.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Another benefit, and one whose value should not be underestimated, is the job enrichment that results from resolving problems rather than simply logging and forwarding them. For many organizations, the help desk is a group with high staff turnover and early burnout. Both of these problems can be minimized by giving the help desk more responsibility — changing its role from that of a conduit, forced to deal with irate customers but rarely being able to create customer satisfaction, to one of a problem solver, with the concomitant rewards.

Should the help desk provide services other than problem resolution? Some companies have expanded the role of the help desk to include such functions as installing and upgrading software, performing preventive maintenance on personal computers and servers, and providing product training to customers. This expansion of services can have both positive and negative effects on the organization. On the negative side, if the help desk is to assume additional responsibilities, increased staff and more training are required.

From the staff's view, the added responsibilities can be either positive or negative, depending on the personalities involved and the corporate culture. The new functions can be perceived as job enrichment and a way to learn more about the hardware and software the team supports. They might also be viewed as diluting the primary mission of the help desk, which is resolving problems.

Defining Service Levels

After management has decided which services the team should provide, the next task is to determine performance levels. It is not sufficient to state that the help desk can respond quickly to customer inquiries, because "quickly" is a subjective term. Instead, the help desk should have specific, measurable performance targets. Examples include:

- Of all calls, 95% should be answered before the third ring.
- Of all problems, 93% should be resolved with no callback (i.e.,customers receive their answer during their first call to the help desk).

Because help desk performance is measured against the target performance levels, those service levels must be achievable. One way to establish acceptable performance measures is to benchmark other companies, determining the levels they have set and routinely meet. Like all benchmark efforts, several caveats are characteristic of this one:

- The companies' environments should be similar. A help desk that supports dumb terminals attached to a mainframe running core transaction processing systems has different performance expectations from one that supports personal computers attached to LANs running personal productivity tools.
- The companies being benchmarked should be recognized as providing above-average or excellent service levels. Learning how a mediocre performer sets its goals does not facilitate developing a world-class help desk.

Although it is important for service levels to be achievable, it is even more critical that they meet customer expectations. Too often, management has developed targets in a vacuum without consulting the end user. Although some companies are reluctant to ask for customer input to the goal-setting process, this is an essential step toward improving satisfaction. One way to obtain end-user requirements is with a customer survey. A sample survey section is included as Exhibit 2.

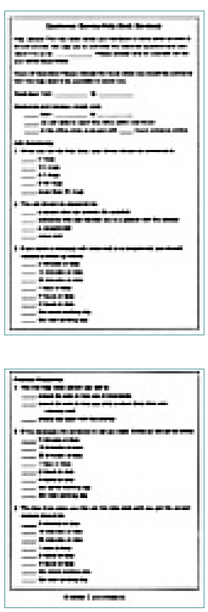


Exhibit 2 Sample help desk customer survey.

ESTABLISHING AND EVALUATING THE BASELINE

After the first step is complete, the organization knows what it considers to be a fully functional help desk and is typically anxious to proceed with the implementation of that ultimate vision. An essential intermediate step exists, however, that must not be ignored. The organization must measure its current performance and compare that to the goals established in the setting of the vision. Without this step, the organization is proceeding blindly, making changes without being certain of their impact or knowing whether those changes are treating symptoms rather than resolving underlying problems.

Before an effective implementation plan can be developed, the organization must complete three key tasks:

- Establish a baseline of current performance.
- Determine how large a gap exists between the present help desk and the ideal that was identified.
- Perform a gap analysis (i.e., determining the underlying causes of the discrepancies).

For those companies that have implemented formal metrics programs for their help desk, measuring current performance presents no challenge. They already know how many calls they handle per day, how long the average caller waits for a response, and what level of satisfaction their customers have with the answers they are given. Many other companies, however, have few if any statistics about the help desk and must compile them.

Although automated tools can simplify the collection of metrics and increase the accuracy of the data, statistics can also be gathered manually. At a minimum, each member of the help desk staff should record the number of calls answered each day, how many are forwarded to someone else for resolution, and how many require follow-up. Although it is more difficult to compute manually, it is also helpful to calculate the minimum, average, and maximum length of time before a call is fully resolved. These are important metrics that are used to measure future performance.

Customer satisfaction, which is a key element of current performance, can be obtained either by interviews (e.g., in person or by telephone) or through a survey. One approach is to expand the customer requirements survey shown in Exhibit 2 to include two columns, one to record the current situation and the other to indicate the customer’s needs.

For organizations that have not yet implemented metrics programs, several caveats apply:

- Be honest with the staff. Measurement is frequently perceived as an attempt by management to identify and punish poor performance. Although it is impossible to completely allay these fears, management should explain that the statistics are used to measure improvements for the group as a whole, rather than for individuals, and that nothing is inherently good or bad about the baseline. It is simply a measurement of the current situation. Future performance and widening the gap between that and current performance are key.
- Do not rush. A meaningful baseline cannot be established in a day or a week. To ensure that seasonal swings in activity (e.g., the implementation of a new software release) do not skew the statistics,

they should be gathered over a period of at least a calendar quarter.

Establishing a baseline is not easy, but it is essential to the success of the help desk's fine-tuning program.

After the goals and the baseline have been established, it is relatively simple to determine the gap between them. The third task, determining the causes for these discrepancies, requires substantially more time and effort. The proper focus, however, can minimize the length of time needed.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Because of the type of work the help desk performs, only three key variables in performance exist: the staff, the automated tools they use, and their customers. Because the help desk cannot mandate changes in its customer base, the fine-tuning program should focus on the help desk and its tools. By analyzing each of these, the manager can identify areas for improvement.

Specific questions that can help pinpoint problems concerning the help desk staff include:

- Has everyone been fully trained on all supported products? Without adequate technical training, the staff requires more time to resolve problems or may be forced to refer calls to a second-level support person.
- Does the staff have fully developed customer service skills? Although most companies recognize the need for technical training, few provide basic customer service education to their help desk staff. If customer satisfaction is less than desired, one cause could be inadequate training. Commercially available courses (in both instructor-led and audio-video tape versions) demonstrate the correct way to answer the phone, to deal with unhappy callers, and to elicit the key information for problem resolution without annoying the caller.
- Is the staff skilled at providing training? In addition to resolving problems, the help desk staff is also expected to provide informal training in the use of both hardware and software products, yet they are frequently given no guidance in the best ways to help adults learn. Like all other skills, this one can be enhanced by formal instructions — in this case, train-the-trainer courses.

- Is the staff large enough? Although this is frequently the first staff-related question a manager asks, it should be the last, because the inability to respond quickly enough to calls can be caused by lack of training as well as a shortage of people. After the manager is convinced that the staff is fully trained and operating at peak efficiency, if response targets are not being met, the probable cause is too few people in the help desk group.

Specific questions that can help pinpoint problems with tools include:

- Is the staff using automated tools for problem tracking and resolution? A comprehensive system is a key element in establishing the help desk of the future. With the increased demands being placed on the help desk, manual methods of tracking problems and documenting their resolution are no longer sufficient. Although they can accomplish the basic goals of measurement, manual methods are labor-intensive, diverting the help desk from responding to customers.
- Is the automated tool a complete one? Simply tracking the number of calls and calculating the time required to resolve them is not enough. A comprehensive help desk system should:
- Provide for and require documentation of all steps taken to resolve a problem, the person performing them, and the time required for each. This establishes accountability and allows the manager to identify potential areas for increased training by quantifying the length of time each individual needs to resolve a specific type of problem.
- Facilitate the management of solutions-in-progress by providing automated reminders of open calls. This feature eliminates neglect of unresolved problems caused by the constant stream of new calls.
- Escalate unresolved problems. In a comprehensive call tracking system, each problem is assigned a severity code, which determines the length of time it can remain unresolved without being escalated to the next level within the organization. This ensures that critical problems are given priority and alerts management before they become disasters. Customer satisfaction and confidence in the help desk typically increase when the system provides automated escalation.
- Is problem resolution aided by maintaining an easily accessed history of previous problems and their solutions? The goal of such a system is to share expertise and expedite problem handling, thus increasing both the help desk’s productivity and customer satisfaction.

MOVING FORWARD

The primary responsibility of any help desk is to efficiently and economically service the needs of the computer- and communications users of an organization. Over time, however, the help desk must be fine tuned to keep pace with the ever-changing needs of those users. After the need to make changes in the present structure has been established, the processes of evaluation and refinement can begin.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter II-10
Milestones in Implementing Help Desk Automation

Marc Tower

For many help desks swamped with too many calls and too few support agents to take them, help desk automation glimmers like an oasis in the desert heat. In fact, help desk software and other automation tools offer organizations the ability to increase the productivity of their agents, improve customer satisfaction, and store problem resolutions for reuse.

Two support organizations with similar characteristics can, however, select the same help desk automation tools and have vastly different results from their use. What makes the difference in successful implementation of help desk automation?

If you examine the successes and failures help desks have with support tools, a common theme would emerge as the likely reason: expectations. Those organizations that implement tools successfully view them as a means to an end, and their support processes are characterized by careful planning and discipline in following a process. Those who have failed in their efforts view tools as the end in and of themselves, expecting to realize the benefits of automation immediately upon the successful installation of those tools.

PROCESS IS EVERYTHING

As advanced and easy to use as most help desk software solutions are today, they still require a high level of discipline in following a process — in this case, the process of answering and responding to help desk calls — as well as an organized approach and project management skills to solve the problems called in. To put it another way, the product is not the process.

Many factors outside the functionality of the tool set impact its success in a support organization. In the course of attempting to automate their help desk, most organizations follow the same path (see Exhibit 1), which is characterized by three milestones:

- Call logging and tracking, which entails the recording and management of requests for support.
- Knowledge capture and reuse, which involves storing and retrieving problem resolutions.
- Measurements and reporting, which includes defining, tracking, and improving key measurements.

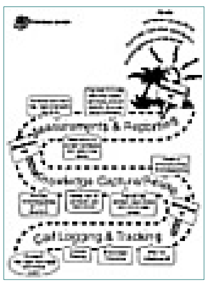


Exhibit 1 The road to help desk automation.

MILESTONE #1: CALL LOGGING AND TRACKING

Support organizations trying to manage calls without the benefit of support automation are sure to have problems. Help desk software solutions can greatly improve responsiveness by allowing support agents to log calls and track them through to closure.

Most call logging and tracking tools have finely crafted, intuitive interfaces that make using the tool far easier than manual methods. This aspect of help desk automation is the first and most easily implemented milestone in the journey toward a fully automated help desk.

Expectations for call logging and tracking automation are usually right in line with the outcome — the elimination of all the yellow sticky notes tacked up all over the workstation of the support agent, and their replacement with records in a database. Call logging and tracking can begin almost immediately after successfully installing the help desk software. Little training, process discipline, or noncomputer project management skill is required to learn how to use the tools for this function or sustain their effective use.

Most help desks successfully arrive at the call-logging and -tracking milestone. Disillusionment sets in, however, when they go no further. Failure to plan halts some automation efforts at this point. Too often, assumptions are made that all the other desired outcomes are just foregone conclusions once help desk software for logging calls is installed. To get past this milestone and achieve success in fully automating the help desk, help desk managers must know what all the milestones are and have a plan to reach them.

MILESTONE #2: KNOWLEDGE CAPTURE AND REUSE

What sells more help desk software than anything else is the promise of capturing knowledge for reuse. The expectation for this milestone is to have a rich and current knowledge base that allows inexperienced support agents to:

- Deliver support in areas where they do not have sufficient technical expertise.
- Avoid resolving the same problem with multiple support agents in the organization.

Capturing problem-solving expertise in a system instead of relying on the "gray-matter database" can improve callers' perception of the help desk staff's competence. If the organization's star support agent should be suddenly unavailable, the advisability of making knowledge available to newer, less-skilled support agents would become very clear.

A concerted, up-front management effort is needed to implement this part of the process. Support agents with the expertise to solve problems may feel that they do not need to put that knowledge into a database for reuse. They already know how to solve most of the problems they encounter. For them, the process they use to manage the call (the help desk software) and the process they use to solve the problem (what is in their head) are not the same. It seems redundant to them to take the time to store the problem and its solution in the knowledge base. This leaves the new or inexperienced support agent out in the cold.

The help desk automation project has to be managed to the desired outcome: the creation of a rich, current knowledge base of solutions. Training and the establishment of a process to populate the knowledge base are therefore necessary steps to reaching this milestone.

Soliciting the Cooperation of Experienced Staff. The first step is to obtain the buy-in of experienced support agents. Help them understand the ramifications of having a comprehensive knowledge base: it will mean fewer interruptions for them as less-experienced agents find answers in the knowledge base instead of tapping them on the shoulder. It is also important to create an expectation that experienced support agents are, in fact, key suppliers of information needed for the knowledge base.

Populating the Knowledge Base. The next step is to "preload" some key problem resolution data into the knowledge base. Too often, organizations plan to do this populating later, after giving their support agents time to become familiar with the new software and log a few calls. The problem then becomes one of negative reinforcement. The staff will have heard the hype about having a knowledge base so their curiosity naturally leads them to try it out. If they find nothing when they do their first few searches, they conclude the knowledge base is of no value to them.

One suggestion is to consider interviewing all experienced problem solvers before installing the help desk software. Have each of them create their top 10 list of problems and solutions, and then ensure that this information is captured in the knowledge base so that it will contain relevant content from the moment it first comes online.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Keeping the Knowledge Base Up-to-Date. Finally, remember that it is human nature to do what is inspected, not what is expected. Set forth expectations for capturing knowledge in the form of a measurements; likewise, set goals for support agents.

From a management perspective, it is useful to think of the knowledge base as if it were an inventory to be managed. There should be regular turnover, with obsolete solutions moving out of the knowledge base and new solutions taking their place. A static knowledge base that rarely grows indicates that it is not a key part of the support process.

Overall, the size of the knowledge base should grow at a modest but predictable rate. The help desk automation tool should include a reporting facility that monitors the status of the knowledge base. Make a habit of measuring the size of the knowledge base, the average age of the solutions stored in it, the date of the last update, the frequency that the knowledge base is referred to, and who is (or is not) contributing solutions routinely.

To gain a competitive advantage and begin to achieve a return on investment from the help desk automation effort, it is imperative to succeed in reaching this milestone. An organization will have made significant progress toward having a fully automated help desk when a solutions knowledge base is in place. However, to completely realize a payoff from the investment, the measurements and reporting milestone has to be the next goal.

MILESTONE #3: MEASUREMENTS AND REPORTING

The failure to establish and track key measurements can label any support operation ineffective. This differs sharply from the pre-implementation expectations for help desk automation; the output at this stage of the implementation should direct the help desk organization to all the appropriate actions needed to improve customer satisfaction.

Planning and discipline in adhering to a process are perhaps most important for reaching this milestone successfully. Measurements are a report card that reveal how effectively and efficiently the help desk responds to calls. Measurements point to areas where improvements are needed and also highlight where the process is working well.

Defining and Tracking Key Measurements. A meticulous planning effort is necessary to determine all the key measures of the support operation. Think of these measures as the vital statistics that must be checked regularly to maintain the health of the call support process. Start by answering these basic questions about the support process:

- What is the monthly call volume?
- What is the average time to close a call?
- What is the oldest open call?
- What is the most frequently occurring problem?

The answers to these questions cannot be speculative, but must be based on actual data. If a help desk software package is used to track calls, then all the raw data needed to report on these key measurements should be available.

Distinguishing Effectiveness and Efficiency Measurements. Determining where the organization wants to be in terms of help desk automation requires identifying the measurements that are indicators of the help desk's effectiveness and efficiency — and making sure that data gathered during call logging and tracking supports these performance criteria.

Efficiency measurements are important, but they are based on factors that internal management cares about and needs to see: cost, workload, and productivity. By contrast, measurements of effectiveness are those that are externally focused and customer-based and reflect what customers care about: responsiveness, accuracy, empathy for their problems, reliability, and deliverables.

Some of these measurements are straightforward and easy to derive; others are more difficult to measure. For example, how is empathy for callers' problems measured? It poses a challenge for sure, but if it is an important performance characteristic for the help desk, it is imperative to know that early on, because the organization's measurement needs should determine the way data is gathered during the initial call logging and tracking process.

If help desk software is already in place to log and track calls, take a look at the data that is actually being gathered during this process. Does it support the management- and customer-focused performance measurements identified as important for the help desk? If the answer is no or not completely, then it will be necessary to change the way data is gathered during call logging and tracking. Remember, start by knowing the destination — that key set of measurements needed to evaluate and improve the performance of the help desk function.

Reports Analysis and Required Skill Sets. With measurements defined, the next step is to produce reports that regularly track key measurements. Depending on the type of measurement, reports may be required hourly, daily, weekly, monthly, or quarterly.

Reports are used to determine if results meet or exceed objectives. Reports are used to group problems by type as well as their frequency of occurrence so that actions can be taken to prevent future occurrences. Most organizations will not have the resources to solve all problems identified all at once, but will attempt to direct their efforts toward solving the most frequently occurring ones first.

In addition, the tracking of measurements from period to period can be used to establish a trend, which may be positive, substantiating improvements, or negative, pointing to flaws in the support process. It may take six or more sets of data to reliably predict trends, identify root causes of all problems, and identify meaningful corrective actions. Knowing how to interpret measurements is simple; doing so faithfully to produce an improvement plan and monitor its progress over time takes dedication but produces a world-class support organization.

| Previous | Table of Contents | Next |

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

A person with analytical skills and a statistical background can be very helpful here. Assign these measurement and reporting duties to someone who is comfortable working with data — analyzing, interpreting, organizing, and communicating it to the organization. Furthermore, this person needs to have firm convictions about the conclusions drawn from the data and must demonstrate assertiveness, to make sure such conclusions are not ignored but rather actually implemented as part of an improvement process.

The measurement and reporting milestone is the one that eludes so many organizations that aspire to a fully automated help desk. Companies frequently put the first two milestones behind them, but fail to grasp the importance of measurements and reporting. Trying to manage a support operation without identifying, tracking, and reporting on key measurements is like trying to drive a car without the benefit of the dashboard.

Making Time to Do Adequate Planning. Some organizations may be about to embark on an initiative to automate their help desk. Or perhaps their earlier attempts to automate stalled after the first or second milestone. No matter where your organization stands, there is a way to ensure the future success of a help desk automation effort.

The single most important action to take when implementing help desk software is to start with the end in mind. The road map metaphor applies here. To reach a destination, a driver would unfold the road map, locate the destination, then begin plotting a course to get there. As often as not, support organizations fail to apply this wisdom to their automation efforts.

These failures are not because technical support managers lack common sense. On the contrary, they are some of the most resourceful professionals in the organization. However, they frequently do not have time to deal with anything that is not urgent. Therefore, the greatest obstacle in any efforts to reach the third milestone of help desk automation implementation is to find the time to adequately plan.

A MODEL FOR QUALITY IMPROVEMENT

A fully automated support process makes delivering support to users look very easy, but getting the support process in place is a significant undertaking that requires project management, planning skills, and discipline in monitoring progress. Although help desk software solutions are the backbone of these processes, selecting the right package for your organization is only the first step leading to an automated help desk.

Data on problem resolutions that has been captured in a help desk knowledge base can be used to justify resources and improve responsiveness. In turn, improvements implemented at the help desk will ripple through the entire enterprise. A help desk that tracks key measurements can become a model in an organization’s overall efforts to improve quality.

Previous	Table of Contents	Next

HOME ACCOUNT INFO SUBSCRIBE LOGIN SEARCH MY ITKNOWLEDGE FAQ SITEMAP CONTACT US

SEARCH ITKNOWLEDGE

Go!

Keyword

Brief Full

- Advanced Search
- Search Tips

BROWSE BY TOPIC

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Bookmark It

Search this book:

Go!

Previous | Table of Contents | Next

Chapter II-11
Automating the Data Center

John L. Connor

Data center automation packages are designed to help data center personnel manage their center’s operations. These systems handle such ordinary operations as operator messages, system start-up and shutdown, and simple time-initiated processing. Many have additional features.

Packages that solve the most complex problems are also generally the most difficult to use. However, if they have been properly selected and implemented, such automation products are still reasonably easy to use.

Exhibit 1 lists the advantages of automating as many operations as possible. Manual tasks that have been automated are performed with increased speed and accuracy. Consequently, automating the manual tasks required to run a company’s data center hardware increases the availability, speed, and accuracy of the applications that run on it.

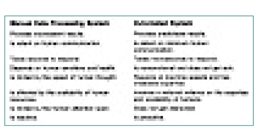


Exhibit 1 A comparison of manual and automated systems.

PLANNING FOR AUTOMATION

Many organizations make the mistake of selecting an automation tool and then trying to make their automation needs fit this product. This process is backward; the objectives and other key requirements of the data center should be defined first. Then, the data center manager should select a product that can accomplish the required objectives in the appropriate time frame. By following the nine-step plan described in this chapter, data center managers can ensure that they select the correct product and implement it properly.

The amount of time they take to complete each step in the plan depends on several factors, including how many people can work on the step, the coding requirements of any automated procedures involved in the step, and the automation process selected. The data center manager need not complete all nine steps without pause. Rather, the manager may consider steps one through four as one phase and go on to subsequent steps at a later time.

The Importance of a Champion

A critical component of the success of an automation project is the project manager who implements it. This person must inform and persuade the company's decision makers that automated operations will greatly benefit the company. Who should this champion be? Ideally, it should be someone in management, but anyone can initiate interest in automated operations.

The project manager must clearly explain all the possible problems involved with the automation project. The manager must try to ensure that everyone who will be involved in the project looks forward to its challenges and the opportunity to work together it presents for management, applications, operations, and systems personnel at all the company's information processing locations.

STEP ONE: CREATING THE IMPLEMENTATION PLAN

The first step of introducing advanced automation is the development of an automation plan. This plan should be created and reviewed by all who will participate in the automation project. Because the project will demand human and capital resources from many areas of the company, both management and technical personnel should contribute to the plan. However, the plan should be managed by only one person.

Because an automation project is not completed overnight, the key groups involved in it must communicate regularly to stay abreast of the project's progress. The main reason automation projects fail is usually that plan participants have failed to communicate such matters as:

- What each phase of the project is trying to accomplish.
- Why the tasks in each phase are being performed.
- How the plan's progress will be measured.
- What task comes next.
- What the benefits of the automation project are to a particular department or organization or to the entire company.

Plan Contents

The automation plan must include the items discussed in the following paragraphs.

A Mission Statement. The mission statement is a broad statement by senior management that reflects the intent of the automation project. This statement sets the general aims of the project and demonstrates that the project has the commitment of senior management, which is fundamental to its success. The following is a sample mission statement:

> To remain competitive in the industry, the company must maintain online service times and availability under increasing transaction volumes. This must be attained within the current budget for hardware and software.

The mission statement should provide only company direction and commitment. The project's objectives and how-to information are provided by other levels of management and technical personnel in other steps of the plan.

Data center managers who do not have a mission statement from senior IS management must create one. They must also inform senior IS managers of the project's status as it progresses. One way or another, the data center must gain senior management's active support, or the automation project will turn into just another operations or systems task—such as performance monitoring—instead of a corporatewide mandate to use automation to improve all the corporation's information processing functions.

Plan Objectives Ranked by Priority. The automation plan must specify what tasks must be accomplished and the order in which they must be performed. These objectives can include:

- Improvement of the availability of data center services.
- Maintenance of data center personnel at current levels.
- Elimination of the center's third shift or of a weekend shift.
- Reduction of the number of consoles.
- Remote operation of a data center.
- Reduction of console message traffic by 80%.
- Standardization of policies.
- Creation of a focal point for operations.
- Simplification of command structures.
- Balancing of workloads to make use of all computer resources.
- Integration of network management system and subsystem automation.
- Automation of the management of IMS or CICS.

| Previous | Table of Contents | Next |

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Project Audit. To achieve a concrete sense of the necessary scope of the automation project, plan developers must conduct an audit to determine the following:

- What is the current status of data center automation?
  - — What hardware and software does it have?
  - — What has been done to automate data center operations?
- What are the goals of the current automation project?
- How much effort and time can be spent on this project?
- How can the benefits of this project be maximized with minimal effort?

Exhibit 2 is an audit checklist of one of the major areas for automation in data centers using MVS systems. A data center manager should use this checklist or others modeled on it to audit each system and subsystem that is a candidate for automation. The manager should then rank these systems and subsystems according to the order in which they should be automated. Viewed together, the information in these checklists gives the data center manager a broad picture of what in their facilities should be automated, the order in which this automation should occur (both among systems and among functions within systems), the effort necessary to automate, and what the result of automation will be.

Exhibit 2 Sample automation audit checklist for MVS systems.

The data center manager may someday decide to automate functions that were not chosen for automation on these checklists. This is not a problem; such a decision simply takes the data center to a more advanced stage of automation.

Skills Audit. Many operator skills used today to manage a data center will not be needed when data center operations are automated. New skills will be required instead. The data center manager must therefore plan now to train, outsource, or hire personnel with these new skills.

Exhibits 3 and 4 are sample forms that data center managers can use to take stock of those skills that will and will not be required and those that will be needed to a reduced extent after automation. Before starting an automation project, the data center manager should audit the skills and people in the data center organization—both those that are available and those that will be needed. This information is necessary to complete the automation plan.

Exhibit 3 Sample skills audit form for skills or tasks that will change.

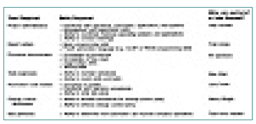


Exhibit 4 Sample skills audit form for tasks requiring new skills.

STEP TWO: LEARNING THE PRODUCT

An automation product usually has many capabilities. To become familiar with them, the data center manager should review the product's documentation. Following a demonstration script is also a useful way to learn abut a product's features.

A data center manager who decides to evaluate two products should create a detailed list of the tasks that each product must automate. (For example, a data center manager who highly values IMS may include questions about how each product's IMS automated operator interface exit works, the operator productivity tools that the products offer, and how the tools interface with the data center's performance tools.) The manager should give this list to each product's vendor along with a time by which the vendors must respond to the manager. The vendors' responses to the list may help the manager choose between the products.

Once the data center manager has selected the product to accomplish the objectives defined in step one of the automation project, the manager should install the product in a test system and train all personnel on the product. Then, the manager should move on to step three.

STEP THREE: ANALYZING AND AUTOMATING MESSAGES

Message rates have been skyrocketing. Each additional million instructions per second involves an additional two or more messages per second on the console. Fortunately, automation can eliminate most of this traffic.

To understand the company's need for message management and simple message automation, the data center manager should perform the following tasks:

- Analyze messages for suppression.
- Analyze messages for the automation of simple replies.
- Analyze messages for message rewording.
- Analyze network messages.
- Analyze messages for console consolidation (i.e., alert management).
- Analyze command procedures to simplify them.

During each of these tasks, the data center manager should seek the input of various sources, such as operators, operations supervisors, applications personnel, technical support, technical services, and the help desk.

Because of the number of people who should be consulted about the nature and sources of company messages, the data center manager should consider setting up a series of meetings to discuss messages and their automation. Face-to-face meetings can strongly contribute to the success of automation projects. The manager must determine the attendance and format of each meeting.

Once step three has been completed, the amount of traffic coming across the console should be reduced by 60% to 90%. Now the manager can move on to step four.

STEP FOUR: AUTOMATING PROCEDURES

During this step, the data center manager examines the procedures by which data center activities are conducted in order to eliminate or automate them. Automating the execution of routine procedures can:

- Reduce the need for personnel during certain periods of activity.
- Reduce system and subsystem outages.
- Reduce the occurrence of problems stemming from operator errors.

Operator errors result from complex procedures followed incorrectly, syntax errors, and slow responses to error messages. Putting routine procedures into a precoded procedure and making them readily accessible or executing them automatically can solve these problems.

An operator's procedure or run book is an excellent source of information on these procedures. If no such book exists, a little more work may be required. Someone must compile the unofficial procedures being used to manage the data center.

Next, the data center manager must eliminate unnecessary procedures, identify procedures that can benefit from automation or from reduction to a few keystrokes, and match those procedures requiring automation (including those procedures yet to be developed) with the people responsible for coding and documentation.

Procedures can be automated in three ways. First, if a sample product exists in the automation product's solution offering, the data center manager can choose to customize the procedure, modify the procedure, designate an employee to modify it, or ask outside support to call in. Second, if the product offers no similar procedure, the data center manager can choose to convert this procedure from the operations procedure book into an automation procedure, ask the employee responsible for developing the procedure to automate it, or ask outside support to automate it. If a manual procedure is not documented and no similar procedure exists in the automation product's solution offering, the data center manager can designate an operator to automate it, ask the employee responsible for developing this procedure to automate it, or call in outside support.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

STEP FIVE: AUTOMATING FUNCTIONS THAT CANNOT BE ACCOMPLISHED FROM THE HOST

Some aspects of automated operations cannot be performed from the host because it is not running or because necessary resources are not readily available on it. These functions are performed by an outboard workstation running on a microcomputer. An outboard can provide one or more of the following automation functions:

- Automation of IPL and IML.
- Heartbeat checking.
- Pager support.
- Remote operation.
- Environmental monitoring.
- Supplemental workstation functions.
- Remote dial-up for support purposes.

A data center manager who wants to automate functions on an outboard must understand its hardware needs. Many of the tasks in the foregoing list require the use of microcomputer boards. The cost of an outboard with a full complement of microcomputer boards can range from $5,000 to $10,000, and the price of two or three outboards can exceed that of mainframe automation software.

Before acquiring an outboard, data center managers should research their needs carefully. Important questions to consider include the following:

- What are the general hardware requirements for the outboard (e.g., CPU, memory, disk space)?
- How is code created on it?
- What language is used?
- Where is the code executed?
- How easy is it to use?
- How many connections does it have; what is its growth potential?

STEP SIX- REDUCING OUTAGES

Because it may be impossible to completely eliminate system or subsystem outages, a realistic data center manager may choose simply to try to detect and recover from them as quickly as possible. To meet this goal, the manager needs an automation product with a continuous state manager.

The continuous state manager allows the data center manager to build a model of the expected state of subsystems at any moment in time. When the state of a system in operation does not match the expected state of its model system, the automation product compares the two and takes a previously defined action. A continuous state manager thus frees operations staff from manually monitoring CICS, IMS, or other subsystems. The manager performs this task automatically according to the continuous state manager’s preset instructions.

In addition to implementing an automation product with a continuous state manager, data centers managers who wish to reduce system outages should understand what problems in their organizations cause outages. Managers who do not know should try to find out.

STEP SEVEN: AUTOMATING PERFORMANCE

Although a system may issue warning messages for such problems as a console buffer shortage, it does not warn of all potential performing problems. A better source of this information is a performance manager. Not only do performance managers issue messages that warn of potentially critical problems, they can also authorize outside services to collect the additional information required to accurately diagnose them. By using a performance manager along with an automation product, the data center manager can therefore automate the analysis and solution of critical and complex performance problems.

Performance monitors provide two types of messages that notify the automation product of performance problems. The first indicates when a system does not meet a response time objective. The second indicates when resources are not performing according to expectations. The implementation of performance monitors, therefore, allows automation products to increase throughput and manage service levels proactively.

One Task at a rime

A message indicating that a response time objective has been exceeded usually indicates a performance problem. Data center managers should therefore automate the response to messages indicating resource performance problems first. By doing so, managers minimize their messages for service-level exceptions, because they have automated the resolution of a problem that might have delayed response times.

Because some solutions to performance problems are quite complex, data center managers should consider implementing performance automation in two stages, especially if their time or resources are limited. The first stage is to ensure that warning messages are very visible so personnel can immediately respond to them.

The second stage is to install automation that reacts to messages. This reaction could be as simple as alerting a specific person responsible for the problem area or automatically starting data monitoring of the problem area to collect information for later review. Or the reaction can be much more sophisticated, involving additional analysis and action based on this analysis.

The next task is to identify all the important messages that are generated by the performance manager warning services and the course of actions that should be taken in response to each message. Possible courses of action for any of these messages include the following:

- Variable modifications to a solution in the solution pack.
- Development of a new solution.
- Creation of an alert to be routed to a central terminal.
- Creation of an alert and a follow-up procedure or solution.
- Routing a message to a TSO user.
- Phoning a beeper.

Of course, some combination of all of these actions is also possible.

When reviewing possible courses of action, data center managers should identify those that require too much effort to be implemented at the current time. Managers should schedule the implementation of these solutions in the future. They should also identify messages that indicate problems for which they do not anticipate automating solutions for now but on which they would like to have documentation for further reference.

Clearly, automating performance is a complicated, multistage process. A form that data center managers can use to organize this process is presented in Exhibit 5.

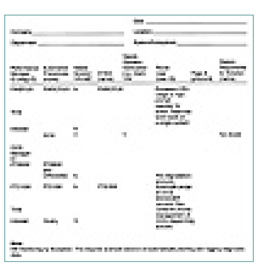


Exhibit 5 Sample performance automation task form for increasing throughput and managing service levels.

Many performance problems can be minimized or prevented from becoming critical if prompt action is taken when a performance monitor product’s warnings are detected. On this form, the data center manager should list the messages that the data center’s performance monitor uses to indicate potential service degradation.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Then for each message, the manager should review all vendor documentation to see whether an automated solution has already been provided. If it has not, the manager should identify what automated action should be taken. To decide this, the manager should ask the following questions:

- Does the message indicate a problem that should be solved?
- Does it need only an alert?
- Does it need a solution and an alert?
- What system ID should the alert be routed to?
- Should a user be notified?
    - — At the user's working terminal?
    - — By beeper?
- Who should develop the solution?

STEP EIGHT: AUTOMATING RECOVERY

There are two important points to consider when planning for systems recovery. The first is that automation does fail. People forget this fact until there is an automation failure. Then panic sets in as personnel rush around trying to recover from it.

The second point is the loss of operating skills a data center suffers when it becomes automated. An operator in a nonautomated environment may perform an IPL several times a month. In an automated environment, the operator may never do an IPL unless there is a failure. How can an operator who never does IPLs be expected to then do one?

The solution to these problems is for the data center manager to have an operations automation development methodology in place before writing automation routines, implementing a vendor's solutions, or building rules.

Oddly enough, advanced automation requires more commitment to documentation and audit requirements than do traditional methods of operating a data center. Automation has been touted as eliminating the need for operator documentation, but in reality, the information maintained in programmed solutions is much more complicated than the simple human procedures it was written to replace. Luckily, the audit requirements are much simpler to enforce in a well-automated data center than they are in a center managed exclusively by human operators.

STEP NINE: ENTERPRISEWIDE AUTOMATION

The last step in this process is the biggest and the most complex to complete. In it, data center managers apply the steps they took to automate their data centers' mainframes to the automation of all data center systems.

To do this, a manager needs an automation product different from the one that automates the mainframe. This product should allow the manager not only to manage hardware and software by such important vendors as IBM Corp., Digital Equipment Corp., and Hewlett-Packard Co., but to integrate the management of these platforms from one central location.

The implementation of such an advanced automation solution requires additional capabilities in the automation package. They include the following:

- The ability to communicate between systems.
- An alert management facility to facilitate the collection of alarms or warnings.
- The means to automate processes without using a mainframe.
- An easy-to-use, open systems procedural interface to other systems. Scalability for growth.
- Sufficient processing power to manage large networks and the devices on them.
- Flexibility and ease of use.

SUMMARY

In recent years, information systems have grown in size, speed, and complexity. They can perform an incredible variety of work on many different platforms. The operations environment has therefore become increasingly complicated and prone to operating errors.

However, major advances have been made in the area of proactive systems automation, and automation software is available for most environments that run in a data center. The next stage of preventive automation will be yet another step toward predictive systems that act as automated systems managers. Vendors will have to ensure that these automation solutions can be implemented without the need for major programming efforts on the part of the implementation staff.

Every day, increasingly complex forms of automation are introduced to the marketplace, and companies find additional sources of information to process. The combination of large amounts of data and complicated installation requirements offers a new challenge to the vendors of automation products—a challenge they must accept to survive if they want their products to be selected for the sophisticated automation tasks of the next decade.

Previous	Table of Contents	Next

Products | **Contact Us** | **About Us** | **Privacy** | **Ad Info** | **Home**

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Section III
Enterprise Systems Planning and Management

In the past 10 years IT computing has undergone a transformation from a centralized function to one that is distributed throughout an organization. This transformation began with the advent of client/server computing and has accelerated with the development of Internet, World Wide Web, and intranet technology.

As computing has changed, so has the data center. It once was where the mainframe was kept. Mainframe systems are still in the data center, but so are minicomputers and servers that connect the systems distributed throughout the enterprise. The transformation from data center operations to enterprise systems operations has greatly changed data center managers' responsibilities. To meet these responsibilities, data centers must be fluent in many different technologies so they can plan for them and subsequently manage them.

The first chapter in this section "Integrating the Web and Enterprisewide Business Systems" discusses the component Web technologies needed to integrate the World Wide Web into existing enterprise systems. As this chapter explains, the World Wide Web can be a standard platform for integrating all enterprise systems, but Web technology has not matured to the point where it can handle mission critical systems.

Along with Web and Internet systems, client/server systems make up enterprise computing. Chapter III–2, “Enterprise Client/Server Planning,” explains how to best integrate client/server systems into the enterprise architecture. This type of planning ensures the client/server systems are of high quality as well as flexible in meeting end-user demands.

What holds enterprise systems together is networks, which are the backbone of enterprise systems. Chapter III–3, “Enterprise Network Design Technology,” discusses the many facets of a project that involves developing an efficient enterprise network design. Chapter III–4, “Planning, Designing, and Optimizing Enterprise Networks,” deepens the discussion begun in the previous chapter by getting into performance and optimization issues. This chapter concludes with a case study about an enterprise network that connected 17 sites located across the United States.

Microsoft’s Windows NT is becoming the networking platform in a growing number of enterprises. Chapter III–5, “Windows NT Project Planning,” examines choosing NT as a networking standard and then migrating systems to it. Issues discussed in the deployment of NT include evaluating corporate needs, the installed base, managing vendors, managing a pilot roll-out, and full-scale deployment.

Network technology is constantly increasing network bandwidth, which allows for more-complex types of information to be distributed throughout an enterprise. Imaging systems are benefiting from the increased bandwidth of enterprise networks, which can accommodate multI–megabyte image files. Chapter III–6, “Networked Imaging Systems Requirements,” discusses network types needed to run these systems as well as associated implementation issues.

Mainframes are often the workhorses that power many enterprise systems. A developing alternative to mainframe computing is massively parallel processing (MPP). MPP is becoming the processing architecture of choice in data warehousing applications. Chapter III–7, “Implementing Massively Parallel Processing,” explains why an enterprise should choose MPP and the platforms and hardware configurations used in its implementation.

If networks are the backbone of enterprise computing, data bases are the foundation. Ensuring the availability of data is a crucial responsibility for data center managers. This responsibility is greater in a distributed environment, where the possibility of losing data increases with the number of systems that can access it. Chapter III–8, “How to Handle Data Loss and Backup,” offers some practical advice and guidelines on how to address this critical issue in enterprise computing.

Developing plans to move the enterprise from legacy systems to client/server is an extremely challenging process. One critical issue in this move is migrating the legacy data base to a relational data base. Chapter III–9, “Legacy Data Base Conversion,” concludes this section by providing data center managers with detailed information to help ensure that this type of move is successful.

Chapter III-1
Integrating the Web and Enterprisewide Business Systems

Chang-Yang Lin

The World Wide Web, or the “Web,” has experienced phenomenal growth since the popular multimedia Web browsers Mosaic and Netscape became available in 1994. Some 1996 figures show that 50% of the Web sites are used by commercial corporations, as compared with 13.5% in the previous year, and that percentage is increasing monthly, as more corporations discover the advantages of maintaining an online presence. Commercial Web sites are still mainly used as marketing tools to provide information about company history, locations, and products, as very few Web sites can respond to information inquiries on enterprise or legacy data that is mostly stored in mainframe computers. As for mission-critical application (e.g., customer-order entry, customer invoicing, billing, and accounts receivable), they are almost nonexistent among the Web sites. This is partially because current Web technology is not mature enough to facilitate effective and risk-free transaction processing over the Internet.

With the remarkable growth of the Web, users, both customers and employees, will inevitably demand the ability to access enterprise data via the Web, as well as to run Web-based enterprisewide applications. Systems-development managers have many issues to consider before setting a plan in motion to satisfy these users’ needs. In addition to providing background information on the Web and its capabilities, this chapter describes the four primary components of the Web. Key terms are examined, including HyperText Transfer Protocol (HTTP), HyperText Markup Language (HTML), and Uniform Resource Locator (URL). The chapter also discusses Web limitations and some “unanswered” business questions, and the limitations of current technology are identified. Intranets and their use in corporate settings are examined. Approaches for tying the Web and enterprise systems into a coherent system are introduced. The development tools and products for the integration are also identified. Finally, the chapter presents planning issues useful in preparing for Web-based enterprisewide business applications.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

THE WEB AND ITS CAPABILITIES

The World Wide Web is a way of organizing the Internet that allows users to search for and retrieve information quickly and easily in a nonlinear way. This information is structured into small chunks, called pages, and it can be displayed page by page through electronic links. Pages can store information in a variety of formats, including numbers, text, graphic images, video, audio, and programs. Essentially, the Web is a collection of independent, yet interrelated, pages wired together by hypermedia links.

Technically, the Web is a kind of client/server networking technology for the purpose of requesting and providing services. It is composed of four components: clients, servers, publishing tools, and communication protocols.

Web Clients

A Web client acts as a front-end browser for requesting service from the servers. Popular Web browsers include Netscape Navigator, Mosaic, and Microsoft's Internet Explorer. These browsers are generally equipped with graphical user interfaces (GUIs) that make Internet navigation relatively easy.

Web Servers

A Web server is the back-end distributing system that processes and manages requests for service from the clients. Popular Web servers include Netscape's Commerce Server, Microsoft's Internet Information Server, Process Software's Purveyor, and O'Reilly and Associates' WebSite. These Web servers can be evaluated in terms of such factors as performance, security, and manageability.

Publishing Tools

HyperText Markup Language (HTML) is an open-platform language used to define Web pages. This language includes a set of tags that must be embedded in the text to make up a hypertext document. Thus, creating an HTML page primarily involves the process of tagging documents; HTML encoding can be done by inserting the code in a standard ASCII text file, inserting tags in a word processing program, or using special software programs that build the code for the user. Such programs allow the user to select, through menus and interactive commands, the desired effects; the program then builds the appropriate HTML code.

Although word processors and other text editors can be used to create Web pages from scratch, tools specifically designed to publish Web pages are available to make working with HTML easier. Examples of these publishing products include Interleaf's Cyberleaf, SoftQuad's HotMetal Pro, InContext Systems' Spider, HTML Assistant Pro, HTMLed, and HotDog. All these products automate at least the tagging process by supporting intuitive what-you-see-is-what-you-get screens(WYSIWYG), menu, toolbar, and drag-and-drop interfaces. In addition, some products such as Cyberleaf are equipped with utility programs able to convert Microsoft Word or WordPerfect documents into HTML pages. The capabilities of these Web publishing tools can be classified loosely into four groups:

- *HTML Editing*. These features are used to enforce HTML syntax rules and to manage the HTML tags for formatting text, designing forms, inserting Universal Resource Links (URLs), and calling up photos, video clips, or sound files.
- *Fundamental Word Processing*. These features are used to create and edit the text.
- *Previewing and Testing*. These features invoke any Web browser to preview or test HTML pages in WYSIWYG form.
- *Document Conversion*. These features convert documents from plain ASCII text files or specific software-dependent files into HTML formats.

Whereas creating simple pages using these publishing tools requires no specific skills, extensive knowledge and skills to integrate hyperlinks, multimedia, and embedded objects will be required to create rich and interactive online pages.

Communication Protocols and URLs

The Web depends on three protocols to facilitate communications. Internet protocols include TCP/IP, HyperText Transfer Protocol (HTTP), and Universal Resource Locators (URLs) to communicate over the multiple networks. HTTP is the method that Web servers and Web clients use to exchange HTML pages. This method is built on the concept of hypertext/hypermedia that permits the nonlinear accessing of the pages.

URLs define the unique location where a page or service can be found. An example of a URL would be http://home.netscape.com/comprod/index.html. This URL begins with the letters http as the transfer format, which indicates that the last portion of the address(i.e., index.html) is an HTML page. The section after "://", in this case, home.netscape.com, represents the host computer where the information is stored. This is also referred to as the "home" page or the web site of the Netscape Communications Corporation because it can be used as the starting point to explore other pages in detail. Anyone can publish a home page or start at someone's home page. The rest of this URL is a path name to the file.

URLs do not always begin with the letters http. Other formats are also available, including ftp and News. Together, URLs and Internet protocols enable users to reach, in addition to the Web, other Internet resources such as e-mail, ftp, gopher, telnet, and discussion groups via Web browsers.

Search Engines

In addition to the above four components, search engines are constantly being created that help users find the Web sites that store desirable information. WAIS (http://www.wais.com), InfoSeek (http://www.infoseek.com), Yahoo (http://www.yahoo.com), WebCrawler(http://www.webcrawler.com), Lycos (http://lycos.cs.cmu.edu) and SavvySearch (http://guaraldi.cs.colostate.edu:2000) are often used for Web searches. These search engines organize their own databases, start their own search mechanisms to support queries ranging from simple query statements to complex formations and even natural-language queries, and they return a list of URLs. Without these searching machines, finding a list of desirable URLs from the vast, unstructured, uncoordinated Web resources is time consuming and could take the users months of point-and-click navigation to assemble.

Previous | Table of Contents | Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

WEB LIMITATIONS: UNANSWERED BUSINESS QUESTIONS

With its capabilities, the Web has been able to facilitate electronic business transactions. Product promotion, customer support, and electronic publishing are a few examples of functions in which Web technology has been successful. Nevertheless, from a business perspective, four fundamental questions remain unanswered. These questions have prevented many corporations from carrying on business via the Web.

- *Is Web navigation mechanism effective?* The Web employs the hypertext mechanism for navigation. For a typical query, the user is often required to do some clicking on the mouse to reach a desirable Web site. Once arrived, more clicking is required before information can be obtained. Although search engines have alleviated some of the difficulty in reaching Web sites, users are still required to do a lot of clicking and bouncing around from page to page following pre-designed links. Such a simple navigation mechanism is not flexible enough to give users more specific information and quicker responses to business queries.
- *Is the Web data structure adequate to support information reporting and query responses?* The Web employs a hypermedia data structure in which information is stored in small chunks, called pages. Text documents and other object-oriented data are fitted into these pages. However, traditional record-based business data and numerical data are not suitable for storage in pages, partially because business data, if stored in this format, cannot be easily accessed on a record-by-record

basis. In addition, HTML is just not powerful enough to handle record-oriented business data, nor does it allow user-controlled queries to be easily formulated. Consequently, key information cannot be provided under present Web-based data structure.

- *Can enterprise data or legacy data be available on the Web?* To date, enterprise data — mostly transaction oriented — is stored mainly in mainframe computers. Security and performance concerns are two major reasons that enterprise data is mostly inaccessible from the Web. Methods and techniques are being developed to bring mainframe-based data into the Web. At present, these methods and techniques are not feasible and therefore transaction-related information on order status, invoice, bill of lading, and payment will mostly remain unanswered.
- *Is the Web suitable for mission-critical business applications?* The Web is not set up for on-line transaction processing and has failed to meet the standards of security, performance, backup, and user management. For example, Web technology is inadequate to perform the five security-related tasks (i.e., authentication, certification, confirmation, nonrepudiation, and encryption). Therefore, an interactive transaction between trading partners is not reliable. Besides the security concern, other key factors have also contributed to a lack of Web-based mission-critical business applications. These factors include stateless conditions during transaction processing, questionable bandwidth to handle real interactive transactions, and lack of user preparedness for electronic commerce.

INTRANETS

Despite a dearth of legacy data on the Web and immature Web technology for effective transaction processing, an increasing number of corporations are now turning to the Web as their IS solution for addressing business problems within corporations. It is predicted that internal Web or intranet usage will surpass external Internet usage by the year 2000. The key factors for adopting intranets are open-platform standards (e.g., HTTP and HTML), ease of installing Web servers and using Web clients, and multimedia capabilities.

The range of intranet applications that can be developed is virtually unlimited. Currently, corporations are deploying intranets as a way to organize their internal communications. Examples of these intranets are:

- Web-based internal e-mail systems
- project planning, monitoring, and reporting
- forums for brainstorming, collaborations, or problem solving
- delivering marketing materials, training materials, or software products
- online customer information queries
- online human resource queries on employee benefits, company policies, personnel information, employee and telephone directories, job listings, and training and education workshops

One main concern in deploying intranet applications on the Web is security.

Currently, several measures are being installed, including "firewalls." Most firewall products focus on keeping external Internet users from getting into intranet applications. Others ensure that users are authorized to access the information they seek.

INTEGRATING THE WEB AND ENTERPRISE SYSTEMS

The process of integrating the Web and enterprisewide systems or building some intranet applications can be approached from two directions. One involves converting enterprise data into hypermedia pages. The other involves building a link between these two systems. Regardless of which approach is used, the goal remains the same; that is, making enterprise data and the various business applications accessible through Web browsers. The use of Web browsers eliminates concerns about heterogeneous hardware and various operating systems over the Internet and intranets as well.

Building links to tie the Web and enterprise systems into a coherent network is much more feasible than converting to hypermedia pages. This is partially because the linkage programs will not interfere with the normal operations of enterprise systems for supporting day-to-day business activities and management decisions. Both researchers and vendors have been placing their emphasis on developing architectures and tools to support the construction of the linkage programs.

Converting to Hypermedia Pages

Enterprise systems are characterized by a variety of data structures, including traditional flat files, relational databases, IMS databases, object-oriented databases, and special package-related files (e.g., spreadsheet files, song clips, and photo images). Theoretically, this data can all be converted into hypermedia pages to support applications ranging from information inquiry to transaction processing over the Web.

Although current technology is not mature enough to support certain tasks effectively over the Web (e.g., complex interactive transaction processing), migrating key enterprise data to the Web will certainly give customers speedy query responses for such applications as marketing and electronic cataloging.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Building Linkage Programs

Building the linkage programs to tie the Web and enterprise systems into a coherent system involves two similar approaches: augmenting HTML programs and augmenting enterprise programs.

Augmenting HTML Programs. The augmented HTML programs include a data-access subprogram. In addition to the data-access function, many augmented programs might include programs to facilitate interactive input and to merge the enterprise data into pages for presentation. These subprograms may contain SQL statements or procedure codes, called scripts. Examples of these tools or products include DECOUX, SWOOP, OpenUI and OpenWeb, WebDBC, and Open Horizon's Connection for Java.

DECOUX supports an augmented form of HTML that includes embedded SQL statements. SWOOP supports the generation and maintenance of Web systems that store information in an ORACLE relational database. The development tools OpenUI and OpenWeb, WebDBC, and Open Horizon's Connection for Java are based on the function-call models that let developers integrate prebuilt, vendor-driven key components together using C++ or other nonprogramming tools. These tools are now being investigated for applications such as hotel reservation, payroll, and human resources.

Augmenting Enterprise Programs with Embedded HTML Statements. Advanced features of HTML, such as forms, are embedded into enterprise programs and are used to capture input transaction data from Web clients. The input data are then fed into enterprise programs for processing. For example, Visual Object COBOL 1.0 by Micro Focus uses CGI to link HTML forms to COBOL programs and therefore let COBOL programs take input from HTML forms.

Besides using the above tools, Java, Sun's object-based open-system language, can be used to create the linkage programs to tie key components together. Furthermore, Java is said to be able to create Web-enabling interactive applications from scratch.

CHALLENGES AND STRATEGY ISSUES

As commercial Web sites and users continue to grow at an incredible rate, corporations are faced with an opportunity: Incorporating Web technology into enterprisewide applications to improve their competitiveness in the global market. The following is a list of questions and suggested solutions that address this opportunity:

- *How do corporations attract potential customers via the Internet and the Web?* They can build a presence on the Web, and then expand and enhance their Web pages.
- *How do corporations make enterprise data accessible via the Web to enhance service effectiveness for both employees and customers?* They can move enterprise data into a HTML format, use Web technology to connect legacy data, build search and index mechanisms to enterprise data, or develop intranet applications.
- *How do corporations deal with the barriers that slow down the implementation of enterprisewide systems, such as multiplatforms, security, bandwidth, and multiple development tools?* Organizations can plan both external and internal Web as an ideal solution for multiplatforms, or make intranets a solution for addressing the internal communication concerns. They can also install security tools or firewalls to prevent unauthorized users from reaching vital legacy data or applications and implement systems to track appropriate technologies, such as Web-development tools, Web servers, and security tools.
- *What strategies must corporations develop to remain competitive?* They can recognize the Web as one part of IS solution, and integrate traditional and Web-based systems. Systems-development managers can support a new intranet development environment. Organizations can prepare for electronic commerce and provide staffing and training for Web technology.

Regardless of Web technology's effectiveness for certain tasks, the rapid growth of the Web and its impact in the global market should not be viewed lightly. Facing these challenges and thus effectively deploying the Web to empower users requires planning. The following sections expand on the previous suggestions, for better planning.

Building a Presence on the Web. Corporations should position themselves on the Internet's Web by building home pages without any delay. As competitors' presences on the Web increase, one way to guarantee the failure of the above challenges is to adopt a "wait-and-see" approach.

Expanding and Enhancing the Pages. Simply migrating paper-based product catalogs to the pages and recording a CEO's welcome messages will not entice potential customers to visit the organization's Web sites repeatedly. Corporations need to think of new ways to both enhance and expand the pages. These may include:

- *Making key enterprise data accessible via Web browsers.* Enterprise data always serves as a foundation from which information can be derived. Both predesigned and ad hoc queries on key enterprise data must be considered to reflect friendliness and flexibility.
- *Providing additional services and facilities from the pages.* Examples of these services include customer and technical support; downloading reports, forms, policies and procedures, or software products; and on-line documentation. Examples of facilities could include a registration form to collect users' information and interests, a special form to allow the users to comment on products, and a platform to facilitate interactive communications.

Plan Intranet Applications. How the Web is used within a corporation must also be planned. Although many applications can be developed based on Web technology, those that involve communication, information sharing, and information distribution should be built first.

Prepare for Electronic Commerce. As Web technologies continue to mature, the solutions designed to prevent security breaches, stateless transactions, and performance concerns will gradually become available. Thus, corporations must prepare for electronic commerce by making enterprisewide applications — including mission-critical applications — Web capable. This might include building Web-capable applications from scratch, linking the enterprise data to the Web, and building the linkages between existing enterprise applications and the Web.

Corporations should identify and plan the projects for electronic commerce. Information reporting or inquiry projects should be built first, because linking SQL databases to the Web will be easier to do. Designing special searching mechanisms on enterprise data will also be necessary for fast inquiry response.

Building the linkages between existing enterprise applications and the Web can be performed next. The proven tools and techniques necessary for building such linkages should be evaluated and selected. Depending on the specific needs of the individual corporations, applications to be linked are ranked.

Education and Training. Both developers and users must undergo proper training for the emerging Web technology. Overall, developers and users should understand how the Internet and the Web can be accessed, used to gather information, and implemented to create business opportunities. The users who are responsible for publishing must learn HTML tools to create pages. Developers must learn the development tools to reengineer applications on the Web. Developers mastering the tools, including such programming languages as C++ and Java, will be essential for successful Web-enabled transformation.

CONCLUSION

Web capabilities are extensive and growing more complex and sophisticated at a rapid rate. To keep abreast of such changes, systems developers must consider such factors as security, transfer protocols and languages, and development tools and environments. All capabilities must be evaluated in context of the enterprise — its goals as well as its propensity for risk taking. Only with a careful weighing of the pros and cons can an organization take full advantage of the technology of the World Wide Web.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter III-2
Enterprise Client/Server Planning

Diane M. Miller

Providing high-quality enterprisewide computing is a demanding task that requires careful planning of the data, hardware, software, and communications resources. Comprehensive business-planning methods such as enterprise-architecture planning (EAP) have been successfully used to achieve overall quality. These thorough, controlled methods foster sound management practices in complex environments, for both the information system and the organization it serves. Because they are time consuming, such methods have generally been used in stable environments, where the resulting system remains operable for some time.

Contemporary organizations, however, are typically characterized by rapid change. To provide ongoing quality, the IS organization should be responsive to changing internal and external conditions and remain strategically positioned for continuous improvement. Many of today’s organizations, for example, are attempting to achieve rightsizing, accompanied by reorganization of mission components. Multiple successive organizational configurations may be required before an optimal arrangement is identified. This changeable environment necessitates that organizations be able to periodically realign information needs and technology resources.

To provide for organizational rightsizing and expected change and to take advantage of new capabilities in data communications and powerful downsized computing hardware, many firms have developed client/server systems. Although these systems are positioned to take advantage of shared data resources and achieve flexibility through appropriate use of contributing components, they have sometimes focused too closely on local needs, creating so-called “stovepipes” primarily used for retrieval. Such systems are not well integrated into the overall functions of the firm because they limit concurrent use, yield unacceptably slow response times, ignore complexities, or fail to provide for cross-departmental functions.

Providing quality services in a changeable environment thus poses a dilemma. On the one hand, there is a need for flexibility; but on the other, control and integration are necessary. Although approaches such as EAP and client/server computing each tend to serve one of these two needs well while precluding the other, it is possible to use these dissimilar approaches in tandem. Following brief descriptions of terms, this chapter presents a planning approach that helps ensure quality service by integrating enterprise-architecture planning with client/server technology.

ENTERPRISE-ARCHITECTURE PLANNING

In the context of EAP, the enterprise refers to the full range of an organization’s mission functions. Enterprise planning typically includes a comprehensive survey of the functional requirements of the business that focuses on the data, applications, and technology needed to accomplish the plan. An enterprise plan comprises a variety of domains of human activity related to information systems, in addition to plans for technological support, data and information maintenance, and documentation standards.

The enterprise architecture is a blueprint of what is needed for success in accomplishing the IS mission. It is a comprehensive top-down view that considers the environment and allows for lateral interrelationships among applications. The architecture defines the requirements; the plan that follows the architectural definition sets forth the design for implementation. The plan is thus the practical mechanism that enables the enterprise architecture to succeed.

The first step in EAP is to define the nature of the business and then to proceed with detailed architectures describing the data, applications, and technology to be used in the system, in the order given. Once the architecture is defined, the plan is developed, implemented, and maintained.

CLIENT/SERVER COMPUTING

Client/server systems attempt to distribute computing functions and data between networked clients and servers in a way that maximizes the capabilities of each. Clients are computers in a network that request computing services from other computers known as servers. Clients are not dumb terminals but usually microcomputers or workstations capable of independent processing tasks. Servers, which can be any computers, are used to provide file and database services, computation, communication, or printing services as requested. Servers can double as clients.

The three computing elements distributed across clients and servers are:

- presentation services, comprising output generation and the user interface
- application processing logic, which handles the functions of the user's program
- back-end services, which provide the resources required for the application processing to occur. Examples of back-end services are data base manipulation, communications, and intense computational functions.

Client/server computing is distinctive because it distributes parts of a single business process between two or more computers; in contrast, ordinary workgroup computing distributes entire processes to various machines.

The three elements of presentation, processing logic, and back-end services are distributed according to different strategies. The simplest form is a distributed processing (DP) strategy, wherein only a graphical user interface (GUI) is processed on the client and all other processing is provided by the server. At the other extreme is a distributed data management (DDM) strategy, which locates all presentation and logical processing functions on the client, along with portions of a segmented or replicated data base. Other strategies fall between these two.

Client/server computing emphasizes practicality by using existing and planned equipment in a way that maximizes the potential for physical access to data and applications. Enhanced user interfaces increase utility by making the system more appealing as well as easier to access.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

COMPARISON OF APPROACHES

Enterprise architecture planning is a data-driven methodology that represents an idealized approach to planning. Client/server computing is thoroughly practical, using event-driven code for program development. EAP is a quality-control approach that focuses on analysis and design, and client/server computing is an optimal-use technique for hardware, communications, and actual implementation.

Idealized notions are worth pursuing, because as patterns for what is potentially attainable, they can raise the quality of what is actually attained. The practical reality of the IS environment, however, is that complex systems rarely reflect the full wishes of the designers, rarely can be developed top-down only, and, once in place, almost certainly need to change. Because the EAP and client/server approaches are both resource-optimization methods emphasizing different parts of the systems life cycle, their use can be complementary in spite of their differences in philosophy. Juxtaposing the two approaches yields a set of principles and a plan of action for implementing quality systems.

IMPLEMENTING QUALITY SYSTEMS

Contemporary systems projects are usually efforts to reengineer existing processes, so planning should focus on the issues that affect the success of such efforts. These issues involve the following:

- commitment of resources
- formulation of the enterprise model

- design of the data, applications, and technology architectures
- migration of legacy systems
- organizational structure

The following sections present the principles behind each of these issues and the rationale for them.

Commitment of Resources

Principle One: Communication. The nature and benefits of enterprise-architecture planning and client/server computing should be clearly communicated to both management and rank-and-file employees before the reengineering project is undertaken. In the case of EAP, managers must understand and accept the need for long-range information-resource management (IRM) so that they will be willing to commit time, resources, and interest to the effort. Client/server computing, in turn, is an evolutionary process that, with no real end in sight, continuously demands high-end resources to achieve quality. Under all circumstances, workers deserve to know why their input is being sought and how a particular effort will affect them. Their enthusiasm for the project may yield creative ideas; their lack of support may mean ultimate failure.

Principle Two: Resource Allocation. Sufficient resources should be allocated to the project. Typical EAP studies require a team of four to six people working six to eight months; client/server projects can only overcome the stovepipe effect if they are broadened in scope to encompass more-complex processes. To be usable, a system must have quality and thoroughness. If adequate resources are not available, EAP will not succeed and client/server computing will remain specialized.

Resources dedicated to the project are not wasted for the following reasons:

- Deliverables resulting from the steps of the procedure are useful in themselves.
- A successful implementation pays for itself shortly and provides ongoing cost savings.
- New systems can be developed faster with less maintenance when an architecture is used.
- Overall, the resulting system will provide considerable strategic potential.

Formulation of the Enterprise Model

Principle One: Functional Analysis. A functional analysis should first be conducted to determine how departments operate and interrelate. Because the intent is to establish a comprehensive, integrated IS plan, a full understanding of what the organization does is prerequisite to determining what information is required for accomplishing its mission.

Principle Two: Focus of the Enterprise Model. To serve as the basis for strategic IS planning, the functional analysis should produce an enterprise model reflecting the mission and objectives of the business. This focus is important because departmental information priorities could change rapidly in response to environmental factors. The basic functions of an organization are less volatile because they reflect the nature of the enterprise and are less affected by fluctuating budgets and organizational boundaries.

Design of the Data Architecture

Principle One: Data Analysis. A comprehensive data analysis based on the functional analysis of mission activities should precede consideration of applications for the data. The data analysis is used to design the data architecture, which provides for data availability, consistency, and quality. Corresponding architectures for applications and technology follow.

Principle Two: Data Capture and Administration. Data should be captured once, at its source. Diverse forms of data from multiple sources should be integrated and made available locally. Data should be administered centrally and stored close to the primary user. Information stored online should be continuously available. This approach eases data-maintenance tasks, ensures appropriate data availability, and reduces response time.

Principle Three: Standards. Archival and data compression/retrieval standards should be formulated for massive data stores, because users must be able to determine what data is available and retrieve it selectively. EAP assumes an organizational scale and complexity sufficient to produce diverse data types and sources.

Design of the Applications Architecture

Principle One: Reuse. In general, applications should employ reusable components exhibiting the characteristics of system objects. Specifically, these characteristics are: the definition of each object includes its data attributes and processing methods (i.e., encapsulation); each classification derives data and procedures from previously defined classes (i.e., inheritance); and the same set of operations works on different objects (i.e., polymorphism). As the basis for software development in client/server computing, objects are particularly suited for transitional systems.

Principle Two: GUIs. A common GUI with cross-platform capability should be used by all applications. Although the interface should be constructed using a standard tool whenever feasible, its appearance to the user should be the same regardless of its method of construction. A familiar GUI enhances utility of the system and comfort of the user and obviates the need for system builders to learn multiple methods.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Principle Three: Data Dependency. Data dependency should determine the implementation plan. In other words, applications that create data should be implemented before applications that use data. Although high-visibility applications seem more critical, it is not only wasteful, but dangerous, to attempt their development without careful attention to the source and quality of contributing data.

Principle Four: Surveying Existing Applications. Existing applications should be functionally surveyed to identify opportunities, redundancies, and omissions. The timing of this activity (i.e., following the business functional analysis and the data analysis) is important. Although existing applications provide the basis for comparisons and assessments of impact and may contribute portions of reengineered systems, they are based on environmental factors that may no longer apply and should not be allowed to inappropriately channel the thinking of planners.

Principle Five: Applications Review. Applications should be periodically reviewed to ensure that they continue to meet functional needs, expand to include new opportunities, and evolve to avoid technological obsolescence. This approach protects the long-term value of a system by ensuring operational efficiencies and strategic effectiveness.

Design of the Technology Architecture

Principle One: The Data Environment. A standard query language (SQL) interface should provide access to multiple database-management systems (DBMSs) and also to nonrelational file structures. Unless the total data environment is considered, use of a DBMS will not yield the benefits of physical and logical data independence, reduced redundancy, data integrity, security, rapid development, and adaptability.

Principle Two: The Distributed Data Model. A distributed-data management model should be used that locates all logical and presentation functions on the client, along with portions of segmented or replicated data bases. This model takes advantage of the hardware tiers commonly found in large organizations and supports large-scale data architectures.

Principle Three: The Systems Development Methodology. The systems development methodology should use information-engineering methods and be supported by computer-aided software engineering (CASE) and repository tools. Besides positioning systems for the future, this approach best accommodates knowledge bases, graphs, images, and other forms of data appearing in contemporary systems.

Principle Four: Connectivity. A high degree of connectivity should be sought: locally, between sites, within the industry, and globally. This is because the need for access to external data and shared software, along with the data synchronization required for the replicated and partitioned databases of the client/server DDM model, produces heavy traffic.

Principle Five: Network Standards. Communications should adhere to national and international standards for network protocols and messages through use of the Open System Interconnection (OSI) network management framework.

Principle Six: Open Systems. Open systems should be used to achieve portability, scalability, interoperability, and compatibility. In other words, they ensure that systems software can run across multiple vendor platforms, across a power range from small to large systems, and in a heterogeneous environment, and they allow existing software and hardware to be integrated into the system.

Principle Seven: Security. Security of data, software, hardware, and communications should incorporate automatic error checking and recoverability, use appropriate redundancies for critical operations, and be as transparent as possible.

Migrating Legacy Systems

Principle One: Interim Adaptations. During transition, mission-critical processing may require interim data adaptations within the existing hardware context. These adaptations ensure continuity of data availability.

Principle Two: Hardware. Powerful clients should be procured as early as financial resources permit, but existing mainframe equipment can be depreciated fully and gradually replaced by servers that complement the clients. Depending on the available hardware, the distribution strategy for the client/server computing elements at each stage can depart from the DDM model during the transition. This approach provides stability for mission-critical processing and overall continuity while observing cost effectiveness.

Principle Three: Network Accessibility. As the system evolves, networks should continue to provide the means to accommodate access by various hardware platforms. Although much shared hardware can be standardized, the IS organization should position itself to take advantage of new technologies. In addition, some existing applications are standalone projects using hardware chosen for unique user needs, and future special projects can be expected to be similarly diverse. Users of these applications require access to data and support functions.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Organizational Structure

Principle One: The Planning Team. To conduct the EAP project, a planning team should be formed with members who perform the roles of coordinator; promoter; analysts for business functions, data, applications, and technology; tool set administrator; librarian; and quality control consultant. One person could play multiple roles, share a role, or perform duties in addition to team membership. The coordinator, however, should devote full time to the project and remain focused on the principles regarding commitment of resources. While the project is ongoing, responsibility for the project's direction may rest with an advisory council composed of IS and non-IS managers, but the coordinator should control the day-to-day activities of the venture.

If efforts are well coordinated, team specialists can provide the most valid insights and execution, thus creating the most effective approach. Any conceptual difficulties that arise can be resolved by the council while daily activity proceeds and without damaging the credibility of the EAP team.

Principle Two: Specialists. Specialists in various phases of operations should be heavily involved in the project during design and development of systems covering their own spheres of responsibility. They too should focus on the principles regarding commitment of resources.

Specialists are necessary because EAP and client/server development each require intensive efforts that fall within different phases of the systems life cycle. EAP assumes a structured systems-development life cycle (SDLC) to achieve necessary control. Client/server computing ordinarily uses alternatives to an SDLC, typically substituting joint application development (JAD), rapid application development (RAD), prototyping, or a combination of these methods. In an ambitious project involving EAP, however, it is appropriate to use all of these methods to provide both immediate focus and global integration.

Principle Three: Reporting Structure. After implementation, all IS professionals within the organization should report either directly or indirectly to the person responsible for the IT function. Clean lines of responsibility foster continuity and consistency and promote equity in salaries and career opportunities.

Principle Four: Ongoing Cross-Functional Interaction. Representatives of customers within functional areas should continue to meet regularly as a council. These representatives should include both IS professionals and functional managers, because the individuals working on a given application should be close to the function. Interaction between people knowledgeable about technology and people well versed in functional areas permits strategic exploitation of the IS potential. Collaboration among departments enhances communication, prevents artificial competition, and provides the best environment for excellent service. Representative communication also fosters the adoption of appropriate end-user standards that permit flexibility but avoid chaos.

Principle Five: User Training. User training should be conducted shortly before installation of each system phase occurs. Ongoing educational programs should be available to all levels of professionals.

Principle Six: Control of PCS. Control measures should be formulated for personal computers as well as for large systems with many users. The move away from stovepipe systems means that small computers will become embedded in the work flow. Increased reliance on the productivity of small machines means that local problems can bring the entire global system to a standstill.

Principle Seven: Measuring IS Value. Finally, systems should be managed as assets, because they have value in themselves. Metrics should be developed to appropriately value various systems aspects, and accounting methods should be used to track their activity. Cost/benefit comparisons should be performed.

These measures are necessary because in IS, as elsewhere, assets are created by investment and sound management, and the performance of the custodian of the asset is judged by the return the asset generates. Value received must be clearly identified.

ACTION PLAN

IS managers can take the following several steps to establish a client/server environment using enterprise architecture planning:

Step One: Initiate Planning

Step one involves the following planning activities:

- determining the scope and objectives of the planning effort
- creating understanding within the organization
- adopting a planning methodology
- arranging for project management computer resources
- assembling the planning team
- preparing the work plan
- obtaining commitment and funding

The deliverables of step one include the planning scope, objectives, project description, methodology description, tools designation, planning-team definition, presentations, and work plan.

Step Two: Develop the Enterprise Model

This involves the following four tasks:

- documenting the organizational structure
- identifying and defining the organizational functions
- documenting the preliminary enterprise model and circulating it for comments
- revising the enterprise model

The deliverables of this step include the organization structure and preliminary and final functional enterprise models.

Previous	Table of Contents	Next

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Step Three: Define the Data Architecture

In step three, the following four activities define the data architecture:

- determining the scope and existing potential sources of system data
- listing candidate data entities
- defining the entities, attributes, and relationships
- relating the entities to the information functions

Step three yields these deliverables: entity definitions, entity-relationship diagrams, an entity-to-function matrix, and a data architecture report.

Step Four: Validate Against Existing Functions and Technology

Step four involves six activities that produce system schematics, catalogued resources, and a validated data architecture. These activities are:

- determining the scope, objectives, and work plan for the survey
- collecting systems performance data and cataloging the results
- drawing schematics
- validating and reviewing the draft of the resource catalog and data architecture
- revising the data architecture and updating the catalog
- disseminating the data architecture and the resource catalog
- administering and maintaining the catalog as planning proceeds

Step Five: Define the Applications Architecture

To define the applications architecture, the following tasks should be performed:

- listing candidate applications
- defining the applications
- relating the applications to functions
- analyzing the impact on current applications
- determining security requirements
- disseminating the applications architecture

The deliverables for step five include a definition of applications, application matrices, an impact analysis, and an applications architecture report.

Step Six: Define the Technology Architecture

The technology architecture is defined through three activities:

- determining a strategy for distributing data and applications
- defining technology platforms
- disseminating the technology architecture

These activities yield the data/applications distribution plan and a technology architecture report.

Step Seven: Formulate an Implementation Strategy and Plans

Step seven comprises the following:

- sequencing the applications
- identifying interim transitions for hardware and data
- estimating the effort and resources and producing a schedule
- estimating the costs and summarizing the benefits
- determining the success factors and making recommendations
- finalizing and presenting the analysis report

These activities yield an applications sequence, migration plan, costs and benefits, success factors, recommendations, and an analysis report and its presentation.

Step Eight: Transition to Implementation

By this point, actions consistent with a structured approach to the SDLC have produced an environment that allows for orderly change. Within this environment, remaining actions have the latitude to use prototyping and JAD to exploit the responsiveness and flexibility of client/server computing without exposing the enterprise to the pitfalls of localized computing. Deliverables that become a part of the resource catalog are policies, standards, procedures, revised descriptions of architectures and technological resources, and ongoing detailed plans for enterprise projects. Additional deliverables are training sessions and an evolving working system.

CONCLUSION

The principles and actions described combine enterprise computing with client/server computing in a way that leverages the strengths of each approach and counters its weaknesses. Taken together, they describe a program for total IS quality. Enterprise planning contributes order, security, credibility, and inclusiveness. Building on that foundation, client/server computing provides flexibility and rapid adaptation to change without compromising system quality.

Recommended Reading

Hoffman, T. Legacy Systems: Getting There from Here. *Computerworld Client/Server Journal*, May 1994.

@FN:Spewak, S.H., and Hill, S.C. 1993. *Enterprise Architecture Planning: Developing a Blueprint for Data, Applications, and Technology*. QED Publishing Group, Boston.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter III-3
Enterprise Network Design Technology

Roshan L. Sharma

Network design is an important but neglected component of the network management process. Part of the reason for its neglect is that it is a highly multidisciplinary field, and it can intimidate even the most experienced manager. All network design efforts involve several distinct, quite different tasks. A data base is created that describes the customer premise equipment and the communication facilities serving all network locations. The end-to-end multi-hour traffic flows between all network locations are modeled. End-to-end performance requirements for all forms of communication are defined. Traffic growth for the life cycle of the network system is modeled. Several networks, all using available technologies, are designed. The best network design, based on cost, cut over considerations, and performance, is selected. The performance of the network after cutover is tested. The analytical tools are updated and preparations are made for the next design cycle. The results of the process are documented.

Consideration of these tasks suggests that a data center manager needs a combination of perseverance, traffic engineering knowledge, an understanding of (and access to) user-friendly, interactive design tools, electrical engineering skills, and marketing savvy (to convince senior management of the new network's value). The knowledge required to perform these tasks far exceeds that demanded of any other corporate officer.

DATA BASE

The first task is by far the most time-consuming. A relational data base should at least list the 10-digit telephone numbers, associated vertical and horizontal coordinates, all customer premises equipment (each record should include the vendor's name, date of installation, and single point-of-contact), the use level of each piece of equipment, type and number of communications facilities serving each location, and the points-of-presence of local exchange carriers and other common carriers along with their vertical and horizontal coordinates.

TRAFFIC FLOW, PERFORMANCE REQUIREMENTS, NETWORK GROWTH

The next three tasks quantify traffic for all hours and all types of communications between all significant pairs of locations, based on the principles of traffic engineering. This set of tasks also includes the creation of meaningful models of traffic growth for the next applicable life cycle.

NETWORK DESIGN

The design of several possible networks requires the availability of a user-friendly, state-of-the-art, computer-based network design tool that can generate strategic network alternatives in an iterative fashion and that can accommodate emerging communications technologies. The lack of such a tool is probably the most significant reason that managers neglect the network design effort. Consequently, the design of many enterprise networks is based on engineering dogma or the designer's hunches, or a design inherited from the previous management is used. In addition, many enterprise networks are transformed slowly during their lifetime into unrecognizable and severely unoptimized entities, particularly compared with the networks that could result from the use of a new communications technologies. Strategic and tactical network planners estimate that enterprises can reduce monthly network management costs by 40% if networks are designed with modern design software and the managers consider cost-effective emerging technologies.

SELECTION, SELLING THE NETWORK, POSTCUTOVER PERIOD

Selecting and selling the best solution to management is an art mastered primarily through experience; an emphasis on life cycle costs and savings analyses is generally effective. Testing the performance of the system during the postcutover period and documenting the results validates and can also improve the software package's analytical tools.

MODERN ENTERPRISE NEEDS

An enterprise must introduce new products or services just to stay abreast of its competitors. Some enterprises (e.g., the regional Bell operating companies and state and federal governments) must maintain cost-effective networks whose administration and management satisfy legislative requirements and the scrutiny of public utility commissions.

Digital transmission technology is making available new communications facilities (e.g., T1, fractional T1, fractional T3, privately owned digital microwave, and fiber optics). Architectures for hybrid networks (employing, for example, frame relay, switched multimegabit data service, asynchronous transfer mode) are either available now or soon will be. Reasonably priced customer premises equipment based on digital multiplexers for access control either at all customer locations or on nodes attached to the backbone network can now be purchased. The hybrid networks that can be implemented using this equipment and the new services and architectures can lead to considerable cost savings over existing networks.

Corporate strategic planners engaged in network design should understand the needs of a modern enterprise. A typical strategic planning cycle lasts three to five years and entails the continual evaluation of the goals and the needs of the enterprise and its competitors and an understanding of new communications technologies, customer premises equipment, and networking standards. The goal of a communications strategy is to deploy these technologies cost-effectively and enhance the productivity of a network's users. A strategic planning group should work closely with the IS department and should have access to all the necessary resources and tools.

OUTSOURCING

Experience shows that a strategic planning group can succeed only if the enterprise does not use outsourcing. No outside consultant or company can fully understand the demands of the marketplace, the structure of the IS department, and network operations. Only the internal strategic planning group can do that. Outsourcing companies have their own bottom line considerations, and these may run counter to the interests of their customers.

HOME ACCOUNT INFO SUBSCRIBE LOGIN SEARCH MY ITKNOWLEDGE FAQ SITEMAP CONTACT US

SEARCH ITKNOWLEDGE

Go!

Keyword

Brief Full

Advanced Search

Search Tips

BROWSE BY TOPIC

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Bookmark It

Search this book:

Go!

Previous | Table of Contents | Next

NETWORK DESIGN TOOLS

A network design tool should enable the designer to work in a manner similar to the pencil and paper (sometimes known as back-of-the-envelope) methods used in the past for simpler voice-only networks. A user-friendly tool should help the strategic planning group obtain feasible network approximations iteratively and interactively. The following examples illustrate situations that require the design tool to deliver either rapid modeling and designing or quick conceptual solutions.

Example I

An enterprise has 25 locations interconnected by a single-level T1 voice network, as depicted in Exhibit 1. What will happen if these same locations are served by a two-level hierarchical network with uniform routing over T1 links? Network design software shows that for the two-level hierarchical network (see Exhibit 2) a three-switch design is the optimum solution. Monthly transmission costs fall from $248,162 with the existing single-level network to $151,239 for the new configuration.

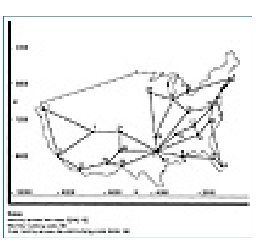


Exhibit 1 Existing 25-node T1 network topology.

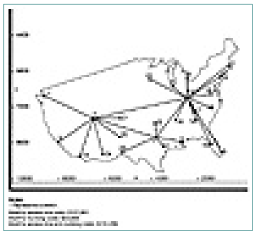


Exhibit 2 Proposed two-level T1 network with three switches.

Example 2

An enterprise has an efficient, optimized two-level, 41-node data network with three switches arranged in a star topology (see Exhibit 3). The existing traffic loads are accommodated on 9,600-bps lines. Can any cost saving be realized by substituting a three-level network with the same topology, seven concentrators, and two tandem switches? The configuration solution is illustrated in Exhibit 4, where it is shown that monthly costs can be reduced from $29,616 to $23,169.

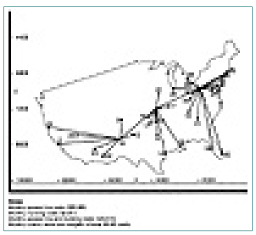


Exhibit 3 A two-level data network with three switches.

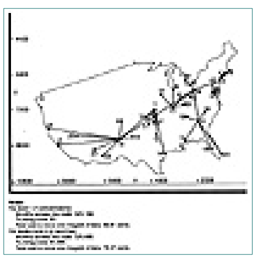


Exhibit 4 A three-level proposed data network with two switches and seven concentrators.

Example 3

An enterprise anticipates a 20-fold increase in traffic over a 41-node 9,600-bps link data network (e.g., either network depicted in Exhibits 3 and 4) when computer-aided design and computer-aided manufacturing (CAD/CAM) and electronic data interchange (EDI) applications are introduced. What are the optimum network topologies if the 9,600 bps links are replaced by 56,000-bps facilities? Exhibit 5 shows a star topology configuration to handle the increased line speed, and Exhibit 6 arranges the network in a multidrop topology.

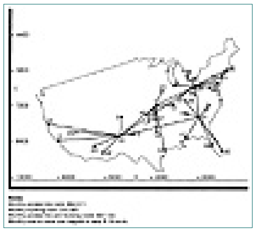


Exhibit 5 A proposed two-level data network with four switches in a star topology.

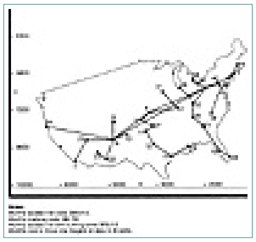


Exhibit 6 A proposed two-level data network with three switches in a multidrop topology.

MAJOR NETWORK DESIGN ISSUES

Current literature dealing with networking is often confusing, being characterized by buzzwords, hype, and the presentation of too many solutions with no single solution for a particular class of user. Some articles recommend that an enterprise network employ only virtual lines; other articles advocate the use of both private leased lines and virtual facilities. LAN design articles do not clarify the merits of optimizing a LAN for its intended department's use versus optimizing it for interconnection, through frame relay or other WAN services, with other LANs. New types of technologies and new network architectures are described, but the readers are left to assess the usefulness of the new facilities in respect to their own enterprises.

The network design process is basically concerned with two issues: topological optimization (the way network nodes are connected to one another while satisfying a limited number of design and performance constraints) and system performance (end-to-end connection and response times, path congestion, and link availability).

Issues related to system performance are most affected by network topology. Recurring network cost is another important consideration and is also determined by the network's topology.

Network design tools provide approximations to performance questions. Precise, closed-form solutions for end-to-end system performance are very difficult to obtain. Solutions based on computer simulations apply only to a specialized set of initial network conditions and then only when calculations reach statistical equilibrium. Proving out the network by testing is becoming difficult because of the extremely low error rates and fast response times the tests must document. Articles that discuss LAN interconnection ignore the difficulties that the interconnection of voice networks caused network designers during the 1970s.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

THE OLD NETWORK DESIGN TECHNOLOGY

Previous generations of network design technology are characterized by the use of specialized tools for voice, data, and video networks. They usually run on mainframes, distancing the network designer from the host processor, and making the software expensive to license and use. The jobs are invariably batch processed and the results become available, after a processing delay, in printout form. Each change of a design parameter or study of a new technology entails another delay. The absence of graphics causes additional delays in interpreting results.

Tariffs

The old design technology also requires the use of an extensive tariff data base. These tariffs have been increasing in number and changing quite rapidly since the divestiture of AT&T in 1984. The complexity of the tariff data base is the main reason mainframes are used. If such a data base is incorporated into a minicomputer or a minicomputer-based workstation, sizable processing delays occur.

Because enterprise networks are designed or planned for a future period, exact tariff accuracy is unnecessary. In addition, because network topologies do not change with changes in any given tariff (they change only with varying design parameters and technologies), a simplified set of existing or new tariffs can be used and excellent network optimization results obtained quickly even on a desktop workstation. At a later time, these topologies can be analyzed for cost using one of the many available personal computer line price products. This two-step approach separates the network design algorithms and the tariff data base. The network design package should not require updating just because a tariff changed slightly.

Design Tools

A useful design package should be an intuitive tool in the hands of an expert network designer. Network design tools based on the older technology are not interactive or user friendly. Lacking an acceptable graphical user interface, they require a long training and familiarization period.

Network design tools based on older technology are not flexible enough to handle the multilevel network hierarchies made possible by new customer premises equipment (e.g., low-cost multiplexers, voice and data compression units, tandem switches, and bridges, routers, and gateways), leased or fully owned digital transmission facilities (e.g., T1, fractional T1, T3, fiber, and digital microwave), and network architectures based on virtual services (e.g., X.25, fast packet, frame relay, and SMDS). Handling of mixed topologies through the process of topological superimposition is also generally not allowed with the older design tools.

NEWER NETWORK DESIGN TECHNOLOGIES

Older network design packages did not produce detailed end-to-end performance models for the various architectures previously mentioned, but neither do the newer ones. End-to-end performance evaluation for controlled situations have always been the province of university or private research laboratories. Network design packages use simplifications of operational environments. Problems occur only when users of these network design packages treat the resulting performance metrics as if they were exact descriptions of real networks.

Software packages are available for simulating system performance and can evaluate, for example, LANs (voice or data) and WANs (which consist of interconnected LANs). However, the simulation program has to reflect the exact network topology and the underlying communications protocols, and it must be completely debugged before the performance metrics (e.g., through-put and end-to-end response times) can be evaluated. Simulation of a typical enterprise network requires considerable run-time; so simulation tools are not interactive. For the same reason, computer simulation is not an ideal method for discovering an optimum network topology.

Iterative, user-friendly, interactive design tools are available for several current workstation platforms. Many of these tools can compute end-to-end connect and response times for unusual operational conditions; some provide special tools for analyzing system security and reliability.

Even the use of the intuitive graphical user interfaces does not eliminate the need for an expert network designer or architect. Because the designer works by repeatedly changing system variables in order to observe the effect on the network, the design package is useful only if it provides solutions rapidly. To allow this during the topological optimization phase, it is necessary to reduce or eliminate reliance on the mainframe or a detailed tariff data base.

Some vendors make available to their customers remote access to the vendor's host-resident network design software. However, delays and high communications costs tend to discourage this practice. Exhibit 7 lists some well-known network design tools available today, along with some advantages and disadvantages of each.

Exhibit 7 Currently available network design tools.

SUMMARY

The time-consuming tasks involved in the process of network design are often the reason it is overlooked or neglected by many managers. Consequently, companies that do not maintain efficient, cost-effective networks capable of handling constant change may have trouble staying abreast of their competition. To facilitate network design, today's strategic planners can use a variety of interactive desktop tools to help them optimize their network's topology.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter III-4
Planning, Designing and Optimizing of Enterprise Networks

Roshan L. Sharma

Network planning, design, and optimization are important components of the network-management process. Traditionally, these functions have been performed through the use of powerful mainframe computers. Because these computers required the use of large tariff-related databases, a great deal of time was spent inputting data and interpreting the output that invariably came in the form of a thick stack of computer printouts. No graphics were available to illustrate the network topologies. Furthermore, the user was always kept out of the design process. However, advances in very large-scale integration (VLSI) technology have made available powerful Pcs that have opened the door to the development of better network planning and design tools.

The network planning and design effort can be broken into the following distinct tasks:

- Creating an enterprisewide database of dumb terminals (e.g., telephones), intelligent workstations, customer-premises equipment (CPE), such as private automatic branch exchange (PABX) or data LANs, and communications facilities serving those locations
- Modeling all types of end-to-end multi-hour traffic flows between all locations
- Modeling traffic growth during a life cycle of the network system

- Defining end-to-end performance requirements for all forms of communications
- Designing strategic and tactical network alternatives using available technologies
- Selecting the best alternative network based on cost, cutover, and performance
- Testing the performance of a post-cutover network implementation
- Updating the analytical tools and preparing for the next design cycle
- Documenting the results

THE ENTERPRISE DATABASE (EDB)

Creating the enterprise database is by far the most time-consuming of all network design tasks. An enterprise database (EDB) should at least list:

- all vendors' exact mailing addresses
- all vendors' 10-digit telephone numbers
- all pieces of CPE with vendor's name, date of installation, and single point of contact for maintenance
- the usage level of each CPE
- the type and number of communication facilities serving each location, associated point-of-presence (POP) central offices of local exchange carriers (LECs), and interexchange carriers (IECs) with their vertical and horizontal coordinates

The list can grow into a very large one when the database must also classify the users at each location and their communications needs. However, the tasks involved in network planning, design, and optimization are impossible without the availability of an EDB. The table that appears later in this chapter illustrates a sample EDB.

TRAFFIC ENGINEERING TECHNIQUES AND TOOLS

The next three network planning tasks demand a capability for traffic modeling and analysis. Before defining the traffic engineering efforts, some basic traffic-related concepts should be introduced.

There are two types of traffic encountered in enterprise networks:

- well-behaved voice and video traffic
- bursty data traffic

It is always assumed that connection-oriented voice traffic behaves in a predictable fashion, which implies that:

- The call-holding times can be expressed by at least two moments (i.e., an average and a variance)
- The finer structures of traffic flows do not require rapid changes in network resource deployment

But a close observation of speech energy over the duration of a conversation will show that there are many pauses. Furthermore, two of the four-wire access lines (ALs) and trunks are always idle since only one party can talk at a time. These facts have helped long-distance carriers send more calls over expensive ocean cables than are possible over available trunks using pure circuit switching by using the time-asynchronous speech interpolator (TASI) technology. Such a technology was never cost effective over cheaper land-based leased lines. With the availability of asynchronous transfer mode (ATM) and Broadband Integrated Services Digital Networks (B-ISDN), users can get the same benefit through the use of variable bit rate (VBR) capability.

The data traffic between two CPEs is always bursty because of the complex rules of data communication protocols. Very small control messages might be involved in both directions before user information can flow. Although a full-duplex connection can be maintained, shared transmission lines in a packet-switched network can carry variable-length packets from many sources concurrently, thus muddying the picture. The faster the transmission lines, the burstier the transmission will appear.

Circuit-Switched Voice and Video Traffic. Circuit-switched voice and video-traffic intensity is measured in erlangs, which is equal to the average number of circuits busy during a "busy hour" between two network nodes. For example, if 15.5 conversations are observed concurrently between two network nodes (e.g., between a PABX and a voice switch or over an access line bundle) during a busy hour, then the voice traffic intensity is 15.5 erlangs.

Packet-Switched Data Traffic. Packet-switched data traffic intensity can be measured as the traffic rate in bits per second (bps) during a busy hour. Only the data rate in bps can describe the bursty nature of data traffic. Experienced network specialists have been using the concept of data erlangs for many years in defining the average data traffic intensity between two network nodes. This is obtained by dividing the observed busy hour data rate (R) by the capacity (C) of each separate transmission line. For example, if the busy hour data rate between two nodes is 392,000 bps and the capacity of a transmission line is 56,000 bps, then the data traffic intensity is 7 erlangs.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

MODELING TRAFFIC FLOWS IN A BRAND-NEW ENTERPRISE NETWORK

It is sometimes difficult to model traffic flows for a brand-new system. Many approximate methods have been devised for predicting traffic intensities (TIs) between all major CPEs. For example, a voice LAN (or PABX) generates about 0.1 * Ns erlangs of busy-hour traffic, where Ns is the number of active subscribers served by the PABX.

A breakdown of these traffic expressions into intranodal and internodal traffic should be determined by the known pattern observed at each enterprise. Some network designers use the 70/30 breakdown — 70% of the traffic remains within the site (voice/data LAN) and 30% of the traffic goes to other CPEs as internodal flows. These TI values can then be entered into an input file that defines each site ID, the related vertical and horizontal coordinates, and the total traffic intensity handled by the site.

The next task is to model the internodal traffic flows (i.e., exact traffic intensities handled by all the nodes and links in the path of a CPE–CPE connection). These computations are generally performed by the network design software for each assumed network topology (i.e., number of network switches and the link types employed at each network hierarchy). Some tools use critical design parameters to determine the fraction of traffic handled by access lines (connecting CPE and a switch) and trunks (connecting two switches). Eventually, the tool provides the total traffic intensity handled by each resource (node or link) of each network topology considered during a typical busy hour.

MODELING TRAFFIC FLOWS IN AN EXISTING ENTERPRISE NETWORK

Exact traffic flows can be modeled using the detailed traffic data gathered by intelligent network nodes (e.g., PABX or LAN). The source ID, destination ID, call-originating time, and call duration for each connection is recorded in station message data recording (SMDR) tapes of the voice network. Similar data is recorded by the data LAN for the packetized traffic. Simple traffic-analysis packages are obtainable for analyzing the exact internodal traffic patterns between all pairs of CPEs. Such data can then be entered in a from–to data file (FTF) to define CPE traffic as simple vectors (i.e., From–Node ID, To–Node ID, and the BHR traffic intensity) for each CPE–nodal pair.

This effort eventually provides actual traffic flows (i.e., the actual traffic intensity handled by all resource, nodes, and links) of each network topology studied during a typical busy hour.

Modeling Time-Consistent Averages (TCAs) of Traffic Flows

Choosing a "busy" hour is an important task. Networks are not cost effective when they are operating during the hour with the highest traffic. A network might provide the required grade-of-service (GOS) during the busiest hour, but at all other hours of the day (especially during the evening and night hours), the GOS level would be overkill. No organization can afford such a network. Network managers who select an hour with the least traffic during the day will hear complaints all day long. Therefore, a proven methodology is needed to select the average traffic intensity for network design. There are two methodologies — one used in North America and one used in all other countries.

The first methodology requires the selection of a typical month and the creation of a matrix (30 × 24) of traffic intensities (TIs) for each network resource for that month. Next, the average traffic intensity for each hour of the day over all 30 days of the month is computed. This process is repeated for each hour of the next 24. The TCA traffic is the maximum value of all 24 TCA values. This value determines the size of the resource (i.e., number of AL and trunks in the bundle connecting two nodes or the computing power of an intelligent node). It is helpful to have a software package for computing TCA traffic intensity (TI) values.

The second methodology requires that the 36 highest TI values be observed over an entire year and then the average computed to get a TCA value. This must be done for all resources.

Both of these methodologies result in more economical networks. However, no single methodology can predict an exact traffic pattern. Traffic values behave like the stock market. A single catastrophe, such as an earthquake, can also change the traffic patterns drastically. The objective of an effective traffic engineering practice is to synthesize an economical enterprise network using a consistent approach.

Modeling Traffic Growth During the System Life Cycle

To estimate the total costs incurred during the life cycle of a network system, the traffic intensities for each year of the life cycle should be modeled. The Delphi approach often works best. In this method, all general managers are interviewed and realistic models of traffic growth during every year of the life cycle can be built. Some divisions might disappear through divestiture or attrition. The data from all of the interviews must be collected, weighed, and processed to create a meaningful model.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

PERFORMANCE ISSUES

Before performance requirements for all the communication needs of an enterprise can be defined, network managers must first study the available concepts of performance and identify the exact enterprise needs in each business area.

Network-System Performance

Many network systems are implemented without any regard to performance. As long as they satisfy the basic needs for communications, everyone is happy. Often, no effort is expended in predicting or measuring the actual performance of the system or in making any measured systemwide improvements after the system is operational.

The lack of any concerted effort in defining and measuring performance of network systems possibly lie in an ignorance of certain system-performance concepts. The performance of a network system can be defined in four ways:

1. total system costs computed on a monthly basis
2. system throughputs in terms of all types of transactions handled during a unit time
3. systemwide quality of service (QOS)
4. systemwide grade of service

Total Monthly Costs. Transmission facilities determine the majority of the total monthly cost of MANs and WANs paid to the local exchange carrier (LEC), interexchange carriers (IECs), and other common carriers. The other major costs are for hardware and the recurring price of network management and control (NMC). Financing the hardware can turn a large one-time cost into an affordable monthly cost. The NMC costs related to spares can be handled just like one-time hardware costs. Some companies hire in-house NMC specialists; others prefer to outsource.

System Throughput. System throughput is measured by the rate at which the various types of transactions are handled per unit time (usually second or minute). Throughput is defined by the number of call attempts or calls completed per second for a voice network. In a data network, throughput is defined by the number of packets or bits handled per second. The throughput capability of each node is generally defined by the equipment vendor. The challenge lies in measuring the system throughput. System throughput can be estimated by enumerating the exact paths of each transaction.

System Quality of Service. Performance aspects dealing with transmission quality, perceived voice quality, error-free seconds, data security, and network reliability (mean time between system failures) fall into the QOS criterion. Most of these parameters are very hard to compute for the entire system. Performance aspects of a critical resource can be estimated to get a feel for the quality of service of the entire system.

System Grade of Service. The GOS criterion deals with end-to-end blocking for a voice network and average response time (measured as the elapsed time between the moment the send key is pressed and the moment the return reply is discerned by the user) for data communications. Analytical tools are available for estimating GOS parameters for voice, data, and integrated networks.

Defining Enterprise-Performance Goals

Performance goals for enterprise networks are generally developed by corporate strategic planners. A typical strategic planning cycle lasts several years and entails:

- Continuous evaluation of the needs of the enterprise and its competitors. This activity defines the relationship of system response times to user productivity for each transaction.
- Study of evolving new technologies, CPE, and networking standards. The most effective way of deploying these new technologies should also be investigated. This study should establish the cost and performance attributes of new hardware (e.g., ATM and LAN switches).

A network-planning group should work closely with the IT department. It is better not to outsource strategic planning because an outside group cannot fully understand the close synergy between the demands of the marketplace, corporate IT, user productivity, and network operations.

Network managers today have to deal with ever-increasing demands for:

- voice, video, image, and data communications
- multiplexing of digitized voice, image, and video signals with regular data traffic at all hierarchies of enterprise locations through switches

(e.g., ATM switches)

- unscheduled or varying demands for digital bandwidth at all hours of a day on a dynamic basis

To design an integrated enterprise network, the strategic planning group needs a user-friendly tool for quickly evaluating solutions that take user demands into account. The right tool should help the strategic planning group reach solutions iteratively and interactively.

MAJOR NETWORK-DESIGN ISSUES

No single approach to network design is ideally suited for all enterprises. Network design is basically concerned with two issues:

1. Topological optimization, which determines the way network nodes are connected to one another (including the type of connections) while satisfying a set of critical design and performance constraints.

2. System performance dealing with end-to-end response times, path congestion, and availabilities. Recurring network cost is generally the most important performance criterion and it is mainly determined by its topology. Network topology also determines the remaining performance issues such as response times and availability. Each network-design package analyzes these performance issues in only an approximate manner.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Previous Network-Design Technology

Many older network-design tools handled only voice or multidrop data networks. Some of the tools that came later handled only interconnections of data LANs to achieve an enterprise data WAN. Furthermore, most of these tools required mainframes. The use of a mainframe introduced an unnecessary curtain between the network designer and the host processor. The network-design jobs were entered invariably via the “batch” approach, and the outputs came in the form of large printouts after a good deal of delay. Each change of a design parameter or study of a new technology required a new non-interactive delay. The absence of network-related graphics from outputs caused additional delays in interpreting the significance of results.

The old design technology also required the use of an extensive database of tariffs. The complexity of the tariff database was probably the main reason behind the need for mainframes. If such a database were incorporated into a desktop minicomputer or a PC-based workstation, users would experience significant processing delays.

Because network topologies do not change with perturbations in any given tariff (they change only with varying design parameters and technologies), using a simplified set of existing or new tariffs is sufficient for designing an optimized network. These topologies can be studied for a detailed cost analysis using one of the many available PC-Line Pricer (PCLP) units. This two-step approach should create a separation between the network design algorithms and the ever-changing tariffs. There should be no need to update the network-design package just because a tariff changed slightly.

Simulation Tools. Some vendors market software packages based on computer simulation for evaluating system performance. LANs (voice or data) and WANs consisting of interconnected data LANs can be evaluated for performance through computer simulation. A good deal of time must be spent on writing the simulation program based on the exact network topology and the underlying communication protocols, and on debugging the software before one can evaluate all of the performance metrics, such as throughput and end-to-end response times.

Because typical enterprise networks require exorbitant run-times, a simulation tool is no longer an ideal way to synthesize an optimum network topology. A network topology optimization package based on analytical tools is always the best approach. The resulting topology can be evaluated for studying detailed system response times and availabilities using an expensive simulation tool.

NEW NETWORK DESIGN TECHNOLOGY

New network design tools are user-friendly, interactive, and can optimize network topology in an iterative fashion while quickly varying the values of critical design parameters. Many of these tools provide special menus for computing end-to-end response times for unusual operational conditions. Some packages even provide special tools for analyzing subsystem security and reliability.

Many new tools based on the graphical user interface (GUI) can evaluate any mix of CPEs, transmission facilities, and network topologies very rapidly in an intuitive manner. Today's design tools also allow the entry of approximate tariffs. But in no way can this new technology eliminate the need for an expert network designer or an architect. Because the expert designer is always involved with "what-if"-type analyses, the potential solutions are meaningful only if the network design tool provides them quickly.

ONE EXAMPLE: THE ECONETS NETWORK PLANNING AND DESIGN PACKAGE

Inputs into this network design package are in the form of flat, sequential files. Results are provided in the form of graphics illustrating a network topology with summary costs of communications facilities and response times and in output files containing detailed cost distributions and critical performance data.

The most important input file, the VHD file, lists the site/node ID, vertical and horizontal coordinates, and total busy hour, time-consistent traffic intensities in bits per second (for data) or millierlangs (for voice) for each location of the enterprise. A from–to data file can also be used to represent exact traffic flows. Another file called the daily traffic profile relates the busy-hour intensities to the other 23 hours of the day for computing the costs on a daily/monthly basis. For an enterprise with many time zones, several busy-hour models can be used.

The second most important input file, the link file, defines the link type that serves each location. Another important input file, the NLT file, defines the link type, capacity, allowed maximum data rate, multiplexing factor, corresponding tariff number, and the multiplying factor for a privately owned facility, if applicable. Up to 10 link types and corresponding capacities, allowed maximum data rates, multiplexing factors, corresponding tariff numbers, and multiplying factors, can be defined by the NLT file. The tariff file can define up to 10 manually entered tariffs, each modeled on 17 parameters. Several link, NLT, and tariff files can be prepared to model many combinations of links and tariffs at all levels of the network hierarchy.

The system-design file defines the busy hour, from–to traffic for all significant pairs, if such data is known. Other input files are also used for modeling/designing ACD networks using a mix of virtual facilities and leased FX lines.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

The File menu allows the creation and viewing/updating of all input/output files. The Networking menu allows the modeling/design of multilevel voice, data, and IV/D networks using the appropriate star data, directed link, and multidrop data-network topologies and voice networks based on star topologies. Network managers can also model, design, and optimize backbone networks in an iterative manner.

The Networking menu also allows the designer to find optimum locations for concentrators/switches by starting with effective solutions and improving these through a fast interative process. By specifying the design parameters, network managers can model and design data networks based on IBM's SNA, packet-switched networks based on CCITT's X.25 standard, and fast packet-switched networks based on frame relay and ATM technology.

By specifying the design parameters, hybrid voice networks can be modeled using all types of leased and virtual facilities with or without multiplexing. Network managers can also optimize a backbone network topology and model any given topology (for cost and routes).

The Analysis menu allows the designer to model/analyze any point-to-point and several multilink paths for congestion/queuing delays, LAN performance, and reliability. Another Analysis menu item allows the computation of the equivalent monthly cost of hardware and payoff periods for privately owned hardware and transmission facilities. The following section outlines a case study of an EcoNets implementation.

AN ENTERPRISE NETWORK PLANNING AND DESIGN CASE STUDY — ECONETS

The enterprise in this case study manufactures, distributes, markets, and maintains highly specialized intelligent workstations. It has 17 sites scattered across the U.S., with headquarters in Las Colinas, TX. Two separate networks serve the enterprise. A voice network connects all 17 locations (or PABXs) to a voice switch located in Las Colinas with leased voice-grade lines (VGLs). A separate data network connects workstations located at all of its locations to a host using the SNA-BSC protocol and 9600-bps lines. The newly appointed network manager wants to study the feasibility of a new network architecture, so a consultant is engaged to study the problem.

A database (a subset of the EDB) for network design was created and is outlined in Table 1.

Table 1 Enterprise database (EDB) for a 17-node network design (voice/data applications).

*********** NODAL DEFINITION DATA ***********

N#	-V-	-H-	LOAD (BPS/MEs)	LATA	LINK	NAME
1	8438	4061	40000	552	0	LCLNTX
2	8436	4034	5000	552	0	DALLTX
3	8296	1094	1300	952	0	SRSTFL
4	8360	906	1300	939	0	FTMYFL
5	6421	8907	1300	674	0	TACMWA
6	6336	8596	1300	676	0	BELVWA
7	4410	1248	1400	128	0	DANVMA
8	6479	2598	1300	466	0	VERSKY
9	9258	7896	1300	730	0	TOAKCA
10	9233	7841	1400	730	0	NORWCA
11	9210	7885	1400	730	0	WLAXCA
12	7292	5925	1400	656	0	DENVCO
13	7731	4025	1300	538	0	TULSOK
14	7235	2069	1300	438	0	NORCGA
15	5972	2555	2500	324	0	COLMOH
16	9228	7920	2500	730	0	STMNCA
17	8173	1147	2500	952	0	TMPAFL

Tot. BHR Traffic = 68500

*********** Node (N) Link (L) Type (T) [NLT] FILE PRINTOUT ***********

******* LEGEND *******

{ C = Link Cap.: Max R = Max. Allwd. Rate (Wm): MF = VM pxg. Fact.: FPF = Priv. Fac. Fact.}

L Type	Link C	Max Link R	MF	Tariff#	FPF
1	9600	6300	1	1	1
2	56000	48000	8	2	1

3	1544000	1440000	24	3	1
4	45000000	40000000	672	4	1

******************** TARIFF DATA PRINTOUT ********************

TARIFF # = 1 AVG. LOCAL LOOPS CHARGES ($) = 294

MILEAGE BANDS:					
	50	**100**	**500**	**1000**	**10000**
FIXED COSTS ($):					
	72.98	149.28	229.28	324.24	324.24
COST PER MILE ($):					
	2.84	1.31	0.51	0.32	0.32

TARIFF # = 2 AVG. LOCAL LOOPS CHARGES ($) = 492

MILEAGE BANDS:					
	50	100	500	1000	10000
FIXED COSTS ($):					
	232	435	571	1081	1081
COST PER MILE ($):					
	7.74	3.68	2.32	1.3	1.3

TARIFF # = 3 AVG. LOCAL LOOPS CHARGES ($) = 2800

MILEAGE BANDS:					
	50	100	10000	10000	10000
FIXED COSTS ($):					
	1770	1808	2008	2500	2500
COST PER MILE ($):					
	10	9.25	7.25	7.25	7.25

TARIFF # = 4 AVG. LOCAL LOOPS CHARGES ($) = 8000

MILEAGE BANDS:					
	10000	10000	10000	10000	10000
FIXED COSTS ($):					
	16600	16600	16600	16600	16600
COST PER MILE ($):					
	47	47	47	47	47

***************** SYSTEM DESIGN PARAMETERS *****************

= 0

ATP/D = 3	UPR/D = 56000	HPR/D = 56000	IML/D =2 8	RML/D = 300
Ncu/D = 4	Rmph/D = 100	HTT/D = 0.001	Fopt/D = 0	Tnp/D = 10
Thm/D = 4	Kpg/D = 0.01	BKL/D = 64	ICPB/D = 56	TGF/C = 1
Flk/C = 0	Fnn/C = 1	Flt/C = 1	Fftd/C = 0	NA = 0
ALT/V/D = 1	NA = 0	Bal/V/A = 0.1	ECC/V = 13.33	ECD/V/A = 300

DREQ/A = 60	PEXD/A = 0.15	Clbr/A = 23	Frst/A = 1	ACDT/A = 2
TKLT/V/D = 1	NA = 0	Btk/V = 0.1	Ffdx/D = 1	MTKU/D = 0.8
BBTF/C = 2	Vmin/C = 3000	Vmax/C = 10000	Hmin/C = 0	Hmax/C = 10000
Fvc0/C = 0	Fvc1/C = 0	Fvc2/C = 0	Fvc3/C = 0	Fvc4/C = 0
Fvc5/C = 1	Fvc6/C = 30	Fvc7/C = 0	Fsh/D = 0	Fnp/C = 1
DPM/A = 30	Fdis/C = 1	NA = 0	TFXC/A = 1	NDEC/C = 7
DECT/C = 1	/A = ACD = 0	/C = Common = 0	/D=Data = 0	

**************** NAMES OF INPUT FILES ****************

VHD17* LINK17* MAPusa* NLT* TARIFF* SDF* NAME17* FTF1* LATA17*

FILES.TXT* CSABDS* UTBL* WUTBL* MUTBL* RSTBL* DTP8* Swf2*

**************** DAILY TRAFFIC PROFILE ****************

Hour Numbers & Corresponding Fractions of Daily Traffic are as follows:

1	0	2	0	3	0	4	0	5	0	6	0
7	0.05	8	0.1	9	0.1	10	0.1	11	0.1	12	0.1
13	0.1	14	0.1	15	0.1	16	0.1	17	0.05	18	0
19	0	20	0	21	0	22	0	23	0	24	0

****** Switch File Definition ******

Number of Switches = 2 @ 11, 1

Networking Menu Item No. Employed = 6

The 17 sites, their vertical and horizontal coordinates, and busy-hour TCA of traffic intensities are shown for both voice (in millierlangs) and data (in bps). Also shown are their names according to a six-symbol city-state (CCCCST) code. Next, an NLT file is defined for these link types. The various design parameters are defined in the SDF. The design parameters for the voice network define the access link type, desired blocking on access lines, trunk-line type, and desired blocking on trunks. The major design parameters for the data network are ATP (analysis type is equal to 3 for response time modeling for an SNA-BSC network), user port rate, host port rate, nodal processing time in ms for each transaction, and half-modem time in ms spent in going through the modem in one direction.

The consultant first modeled the existing voice and data networks. The monthly costs for these two separate networks were $60,930 and $10,017, respectively. The EcoNets tool was then used to study various topologies consisting of switches and three link types for voice and only the 9600-bps line for data (higher-speed lines resulted in no improvements). The results are shown in Exhibit 1. The optimum voice network topology (see Exhibit 2)consisted of two switches (as determined by the EcoNet's center-of-gravity finding item on the Networking menu) and 56K-bps lines, each of which carries eight digitally encoded voice conversations.

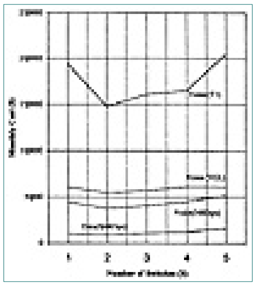


Exhibit 1 Costs versus number of switches and link types.

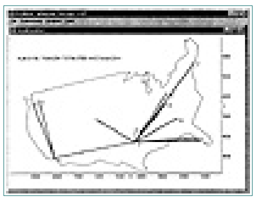


Exhibit 2 Optimum star data network topology for IVD application.

The one-time cost of 17 special hardware boxes that perform voice encoding and multiplexing in the same box did not influence the optimum network topology. The optimum data network topology (see Exhibit 3) consisted of the same two switches as were used for the voice network and 9600-bps lines. The costs of these optimum networks were $37,546 and $9147, respectively. This represented a monthly savings of $23,254 (or about 32.8% of existing costs). No matter how the figure is examined, it amounts to a substantial savings.

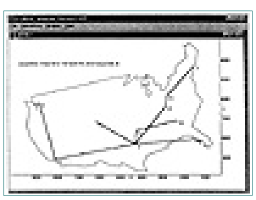


Exhibit 3 Optimum MD-data network topology with two switches.

Additional savings can be achieved by computing the total data rate (in bps) of voice conversations from each site and adding the regular data traffic and constructing a new VHD file. An optimum star-data topology consisting of two switches and 56K-bps lines can be achieved. The topology is identical to that of the optimum voice network (see Exhibit 1) and the monthly cost is about the same. The cost of the separate data network disappears completely. The new monthly savings of $33,392 represent 47.1% of existing costs. These additional savings resulted from the fact that the 56K-bps line used in the integrated voice/data network had enough excess capacity to handle the data traffic. Such a phenomenon is similar to the one experienced by network managers working with larger T1 networks in the 1980s. Those networks had enough excess capacity in the T1 trunks to handle the data traffic. The broadband data networks of the future should have enough excess capacity to handle voice traffic.

This example illustrates only a small enterprise network. Bigger savings can be achieved through optimization of larger enterprise networks. Savings result because integrated networks make use of excess capacity, and aggregation of many separate applications allows the deployment of transmission facilities with higher capacities that generally cost less on a per-transaction basis. These types of network planning and design tools provide network managers with many more opportunities for providing a cost-effective, integrated network to the enterprise.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous	Table of Contents	Next

Chapter III-5
Windows NT Project Planning

Bill Camarda

This chapter, intended for organizations that are planning a migration to Windows NT, covers:

- identifying the organization's goals for deploying NT
- identifying resources that can assist in the migration process
- evaluating the organization's installed base
- choosing vendor partners to assist in the migration
- planning a pilot rollout
- running a full-scale rollout

ESTABLISHING BUSINESS AND TECHNICAL GOALS FOR NT DEPLOYMENT

Before beginning a migration to Windows NT, it is important to understand what the business and technical goals are.

Business Goals for NT Deployment

Exhibit 1 lists some possible business goals for an NT project, along with the deployment issues that will need to be addressed to reach those goals.

Exhibit 1 Business goals and deployment issues.

Technical Goals for NT Deployment

Exhibit 2 lists some possible technical goals for an NT project, along with deployment issues related to each of these goals.

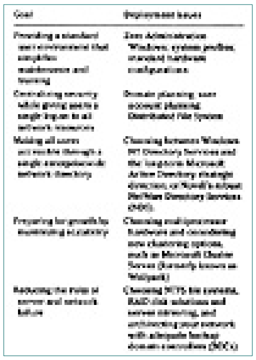


Exhibit 2 Technical goals and deployment issues.

Building the Migration Team

Once the goals have been established, the next step in planning a smooth migration to NT workstation and/or NT server is to divide the responsibilities. Most large organizations identify several teams, each with a leader and a specific role in planning and deploying Windows NT. These teams may include:

1. A planning and coordination team that includes the project leader and representatives of each other team

2. An executive team that includes the IT organization's leader (or project manager responsible for the NT deployment), others with authority over relevant IT procedures, finance management, and executives from lines of business that will be impacted by the NT deployment

3. An installation team consisting of technicians who will actually install NT, as well as technical experts who can evaluate and test configurations for performance and compatibility

4. Training and/or support team(s) that may include help-desk representatives, internal trainers, those responsible for hiring external trainers, and decision-makers responsible for providing adequate resources to frontline support staff

In organizing these teams, it is all too common to disregard the central role of users — the people who must ultimately be productive with NT on a day-to-day basis. Bringing user representatives into the process early improves the likelihood of achieving wholehearted buy-in, and substantially improves a company's chances for success.

IT Qualifications Needed for the Deployment Team

One way to identify the right internal and external resources for the NT deployment team is to work with Microsoft Certified Professionals (MCPs). In addition to hiring certified professionals, an organization might decide that certifying more of its existing IT staff as NT experts should be an important element of the deployment process. A company could consider providing training for installers, system administrators, support staff, and anyone else with day-to-day responsibilities for Windows NT systems.

There are currently four MCP certifications, and it is important to understand the differences among them:

- Microsoft Certified Systems Engineers (MCSEs) have passed four operating-system exams, at least two of them related to Windows NT. There are currently two tracks: one for Windows NT 3.51 and another for Windows NT 4.0. The NT 4.0 core requirements contain deeper coverage of NT 4.0 Server deployment issues.
- Microsoft Certified Solutions Developers (MCSDs) have passed two core exams covering Windows 32-bit architecture, OLE, user interface design, and Windows Open Services Architecture components, along with two elective exams covering Microsoft development tools and/or SQL Server.
- Microsoft Certified Product Specialists (MCPSs) have passed one detailed operating-system exam — either Windows NT Workstation, Windows NT Server, or 16-bit Windows.
- Microsoft Certified Trainers (MCTs) teach (or intend to teach) at Microsoft Authorized Technical Education Centers; Microsoft certifies both their subject-matter expertise and exposure to some basic training techniques.

Gathering Information Resources

It is helpful to gather as many deployment resources as possible early in the planning process. Not surprisingly, Microsoft is a prime source of free and low-cost information on deploying Windows NT. In addition to the Windows NT Workstation 4.0 Resource Kit and Windows NT Server 4.0 Resource Kit, Exhibit 3 lists many of the free resources available on Microsoft's Web site.

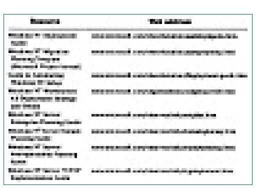


Exhibit 3 Selected resources available on Microsoft's web site.

EVALUATING THE INSTALLED BASE

Now that the teams are in place, it is important to thoroughly understand the computing environment into which Windows NT is being deployed. Consider desktop PCs and servers that may be upgraded to, replaced with, or served by Windows NT systems; and other equipment, especially mainframes, minicomputers, UNIX servers and workstations, and NetWare servers that Windows NT must coexist with.

It is extremely helpful if a company has a detailed inventory of the desktop and server systems in use throughout the organization. If no inventory exists, the company will need one before initiating a full-scale rollout. For the moment, however, the IS department can identify a representative sample of systems begin creating standard configurations, testing them, and making “upgrade vs. replace” decisions.

Planning for New Workstations

If the organization is purchasing new workstations, IS should seriously consider standardizing on one brand of PC throughout the organization. This has several benefits, including:

- A single point of contact for technical support (both hardware and Windows NT), troubleshooting, upgrades, and accountability
- A single approach to manageability. While the new NetPC and Intel’s proposed “Managed PC,” may standardize hardware management, each leading vendor currently has its own approach — not necessarily compatible with anyone else’s.
- Pricing leverage associated with quantity purchasing

If at all possible, test proposed new workstation configurations on both Windows NT 4.0 and the betas of Windows NT 5.0 that have become increasingly available throughout late 1997 and early 1998. While early betas will not fully reflect the performance of the final product, preliminary NT 5.0 benchmarks can help an organization ensure that its systems will be useful well past the year 2000.

| Previous | Table of Contents | Next |

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Upgrade Planning for Existing Workstations

The installed base and normal upgrade schedule will play an important role in determining which existing systems are worth upgrading. Within this context, IS should start by determining the lowest-performance PC worth upgrading to NT Workstation. While NT Workstation can theoretically run on a low-end 486, most companies restrict upgrades to systems with substantially more power.

For example, where processing requirements are modest, a company might use Pentium 75 or Pentium 90 systems as a preliminary cutoff point. Systems slower than this would rarely be considered for upgrades to Windows NT. Starting from this baseline, the IS department would then test borderline systems to determine whether they will deliver adequate performance.

In most cases, these systems will require memory upgrades — probably to at least 32 MB. To determine whether these systems are in fact worth upgrading, IS must price the time and cost of these upgrades and consider the remaining useful life of the hardware.

Where processing requirements are more substantial, IS might want to start with Pentium 133, 150, or 166 systems as a cutoff point. Again, it is important to take memory upgrades into account, although it is possible that some of your Pentium 133–166 systems are already configured with adequate memory to run Windows NT Workstation.

Whatever level is selected, it is helpful to check whether the organization’s representative systems appear on Microsoft’s Windows NT Hardware

Compatibility List. Many major vendor systems intended for business use do appear on the list, although not all. While many systems that have not been certified by Microsoft will run Windows NT successfully, Microsoft will not support these configurations. If a company owns systems that do not appear on the Hardware Compatibility List, IS must decide whether to take full responsibility for supporting these workstations running NT. If these systems cannot be replaced, the organization may wish to identify third-party or vendor resources that can assist in maintaining them.

Before beginning testing, make sure the version of Windows NT selected reflects the latest Service Packs introduced by Microsoft. To check for the latest Service Pack — and to download it — visit: www.microsoft.com/NTServerSupport/Content/ServicePacks/Default.htm.

It is not enough to test Windows NT on standalone systems. IS should set up a test network that is as representative of the planned network as is practical.

Evaluating and Testing Software

An organization does not need to test software carefully before deploying it widely on Windows NT systems. The following sections describe some of the issues to take into account.

Win16 (Windows 3.x) Applications. Older 16-bit Windows applications will run in the new Windows 4.0 16-bit subsystem, a Windows 3.x emulator sometimes called “Windows on Windows” or WOWEXEC. IS will need to test both the reliability and the performance of applications running on this emulator. IS could also have to determine whether to run these applications in their own memory space (the default setting) or in a shared memory space with other Win16 applications (potentially faster, but one failed Win16 application can crash all Win16 applications running at the same time).

DOS Applications. Many people have discovered that DOS programs will not run in Windows NT because they must address hardware directly. Many other DOS applications will work via Windows NT Workstation’s DOS emulator. If the company still depends on DOS applications, they should be tested carefully. IS may also have to experiment with settings in each DOS application’s Properties dialog box to maximize performance — especially settings in the Memory tab that control the amount of conventional, expanded, extended, and MS-DOS-protected-mode memory available to an application.

Windows 95 Applications. Even if all the applications are 32-bit Windows applications that utilize the Win32 API, there are a few “gotchas,” including:

- APIs specific to Windows 95 that are not available in Windows NT Workstation 4.0, such as Direct3D, Independent Color Matching, Plug-and-Play, “Flat Thunks,” and the Pen API, as well as these Windows 95 OSR 2-specific APIs: FAT32 File System, DirectX 2, ActiveMovie, and Windows Internet Extensions API
- APIs that are common to both operating systems but may work differently, including Unicode and some security attributes

Upgrading Custom Applications

Many companies depend on custom applications originally written for 16-bit Windows 3.x environments. For performance and compatibility reasons, organizations will usually want to port these applications to the Win32 API rather than running them in the “Windows on Windows” emulator.

This is, as programmers say, a nontrivial task. There are significant differences between Win16 and Win32 applications. In the Visual Basic environment, these include differences in naming, treatment of integers, and string routines; deserialized input; changes due to preemptive multitasking; and changes to DLLs. Bottom line: if the company wants to roll out revised Win32 custom applications when it rolls out Windows NT, IS should start updating those programs now.

Identifying, Evaluating and Testing Peripherals

In addition to the PC hardware itself, IS must systematically identify all the peripherals and other devices the organization expects to use with NT Workstation or NT Server. Once this is done, IS should review Microsoft’s Hardware Compatibility Web Page for Windows NT (http://www.microsoft.com/ntworkstation/hwtest.htm) to determine whether these devices have NT 4.0 compatible drivers. If in doubt, visit the vendor’s Web site. IS should consider each of the following:

- video cards
- video-capture cards
- audio cards
- SCSI host adapters and devices, including CD-ROM drives, tape drives, removable media and scanners
- other (non-SCSI) CD-ROM drives and tape drives
- network interface cards (Ethernet, Fast Ethernet, ATM)
- ISDN adapters
- modems and multiport serial adapters
- printers
- PCMCIA (PC Card) devices
- uninterruptible Power Supplies
- mice and other pointing devices

If IS is deploying NT throughout the entire organization, then the availability of final-release (not beta) NT drivers should be considered as a prerequisite for future purchases. NT drivers should be tested carefully — especially video drivers that now run in the Windows NT kernel, where they can potentially impact Windows NT’s stability.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Integrating Macintosh Desktops

If an organization has an installed base of Macintoshes that it does not intend to replace with Wintel systems, its testing must encompass Windows NT Server Services for Macintosh. Services for Macintosh, a standard component of Windows NT Server, makes it possible to:

- create Macintosh volumes on an NT Server system
- support AppleShare, allowing NT disks and folders to appear on Macintosh client desktops
- support AppleTalk networks, including AppleTalk Internet Routers — eliminating the need to purchase additional routers to support Macintoshes
- provide file sharing and print services to Macintosh clients

Microsoft even provides a Web administration tool (available at www.microsoft.com/ntserver/webadmin/webadmindl.htm) that makes it possible to administer an NT Server system from a Macintosh (or any other) client, using a Web browser front-end.

There are some limitations to Windows NT Server's Macintosh support. Macintoshes cannot access other Windows clients, and neither the server nor other Windows clients can access files stored locally on a Macintosh. If these limitations will be a problem in an organization's environment, then IS might consider third-party products such as Dave (Thursby Software Systems, Inc., www.thursby.com, 1-817-478-5070) or MacLAN Connect (Miramar Systems, www.miramarsys.com, 1-800-862-2526).

UNIX Servers and Workstations

Windows NT is increasingly being introduced into UNIX environments. Windows NT can integrate with UNIX workstations and servers, but users will probably have to rely on third-party products to accomplish their goals. Most companies need one or more of the following elements of UNIX/NT interoperability:

- Network File System (NFS) file and printer sharing
- X terminal access from Windows NT workstations to run X applications hosted on UNIX systems
- X terminal access from UNIX hosts to Windows NT, so UNIX systems can display Windows applications

As part of the project planning and testing, IS must identify the company's needs for UNIX interoperability and compare the products available to provide it. Of course, this means that project teams will need individuals with significant UNIX experience.

Mainframes/Hosts

Some organizations may be planning to use Windows NT as a client/server platform that supplements mainframe-based legacy systems, or helps to migrate away from them. If so, IS should consider SNA gateways and associated hardware designed to:

- improve desktop workstation response time
- support query-intensive and communications-intensive applications
- add redundancy and load balancing
- support advanced groupware and intranet solutions cost effectively

As traditional dumb terminals have been replaced by PCs running terminal emulation software, traditional cluster controllers are also being replaced by SNA gateway software and hardware. These solutions, such as Microsoft's SNA Server and IBM's Comm Server, are typically much less expensive to purchase, install, and support than controllers were. They typically offer better performance as well.

SNA Server offloads network traffic from mainframes and IBM AS/400 midrange systems, freeing up host resources for line-of-business applications. It serves as a TCP/IP-to-SNA gateway, helping companies migrate to TCP/IP while retaining the reliability and securities advantages of SNA.

SNA Server also extends NT Server's existing domain-based unified sign-on capabilities to mainframe and AS/400 systems, so users who have been authenticated by a Windows NT domain controller can gain access to files, printers, databases, messaging systems, and other applications running on hosts — consistent with security restrictions that IS establishes.

To take full advantage of SNA Server or products like it, an organization needs robust server hardware. It may be necessary to integrate third-party mainframe channel adapters as well. For example, companies such as General Signal

(888-GSN-DATA, www.gsnetworks.com) and Polaris Communications (1-800-353-1533, www.polariscomm.com) deliver PCI-based boards that support IBM's Enterprise System Connection (ESCON) high-speed connectivity.

Before deploying SNA Server, IS should check with its Microsoft account representative to understand Microsoft's strategic direction for this product. It has been rumored that Microsoft may eventually fold SNA Server functions into NT Server and SQL Server.

Planning Issues to Handle Concurrently with Evaluation and Testing

As IS evaluates its installed base of hardware and software, the NT deployment teams can concurrently consider several other important issues. For example, they can:

- Create budgets and timetables for the deployment and rollout
- Determine which NT capabilities to deploy, which to disable, and which to deploy only on selected workstations
- Decide whether to deploy NT using Microsoft System Management Server (SMS) or third-party software-delivery tools
- Plan for training installers, Help Desk personnel, and trainers

CHOOSING AND MANAGING VENDORS

Among the most critical decisions IS will make is the choice of vendors to partner in the deployment of Windows NT. Many companies want the business; this is not surprising because services tend to deliver much higher margins than commodity hardware sales.

Given the rapid growth of Windows NT in the enterprise, many companies are focusing on delivering Windows NT services. These include major consultancies and system integrators such as EDS and Entex, as well as the services organizations of traditional hardware suppliers such as Digital and IBM. If NT is being deployed in a smaller company, that organization might choose a local or regional systems integrator or client/server developer to assist.

Previous | Table of Contents | Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Vendors with Strategic Microsoft Relationships

Microsoft maintains especially close relationships with some suppliers of PC and server hardware. These relationships certainly do not preclude users from choosing other suppliers, but they may be worth considering when making vendor decisions. At minimum, IS may want to question Microsoft's strategic partners on how they are delivering the benefits their alliances are supposed to provide — and question competitors on how they can deliver comparable benefits.

On the server side, Microsoft has announced especially close partnerships with Digital (DEC) and Hewlett-Packard. To varying degrees, these relationships have led these vendors to deliver more fully integrated solutions for NT deployment. For example, Digital already has 1400 engineers with NT certification from Microsoft. Digital's service offerings for NT include:

- Legacy NOS Migration to Windows NT services, including methodologies for migrating user files, access privileges, print services, and other features of legacy NOS environments, including NetWare, Banyan VINES, DEC PATHWORKS, IBM LAN Server
- Building NT Applications for the Enterprise, services to plan, design, and implement client/server computing built on Windows NT
- Software Support for Windows NT, services to provide support for Windows NT environments, with a selection of response times and problem-resolution capabilities
- Installation and Startup for Windows NT and Windows NT Clusters, services to rapidly install, configure and implement NT servers and

clusters

Hewlett-Packard's recent partnership with Microsoft has thus far led to improvements in scalability on Intel-based servers, as well as a new family of Business Recovery Services intended to help companies prevent and recover from failures associated with NT servers.

It is important to note that while Microsoft's NT partnerships with Digital and Hewlett-Packard have been the most prominent to date, other leading vendors now offer extensive support for Windows NT. To cite just two examples, Unisys recently established the Enterprise NT Services organization, intended to offer a full suite of services for deploying mission-critical applications on NT systems; NCR also offers substantial Windows NT consulting services.

On the workstation side, Microsoft and Intel lead the NetPC effort intended to lower the cost of administration; partners in this effort include Compaq, Dell, Digital, Gateway 2000, Hewlett-Packard, Packard Bell NEC, and Texas Instruments.

As already mentioned, standardizing on a single provider of PCs, a single provider of servers, and a single provider of network-interface cards can simplify management tasks for years to come. Whether preparing a formal RFP, or requesting proposals on a less formal basis, above all it is important to be explicit about what is expected from the vendor. In addition to cost, the following items should be considered when making choices:

- a track record with your company or companies like yours
- availability of a trustworthy single point of contact
- ability to deliver an end-to-end solution
- support commitments, both for hardware and for Windows NT
- availability of specific technical resources where and when you need them
- product delivery dates and a vendor's track record in meeting them — especially if notebook PCs are involved
- contract flexibility (e.g., the ability to substitute more-advanced technologies for those covered in the contract)

Testing Standard Configurations

Once IS has established standard configurations of both existing and new equipment, they should run detailed tests of:

- the NT installation process and automated batch scripts
- network connectivity
- applications software
- the uninstall process (restoring previous operating systems on upgraded computers)
- disk space variables (both for installation and swap files)
- local and server-based administration tools

IS may find that it needs to make adjustments to the standard client configuration to improve performance, compatibility, stability, or user convenience.

PLANNING AND MANAGING A PILOT ROLLOUT

In most organizations, the next step is to perform a pilot rollout in a small department or division. Ideally, choose an organization that is open to new technology and not stressed by a major project deadline or recent downsizing. (Obviously, some organizations will not have this luxury.) A typical pilot rollout could include the following steps:

1. Prepare a detailed logistical plan for the pilot rollout, including tested scripts for automated installation from distribution servers.

2. Prepare and implement a support plan, so users have immediate access to assistance when they need it.

3. Plan a schedule (e.g., how many systems can be upgraded per day, and how long the pilot rollout will take).

4. Notify users well in advance of the installation.

5. Develop training materials that reflect both the performance of standard NT tasks and concerns unique to your company (e.g., logons, file locations, custom applications, etc.).

6. Schedule training.

7. Perform a verified backup, a virus check, and disk defragmentation on all pilot machines that will be upgraded to Windows NT.

8. Virus check all pilot machines prior to installation.

9. If necessary, upgrade BIOSes, memory, or other hardware prior to installation.

10. Make sure NT driver software is available for installation wherever needed.

11. Run the installation at a time least likely to interfere with deadlines.

12. Once the pilot installation has taken place, IS should:

13. Follow up to ensure all systems are working properly.

14. Stay in close contact with all members of the pilot group to identify problems, questions, and other issues.

15. Respond to user concerns and carefully track the changes that users request or require.

16. Assign technicians to check real-world performance against expectations, so adjustments can be made, if needed, before a full-scale rollout.

17. Compare schedules and costs against expectations, so budgets and timeframes can be adjusted for the full-scale rollout later.

PREPARING FOR FULL-SCALE DEPLOYMENT

Now that the pilot rollout has been conducted, IS can begin to prepare for the organizationwide rollout by:

- Creating budgets and schedules that reflect the actual experience.

- Revising the company's IT procedures to reflect the changes that NT will require, and notifying users where necessary.
- Performing a complete systems inventory and storing the information in a centralized database that can be updated to reflect changes to individual systems.
- Hiring or reassigning any additional staff needed for the full-fledged rollout.
- Rolling out Windows NT using the procedures used in the pilot rollout — adapted, of course, to reflect any necessary changes.

CONCLUSION

This chapter has discussed many of the issues involved in successfully rolling out Windows NT Workstation. But Windows NT is inherently a networked operating system, designed for use in highly distributed environments. It is important to review the critical networking issues associated with planning an NT deployment — including what is needed to architect Windows NT domains that will serve your company well for years to come.

| Previous | Table of Contents | Next |

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter III-6
Networked Imaging Systems Requirements

Nathan J. Muller

Computerized document imaging systems allow businesses and government agencies to automate paper-based workflow processes by scanning documents into digital form and storing them in massive image databases for instant access. When implemented over a LAN, storing documents on an image server can solve many of the problems commonly experienced with paper files, including that of multiple access. The digitized document images can be retrieved and processed in sequence by multiple workstation operators until the entire transaction is complete.

The automation of document workflow can be compared to an assembly line where repetitive, paper-intensive tasks are performed at image-enabled workstations. There, resource allocation can be monitored and productivity improved through automatic document distribution and routing. Special workflow management software controls the flow of documents from station to station, typically eliminating many intervening document processing stages and streamlining others. Some imaging systems even allow the documents to be updated by different users, whereupon the updated document is filed in order of last update first. That way, a complete history can be maintained in an electronic case file.

When the WAN is used to connect geographically separate LANs, the benefits of imaging can be extended throughout the entire enterprise. The size of image files vs. other types of traffic normally traversing LANs and WANs, however, can affect existing networks in terms of slowing access and response time unless such networks are upgraded or optimized to handle the increased load.

A full-time image application involving document scanning and distribution, file retrieval and annotation, and case file assembly and transfer, can easily dominate a LAN and possibly overwhelm it. For this reason, many companies should modify their networks to accommodate imaging applications.

Network upgrades can be accomplished in a number of ways, including the addition of higher-performance equipment and transmission media to build higher-speed backbone networks, the use of file compression and forms reduction software, or subscription to carrier-provided digital services. Understanding the impact of imaging on existing networks and carefully evaluating the available alternatives will permit maximum levels of efficiency and economy to result from the imaging application, while ensuring the highest levels of availability and performance for the entire corporate network.

PROCESSING REQUIREMENTS

The success of document imaging often hinges on how well the system's implementors understand business processes and workflows. For them, the most obvious task is to learn how employees execute their responsibilities and solicit their input on how things can be improved. The participants should settle procedural problems early so that workflows can be properly scripted for automation and documents routed to the appropriate workstation operators. Mapping existing departmental procedures usually reveals processes that duplicate effort, employees working at cross-purposes, and unnecessary paperwork and filing requirements.

The key planning undertaking is an evaluation of document-processing needs. At minimum, this should include the following:

- Listing all documents that are currently being processed in the course of business
- Determining how many of each type of document arrives at each business location daily, weekly, monthly, and yearly
- Finding out who provides the data in the documents, the purpose of the data, and what information systems or applications currently use the data provided in the documents
- Establishing which information in each document is most frequently used
- Determining appropriate index fields for each type of document
- Preparing a flowchart for each type of document that shows the path it follows when it is received in the office — the stops it makes, what happens at each stop, and what alternative paths exist

Company structures adapt to changing customer needs and market directions. Imaging tasks, to be implemented effectively, should be critiqued against the same standards. Automating an inefficient process only wastes corporate resources. The best applications for imaging technology are:

- those whose contribution to the company's core business function is widely understood and those whose need for improved quality and timeliness is readily apparent

- those whose workflows involve repetitive tasks, which lend themselves to automation
- those in which the time spent in paper handling can be dramatically reduced.
- those for which a significant positive return on investment is likely
- those in which early implementation mistakes will not jeopardize mission-critical functions and place the company at risk

Merely overlaying imaging technology on an existing workflow will almost always produce the smallest benefit. There are several levels of technology implementation that merit consideration, categorized by increasing levels of process reorganization:

- *Pilot projects*. These are created for limited use at one location. Instead of fostering commitment, they often promote a wait-and-see attitude that more often than not guarantees failure.
- *Internal integration*. Using imaging to transfer information between processes may yield significant benefits. If existing processes are not made more efficient first, long-term gains will be limited and return on investment prolonged.
- *Process redesign*. An entire organization or discrete departmental process can be restructured to take advantage of imaging technology. Although this method can produce noteworthy improvements in efficiency, it can also be difficult and time-consuming to execute.
- *Network engineering*. Extends the reorganization process to locations outside the company's main location. This method can produce enterprisewide benefits but is more difficult to implement.

Previous | Table of Contents | Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

To properly implement a technology at the right level of process reorganization requires that companies strive to understand their business processes, which is not as easy as it sounds because it involves a commitment of staff, effort, and time. If many corporate managers have a stake in preserving the status quo, this complicates matters. Having an outside consultant evaluate various business processes and workflows may render a more objective and accurate assessment. This, in turn, helps ensure that the investment in corporate resources can be targeted wisely. Internal evaluations of business processes and workflows can be performed, with third-party assistance or third-party review to validate, or invalidate, the conclusions. The high cost of imaging system implementation justifies these extended measures.

LAN CONSIDERATIONS

The most popular LAN types are Ethernet and Token-Ring, and although most companies already have LANs by the time they are ready to add image applications, they often have a choice between Ethernet and Token-Ring because both are available among various workgroups and departments. The question is not which LAN to build, but which one to extend to accommodate the imaging application. Additionally, there are several standards to choose from when building high-speed backbones that are used to interconnect multiple LANs or high-performance workstations, or both. These standards include FDDI 100Base-T Fast Ethernet, and VG-Any LAN — all of which offer a data rate of 100M bps.

Depending on the existing traffic load, adding an imaging application may require that the legacy LAN be reconfigured or upgraded to maintain an acceptable level of performance for all the applications. If the imaging application involves continuous use of the LAN for large files, a dedicated LAN may be required. Alternatively, the workgroup running image applications may be partitioned from the rest of the LAN so as not to affect the performance of other applications.

Ethernet

If the bus network remains fairly static, traffic is uniform over time, and users can tolerate some delay from retransmissions caused by collisions, an existing Ethernet may be an economical choice for supporting imaging applications. Networks have a tendency to grow, however, as organizational needs become more varied and sophisticated. As the number of workstations increases and the volume of traffic grows on the Ethernet, so does the possibility of collisions, which slow down the network with retransmission attempts.

Performance Tuning. Using LAN management systems, analysis tools, and utilities, the network manager can accurately measure performance and take immediate steps to make improvements, such as segmenting the network into subnets, caching images, using file compression, and adding higher-performance peripherals. Since performance demands may vary on a daily basis, the ability to respond quickly constitutes a key benefit of LAN-based imaging systems over mainframe-based systems.

Performance on the main imaging network can be maintained by putting resource-intensive services, such as scanning, printing, and faxing, on subnetworks. These subnets can be selectively isolated from the rest of the network using bridges or routers. This allows a large accounts-payable department, for example, to scan 10,000 invoices a day without bogging down the main network where users are trying to retrieve data. Scanned images are stored on disk and registered in the image system index using a batch process. Similarly, when a large image print job is scheduled, the data can be dumped off to a subnet for printing, instead of tying up the main network.

For those who want to integrate standard data and images on the same Ethernet, there are several issues that must be addressed, among them, packet size. The maximum packet size for the IEEE 802.3 standard is about 1,500 bytes. Larger packet sizes allow an 802.3 network to approach the theoretical maximum throughput of the network, which is 10M bps. This is because larger packet sizes decrease the likelihood of information-destroying packet collisions.

It is possible to implement networks with packet sizes that exceed the 802.3 standard to improve LAN performance during image transfers. Implementing non-standard packet sizes has several implications, however. Larger packet sizes violate the 802.3 standard, so steps must be taken to ensure that this violation does not negatively affect network hardware and software designed to the standard. This can be accomplished by ensuring that network interface boards on all network stations have enough buffer memory to support non-standard packets. Such interface boards are usually more expensive, however. If adjacent networks are not configured to support larger packet sizes, internetworking can become more complex. This is because a packet-restructuring function must exist at network interface points and, due to the increased processing that must be done, network performance may degrade.

Shared Versus Dedicated Networks. Image planning becomes quite difficult when there are other types of data flowing on the network, so it becomes necessary to consider whether a dedicated imaging network or a shared network is better. The main issues are the predictability of the traffic pattern and the performance required. Transmitting mixed data types across a network does not necessarily cause problems. If, however, problems occur, performance optimization becomes a requirement, and this may entail compromises that affect all users.

Although due consideration should be given to all users in the mixed-usage network, image-intensive production environments, such as an insurance claims-processing application, are likely to require dedicated networks. In these instances, network parameters can be more easily tuned to optimize performance for predictable traffic patterns. To be on the safe side, a pilot image network can be segmented from the production network to test all of the assumptions about LAN performance under varying load conditions.

Previous	Table of Contents	Next

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

The Impact of Protocols. Protocols, too, have an impact on LAN performance. Deterministic protocols, characteristic of Token-Ring, provide a more predictable worst-case response time if many stations have information to send. In low-traffic situations, however, deterministic protocols also impart a fixed-delay overhead. Contention media-access protocols, characteristic of Ethernet, provide immediate access to the transmission medium when there are few stations with information to send, or when a station infrequently sends information. The performance of a contention medium-access protocol degrades significantly if there are many stations attempting to generate traffic at the same time.

If delay proves to be a persistent problem, consideration should be given to putting the image application on its own LAN. This can be accomplished very economically over the same type of unshielded twisted-pair that already connects telephones to the corporate PBX. Enough excess wiring exists in most offices that no additional installation is usually required. The reliability of telephone wire for LANs has been improved with the adoption of the 10Base-T standard.

Among the advantages of 10Base-T is that it does not require existing Ethernet installations to be retrofitted with special adapters at the wall and wire closet; it works with standard IEEE 802.3 Ethernet interfaces and the eight-pin modular jacks typically found on most business telephones. Centralized 10Base-T hubs can be created to compress and route image traffic, provide network management capabilities, and provide access to the WAN via optional bridge router modules.

Token-Ring

Among the advantages of Token-Ring over Ethernet for imaging applications is that access is not contention-based, but deterministic, meaning that each station has guaranteed access to the ring network in turn. Therefore, a higher throughput rate is possible on the Token-Ring in heavily loaded situations, even when image transfers are factored into the equation.

Overall, Token-Ring has a lot to recommend it for image applications. At 16M bps, it exceeds Ethernet performance by 63% and overcomes the distance limitations of Ethernet, about 2.5 miles, through token regeneration at each station. Unlike Ethernet, Token-Ring provides every station with guaranteed access to the network instead of making each station wait for access until the network is idle, as in Ethernet. Token-Ring traffic can even be ranked according to priority, so that workstations can swap document images, for example, before other types of applications can use the network.

Fiber Distributed Data Interface

Legacy LANs may not have enough bandwidth capacity and growth potential to meet an organization's long-term imaging needs, especially with the addition of image applications. FDDI is a 100M bps fiber optic LAN that addresses the bandwidth distance, and fault-recovery issues that limit conventional LANs. FDDI provides the needed capacity and added throughput by a deterministic token-passing access method, which makes for a high level of network availability.

FDDI uses a timed token-passing access protocol for passing frame as large as 4,500 bytes. The standard permits support of up to 1000 connections over a fiber path of 200 km in length. Each station regenerates and repeats frames sent to it, which also serves as the means for identifying newly attached devices on the network. FDDI includes built-in management capabilities that detect failures and reconfigure the network automatically.

Although FDDI boards for connecting microcomputers and workstations to the backbone have been available for several years, they are still too expensive to allow most organizations to exploit FDDI fully as a super Token-Ring LAN. Currently, the most economical use of FDDI is as a backbone between hubs through which LANs may be interconnected in a campus environment.

100Base-T Fast Ethernet

100Base-T is a 100M bps version of 10Base-T, the dominant LAN technology in the industry. It provides 10 times the performance of 10Base-T for less than twice the price. Like 10Base-T, 100Base-T is being standardized by the IEEE 802.3 committee, the group responsible for all Ethernet standards. 100Base-T uses the same media access-control method, contention with collision detection, that is used on all Ethernet products.

In addition, Fast Ethernet allows the use of the same unshielded twisted pair wiring that is already installed for 10Base-T networks and is implemented using the same star topology as a 10Base-T network. This allows Fast Ethernet to leverage proven Ethernet technology, as well as the existing base of

Ethernet hardware and software products. Many hub vendors support both 100Base-T and 10Base-T, providing users with a smooth migration path to higher-speed networking in accordance with application needs.

One difference between Ethernet and Fast Ethernet is that the maximum network diameter of a 100M bps Fast Ethernet network is roughly 210 meters, whereas the maximum network diameter in a 10M bps Ethernet network can be up to 500meters. This may appear to be somewhat restrictive. However, distance is rarely a limiting factor in today's collapsed backbone networks implemented by intelligent wiring hubs, with interconnections accomplished by bridges and routers. This makes Fast Ethernet easy to implement for large-scale enterprise networks in support of heavy-duty imaging applications.

VG-AnyLAN

The other 100M bps standard is 100VG-AnyLAN. Like 100Base-T, 100VG-AnyLAN uses the Ethernet-frame format, which means that compatibility between 10M bps and 100M bps networks can be achieved with a speed-matching bridge between the networks. The key difference between the two 100M bps networks is that Fast Ethernet preserves Ethernet's CSMA/CD mechanism, whereas VG-AnyLAN dispenses with it entirely. In using a collisionless access method, VG-AnyLAN is able to consistently deliver 95% of the available bandwidth to users, assuming a large packet size. The collisionless protocol ensures fairness by not permitting bandwidth hogging and offers strong performance in very heavy loading situations, making it very suitable for document imaging.

100VG-AnyLAN supports a variety of media, including four-pair Category 3, 4, and 5 cable. It can support a mix of voice and data on each pair. VG-AnyLAN also supports two-pair shielded twisted-pair and optical fiber. VG-AnyLAN provides more robust signaling than 100Base-T, making it virtually insensitive to cable quality. These characteristics make VG-AnyLAN worthy of consideration for upgrading legacy LANs for document image transfers.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Asynchronous Transfer Mode

With document imaging and multimedia applications consuming an ever-larger portion of the available bandwidth on legacy LANs, the performance of all applications diminishes accordingly. ATM technology promises to address this and other problems by providing:

- *Increased bandwidth*. ATM goes considerably beyond the 10M bps and 16M bps offered by legacy Ethernet and Token-Ring LANs. To date, there are ATM standards for transmission at 25M bps, 52M bps, 100M bps, 155M bps and 622M bps.
- *Low delay*. Because ATM uses very short fixed-length packets that can be switched in hardware (instead of software), there is very little delay in transmission.
- *Improved network availability*. With ATM's increased bandwidth and low delay, the network is always available to run applications on demand. Control mechanisms at various levels in the network ensure that congestion does not become a problem.
- *Scalability*. ATM is easily scalable, meaning that the amount of bandwidth can be tailored to the needs of expanding application complexity and faster host processors without adversely impacting the performance of other applications and hosts on the network.
- *Overcoming limits on cable length*. ATM overcomes the distance limitations of Ethernet and Token-Ring. In fact, there is no practical limit on the distance of ATM transmission links.

For these reasons, ATM may emerge as the technology of choice for LAN backbones in the not-too-distant future. Both 100Base-T and 100VG-AnyLAN offer a migration path to ATM. In addition, ATM is also used on the WAN, so LANs can be seamlessly internetworked over great distances. The problem with ATM, however, is that a totally new infrastructure must be overlayed onto the existing network. This is still too expensive for most companies as well as the carriers.

Wireless Transmission

To convey document images between buildings, laser-optic infrared technology can be used to connect Ethernet or Token-Ring segments in each building, when laying fiber lines or leasing private lines is impractical due to cost or physical barriers. The wireless transceivers attach to a hub, switch, or router at one building and transmit the document images to a similar setup at a second building.

The wireless link derives its speed from the device to which the infrared transceivers are connected. An ATM switching hub, for example, could transmit data at 155M bps. The distance between the transceivers varies according to the data rate. When running at standard Token-Ring or Ethernet speeds, for example, the transceivers can be spaced about 3,300 feet apart. When attached to an ATM hub running at 155M bps, the transceivers can be only 900 feet apart.

Normally, laser-optic transmission is subject to signal interference from environmental factors. Vendors are employing new techniques, such as using a wider transmitting beam and a larger receiving lens, to provide more immunity to signal interference, which results in more reliable transmission. To gain this increased reliability, however, users of these products must be prepared to sacrifice transmission distance, which is usually limited to 1,000 feet.

STORAGE CONSIDERATIONS

After documents have been scanned and indexed, they must be moved to a storage facility, where they can be retrieved as needed by multiple users. Image files must be moved from the scanning station’s own hard disk to a server on the network, where higher-capacity magnetic or optical storage is available. In a small office, the “server” may be just another PC or workstation configured with a high-capacity hard disk. In a large IBM host-based environment, however, this arrangement may be inadequate, and DASD would be used instead. For longer-term storage, there are several optical disk types, including WORM drives, and equipment configurations to choose from, as well as tape drive and library systems.

Optimization Techniques

Although the source of performance bottlenecks can be any one of the many system components, in an imaging system, the I/O devices are usually the culprits. Of these, an optical disk jukebox’s robotics and disk spin-up and spin-down time can constitute a serious bottleneck. The situation worsens when a few large documents, stored across multiple optical platters, are requested at the same time.

There are several ways to alleviate this type of problem. The first entails prefetching the images through workflow software. This allows the system to batch-retrieve images overnight, when the system is used the least. The images are moved from the server's optical media, where images are permanently stored, to local magnetic media, which provides a faster access time for workflow operations. Cache is implemented based on an understanding of which images are likely to be required next. This technique is effective in workflow applications, where there are queues of images to be worked on. The vendors' storage management utilities are used to implement image caching.

A second solution is to employ an optical-storage manager that writes all the images that belong to one folder to a single optical disk. This prevents the folder from being fragmented over multiple optical disks, in case it grows.

Another method is to implement a hierarchical data-storage scheme. This involves understanding the use and life cycle of a document. When a document is in the active part of its life cycle, for example, it is stored on magnetic storage, which has fast retrieval capabilities. When the document becomes inactive, it should be automatically moved to optical media in a jukebox. When the document is archived, it should be moved to an archive medium, such as helical scan tape. A document should be able to move back and forth through this hierarchy.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

The Role of Data Compression

Image documents are much larger than ordinary text or binary files, so most vendors' imaging systems support one or more compression schemes, including their own proprietary method. Compression allows images to be stored on the PC's hard drive until the increasing volume justifies moving to more expensive optical WORM units or tape storage systems. Compression ratios of 25:1 are common for most types of document images. For graphics and photos, the compression ratios are as high as 50:1 without a noticeable loss of image quality. The latest algorithms promise compression ratios of up to 300:1.

Compression is typically a function of software, which is usually bundled with the imaging software. Compression may also involve the addition of a compression board to offload the extra processing from the workstation's main CPU and provide a substantial performance boost. Document images and text are stored in compressed form and transparently restored when accessed. A service is said to be "transparent" when it is provided without an explicit request from the user, operating in the background and having no noticeable effect on the requesting program or any other application, including the operating system.

Compression typically works as follows:

- The data is examined to identify patterns.
- To each identified pattern, a unique code is assigned, which consists of significantly fewer bytes.
- The unique codes are substituted for the corresponding patterns.

- The data is then stored in its new format.
- Upon retrieval, the encoded data is replaced with the original strings of data, completely restoring the image.

A process called forms removal can improve compression ratios, no matter which algorithm is used. When scanning forms, the data, rather than the form itself, is important. In eliminating the repetitive forms and saving only the data, compression ratios can be vastly improved.

A dental claim form, for example, scanned at 300 dots per inch (dpi) and compressed according to the international Group 4 standard, typically uses 73K bytes of disk storage. When the form is removed, the same image file occupies only 13K bytes in compressed form. When the form is called up for display on a workstation monitor, the data is overlayed on an image of an empty form. The file containing the empty form can be stored locally, while the data can be retrieved from a data base server.

WAN CONSIDERATIONS

To transfer image files to remote corporate locations, carrier-provided services or private backbone facilities can be used to interconnect LANs. Because of the huge size of image files and the need to support myriad other applications, however, not just any carrier service or backbone facility will do, even if the image files are compressed before transmission.

Dialup and Analog Leased Lines

For occasional image transfers to remote locations, modems operating at up to 28.8K bps over dialup or leased analog transmission lines may suffice. However, if image transfers are frequent or continuous, the use of analog lines, dialup or leased, is fraught with problems. In addition to creating intolerable transmission bottlenecks, analog lines are commonly affected by voice frequency impairments that can corrupt data. This means modem transmissions can fall back to 14.4K bps or lower until line quality improves, or the connection can drop altogether, forcing a new connection to be established and the imaged documents to be retransmitted.

To minimize the effects of line impairments on analog leased lines, extra-cost line conditioning may be requested from the carrier. No performance guarantees come with line conditioning, however; the carrier promises higher quality lines on a best-effort basis. The low-speed and uncertain line quality of dialup and analog leased lines renders them unreliable for carrying any type of LAN traffic, let alone image traffic.

Packet Switched Services

Packet switched networks thrived in the 1970s and 1980s as a way for asynchronous terminals to access remote computers over noisy dialup and analog leased lines. The driving force behind the acceptance of packet switching networks was the adoption of the international X.25 standard. However, carrier-provided packet data services based on the X.25 standard have been limited to 56K bps, mostly because of the X.25 protocol's overhead burden that, among other things, provides error-checking and correction at every network node to ensure error-free transmission.

If a receiving node detects a transmission error, it requests a retransmission from the sending node. While this is valuable for point-of-sale applications requiring the accurate transmission of credit card and other financial information, it is a cumbersome and unnecessary process when image traffic is concerned, especially when large image files are being transferred. Consideration must be given to more reliable digital services, therefore, starting with Digital Data Services (DDS).

Digital Data Services

DDS offers a range of speeds, from 2.4K bps to 56K bps. DDS does not require a modem, but a digital termination device called a digital service unit. The appeal of DDS for data is the higher quality of digital versus analog transmission. At one time, DDS at 56K bps was considered a good way to connect LANs via remote bridges. Today, the 56K bps line rate is viewed as a bottleneck to LANs that operate from 10M bps to 16M bps.

Fractional T1

FT1 entails the provision and use of bandwidth in 56/64K bps increments without paying for an entire T1 facility. FT1 allows users to order only the amount of bandwidth needed to support the image application. This saves the user the expense of leasing a full T1 line for partial use, while easing the WAN bottleneck between LANs, which is a problem with DDS and X.25 at 56K bps. FT1 is widely available among local and interexchange carriers, so there is less of a back-haul problem to contend with, as in DDS, and because it is derived from a T1 facility, FT1 offers high reliability and availability.

To avoid the cost of leased lines when the amount of image traffic is relatively low, an inverse multiplexer or bandwidth controller can be used to dialup as many 56/64K bps channels as needed. The inverse multiplexer aggregates multiple channels to achieve a single higher-speed channel. Upon completion of the transmission, the channels are taken down. Instead of incurring a fixed monthly charge for an underutilized leased line, the user is billed for the dialup channels on a usage basis, which can be more economical.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Integrated Services Digital Network

Primary rate ISDN (23B+D) is another wide-area internetworking option that may be used to support image applications. ISDN is a carrier-provided switched digital service that is billed for on a time and distance basis, just like an ordinary phone call. ISDN channels are available in bandwidth increments of 56/64K bps, 384K bps, and 1.536M bps.

The billing method is both an advantage and a liability, depending on how many hours a month the ISDN channels are used to support the image application. When used continuously, ISDN channels in the US are considerably more expensive than using the bandwidth equivalents offered over dedicated leased lines (FT1/T1), which entail fixed monthly charges determined by distance, no matter how long the line is used. However, if the ISDN channels are used only 20 hours a month for image transfers, ISDN can be more economical than paying for full-time leased lines.

Dedicated T1 Lines

T1 digital lines are an ideal medium for interconnecting LANs from point to point, especially LANs carrying high-volume image traffic. They offer excellent reliability and availability, in addition to high capacity. An increasing number of bridges and routers offer T1 interfaces to facilitate LAN-to-WAN connectivity.

With imaging applications already putting a strain on today's Ethernet and Token-Ring LANs, however, running image applications over T1 lines may not be an adequate long-term strategy. Complementary LAN interconnection strategies may be called for, such as frame relay, which make more efficient use of the available bandwidth.

Frame Relay

To prevent the network from becoming the bottleneck, it is advisable to avoid protocols that feature a high degree of error-checking overhead and acknowledgments and instead seek protocols that feature no acknowledgments at all, or offer multiple-packet receipt acknowledgment. Of the former, frame relay is a good example.

The technical concept behind frame relay is simple: To eliminate protocol overhead and unnecessary processing to speed up network throughput. Error correction and flow control already exist at the upper layers of most modern computer communication protocol stacks, and thus may be relegated to the "edges" of the network rather than performed at every node within the network, as in X.25. Frame relay eliminates 75% of the protocol overhead of X.25, including error correction.

Because frame relay operates over high-quality digital facilities, there is no need for error correction. This function can be relegated to customer premises equipment. Frame relay also offers greater efficiency, since an application can momentarily seize the entire amount of allocated bandwidth to transmit information in bursts. Upon completion of a duty cycle, it relinquishes the bandwidth to other applications.

T3 Services

T3 service is typically offered over fiber facilities. The applications touted by T3 advocates include LAN interconnection, multiple T1 line replacement, and high-speed backbones that integrate voice, data, video, and image traffic.

T3 service entails special construction of access lines from the customer premises to the carrier's serving office. Special construction costs at each end differ widely from region to region, from a low of approximately $8,000to a high of approximately $150,000. These costs are almost never factored into the crossover comparisons with T1, so the true cost of T3 makes it difficult for even the largest companies to justify.

In the absence of optical standards for T3, proprietary interfaces have proliferated that, in turn, restrict the ability of users to mix and match different manufacturers' equipment end to end. T3 services, such as AT&T's T45, require the customer to negotiate the type of optical interfaces to be placed in the various serving offices of the interexchange carriers.

Some carriers are capitalizing on the appeal of Fractional T1 by extending the "fractional" concept to T3. Under this concept, the user can order bandwidth in T1 increments up to the full T3 rate of 44.736M bps. This service is designed for users who need more than the 1.544M bps offered by T1 but less than the full bandwidth offered by T3, to support the interconnection of Token-Ring or Ethernet LANs carrying image traffic. A bridge is used to connect each LAN to the public network. From the user's point of view, the public network appears as an extension of the LAN.

Current FT3 offerings are not intended as a migration path to more advanced broadband services, such as Broadband ISDN, which will be based on Synchronous Optical Network (SONET) and Asynchronous Transfer Mode (ATM) technologies.

Switched Multimegabit Data Service (SMDS)

SMDS ultimately may offer a better solution than frame relay for linking LANs in a metropolitan area so that corporate locations can share image files. SMDS is a high-speed data service, offering customers the economic benefits of shared transmission facilities, combined with the equivalent privacy and control of dedicated networks.

Access to SMDS is provided via dedicated lines. On each access line, the required customer premises equipment consists of a router with an SMDS interface and a CSU/DSU (channel service unit/data service unit) with an SMDS interface. A DXI provides standardized connectivity between the two devices, while a LMI permits the CSU/DSU to pass performance information to the router so it can be sent to a SNMP-based management workstation.

Carrier-Based ATM Services

Only a handful of carriers offer ATM services. The carriers that do, provide only PVC. These are ATM connections that are set up between a sending and receiving station on the network. They are set up by a network administrator and remain up until torn down manually. Another type of ATM connection, which will be offered in the future, is called the switched virtual circuit. This type of connection is set up and torn down by the ATM network on demand without manual intervention.

PVC connections are more reliable for certain types of applications, such as document imaging, whereas SVC connections are intended for routine types of applications. SVCs can time out and tear down after a brief period without traffic between resources. When a replacement connection is requested, the network may not be able to provide the same guaranteed bandwidth it had previously allocated to that application. PVCs stay in place, even if there is no traffic for a prolonged period of time.

Previous | Table of Contents | Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Bookmark It

Search this book:

Go!

Previous | Table of Contents | Next

SUPPORT ISSUES

A variety of support issues deserve attention with a LAN-based document imaging system. Among them are systems reliability, systems integration, and training.

System Reliability

When documents critical to a business's operation are committed to imaging technology, the systems and networks involved must be reliable and stable. Specifically, these systems and the networks they run on must be protected against failure. The possible solutions include but are not limited to:

- uninterruptible power supply
- redundant components and subsystems
- alternate WAN routes between corporate locations and bypass circuitry between LAN hubs and major subsystems
- arrangements for local service and support from vendors to minimize system downtime

A determination must be made as to what protective measures are already available and to what extent they can be applied to the imaging system. If new protective measures are needed, such considerations must be entered into the planning process and factored into the purchase price of the imaging system.

Systems Integration

When an imaging system is being designed from scratch and is composed of products from different vendors, the task becomes one of tying together these diverse elements to provide users with transparent access to every other element on the network. Only rarely do in-house staff have the expertise to accomplish this alone. Systems integrators can help.

A systems integrator's qualifications can include the following:

- specialized industry knowledge, awareness of technology trends, and experiences gained from a broad customer base
- knowledge of specific protocols, interfaces, and cabling requirements
- a reputation for fast, high-quality work
- the ability to train existing systems staff to take over responsibility for ongoing management, administration, and control
- the need for an outside party to act as a catalyst in implementing change and to validate (or fine-tune) in-house plans

Integration firms provide several services, some of which are listed here. No single firm can provide all of these services, which is why they frequently form working partnerships to propose a total solution. In-house data center professionals may also have experience in one or more of these areas:

- *Design and development*. This area includes such activities as network design, facilities engineering, equipment installation and customization; hardware, software, network integration; acceptance testing, and network management.
- *Consulting*. This area includes business planning, systems and network architecture, technology assessment, feasibility studies, request-for-proposal development, vendor evaluation and product selection, quality assurance, security auditing, disaster recovery planning, and project management.
- *Systems implementation*. This area includes procurement, documentation, configuration management, contract management, and program management.
- *Facilities management*. This area includes operations, technical support, hot-line services, change management, and trouble ticket administration.
- *Systems reengineering*. This area includes maintenance, systems and network optimization, remote monitoring and diagnostics, and the application of automated design tools.
- *Training*. This area includes hands-on user training, technical and management courses, executive briefings, and seminars that address industry trends.

Training

A LAN-based imaging system can be quite complex, especially when configured for workflow applications; therefore, extensive operator and supervisor training may be required. A reputable imaging system vendor will offer a full line of instruction about its products and technology and offer dedicated training staff and facilities.

Many times, formal classroom training at the vendor's facilities is not sufficient, especially if the imaging system requires a high degree of customization. In this case, the vendor or systems integrator should be willing and capable of offering on-site training at the customer's location. During the planning phases, extra costs, if any, of additional training for new employees hired after the original training period, as well as the cost of follow-up training to fine-tune user skills, should be identified.

It is always a good idea to ask about the experience and qualifications of the trainers. The vendor should not simply send technicians to provide training; generally, technicians lack a user's point of view and do not always make effective instructors, unless they have been specifically trained for that responsibility.

The nature and scope of training can usually be ascertained by reviewing copies of the training materials before committing to a specific imaging system vendor. The materials should provide clear and comprehensive learning objectives supported by well-organized lesson structures and descriptions which can be used as reference material after the training sessions. If the vendor does not provide this kind of depth in its training package, the customer may be getting less out of the capital investment than anticipated at the time of purchase.

CONCLUSION

Despite the large investments companies have made in office technologies over the years, productivity gains have been hampered by the huge quantity of paper these technologies have tended to produce. Now that LAN-based imaging has arrived to address the problems of paper overload and workflow, planning and implementing such systems can provide significant opportunities for reengineering work processes and streamlining management structures within organizations.

With proper planning, the benefits of implementing LAN-based imaging systems include better document control, faster retrieval of vital information, and multiuser access. These benefits can improve the quality and timeliness of customer service, a strategic goal in many businesses, and are likely to have a substantial impact in the data center. Data center managers can benefit themselves and their companies by applying their experience and expertise to helping departmental managers in evaluating, planning for, and implementing this relatively new technology.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter III-7
Implementing Massively Parallel Processing

Prem N. Mehra

Massively parallel processing, once exclusively limited to scientific and technical applications, offers organizations the ability to exploit vast amounts of data for both strategic and operational purposes. Because of its newness in the parallel processing arena, the deployment of massively parallel processing raises several important implementation considerations. This chapter discusses only the considerations and issues that are new or have become important with this technology. Exhibit 1 is a high-level framework used to organize the discussion.

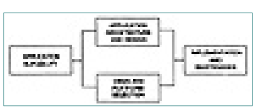


Exhibit 1 Implementation considerations in deploying massively parallel processing.

DETERMINING APPLICATION SUITABILITY

Assuming that a new application has been identified and justified, one of the first considerations is to assess if there is a match with the massively parallel processing (MPP) technology. Most vendors of MPP hardware and databases have positioned this technology to information delivery facility (i.e., data warehousing) applications only. Database-management systems (DBMSs) and massively parallel processors enable intraSQL parallelism, which benefits complex and long-running queries associated with online analytical processing (OLAP). A simple and quick OLAP SQL is not likely to benefit from intraSQL parallelism.

Because the vendors' focus has been on information delivery facility (IDF) applications, they have not concentrated on online transaction processing (OLTP) applications, which are generally destined for symmetric multiprocessors (SMPs). It is possible, however, that things will change over the next few years, as new technology components become available.

For example, one indication of this emerging change is the use of IBM RS/6000 SP MPP nodes as SAP R/3 application servers. The motivation for use is to minimize system administration costs and not to exploit parallelism. Similarly, print and file-server consolidation offers another usage scenario. Another emerging use for RS/6000 SP nodes is to host Oracle parallel server instances to provide added capacity to handle the database requests issued by the SAP R/3 application servers. Again, the motivation is to get more capacity from additional nodes, as opposed to exploitation of intraSQL parallelism.

However, with these exceptions, MPP use is currently limited to OLAP IDF processing.

SELECTING A DBMS AND PLATFORM

The choice of a DBMS and hardware platform for a new application should be made based on many considerations, not just on parallelism. In fact, parallelism is merely one aspect of technology, whereas technology itself is just one of many considerations. This section examines the bigger picture of selection considerations and then focuses on parallel processing technology as if the decision is to be based solely on such features.

DBMS Selection

With all the discussion about parallelism, it might be tempting to select a DBMS based on its software architecture and technical functions and features. However, the DBMS is only a piece and must fit within the whole enterprise information architecture. In addition, vendors constantly try to outdo each other in offering exclusive features, which then also become available sooner or later on other DBMSs.

The selection process should be based on an organization's guiding principles for information technology, which reflect the corporate values and culture. They include such factors as:

- the role technology plays in the corporate competitive positioning
- the desire and ability to deploy leading-edge, but unproved, technology the purchase of technology from a market leader vs. one from a supplier who might not be a leader but can provide innovation

Thoughtful analysis in decision making is beneficial, but it should not lead to the analysis/paralysis syndrome, which may negate a competitive marketing opportunity. The advantage of basing the decision process on the technology guiding principles is that appropriate, but not unnecessary, time and energy are spent in analysis based on the planned use of the technology.

Keeping this background in mind, the analysis for selection should be driven by the added value provided by the DBMS vendor. Factors contributing in the value proposition are:

- vendor technical and marketing support
- the working relationship
- the availability of application packages using the DBMS
- the availability of skills and education

In other words, the current and potential benefits of a long-term relationship can overcome several short-term technical considerations.

The technical criteria for DBMS selection change every three to four years. In the mid-1980s, considerations such as compliance to the relational model, referential integrity, support of views, and online utilities were paramount. In the early 1990s, connectivity to other products, end-user query tools, triggers, stored procedures, and row-level locking took the center stage. In the 1990s, the emphasis has shifted to replication, user-defined data types and functions, and exploitation of parallel processing. One can assume that by the time the 21st century arrives, new criteria will have emerged.

Exhibit 2 depicts a framework that graphically summarizes the previous discussion.

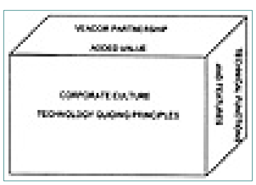


Exhibit 2 A framework for DBMS selection.

| Previous | Table of Contents | Next |

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Looking strictly from the perspective of massively parallel processing, the technical evaluation considerations should include the following:

- *How will the DBMS interact with legacy and other transaction-oriented systems with respect to data and metadata?* It is very likely that the use of massively parallel processing in an installation will be first introduced for OLAP in an information-delivery facility (i.e., data warehouse) environment that involves large quantities of data that is likely to be sourced from these systems.
- *What tools are available to assist in designing and implementing databases and data partitioning?* Partitioning tools are of great value in implementing partitioned data DBMSs. Less obvious is the significance of partitioning for shared-data architectures. Data-partitioning data offers ease in management from a utilities perspective and is required for either architecture when large databases are involved.
- *What SQL functions are parallelized?* Does the DBMS support all appropriate SQL statements or only a subset? Is the optimizer sophisticated enough to choose an appropriate level of parallelism on its own, or does it require user assistance? An SQL programmer should not be required to bear the burden of coaxing the optimizer to use parallelism.
- *Does the DBMS support the installed or planned hardware platform?* The considerations for choosing a hardware configuration are discussed in more detail in later sections.

Finally, the parallelization of database administration utilities, perhaps the most important consideration, must be deliberated. Will the DBMS perform the following in parallel:

- New table loading, appending data to existing tables, and purging tables of old rows?
- Redistribution of data if the system load is unbalanced? Can it be accomplished granularly, or must all the redistribution be done en masse?
- Can the backup and recovery utilities be executed in parallel against partitions of data? This is important regardless of the software architecture used.
- Are performance and monitoring tools available in the parallel environment? Can the tools help in isolating poor performing queries in a multiuser environment? What tools are available to project capacity requirements?

All these facilities not only must perform with small amounts of data and a limited number of users, but must also scale up when data size increases. In addition, they must speed up if more resources are added and the work load is maintained constant. These tools should be able to deliver their functionality without adversely affecting user data availability.

This is clearly a tall order for any DBMS to satisfy, but progress is being made. Commercial massively parallel processing is possible today, as seen in implementations by such companies as MCI and Sears.

As DBMS evaluation and selection is being made, the next consideration is which hardware configuration to deploy.

SELECTING A HARDWARE CONFIGURATION

As discussed earlier, the choice of hardware configuration is also quite complex. Of the three options discussed — shared everything, shared nothing, and shared disk — there are no clear-cut winners or losers. They all have applicability, depending on the type of work load.

Information-Delivery Facility

All these configurations have a place in the computation of the OLAP workload. Attempts to define their positioning have led to identification of database size and CPU resource consumption by the SQL statement as two significant parameters. Exhibit 3 shows how these can be used to divide the solution domain into four quadrants, using the parameters of required CPU resource and data size.

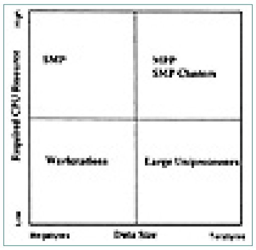


Exhibit 3 The information delivery facility (data warehouse) solution domain.

Required CPU Resources. Assuming that the disk and other costs are the same, the required CPU resources parameter (assigned to the Y-axis from low to high) represents the major differentiating execution time cost variable. One might use SQL complexity as a proxy for this parameter, and that is valid if care is exercised. At times, however, complex and long SQL statements with lots of predicates are inexpensive to run, because they narrow the data-selection criteria. On the other hand, very simple-looking SQL statements can be extremely expensive, because they require scanning and joining large tables without any filtering. On the Y-axis, any parameter that represents the differentiating run-time CPU cost is acceptable.

Data Size. Data size, on the X-axis from megabytes to terabytes, represents the total size of the database. However, this is really a metaphor for the ability of the hardware platform (along with a suitable DBMS with its utilities for loading, backup, recovery, reorganization, and management of data) to support user databases.

Computational solutions that reside in the lower left quadrant can be addressed by increasingly powerful desktop or deskside workstations available today.

When the data size grows, the systems management sophistication and I/O bandwidth requirements lead to the solution domain of large uniprocessors in the lower right quadrant. Water-cooled technology-based (known as BIPOLAR technology-based) SMP mainframes are also possible options here. Sophisticated systems-management disciplines, learned over decades to manage hundreds of gigabytes of data, position them very well here. This is evidenced by installations that use DB2 for MVS successfully for OLAP in this domain.

The upper left quadrant is the domain of comparatively small databases, but with those OLAP workloads that put large demand on the CPU resources. SMPs, including the new air-cooled (CMOS technology-based) mainframes, are solutions in this domain because the CPU resource is available less expensively on these platforms than on alternative configurations.

When the database is large and the required CPU resources are high, MPP offers the solutions in the upper right quadrant. Such configurations, along with the partitioned data and shared data software, have already proven their capabilities to handle problems in this domain. SMP clusters also have the potential to be players in this arena. However, there are no widely known successful SMP clusters installations that can be used for illustration.

The size of database that divides the X-axis is an interesting consideration. A wide variation in valuation is possible. Also, as time passes and the programmer gains more experience, the value seems to move to the right. A few years ago, 25G bytes to 50G bytes would have been considered the top end for SMP configurations. Today, 100G bytes to 150G bytes seems to be the consensus, and the value is likely to go higher in the next few years.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

OLTP and Batch Environments

For this workload, the use of SMP configurations is well established. Also, SMP clusters have been successfully employed. Introduction of MVS Parallel Sysplex by IBM (IBMSY) is an example of the increasing popularity of this configuration for OLTP and batch work load. What is new here is that MPP configurations are also starting to be deployed as application and database servers for OLTP and batch. SAP R/3 implementation using Oracle 7 is an example. Interestingly, there is no intraSQL parallelism exploitation here. After all, OLTP work load, by design, issues SQL statements that perform direct access and update of few rows, which do not gain significant benefit from the intraSQL parallelism.

The motivation for deployment in such cases is the ease in systems management. The packaging of MPP simplifies the problems associated with the maintenance and distribution of software to the equivalent number of distinct nodes that would be required for the same workload. A few operational considerations also favor MPP. This chapter does not discuss these considerations, because the focus has been parallelism. Whether this deployment is a flash in the pan or will become more common depends on several factors, including pricing schemes, marketing considerations, and enhancements made in the systems management arena by the alternative solutions of physically distributed nodes.

APPLICATION ARCHITECTURE AND DESIGN

Currently, the primary use of massively parallel processing is for the information delivery facility, and the architecture and design consideration for those kinds of applications should be followed when implementing massively parallel processing. From the application–architecture point of view, care should be taken to ensure that the application does not prevent the exploitation of the parallelism features. From an SQL coding perspective, the use of parallelism is transparent to the application programmer.

Database design requires careful consideration. In addition to the traditional design issue (e.g., data normalization and denormalization, indexes, and design issues), partitioning and load-balancing considerations need special attention. To some extent, these considerations depend on the architecture and features of the selected DBMS and the hardware platform. For this reason, a thorough understanding of the underlying DBMS architecture and the vendor's implementation of the architecture is essential.

The DBMS architecture section mentioned the use of data parallelism to exploit parallelism. Data partitioning enables data parallelism. How data is partitioned depends on its planned use and the available partitioning features of the selected DBMS. Some popular techniques are:

- *Key range*. Data is partitioned based on key value ranges. With this scheme, queries that scan a limited range of data, such as those with BETWEEN, LESS THAN, or GREATER THAN predicates, can be isolated to certain partitions.
- *Expression-based*. This is similar to, but more sophisticated than, the key range partitioning scheme. SQL-like logic is used to partition data and to direct some user queries to specific partitions. It is more flexible than the key range partitioning scheme and provides fine-grained control but also incurs a slightly higher cost.
- *Hash-based*. Data is partitioned based on a hashing scheme and permits more-uniform distribution of data. It is the most widely used data-distribution scheme and permits the SQL query work load to be widely distributed among the nodes. If the values in the distribution key are skewed, hashing does not result in uniform data distribution. It can be the best-performing partitioning scheme for row-level retrieval, because the request can be directed to a single partition.
- *Round robin*. Data is distributed equally among all nodes without any regard to primary or other column values. This results in uniform distribution of data and processing; however, the retrieval can be expensive for key value access, because all the partitions are accessed.
- *Schema partitioning*. This implies no partitioning at all and generally is used for small tables. The whole table belongs to a single DBMS instance and is located on one node. When needed by other DBMSs at other nodes, the entire table is transmitted.

Not all DBMSs support all of these techniques.

Such products as DB2 Parallel Edition, Informix XPS, and Sybase MPP require that data be partitioned for exploiting parallelism. There is one major exception: Oracle7 does not require data partitioning. During execution, data blocks are directed to the different Oracle7 instances to implement the notion of data parallelism. On the one hand, this results in increased MPP interconnect traffic, but on the other hand, this helps Oracle7 in balancing load among different nodes of an MPP.

Load balancing can be viewed from two points of views: physical and logical. The physical involves balancing loads among physical resources, primarily multiple processor nodes. The logical involves with load balancing different tasks accessing a logical resource such as a table.

HOME ACCOUNT INFO SUBSCRIBE LOGIN SEARCH MY ITKNOWLEDGE FAQ SITEMAP CONTACT US

SEARCH ITKNOWLEDGE

Go!

Keyword

Brief Full

Advanced Search

Search Tips

BROWSE BY TOPIC

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Bookmark It

Search this book:

Go!

Previous | Table of Contents | Next

Physical Load Balancing

A partitioned data software architecture executing on a shared-nothing configuration owns its fragment of the data and executes operations associated with the SQL plan assigned to it by the coordinator task, as shown in Exhibit 4.

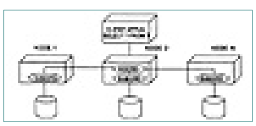


Exhibit 4 A partitioned data architecture executing on a shared-nothing configuration.

A DBMS instance executing on a given node controls its own I/O, performs locking, applies the local predicates, extracts the rows of interest, and transfers them to the next stage of processing, which may reside on the same or some other node. A node can become overloaded if there is too much demand for data managed by the DBMS executing on the particular node. Redistribution of data is the remedy to reducing the load.

Data can be redistributed by selecting another distribution key or by using a different partitioning scheme. This assumes that the use pattern is fairly consistent and does not change frequently. Much more important, it requires a priori knowledge of usage of data, which can be difficult to anticipate in an OLAP environment.

A data distribution approach that can balance the node load is known as a round robin. As discussed, this scheme also avoids the problems associated with skewed data. There are, however, many negatives associated with the round-robin scheme. Any query that otherwise might benefit from data clustering, such as BETWEEN or GREATER THAN, are forced to execute on multiple nodes. A query that might return only two to three rows requires perhaps 100 tasks in a 100-node example.

The shared-data software implementation does not suffer from this node overload problem. Here data is not partitioned, but shared among all the nodes. At run time, all rows are visible to all nodes. This point is frequently emphasized by implementers of databases based on this architecture, such as Oracle. (From a technology perspective, an installation database and system administrator must consider these issues in making a DBMS selection.)

Assuming that node-level balancing has been obtained in a partitioned data implementation, a second-level partitioning scheme within a node can lead to minimization of I/O. For example, Informix XPS offers a hybrid partitioning scheme, in which hash partitioning leads to distribution of data to all the nodes, followed by a second level of range partitioning within the node to limit I/O to a few devices.

Logical Load Balancing

This is concerned with attaining a balanced load for a logical resource such as a table. Contrast this with physical load balancing, in which the emphasis is on load balancing of physical resources, such as nodes and devices. Logical load balancing requires execution-time algorithms that can spawn tasks based on the work characteristics. An example is shown in Exhibit 5.

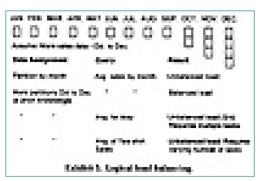


Exhibit 5 Logical load balancing.

Assume a shared-nothing configuration, partitioned-data DBMS, and a sales table that is partitioned 12 ways, one partition for each month, without realizing that the last three months partitions have more data than others. A query for computing average sales per month results in unbalanced load for the last three months. This can be remedied by using several redistribution techniques. As discussed earlier, the new partitioning scheme of data to achieve a balanced load requires a priori knowledge of data usage, which can be a challenge.

Next, assume a need to compute average sales for items sold in the month of May. This again provides a nonuniform load situation, with one partition and the corresponding node being very busy. To achieve parallelism, the DBMS must have the sophistication, during execution, to logically partition the data for the month of May into four parts, one for each week. The DBMS must also spawn four tasks, or threads, to process the parts using shared data and buffer software architecture within the node.

Assuming this can be achieved, the next query computes average of tee-shirt sales. This presents an additional level of challenge in that the number of tasks needed for the summer month partitions, when tee-shirts are sold, will need different numbers of tasks.

Current commercial DBMS offerings are attempting to address the challenges encountered in exploiting massively parallel processing. The major vendors have selected either the partitioned data or shared data as the underlying architecture of their choice. But, their implementations are hybrid in nature and use features that make classification difficult. With each new release, one vendor leapfrogs over its competitors in several functions. Under these conditions, how one proceeds to choose a DBMS and appropriate hardware configuration is quite complex.

CONCLUSION: IMPLEMENTATION AND MAINTENANCE

The best practices developed over the years for system implementation and maintenance continue to apply. Some of the considerations that warrant particular attention include procedures for redistribution of data to account for skew and system management. The availability of tools to facilitate these operations should be a major consideration in the DBMS and platform selection. Training and support should be made available to permit their use.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter III-8
How to Handle Data Loss and Backup

William Krouslis

Data loss occurs for many reasons. In the order of the frequency of their occurrence, the 12 most common causes of data loss are:

- The unintentional overwriting of a data file.
- The failure to back up microcomputer or workstation records on a diskette or magnetic tape because of time pressures or other tasks that appear to take precedence.
- The crash of a hard disk on a workstation or a network file server, causing not only the loss of information but also the damage to or loss of both the operating systems and the bindery of affected network management systems.
- The malfunction of diskettes or magnetic tapes that prevents them from recording data properly. This is a significant problem with tapes because they can write errors during a backup or read errors when trying to read data back. These problems are exacerbated when the read heads are not cleaned regularly.
- The inability to locate the diskettes or tapes on which desired data is stored. This usually occurs when material is stored in inappropriately or unmarked boxes or, equally likely, in boxes that were indexed in a fashion that nobody in the organization now understands.
- The loss of data because of an unexpected electrical power supply

failure.

- Damage to diskettes or tapes in transit or at a secondary storage site.
- A computer virus that damages or destroys data files and operating systems.
- The theft or misuse of data, including its removal or use by unauthorized persons.
- The degradation over time of the magnetic signal — perhaps by just one byte or even a bit — on a diskette or tape. Such changes can make the content of a disk or tape unreadable without the performance of some sophisticated, costly, and time-consuming retrieval activities.
- The changing of one or more versions of an operating system or application previously used to create and record data so this data cannot be read without reloading the original software (if it can be found at all) or running routines of varying complexities to convert the data into compatible formats.
- The destruction of a building or a work area within it by fire, earthquakes, temperature extremes, flood, or other such catastrophes that render diskettes or tapes unreadable or unavailable.

Once lost, data cannot be replaced. Accepting this fact, the prudent user routinely backs up data to a medium and location that together guarantee that no single disaster can destroy both the primary electronic source record and the backup copy simultaneously. Choosing a procedure that protects an organization from anything from accidental data erasure to data destruction in a building fire is the first step in creating a workable backup routine.

SIX BACKUP OPTIONS

There are six principal backup options. They are discussed in the following sections.

Do Nothing

Many now follow this course. They risk the complete loss of their data files.

Use Magnetic Tapes and Diskettes and Carry Them Off-site

This method is the cheapest and the most commonly used. However, it involves many potential problems in cataloging, maintaining, and recovering data.

The traditional method of assigning data owners the responsibility of regularly backing up data and storing it on a diskette does not work for two main reasons. First, users take diskettes containing their backup copies home at the end of the work day, compromising the security of these backups. Second, users stop backing up data regularly when their files grow to require multiple diskettes. When the backup process annoys the data owner, the manual backup procedure is dead.

Centralizing all backup activities and having them performed by specific individuals does not work either. The backup records become sloppy, and labeling them becomes an administrative nightmare. The people responsible for making backups leave the organization, or they have new responsibilities that interfere with the backup process. As a result, the backups are made less frequently, and eventually the backup process stops altogether.

Tape backups can be made automatically. But tape drives can be expensive, and the time necessary to write to tape can be long, especially if the backup is done in the verify mode. Tape drive heads also become dirty, diminishing the quality of backup information.

Using a typical tape rotation schedule results in a reasonable backup, as long as the tape containing the backup is moved to a secure off-site location. Unfortunately, most backup tapes never make their final journey. Left unprotected in the same area where servers are, they are subject to the same event that might cause the primary data source to be lost. A vigorous off-site storage transfer and management procedure is therefore necessary.

Employ a System That Can Be Connected to Multiple Disks

These disks must be adapted to provide limited random access. Organizations using this arrangement must have sophisticated on-site management, and even with it, they face many of the problems common to tape systems. This arrangement also does not aid in disaster recovery.

When data is written to disks simultaneously over the same channel, a fault-tolerant mechanism protects the information. Although this approach is better than nothing, a failure encompassing more than just the storage device would render both sources of data unavailable. Adding channel extension and disk duplexing to this arrangement provides added redundancy. However, if both disks are in the same server, a single event could still compromise both sources of data.

Another option may be the use of a RAID (redundant array of inexpensive disks). Spreading data across multiple drives obviously reduces the risk associated with the failure of a single drive. However, the use of RAID does not eliminate the need to backup.

| Previous | Table of Contents | Next |

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Use an Off-Site, Electronic Vaulting Service

Using these services eliminates the maintenance and recovery problems associated with backing up data files. Although this type of service has been available for many years for mainframes, it is just becoming available for local area networks and microcomputers. For example, Surefind, using technology developed by AT&T Commvault Systems, can back up UNIX, DEC-VMS, Novell Netware, and MS-DOS and can transmit data over ordinary telephone lines to a secure location containing optical storage devices.

These services usually incorporate optical media with either writeonce read many (WORM) or erasable optical disks. Because the optical disk is a random-access unit (as opposed to sequential tape), lost files are available directly, and because the storage mechanism is located at a distance from the primary place where data is stored, this type of service offers real data processing disaster recovery capabilities. Outsourcing the backup mechanism to a third-party source may therefore be the most reasonable solution for the small and medium user.

Purchase Mass Storage Equipment and Install it In-House

This arrangement is useful in organizations with several local area networks that generate sufficient amounts of data. Off-site disaster recovery is provided by placing the storage devices in a location different than the one that houses network hubs and servers. This solution is reasonable for large users with large amounts of data. The basic technology employed in Surefind was developed initially in this type of setting.

Funnel Data Through an In-House Network for Storage by a Mainframe and Peripherals

This approach provides reliable data storage and retrieval and, often, prompt data access. However, unless an organization has excess storage capability in the data center housing the mainframe, using this mainframe and its peripherals to store and retrieve microcomputer data is relatively costly. Organizations using this approach must also put their network administrators under the control of the their information services departments with respect to matters of scheduling and systems development.

BACKUP SYSTEM EVALUATION

When evaluating a backup system, the Data Center manager should ask these questions:

- Does the system daily and automatically back up all necessary data that the organization produces?
- Does it maintain duplicate, verified, and separately stored data and applications that cannot be erased or destroyed by such catastrophes as fire, theft, or computer viruses? Can the content of data files be changed or manipulated once recorded?
- Does the backup system limit access to files to only authorized persons, and does it routinely encrypt all data flowing to and from the storage array?
- Does the backup system permit round-the-clock access to files, according to various search criteria, and can files be restored, if necessary, within minutes?
- Does the system restore files and support software, including binderies, after computer equipment malfunctions, is destroyed, or becomes inaccessible to users? Can the system return lost files and software to the repaired equipment or send them to another authorized location?
- Can the system write all versions of all files as they are received from each local area network so that other files and software already stored remain separate and isolated, thus eliminating the spread of any computer viruses? Does the system permit these files to be restored to their original state before a virus occurred?
- Does the system bar manipulation of stored data? Is an electronic audit trail of what happens to this data available to the organization’s internal and external auditors?
- Does the system automatically retain specified data elements online, while providing quick access to off-line items over any period of time (but generally for a minimum of seven years)?
- Is the system cost-effective? Does it eliminate both the obvious and hidden costs of conventional backup procedures, including the cost of on-site storage equipment, off-site storage packaging, as well as the personnel required to back up, catalog, transport, store, and retrieve data?

A key point to consider in evaluating any backup routine is whether the procedure facilitates disaster recovery. To ensure recovery, many small and medium users consider outsourcing.

BACKUP METHOD COST COMPARISON

The costs associated with any backup mechanism cannot be overlooked. Those organizations that back up data consistently incur measurable costs to maintain their backup procedures. These costs include not only the time and cost of the people, equipment, and storage space needed for a backup facility but, more critically, expenses associated with the time to retrieve stored data. Retrieval, the reason organizations back up data, is always labor intensive — an administrator must load and unload tapes, label and store them, and retrieve them as necessary to restore data requested by a user.

Organizations should compare the costs of outsourcing backup to that of an in-house solution. In doing so, they should examine the following factors:

- Maintenance of data integrity (over any length of time).
- Safety of the storage facilities.
- Desired levels of control.
- Convenience of data access.

In addition, consideration must be given to the fact that an organization cannot replace lost data, even if it receives significant insurance payments, and the untimely delivery of data can be useless, if not disastrous, to an organization if crucial information is unavailable when required for critical decision making.

SUMMARY

Information is the heart and soul of an organization. Accepting the importance of data backup is the crucial first step in allocating the resources necessary to ensure that data remains accessible, readable, and retrievable.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter III-9
Legacy Data Base Conversion

James Woods

Converting a legacy, nonrelational computer system to a relational data base system can be a difficult undertaking fraught with uncertainty and trouble areas, but proper planning on the part of the data center manager can identify the risks and accordingly significantly reduce them.

The material presented in this chapter aids the data center manager in planning the move. Encompassing more than the traditional information (e.g., table normalization and project organization), the article examines managerial, political, and other considerations and identifies and discusses the more technical considerations.

SYSTEM CAPABILITIES

Before any project is begun, certain questions must be answered. Why is the organization going to a relational data base? What are the benefits that managers hope to gain? Now, if the list of benefits comes solely from the vendor's representative, the organization may not get a complete and accurate representation of what is to be gained. Instead, managers need to consider what new capabilities mean specifically to the way the organization does business, for example, how the new system will make it easier or faster to do business, thereby lowering overhead?

One of the things to be considered are the capabilities of the old system as opposed to the capabilities of the new. If the abilities of the new system are drawn as a circle or a set, and the abilities of the old system are likewise drawn, the two should have an area where they overlap, representing the union of the two sets. If project managers target this union to be the result of the conversion, they are loosing many of the advantages of the new system. The most desirable objective, therefore, is to gain the whole second set, rather than just the union of those sets. Project managers should not limit the new system by thinking only in terms of the old system.

The legacy system was thought of in terms of applications. The new system should be thought of in terms of models. There is a paradigm shift involved. The most severe shift is to be expected at the technical level. As far as the end user is concerned, there should not be a great difference in the content of the information at the first stage. There certainly may be, after the initial conversion, because at that point, it will be possible to implement the wonderful features that have been talked about for years but were never cost-effective to add.

Generally, one of the benefits of moving to a modern data base system is the facility of the tools. They are better, faster, and more complete. COBOL, for example, may indeed be the mainstay of business, because of the legacy systems, but it is not more powerful than a visually oriented diagramming tool that will automatically set up the users screens, filter the data, and so forth.

PRECONCEPTIONS AND MISCONCEPTIONS

The announcement of an implementation of a legacy-conversion project gives rise to certain predictable reactions within an organization, and the data center manager needs to be aware of the preconceptions members may hold. Two common expectations are:

1. There will be no problems with the new system, or at least the new system will present fewer problems than historically encountered with the legacy system. Human nature is such that staff will expect the new system to be without challenges. Everyone hopes to move from the old, patched system to the new, improved system, one that will not have any problems. This, however, is seldom the case. Whereas the likelihood is that the new system will offer many advantages over the old system, those advantages do not exhibit themselves without effort.

2. The new system will be more efficient. On the contrary, data base performance could very well be lowered when performing the same tasks using a relational system. A data base that must make access path decisions at query time is inherently slower than a system that is preconfigured only to retrieve data in a particular way; such decision-making takes time. However, if a computer hardware upgrade is also involved in the conversion project, the increased demand for central processing unit (CPU) cycles is more than compensated for by the increased power of the new machines. The data center manager should note that if the organization is changing only data base systems, rather than changing data base systems as well as moving to a new, more powerful computer platform, users could most likely suffer a

> performance hit for at least part of the system, perhaps even a major part. This, of course, depends on the efficiency of the existing system. A broad rule is that generalized solutions cost more CPU cycles than specific solutions do. The system does, however, gain great flexibility in return for the additional CPU cycle cost.

To identify and isolate the potential problem areas, the safest route is to perform benchmarks for both the old and new systems. In fact, many organizations make it a condition of sale. The managers can choose samples of transaction data and run them on both systems. At least one of the sample sets should be large, because the response of a data base system is seldom linear. Managers should also be sure to include critical applications in the benchmark set. These are the applications that must fly in order for the new system to be a success.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

BEYOND NORMALIZATION

Any thorough textbook on relational data bases outlines clear instructions on applying standard normalization rules to non-normalized data. (Normalization refers to a procedure used to ensure that a data model conforms to standards that have been developed to avoid duplication of data and to minimize create, update, and delete anomalies of data. Normalization involves the decomposition of a large relation into smaller relations; this process generally improves data integrity and increases the effectiveness of a data base system's long-term maintenance.) Textbook instructions on applying normalization rules generally do not cover some of the difficulties that can be encountered in the conversion of a legacy system. The information content must be reverse-engineered from the legacy data base, at which point the actual normalization can begin. The exceptions that are almost never examined within the text books fall into two categories:

1. *The Data-Definition Shift*. The definition of the data originated, at one point in the system's history, in one form and has evolved into another. One reason is that the data usage is subject to change over the course of many years. For example, what used to be a facility location may now really be labeled "Material Storage Location." This type of shift has important ramifications when deciding how to represent this data in a relational data base.

2. *Incognito Data*. In this situation, the data's name is not necessarily indicative of its function. It is, instead, a statement of the data's original intent. In fact, the name of a data item reflects the understanding of the programmer involved at the time that the first program using that data was written.

DATA REDUNDANCY

On occasion, the conversion process uncovers two, or more, items of data that conflict. Possibly, they have different names, but they serve similar functions. Under the old system, these two or more items do not come into contact, but they may in the new. The function of each piece of data must be clearly understood before a correct model of that data can be made in the relational data base.

Summary Data Redundancy

Many systems store summary data, because the cost, in terms of time/ CPU cycles, is too high to perform the calculations in real time. However, this stored summary data may not match the actual counts or sums. This causes a difficult and embarrassing situation: The first report on the new system does not balance with the report on the old system. This can be distressing because, if the new system is correct, the old system has been wrong for an undetermined amount of time. In any case, the summary data should be discarded in favor of direct calculations from the data base.

Data Conflicts

It is not unusual to have redundant data conflict. For example, in one system, the vendor record was duplicated for each product line that was supplied by that vendor. There was a bug in the update program that caused the system to update the records of only the active products. During the conversion project, when the data was brought over, sometimes one of the old records was picked up and the demographic data was taken from there. There must be a standard decision reached, to apply to all redundant data within the system, as to which data will be considered true. The other data must be discarded during the transfer. However, the data center manager should be advised that this could cause the users to see differences between their old reports and new.

HISTORICAL ERROR TRACKS

A scenario common to organizations going through a conversion project is the existence of hidden, damaged data. What very often has happened is this: At one time in a company’s history, an error occurred in an update program; the program was fixed, and the data was corrected. However, some of the damaged data still lingers in the system. It may never actually show up in user reports, but it will stop the new systems’ data transfer cold, because it violates the very rules that were culled from the program that was supposed to guard that data.

The precaution is simple: Programs must be audited; so must data. For example, a certain field is supposed to contain the groupings of letters INC or SER, which indicate the type of record. Before any transfer attempt is made, a simple program should be written to look at all the records, including historical records, if they are to be transferred to the new data base, to ascertain that indeed those are the only two codes embedded in the data.

If the data base involved employs dictionaries, staff members can use them as a source for the data item name and function, depending on how well the code has been documented. However, if the legacy file or data base system does not have a centralized dictionary, then staff members are dependent on the program code to provide the name and function and thereby the implied function of the data item.

AVOIDING HIDDEN PITFALLS

The larger the number of programs, and the more extended the lifetime of the system, the more likely it is that the data items involved conflict in intent and purpose and perhaps even form and function. At this point, it might be time to start thinking about the planning of the conversion. Even though the legacy system is presumably well understood and the relational data base is thought to be well understood, no manager should assume that the translation from the legacy system to the relational data base will be the simple matter of applying normalization rules to the legacy system.

The first assumption that leads to numerous problems is that the current staff understands the intricacies of the legacy system. Unless their numbers include at least a few members who originally helped to build the system, the assumption should be otherwise. Each staff member working on the conversion has a specific, applications-oriented view of the data, as opposed to a systemwide view, and the conflicts and the anomalies that staff members in systems development have lived with and accommodated within the application code over the years will not be able to be tolerated easily within the new system. The situation calls for a solution to be found, finally.

The second assumption is that relational data bases are well understood. In academic and theoretical circles, this is a true assumption. It is not, however, necessarily true of the organization’s staff, and this staff must be able to support the system. Sending them to the vendors school is a starting point, but it is not a finish line. They must understand relational data bases, but they must also see the need and understand the benefits for the organization.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

THE COMPONENTS OF THE CONVERSION PROCESS

The conversion is not a technical process; rather, it is a managerial process with major technical components. The following sections describe two key considerations in the project.

Defining Documentation

Certain preliminary decisions have been made, so the project has loosely been defined. For example, the organization is determined to move to a relational data base; managers have chosen which data base is to be the replacement system, and a clear picture of the current system has been created. The staff training has been arranged for, and the project is set to go in a couple of months. At this point, what is the procedure?

The first step is to document the current legacy system. If the system has inadequate documentation, the project will be besieged by last-minute surprises during the conversion process and while bringing up the new system. If the system is over-documented (if, indeed, such a thing is possible), the project will be assured of no surprises and a smooth transition. Therefore, logic would dictate that, if the staff does err, they should err on the side of too much documentation.

The term *documentation* requires some definition, because what the manager means by *documentation* and what the programmer means are not necessarily the same things.

To the programmer, documentation means materials that answer such questions as, “When I get ready to make an application that asks for the insured’s middle name, what data item, in what file, will give it to me? ” and “Is there an index on that item?”

What managers mean when they ask for documentation is material that answers such questions as, “When I ask you to modify a particular application, is there some documentation that you can use to find out what that application currently does and what the factors are that will be involved in your modification of that process?”

What end users mean when they ask for documentation, of course, is how they “drive” that application. In the context of this article, users could include either the end user for terminal systems or the operator for batch systems.

At least three different definitions exist for documentation. For the purposes of the conversion, the term actually refers to a combination of all of three, to some degree. Technically, yes, the programmer level documentation must be complete. The interrelationships that the manager wants must be completely documented. The user information, however, does not need to be complete for the purposes of the conversion, but it still needs to be noted and understood.

Determining the Level of User Impact

One of the determinations to make in the labyrinth of management decisions for a conversion effort of this type is estimating the desired degree of impact on the current organization and end users? Questions to consider are: Will applications look the same? Will they act the same? It may be a highly desirable motive, politically and even sociologically, to keep the impact as small as possible. However, minimizing the effects is not a desirable goal, technically. That would mean the project is simply putting new milk in an old bottle, limiting the benefits of the new system by trying to make the system appear as it always did.

One of the things that users usually insist on, of course, are accurate paper reports. Those reports have, over the years, become a definition of their work. Even though there has been much crowing about the benefits of the paperless office for some time now, it has not yet materialized. This does not mean, however, that the office has to be one or the other, entirely based on paper reports or entirely paperless; it is not an all-or-nothing kind of deal. The new system can reduce the amount of paper and still come out way ahead, and the biggest deterrent to being able to achieve great savings in information acquisition and turnaround is the end user who may be emotionally tied to the reports. It has been their private data base; they have been able to mark it up, highlight it, and in general own it. Now, all of a sudden, the new system threatens to take that away from them, and the new data base is in a magic box that the end user does not yet know how to access.

Managers must sell the benefits of online information as opposed to printed information. It is to the corporation’s benefit to head in this direction, as many of the modern data base systems are oriented toward online information retrieval, as opposed to printed information. True, the new system can be created to replicate the old reports, but this approach misses one of the major benefits of an online data base.

DATA HISTORY IN THE NEW SYSTEM

Legacy systems typically have a particular way of trapping the data's history. Some remove the record, or a copy of it, to another file. Others record the history within the record itself. A relational data base, however, is designed to model the current data flow. It is a model of the current data within the organization. The model reflects the data as it is rather than as it was. Usually, the plan should be to trap the information in a number of historical tables, which must be designed at the outset.

SUMMARY

In general, time for planning is crucial. Conversions succeed or fail in the planning stage. The management challenge is most often seen as technical, but there are many areas to manage in such an endeavor. The technical planning is a critical activity, but so are managing expectations of the new system, selling the capabilities of the new system, and providing a plan to implement those capabilities into company strategic tools that help put the organization ahead of the competition.

There has never been a conversion that was "over planned"; however, many that have not been planned in sufficient detail to succeed.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Bookmark It

Search this book:

Go!

Previous | Table of Contents | Next

Section IV Networking

Data center operations managers at one time had to be experts only in mainframe technology. They still must be experts in computing technology, and today computing technology means networking technology. This section brings together overviews in some of the latest networking technology, plus gives data center managers tips on how to keep networks up and running.

For many users, their first contact with the network is their modem. To data center managers, their interface with these users is the line these modems use to connect with the mainframe or server. Chapter IV-1, “New Modem Technologies and Trends,” explains the latest modem technologies, how they operate, how they are best applied, and how they connect to the enterprise networks.

Ethernet has become the de facto LAN standard, and though its obsolescence has been predicted several time during the past 10 years, it remains a vital technology. Chapter IV-2, “An Ethernet Technology Tutorial,” gives data center managers a solid base in this technology and shows how to reduce Ethernet network congestion and increase performance. As Chapter IV-3, “Introduction to 1000BASE-T: Fast (and Faster) Ethernet,” explains, Ethernet technology has been revitalized by new standards that have given the capability to support service at 100 Mbps. The chapter explains how the standard works and how it affects network design.

Another important trend in networking has been the convergence of voice, video, and data over local area networks (LANs), which has been made possible by such technology as Fast Ethernet. Chapter IV-4, “Voice and Video over the LAN,” discusses when this type of networking application is appropriate for an organization and then explains network requirements for implementing this application.

The last leading-edge technology covered in this section is computer telephony integration (CTI). This technology ranges from voice mail to computer/fax capabilities to advanced, telephone-based, menu-driven systems. CTI can help reduce an organization’s wiring infrastructure as well as help reduce the workload at the help desk. Chapter IV-5, “Computer Telephony: An Evolving Technology,” explains how CTI is being integrated with LANs and the Internet.

Providing security and the ability to reconfigure to meet changing needs are two pressing concerns of network managers. Virtual networks dedicate bandwidth to users, which makes data transmissions more secure, and such networks allow users connect to them regardless of the users’ location, which obviates many reconfiguration problems. Chapter IV-6, “Virtual Networking Management and Planning,” explains these and other features of virtual networking.

Because many businesses depend on networks for their existence, this section would not be complete without a chapter on how to ensure network availability. Chapter IV-7, “The Data Center Manager’s Guide to Ensuring LAN Reliability and Availability,” examines reliability and availability in different LAN configurations and how they are achieved with the components that make up a LAN.

Chapter IV-1
New Modem Technologies and Trends

Nathan J. Muller

Traditionally, modems have been used for transferring files, accessing bulletin boards, and connecting users to the Internet. Today, there are multifunction modems that include sound boards, message centers, and fax capabilities. There are also modems that work over wireless and cable television (CATV) networks, and modems that are capable of supporting a voice conversation and data transfer simultaneously over the same line.

Even with all these innovations, modems still perform the same basic functions — modulation and demodulation. They convert (i.e., modulate) the digital signals generated by standalone or networked PCs into analog signals suitable for transmission over dialup telephone lines or voice-grade leased lines. Another modem, located at the receiving end of the transmission, converts (i.e., demodulates) the analog signals back into their original digital form.

THE EVOLUTION OF PACKAGING

For years, modems have been available in external, rackmount, or internal versions. But even in packaging, there is plenty of room for innovation.

External Modems

External modems are standalone hardware devices that connect to a microcomputer's communications port via telephone cabling. They are equipped with front-panel status indicators that inform users of modem activities. Rackmount modems are full- or half-cards that reside in an equipment frame. From there, the individual modems connect to the various PCs. External modems typically reside on an office desk, while the rackmount versions are located in a convenient equipment cabinet or wiring closet for easy troubleshooting and maintenance.

Internal Modems

Internal modems, which insert into an available expansion slot inside the computer, are best suited for users who rely extensively on the wide area networkwide area network(WAN) to do their jobs. For users who require only occasional access to the WAN, companies can save money by equipping a communications server with a pool of modems that can be shared by many users on a first-come, first-served basis.

PCMCIA Cards. Another type of internal modem is the size of a credit card and standardized by the Personal Computer Memory Card International Association (PCMCIA). Both wireline and wireless modems are available as plug-in cards to the PCMCIA Type II slots in portable computers. There are even multifunction PCMCIA cards that combine the modem and LAN adapter (Ethernet or Token Ring) on the same card, giving users more connectivity options without requiring an additional card and using up the second PCMCIA slot included with most laptop and notebook computers.

There are also PCMCIA cards that connect to various wireless messaging services. They have a built-in antenna and can even act as standalone receivers when the computer is turned off. Some cards are programmable, allowing users to access or receive messages from different wireless services, including those based on cellular digital packet data (CDPD) and packet radio services such as ARDIS and RAM Mobile Data. Some wireless modems can automatically identify the type of modem protocol used at the receiving end and adjust their own operation and speed accordingly.

CDPD Modems. Like wireline modems, wireless modems are packed with functionality. There are CDPD modems, for example, that work with any DOS- or Windows-based computer, supporting V.22bis, V.23, V.23bis, V.42bis, Group 3 fax and V.17 wireline fax and data protocols, plus Microcom's MNP-10 cellular protocol or Paradyne's Enhanced Cellular Throughput (ECT) protocol.

| Previous | Table of Contents | Next |

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Other Multiport Modems. Another new way vendors are packaging modems is by combining them with integrated services digital network (ISDN) terminal adapters. This allows users to communicate with conventional dialup services at up to 28.8K bps and also take advantage of ISDN when possible — all without cluttering the desktop or having to use up scarce slots in the PC.

Multiport modems of this kind are being introduced for the corporate environment. U.S. Robotics, for example, offers two models that dynamically support both digital and analog connections. Aimed at telecommunications managers who use a lot of analog lines on their networks, the company's 8-port MP/8 I-Modem and 16-port MP/16 I-Modem can handle analog and ISDN calls, making them suitable for companies that are rolling out ISDN connections to analog users. The modems actually determine whether they are connected to an analog or digital line and operate in that mode automatically. The MP modems act as a front end to existing terminal servers. With a channel-aggregation feature, users can add up to 64K bps of bandwidth over each ISDN link during Internet access or other tasks that call for additional bandwidth.

Multimedia Extensions (MMX) Technology. The next step in modem packaging is to eliminate the need for dedicated hardware altogether. Intel is trying to do precisely this with its native signal processing (NSP) initiative that embeds software emulation of modem and sound-card hardware in the Pentium chip. Known as MMX technology, the idea is to allow any Windows application to have access to features implemented in a special driver — to send and receive e-mail and faxes or to play sound files — with no modem or specialized hardware aside from the main CPU. In effect, NSP will give every Pentium PC the ability to manage basic communications and multimedia tasks, which in turn will let software developers add these features to their applications without having to worry about whether customers have the necessary hardware.

Intel’s MMX technology is what makes native signal processing on the Pentium platform feasible. A 100-MHz Pentium would be required to dedicate 60% of its resources to V.34 modem processing. With MMX, a 200-MHz Pentium would need to use only 20% of its resources, which is insignificant to most users.

Host Signal Processor Modems. Motorola is about to enter the emerging market for host signal processor (HSP) modems, which rely heavily on software and the processing power of the host PC. Rather than relying on its own digital signal processor (DSP), this type of modem is based on a less-expensive application-specific integrated circuit (ASIC) and takes advantage of a PC’s central Pentium chip for processing power. Motorola plans to introduce an HSP product line geared toward remote access, telecommuters, and mobile workers. The product line will eventually go beyond analog modems to include ISDN terminal adapters and digital subscriber line modems. HSP technology has been around for several years, but has only recently become feasible because of the growing availability of faster desktop computers based on the Pentium chip.

MODEM FEATURES

In an effort to distinguish theirs from others on the market, modem manufacturers are continually redesigning their products to incorporate the latest standards, enhancing existing features, and adding new ones. Advancements in modulation techniques, error correction, data compression, and diagnostics are among the continuing efforts of modem manufacturers.

Modulation Techniques

The modulation technique has a lot to do with the speed and reliability of data transmission. Modems convey information by exchanging analog symbols, each of which represents multiple bits.

The symbol rate for modems operating over ordinary phone lines is limited to about 3,400 bps. When a modem has data to transmit, a bit sequence is selected from a pool of available symbols to represent that particular sequence. By packing more data bits into one symbol, modems can achieve higher bit rates. For example, the V.34+ specification squeezes up to 9.8 bits per symbol, versus only 8.4 bits for V.34. Under ideal line conditions, this equates to 33.6K bps for V.34+ modems and 28.8K bps for V.34 modems.

Increasing the size of the symbol pool allows the modem to adjust to a range of noise conditions, which results in an overall higher speed. The V.34+ offers 1,664 symbols, whereas V.34 offers only 960. The higher number of symbols makes it easier for the receiving modem to differentiate data from noise, resulting in more-reliable transmission.

Both V.34+ and V.34 modems use adaptive techniques that enable them to learn about the quality of the line and make adjustments. For example, the sending and receiving modems exchange a set of signals to determine the maximum transmission rate on a particular circuit before user data is actually sent. They also compensate for signal loss detected on a line. If the line exhibits signal loss, the modem can guess how a signal is likely to be degraded across a circuit and boost the signal accordingly to offset the impairment.

These techniques do not guarantee higher speeds and error-free transmission. They only optimize performance on a line-by-line basis.

Speed

Regardless of manufacturer or standards, the advertised data rate of most modems does not always coincide with the actual data rate. This is because the quality of the connection has a lot to do with the speed of the modem.

If the connection is noisy, for example, a 28.8K-bps modem may have to step down to 24K bps (i.e., "fall back") to continue transmitting data. Likewise, a 19.2K-bps modem more frequently operates at 9.6K bps when the line gets too noisy.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous Table of Contents Next

Error Correction

Networks and telecommunications carriers often contain disturbances with which modems must deal or, in some cases, overcome. These disturbances include attenuation distortion, envelope delay distortion, phase jitter, impulse noise, background noise, and harmonic distortion — all of which negatively affect data transmission. To alleviate the disturbances encountered with transferring data over leased lines (without line conditioning) and dialup lines, most products include an error-correction technique in which a processor puts a bit stream through a series of complex algorithms before data transmission.

The most prominent error-correction technique has been the Microcom Networking Protocol (MNP), which uses the cyclic redundancy check (CRC) method for detecting packet errors, and requests retransmissions when necessary.

Link access procedure B (LAP-B), a similar technique, is a member of the high-level data link control (HDLC) protocol family, the error-correcting protocol in X.25 for packet-switched networks. LAP-M is an extension to that standard for modem use and is the core of the International Telecommunications Union (ITU) V.42 error-correcting standard. This standard also supports MNP Classes 1 through 4. Full conformance with the V.42 standard requires that both LAP-M and MNP Classes 1 through 4 are supported by the modem. Virtually all modems currently made by major manufacturers conform with the V.42 standard.

The MNP is divided into nine classes. Only the first four deal with error recovery, which is why only those four are referenced in V.42. The other five classes deal with data compression. The MNP error-recovery classes perform the following functions:

- MNP Classes 1 to 3 packetize data and, the manufacturer claims, ensure 100% data integrity.
- MNP Class 4 achieves up to 120% link throughput efficiency via Microcom's Adaptive Packet Assembly and Data Phase Optimization, which automatically adjusts packet size relative to line conditions and reduces protocol overhead.

Data Compression

With the adoption of the V.42bis recommendation by the ITU in 1988, there is a single data-compression standard — Lempel-Ziv. This algorithm compresses most data types, including executable programs, graphics, numerics, ASCII text, or binary data streams. Compression ratios of 4:1 can be achieved, although actual throughput gains from data compression depend on the types of data being compressed. Text files are the most likely to yield performance gains, followed by spreadsheet and database files. Executable files are most resistant to compression algorithms because of the random nature of the data.

Diagnostics and Other Features

Most modems perform a series of diagnostic tests to identify internal and transmission-line problems. Most modems also offer standard loopback tests, such as local analog, local digital, and remote digital loopback. Once a modem is set in test mode, characters entered on the keyboard are looped back to the screen for verification.

Most modems also include standard calling features such as automatic dial, answer, redial, fall back, and call-progress monitoring. Calling features simplify the chore of establishing and maintaining a communications connection by automating the dialing process. Telephone numbers can be stored in nonvolatile memory.

Other standard modem features commonly offered include fall back and remote operation. Fall back allows a modem to automatically drop, or fall back, to a lower speed in the event of line noise, and then revert to the original transmission speed after line conditions improve. Remote operation, as the name implies, allows users to activate and configure a modem from a remote terminal.

SECURITY

Many businesses have become increasingly aware of the importance of implementing a thorough network security strategy to safeguard valuable network data from intruders. Modems that offer security features usually provide two levels of protection: password and dial-back. Password protection requires the user to enter a code, which is verified against an internal security table. Many modems can store multiple passwords.

The dial-back feature offers a higher level of protection. Incoming calls are prompted for a password, and the modem either calls back the originating modem using a number stored in the security table or prompts the user for a telephone number and then calls back.

Security procedures can be implemented before the modem handshaking sequence, rather than after it. This effectively eliminates the access opportunity for potential intruders. In addition to saving connection establishment time, this method uses a precision high-speed analog security sequence that is not even detectable by advanced line-monitoring equipment.

For the highest level of security, some modems even support the Data Encryption Standard (DES). Although DES has been around since 1977, it is still one of the most effective means of protecting data. DES-based encryption software uses an algorithm that encodes 64-bit blocks of data and uses a 56-bit key. The length of the key imposes a difficult decoding barrier to would-be intruders because 72 quadrillion (72,000,000,000,000,000) keys are possible.

TRANSMISSION TECHNIQUES

Modems use two types of transmission techniques: asynchronous or synchronous. The user's operating environment determines whether an asynchronous or synchronous modem is required.

During asynchronous transmission, start- and stop-bits frame each segment of data during transfer to distinguish each bit from the one preceding it. Synchronous transmission transfers data in one continuous stream; therefore, the transmitting and receiving data terminal equipment (DTE) must be synchronized precisely to distinguish each character in the data stream.

PC-to-PC or PC-to-Mainframe

Most mainframes and minicomputers use synchronous protocols, whereas PC-to-PC communications are typically asynchronous. Users who require both PC-to-PC and PC-to-mainframe communications can purchase modems that support both types of transmissions. The software that comes with the modem usually supports several emulation techniques for file transfers between hosts and PCs.

Those who require more out of the PC-to-host link than simple file transfer can look to such software as Attachmate Corp.'s Extra Personal Client 6.1and Wall Data Inc.'s Rumba Office 95/NT 5.0. Both products offer advanced data-query capabilities. These and other host-access suites offer a variety of connection types and methods within one product, usually installed off a single CD-ROM. At a minimum, such products can connect to an IBM 3270 or AS/400 host without having to depend on IBM's DOS-based drivers.

Previous | Table of Contents | Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

WIRELESS MODEMS

Wireless modems are required to transfer data over public wireless services and private wireless networks. These modems come in a variety of hardware configurations: standalone, built-in, and removable PCMCIA card.

Newer modems are programmable and therefore capable of being used with a variety of wireless services using different frequencies and protocols. There are even modems that mimic wireline protocols, allowing existing applications to be run over the wireless network without modification.

Private Wireless Networks

Private wireless networks operate in a range of unique frequency bands to ensure privacy. Using radio modems operating over dedicated frequencies within these frequency bands also permits the transmission of business-critical information without interference problems. Furthermore, the strategic deployment of radio modems can provide metropolitan area coverage without the use of expensive antenna arrays.

Such modems are designed to provide a wireless, protocol-independent interface between host computers and remote terminals located as far away as 30 miles. Most provide a transmission rate of at least 19.2K-bps point-to-point in either half- or full-duplex mode. Some radio modems even support point-to-multipoint radio network configurations, serving as a virtual multidrop radio link that replaces the need for expensive dedicated lines (see Exhibit 1). In this configuration, one modem is designated as the master, passing polling information and responses between the host and terminals over two different frequencies.

Exhibit 1 Radio modem configuration.

In multidrop configurations, a radio network is capable of supporting one type of asynchronous or synchronous polling protocol. Because such modems perform no processing or interpreting of the protocol, the host (or front-end processor) must generate all required protocol framing, line discipline, node addressing, and data encapsulation. Depending on the vendor, these modems may be equipped with an integral repeater to maintain signal integrity over longer distances.

MULTIFREQUENCY MODEMS

Regardless of the transmission technology or the hardware configuration used, to operate properly, the modem must be tuned to the frequency of the service provider's wireless network. Until recently, modems were offered in different versions according to the wireless network the modem would connect to. This delayed product development and inflated the cost of manufacturing, which was passed on to users in the form of higher prices for equipment.

To overcome these problems, chip manufacturers have developed programmable chipsets that are not limited to a specific network's radio frequency. Newer wireless modems are computer configurable. Within specified frequency ranges, the transmit and receive frequencies are independently selectable via software.

MULTIMEDIA (HYBRID) MODEMS

Not only can modems be programmed for multifrequency use, they can provide seamless integration of multiple media — wireline and wireless — through a common programmable interface. This is accomplished with a chipset that supports both wireline and wireless communications. Special software used with the chipset provides a method for connecting cellular phones to modems, which is important because cellular phones lack dial tones and other features used by modems on the wireline phone network. The software makes it appear that those features exist.

CABLE MODEMS

Although today's computers powered by Pentium and PowerPC chips are better equipped than ever to handle multimedia and video, they face a bottleneck that, in most cases, offers top speeds of no more than 28.8K bps or 33.6K bps over dialup lines. A new type of modem, the cable modem, has emerged for delivering entertainment and information services, including Internet access, to television sets and PCs over the installed base of ordinary twisted-pair wiring. Using traditional coaxial cable installed by CATV operators, these modems can deliver speeds of up to 1,000 times that of today's analog modems.

Cable modems are really not modems in the conventional sense. They modulate and demodulate signals like a conventional modem, but otherwise they are more like routers that are designed for installation on CATV networks, which themselves operate much like Ethernet LANs. These cable modems, as well as the cable operators' plant equipment, are even being managed using the familiar simple network management protocol (SNMP).

Upstream Interference

Typically, a cable modem sends and receives data in two slightly different fashions. In the downstream direction (from the network to the user), the digital data is modulated and then placed on a typical 6-MHz television carrier. There are several modulation schemes, but the two most popular are quadrature phase-shift keying (QPSK), which provides up to 10M bps, and quadrature amplitude modulation (QAM), which can provide up to 36M bps. This signal can be placed in a 6-MHz channel adjacent to TV signals on either side without disturbing the cable television video signals.

The upstream channel (from the user to the network) is more complicated. Typically, in a two-way activated cable network, the upstream channel is transmitted between 5 and 40 MHz. This tends to be a noisy environment with lots of interference from ham radio, CB radios, and impulse noise from home appliances or office machines. In addition, interference is easily introduced in the home, due to loose connectors or poor cabling. Because cable networks are tree and branch networks, all this noise gets combined, and increases as the signals travel upstream.

Most manufacturers will use QPSK or a similar modulation scheme in the upstream direction because it is a more robust scheme than higher-order modulation techniques in a noisy environment. The drawback is that QPSK is slower than QAM.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous Table of Contents Next

ROLE OF DIGITAL SIGNAL PROCESSING

Multifunction modems use programmable digital signal processing (DSP) technology to turn a computer into a complete desktop message center, allowing the user to control telephone, voice (recording and playback), fax, data transfers, and e-mail. Typical features include multiple mailboxes for voice mail, caller ID support, call forwarding, remote message retrieval, phone directory, and contact data base.

In some cases, the modem is actually on a full-duplex sound card. By plugging in speakers and a subwoofer, the user can even enjoy a stereo-sound speakerphone. A separate connection to a CD-ROM player allows the user to work at the computer while listening to music. However, these DSP-based products cannot be used as modems and sound cards simultaneously because the processor can take on only one identity at a time.

With DSP, the modem can be easily upgraded to the latest communications standards and new capabilities can be added simply by loading additional software. For example, a 14.4K-bps modem can be upgraded to 28.8K bps by installing new software instead of having to buy new hardware. Likewise, a 28.8K-bps modem can be upgraded to 33.6K bps in the same way, often at no extra charge from the vendor.

Digital Simultaneous Voice and Data

A new type of modem — digital simultaneous voice and data (DSVD) — allows the user to send voice and data at the same time over a single telephone line. Interference is avoided by having voice and data use different frequencies.

The biggest advantage of DSVD is that users no longer need to interrupt telephone conversations or install a separate line to transmit data or receive faxes. Multimedia modems typically include full-duplex speakerphone, fax, and 16-bit stereo audio capabilities, in addition to advanced modem functionality.

56K-bps MODEMS

Until now, 33.6K was thought to be the practical modem speed over standard phone lines. A new class of modems that can transmit data at 56K bps employs technology that takes advantage of the fact that for most of its length, an analog modem connection is really digital.

When an analog signal leaves the user's modem, it is carried to a phone company central office, where it is digitized. If it is destined for a remote analog line, it is converted back to an analog signal at the central office nearest the receiving user. However, if the receiving user has a digital connection to the carrier's network, the modem traffic is converted at only one place, where the analog line meets the central office. When the traffic is converted from analog to digital, noise is introduced that cuts throughput. But the noise is less in the other direction, from digital to analog, allowing the greater downstream bandwidth (see Exhibit 2).

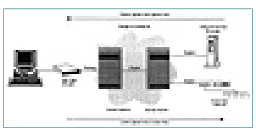


Exhibit 2 56K-bps modem.

Because it is already in digital form, the traffic is free of impairments from noise introduced when an analog modem signal is made digital within the carrier network. From an analog source, the top speed is quite a bit lower than 56K bps because the traffic is subject to impairment from noise.

A fully digital ISDN basic rate interface line offers up to 128K bps in both directions, but the service is far more expensive than an analog phone line. In addition, ISDN is not universally available and can still be difficult for users to set up with their equipment. Although the new 56K-bps modems are not meant to challenge ISDN, they will fulfill a role as a more economical alternative and significantly cut into the ISDN market.

Among the vendors of 56K-bps modems are U.S. Robotics and Motorola. As a key contributor to many past standards, U.S. Robotics plans to lead the effort to standardize 56K-bps technology worldwide. However, because of the typically lengthy time to bring a standard to fruition, U.S. Robotics is bringing 56K-bps modems to market immediately using its x2 proprietary technology.

Motorola also announced plans for standards-based modems as part of its broader line of ISDN, cable modem, and other high-speed communications products. Like U.S. Robotics, Motorola has been a key developer of the technologies contained in virtually all of the earlier "V" modem standards. The company intends to add significant portions of its own technology to the new

56K-bps modem standard.

CONCLUSION

Once thought to be an outdated technology that would be supplanted by ISDN terminal adapters connected to digital lines, modems are not only growing in use, they are undergoing a surge in innovation as well. Not only has ISDN not met industry projections in terms of availability, but the carriers appear not to be encouraging its use, as evidenced by significant price hikes.

Higher-speed modems, the advent of cable modems that work over CATV networks, and new technologies that rely more on a computer's CPU for carrying out modem functions have all combined to breathe new life into this market segment. Although today's new-generation modems will not replace the need for ISDN in all cases, they certainly will give many potential ISDN users reason to consider the modem as a practical and economical near-term alternative.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter IV-2
An Ethernet Technology Tutorial

Jeff Leventhal

Ethernet networking has come a long way since it was pioneered more than 20 years ago by Robert Metcalfe, a scientist at Xerox's Palo Alto Research Center (PARC) who envisioned computers connected locally. Messages would be broadcast onto a coaxial cable (what Metcalfe would eventually call the "ether," after the medium once postulated as the carrier of electromagnetic radiation). Each intended destination would pick up the data packets carrying its address. When packets collided, their nonarrival would be detected by transmitters, which would back off and, after a random interval, try again until the packets went through.

Metcalfe did not believe that 100% reliability was vital to a workable local area network (LAN). He reasoned that it was logical to purchase cheaper interfacing hardware and software in spite of errors and inefficiencies, provided the system could guarantee that mistakes would be identified and corrected at a higher level of protocol software.

In 1980, Metcalfe recruited Digital Equipment Corp. and Intel Corp. to join Xerox in submitting his Ethernet specification to the IEEE 802 committee to make it a standard. The 802.3 Ethernet standard that emerged differed only slightly from his original submission. Metcalfe launched a company in the same year to build Ethernet interfaces.

This chapter presents an overview of basic Ethernet technology, the most widely used LAN-access method in business settings today. Ethernet networks consist of many different pieces of hardware and software that must adhere to several different industry standards specifications. The following sections describe what each component does, discuss how collisions are caused on Ethernet networks, and propose ways to reduce LAN congestion.

ETHERNET ADDRESSES

The media access control address is the unique hexadecimal serial number assigned to each Ethernet network for identification. This address is permanently set when it is manufactured. Although the serial number can be changed through software, this is not generally recommended.

Exhibit 1 illustrates the components of an Ethernet packet. Each card has a unique MAC address so that it can exclusively grab packets that are meant for it off the wire. If MAC addresses are not unique, there is no way to distinguish between two stations. Devices on the network watch network traffic and look for their own MAC address in each packet to determine whether they should decode it. Special circumstances exist for broadcasting to every device.

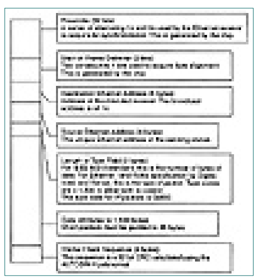


Exhibit 1 An ethernet packet.

The MAC addresses are exactly 6 bytes long and are usually written in hexadecimal (i.e., 12:34:56:78:90:AB). Each manufacturer of Ethernet devices applies for a certain range of MAC addresses. The first three bytes of the address determine the manufacturer. A current listing of vendor MAC address assignments is available on ftp.lcs.mit.edu under pub/map/Ethernet-codes.

The Preamble. The preamble of an Ethernet packet is a seven-octet field of alternating 1 and 0 binary bits sent prior to each frame to allow the circuitry to reach its steady-state synchronization with received-frame timing. The start-frame delimiter of the Ethernet packet is a binary sequence (i.e., 10101011) immediately following the preamble and indicating the beginning of a frame.

Cyclic Redundancy Check. A cyclic redundancy check (CRC) is a method of detecting errors in a message by performing a mathematical calculation on the bits and then sending the results along with the message. The receiving workstation performs the same calculation on the message data and then checks the results against those transmitted at the end of the message. If the results do not match, the receiving end asks the sending end to send again.

A broadcast address is the unique address that identifies a packet as appropriate to all receiving stations. According to the IEEE 802.3 standard, this is any address in which the second byte is an odd number.

IEEE STANDARDS

The Institute of Electrical and Electronics Engineers (IEEE) established a naming convention for the different physical types of Ethernet. For example, in 10BaseT:

- The 10 stands for the signaling speed of 10 MHz.
- Base refers to baseband.
- The letter following the speed indicator and Base refers to the media type. T means twisted pair and F means fiber.

Other examples of this naming convention are:

- 10Base2 is 10-MHz Ethernet running over thin, 50-ohm baseband coaxial cable. 10Base2 is sometimes referred to as Thin Ethernet or CheaperNet.
- 10Base5 is 10-MHz Ethernet running over standard (thick) 50-ohm baseband coaxial cable.
- 10BaseF is 10-MHz Ethernet running over fiber-optic cabling.
- 10BaseT is 10-MHz Ethernet running over unshielded twisted pair cable.
- 100BaseTX is 1,000-MHz Ethernet running over two pair of category 5 cabling.
- 100BaseT4 is 100-MHz Ethernet running over four pair of category 3, 4, or 5 cabling.
- 100BaseFX is 100-MHz Ethernet running over two strands of fiber-optic cabling.

ETHERNET COMPONENTS

A network segment, sometimes referred to as a subnet, is a piece of network wire bound by bridge, router, repeater, or terminators.

A transceiver allows a station to transmit and receive to and from the common medium. In addition, Ethernet transceivers detect collisions on the medium and provide electrical isolation between stations. The IEEE term for a transceiver is medium access unit (MAU). An attachment unit interface (AUI) is an IEEE term for the connection between a controller and the transceiver.

A repeater acts on a purely electrical level to connect segments. It amplifies and reshapes (and, depending on the type, possibly retimes) the analog waveform to extend network segment distances. It knows nothing about addresses or forwarding, thus it cannot be used to reduce traffic.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Hubs

A hub connects servers, workstations, and other devices to form a network. When a station sends a request for a file from the server, the request, contained in Ethernet packets, first travels through the hub. The hub then broadcasts the packet to each of its ports, regardless of its actual destination. All stations connected to the hub must take turns sending packets on the shared-bus backbone of the hub, sharing its bandwidth. Some hubs (e.g., as 10BaseT and Standard Microsystems Corp.'s Active Arcnet) include electronics to regenerate and retime the signal between each hub port. Others (e.g., as 10BaseF or Passive Arcnet) simply act as signal splitters, similar to the multitap cable-TV splitters that might be used on private home-antenna coaxial cables.

Stackable Hubs. A relatively new twist on hubs is the ability to stack them. Instead of adding Ethernet ports in a hub by buying standalone hubs, the stackable hub offers a less expensive and complete solution with options to add more features and more hubs to the stack.

Hubs in a stack are connected with connector cables that enable them to act as one hub. The connector cable keeps the entire group of hubs on the same collision domain while allowing management of the stack from one Simple Network Management Protocol agent. The entire stack counts as a single logical repeater and therefore avoids the previous repeater hop limitations of joining standalone hubs.

Some stackable hubs can be made more efficient by segmenting their single collision domain into smaller collision domains. Each of these smaller domains can be connected to a port on a 10M-bps or 100M-bps switch to alleviate network congestion.

Bridges

A bridge connects distinct segments (usually referring to a physical length of wire) and transmits traffic between them. The user can thus extend the maximum size of the network without breaking the maximum wire length, attached device count, or number of repeaters for a network segment.

A remote bridge has an Ethernet interface on one side and a serial interface on the other. It connects to a similar device on the other side of the serial line. It is most commonly used in wide-area network (WAN) links, where it is impossible or impractical to install network cables. A high-speed modem and intervening telephone lines or public-data network would be used to connect the two remote bridges.

According to the IEEE 802.1 specification, the maximum number of concatenated bridges in a bridged LAN is seven. This number is arbitrary, however, and is based on simulations of application performance with expected bridge delays. It is helpful to limit the number of bridge hops to four, with no more than one of these being a WAN-linked remote bridge.

Routers

Routers work much like bridges, but they pay attention to the upper network layer protocols (OSI, or Open Systems Interconnection, layer 3) rather than physical layer (OSI layer 1) protocols. A router decides whether to forward a packet by looking at the protocol-level addresses (i.e., Transmission Control Protocol/Internet Protocol, or TCP/IP, addresses) rather than the MAC address. Because routers work at layer 3 of the OSI stack, it is possible for them to transfer packets between different media types (i.e., leased lines, Ethernet, Token Ring, X.25, frame relay, and FDDI). Many routers can also function as bridges.

Repeaters work only at layer 1, bridges at layers 2 and 1, and routers at layers 3, 2, and 1. Some confusion stems from the fact that some bridges do routing, although that routing is not based on network or higher-layer header information. Likewise, routers also perform bridging when they deal with a protocol that has network layer bytes such as DEC LAT and NetBIOS.

Why Can’t a Router Be More Like a Bridge? An often-repeated question when designing a network is whether to use a router or a bridge. There is no absolute answer to this. The network traffic, layout, type, and the number of servers must be considered. Routing is always preferable to bridging; however, routers are slower and usually more expensive because of the amount of processing required to look inside the physical packet and determine which interface to send the packet to. Also, many applications use nonroutable protocols (i.e., NetBIOS, DEC LAT).

Bridges are usually good choices for small networks with few, if any, slow redundant links between destinations. Bridges may be the only choice for certain protocols, unless the user has the means to encapsulate (i.e., tunnel) the unroutable protocol inside a routable protocol.

Routers are usually much better choices for larger networks, particularly where the user wants to have a relatively clean WAN backbone network. Routers are better at protecting against protocol errors (such as broadcast storms) and bandwidth utilization. Because routers look deeper inside the data packet, they can also make forwarding decisions based on the upper-layer protocols. Occasionally, a combination of the two devices is the best way to go. Bridges can be used to segment small networks that are geographically close to each other, between each other and the router to the rest of the WAN.

Drivers

A driver is the software that allows an Ethernet card to decode packets and send them to the operating system, and encode data from the operating system for transmission by the Ethernet card through the network. By handling the nitty-gritty hardware-interface chores, the driver provides a device-independent interface to the upper-layer protocols, thereby making them more universal and easier to develop and use.

The network driver interface specification (NDIS) created by Microsoft Corp. allows the “stacking” of multiple protocols on a single underlying driver. A single Ethernet card in a PC can speak many different network “languages,” usually at the same time.

A packet driver is another method of allowing multiple protocols to access the network interface at the same time and provides a device-independent interface to various TCP/IP applications. The open data-link interface (ODI) driver is Novell Inc.’s and Apple Computer Inc.’s equivalent of NDIS. Users should select the driver most commonly required by their software.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

AVOIDING PACKET COLLISIONS

Ethernet is a type of baseband network. A baseband network provides a single channel for communications across the physical cable so only one device can transmit data at a time. Devices on a baseband network are permitted to use all the available bandwidth for transmission. An analogy is a single phone line that runs into a private home — only one person can talk on the line at a time. If more than one person wants to talk, everyone has to take turns.

Ethernet devices take turns transmitting packets of information using a method called carrier sense, multiple access with collision detection (CSMA/CD). Before an Ethernet device puts a packet "on the wire," it listens to determine whether another device is already transmitting. Once the device finds that the wire is clear, it starts sending the packet while also listening to hear if another device has started sending at the same time (which is called a collision).

A collision occurs when two devices detect that the network is idle and try sending packets at exactly the same time. Because only one device can transmit at a time, both devices must back off and attempt to retransmit again. The retransmission algorithm requires each device to wait a random amount of time, so that the two are very likely to retry at different times, and thus the second one will sense that the network is busy and wait until the packet is finished. If the two devices retry at the same time (or almost the same time), they will collide again, and the process repeats until either the packet finally makes it onto the network without collisions, or 16 consecutive collisions occur and the packet is aborted.

The number and frequency of collisions has a significant impact on network performance and often depends on the applications and protocols running on a network. In many cases, collision rates of 50% will not cause a large decrease in perceived throughput. If a network is slowing down and the percentage of collisions is high, users may try segmenting their networks with either a bridge or router to see if performance improves.

Another means of reducing collisions is to reduce the number of devices on the network segment. This is accomplished by splitting the segment into two pieces and putting a bridge or router between them.

Late Collisions

A late collision occurs when two devices transmit at the same time, but because of cabling errors (most commonly, excessive network-segment length or repeaters between devices) neither device detects a collision. This happens because the time to propagate the signal from one end of the network to another is longer than the time to put the entire packet on the network, so the two devices that cause the late collision never see that the other is sending until after it puts the entire packet on the network. Late collisions are only detected during transmissions of packets longer than 64 bytes. Detection is exactly the same as for a normal collision; it just happens too late. Ethernet controllers do not retransmit packets lost to late collisions.

Typical causes of late collisions are segment cable lengths in excess of the maximum permitted for the cable type, faulty connectors or improper cabling, excessive numbers of repeaters between network devices, and defective Ethernet transceivers or controllers.

When a workstation receives a collision and it is transmitting, it puts out a jam so all other stations will see the collision. When a repeater detects a collision on one port, it puts out a jam on all other ports, causing a collision to occur on those lines that are transmitting, and causing any nontransmitting stations to wait to transmit.

Broadcast Storms. A broadcast storm is an undesirable network event in which many broadcasts are sent simultaneously and across all network segments. A broadcast storm uses substantial network bandwidth. The inevitable result is severe degradation of performance or complete loss of the network as the devices continue to generate more traffic. This can be related to the physical transmission or to very high-level protocols.

Measuring Network Performance

Parameters for measuring network performance are discussed in the following sections.

Channel Utilization. Channel utilization is the percentage of time in which the network cable is busy carrying data. A general rule of thumb is that if utilization exceeds an average of 50% during the busiest time of the day, then it is time to look at solutions to ease the traffic load.

Application Software Degradation. Application software degradation manifests itself when file transfers take a very long time or fail completely.

LAN congestion can cause this condition, but several other variables can contribute as well, including the performance of components in the network server or client workstations, such as the CPU, memory, disk drive, or number of users on the LAN.

User Dissatisfaction. User dissatisfaction is the final arbiter of network performance. Regardless of whether the network is performing within reasonable limits, if users are dissatisfied, then it is time to look for alternatives to reduce congestion.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

INCREASING NETWORK CAPACITY

Segmenting a network using routers is often a low-cost solution to easing network traffic, as illustrated in Exhibit 2. Routing between two segments is a practical option only as long as most of the load generated in workgroup

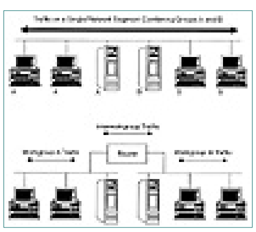


Exhibit 2 Segmenting a LAN with a router.

A remains within the workgroup A segment. If most of the load generated by workgroup A is directed at workgroups B, C, and D, then the intersegment load may cause congestion at the router, and an alternative solution must be sought. A frequent alternative solution is the introduction of a switch.

Switches and Switching

The LAN switch allows information on any port to simultaneously pass to any other port. The switch inspects each incoming Ethernet packet for the destination address and quickly determines the appropriate output port by consulting its internal address map. If the output port on the switch is available, then the switch begins to immediately forward the packet to its destination.

This cut-through approach to switching generally reduces the latency (or waiting period) that it takes to send the packet to its destination and is faster than the store-and-forward approach to switching, which first requires the switch to receive the entire packet before determining its destination and then forwarding it on. In the event that the output port is not available, then the switch buffers the incoming packet and forwards it when the port becomes available.

Cut-through Switching. Cut-through technology is based on the premise that there is no need for a hub to await the arrival of a complete packet before forwarding individual packets to their destinations. These switches wait only long enough to read the destination from the packet's header and immediately begin forwarding it to its destination. The delay in forwarding is limited to approximately 4.8 microseconds (equivalent to the six-byte time frame that contains the header information). Depending on the protocol and packet sizes, this technology can significantly enhance network performance.

Protocols that require an acknowledgment for every packet sent benefit from this technology. Other types of protocols, such as burst-mode IPX, which do not follow the one-request-elicits-one-response method, do not benefit as much from this technology. Also, networks configured to use packet sizes larger than 1,024 bytes will not see as much of an increase in throughput. The larger the packet size, the smaller the aggregate throughput increase in a cut-through switching environment.

A concern to many network managers with cut-through switches is the lack of error-checking capabilities. Because the switch does not wait for the entire packet to arrive before forwarding frames to their destinations, malformed packets, bad packets, and excessive collisions could be passed on to the destination port or segment. In large environments where the switch functions as a central hub or network backbone, propagation of errors should be considered.

Store-and-Forward Switching. Store-and-forward switches maintain address tables for each port or segment. The switch awaits the arrival of each packet entirely, checks the packet for CRC errors, and forwards the packet to its destination only if it is free of errors. The switch flushes and rebuilds the address tables periodically so that it constantly knows where each device is located.

Switches can be placed at different locations in a network configuration depending on the bandwidth needed. Workgroup or departmental switches connect one or more servers and workstations to one switch, thus allowing the workstations to have multiple 10M-bps paths into the server. Enterprise switches connect multiple network segments, other hubs and switches, and common resources such as backup and remote-access systems.

An enterprise switch is sometimes referred to as a collapsed backbone because the switch logically becomes the backbone for the network. Each segment has its own connection, along with each of the workgroup and departmental switches and hubs, thereby greatly increasing the aggregate bandwidth available on the network.

Half- and Full-Duplexing

Ethernet normally communicates at half-duplex. That is, while data can be transferred in any direction at any given time, a workstation can either transmit or receive data, but not both. Given the original physical media used with Ethernet in the late 1970s and early 1980s — coaxial cable — this was the only communications possible because the same wire was used for transmission and reception.

The twisted-pair wires used with 10BaseT and 100BaseT technology include separate wire pairs for transmission and reception. However, it is still difficult to transmit and receive data over twisted-pair wires as long as collisions are likely and the Ethernet adapter will back off from sending data the second it senses that it is receiving data. However, the situation is different in a microsegmented switch environment.

In a microsegmented switch, there is only one device assigned to use a pair of wires because there is only one workstation connected to each port of the switch. Only the attached workstation ever speaks to the switch (using the transmit pair of the cable), and only the switch ever talks to the attached workstation (using the receive pair of the cable). With the removal of any possibility of contention, there is no longer a need for the Ethernet adapter to perform collision detection and back off when a collision occurs. With the elimination of collision detection, the workstation and the switch can transmit in both directions simultaneously and achieve full-duplex communications.

The network requirements for duplex channels are that there are separate wire pairs to transmit and receive data, only one device is connected to each port of the switch, and that the central hub is a switch, not a repeater hub. Even if these requirements are met, only appropriate software applications will be able to take advantage of full-duplexing.

Most applications running on LANs today do not use bandwidth symmetrically. For example, file transfers are asymmetrical in nature; large amounts of data are transferred in one direction, with short acknowledgments being returned in the other direction (e.g., more users download information from the Internet than upload). However, videoconferencing, an application that is gaining popularity, is expected to benefit from full-duplexing.

However, the benefits of full-duplexing are not limited to applications such as videoconferencing. Server operating systems support full-duplex and as a result can send a file to a workstation and receive acknowledgments that the file is being properly received at the same time. These benefits of full-duplexing can be fully realized on the network backbone on a switch-to-switch connection because traffic travels between two switches in both directions simultaneously. Thus two 100M-bps switches can communicate at 200M-bps.

CONCLUSION

The rapid growth of the installed base of Ethernet networks has fueled efforts to improve their efficiency and ease of administration. The need for higher speeds has led to the introduction of gigabit Ethernet adapters and switches. The need for easier ways to administer rapidly expanding networks has led to the purchase of smaller vendors by established enterprise-management-systems vendors (such as the purchase of Preferred Systems by Computer Associates, Hewlett-Packard's purchase of Symantec's Peter Norton Group, and Intel's investment in SMC). The constant requirements for performance and ease of administration are expected to further drive the industry to create new technologies.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter IV-3
Introduction to 100BASE-T: Fast (and Faster) Ethernet

Colin Mick

Fast Ethernet (100BASE-T) is an extension to the IEEE802.3 Ethernet standard to support service at 100M bps. It is virtually identical to 10BASE-T, in that it uses the same media-access control (MAC) layer, frame format, and carrier-sense multiple access with collision-detection (CSMA/CD) protocol. This means that network managers can use 100BASE-T to improve bandwidth and still make maximum use of investments in equipment, management tools, applications, and network support personnel.

100BASE-T is designed to work transparently with 10BASE-T systems. Switches (high-speed, multiport bridges) are used to connect existing 10BASE-T networks to 100BASE-T technology. By building networks with 100BASE-T and 10BASE-T linked with switches and repeating hubs, network designers can build networks that provide four levels of service:

1. shared 10M-bps service
2. dedicated (switched) 10M-bps service
3. shared 100M-bps service
4. dedicated 100M-bps service

Operating at higher speeds with the same frame size and the CSMA/CD protocol requires that 100BASE-T collision domain diameters be smaller —

typically about 200 meters. In 100BASE-T, larger networks are built by combining collision domains by way of switches. Fiber (100BASE-FX) links are used to support long (i.e., 412 meters in half-duplex, 2 kilometers in full-duplex) cable runs. Within a single collision domain, port density is increased by using modular or stacking hubs.

The 100BASE-T standard (IEEE802.3u, 1995) currently defines four physical layer signaling systems:

1. 100BASE-TX supports operation over two pairs of Category 5 unshielded twisted-pair (UTP) or shielded twisted-pair (STP) cables.

2. 100BASE-T4 supports operation over four pairs of Category 3, Category 4, or Category 5 UTP or STP cables.

3. 100BASE-T2 supports operation over two pairs of Category 3, Category 4, or Category 5 UTP or STP cables.

4. 100BASE-FX supports operation over two 62.5-micron multimode fibers.

HOW IT WORKS: AN ISO VIEW

Exhibit 1 depicts an ISO seven-layer diagram comparing 10BASE-T and 100BASE-T. Both 10BASE-T and 100BASE-T defined operations at the lower half of the data link layer (known as the Media Access or MAC layer) and the physical layer. Extension of the Ethernet standard to 100M-bps operation required one small change to the MAC layer operation specified in the IEEE802.3 standard. Originally, timing was defined in absolute terms (i.e., an external reference clock). As a result, timing specifications were defined in milliseconds, nanoseconds, and picoseconds. To support 100M-bps operation, timing was respecified relative to the internal clock of the MAC. This meant that specifications were defined in bit times.

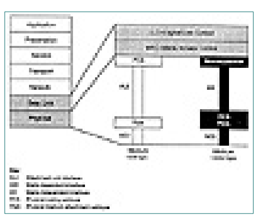


Exhibit 1 100M bps standards model.

Several changes were made at the physical layer. In 10BASE-T, coding (i.e., conversion of data bits to symbols) is done in the PLS layer, directly below the MAC. A mechanical interface called the attachment unit interface (AUI) is situated directly below the PLS. Below the AUI is the PMA layer, which converts the digital symbols into analog symbols that can be sent across the wire and a media-dependent interface (MDI) — a socket for connecting the cable.

100BASE-T puts the coding, called the physical coding sublayer (PCS), below the mechanical interface. This was done to make it possible to offer a variety of coding systems that could be packaged in a transceiver along with the analog/digital circuitry for connection via the mechanical interface. The mechanical interface used for 100BASE-T is called the Media Independent

Interface (MII). It is similar to the AUI, but offers a larger data path and the ability to move management information between the PHY and the MAC. A simple mapping function, called the Reconciliation Sublayer, handles linking the MII to the MAC. As noted previously, 100BASE-T currently supports four signaling systems (see Exhibit 2): 100BASE-TX, 100BASE-T4, 100BASE-T2, and 100BASE-FX.

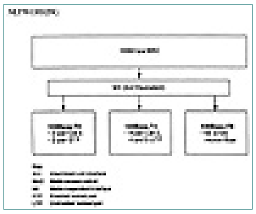


Exhibit 2 100BASE-T physical layers.

Two 100BASE-T signaling systems — 100BASE-TX and 100BASE-FX — are based on the transport protocol/physical medium dependent (TP/PMD) specification developed by the ANSI X3T12 committee to support sending fiber distributed data interface (FDDI) signals over copper wire (see Exhibit 3). TP/PMD uses continuous signaling, unlike the discrete signaling used with 10BASE-T. In 10BASE-T, when a station is finished sending a frame, it sends a few idle signals and then goes quiet, except for a link pulse, which is sent every 16 ms to indicate that the link is still good.

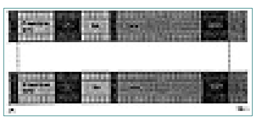
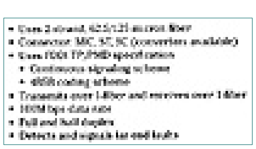


Exhibit 3 100BASE-T (TX and FX) frames.

In TP/PMD, a continuous stream of idle symbols is sent when data is not being transmitted. To ease the transition between data and idle signals, a JK symbol sequence is added to the front of a data frame and a TR symbol sequence is added to the end of the frame before transmission of idle symbols begins. The JK, TR, and idle transmission patterns must be added to Ethernet frames when they are transmitted via the TP/PMD specification.

Both 100BASE-TX and 100BASE-FX use 4B5B coding. This means it takes 5 baud (signal transitions on the wire) to transmit 4 bits of information. This is vastly more efficient than the Manchester coding used for 10BASE-T, which requires 2 baud to send each bit across the wire.

Exhibit 4 summarizes the attributes of 100BASE-FX. It uses two strands of 63.5-micron fiber. All standard connectors are listed in the specification — different manufacturers support different types of connectors. 100BASE-FX uses the FDDI TP/PMD specification with continuous signaling and 4B5B coding. The data clock runs at 125 MHz, providing a signaling rate of 100M bps with the 80% efficiency of 4B5B coding. One fiber is used for transmitting data, the other for receiving data. It can support both half-duplex and full-duplex operation and has automatic link detection.



- Uses 2 strand, 62.5/125 micron fiber
- Connector: MIC, ST, SC (converters available)
- Uses FDDI TP/PMD specification
  - Continuous signaling scheme
  - 4B5B coding scheme
- Transmits over 1 fiber and receives over 1 fiber
- 100M bps data rate
- Full and half duplex
- Detects and signals far end faults

Exhibit 4 10BASE-T-FX.

100BASE-TX

Exhibit 5 summarizes the attributes of 100BASE-TX. It operates over two pairs of Category 5 UTP or STP, and uses Category 5-certified RJ-45 connectors. It uses the 125-MHz data clock, continuous signaling, and 4B5B coding of 100BASE-FX, but adds signal scrambling and MLT-3 conditioning to deal with noise problems associated with sending high-frequency signals over copper. 100BASE-TX uses exactly the same connector pinouts as 10BASE-T. It transmits over one pair and receives over the other. It supports half-duplex and full-duplex operation.

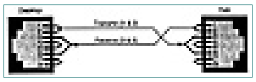


Exhibit 5 100BASE-TX.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

100BASE-T4

100BASE-T4 (see Exhibit 6) is a more complex signaling system because it must support a 100M-bps data rate over cable certified for operation at 16 MHz. This is accomplished by increasing the number of cable pairs used for data transmission and using a more sophisticated coding system. 100BASE-T4 starts with the two pairs used for 10BASE-T — one for transmit and one for receive — and adds two additional pairs that are used bidirectionally. This means that, when transmitting, 100BASE-T4 always transmits over three pairs (one dedicated and two bidirectional) while listening for collisions on the remaining pair. It uses a much more sophisticated coding system called 8B6T.

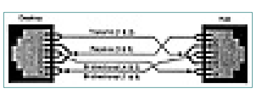


Exhibit 6 100BASE-T4.

Unlike other coding systems that use binary (0, 1) codes, 100BASE-T4 uses ternary (+1, 0, –1) codes, which enable it to pack 8 bits of data into 6 ternary symbols. By using 8B6T coding and three wire pairs for transmission, 100BASE-T4 provides a 100M-bps data-transmission rate with a clock speed of only 25 MHz (8 bits transmitted as 6 ternary symbols over three wire pairs at 25 MHz.)

This process is diagrammed in Exhibit 7: 1 byte (8 bits) of data is encoded into 6 ternary symbols, which are transmitted sequentially across three wire pairs. Unlike 100BASE-TX and 100BASE-FX, 100BASE-T4 does not support full-duplex operation.

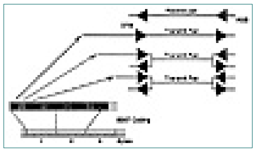


Exhibit 7 100BASE-T4 signaling.

100BASE-T2

100BASE-T2 provides a more robust and noise-resistant signaling system capable of operating over two pairs of Category 3, Category 4, or Category 5 UTP, or over STP links and supporting both half-duplex and full-duplex operation. It uses an extremely sophisticated coding system called PAM5X5, which employs quinary (five-level — +2, +1, 0, –1, –2) signaling. In addition, it uses hybrid circuitry to enable simultaneous bidirectional transmission of 50M-bps data streams over each of the two wire pairs (see Exhibit 8).

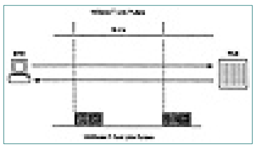


Exhibit 8 Media-independent interface (MII).

Because of its robust encoding, 100BASE-T2 emits less noise during use and is less susceptible to noise from external sources. When used with four-pair Category 5 cable bundles, it can coexist with other signaling systems. A single four-pair bundle can carry two 100BASE-T2 links, one 100BASE-T2 link, and one 10BASE-T link, or one 100BASE-T link and one voice (telephone) link.

Media-Independent Interface (MII)

The Media-Independent Interface is a mechanical interface to the Ethernet MAC, similar to the AUI, which is used to connect transceivers (see Exhibit 9). The MII supports a nibble-wide data path, a station management interface, and command and status registers. It uses a 40-pin connector, similar in appearance to mini-small computer systems interface (mini-SCSI) connectors.

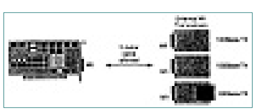


Exhibit 9 100BASE-T auto negotiation (2).

Auto-Negotiation

- Auto-Negotiation provides automatic link testing and configuration for UTP signaling systems. All 100BASE-T systems using UTP or STP go through Auto-Negotiation prior to establishing a link. During this start-up process, 100BASE-T systems on each side of a link:
- Check the link.
- Exchange coded information defining the abilities of each link partner (e.g., 10BASE-T half-duplex operation, 10BASE-T full-duplex operation, 100BASE-TX half-duplex operation, 100BASE-TX full-duplex operation, 100BASE-T2 half-duplex operation,

1000BASE-T2 full-duplex operation or 100BASE-T4 operation).

- Go to an internal lookup table to determine the highest common operation mode.
- Configure themselves as per the table.
- Turn off Auto-Negotiation.
- Open the link.

If one end of the link is a 10BASE-T system that does not support Auto-Negotiation, the partner is automatically configured for 10BASE-T half-duplex operation (default mode). When confronted with another networking technology that uses the RJ-45 connector (e.g., Token Ring), Auto-negotiation will automatically fail the link.

Auto-Negotiation is based on the link pulse used in 10BASE-T. For Auto-Negotiation, the link pulse is divided into 33 fast link pulses that are used to carry pages of coded information between link partners.

Full-Duplex Operation

Full-duplex operation supports simultaneous signaling in both directions over dedicated links by turning off the CSMA/CD collision-detection circuitry. It provides some increase in bandwidth over links that have a high proportion of bidirectional traffic, such as switch-switch and switch-server links. In addition, full-duplex operation increases the maximum length of fiber links. Whereas a half-duplex link is limited to 412 meters by the need to detect collisions, full-duplex operation supports links of up to 2 kilometers because no collision detection is required. This increased link length is only useful for fiber links, signal attenuation limits, and copper link length to 100 meters for both half- and full-duplex operation.

Previous | Table of Contents | Next

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Flow Control

Flow control provides a method for controlling traffic flows between intermediate devices (primarily switches and routers) and between intermediate devices and servers to avoid dropping packets. Currently, two-speed (10/100 or 100/1000) operation requires large buffers to reduce the probability of dropping packets when a continuous stream of packets is sent from a high-speed to a low-speed device (e.g., 100M bps to 10M bps, or 1000M bps to 100M bps). In such a scenario, when the buffers fill, the intermediate device drops the unbuffered packets.

Flow control provides a management alternative to having large buffers. When a buffer approaches full, the receiving device can send a flow control packet back to the sending device to stop the incoming packet stream. When the buffers of the receiving device empty, packet transmission starts again. This eliminates dropped packets and allows manufacturers to build switches with smaller buffers, which reduces costs.

Repeaters and Repeater Connections

Repeaters provide for shared media operation in 10BASE-T and 100BASE-T via the CSMA/CD protocol. 10BASE-T networks have a collision domain diameter of 1000 meters. This permits building large, single-collision domain networks using hierarchical, cascaded repeating hubs to increase port density. 100BASE-T does not permit hierarchical cascading of hubs because the maximum collision domain for UTP is slightly more than 200 meters (see Exhibit 10).

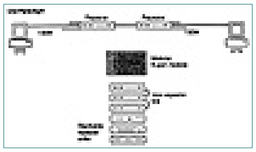


Exhibit 10 Repeater connection styles.

Two techniques can be used to build large, single-collision domain networks (i.e., increase port density). One technique is to use modular hubs, where ports can be added by inserting additional multiport cards into the hub chassis. A second is to use stackable hubs — standalone repeaters that can be connected via high-bandwidth stacking ports that do not impact the collision domain.

Topology Rules

Topology rules for half-duplex 100BASE-T networks are shown in Exhibit 11. Copper links are limited to 100 meters by the U.S. cabling standard EIA/TIA-568-A. A collision domain containing two copper links can contain one class I repeater and two 100-meter copper links; or two class II repeaters, two 100-meter copper links, and a 5-meter copper inter-repeater link. A collision domain containing a class I repeater with two fiber links can support two fiber links of 136 meters, for a collision domain diameter of 272 meters. A collision domain containing a class I repeater can also support one copper link of 100 meters and a single fiber link of 160 meters.

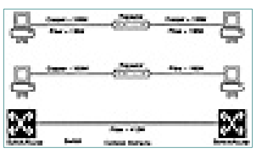


Exhibit 11 100BASE-T topologies.

A fiber DTE-DTE half-duplex collision domain (e.g., a switch-to-switch or switch-to-server) can support a 412-meter fiber link. Links of up to 2 kilometers can be supported over fiber by operating in full-duplex mode, which turns off the CSMA/CD portion of the protocol and requires a dedicated link (see Exhibit 12).

Exhibit 12 Full duplex.

Gigabit Ethernet

Work to extend the Ethernet family to 1000M-bps (gigabit) operation is well underway. The first products using the new technology were demonstrated at Networld + Interop Las Vegas in May 1997, and the first products started shipping during the summer of 1997. Initial products will support operation over 62.5-micron multimode fiber (1000BASE-SX), 50-micron single-mode fiber (1000BASE-LX), or short lengths (to 25 meters) of coaxial cable (1000BASE-CX). The operation of these products is being defined in a supplement to the IEEE 802.3 standard entitled 802.3z

It was scheduled for completion in early 1998. A second supplement, entitled 802.3ab, will define gigabit Ethernet over 100-meter, four-pair Category 5 copper links (1000BASE-T). It is scheduled for completion in late 1998.

IMPACTS ON NETWORK DESIGN

Fast Ethernet is a family of 100M-bps signaling systems for use with the standard Ethernet MAC layer. The family consists of four signaling systems (100BASE-TX, 100BASE-T4, 100BASE-T2, and 100BASE-FX) and technologies that support automatic startup (Auto-Negotiation), shared media operation (Repeaters), full-duplex operation, and flow-control to manage traffic flow.

Fast Ethernet devices work seamlessly with legacy Ethernet systems. They have the same MAC layer, the same frame format, and the same CSMA/CD protocol for shared media operation. Auto-Negotiation ensures that all 100BASE-T devices operating over copper links automatically configure themselves to operate with link partners. This makes 100BASE-T a very economical technology for adding high-bandwidth links to legacy systems.

Higher-speed operation reduces the diameter of 100BASE-T shared-media collision domains to approximately 200 meters for copper. Collision domains can be extended through the use of fiber and connected via switches to build large, complex networks. Full-duplex operation improves bandwidth for bidirectional links and increases the maximum length of fiber links to kilometers. Port density within a single collision domain is expanded through the use of modular and stackable hubs.

100M bps is not the endpoint for Ethernet. 1000M-bps (gigabit) devices were demonstrated in the spring of 1997 and began shipping in the summer of 1997. Targets for gigabit Ethernet operation are 700 meters for full-duplex single mode fiber links, 25 meters for short-haul copper coax links, and 100 meters for Category 5 copper links.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter IV-4
Voice and Video on the LAN

Martin Taylor

Most desktops in enterprises today are equipped with two network connections: a LAN connection to the PC or workstation for data communications, and a phone connection to the PBX for voice communications. The LAN and the PBX exist as two separate networks with little or no connectivity between them. Each has evolved to meet the very specific and differing needs of data and voice communications, respectively.

Despite much talk in the industry about the convergence of computers and communications, LANs and PBXs have not really moved any closer together during the last decade. In the mid-1980s, some PBX vendors sought to bring data services to the desktop via ISDN technology, but the advent of PCs requiring far more than 64K-bps communications bandwidth favored the emerging LAN standards of Ethernet and Token Ring. So far, most LAN vendors have not attempted to support voice communications on the LAN. But all this is about to change.

There are three key factors at work today that suggest that voice and data convergence in the LAN is about to become a hot topic in the industry:

1. widespread acceptance of advanced LAN-switching technologies, including ATM, which makes it possible for the first time to deliver reliable, high-quality, low-delay voice transmissions over the LAN
2. emergence of the first standard for LAN-based videoconferencing

and voice telephony, H.323, which removes objections about the use of proprietary protocols for voice and video over the LAN

3. deployment of the latest generation of Intel processors, featuring MMX technology, which makes high-quality software-based, real-time voice and video processing feasible for the first time, and the new PC hardware architectures with Universal Serial Bus that permit voice and video peripherals to be attached without additional hardware inside the PC

This chapter first looks at the business reasons for considering the deployment of voice and video over the LAN, and then discusses the technical issues and requirements.

THE VALUE OF VOICE AND VIDEO ON THE LAN

There are essentially two main kinds of motivation for considering voice and/or video on the LAN: the need to support new types of application that involve real-time communications, and the desire to improve the overall cost-effectiveness of the local communications infrastructure.

New Types of Applications

Desktop videoconferencing, real-time multimedia collaboration, and video-based training are all examples of new kinds of applications that can benefit from the delivery of voice and video over the LAN.

The uptake of desktop videoconferencing has been held back by a combination of high costs and the difficulty of delivering appropriate network services to the desktop. Standards-based H.320 desktop videoconferencing systems require costly video compression and ISDN interface hardware, as well as the provision of new ISDN connections at the desktop alongside the LAN and the phone system. To lower cost and simplify deployment dramatically, new systems based on the H.323 standard and designed to run over the LAN will leverage the processing power of the latest PCs and the existing switched LAN infrastructure.

Desktop videoconferencing can be used either to support internal meetings and discussions between groups located at remote sites, or to support direct interaction with customers and clients. For example, some enterprises in the mortgage-lending business use videoconferencing to conduct mortgage-approval interviews with potential borrowers, so as to greatly reduce the overall time to complete a mortgage sale.

Real-time collaboration applications, involving any mix of video and voice with data conferencing to support application sharing and interactive whiteboarding, provide a new way for individuals and small groups to collaborate and work together remotely in real time. This emerging class of applications, typified by Microsoft NetMeeting, is being evaluated by many enterprises, particularly for help desk applications.

By contrast, video-based training is already widely used in enterprise LANs. By delivering self-paced video learning materials to the desktop, training needs can be met in a more timely and less disruptive fashion than traditional classroom methods.

The growing popularity of these kinds of applications should be noted by network planners and designers. To obtain the business benefits with the least possible delay, a preplanned strategy for local LAN upgrades to support voice and video will reduce the lead time for the deployment of these applications and enable the enterprise to move swiftly when the application need has been identified.

Infrastructure Efficiencies

A single local communications infrastructure based on a LAN that handles data, voice, and video has the potential of costing less to own and operate than separate PBX and data-only LAN infrastructures.

The average capital cost of a fully featured PBX for large enterprises is between $700 and $750 per user, according to a leading U.S. telecommunications consultancy, TEQConsult Group. Furthermore, this is expected to rise slightly over the next few years as users demand more sophisticated features from their phone system. It is not difficult to see how a switched LAN that has been enhanced to handle voice could provide a solution for telephony at a fraction of this cost.

Most large PBX installations are equipped with additional facilities such as voice-mail and Interactive Voice Response systems for auto-attendant operation. These systems are typically connected directly to the PBX via proprietary interfaces, and they too represent major capital investments. With voice on the LAN, such voice-processing applications could be based on open server platforms and leverage the low-cost processing power and disk storage that is a feature of today's PC server market, thereby lowering the system's capital cost still further.

Separate PBX and LAN infrastructures each incur their own management and operational costs. For example, moves, adds, and changes require separate actions to patch physical LAN and voice connections, and to update LAN logon and voice directories. With telephony provided over a voice-enabled LAN supporting combined directory services, the management effort required to administer moves and changes would be substantially reduced.

These cost-of-ownership benefits come with a raft of usability improvements for telephony. The PC (with phone handset attached) becomes the communications terminal for making and receiving phone calls, and the processing power and graphical user interface of the PC can be leveraged to provide point-and-click call launch and manipulation. Features of PBXs such as call transfer, divert, and hold, which are hard to invoke from a phone keypad, become very easy to use from a Windows interface.

Incoming callers can be identified on the PC display by matching Calling Line Identifier with directory entries. And with voice mail and e-mail supported on a unified messaging platform such as Microsoft Exchange or Lotus Notes, all messages are accessible and manageable via a single user interface.

These usability benefits for voice telephony over the LAN extend also to videoconferencing — a single consistent user interface can be applied to both video and voice-only calls.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

LAN TECHNOLOGIES FOR INTEGRATED VOICE AND VIDEO

The LAN technologies in widespread use today — Ethernet, Fast Ethernet, FDDI, and Token Ring — were not designed with the needs of real-time voice and video in mind. These LAN technologies provide "best-effort" delivery of data packets, but offer no guarantees about how long delivery will take. Interactive real-time voice and video communications over the LAN require the delivery of a steady stream of packets with very low end-to-end delay, and this cannot generally be achieved with the current LAN technologies as they stand.

Asynchronous Transfer Mode (ATM)

At one time, there was believed that ATM networking to the desktop would be embraced by LAN users to solve this problem. ATM is a networking technology that was designed specifically to handle a combination of the low-delay, steady-stream characteristics of voice and video and the bursty, intermittent characteristics of data communications.

The ATM Forum, the industry body responsible for publishing ATM specifications, has developed a number of standards that enable desktops connected directly to ATM networks to support existing LAN data applications as well as voice telephony and videoconferencing. The ATM Forum standards for the support of voice and video over ATM to the desktop typically avoid the use of traditional LAN protocols such as IP, and instead place the voice or video streams directly over the ATM protocols.

While it is clear that ATM to the desktop provides an elegant and effective solution for combining voice, video, and data over the LAN, this approach does imply a “forklift” to the LAN infrastructure and the end station connection. The cost and disruptive impact of such an upgrade tend to limit its appeal, and as a result desktop ATM is not expected to be widely adopted.

However, the ability of ATM to provide Quality of Service, — that is, to deliver real-time voice or video streams with a guaranteed upper bound on delay — makes ATM an excellent choice for the LAN backbone where voice and video over the LAN is needed.

Shared and Switched LANs

It is generally accepted that shared LANs are unsuitable for handling real-time voice and video because of the widely varying delays observed when multiple stations are contending for access to the transmission medium. The CSMA/CD access method used in shared Ethernet is particularly poor in this respect. Token Ring, on the other hand, is based on a token-passing access method with multiple levels of priority. Stations waiting to send data packets can be pre-empted by other stations on the ring with higher priority voice or video packets to send. As a result, Token Ring has excellent potential to handle real-time voice and video traffic, although this potential has yet to be realized in currently available networking products.

LAN switching does much to overcome the limitations of shared LANs, although today’s products are still a long way from providing an answer for voice and video over the LAN. It is now cost-effective to provide users with dedicated 10M-bps Ethernet connections to the desktop, and 100M-bps Fast Ethernet uplinks from the wiring closet to the backbone.

However, despite the vast increase in bandwidth provision per user that this represents over and above a shared LAN scenario, there is still contention in the network leading to unacceptable delay characteristics. For example, multiple users connected to the switch may demand file transfers from several servers connected via 100M-bps Fast Ethernet to the backbone. Each server may send a burst of packets that temporarily overwhelms the Fast Ethernet uplink to the wiring closet. A queue will form in the backbone switch that is driving this link, and any voice or video packets being sent to the same wiring closet will have to wait their turn behind the data packets in this queue. The resultant delays will compromise the perceived quality of the voice or video transmission.

The only way to overcome this problem is to find a way of treating real-time voice and video packets differently from data packets in the network, and to give them preferential treatment when transient data overloads cause queues to form on busy network links. In practice, this means that LAN packets must be tagged with some kind of priority information that enables switches to identify which packets need to jump the queue.

The IEEE 802, which oversees standards for LAN technologies, has initiated a project identified as 802.1p, which is concerned with Traffic Class Expediting in LAN switches.

The principal problem faced by 802.1p is that there is no spare information field in the standard Ethernet packet format that could carry the required priority tag. As a result, it has been necessary to propose a new Ethernet packet format with an additional 4 bytes of information in the packet header that can contain a 3-bit priority tag field (offering 8 levels of priority), together with some other information concerned with Virtual LANs.

With the new Ethernet packet format containing a priority tag, end-station applications can identify real-time voice or video packets by assigning them a high priority value in the tag. LAN switches that have been enhanced to process the priority tags can separate high- and low-priority traffic in the switching fabric and place them in separate queues at outgoing switch ports. The LAN switches need to implement a queue-scheduling algorithm that gives preference to the higher-priority queues on outgoing ports, and by this means it is hoped that real-time voice and video can be carried over the LAN without incurring unacceptable delays during periods of heavy data traffic.

As of July 1997, the 802.1p standard was still in draft form and the standard is not expected to be completed until 1998. Ethernet switches that support the 802.1p priority tags with multiple internal queuing structures will require a new generation of switching silicon, and the earliest we could expect to see products that conform to the standard would be late 1998 or into 1999. Surprisingly, we may see Token Ring switches that handle multiple priority levels before that time, leveraging the capabilities of the existing Token Ring standard that supports 8 levels of priority.

Previous | Table of Contents | Next

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Hybrid ATM Networks

The discussion of ATM described how it offers guaranteed Quality of Service for real-time voice and video streams. Today, ATM is increasingly used as a LAN backbone for pure data applications because it offers greater scalability and fault tolerance than other LAN technologies. Ethernet and Token Ring LANs are connected to ATM via “edge switches” equipped with ATM uplinks, typically supporting the ATM Forum standard for carrying LAN traffic over ATM, know as LAN Emulation.

It is possible to enhance ATM edge switches to enable desktops connected via Ethernet or Token Ring to enjoy the benefits of ATM Quality of Service across the LAN backbone. Two techniques have been proposed to achieve this.

The first technique, known as “Cell-in-Frame,” extends the native ATM signaling protocols over dedicated Ethernet connections from the edge switch to the end station. The voice or video application in the end station places the voice or video stream in ATM cells using the ATM Forum standards for native ATM transport, and then encapsulates the ATM cells in Ethernet packets for transport to the edge switch for onward transmission onto the ATM network. Effectively, this is ATM to the desktop, but using physical Ethernet with standard Ethernet adapter cards as a kind of physical transport layer for ATM traffic.

The second technique makes use of an emerging standard protocol for end stations to request Quality of Service for IP-based voice or video applications, known as the Resource Reservation Protocol, or RSVP. The enhanced edge switch intercepts RSVP requests originated by end stations and converts them into ATM signaling to request the setup of connections across the ATM backbone with the appropriate Quality of Service. The edge switch then distinguishes between IP packets containing data and those containing voice or video, using the information provided by RSVP, and steers voice and video packets onto ATM connections that have Quality of Service.

At the time of writing, the technique described here for RSVP-to-ATM mapping enjoys somewhat broader industry support than Cell-in-Frame, perhaps because of its relationship with Internet technology.

Until LAN switches supporting 802.1p priority tagging have proven themselves capable of meeting the very stringent end-to-end delay requirements for real-time voice and video communications, hybrid approaches based on ATM in the backbone and switched Ethernet or Token Ring to the desktop are likely to find acceptance as the solution of choice for voice and video over the LAN.

Standards for LAN-Based Voice and Video Applications

Standards for voice and video over the LAN fall into two categories: those designed for native ATM protocols, and those intended for general-purpose LAN protocols, particularly IP.

Standards for native ATM protocols, such as the ATM Forum's Voice Telephony over ATM (VTOA) are appropriate only for ATM-connected desktops, or desktops running Cell-in-Frame over Ethernet.

Standards for applications that run over IP are applicable both to ATM-connected desktops and desktops in general Ethernet or Token Ring environments. The most important standard in this space is H.323, which was developed by the International Telecommunications Union. While H.323 is designed to be independent of the underlying networking protocol, it will most often be deployed running over IP.

H.323 references other existing standards for the digital encoding and compression of voice and video signals, and describes how audio and video streams are carried in the payload of IP packets with the aid of the Real Time Protocol (RTP), which provides timing and synchronization information. H.323 also covers the handling of data streams for application sharing, shared whiteboarding, and real-time file transfer (referencing the T.120 standard), and includes signaling based on ISDN messaging protocols for call setup and teardown.

The H.323 standard is flexible and accommodates any combination of real-time voice, video, and data as part of a single point-to-point or multipoint conference call. It can be used with a voice stream alone as the basis of a LAN telephony solution. H.323 enjoys the broadest support in the industry as a proposed standard for Internet telephony.

Additional Components: Gateways and Gatekeepers

Creating a LAN infrastructure that can consistently deliver voice and video streams with sufficiently low delay is an absolute prerequisite for integrating voice and video on the LAN, but it is by no means the complete answer to the problem. There are two other key components of a complete voice and video solution, which in H.323 parlance are known as the *gateway* and the *gatekeeper*.

An H.323 gateway provides interconnection between voice and video services on the LAN, and external voice and video services typically provided over circuit-switched networks such as ISDN and the public telephone network. The gateway terminates the IP and RTP protocols carrying the voice and video streams, and converts them to appropriate formats for external networks. For videoconferencing, the conversion is most likely to be to H.320, another ITU standard that specifies how voice and video are carried over ISDN connections. For voice-only connections, the conversion will be to the G.711 standard for digital telephony. This allows voice interworking with any phone on a public network or connected to a PBX.

An H.323 gatekeeper is a pure software function that provides central call-control services. While it is possible to run H.323 voice and video communications over the LAN without a gatekeeper, in practice this function is extremely useful. At the most basic level, the gatekeeper provides directory services and policy-based controls applied to the use of voice and video communications. For example, the gatekeeper can bar stations from accessing certain types of external phone numbers at certain times of day. The gatekeeper can be thought of as the “server” in a client/server model of LAN-based telephony and videoconferencing.

At a more sophisticated level, the gatekeeper may be able to support supplementary services, including call transfer, hold and divert, hunt groups, pick-up groups, attendant operation, etc. — features that are typically found in high-end PBXs for controlling and managing voice calls. While the H.323 standard does not explicitly describe how supplementary call control features can be supported, the standard does provide a framework for the addition of these advanced capabilities.

CONCLUSION

This chapter has explained the value of voice and video integration on the LAN in terms of both application-driven needs and the desire for infrastructure efficiencies. It has looked at the technology issues surrounding the transport of real-time voice and video streams over LAN infrastructures, and concluded that ATM backbones provide a solution in the near term, with the possibility later of a solution based entirely on switched Ethernet or Token Ring.

Finally, some additional functional elements have been described, such as gateways and gatekeepers, that are an essential part of a complete solution for voice and video over the LAN. Over the last decade, the open standards-based environment typified by PCs and LANs has revolutionized the way data is handled and processed in enterprise environments. Now, this open and standards-based approach is set to tackle the challenge of voice and video, formerly the exclusive domain of the PBX. The history of LAN evolution is set to repeat itself, and one can expect the traditional proprietary mainframe PBX to diminish in importance to the enterprise, giving way to client/server telephony and videoconferencing, just as the mainframe computer has been pushed into the background by client/server techniques for data processing.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Bookmark It

Search this book:

Go!

Previous | Table of Contents | Next

Chapter IV-5
Computer Telephony: An Evolving Technology

Gilbert Held

The most recognized computer telephony application is voice mail. Through the integration of a computer and a private branch exchange (PBX), calls can be automatically answered after a predefined number of rings, forwarded to predefined telephone stations, and even digitized, stored, and forwarded to another corporate location for retrieval by a remote user. Although the primary effort of computer telephony focused on the integration of computer-based systems to support voice-mail applications, the emergence of facsimile as a popular mechanism for distributing information resulted in the development of fax-back systems and local area network (LAN)-based fax servers.

Computer telephony has expanded to take advantage of the distance-insensitive pricing associated with the Internet. In fact, two rapidly emerging technologies that have the highest level of growth of all Internet applications are the transmission of digitized voice via the Internet and its use as a mechanism to distribute faxes on a worldwide basis. The latter mechanism eliminates a majority of the cost associated with the public switched telephone network (PSTN), which was previously used almost exclusively for those tasks.

In the PC desktop area, the introduction of Novell and Microsoft telephony application programming interfaces permits NetWare and Windows NT-based computer systems to operate programs that manage a user’s voice, fax, and electronic mail messages.

Because of the emerging role of computer telephony and its migration into the organizational LAN environment, network managers must investigate potential network architectures that may be required to support this technology.

USING THE INTERNET TO SUPPORT COMPUTER TELEPHONY

Use of the Internet to support computer telephony holds potential promise for significant economic savings, although performance and security-related problems will need to be addressed to make it a viable communications mechanism. The rationale for using the Internet to transport voice and fax is primarily one of economics because many, if not most, organizations are already connected. Because almost all Internet service providers (ISPs) charge a flat monthly connection fee based on the operating rate of the connection and not the activity on the line, the addition of voice and fax transmission between locations connected to the Internet can be accomplished for the one-time cost associated with acquiring hardware.

For example, MICOM Communications of Simi Valley, CA recently introduced a communications server that supports voice and fax over the Internet Protocol (IP). Because the server can be directly connected to a LAN running IP, which can in turn be connected via a router to the Internet, transporting digitized voice and fax over the Internet is simply a matter of configuring hardware and software.

Exhibit 1 illustrates the use of a MICOM communications server and its interconnection to a PBX and LAN to transmit and receive voice, data, and faxes via the Internet. As indicated in Exhibit 1, workstations connected to a bus-based Ethernet LAN can communicate via the router to transmit and receive Internet-related communications. Through the use of the MICOM communications server, voice conversations initiated through a PBX can be routed to the server and digitized for transmission via the Transmission Control Protocol/Internet Protocol (TCP/IP). The MICOM communications server contains circuitry that accepts a normal pulse code modulated (PCM) 64K-bps digital data stream produced by a PBX that can be compressed into an 8K-bps data stream through the use of adaptive differential pulse code modulation (ADPCM) and idle time compression.

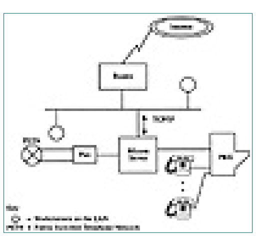


Exhibit 1 Using a MICOM voice-over-IP communications server.

ADPCM is a predictive compression technique standardized by the International Telecommunications Union-Telecommunications Standardization Sector (ITU-T) body. Through the use of ADPCM, the 64K-bps PCM data stream can be reduced to 16K bps. Because no human voice is continuous, a high proportion of idle time exists in human speech that can be replaced by codes indicating the gap in time when there is no talking. By suppressing idle time, MICOM is able to further reduce the digitized voice data stream to 8K bps. Unfortunately, there are no standards governing the suppression of idle times in speech, resulting in the requirement of another MICOM communications server installed at the remote location to take advantage of the idle-tim-suppression capability of the server.

Besides digitizing and transporting voice via TCP/IP, the MICOM voice-over-IP server can also be used to transport faxes. The server can accept ITU-T Group 3 fax connections, providing the capability to route faxes received via the PSTN to a predefined IP network address. To support an outdial capability, a user can transmit documents to the fax connected to the communications server and enter two numbers. The first number would represent another MICOM communications server; the second would represent the telephone number the server would outdial on receipt of the fax.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Search this book:

Go!

Previous | Table of Contents | Next

ECONOMICS OF TELEPHONY OVER THE INTERNET

A MICOM communications server capable of supporting four simultaneous digitized telephone conversations via a TCP/IP network can be obtained for less than $10,000. Thus, an investment of approximately $20,000 enables two sites to communicate via the Internet. Assuming an organization already has some available bandwidth on its Internet connections to accommodate at least one 8K-bps digitized voice connection along with existing data communications, the potential payback can be computed relatively easily.

A prime-time, long-distance call within the continental U.S. currently averages 15¢ per minute. If an organization can use the Internet to replace a series of single calls that average one hour per day between two locations, it would save $9 per day in long-distance telephone charges. On the basis of an average of 22 workdays per month, the organization would save $198 per month, or $2,376 yearly. Because two communications servers require an investment of approximately $20,000, this would result in a rate of return of ($2376/20,000) * 100, or approximately 12% on the company's equipment investment. Although this is a reasonable rate of return, it is far from the 20 to 30% network managers and administrators are accustomed to.

Suppose an organization uses the Internet to connect two locations in different countries, such as the U.S. and England. During the business day, an international call between those countries typically costs 70¢ per minute. Thus, a pair of communications servers used to replace one hour per day of long-distance telephone traffic would reduce the cost of international voice communications by $42 per day, or $924 monthly, again assuming 22 business days per month. On a yearly basis, the organization would save $11,088, which would provide a rate of return of ($11,088/20,000) * 100, or approximately 55% on its equipment investment.

Thus, the use of the Internet as a mechanism to replace the PSTN network can provide a reasonable, and in some cases outstanding, rate of return on equipment investments. Once network managers appreciate the economics associated with using the Internet as a mechanism for transporting voice and data, they must focus on two problem areas: predictability and security.

UNPREDICTABLE TRANSMISSION

Currently, no mechanism guarantees bandwidth availability when transmitting information on the Internet. This means that a large file transfer, a series of interactive query responses, or similar activities that occur between Internet network access locations can have an adverse effect on the ability of real-time voice to be heard clearly. Of course, if network managers use the Internet to transmit voice mail or faxes, short random-time delays will not adversely affect the receipt of such applications.

Some users have implemented a partial solution to the network predictability problem with a degree of success — they overscribe their local loops, installing a T1 connection if a fractional T1 connection is needed. This solution usually removes bottlenecks at the ISP point of connection; however, it does not affect the flow of data through the Internet and does not tackle the heart of the problem, which is the unpredictable transmission on the Internet.

Fortunately, a new protocol known as RSVP (ReSerVation Protocol) is being developed to provide guaranteed bandwidth through the Internet. Unfortunately, RSVP is several years away from implementation, meaning that the ability of the Internet to transport real-time voice will continue to be randomly affected by other network activity.

Until RSVP is available, the ability to use the Internet to transport voice for business applications will rely primarily on an organization’s Internet connection rate and the method of voice digitization used. The higher the digitization compression ratio, the lower the resulting operating rate required to transport digitized voice. Thus, highly compressed voice that requires less bandwidth to transport will have a lower probability of being affected by other network traffic than lesser compressed voice that requires a higher bandwidth to transport. Communications equipment, such as the MICOM communications server, that takes advantage of gaps in human speech to encode voice at a relatively low data rate can move digitized voice through the Internet without adverse delays affecting regenerated speech faster than other products that result in higher voice digitization rates.

SECURITY

As most users are well aware, the Internet appears to be a breeding ground for hackers and crackers. Thus, the network configuration previously illustrated in Exhibit 1 should be modified by the addition of a firewall to protect the organizational LAN from possible intruders.

Recognizing the limitations associated with the unpredictability and security issues of connecting networks to the Internet, many companies are deciding not to do so. Many organizations are constructing corporate intranets based on the use of Web browsers, Telnet servers, and the TCP/IP protocol stack. The network configuration shown in Exhibit 1 can also be considered for implementation on the corporate intranet. Doing so helps avoid bandwidth predictability problems associated with the use of the Internet because network managers and administrators have far better control over bandwidth on an internal intranet.

In addition, if the intranet is not connected to the Internet, an organization can avoid the cost and administration effort required to manage firewalls. Thus, the use of a corporate intranet for voice, fax, and data transmission can be expected to be a more viable mechanism for the application of computer telephony to LANs than the Internet.

LAN-BASED COMPUTER TELEPHONY INTEGRATION

Today, almost all organizations that operate LANs have two separate wiring architectures, one for the LAN and a second to support desktop telephones. Exhibit 2 illustrates this conventional corporate communications architecture. Not only is the establishment of a dual wiring infrastructure expensive, but it also requires a corporate PBX that may or may not support voice mail and fax server functions.

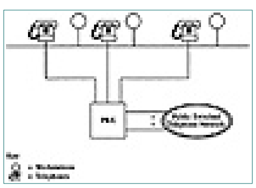


Exhibit 2 Conventional organizational communications architecture.

Using a Communications Server

Over the past two years, several hardware vendors introduced Novell and Windows NT-based communications servers that provide a variety of corporate voice communications functions, including the ability to provide partial or total integration into a single wiring infrastructure. Some products, such as PhoneNet, from the company of the same name located in Hertzalia, Israel, permit voice to be transported on a lightly loaded Ethernet segment, enabling a microphone, sound card, and speakers attached to a LAN-based PC to replace separate telephones and their wiring to a PBX.

In addition to enabling a single wiring infrastructure to support voice, data, and fax, PhoneNet and similar systems include voice-mail facilities and a fax-server capability that allows word processing files to be transmitted as faxes. When used as a partial integration mechanism, the communications server does not replace individual telephone stations. Instead, the server is connected to the LAN to provide a direct fax capability to individual workstations while each telephone continues to be connected to the communications server. Exhibit 3 illustrates the partial integration of a computer-based telephone into a corporate LAN.

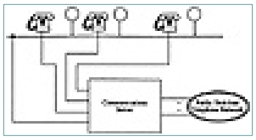


Exhibit 3 Partial integration of computer-based telephone onto an organizational LAN.

CONCLUSION

With more than 700 million telephones and 100 million fax machines installed worldwide, the integration of voice and fax into LAN-based networks via computer telephony is a viable method of reducing an expensive multiple wiring infrastructure. In addition, because many organizations already interconnect their geographically separated LANs via internal networks, corporate intranets, or the Internet, any mechanism to move voice and fax transmission off the PSTN and onto existing digital networks may produce significant cost savings.

Network managers and LAN administrators should carefully review their public and private network use to determine whether it is economically feasible to migrate voice and fax to TCP/IP-based networks. If so, they should weigh the economic savings against potential performance and security problems to determine whether to launch the migration effort.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter IV-6
Virtual Networking Management and Planning

Trenton Waterhouse

The switch could be considered a third-generation internetworking device. First-generation devices, or bridges, offered a high degree of performance throughput but relatively little value, because the bridge's limited decision intelligence resulted in broadcast storms that produced network instability. Routers, the second generation of internetworking devices, increased network reliability and offered great value with fire-walling capabilities, but the trade off was in performance. When routers are used in combination with each other, bandwidth suffers, which is detrimental for delay-sensitive applications such as multimedia.

The switched virtual network offers all the performance of the bridge with the value of the router. The constraints of physical networking are removed by the logical intelligence that structures and enforces policies of operation to ensure stability and security. Regardless of access technology or geographic location, any-to-any communications is the goal.

THE BUSINESS CASE FOR VIRTUAL NETWORKING

Both the business manager and the technical manager should find interest in this new virtual networking scheme. The business manager is usually interested in cost-of-ownership issues. Numerous studies from organizations such as the Gartner Group and Forrester Research have indicated that only 20% of networking costs are associated with capital equipment acquisition. The other 80% of annual budgets are dedicated to items such as wide area networking charges, personnel, training, maintenance and vendor support, as well as the traditional equipment moves, adds, and changes.

It is important for network planners to remember that capital expenditure happens in year one, even though the equipment may be operating for another four years. Wide area network (WAN) charges can account for up to 40% of an organization's networking budget. For every dollar that the technical staff spends on new equipment, another four dollars is spent on the operation of that equipment. Therefore, focus should be on the cost-of-ownership issues, not necessarily the cost of the network devices.

Network Reliability. Business managers are also looking for increased reliability as the network plays a major role in the core operations of the organization. Networks have become a business tool to gain competitive advantage-they are mission critical and, much like a utility, must provide a highly reliable and available means of communications. Every office today includes an electrical outlet, a phone jack, and a network connection. Electrical and phone service are generally regarded as stable utilities that can be relied on daily. Networks, however, do not always provide such levels of service.

Network Accountability. Managers also can benefit from the increased accountability that virtual networks are able to offer. Organizational networking budgets can range from hundreds of thousands of dollars to hundreds of millions or even billions per year. Accounting for the use of the network that consumes those funds is a critical issue. There is no better example than WAN access charges. Remote site connectivity can consume a great deal of the budget, and the questions of who, what, when, and where in regards to network use are impossible to determine. Most users consider the network to be free, but the tools to manage and account for its use are increasingly a requirement, not an option.

THE TECHNOLOGY CASE FOR VIRTUAL NETWORKING

The technical manager's needs for higher capacity, greater performance, and increased efficiency can be met through the deployment of switched virtual networks. Each user is offered dedicated bandwidth to the desktop with uplinks of increasing bandwidth to servers or other enterprise networks. Rather than contending for bandwidth in shared access environments, all users are provided with their own private link. This degree of privacy allows for increased security because data is sent only to intended recipients, rather than seen by all.

The most attractive feature to the technical manager, however, may be the benefits gained through increased ease of operation and administration of virtual networks. A long-standing objective has been to deliver network services to users without continually having to reconfigure the devices that make up that network, furthermore, many of the costs associated with moves, adds, and changes of users can be alleviated as the constraints of physical networking are removed. Regardless of user location, they can remain part of the same virtual network. Through the use of graphical tools, users are added and deleted from work groups. In the same manner, policies of operation and security filters can be applied. In a sense, the virtual network accomplishes the goal of managing the individual users and individual conversations, rather than the devices that make up the network.

VIRTUAL NETWORKING DEFINED

From the user's perspective, a virtual network is a data communications system that provides access control and network configuration changes using software control. It functions like a traditional network but is built using switches.

The ideal virtual network does not restrict access to a particular topology or protocol. A virtual network that can only support Ethernet users with transmission control protocol /Internet protocol (TCP/IP) applications is limited. The ultimate virtual network allows any-to-any connectivity between Ethernet, Token Ring, fiber distributed data interface (FDDI), asynchronous transfer mode (ATM), Internet protocol (IP), Internetwork Packet Exchange (IPX), AppleTalk, or Systems Network Architecture (SNA) networks. A single virtual network infrastructure under a single management architecture is the goal.

Network management software becomes a key enabling requirement for the construction of switched virtual networks. The greatest challenge network designers face is the separation of the physical network connectivity from the logical connection services it can provide. Many of the design issues associated with networks can be attributed to the physical parameters of protocols and the routers used as the interconnection device. A challenge for any network manager is to remain compatible with existing layer 3 protocols and routers and still preserve the investment in existing local area network (LAN) equipment to the greatest extent possible.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Using Telephony as a Model

The principles of operation for switched virtual networks are concretely founded in the success of the global communications systems. Without doubt, the phone system is the world’s largest and most reliable network. Built using advanced digital switches controlled by software, extensive accounting and management tools ensure the success of this highly effective means of communication. The connection-oriented switch is the key. End-to-end connections across multiple switches and various transmission types ranging from copper to fiber optics to microwave to satellites allow millions of calls per day to be successfully completed, regardless of the type of phone or where the user is calling from. The telephony model is used throughout this article to help illustrate the workings of a virtual network.

SWITCHING DEFINED

One of the more confusing terms in the networking industry today is the word *switch*. For the purpose of this article, switching can be broken down into three fundamental areas:

- Configuration switching.
- Packet switching.
- Cell switching.

The earliest form of switching enabled the network manager to assign an individual port or an entire group of ports to a particular backplane segment within an intelligent hub device. This port configuration switching allowed the logical grouping of users onto a particular segment without the need to physically travel to the wiring closet to move cables or connectors. In a sense, this offers an electronic patch panel function. Although the benefit is a reduction of moves, adds, and change costs, this advantage can only be realized within the confines of a single hub. The application of this type of switching is limited because it cannot extend beyond one intelligent concentrator. Although beneficial in the work group, the enterprise needs cannot be met.

Phone system operators in the 1940s manually patched user connections through to destinations and recorded call time and duration. Using configuration switching is similar to patching phone lines together. Just as the phone network grew at a pace that required the switching to be performed automatically without operator intervention, so too have data networks outgrown the limitations of configuration switching.

Packet switching isolates each port to deliver dedicated bandwidth to each user in the network. Fundamentally, a packet switch is any device that accepts an incoming packet on one port and then makes a decision whether to filter or forward the packet out another interface. There are two types of packet switch transports: connectionless and connection-oriented.

Connectionless Packet Switching

Connectionless devices are probably more familiar to network professionals when described as bridges or routers. A bridge is a layer 2 (of the Open Systems Interconnection [OSI1 reference model) switch that bases its decisions on the media access control (MAC) address of attached workstations. What many vendors describe as a switch is actually a wire-speed MAC layer bridge. Three methods of decision making in these types of devices are cut-through, modified cut-through, and store-and-forward.

The Cut-through Switch. This switch reads a packet to the destination address before it starts forwarding to the outbound interface. The benefit is an extremely low latency or delay in the forwarding of packets. The penalty is the propagation of errors, because the frame is being forwarded before it can be verified as valid, and the inability to support interfaces of different speeds that prevents high-bandwidth uplinks of FDDI or ATM on these type of devices.

The Modified Cut-through Switch. This switch reads the first 64 bytes of a frame and then starts forwarding to the outbound interface, which greatly reduces the chances of propagating errored frames throughout the network. However, this method still requires all ports to be of the same type and speed.

Store and Forward Switch. The most flexible switch design uses a store-and-forward methodology that reads the entire frame before any filtering or forwarding decisions are made, thus ensuring that only packets that are not errored are forwarded on the network. This method also allows packets to be buffered when transferring data between networks of different types, such as Ethernet to FDDI or ATM.

Bridges and Routers. A router is a layer 3 switch that bases its decisions on the network protocol address of attached workstations. Bridges and routers are considered connectionless because they forward and forget, requiring a decision to be made on every single inbound packet. The performance implications are that even though two communicating nodes on opposite sides of a bridge or router may be the only devices on their respective networks, the bridge or router must continuously make filter or forward decisions on every packet sent between the two nodes. If the phone network were built using bridges or routers, users would have to hang up and redial their destination after every word, which is not a very practical proposition.

A connectionless transport is not capable of defining which path its payload will take, cannot guarantee delivery, and is generally slower than a connection-oriented system. When a node sends a packet through a bridged or routed network, it is analogous to dropping a letter into a mailbox. It is not apparent how the letter got to its destination. The arrival of a letter cannot be guaranteed (protocol prioritization techniques are comparable to sending a letter by express mail). If a letter is lost (or a packet dropped), determining where it was lost is often difficult. The only way the sender knows that the letter was received is if the recipient sends another letter back to the sender (i.e., frame acknowledgment).

In a sense, today’s shared-access networks are like the party lines of the early telephone network. But just as the phone network evolved from party lines to dedicated lines as usage and deployment grew, so too must the data networks offer this same level of service guarantee and broad adoption.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Connection-Oriented Switches

The connection-oriented switch that the phone systems use offers immediate acknowledgment of communications when the person picks up at the other end. The exact path the call took as well as its time and duration can be logged. The destination only needs to be dialed only once and information is exchanged until both parties hang up.

The idea of connection-oriented communications is not new. This type of switching provides a high degree of reliability and reduces operational costs. Multiple classes of service can be defined to support voice, video, and data transfer. Excellent bandwidth management through congestion control techniques are possible and security and access control are greatly improved. Connection-oriented switching, along with easy-to-implement policy-based management and accounting facilities, have enabled the phone system to become universally accessible.

Frame relay technology is centered around connection-oriented communications, as is the most promising future networking technology—ATM. ATM is the most desirable networking technology because it offers dedicated, scalable bandwidth solutions for voice, video, and data.

ATM Switching. ATM switching is connection-oriented. Communications in an ATM network can be broken down into three phases: call setup (analogous to dialing a phone), data transfer (talking on the phone), and call teardown (hanging up the phone). The use of fixed-length 53-byte cells for data transfer delivers fixed latency transfer times for constant bit rate applications such as voice and video. ATM addressing schemes are similar to a telephone number. In fact, the original designers of ATM technology had their roots in the telephony arena at BellCore, so many analogies to the operation of the phone system can be made when referring to an ATM network.

Although the benefits of ATM networking are attractive, there are currently nearly 100 million networked personal computers that do not have ATM interfaces. Few organizations can afford to replace all of their existing desktop and server interfaces, not to mention network analyzers and troubleshooting equipment.

Through the preservation of existing interface technology, by merely changing the internetworking devices from being connectionless to connection-oriented, many of the benefits of ATM may be realized without requiring the investment in all new ATM equipment. If LANs were designed to operate using the same principles as ATM, rather than making ATM compatible with LANs, users would benefit without significant capital investments in new equipment. By adding switch technology to the middle of the network, network administrators can be spared the trouble of upgrading numerous user devices, and users can be spared the inconvenience of rewiring and disruptions at their work site during an upgrade.

FEATURES OF SWITCHING SOFTWARE

The software that runs on switches is just as important as the switches themselves. A salesperson from AT&T, Fujitsu, or Northern Telecom does not focus the potential customer on the hardware aspects of the telephone switches. On the contrary, the salesperson conveys the benefits of the call management software, accounting, and automatic call distributor (ACD) functions. Switched virtual networks should also be evaluated for their ability to deliver value because of the software features.

The Virtual Network Server

Network management software has traditionally been thought of as software that passively reports the status and operation of devices in the network. In the switched virtual network, the network management software takes on a new role as an active participant in operations as well as configuration and reporting. A new middleware component known as the virtual network server (VNS) enforces the policies of operation defined by the network administrator through management software applications. The switches provide the data transport for the users of the network.

Directory Service. One of the software features in the VNS is the directory service. The directory service allows the identification of a device by logical name, MAC address, network protocol address, and ATM address, along with the switch and port that the user is connected to within the virtual network domain. The directory listing could be populated manually or dynamically as addresses are discovered. To fully realize the benefits of switched virtual networking, automatic configuration is absolutely essential. The directory service allows end nodes to be located and identified.

Security Service. The VNS security service would be used during call setup phases to determine whether users or groups of users were allowed to connect to each other. On a user-by-user and conversation-by-conversation basis, the network manager would have control. This communications policy management is analogous to call management on a telephone private branch exchange (PBX) where 900 numbers, long-distance, or international calls can be blocked. Users could be grouped together to form policy groups in which rules could be applied to individual users, groups, or even nested groups. Policies could be defined as open or secure, inclusive or exclusive.

A sample default policy can ensure that all communications are specifically defined to the VNS in order to be authorized. Policy groups can be manipulated either through drag-and-drop graphical user interfaces or programatically through simple network management protocol (SNMP) commands.

Finally, and most important, the directory service can work in conjunction with the security service to ensure that policies follow the users as they move throughout the network. This feature alone could save time spent maintaining a router access list, as occurs headaches when a user changes location in the traditional network. However, it is important to realize that switched virtual networks ease administrative chores, they do not eliminate them.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Connection Management Service. The VNS connection management service is used to define the path communications would take through the switch fabric. A site may be linked by a relatively high-speed ATM link and a parallel but relatively low-speed Ethernet link. Network connections with a defined high quality of service QOS) could traverse the ATM link and lower QOS connections could traverse the Ethernet. This connection management service allows for the transparent rerouting of calls in the event of a network fault. Connection management could also provide ongoing network monitoring in which individual user conversations could be tapped or traced for easy troubleshooting.

Bandwidth Service. The VNS bandwidth service is used during the call setup when a connection request is made. Video teleconferencing users may require a committed information rate (CIR) of 10M bps whereas the terminal emulation users may only require 1M bps. This is where ATM end stations and ATM switches negotiate the amount of bandwidth dedicated to a particular virtual circuit using user-to-network interface (UNI) signaling. Ethernet, Token Ring, and FDDI nodes do not recognize UNI signaling, but the switches they attach to could proxy signal for the end station, thus allowing a single bandwidth manager for the entire network, not just the ATM portion.

Broadcast Service. The VNS broadcast service uses as its base the concept of the broadcast unknown server (BUS) that is part of the ATM Forum's LAN emulation draft standard. This is how broadcasts are flooded through the network to remain compatible with the operation of many of today's protocols and network operating systems. A degree of intelligence can be assigned to the VNS that would allow for broadcasts or multicasts based on protocol type or even policy group.

Virtual Routing Service. The VNS virtual routing service is one of the most critical components of a virtual network. Just as traditional networks required traditional routers for interconnection, virtual LANs will require virtual routers for internetworking between virtual LANs. In other words, routing is required, but routers may not be. Some protocols such as TCP/IP actually require a router for users on two different subnetworks to speak with each other. In addition, most networks today are logically divided based on network layer protocol addresses with routers acting as the building block between segments.

The difference in operation between a virtual router and a traditional router goes back to the connection-oriented versus connectionless distinction. Routing allows for address resolution between the layer 3 protocol address and the layer 2 MAC address just as it happens through the address resolution protocol (ARP) process in TCP/IP networks. The VNS virtual routing service performs the address resolution function, but once the end station addresses are resolved, establishes a virtual connection between the two users. Two users separated by a traditional router would always have the router intervening on every single packet because the router would have resolved the protocol addresses to its own MAC address rather than the actual end station's MAC address. This VNS routing service allows the network to route once for connection setup and switch all successive packets.

Accounting Service. The VNS accounting service is beneficial because it allows the creation of the network bill. Similar to the way a telephone bill is broken down, the accounting service details connection duration with date and time stamp along with bandwidth consumption details. This is most directly applicable in the WAN. For many network managers, WAN usage is never really accounted for on an individual user basis, yet it can consume up to 40% of the operations budget.

As usage-based WAN service options such as integrated services digital network (ISDN) gain popularity, accounting becomes that much more critical. Interexchange carriers (IXCs), competitive access providers, and the regional Bell operating companies (RBOCs) continue to deliver higher-bandwidth links with usage-based tariffs. In the future, they could install a 155M-bps synchronous optical network (SONET) OC3 link and only charge for the actual bandwidth used. Unless network managers have tools to control access to and account for usage of WAN links, WAN costs will continue to rise. This service lets network managers know who is using the WAN.

VIRTUAL NETWORKS VERSUS VIRTUAL LANS

Throughout this discussion, words have been carefully chosen to describe the operation of switched virtual networks. Many of the current vendor offerings on the market have as their goal the construction of a switched virtual LAN. These virtual LANs are interconnected using a traditional router device. However, the router has been viewed as the performance bottleneck. Routers should be deployed when segmentation or separation is the need; switches should be used to deliver more bandwidth. The virtual LAN (VLAN) concept is merely an interim step along the way to realizing the fully virtual network.

The ATM Forum's draft LAN emulation standard allows ATM devices to internetwork with traditional LAN networks such as Ethernet and Token Ring. However, it seems ironic that it essentially tries to make ATM networks operate like a traditional shared-access LAN segment. Although it is required for near-term deployment of ATM solutions into existing LAN architectures, its position as an end-all solution is questionable. A more logical approach uses ATM as the model that LANs must emulate.

SUMMARY

Each vendor's approach to virtual networking features will vary slightly in implementation. Most vendors have agreed, however, that the router is moving to the periphery of the network and the core will be based on switching technologies with virtual network capabilities. The three critical success factors that a virtual network vendor must display to effectively deliver on all the promise of virtual networks are connectivity, internetworking, and network management.

Connectivity expertise through a demonstrated leadership in the intelligent hub industry ensures the user a broad product line with numerous options in regards to topology and media types. The product should fit the network, rather than the network design being dictated by the capability of the product. This indicates a vendor's willingness to embrace standards-based connectivity solutions as well as SNMP management and remote monitoring (RMON) analyzer capabilities.

Internetworking expertise ensures that the vendor is fully equipped to deal with layer 2 as well as layer 3 switching issues through an understanding of protocols and their operation. This is not something that can be learned overnight. The integration of these technologies is still unattainable.

Network management software is crucial-virtual networks do not exist or operate without it. The virtual network services provide all the value to the switch fabric. Users should look for a vendor that has delivered distributed management capabilities. Just as the telephone network relies on a distributed software intelligence for its operations, so too must the switched virtual network provide the same degree of redundancy and fault tolerance. Users should also consider whether the vendor embraces all of the popular network management platforms (e.g., SunNet Manager, HP OpenView, Cabletron SPECTRUM, and IBM NetView for AIX) or only one. Finally, users should make sure the vendor has experience managing multiple types of devices from vendors other than itself. It would be naive to think that all of the components that make up a network are of one type from one vendor.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter IV-7
The Data Center Manager's Guide to Ensuring LAN Reliability and Availability

Nathan J. Muller

In many businesses, LANs have displaced older, paper-based methods of doing work, and these organizations now depend on the reliability and availability of their networks for daily operations. A LAN's reliability is a measure of its capacity to operate despite failures in communications links and equipment. Availability is a measure of performance; it refers to the LAN's capacity to meet the communications needs of its users. On a LAN with high availability, services are provided immediately; on a LAN with low availability, users have to wait for network access, or network response is poor. This chapter discusses some of the technologies that contribute to LAN reliability and availability. By applying them, data center managers can ensure that their networks enhance company productivity, not detract from it.

Discussions in this chapter refer to LANs according to their topology. Simple versions of the basic LAN topologies are shown in Exhibit 1.

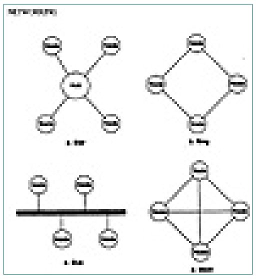


Exhibit 1 Simple versions of the basic network topologies.

RELIABILITY

Star topologies score high in LAN reliability. The loss of a link on a star network prevents communication between the hub and the node attached to that link, but all other nodes operate normally. The hub is the weak link; the reliability of the network depends on the reliability of the central hub, and if the hub malfunctions, all nodes go down. If a peripheral node fails, only that node is affected.

A ring network topology, in its pure form, offers poor reliability during both node and link failures. In ring networks, link segments connect adjacent nodes to each other. Each node passes the data it receives to the adjacent node. The loss of one link not only cuts off one node but can bring down the entire network. To improve the reliability of a ring network, the data center manager can add redundant links between nodes and bypass circuitry in the nodes. This course of action, however, is very expensive.

Although link failures also cause problems on a bus network, a node failure will not bring it down. A redundant link for each segment increases the reliability of the network, but as with ring topologies, it raises the cost.

Mesh topologies are used on internetworks, which connect geographically separated LANs. Mesh networks are the most reliable because they always provide more than one route between nodes. However, this route diversity is achieved by adding more physical links and the equipment (e.g., routers) to support them, which again raises the cost of these networks.

AVAILABILITY

Availability is a measure of a LAN’s capacity to support all users who wish to access it. A network that is highly available provides services immediately to users, whereas a network that suffers from low availability typically forces users to wait for access or degrades overall network performance when it supports too many users at the same time.

Availability on a bus network depends on its load, length, and access control protocol. Under a light load, availability is virtually assured for any user who wishes access. As the load increases, however, so does the probability of information collisions. When a collision occurs, the transmitting nodes stop and attempt to transmit after a short interval. The chance of collisions also increases with bus length.

A mesh topology’s multiple paths make it the most reliable of networks. In addition, because it provides the highest degree of interconnectivity between users, this type of network is always available.

The capacity of a network based on a star topology is limited to the capacity of the central hub. Under heavy load conditions, users can be denied access. However, hubs equipped with multiple processors can provide improved access in high-demand networks.

Although the ring topology does not provide the same degree of availability as a mesh topology does, it is an improvement over the star topology. Ring availability is lower than a mesh network's because each node on a ring must wait to receive permission before transmitting data. As the number of nodes on the ring increases, the time interval allotted to each station for transmission decreases.

ACCESS METHODS

All LANs share media, but the way they do is determined by their access method. A network's access method plays a key role in determining its reliability and availability.

Bus

Ethernet, the major type of bus topology, is contention-based, which means that nodes compete with each other for access to the network. Each terminal listens to the network to determine whether it is idle. Upon sensing that no traffic is currently on the line, the terminal is free to transmit. The trouble with this access method is that several terminals may try to transmit simultaneously, causing a data collision. The more terminals connected to the network, the higher the probability that such collisions will occur.

To avoid a loss of data, transceivers listen as they send, comparing what is sent with what is heard. If they are not the same, a collision has occurred. The transceiver notifies the attached node with a collision signal and sends error messages onto the Ethernet so that all nodes know of the collision.

Because packets travel at nearly the speed of light over coaxial cable, collisions may appear to be unlikely. In fact, during the very brief interval that it takes for a packet to traverse the network, terminals at the far end cannot know that the network is in use. This collision window imposes a practical limit on the length of a bus network that does not use repeaters. In its 802.3 standard for Ethernets, the Institute of Electrical and Electronic Engineers recognizes 2,500 meters as the maximum bus length, regardless of data rate or cable type.

Ethernet is relatively simple to implement; therefore, it typically costs less than other types of network do. Because each node functions independently, the failure of one does not disrupt the operation of the others, and nodes may be added or removed without disrupting the network. Ethernet is media independent, functioning well over twisted-pair wire, coaxial cable, and optical fiber. The choice of media depends on desired data rate, range, and immunity to interference from external sources of electrical noise.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Ring

The token-ring network uses a more efficient topology. Logically, the network forms a ring; however, it may be physically configured as a star. A token consisting of several bytes of data (also referred to as a packet) circulates around the ring, giving each terminal in sequence a chance to put information on the network. This token-passing access method ensures all terminals an equal share of network time.

To transmit, a station captures the token, adds a data packet to it, and then reinserts the token on the ring. Each station in turn receives the packet, but only the addressee retains the data in it, and only the station that put the packet on the ring can remove it. Because each terminal regenerates data packets and the token, token-ring networks do not have the size and distance limitations of Ethernets.

A further advantage of token-ring networks is that traffic can be assigned different priorities. A station can transmit a token only if it has traffic equal to or higher in priority than the priority indicator embedded in the token.

The token ring is not without liabilities. Special procedures must be followed to add a terminal to the network without breaking the ring and to ensure that the new station is recognized by the others and is granted a proportionate share of network time. Because failed repeater circuits in a node can break the ring, bringing down the whole network, each station must be equipped with bypass circuitry.

A logical token ring can be wired as a star. If bypass circuitry is used with this configuration, failures of network equipment can more easily be corrected than if the bypass circuits are in the nodes. Each node is wired to a centrally located panel, which contains the bypass circuits. If a nodal or link failure occurs, the bypass circuit is activated, preventing the ring from being broken. Centralizing the bypass circuits in this way also facilitates moves and changes as well as fault isolation, because identifying connections at a central point is much easier than performing traces between offices and floors.

Anomalies on a token-ring network can tie it up until someone figures out the problem. If a terminal fails before it has passed the token, the whole network goes down until a new token is inserted. The token may even be corrupted by noise so that it is unrecognizable to the stations. The network can also be disrupted by the occasional appearance of two tokens or by the presence of continuously circulating data packets. The latter can occur when data is sent and the originating terminal fails before it can remove the packet from the ring.

To ensure network availability, one terminal is typically designated as the control station to continuously monitor network operations and do such necessary housecleaning as reinserting lost tokens, taking extra tokens off the network, or disposing of “lost” packets. Each station is equipped with control circuitry, so that the first station detecting the failure of the control station assumes responsibility for network supervision. Such protective measures complicate the network and add to its cost.

Star

In a network based on the star topology, network devices are connected to a central hub. This topology is familiar in the office environment, where each telephone is ultimately tied into the PBX. Another example of a star network entails several terminals sharing a single host. The star and the ring network share a key disadvantage in that the failure of a single critical node can result in the failure of the entire network, unless provisions are made for hardware redundancy or bypass. In the star topology, the critical point, of course, is the central node.

An example of a LAN product that uses the star topology is AT&T’s Datakit. In this system, all the network interface units and interconnecting media are contained within a single cabinet to which the individual stations are connected through twisted-pair wiring. The system looks very much like a PBX, but it implements data communications. AT&T also offers StarLAN, which operates at 1M bps, and StarLAN 10, which operates at 10M bps. AT&T originally developed StarLAN to satisfy the need for a low-cost, easy-to-install local area network that would offer more configuration flexibility than IBM’s Token Ring.

Unlike a bus or ring network, in which intelligence is distributed throughout the system, the star network concentrates all of the intelligence required to run the network at a central hub. In the case of AT&T’s StarLAN, the hub is a minicomputer. The failure of one terminal on the star network does not affect the operation of the others unless, of course, the faulty terminal happens to be the hub. Because network intelligence is centralized at the hub, safeguards must be taken to protect it from catastrophic failure. Such measures may include an uninterruptible power supply, an alternative computer on hot standby, or use of a fault-tolerant computer with redundant subsystems built into it.

Other steps may be taken to minimize the effects of a hub outage. For example, file servers in front of the hub may permit limited communication among the terminals connected to it. With such an arrangement, users cannot communicate with terminals not connected to the server, but they can access files stored in the assigned disk area of the server.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Search this book:

Go!

Previous | Table of Contents | Next

EQUIPMENT RESTORAL CAPABILITIES

A variety of types of equipment can help restore service on a disrupted LAN.

Data Switches

Data switches—also known as port-selection or port-contention devices—have been used in mainframe computer environments since the early 1970s. They permit terminals to share a relatively limited number of computer ports on one or more hosts or the hosts' associated peripherals. Data switches have evolved from relatively simple port-selection and port-contention devices to sophisticated communications controllers capable of managing access to LANs and WANs and of initiating restoral procedures.

Unlike matrix switches, which provide permanent connections, data switches establish connections dynamically to users as needed. Such sophistication offers an economical migration path to LANs. For LAN users, the data switch makes an efficient LAN server or gateway to packet and T1 network.

Terminals may be connected to the data switch in several ways. The most reliable connection is a direct connection by RS-232C cabling, at distances as far as 50 feet. For distances of one to two miles, between buildings for example, terminals may be connected to the data switch by line drivers or local multiplexers. Some vendors have integrated these devices into their data switches as optional plug-in cards. Terminal-to-switch connections may also use twisted-pair wiring, whereas remote connections may be achieved over dial-up phone lines.

Today's data switches have enough built-in redundancy to virtually eliminate the data switch as a single point of network failure. Not only is control logic concentrated on a single processor card, but optional redundant logic allows some data switches to automatically activate an optional secondary processor card if the first one fails. Configuration instructions are automatically copied into the standby card on cutover, eliminating the need for manual reentry.

Some manufacturers offer redundant power supplies that plug into the switch and activate automatically on failure of the primary power supply. To prevent unnecessary downtime, the faulty power supply module may even be safely replaced with another plug-in unit while the data switch is in operation.

The redundant or split backplane protects the switch from damage that can occur from the failure of components connected to the bus. In the event of a failure on the bus, the switch automatically cuts over to the spare backplane to maintain uninterrupted operation.

Some data switches perform continuous background diagnostic procedures so that faulty channels can be disabled automatically. If a requested port is out of service, a message notifies the user that the port is unavailable. When the appropriate channel board is replaced, the data switch automatically reenables it. Of course, a terminal keyboard can be used to disable a port from a terminal for any reason.

If the primary route is busy or out of service, the dynamic rerouting capability of some data switches allows users to reach any node on the network without performing manual reroutes. The process is entirely transparent to the user.

Data switches with built-in data rate conversion capability eliminate the need to match terminals with computer ports; each computer port can be set at its highest rate. The buffer in the data switch performs rate conversion for any device that operates at a rate different from its assigned port. Users do not have to be concerned about speed at all, and data center managers do not have to waste time changing the transmission speeds of computer ports to accommodate lower-speed devices. A computer port set at 19.2K bps can send data to a much slower printer.

For reliable data rate conversion (i.e., with no loss of data) connecting devices must provide flow control. When `XON/XOFF` is used for flow control, for example, the switch buffer is prevented from overflowing during data rate conversion. When the buffer is in danger of overflowing, an `XOFF` signal is sent to the computer, telling it to suspend transmission. When the buffer clears, an `XON` signal is sent to the computer, telling it to resume transmission. These settings are also used for reformatting character structures, enabling devices of different manufacturers to communicate with each other through the data switch.

These features and levels of redundancy are especially important for data switches because they are usually configured in the center of star-type networks. The data switch is potentially the only point of failure that could bring down the entire network. The modularity of today's data switches provide an economical method of network protection.

LAN Servers

Distributing LAN resources protects users against the loss of information and the downtime that users of a centralized network experience when it fails. Servers are used to decentralize LAN functions, including security and data protection, network management, and resource accounting.

Moreover, the use of specialized devices as servers permits the integration of diagnostic and maintenance capabilities not found in general-purpose microcomputers. Among these capabilities are error detection and correction, soft controller error detection and correction, and automatic shutdown in case of catastrophic error. Some servers include such management functions as remote console capabilities.

Protecting data at the server has become a critical concern. Some servers store multiple gigabytes of data; loss of this much data or even damage to it can have disastrous consequences. There are several ways to configure a server to minimize data loss, depending on the level of fault tolerance desired and the available budget. Server drives can be unmirrored, mirrored, or duplexed, or a redundant array of inexpensive disks (RAID) can be used.

An unmirrored server configuration entails the use of one disk drive and one disk channel, which includes the controller, a power supply, and interface cabling. This is the basic configuration of most servers. Its advantage is chiefly cost; the user pays only for one disk and disk channel. The disadvantage of this configuration is that a failure in the drive or anywhere on the disk channel can cause the temporary or permanent loss of the stored data.

The mirrored server configuration entails the use of two hard disks of similar size, as well as a single disk channel over which the two disks can be mirrored. In this configuration, all data written to one disk is automatically copied onto the other. If one of the disks fails, the other takes over, thus protecting the data and assuring all users of access to the data. The server's operating system issues an alarm when one of the mirrored disks needs replacing.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

The disadvantage of this configuration is that both disks use the same channel and controller. If either the channel or the controller fails, both disks become inoperative. Furthermore, because the disk channel and controller are shared, writes to the disks must be performed sequentially; that is, after the write has been made to one disk, a write is made to the other disk. This process can degrade overall server performance under heavy loads.

In disk duplexing, multiple disk drives are installed with separate disk channels for each set of drives. If a malfunction occurs anywhere along a disk channel, normal operation continues on the remaining channel and drives. Because each disk drive uses a separate disk channel, write operations are performed simultaneously, a performance advantage over servers using disk mirroring.

Disk duplexing also offers a performance advantage in read operations. Read requests are given to both drives. The drive that is closest to the information answers the request, and the request given to the other drive is cancelled. In addition, the duplexed disks share multiple read requests for concurrent access.

The disadvantage of disk duplexing is the extra cost involved for multiple hard disk drives (also required for disk mirroring) and for the additional disk channels and controller hardware. However, this cost must be weighed against the replacement cost of lost information, plus costs that accrue from the interruption of critical operations and from lost business opportunities. Compared to them, the investment of a few hundred or even a few thousand dollars is negligible.

An emerging method of data protection uses RAID, or many small disks instead of a single large one. Distributing data across disks offers protection from a crash that could cause the loss of all data if it were stored on a single shared disk. Multiple disks also ease disk I/0 bottlenecks, thereby improving information throughput.

For various technical reasons, current RAID solutions are not yet widely viewed as an option. RAID manufacturers, for example, have not yet perfected the means to put enough storage capacity (i.e., in the gigabyte range) onto 5 1/4- and 3 1/2-inch drives. Intelligent controllers that direct the placement and retrieval of data must still be refined, and RAID devices must also come down in price to compete with conventional disk storage.

However, RAID may yet evolve to perform the massive storage chores of the advanced fiber-optic networks that will be introduced by IBM and other computer manufacturers during the 1990s. IBM is a committed proponent of RAID. Its RAID-3 configuration for the RS/6000 workstation includes as many as 40 5 1/4-inch enhanced small device interface disk drives operating at a sustained data rate of 18M bytes per second. Prices for RAID-3 begin at about $113,000.

Bridges

The bridge is a protocol-independent interconnection device that operates at the data link layer. To be more specific, a bridge interconnects at the media access control (MAC) sublayer and routes by using the logical link control sublayer. In working below the communications protocols, the bridge can interconnect LANs that use diverse communications protocols. As long as a bridge operates at the MAC layer, it does not need to perform protocol conversion. It monitors all traffic on the subnets that it links. In reading packets, it looks only for the MAC-layer source and destination address to determine where the packet is going. This means that a bridge can interconnect DECnet, TCP/IP, or XNS networks without concern for higher-level protocols. Unless LAN protocols are the same, however, bridging alone cannot ensure that applications from one network interoperate with applications on another.

As the user population of a LAN grows, performance can suffer because an increasing number of users must contend for the same amount of bandwidth. This can be quite frustrating to the user who merely wants to send a message or print a document. Bridges are useful for partitioning sprawling LANs into discrete subnetworks that are easier to control and manage. Through the use of bridges, similar devices, protocols, and transmission media can be grouped together into communities of interest. This partitioning can eliminate congestion and improve the response time of the entire network. Bridges can also make adding, moving, and changing devices on the network much easier, because the effect on only the subnetwork need be considered. Finally, partitioning makes problems easier to diagnose and isolate while enhancing overall security.

There are two types of bridges: dumb and smart. Dumb bridges must be told which addresses are local and which are remote to filter packets. Smart bridges, also called learning bridges, can figure out address locations by themselves. Such bridges have distinct learning modes for intraLAN traffic and interLAN traffic. IntraLAN traffic requires that the bridge identify each device on the LAN. Some bridges accomplish this within several seconds, even for LANs with several hundred terminals. The locations at remote devices are automatically determined by a process referred to as flooding, A bridge broadcasts to all locations the first packet it receives with an unknown destination address. When it reaches a reply from the remote device, it updates its routing table.

Learning bridges are also used on networks that have many bridges and several paths that traffic can follow between nodes. In this environment, it is possible for some packets to be duplicated, or endlessly looped between bridges. A smart bridge incorporates the spanning tree algorithm or some other proprietary routing algorithm (IBM uses source routing but has indicated it will also offer the spanning tree algorithm) to detect loops and shut down the redundant alternative path. If an active link fails, the smart bridge can detect the failure and activate an idle link automatically.

A bridge that is an integral part of a T1 multiplier is under the control of the integrated network management system. Available bandwidth can be allocated either for LAN-to-LAN communications or for voice or data communications as needed. The integral bridge allows the multiplexer's existing management system to monitor and collect error and use statistics. Furthermore, the bridge's filtering capabilities allow the data center manager to restrict the types of packets that go out over the bridge, thus alleviating traffic bottlenecks.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Routers

The traditional router is a device that is similar to a bridge in that both provide filtering and bridging between subnetworks. Whereas bridges operate at the physical and data link layers, routers join LANs at a higher level: the network layer. Routers take data formatted for LAN protocols, convert it for wide-area packet network protocols, and then perform the process in reverse at the remote location.

Whereas bridges are transparent to network protocols and are used for point-to-point connections between LANs, routers can be used to build complex internetworks. Routers also offer the highest degree of redundancy and fault tolerance, providing congestion control in conjunction with end nodes to ensure that packets traversing large internets do not experience critical errors that can cause host sessions to time out.

Routing is intended to make the most efficient use of a network by sending data over the most available and direct route between nodes. Routers devise their own routing tables, which can adapt quickly to changes in network traffic, thereby balancing the load. Routers can also detect changes in the network to avoid congested or inoperative links. However, not all routers can use more than one path concurrently; some even cause loop problems.

Nevertheless, routers allow the partitioning of networks for tighter access control by eliminating the broadcast requirement of faster, cheaper, but less reliable bridges. Bridges are less reliable than routers because they deliver packets of data on a best-effort basis, which can result in lost data unless the host computer protocol provides error protection. In contrast, routers can provide flow control and more comprehensive error protection.

Recognizing the value of routers for network reliability, vendors of intelligent wiring hubs are now offering router modules that fit into their hub chassis. The router-on-a-card strategy promotes a tight coupling of the device’s network management with the hub vendor’s network management system, usually the simple network management protocol (SNMP). With SNMP, devices from other manufacturers can be managed from the hub; because the router shares the hub’s power supply, it does not introduce another potential point of network failure.

Another advantage of choosing a router module rather than a standalone version is that it eliminates the need to change the physical configuration of an existing LAN. With a standalone router at a central location, the user must run another riser cable from the router and add another port. This may cost much more than adding a module to an existing box.

Hubs now have such advanced management capabilities as protocol and traffic analyses, distributed network management, comprehensive port control, and relational data bases that archive historical information on network performance, store an inventory of interconnected devices, and keep service vendor contact information. These new hubs considerably enhance network reliability. Ultimately, the hub vendors will interface their equipment to enterprise management systems such as IBM’s NetView and AT&T’s Accumaster Integrator.

Some high-end hub vendors have introduced reduced instruction set computing (RISC) architectures to help users avoid the transmission bottlenecks that can result from embedding more and more internet functions into the hub. The move to RISC increases throughput for routing, bridging, and connectivity to high-speed LAN backbones, such as the fiber distributed data interface (FDDI). With such hubs, it is even possible to use a high-speed backbone to connect several FDDI rings. All of this capacity is made available at the port.

Intelligent Wiring Hubs

A fault anywhere in the cabling of a bus or ring network can bring it down, and this weakness is compounded by the inability of these networks to identify the location of a failure from a central administration point. These shortcomings led to the development of the intelligent wiring hub. This device physically rewires bus and ring networks into star networks while logically maintaining their Ethernet or token-ring characteristics. Cabling faults affect only the link’s node; more important, the intelligent hub provides a centralized point for network administration and control.

Installing an intelligent wiring hub saves costs in several ways. Because unshielded twisted-pair wiring is used, there is no need to install new cabling. Redundant links are unnecessary, and bypass circuitry at every drop location is no longer needed to ensure network reliability.

A fully redundant backbone can be installed to interconnect LANs. Backbone redundancy can be achieved at two levels: cable and hub. A secondary physical cable links all of the hubs to protect the network in case one of the cables experiences a break. To protect the network against hub failure, a standby hub must be cabled into the network.

The flexibility of the hub architecture lends itself to variable degrees of fault tolerance, depending on the criticality of the applications being run. For example, workstations running noncritical applications can share the same link to the same LAN module at the hub. Although this arrangement is economical, a failure in the LAN module would put all of the workstations on that link out of commission. A slightly higher degree of fault tolerance can be achieved by distributing the workstations among two LAN modules and links. That way, the failure of one module would affect only half the number of workstations. A one-to-one correspondence of workstations to modules offers an even greater level of fault tolerance, because the failure of one module affects only the workstation connected to it. Of course, this is also a more expensive solution.

Sometimes a mission-critical application demands the highest level of fault tolerance. This can be achieved by connecting the workstation running the application to two LAN modules at the hub, with separate links. A transceiver is used to split the links at the workstation (the ultimate in fault tolerance would be achieved by connecting one of those links to a different hub).

An intelligent wiring hub's subsystems are appropriate points for built-in redundancy. The hub's management system can enhance the fault tolerance of the control logic, backplane, and power supply by monitoring their operation and reporting any anomalies. With the power supply, for example, this monitoring can include hotspot detection and fan diagnostics to identify trouble before it disrupts hub operation. If the main power supply fails, the redundant unit switches over automatically or under management control, without disrupting the network.

Previous	Table of Contents	Next

Products | **Contact Us** | **About Us** | **Privacy** | **Ad Info** | **Home**

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

IMPACT ON SERVICE SELECTION

Interconnecting LANs over a wide area network requires the installation of devices that connect them to a carrier's network. The selection of appropriate hardware can reduce the need for these services; for example, if the hardware is capable of error correction, the carrier's network need not perform this vital function.

X.25 packet networks, often used for LAN interconnection, have such substantial error correction capabilities that any node on these networks can request a retransmission of errored data from the node that sent it. At the time X.25 networks were set up, errors had to be detected and corrected within the network because most end-user equipment did not have enough intelligence and spare processing power to perform this task.

With much of the public network now converted to inherently reliable digital switching and transmission, there is less need for error protection. Today's intelligent end devices are adept at handling error control and diverse protocols. Consequently, the communications protocol used over the network may be scaled down to its bare essentials, permitting an optimal balance of efficiency and throughput. This is the idea behind frame relay and the reason frame relay services are rapidly becoming the preferred means for interconnecting LANs.

SUMMARY

Many companies depend on LANs for much of their processing needs. These networks must be reliable and fast. Fortunately, data center managers with responsibility for the effective functioning of these LANS can choose among many strategies and devices to make this job easier.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Section V
Contingency Planning

Data center managers are responsible for operations 24 hours a day, seven days a week, and 365 days a year. This is a challenge under the best of operating conditions. When disaster strikes an organization, continued data center availability is especially crucial.

Because data center managers have long been responsible for providing continuous and reliable systems operations, the responsibility for contingency planning has often fallen to them. This section presents information by practitioners on how to plan for and implement an effective, enterprisewide contingency plan. It starts with “An Overview of Business Continuity Planning,” which gives data center managers a firm footing from which they can carry out their role as de facto recovery plan managers.

From the overview, the section goes into the details of creating a recovery plan. Chapters V-2 “An Introduction to Data Center Recover Planning,” V-3 “Developing a Data Center Recovery Plan,” and V-4, “Producing a Data Center Recovery Plan,” all by the same authors, walk data center managers through the steps of designing and producing the recovery plan. These chapters provide sample documents and forms that can be easily adapted to develop an actual recovery plan. The authors of the chapters have developed many such plans, and they share this experience in these chapters. One of the authors, Edward S. Devlin, is credited as being the “father of recovery planning.”

Networks have become the backbone of information systems processing today.

They have become huge productivity tools for businesses, but, from a contingency planning viewpoint, this has created a complex situation. Planning a network recovery following a disaster is a monumental task. Chapter V-5, “Network Disaster Recovery Planning,” provides guidance regarding the issues and solutions involved with contingency planning for networks.

Chapter V-1
Overview of Business Continuity Planning

Sally Meglathery

Corporate business continuity planning specifies the methodology, structure, discipline, and procedures needed to back up and recover functional units struck by a catastrophe. Therefore, every functional unit must accept responsibility for developing and implementing the business continuity plan, and the plan must have the total support of management.

Strategically, senior management must ensure the development of a policy stating that the company will recover from any type of outage. Such recovery requires high-level commitment to the policy from all levels of management. Tactically, however, middle management implements the policy and the plan and is responsible for the daily operation of the plan. For management and the functional units to participate, they must have a comprehensive methodology to guide them in their actions and activities. This chapter discusses methods of developing a corporate business continuity plan.

PROJECT PUNNING

There are numerous reasons for developing a total business continuity plan. Some of the most compelling are legal and regulatory requirements. Consideration must be given to the following when developing the plan:

- Are there any federal statutes or regulations applicable to the business which would apply to disasters relating to the business?
- Are there any state statutes or regulations applicable to the business which would apply to disasters relating to the business?
- What contract requirements (e.g., labor contracts, insurance agreements, mortgages, loans, or other financial documents) should be addressed by the plan?
- Are there any common-law considerations, such as claims against directors and officers raised by shareholders and others? Could there be negligence claims against the company for property damage or injuries to customers or business visitors?

Before beginning development of the business continuity plan, management should identify a business continuity project team. The project team is responsible for developing the business continuity plan and designing procedures and reporting techniques to support overall project management. In addition, the project team should identify individuals from senior management to review and approve the work performed by the project team.

Although the makeup of the project team will vary among companies, the following departments should be represented on the team:

- Real estate and facilities.
- Security.
- Human resources.
- Information systems.
- Communications.
- Technology, planning, and development.

Additional departments may also be represented. A business continuity manager should be delegated for the team.

DEVELOPING THE PLAN

The plan that is developed must ensure that any disaster will have a minimum impact on the company. The plan should address the company's reasons for establishing the plan, the functional area of the company's business that the plan will cover, and what staff or materials are in place or should be in place for the plan to function. The following sections discuss the requirements of the business continuity plan, the various elements of the plan, and the scope of the plan.

Previous	Table of Contents	Next

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Plan Requirements

Although most plans address the need to continue data processing operations and to support critical operations during a crisis, most plans fail to consider loss of other functional units within the organization. Data processing generally initiates the need for disaster recovery planning; however, it is now recognized that recovering data centers alone cannot ensure the continuing health of the organization. Companies must address corporate division and department business continuity planning as well. In fact, planning should be done for all essential functional units of the organization.

The plan must be comprehensive; it must deal with the broadest range of disasters possible. There should be a basic plan with additional procedures for specific hazards (e.g., earthquakes, fires, or exposure to hazardous materials). The plan should preserve the integrity of the business, not individual items or goals.

The plan must contain sufficient detail so that its users will know what procedures to follow, how to perform these activities, and the resources that will be available. The plan should contain action steps that have been decided on and agreed to in advance. Both the response to the immediate disaster and the recovery and continuance of business operations and functions must be specified.

The plan must be owned by the organization. Key personnel must participate in identifying priorities, determining alternative strategies, negotiating agreements, and assembling necessary materials. The plan should be reviewed on a periodic basis or when circumstances change. It should be periodically tested with a defined testing program to ensure that it remains effective and up to date.

Plan Elements

The plan itself has five major elements:

- Risk and business impact analysis.
- Alternative analysis.
- Response and recovery planning and plan documentation.
- Plan publication and testing.
- Training and implementation.

These are discussed in the following sections.

Risk and Business Impact Analysis. Before the plan is written, the hazards that may affect the company's facilities must be identified and their potential impact determined. It is also necessary to identify and rank the major business functions and operations. This helps determine the maximum allowable downtime for individual business functions and operations. From there, the minimum resource and personnel needs and time frames in which they will be needed can be identified. Finally, consideration of emergency operating procedures and strategies can begin.

Alternative Analysis. Using the risk and business impact analysis as a base, consideration is given to the internal and external alternatives available for continuation of each function within the necessary time frames. These alternatives should be chosen on the basis of their cost, benefits, and feasibility. The alternatives considered should include not only those that are currently available but those that can be developed.

Response and Recovery Planning and Plan Documentation. This involves the development and documentation of the procedures to be used to activate the plan (by declaration or event). Administrators move specific functions to the alternative or backup facility, maintain operations at that site while the primary site is being restored or a new permanent site prepared, and return operations to the primary site or another permanent location. The plan must identify ways to procure alternative resources to carry out business activities; determine responsibilities and notification procedures for the company, vendors, customers, and others; and detail recovery strategies and responsibilities.

Plan Publication and Testing. The plan must be reviewed and agreed to by senior management and all departments. It must then be documented and distributed to key personnel with additional copies secured off site. Individual sections of the plan should be distributed to those who will be involved with its activation and operation.

The plan should contain a schedule for periodic review and updating. The only way to assess the adequacy of the plan before a disaster occurs is with a program of periodic tests. The tests used will vary from conceptual walkthroughs to actual relocation of specific departments or business functions.

Training and Implementation. Employees should understand what is expected of them in a disaster and what their roles will be in the recovery process. This is achieved with a training and education program, which should be conducted before the plan is implemented.

Previous	Table of Contents	Next

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

The Scope of the Plan

All key personnel should be identified in the business continuity plan and given specific assignments. Common terminology should be defined in the plan document to avoid confusion at the time the plan is put into effect. In addition, the plan should interface with the IS disaster recovery plan. Budgets should be prepared for the initial costs of developing the plan and for the costs of maintaining the plan.

The scope of the business continuity plan should include the features discussed in the following sections.

A Vital Records Program. The plan should help establish an information valuation program to determine which records should be retained and for how long. In addition, there should be a methodology for ensuring that critical records are retained off site.

Security Requirements. The plan defines what security measures must be in place in the event of a disaster and what security measures are necessary for an off-site location. It also states who has access to each location.

Accounting Procedures. Procedures must be put in place to facilitate the acquisition of needed replacement parts and to properly account for the costs of recovery. This in turn facilitates the filing of insurance claims, among other benefits.

Insurance Requirements. The plan should define what insurance claims must be filed and give guidelines on working with risk managers to file a claim. One of the benefits of developing the business continuity plan is that insurance requirements are specifically defined.

Interdepartmental Interfaces. Interfaces between divisions and departments must be defined in the business continuity plan.

Backup, Recovery, and Restoration Strategies. All critical data, files, and documents should be backed up and stored off site. Recovery procedures should be documented in the business continuity plan, defining the steps necessary to recover the information that was lost. Restoration may require recreating the lost data, files, or documents rather than recovering with a backup. Procedures for such restoration must be documented.

Plan Maintenance and Testing. Once implemented, the plan must be tested regularly to ensure that it is up-to-date. The plan should include a maintenance and testing schedule as well as a methodology for testing the plan to ensure that it is operating as expected.

IDENTIFYING CRITICAL RESOURCES

Not all activities within an organization are critical at the time of a catastrophe. The management disaster decision team identifies those operations that it deems critical to the organization. This determination is based on several specific factors, including the time at which the disaster occurs, legal and regulatory requirements, the amount of time that availability is lost, the company's public image, loss of market share, loss of revenue, the type of service loss (e.g., administrative, executive, or financial), and deadline requirements.

In addition, the plan should account for the facilities, equipment, materials, and supplies needed to adequately perform required tasks. Voice and data communications are particularly critical and should be given proper consideration.

For example, personnel are vital to the success of the recovery, and their comfort and support should be given special attention. Supplies and forms should be maintained off site so that a supply is readily available in times of emergency. In addition, transportation can easily be disrupted in times of emergency, and transportation to an off-site location may not be readily available. Therefore, transportation to the main site or an off-site location must be planned if employees are to arrive at the designated stations in a timely manner.

Spare parts and units for power and environmental systems (e.g., air conditioners, fans, and heaters) should be available at the central business location. The engineering staff should have spare parts on hand for replacing broken parts. A backup unit should be available to replace the disabled units. When that is not possible or when the outage is outside the control of the company (e.g., the loss of a telephone company's central office or a power company's power station), the company must be prepared to move to its off-site location.

A vital record is any document that is necessary to ensure the survival of the business. To ensure the preservation and availability of vital records, all corporate documents should be classified as to their importance (e.g., essential, valuable, important, or nonessential). Corporate recordkeeping policies as well as retention requirements based on legal or regulatory requirements should be documented. The source document should be controlled and protected. In addition, there should be backup procedures for the documents, and a copy of them should be maintained at the off-site location.

Documentation, policies, procedures, and standards should be available in hard copy and should be accessible in both main and off-site locations. A disaster recovery plan has no value if the disaster recovery team cannot locate a copy of it.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

ORGANIZING THE PROJECT

The business continuity plan should be prefaced with a mission statement or purpose. This can be incorporated into the introductory section of the plan. All departments and functions involved in the project must understand the need for the plan, agree to participate in its implementation, and be committed to enforcing the plan.

The departments and functions that participate in the project vary among companies. In most companies, however, senior management must be kept up to date and is responsible for making most key decisions. The audit department oversees the entire process, ensuring that controls are enforced. When a disaster strikes, the building and facilities staff determine any losses and necessary repairs, and the public relations and marketing staffs calm customers and reassure them that the company is all right. A legal staff helps protect the company from litigation, negotiates purchase contracts, and enforces contracts.

The human resources department is usually responsible for keeping all employees informed during and after a disaster, particularly in union shops. In addition, this staff often serves as the go-between for employees and management.

When it is necessary to replace equipment or parts, the purchasing department acquires the necessary components at the best possible price, and the financial or accounting department controls costs and purchases. The engineering department ensures that the companies are properly ordered and installed.

At some level, all disasters have an impact on data processing. Therefore, the

IS department must be kept up-to-date and should participate in the recovery procedures. The operations department ensures that the company continues to run as smoothly as possible.

Depending on the company's business, the following departments might also be included in the business continuity planning process:

- Manufacturing.
- Research and development.
- Warehouse and distribution.
- Customer service.
- Field support services.

Representatives from these business areas can identify the functional, management, and support operations of the company in the initial phases of the project, while gathering information for the plan. As a result, critical divisions and departments that support the organization in times of catastrophe are identified.

In any company, the business continuity plan cannot be developed without the commitment and assistance of management and departmental staff. A considerable amount of coordination is also required, both within the company and between any external resources or consultants and company personnel. To facilitate this, it is recommended that different planning teams and functions be created. The size, number, and type of teams used are determined by the size of the company and by the computing environment. The following are various options, ranging from senior-level management teams on down:

- *The management decision-making team*. This team consists of senior management. It is responsible for making major decisions about the continuity plan and about whether or not to move off site after a disaster.
- *The business continuity steering committee*. This committee provides overall management of the project. It establishes and controls policies, standards, and procedures, and it defines the organization of the departments and other participants to ensure cohesive planning groups. This committee should include members of operations, IS, and finance. The actual composition of the team can be agreed on at the initiation of the project.
- *The business continuity planning coordinator*. This individual provides day-to-day coordination of the project and typically works with external resources or consultants. This person must be able to commit sufficient time to the project to ensure that it is completed within the agreed time frame.
- *The management operations team*. This team consists of line managers who are responsible for managing the day-to-day operations after a disaster occurs. They advise the management decision-making team and report decisions down through their respective areas.
- *Department coordinators*. These individuals are responsible for providing information on their department's operations, completing forms, and developing draft plans. Related departments can be grouped

under one coordinator; other departments may have their own individual coordinators. The time required of these individuals increases with each phase of plan development.

- *The emergency operations team*. This team consists of those people who are responsible for ensuring that operations keep running in the off-site environment.
- *The damage assessment and postinvestigation team*. This team is responsible for evaluating damages to the facility and determining the cost to restore operations. It should consist of those people in charge of facilities and operations.
- *The reconstruction team*. This team consists primarily of facilities personnel. It is responsible for managing restoration activities. It is recommended that at least a business continuity steering committee, a business continuity planning coordinator, and department coordinators be appointed.

It is important that departmental employees involved in developing the plan for their departments be aware of the reasons for developing the plan, the project organization, what is expected of them during the project, and the tools and information that will be provided to assist them in their work. This can be achieved by holding one or more group business continuity training meetings to discuss these points. During these meetings, any software that will be used should be demonstrated and all questionnaires and forms to be used in developing the plan should be explained in detail.

The following sections discuss the responsibilities of the various teams that may be involved in business continuity planning.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

The Disaster Decision-Making Team

The disaster decision-making team is primarily responsible for notifying the board of directors, regulatory bodies, regional companies, local companies, international bodies, and the media as required. This team may make these notifications itself or delegate the work.

In addition, members of this team make the final business decisions regarding whether the plan should go into effect, whether to move operations to the off-site location or continue business at the main site, and even whether to continue conducting business at all. Should the plan be put into effect, the team is kept up to date through management operations teams, the business continuity coordinator, and those functional areas reporting to the team that are in charge of handling areas of the disaster.

All recovery activities are submitted to this team for review; however, all disaster containment activities are handled on site as the events take place. Steps taken to contain the disaster are reported back to this team through the management operations team, as they occur if possible or after the fact if not. All major decisions regarding expenditures of funds are made by this team.

The Business Continuity Steering Committee and Planning Coordinator

The business continuity steering committee is responsible for establishing and controlling policies, standards, and procedures and for defining the structure of the project to ensure that the departments and other participants work together cohesively. In addition, the committee reviews, approves, and coordinates the plans developed by the participating groups.

In the event of a disaster, this committee serves as a facilitator, responsible for providing transportation to the backup facilities, if required; notifying affected personnel and families of the status of the disaster; providing cash for needed travel or emergency items; securing the affected areas, the business resumption control center, and the backup site; escorting personnel, if necessary; and presenting a carefully formatted release to the media and affected personnel as to the status of operations and personnel. Several areas are represented on the business continuity steering committee during the disaster, to ensure that basic necessities are made available to support those individuals working to recover the business.

The size of the business continuity steering committee depends on the extent of the disaster and the recovery needs. The following departments should be consulted in forming the committee:

- Purchasing.
- Human resources.
- Communications.
- Auditing.
- Finance and accounting.
- Transportation and amenities.
- Facilities.
- Security.
- Public relations.
- Risk management and insurance.
- Administrative services.
- Operations.
- Information systems.

The business continuity planning coordinator interfaces with the business continuity steering committee to ensure a smooth and successful transition to each phase of the plan. In addition, the coordinator acts as a team manager for the management operations team, discussed in the following section.

The Management Operations Team

The management operations team is responsible for coordinating all emergency operations teams. When management decides that the business continuity plan is to be implemented, these team members (or their alternates) contact the emergency operations team members to advise them of the disaster declaration. They then report to the business resumption control center to begin damage assessment. Once at the disaster site, the management operations team monitors the emergency operations team’s progress and acts as overall manager for all emergency operations teams activated by the operational group.

The management operations team forwards all requests for space, equipment, supplies, and additional human resources support to the department coordinator. The team members report daily on the status of all emergency operations to the business resumption coordinator for the management operations team.

The management operations team is primarily responsible for determining the extent of the disaster, relocating at the business resumption control center, and notifying emergency operations team managers and department coordinators. In addition, the team monitors recovery progress, and compliance with the business resumption plan during recovery and reports on recovery status to the business resumption coordinator, who in turn reports to the company president as required.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

The Department Coordinators Team

The department coordinators team is composed of members from all functional areas. Each department coordinator acts as chairperson for his or her department's emergency operations team. In addition, the department coordinator manages the management disaster decision team and the business continuity steering committee. He or she communicates all of the department's needs and the department's status.

Department coordinators have access to the business resumption control center and attend strategic planning meetings. When a disaster occurs, they contact all emergency operations team managers and coordinate recovery efforts. Department coordinators submit written requests for equipment or supplies as soon as needs are made known to the business continuity steering committee.

Perhaps most important, the department coordinators monitor recovery operations. In this capacity, they receive and communicate status reports, receive daily reports from all emergency operations team managers, request additional human resources support as necessary, and maintain a log of the department's status and progress. In addition, the department coordinators communicate all decisions made by the management disaster decision team to affected managers within the department.

The Emergency Operations Team

The members of the emergency operations team are responsible for the smooth transition to the prearranged emergency backup center, continued operations, emergency procedures, notification of users, requisition of equipment and supplies, and a return to normal processing. Each member of the team should designate an alternate in case the primary team member is unavailable when a disaster occurs.

The size of the emergency operations team depends on the extent of the disaster and operating needs. The responsibilities of the team members include forwarding requests to the business continuity steering committee for transportation to the alternative facilities, if required, and for notification of key employees, affected families, and any employees who were off duty at the time of the disaster. In addition, the emergency operations team makes requests for first aid, supplies, mail or courier service, replacement of software or equipment, temporary workers, additional security or communications measures, backup power, and documentation. Team members also work with the data processing operations and communications departments.

Each emergency operations team has a team manager and a backup manager, who report to the department coordinator. The team manager is responsible for coordinating the recovery effort. The managers participate in the damage assessment meeting to determine the extent of the damage. The manager gives daily status reports regarding recovery and ongoing operations to the business resumption coordinator.

The Damage Assessment and Postinvestigation Team

The damage assessment team reports directly to the management operations team and notifies it of the extent of damage. After damages have been assessed, this team functions as a postinvestigation team to determine the cause of the disaster. In some cases, the cause is obvious (e.g., an earthquake), but in many cases it is not. For example, in the case of a fire, the origin of the fire must be determined as well as how to prevent such a fire from happening again.

The Reconstruction Team

The reconstruction team is composed of those departments required to restore the damaged site. It should include all departments associated with building services as well as representatives from the damaged areas.

The reconstruction team's responsibilities include both temporary and long-term reconstruction efforts. From the initial damage assessment to final reconstruction of the damaged area, the reconstruction team directs and coordinates efforts to bring about a smooth, efficient reconstruction of the damaged areas.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Bookmark It

Search this book:

Go!

Previous | Table of Contents | Next

PREPARING THE PLAN

In preparing the plan, members of the business continuity project team must assemble documentation about their specific functional area and operating environment. In addition, they must identify critical performance requirements and rank the tasks within their jobs according to priority.

Departments that rely heavily on computer processing must explain in detail how their operations interface with each other and are supported by data processing. The needed information can be gathered from:

- Organizational charts.
- Job descriptions.
- Procedures manuals.
- Technical support requirements.
- Existing disaster recovery or business continuity plans.
- Risk analyses.
- Business impact analyses.
- Vulnerability assessments.

Questionnaires can be used successfully to gather information that can provide a foundation for the strategies that must be developed in the planning process. Although questionnaires should be customized for individual projects, they should always provide the basic information presented in Exhibit 1.

Exhibit 1 Checklist of basic information required on business continuity planning questionnaires.

Departments should be asked to complete the questionnaire after the initial training meeting. The completed form should be returned to the department coordinator and any external consultants for review. The department coordinator and external consultants should review the answers with the department manager and the employee who completed the form to clarify, amend, and confirm the information.

The completed questionnaires should be compared to determine the priority of departmental functions, the impact relative to specific time frames, and the minimum resources needed to maintain the company’s critical functions. This information is helpful when considering alternative or backup sites that will be needed.

All of the information obtained in these early phases of plan development is integrated into the business continuity plan. Plan development is designed to integrate or provide interfaces between sections of the data processing plan and the corporate business continuity plan. In addition, the plan incorporates any emergency procedures and provides references to any applicable sections of existing data center and departmental standards and procedures manuals.

The prompt recovery of an organization’s corporate and functional operations from a loss of capability depends on the availability of a broad spectrum of resources. The procedures necessary to restore operations—initially in temporary facilities and later in the original or another permanent location—are detailed in the plan.

Each of the functional units prepares its plan on the basis of the outline provided by the plan coordinators (see the sample outline provided in Exhibit 2). The outline can be modified to suit the needs of the individual units. Although the plan discussed in this section addresses disaster backup and recovery from a worst-case scenario, less severe or even short-term interruptions can also be planned for by using subsets of the overall plan.

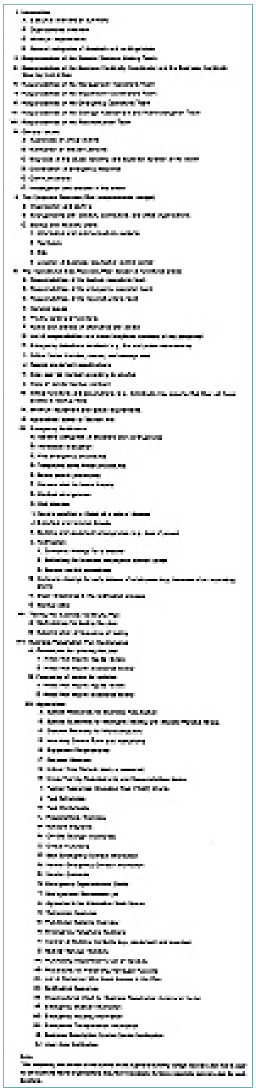


Exhibit 2 Sample outline of business continuity plan.

BUSINESS CONTINUITY PLANNING SOFTWARE

Several contingency planning and risk analysis software packages are currently on the market. It is not practical to list and evaluate them because that list is constantly changing. However, certain criteria should be used during the software package selection process.

For example, ease of use and the number of installations or users are important when the company is selecting any software package, as are the frequency and availability of updates, the quality of documentation and vendor support, the reputation of the vendor, and the amount of training the vendor provides. The usability of output should also be considered. Specific to contingency planning, the software should be evaluated in terms of whether it provides total business continuity planning assistance or simply data center recovery.

ACTION PLAN

For each company, the business continuity plan should cover all types of disaster situations. Procedures should be focused on getting the system running again within an acceptable time frame. The cause of the downtime is not important except in cases of regional disasters (e.g., earthquakes) or such specific hazards as a toxic spill. Special procedures should be included in the plan for these types of disasters.

The recovery strategies and procedures should be organized according to business functions. Strategies and procedures should be sufficiently detailed to enable company personnel to understand what is expected of them and how they should complete their responsibilities. However, strategies and procedures should be sufficiently flexible to permit changes should circumstances warrant them. Procedures should cover the maintenance of critical functions in an emergency mode as well as restoration of the primary facility or relocation to another permanent location.

The plan must specify the priority of recovery activities. It is impractical to determine during an emergency the order in which recovery procedures are to be conducted.

Personnel from the departments covered by the plan should be involved in its development from the start. These departments will be the users of the plan and therefore should play an integral part in its development.

The plan should be reviewed and updated on a regular basis; a plan is only as effective as its maintenance and updating program. Changes in departmental or company operations can quickly render a plan obsolete. A thorough maintenance and updating program prevents this.

Development of a business continuity plan may seem like a long and tedious process with no immediate benefit to the company. However, over the long term, a well-developed and well-maintained plan can help ensure that the company stays in business when a disaster strikes.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter V-2
Introduction to Data Center Recovery Planning

Edward S. Devlin
Cole H. Emerson
Leo A. Wrobel, Jr.

The data center recovery plan (DCRP) consists of documented plans and procedures that will be employed by key data center personnel following a disaster for the express purpose of resuming data center operations in an organized and timely manner. The DCRP lists the documented recovery resources, which include recovery-plan procedures, checklist, and forms.

Disaster

Although a disaster is defined in the dictionary as a sudden, calamitous event bringing great damage, loss, or destruction, a disaster will be defined in the DCRP as any extended interruption to data center operations. Although the interruption might be from physical damage to the building or the data center contents, the term disaster will also be used to indicate any situation in which computer services have suffered an interruption that may extend beyond an acceptable time frame. The following are examples of disasters:

- A hardware failure that causes data center operations to be interrupted over an extended period of time
- A power company suffers a disaster on its premises, causing an interruption of its services and resulting in a data center outage

- A telephone company suffers a disaster on its premises, causing an interruption of its services and resulting in a data center outage

Acceptable Time Frame

The definition of an acceptable time frame varies among companies. To one, the acceptable time frame may be one day; to another, it might be five days. Each recovery planner must identify the acceptable time frame for the specific company.

The DCRP contains the recovery responsibilities for data center personnel primarily, but it also includes information concerning the support roles of both the executive management and the staff departments.

Executive Management Committee

The executive management committee is the group of senior executives that deal with all the major crises that face the company. Although they are not directly involved in the data center recovery operation, they need to be kept informed of its progress throughout. The committee intervenes only when the situation warrants it.

Staff Departments

The staff departments include corporate security, building services/ engineering, public relations, human resources, insurance, legal and purchasing. They are directly involved in assisting the data center during the recovery operation. The DCRP identifies who in each staff department should be notified if there is a disaster and what support they will provide. The DCRP does not document how they will provide support; that information is included in the companywide business-resumption plan.

The Difference Between Data Center Recovery Plan and Business Resumption Planning

The business resumption plan (BRP) consists of documented plans and procedures that will be employed by key company personnel following a disaster for the express purpose of resuming business operation in a timely manner. The BRP contains documented business-resumption responsibilities for all departments in the company (i.e., the revenue-generating department, the staff departments, and the executive management committee). The DCRP is just one component of the overall plan for crisis management (See Exhibit 1).

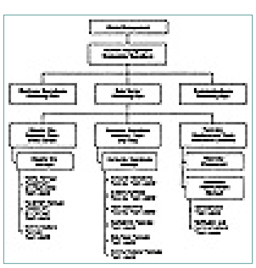


Exhibit 1 Plan components for crisis management.

THE DATA CENTER RECOVERY PLAN MODEL

The data center is one of the divisions of the information systems (IS) department. The responsibilities of the IS department include the data center, as well as communications, personal computers and networks. There is no single IS department organization common to all companies. The organization model that is used in this chapter assumes that the IS department comprises five divisions:

1. System programming
2. Data center operations
3. Applications programming
4. Database
5. Communications

The vice president of information systems manages the department; each of the five divisions is managed by a director (See Exhibit 2).

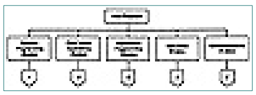


Exhibit 2 Information systems department.

The systems programming division consists of a director and the teams responsible for each of the major operating systems; each team is supervised by a manager. The data center operations division consists of a director and various teams, each supervised by a manager (See Exhibit 3).

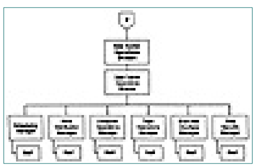


Exhibit 3 Data center operations division.

The applications programming division consists of a director and teams for each major application systems (e.g., a financial systems team). The database division consists of a director and a staff of database specialists. Last, the communications division consists of a director and two teams — the data communications team and the voice communications team — each of which is supervised by a manager. This model IS department is the subject of the DCRP that is discussed in this series of articles.

Previous | Table of Contents | Next

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

History of Data Center Recovery Planning

Data centers throughout the United States started documenting formal recovery plans in 1973. Before 1973, certain segments of the plan were developed, but not complete DCRPs. For example, the data center division of an IS department typically documented its role in a data center recovery operation, but the systems and the applications programming divisions did not. As these divisions began to prepare recovery plans, they often failed to synchronize them with the other divisions' plan. This situation created the demand for documented formal plans so that all divisions could establish a uniform set of recovery objectives.

The data center was the focal point for the development of the disaster-recovery plans during the 1970s, because the revenue-generating departments of most companies needed to provide information to, or obtain information from, the data center to accomplish their business objectives. Because the data center was responsible for the processing of information, it had to develop plans to minimize the length of time the computer would be unable to provide access to this information.

The Evolution of Data Center Recovery Planning: 1970s to 1990s

There are several differences between the DCRPs of the 1970s and those of the 1990s. First, if a data center was disabled by disaster in the early 1970s, this event would not have prevented the company from performing its essential business functions. Because the computer usually acted as an after-the-fact bookkeeper, the business functions would still have been performed; the company could have delivered its product or services, although it would suffer from late billing of customers and late receipt or receivables. During the 1990s, the disabling of the data center can prevent a company from completing many of its essential business functions. The impact of the interruption could result in a significant loss of revenue, profits, and company credibility.

A second difference is that computer environments of the 1990s are much more complex than they were during the 1970s. During the 1970s, data centers operated in a batch environment. The end users' work was carried into the data entry area, where it was keyed; the keyed input was then sent to the schedulers, who submitted it for processing. After the completion of processing, it was sent to quality control, where it was balanced to the control figures. The output would then be picked up by or delivered to the end user.

If there was a disaster, the end users' work could be taken to another data entry location for keying. (If there were no other data entry locations available, it could be taken to a data entry service bureau or to the computer backup site.) The keyed input would be sent to schedulers at the computer backup site, where it would be submitted for processing. After processing, it would be sent to quality control, either at the computer backup site or recovery headquarters, where it could be balanced to the controls. The output would then be delivered to the end user.

Data centers of the 1990s operate in an online environment. End users work directly on workstations; the data is sent to the data center electronically. The end user can call up programs remotely to process the input. After processing, the output is sent back to the end user electronically. Such online environments require that complex technical issues be addressed and implemented before a disaster occurs. When a disaster causes an interruption to the data center's processing, the applications must move to a computer backup site. Such an environment requires several technical issues to be addressed, including ensuring that:

- The backup site has compatible hardware, processors, disk drives, and tape drives.
- There is sufficient disk space to bring up the operating systems and utilities and run the application.
- The operating systems and utilities can be loaded.
- Application data files can be loaded onto the disk space provided by the backup site.
- Application data files can be reconstructed from the time of backup to the off-premises tapes to the time of the disaster.
- Data already processed earlier in the day but destroyed before it could be backed up and sent off-premises can be reentered.
- End-user workstation data-communications lines can be switched into the backup site.
- The security concerns of critical data in a dial-up network environment are addressed.

The good news is there are now more resources available to help deal with these complex issues. First, there are commercial hot sites. Although such sites can meet the needs of most companies today, they cannot meet the disk-space and communications requirements of the largest data centers.

Second, testing procedures can help ensure there is sufficient disk space at the backup hot site for loading of operating systems, utilities, and application data files. During such testing, the reconstruction of data files can also be verified. Third, electronic vaulting can be used to send current transactions off-premises immediately after they are received by the data center.

The final difference in recovery planning between the 1970s and 1990s is that the recovery plans of the 1970s did not address end user, staff departments, and executive management. Data center personnel felt that if a disaster did strike, they would be the only group in the company prepared to respond. (They often complained that they would be up and running, but they would not have anything to process because they were the only group in the company that had done any planning.) The DCRPs of the 1990s include planning sessions with end users, the staff departments that will support the recovery operation, and members of the executive management team that will oversee operation. These planning sessions help the rest of the organization in preparing its overall resumption plan.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Importance of a Data Center Today

With increased numbers of PC-based LANs, the question of the importance of today’s data center is often raised. There is no doubt that local-area networks had an impact on the role of the data center in the organization. As local-area networks become entwined in wide-area networks, however, they often communicate with each other through the data center’s mainframe. In many respects, the data center remains the heart of the organization, sending information that allows the company head to make decisions, and offices and factories to perform specific functions.

REASONS FOR DEVELOPING A DCRP

There are four potential catalysts that can generate a company’s need to develop a DCRP:

1. There is a regulatory requirement for the company to have a DCRP.

2. The company believes that it could be held liable by its clients or customers for not having one.

3. The company is aware of a recent disaster to a company in the same geographic area or industry.

4. The company believes that the impact from an extended business outage to the data center would be too risky.

Regulatory Requirements

First, companies develop the DCRP to be in compliance with regulations governing their industry. For example, the banking industry has had regulations requiring the documentation of a DCRP since May 1983, when BC-177 was issued by the Comptroller of the Currency, requiring all national banks to have a DCRP. This regulation may have been the result of a fire that extensively damaged the headquarters building of Norwest Bank in Minneapolis in November 1982.

The initial version of the BC-177 regulation indicated that the DCRP was needed for any areas of the bank that contained computers. A later version added any areas of the bank that had such electronic equipment as terminals that connect to a mainframe or minicomputer. Currently, this requirement covers any essential business operations of the bank regardless of whether there are any computers.

Liability

Companies have also indicated that a DCRP can help them defend against litigation from customers or client, even if their plan did not work.

Recent Disaster

Many companies have begun a DCRP project just after a disaster occurred in their area or to a company in their business. For example, several insurance companies developed or strengthened their DCRP after the Penn Mutual Insurance Company fire in Philadelphia on May 30, 1989. (A contributing factor might have been the public acknowledgment that the fire was caused by an arsonist who was able to gain access to an area restricted by an access-control system.)

Impact Analysis

Many companies have analyzed what the business impact would be if the data center were to become inoperational for an extended period of time. This is the most common reason executive management is willing to provide a budget for this process. The results of a comprehensive business-impact analysis and a data center application-impact analysis often prompt the executive committee to require the development of a data-recovery plan. The business-impact analysis analyzes the effect on the company if any of the business functions suffer an extended interruption. The application-impact analysis analyzes the impact on the application owner, or any end user, if the data center cannot process a specific computer application for an extended period of time. Typically, other high-profile areas of the company are also required to develop a plan shortly after the data center is incapacitated.

RECOVERY PLAN DEVELOPMENT ASSUMPTIONS

Because potential recovery issues are related to the size of a company, the size of its data center, the type of disaster, and the level of damage caused by the disaster, the following parameters have been used to structure a DCRP.

- The location of the data center is in a separate building from the corporate headquarters.

- The corporate headquarters house the executive, the staff departments, and the data center end users. The executive personnel are the company's key officers, who have final decision-making authority. The staff department personnel manage departments that are responsible for such functions as corporate security, human resources, public relations, insurance, legal, purchasing, and transportation. These departments contribute to the successful operation of the company but do not generate revenue for the company. The data center end users are departments that use the data center mainframe to process information that is used in generating revenue and is essential to their day-to-day operations.
- If the computer center is in the same building as the corporate headquarters, the company must consider companywide planning elements.

The data center organization that the DCRP is being developed for is as presented in this chapter. It is large enough to staff the different teams that make up a data center. For small companies it may be necessary to combine the responsibilities of two or more teams into one team. If a suggested team is not needed, it can be eliminated.

The disaster scenario for the DCRP that is covered is as follows:

- A disaster has occurred causing physical damage to the computer and data communications network equipment, resulting in the inability to use the computer center to support business operations.
- The disaster is isolated in the building housing the data center and communications network. (Regional or local disaster logistics are discussed as exceptions.)
- The data center is inaccessible following the disaster and may remain inaccessible for an extended period of time.
- The data files, supplies, and forms located inside the data center have been damaged or destroyed.
- Key data center personnel have not been injured in the disaster and are available to perform the required recovery actions.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter V-3
Developing a Data Center Recovery Plan

Edward S. Devlin
Cole H. Emerson
Leo A. Wrobel, Jr.

This chapter examines the following seven-step process data center managers can use to develop a recovery plan for their centers.

1. Selecting a plan development committee
2. Establishing the plan's scope
3. Establishing its objects
4. Determining the premise of the plan (i.e., making assumptions about the readiness of the plan and the types of disaster that might occur)
5. Establishing the level of detail for the plan
6. Establishing the plan's format
7. Determining the recovery logistics

STEP 1: SELECT MEMBERS OF THE DEVELOPMENT COMMITTEE

The first step in developing the DCRP is selecting the development committee (see Exhibit 1). The development committee is composed of the personnel who have been selected to work on the development and documentation of the DCRP.

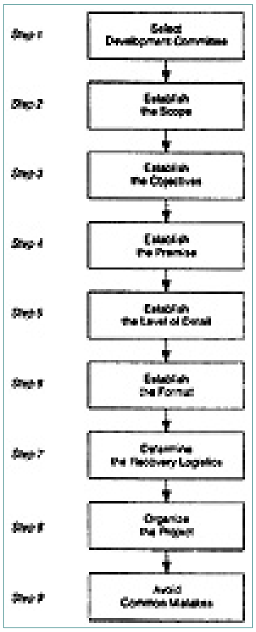


Exhibit 1 The CDRP development steps.

The selection of committee members is important, because this team participates in the development, documentation, implementation, and testing phases of the DCRP project. It decides on the recovery strategies, the resources, and the procedures that will be used in the recovery program.

The DCRP can be thought of as a chain made up of strong links (the recovery team) connected to one another. The recovery teams work together to provide a strong resource during a recovery operation, just as links work together to form a strong chain.

If the person selected to be the representative for a particular division on the development committee is too busy to give the project sufficient time, that team's section of the plan may be a weak link, and, to quote an old adage, a chain is only as strong as its weakest link. Unless all links in the recovery plan are strong, the chain will be likely to break, resulting in a recovery operation that is neither organized or timely.

The DCRP coordinator initiates the process of selecting the development committee. The DCRP coordinator is the representative of the IS department head who has been given the primary responsibility and authority for carrying out the recovery planning project.

The DCRP coordinator should first obtain an organization chart of the IS department, then meet with the head of the IS department and the directors of the divisions to determine who should represent each division on the development committee and what recovery teams will be drawn from each division. The directors are from systems programming, data center operations, application programming, database, and communications divisions.

After the directors have designated their choices for the development committee and selected the recovery teams from each division, they should identify which personnel will be the recovery team leaders — those with primary recovery team leader responsibilities, and those who will be their alternates or backups. An overview of the major recovery teams is presented in

Exhibit 2.

Exhibit 2 Major recovery teams.

The DCRP coordinator and the development committee will meet with the recovery team leaders and their alternates several times during the development phase of the project, for example, during the formal kickoff meeting, during individual recovery team meetings, and during any multiple team meetings. The DCRP coordinator should then meet with the development committee representatives to establish the scope, objectives, premise, level of detail, format, and logistics that will be used in the DCRP.

STEP 2: ESTABLISH THE SCOPE FOR THE DCRP

One of the first steps the development committee takes is to establish the scope of the DCRP. In this project, the term scope refers to establishing the range of operation the DCRP will cover. The DCRP consists of documented plans and procedures that will be employed by key IS personnel following a disaster for the express purpose of resuming data center operations in an organized and timely manner. The critical word is following, as opposed to before or during. Those elements that deal with plans and procedures to be employed before and during are parts of other corporate plans. For example, building evacuation procedures will possibly be used during a disaster.

Previous | Table of Contents | Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter V-4
Producing a Data Center Recovery Plan

Edward S. Devlin
Cole H. Emerson
Leo A. Wrobel, Jr.

In organizing the data center recovery plan (DCRP) project, the development committee should:

1. Conduct the formal kickoff meeting, at which the DCRP coordinator introduces the project to the key players in the data center.

2. Conduct initial individual recovery team meetings for the purpose of developing the first drafts of the team sections of the DCRP manual and for distributing the data gathering forms.

3. Conduct a second round of meetings with the individual recovery teams to review the first draft of their sections of the DCRP and to collect their data gathering forms.

4. Conduct joint recovery-team meetings, with two or more teams working together.

5. Conduct the final round of individual recovery-team meetings to review the drafts to date and include any input or changes required after multiple team meetings. This is the step that produces the final documentation that makes up the DCRP manual.

6. Conduct a formal turnover meeting wherein a presentation of the DCRP, in the form of an overview, is made.

7. Establish a maintenance and exercise program that consists of scheduled updates to the information documented in the plan and periodic exercises of the plan to ensure it is both current and correct.

The organizing steps as summarized in Exhibit 1 provide a game plan that should help the development committee complete the project in an organized and timely manner. The seven steps are explained in more depth in the following sections of this article.

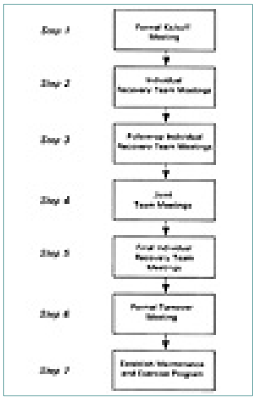


Exhibit 1 Organizing steps for the development project.

STEP 1: CONDUCT FORMAL KICKOFF MEETING

To officially start the project, the DCRP coordinator should conduct a formal kickoff meeting. The individuals who should be invited are the head of the IS department, the IS department representatives who have been selected to be part of the development committee, and all the individuals chosen to be part of the recovery teams, either in a primary, alternate, or supporting role.

The value of the kickoff meeting is that it provides an opportunity for the head of the IS department to introduce the project as a priority. This starts the plan development project on the right foot. The head of the IS department wants the project completed, and the DCRP coordinator is the assigned representative with the authority to carry out the project. The meeting is also valuable because it allows the DCRP coordinator to set the scene, once and for all, for the development project with all the participants. The DCRP coordinator can introduce the scope, the objectives it needs to meet, the premise under which it will be developed, and the level of detail that will be used in developing it. This meeting gives the DCRP coordinator the opportunity to present actual disaster case studies that help illustrate the need for, and value of, the DCRP. At the conclusion of the presentation, attendees should be encouraged to ask questions to further clarify the DCRP project.

STEP 2: CONDUCT INITIAL INDIVIDUAL RECOVERY TEAM MEETINGS

After the kickoff meeting, the DCRP coordinator should call separate meetings with each of the recovery teams that will be needed for the DCRP. The purpose of this meeting is to discuss each team’s recovery logistics and procedures, and the data that must be gathered to document the DCRP.

During the individual recovery-team meeting, the development committee should discuss the recovery logistics that have been assigned to that team. For example, the tape operations recovery team should discuss how the backups will be identified, how the off-premises storage location will be contacted, and how the backups will be taken from the off-premises location to the computer backup site. For the notification and communications team, there should be a discussion of who will make the outgoing phone calls to data center personnel, why these people must follow personnel notification procedure, and how to control outgoing and incoming phone calls.

The team should discuss the recovery checklists, which can be used to check off the actions that have been successfully taken or to account for the status of the inventories of equipment, supplies, and forms. The development committee should also discuss the recovery responsibilities. Finally, this is the time to present the data-gathering assignments to the recovery teams. The data to be gathered will be used in creating the recovery checklists and recovery procedures. The data-gathering assignments are described in the following paragraphs.

ISD Personnel Notification Information. A data-gathering form is used to create the IS department personnel notification checklist. The checklist is, in turn, used to notify the IS department employees of the activation of the DCRP and to advise them of the actions they are to perform.

The information to be gathered includes:

- employee name
- job title
- department name (within the data center)
- home phone number
- car phone number
- beeper number
- address

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter V-5
Network Disaster Recovery Planning

Nathan J. Muller

Although strategies for protecting computer resources from potential disaster are receiving an increasing share of attention these days, the lion's share of concern should instead be given to the links between hosts and terminals. Digital facilities are sensitive to a variety of anomalies that can degrade performance, resulting in transmission errors, retransmission delays, and, at worst, prolonged power failures.

Because there are many more links than host computers, there are more opportunities for failure on the network than in the hosts themselves. Consequently, a disaster recovery plan that takes into account such backup methods as the use of hot sites or cold sites without giving due consideration to link-restoral methods ignores a significant area of potential problems.

Fortunately, corporations can use several methods that help protect their data networks against prolonged downtime and data loss. These methods differ mostly in cost and efficiency.

NETWORK RELIABILITY

A reliable network continues operations despite the failure of a critical element. In the case of a failed link, a reliable network can continue to support applications over an alternative link with unused or spare channel capacity. The mesh topology, for example, is reliable because it provides a diversity of routes; however, route diversity entails adding more physical links to the network, which typically inflates the cost of networking.

Component Reliability

Star Topology. With respect to link failures, the star topology is highly reliable. Although the loss of a link prevents communications between the hub and the affected node, all other nodes continue to operate as before unless the hub suffers a malfunction.

The hub is the weak link in the star topology; the reliability of the network depends on the reliability of the central hub. To ensure a high degree of reliability, the hub has redundant subsystems as critical points: the control logic, backplane, and power supply. The hub's management system can enhance the fault tolerance of these redundant subsystems by monitoring their operation and reporting anomalies. With the power supply, for example, monitoring may include hotspot detection and fan diagnostics that identify trouble before it disrupts hub operation. Upon the failure of the main power supply, the redundant unit switches over automatically or under management control without disrupting the network.

The flexibility of the hub architecture lends itself to variable degrees of fault tolerance, depending on the criticality of the applications. For example, workstations running noncritical applications may share a link to the same local area network (LAN) module at the hub. Although this configuration might seem economical, it is disadvantageous in that a failure in the LAN module would put all the workstations on that link out of commission. A slightly higher degree of fault tolerance may be achieved by distributing the workstations among two LAN modules and links. That way, the failure of one module would affect only half the number of workstations. A one-to-one correspondence of workstations to modules offers an even greater level of fault tolerance, because the failure of one module affects only the workstation connected to it; however, this configuration is also a more expensive solution.

A critical application may demand the highest level of fault tolerance. This can be achieved by connecting the workstation to two LAN modules at the hub with separate links. The ultimate in fault tolerance can be achieved by connecting one of those links to a different hub. In this arrangement, a transceiver is used to split the links from the application's host computer, enabling each link to connect with a different module in the hub or to a different hub. All of these levels of fault tolerance are summarized in Exhibit 1.

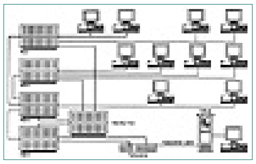


Exhibit 1 Fault tolerance of the hub architecture.

Ring Topology. In its pure form, the ring topology offers poor reliability to both node and link failures. The ring uses link segments to connect adjacent nodes. Each node is actively involved in the transmissions of other nodes through token passing. The token is received by each node and passed onto the adjacent node. The loss of a link not only results in the loss of a node but brings down the entire network as well. Improvement of the reliability of the ring topology requires adding redundant links between nodes as well as bypass circuitry. Adding such components, however, makes the ring topology less cost-effective.

Bus Topology. The bus topology also provides poor reliability. If the link fails, that entire segment of the network is rendered useless. If a node fails, on the other hand, the rest of the network continues to operate. A redundant link for each segment increases the reliability of the bus topology but at extra cost.

NETWORK AVAILABILITY

Availability is a measure of performance dealing with the LAN's ability to support all users who wish to access it. A network that is highly available provides services immediately to users, whereas a network that suffers from low availability typically forces users to wait for access.

HOME ACCOUNT INFO SUBSCRIBE LOGIN SEARCH MY ITKNOWLEDGE FAQ SITEMAP CONTACT US

SEARCH ITKNOWLEDGE

Go!

Keyword

Brief Full

Advanced Search

Search Tips

BROWSE BY TOPIC

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Bookmark It

Search this book:

Go!

Previous | Table of Contents | Next

Section VI
Enterprise Security Management

Part of the data center manager's responsibility for providing 24/7/365 operations is securing enterprise systems so they do not go down to a hack or other type of attack. Ensuring system reliability and integrity also involves securing against other types of unauthorized computer activity. A data center manager's role in securing enterprise systems has become as multifaceted as computing technology has become in today's distributed environments.

In a traditional data center environment, managers mostly ensured that the door to the data center could be opened only by authorized personnel. With the advent of networks and dial-in connections, the number of "doors" to the center's mainframe, minis, and servers has multiplied, and, likewise, so have the security risks. As Chapter VI–1, "Enterprise Security Architecture," explains, effective and efficient security cannot be achieved without an enterprisewide design and a coherent management system.

In a growing number of organizations, enterprise systems means networks and Internet-based systems that accept requests for services from users in various locations, which is the basis of their appeal to organizations. It is also the basis of these systems' security risks, because the systems cannot always verify who is making the requests. Chapter VI–2, "Network and Internet Security," explains why such verification is not always possible and how to implement other forms of security to protect against this risk.

Many mainframes have become the "data base server" in large enterprises.

These mainframe servers send the data across networks to various locations, and thereby expose valuable corporate data to security risks. Chapter VI–3, “Securing Distributed Data Networks,” guides data center managers through various data access controls, ways to secure devices and files, and methods to protect operating systems, as well as present a plan for these systems’ protection.

The Internet is increasingly becoming the platform of choice for many enterprise functions. The Internet was originally invented for the free flow of information between academic and research institutions, where security was not a priority. It was not designed for the distribution of internal corporate information or business transactions, where security is key. Yet, Internet-based systems are increasingly used in mission-critical applications. Chapter VI–4, “Security Management for the World Wide Web,” provides needed guidance in securing Internet-based enterprise systems.

Another platform gaining acceptance as a standard is Microsoft’s Windows NT. Chapter VI–5, “Access Security in a Windows NT Environment,” walks data center managers through the various security features in NT 4.0, and explains when these features are most suitable.

Encoding data is one basic way to secure it. This strategy has received increased attention with the advent of the Internet. Public key cryptology is examined in Chapter VI–6, “Understanding Public Key Cryptology,” which also lists applications of this technology as well as available products.

Viruses and hack attacks are the best-known security threats. Chapter VI–7, “Malicious Software and Hacking,” explains hacking and software-attack techniques and provides measures on how to thwart these attacks.

Securing enterprise systems is very complex, and no number of security measures will guarantee that a system will not be attacked. Network traps are based on this premise. They are set up to give intruders system access, but they monitor the intruder’s activity and often uncover his or her identity. Chapter VI–8, “How to Trap a Network Intruder,” explains the various network traps and how to set them.

Chapter VI-1
Enterprise Security Architecture

William Hugh Murray

Sometime during the 1980s the world of computer users shifted from a majority of multiuser systems to a majority of single-user systems. Now, the trend is toward connecting all computers in the world into the most complex mechanism that humans have ever built. While for many purposes this can be done on an ad hoc basis, for purposes of security, audit, and control, it is essential to have a rigorous and timely design. Effective, much more efficient security cannot be achieved without an enterprisewide design and a coherent management system.

Enterprise

The dictionary definition of enterprise is a project, a task, or an undertaking; or, the readiness for such, the motivation, or the moving forward of that undertaking. But the dictionary does not contain the definition of enterprise as used in this chapter. Here, enterprise is defined as the largest unit of business organization, that unit of business organization associated with ownership. If the institution is a government institution, then it is the smallest unit headed by an elected official. What needs to be understood is that it is a large, coordinated, and independent organization.

| Previous | Table of Contents | Next |

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Chapter VI-2
Network and Internet Security

Steven Bellovin

Why is network security so troublesome, while stand-alone computers remain relatively secure? The problem of network security is difficult because of the complex and open nature of the networks themselves.

There are many reasons for this. First and foremost, a network is designed to accept requests from outside. It is easier for an isolated computer to protect itself from outsiders because it can demand authentication — a successful log-on — first. By contrast, a networked computer expects to receive unauthenticated requests, if for no other reason than to receive electronic mail. This lack of authentication introduces some additional risk, simply because the receiving machine needs to talk to potentially hostile parties.

Even services that should, in principle, be authenticated, often are not. The reasons range from technical difficulty (see the subsequent discussion of routing) to cost to design choices: the architects of that service were either unaware of, or chose to discount, the threats that can arise when a system intended for use in a friendly environment is suddenly exposed to a wide-open network such as the Internet.

More generally, a networked computer offers many different services; a standalone computer offers just one: log-on. Whatever the inherent difficulty of implementing any single service, it is obvious that adding more services will increase the threat at least linearly. In reality, the problem is compounded by the fact that different services can interact. For example, an attacker may use a file transfer protocol to upload some malicious software and then trick some other network service into executing it.

Additional problems arise because of the unbounded nature of a network. A typical local area network may be viewed as an implementation of a loosely coupled, distributed operating system. But in single-computer operating systems, the kernel can trust its own data. That is, one component can create a control block for another to act on. Similarly, the path to the disk is trustable, in that a read request will retrieve the proper data, and a write request will have been vetted by the operating system.

Those assumptions do not hold on a network. A request to a file server may carry fraudulent user credentials, resulting in access violations. The data returned may have been inserted by an intruder or by an authorized user who is trying to gain more privileges. In short, the distributed operating system cannot believe anything, even transmissions from the kernel talking to itself.

In principle, many of these problems can be overcome. In practice, the problem seems to be intractable. Networked computers are far more vulnerable than standalone computers.

GENERAL THREATS

Network security flaws fall into two main categories. Some services do inadequate authentication of incoming requests. Others try to do the right thing; however, buggy code lets the intruder in. Strong authentication and cryptography can do nothing against this second threat; it allows the target computer to establish a well-authenticated, absolutely private connection to a hacker who is capable of doing harm.

Authentication Failures

Some machines grant access based on the network address of the caller. This is acceptable if and only if two conditions are met. First, the trusted network and its attached machines must both be adequately secure, physically and logically. On a typical local area network (LAN), anyone who controls a machine attached to the LAN can reconfigure it to impersonate any other machine on that cable. Depending on the exact situation, this may or may not be easily detectable. Additionally, it is often possible to turn such machines into eavesdropping stations, capable of listening to all other traffic on the LAN. This specifically includes passwords or even encrypted data if the encryption key is derived from a user-specified password [Gong et al., 1993].

Network-based authentication is also suspect if the network cannot be trusted to tell the truth. However, such a level of trust is not tautological; on typical packet networks, such as the Internet, each transmitting host is responsible for putting its own reply address in each and every packet. Obviously, an attacker's machine can lie — and often does.

In many instances, a topological defense will suffice. For example, a router at a network border can reject incoming packets that purport to be from the inside network. In the general case, though, this is inadequate; the interconnections of the networks can be too complex to permit delineation of a simple border, or a site may wish to grant privileges — that is, trust — to some machine that really is outside the physical boundaries of the network.

Although address spoofing is commonly associated with packet networks, it can happen with circuit networks as well. The difference is in who can lie about addresses; in a circuit net, a misconfigured or malconfigured switch can announce incorrect source addresses. Although not often a threat in simple topologies, in networks where different switches are run by different parties, address errors present a real danger. The best-known example is probably the phone system, where many different companies and organizations around the world run different pieces of it. Again, topological defenses sometimes work, but limitations to the actual interconnection patterns remain.

Even if the network address itself can be trusted, there still may be vulnerabilities. Many systems rely not on the network address, but on the network name of the calling party. Depending on how addresses are mapped to names, an enemy can attack the translation process and thereby spoof the target. See Bellovin [1995] for one such example.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter VI-3
Securing Distributed Data Networks

Nathan J. Muller

Protecting vital corporate information from unauthorized access has always been a concern for organizations. Security is a paramount issue with electronic data interchange (EDI) and electronic funds transfer (EFT) applications that involve financial transactions and monetary payments.

Security measures for EDI and other messaging systems are basically the same as for other automated information systems. Major considerations are message integrity and source authentication as well as access control for the microcomputers, minicomputers, and mainframes that handle various portions of the applications.

To the user of such systems, the application and data appear to reside locally; however, each application component and individual user has access to-and sometimes control over—the corporate data residing on various hosts and storage devices. Such openness invites misuse and requires a degree of administration to enforce security procedures.

The ease of use and ready accessibility that have made local area networks (LANs) so popular have also made them vulnerable to security violations. Because LAN resources are distributed throughout the organization, access can be obtained at many points within the network, including unattended wire rooms and closets. Aside from taking precautions to control the physical environment, an organization should periodically evaluate the accessibility of all shared network resources. Distributed environments are usually not controlled by common management structures and authorities. Furthermore, distributed environments constantly evolve in unpredictable directions, further complicating the implementation of centralized security procedures.

ACCESS CONTROLS FOR SHARED DATA BASES

Access controls are needed to prevent unauthorized local access to the network and to control remote access through dial-up ports. The three minimum levels of user access usually assigned are: public, private, and shared access. Public access allows all users to have read-only access to file information. Private access gives specific users read-and-write file access, whereas shared access allows all users to read and write to files.

Perhaps the most difficult application to secure is the shared data base environment offered by LANs. When a company offers network access to one or more data bases, it must restrict and control all user query operations. Each data base should have a protective key; or series of steps, known only to those individuals entitled to access the data. To ensure that intruders cannot duplicate the data from the system files, users should first have to sign on with passwords and then prove that they are entitled to the data requested. Passwords should have a minimum of six or seven characters; passwords with fewer characters are too easily broken by brute-force guessing.

Plaintext passwords are especially vulnerable on LANs because each guess increases the chance of unauthorized entry by a factor of $1 \times n$, where n equals the number of passwords on the LAN. A user ID should be suspended after a certain number of passwords have been entered to reduce the chance of a trial-and-error procedure being used to successfully access the operating system. In addition, to help track any inconsistencies, the IS security function should obtain a daily printout of the keys and passwords used.

Password Protection

To maintain confidentiality, passwords should be difficult to guess or discover with random keyboard combinations. The most effective password is one that is long and obscure yet easily remembered by the workstation user. Changing passwords frequently and using a multilevel password-protection scheme can also help ensure confidentiality.

There are two systems of password protection that organizations can employ to maintain security: hierarchical and specific. With hierarchical passwords, users employ a defined password to gain access to a designated security level, as well as all lower levels. With specific passwords, however, users can access only the intended level. Although specific-level passwords offer greater security, they require that a senior, trusted employee have multiple passwords in order to access the many data bases and associated levels used daily. Password levels, especially specific levels, also make the task of network management more complex.

LAN administrators (who may be from the IS group or a functional department) must select and implement a particular password security method;

they must also ensure that the connected workstations play an active role in supporting password use. For example, when a user enters the password, the monitor should automatically blank out all key entries to minimize the risk of exposing the password to casual observers. The LAN administrator can also install password routines that do not display any information on the screen or that sound an audible alarm and lock the keyboard after a specified number of failed entry attempts. In addition, the LAN administrator should survey the password master file, keep the file on disk, change any infrequently used passwords, and review risks periodically whenever a breach of security is suspected.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter VI-4
Security Management for the World Wide Web

Lynda L. McGhie
Phillip Q. Maier

Companies continue to flock to the Internet in ever-increasing numbers, despite the fact that the overall and underlying environment is not secure. To further complicate the matter, vendors, standards bodies, security organizations, and practitioners cannot agree on a standard, compliant, and technically available approach. As a group of investors concerned with the success of the Internet for business purposes, it is critical that we pool our collective resources and work together to quickly establish and support interoperable security standards; open security interfaces to existing security products and security control mechanisms within other program products; and hardware and software solutions within heterogeneous operating systems that will facilitate smooth transitions.

Interfaces and teaming relationships to further this goal include computer and network security and information security professional associations (CSI, ISSA, NCSA), professional technical and engineering organizations (I/EEE, IETF), vendor and product user groups, government and standards bodies, seminars and conferences, training companies/institutes (MIS), and informal networking among practitioners.

Having the tools and solutions available within the marketplace is a beginning, but we also need strategies and migration paths to accommodate and integrate Internet, Intranet, and World Wide Web (WWW) technologies into our existing IT infrastructure. While there are always emerging challenges, introduction of newer technologies, and customers with challenging and perplexing problems to solve, this approach should enable us to maximize the effectiveness of our existing security investments, while bridging the gap to the long-awaited and sought-after perfect solution.

Security solutions are slowly emerging, but interoperability, universally accepted security standards, application programming interfaces (APIs) for security, vendor support and cooperation, and multiplatform security products are still problematic. Where there are products and solutions, they tend to have niche applicability, be vendor-centric or only address one of a larger set of security problems and requirements. For the most part, no single vendor or even software/vendor consortium has addressed the overall security problem within “open” systems and public networks. This indicates that the problem is very large, and that we are years away from solving today’s problem, not to mention tomorrow’s.

By acknowledging today’s challenges, benchmarking today’s requirements, and understanding our “as is condition” accordingly, we as security practitioners can best plan for security in the 21st century. Added benefits adjacent to this strategy will hopefully include a more cost-effective and seamless integration of security policies, security architectures, security control mechanisms, and security management processes to support this environment.

For most companies, the transition to “open” systems technologies is still in progress and most of us are somewhere in the process of converting mainframe applications and systems to distributed network-centric client-server infrastructures. Nevertheless, we are continually challenged to provide a secure environment today, tomorrow, and in the future, including smooth transitions from one generation to another. This chapter considers a phased integration methodology that initially focuses on the update of corporate policies and procedures, including most security policies and procedures. It also enhances existing distributed security architectures to accommodate the use of the Internet, Intranet, and WWW technologies, and devises a security implementation plan that incorporates the use of new and emerging security products and techniques. Finally, the chapter addresses security management and infrastructure support requirements to tie it all together.

It is important to keep in mind, as with any new and emerging technology, Internet, Intranet, and WWW technologies do not necessarily bring new and unique security concerns, risks, and vulnerabilities, but rather introduces new problems, challenges and approaches within our existing security infrastructure.

Security requirements, goals, and objectives remain the same, while the application of security, control mechanisms, and solution sets are different and require the involvement and cooperation of multidisciplined technical and functional area teams. As in any distributed environment, there are more players, and it is more difficult to find or interpret the overall requirements or even talk to anyone who sees or understands the big picture. More people are involved than ever before, emphasizing the need to communicate both strategic and tactical security plans broadly and effectively throughout the entire enterprise. The security challenges and the resultant problems become larger and more complex in this environment. Management must be kept up to date and thoroughly cognizant of overall risk to the corporation’s information assets with the implementation or decisions to implement new technologies. They must also understand, fund, and support the influx of resources required to manage the security environment.

As with any new and emerging technology, security should be addressed early in terms of understanding the requirements, participating in the evaluation of products and related technologies, and, finally, in the engineering, design, and implementation of new applications and systems. Security should also be considered during all phases of the systems development life cycle. This is nothing new, and many of us have learned this lesson painfully over the years as we have tried to retrofit security solutions as an adjunct to the implementation of some large and complex system. Another important point to consider throughout the integration of new technologies, is that technology does not drive or dictate security policies, but the existing and established security policies drive the application of new technologies. This point must be made to management, customers, and supporting IT personnel.

For most of us, the WWW will be one of the most universal and influential trends impacting our internal enterprise and its computing and networking support structure. It will widely influence our decisions to extend our internal business processes out to the Internet and beyond. It will enable us to use the same user interface, the same critical systems and applications, work toward one single original source of data, and continue to address the age-old problem of reaching the largest number of users at the lowest cost possible.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter VI-5
Access Security in a Windows NT Environment

Gilbert Held

Access security in a Windows NT environment is controlled through the use of the User Manager for Domains program. That program can be directly accessed from the Windows NT 4.0 Start menu button by selecting “programs” from that menu, selecting the “Administrative Tools (Common)” entry in the program pop-up bar, and then selecting the “User Manager for Domain” entry from the next pop-up menu displayed in response to selecting the “Administrative Tools (Common)” program entry. Exhibit 1 illustrates the sequence of previously discussed pop-up menus invoked from the Start button to select the “User Manager for Domains” entry. The latter is shown selected by the highlighted bar in the right pop-up menu.

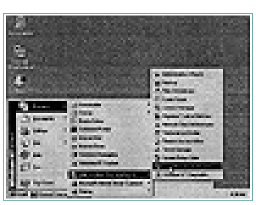


Exhibit 1 Accessing the user manager for domains program from Windows NT 4.0 start button.

The User Manager for Domains is the Windows NT program through which the administrator and authorized users can control access to the server. Exhibit 2 illustrates the initial display of the User Manager program for the computer used by the author. In examining Exhibit 2, note that the screen is subdivided into two horizontal portions. The upper portion of the screen lists the currently configured users' authorized access to the computer while the lower portion of the screen lists in alphabetical order predefined groups and a description of each group. Note that the Administrator's group has the highest level of privileges. Members of that group can add and delete user accounts and group accounts as well as control access to the computer and individual files on the computer. In comparison, the default Guest group contains the fewest privileges and is normally associated with the predefined username Guest, which is included in each version of Windows NT to allow users without an account a limited degree of computer access.

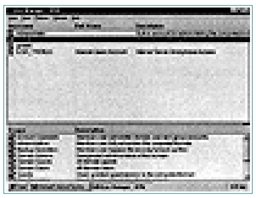


Exhibit 2 Initial display of the Windows NT user manager program.

In examining Exhibit 2, it should also be noted that the Administrator and Guest usernames represent default names established when the server software is installed. The username gxheld was added when the server software was originally configured, while the username *IUSR_TREBLIG* represents a guest Internet account. That account is automatically created by Windows NT 4.0 Server when you install Microsoft's Internet Information Server (ITS) software, and the account provides anonymous access to the Windows NT 4.0 Server's World Wide Web server program. If you wish to restrict access to valid predefined server user accounts, the first thing you should do is to remove the Internet Guest account. In addition, from a security standpoint, it is a good idea to disable the general Guest username to preclude unauthorized users from attempting to use that predefined account as a mechanism to try to breech server security. Since the best way to obtain an appreciation of Window's NT access security is by working with a user account, let's do so. In doing so, we will also note how easy it is to disable a predefined Guest account.

ACCOUNT MANIPULATION

Through the use of the User menu at the top of Exhibit 2 you can add, delete, or modify a user account. In addition, by double clicking on a specific username you can generate a dialog box labeled "User Properties" that provides the basic mechanism for associating different levels of access control to a specific user.

Exhibit 3 illustrates the User Properties dialog box for the username *gxheld* used by this author. After you enter an optional full name and description for the user, you must enter a password twice, with the second entry used to confirm the password. Windows NT is similar to many other operating systems in that it does not provide restrictions on the composition of a password other than its length. Although you could use "rabbit," "bigbill," "termite," or a phrase such as "abadabado," it is suggested that if you do so you should also add some numerics to the word or phrase. Otherwise an unscrupulous person could write a program to use each entry in an electronic dictionary in an attempt to gain unauthorized access to a user account.

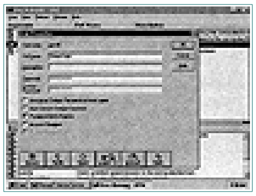


Exhibit 3 User properties dialog box.

Under the second password entry box you will find a series of five small square boxes and labels associated with each box that define its use. By clicking on an appropriate box, a check mark will appear that sets or enables the option. For example, a check mark is shown in the box labeled “Password Never Expires,” which allows this author’s account to remain active with the same password. Note the entry below the “Password Never Expires” entry. That entry, which is labeled “Account Disabled,” provides you with the ability to temporarily disable a previously created account. It is highly recommended that you do this to the Guest account until such time you determine there is sufficient reason to have this type of account. At that time you would select the Guest account and again click on the square associated with the label “Account Disabled” to remove the check mark, which then enables the use of the account. The fifth box, whose label is light gray and barely visible, is “Account Lockout,” set by Windows NT if a predefined number of access attempts fail. You can only deselect or remove a lockout from the User Properties screen.

POLICIES

To provide a degree of control over password composition and account lockout, Windows NT includes an Account Policy dialog box. This box is displayed by selecting the Account entry from the Policy menu located at the top of Exhibit3. Exhibit 4 illustrates the Account Policy dialog box in which this author added a few entries. In examining the available password options note that Windows NT simply allows an administrator to specify a minimum character length, the use of a blank password, which is *not* recommended, an expiration date, and password history that can force the next password to be different from the previously selected password. Unfortunately, Windows NT is similar to other operating systems in that it does not check the composition of the password to prevent the use of common words or phrases susceptible to a dictionary attack.

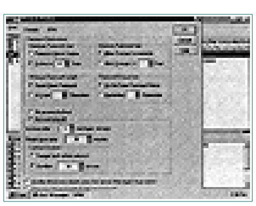


Exhibit 4 Account policy dialog box.

The lower portion of Exhibit 4 shows the settings for an account lockout. In this example, the account will be locked out after five bad attempts and the user will have to wait 30 minutes prior to attempting to reconnect; however, the user can always call the Administrator and ask to have the lockout removed.

Previous	Table of Contents	Next

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Chapter VI-6
Understanding Public Key Cryptology

Gilbert Held

The growth in the use of communications now provides end-users with the ability to perform a variety of functions that just a few years were unimaginable. Today, users can avail themselves of CompuServe, Prodigy, or the services of numerous Internet access providers, to order books, CDs, and even automobiles electronically. Although security has always been a limiting factor, holding end users from transmitting credit card information to on-line sales organizations, the use of public key cryptology represents a mechanism that can overcome that limitation.

The purpose of this chapter is to acquaint systems development managers with the operation and use of public key cryptology to include how this technology is being incorporated into WWW browsers, Web servers, and other computer-based software products. The systems development manager can apply knowledge about the operation of this technology to facilitate secure communications — and understanding the advantages over traditional methods used to secure communications. Because an understanding of the advantages of public key cryptology requires a comparison with traditional cryptology methods, this chapter first examines the general method by which traditional cryptological systems operate.

In a traditional cryptological system, both the person encrypting information and the person who will decrypt the received information use the same encrypting key. The encrypting device uses the key to perform an additive operation, usually a modular 2 (mod 2) operation. In comparison, the decrypting device uses the same key to perform a subtractive operation, which is normally based on a mod 2 operation.

An example can best illustrate the operation of a traditional cryptological system. A one-byte portion of the key sequence used to encrypt data is the binary sequence 10110010. Exhibit 1 illustrates the encryption and decryption of an eight-bit data byte by the one-byte encryption key.

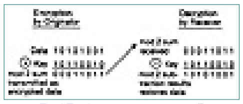


Exhibit 1 Traditional cryptological system operation.

An examination of Exhibit 1 reveals that the mod 2 addition of the key and data that results in a mod 2 sum is transmitted as encrypted data. At the receiver, the same key is used to perform a mod 2 subtraction from the mod 2 sum received, resulting in the reconstruction of the original data. Thus, decryption is performed by the mod 2 subtraction of the key from the received encrypted data.

Traditional Systems Limitations

Although traditional cryptological systems are in wide use and, depending on the key length and method of key selection, are difficult, to essentially impossible, to break, they have a major limitation: The distribution of keys. Exhibit 2 illustrates how the distribution of keys can represent a problem. As indicated in Exhibit 2a, when communications are limited to between two nodes or users, only one key is required. When three nodes or users require communications, the number of keys increases to three, as shown in Exhibit 2b. Exhibit 2c shows that communication among four nodes requires the use of six distinct keys.

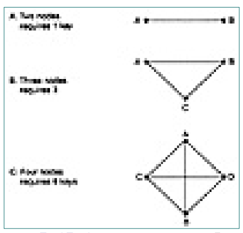


Exhibit 2 Key distribution.

In general form, the equation to determine how many keys are required is:

$$k = n*(n - 1)/2 \text{ where: } k = \text{number of keys, } n = \text{number of users}$$

This means that the use of a traditional cryptological system to secure communications for a large number of users can result in the expenditure of a considerable effort to distribute keys (e.g., for 15 users, 105 keys are required). In addition, each user requiring the ability to communicate with two or more nodes in a large network must be careful in his or her selection of keys, as the selection of the wrong key to encrypt or decrypt data will not produce the desired effect.

In a modern communications environment, such as the World Wide Web, with which the potential exists for tens of thousands of vendors marketing products to tens of millions of users, it becomes obvious that the use of a traditional cryptological system would not be practical for general purpose use. This is due to the massive number of keys that would be required to be distributed as well as the problem users would face in storing, retrieving, and using an appropriate key from an extremely large database of keys. Clearly, an alternative method of key distribution and usage is required. A public key cryptological system is that alternative.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter VI-7
Malicious Software and Hacking

David Ferbrache
Stuart Mort

Since the advent of one of the first computer viruses on the IBM personal computer (PC) platform in 1986, the variety and complexity of malicious software has grown to encompass over 5,000 viruses on IBM PC, Apple Macintosh, Commodore Amiga, Atari ST, and many other platforms. In addition to viruses, a wide range of other disruptions such as Trojan horses, logic bombs, and e-mail bombs have been detected. In each case, the software has been crafted with malicious intent ranging from system disruption to demonstration of the intelligence and creativity of the author.

The wide variety of malicious software is complemented by an extensive range of tools and methods designed to support unauthorized access to computer systems, misuse of telecommunications facilities and computer-based fraud. Behind this range of utilities lies a stratified and complex underculture: the computer underground. The underground embraces all age groups, motivations, and nationalities, and its activities include software piracy, elite system hacking, pornographic bulletin boards, and virus exchange bulletin boards.

CULTURE OF THE UNDERGROUND

An attempt to define the computer underground can produce a variety of descriptions from a number of sources. Many consider it a collection of friendless teenagers who spend their time destroying people's data. To others, it is an elite society of computer gurus whose expertise is an embarrassment to the legitimate bodies that continually try to extinguish their existence. However, the computer underground is really a collection of computer enthusiasts with as varied a collection of personalities as you would experience in any walk of life.

Not all members of the underground are computer anarchists; many use it as an environment in which to gather information and share ideas. However, many are in the following categories:

- *Hackers*, who try to break into computer systems for reasons such as gaining information or destroying data.
- *Malicious software writers*, who create software with a malicious intention. Viruses and Trojan horses are examples.
- *Phreakers*, who hack phones. This is done mainly to gain free phone calls to support other activities such as hacking.

Some have described the inhabitants of the underground as information warriors; this is too glamorous and inaccurate a term. It is true that many individuals' main cause is the freedom of information. These individuals may gain this information by breaking into a computer system, and extracting the stored information for distribution to any person who wants it. Many try to sell the information; these could be termed information brokers. Virus writers are certainly not information warriors, but may be information destroyers.

Thus, we have the person with the computer, surfing the net. An interesting site is stumbled across, with the electronic equivalent of a barbed-wire fence. Behind this fence there must be something interesting, otherwise, why so much security? The site is probed in an attempt to challenge the security. Is this just a person's keen interest in the unknown, or is there a deeper malicious intent?

When security is breached, an assessment of the damage must be made. Was the availability of the system damaged? A virus could have destroyed vital files, crucial to the operation of the system. Has the integrity of data been compromised? An employee's salary could have been changed. Confidentiality lost? A company's new idea stolen. The cost of recovering from a security breach can be major — the time spent by an antivirus expert cleaning up machines after an infection, the time lost when employees could not work because their machines were inoperable. The cost mounts up. It is possible for a company dependent on its computer systems to go bankrupt after a security breach. It could also put peoples' lives at risk. The computer underground poses a significant threat to computer systems of all descriptions all over the world.

| Previous | Table of Contents | Next |

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Chapter VI-8
How to Trap the Network Intruder

Jeff Flynn

The job of securing networks is quite difficult. Probably the most significant reason is system complexity. Networks are complicated. They are so complicated no one person can fully comprehend exactly how they work. The models that govern the designs were developed with this concept in mind and provide a layered view of networks that hide the true complexity. This makes it possible for programmers to work on various layers without understanding all the details of the other layers. Of course, programmers on occasion make mistakes, and these mistakes accumulate. Consequently, the Internet we have come to rely on is vulnerable to a wide variety of attacks. Some of the vulnerabilities are well known. Others are known only to a few or are yet to be discovered.

As the Internet grows, so too does the complexity. The growth of the Internet is still accelerating. Every year, more systems are connected to it than were connected the year before. These systems contain increasing amounts of memory. Larger memories allow programmers to develop larger and more complex programs, which provides the programmers with more opportunities to make mistakes. Larger programs also provide intruders with more places to hide malicious code.

Thus, a good network security manager must be very good indeed. The best network security managers may find themselves performing against the unrealistic expectation that they cannot be overwhelmed. These experts must keep up with all the latest attacks and countermeasures. Attackers, on the other hand, need to know only one or a small combination of attacks that will work against their opponents.

A common response to this situation is to simply fix the known problems. This involves closely monitoring reports from organizations such as CERT or CIAC. As new vulnerabilities are discovered, the system manager responds appropriately. Unfortunately, the list of problems is also growing at an increasing rate. This can be a frustrating experience for the system manager who is forced to fight a losing battle. Likewise, financial managers are caught. They recognize that there are significant risks, yet no investment in safeguards can guarantee immunity from disaster.

It is hard to assess the extent to which tools have improved the situation. The Internet is a highly dynamic environment and does not provide good control samples for making such observations. The commonsense view might be, “However bad it is, it would be worse if we didn’t have these devices.” Unfortunately, the tools are not always applied properly and can lull management into thinking the situation is under control when it is not. In this situation, there is no benefit. The impact on the intruders is also quite difficult to assess. Serious intruders go to great lengths to keep their identities and approaches secret. Assessing the threat is, hence, a difficult aspect of evaluating the effectiveness of tools.

ASSESSING THE THREAT

There are many ways to gain a perspective on the threat. Most professionals in the field of network security use more than one. Some ways are more subjective than others. Yet there are several popular choices.

Reading

Several written information sources are available on the subject of network security. These include books, technical articles, newspaper articles, trade journal articles, newsgroups, and mailing lists. Each of these mediums has its strengths. Each also has its weaknesses. Trade journal articles, for example, can be biased and may attempt to use fear, uncertainty, and doubt to motivate buyers. Newspaper articles, although less biased, are driven by readership and limited in technical detail. Technical articles are many times too technical, sometimes describing threats that were not threats before publication. The information found in books is quickly dated. Finally, newsgroups and mailing lists, while providing timely information, are transmitted via networks that are subject to the same attacks we are attempting to prevent.

Experimentation

One way to see how difficult it is for someone to break into your system is to attempt to break into it yourself. The Self-Hack Audit, sometimes called Penetration Testing, is a useful means for finding weaknesses and is likely to improve awareness. Similarly, information warfare games provide true insight into how sophisticated intrusions can occur. Still, both of these methods are contrived and do not necessarily represent the actual threat.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Section VII Personnel and Career Management

Data center managers do not manage only technology. They also manage a staff, without which no data center can run, and being skilled in dealing with employees is as crucial as being skilled in managing any mainframe or network. Perhaps the most important employee a data center has to look out for is herself or himself. Managing a career as an IS professional and data center manager requires paying as much attention to changing market conditions as managing systems does with constantly changing technology.

In today's economy with its tight job market, skilled employees are scarce. Keeping these workers on staff is a challenge, because there are many other employers eager to hire them away. Often, budgets limit what valued employees can be paid, so data center managers must find other ways to make employees want to stay employed at the same organization. Chapter VII–1, "Fostering Loyal and Long-Term Employees by Raising Organizational Identification," gives techniques for making employees not just hired help but team members, whose loyalty to an organization is valued.

Another way to reward employees that does not involve financial compensation is to increase their responsibility in decision making. This type of empowerment not only makes the employees more valuable but also is part of many organizational styles that promote quality and doing more with less. Chapter VII–2, "IT Organizational Styles" shows ways to organize the data center so that staff members are empowered to do more and deliver higher-quality service to the center's customers. Chapter VII–3, "Implementing Employee Empowerment," shows how to implement these organizational styles that promote employees' roles in an organization.

Just as empowering employees is a trend in today's organizations, data center managers are finding that they have to empower themselves as managers of their own careers. Chapter VII–4, "IT Career Planning," emphasizes that data center managers must take it upon themselves to become more knowledgeable about their organization's business as well as become more skilled as business professionals. Since organizations are increasingly aligning their business strategies with IT strategies, data center managers must become as well versed in management as they are in technology.

As Chapter VII–5 explains, one way to take control of managing a career is to develop a personal business plan that enables self-examination of values, vision, goals, strengths, weaknesses, opportunities, threats, and distinctive competencies. A personal business plan is a tool for data center managers to assess the current stage of their careers, where they want their careers to head, how to take their careers to desired goals, which opportunities or threats may hinder their careers, and industry and market conditions.

Chapter VII–6, "Developing People Skills: New IT Priority," concludes this section by showing the relationship between managing personnel and advancing one's career. Dealing with and effectively managing people may be the single most important skill a data center manager needs today to succeed. Without willing participants to develop and implement technology, technology has no real value to an organization. This chapter emphasizes improving one's communication skills and discusses how these skills can benefit relationships with end users, upper management, and staff, as well as career advancement.

Chapter VII-1
Fostering Loyal and Long-Term Employees by Raising Organizational Identification

Carl Stephen Guynes
J. Wayne Spence
Leon A. Kappelman

The increased organizational reliance on IT has led to the evolution of a group of highly technical IT specialists who display certain distinctive, common characteristics. Among these common traits are:

- High growth needs.
- High professional identification.
- Low organizational identification.

IS professionals tend to like the change and challenge of their work, generally have high standards of professional conduct and performance.

Unfortunately for most employers, they also exhibit low organizational identification — that is, they do not exhibit a strong sense of identification or commitment to their current employer. Although there is little an organization can do about fostering the growth needs of its employees, and can have only a limited, albeit important, influence on their sense of professionalism, an organization can take significant actions to foster their sense of identification. All too often, however, these actions are not taken, and organizations not only experience high turnover and all of its associated problems but, more important and enigmatic, fail to optimally utilize their information assets.

This historical low organizational identification problem is now being manifested by other organizational members as more firms downsize; thus, the prescription for IS personnel may well have a wider context. IS professionals knows that since their skill set is in demand, if they become disenchanted with their current employer they can find a suitable job with another firm by investing a minimum amount of effort. Perhaps because of their employment flexibility coupled with the specialized nature of the work, it is not surprising to discover a peculiar cliquishness associated with small specialized work groups such as that found in IS shops. This exclusiveness of interest and resulting camaraderie fosters a sense of functional identification among some IS staff members at the expense of organizational identification. But these are not mutually exclusive conditions and this does not have to be the case.

IS Employees' Concerns are Not Always the Company's

In many cases, these IS professionals are not really concerned about the organization. They have a job to perform, and it does not matter to whom they report. Organizational problems only secondarily affect them, and they are frequently indifferent regarding the nature of a problem. They do their job in much the same way regardless of whether they are working for a public-sector organization or a large private corporation. The problems that they face may be different depending upon the particular organization they work for, but their approach to solving the problem is handled in a similar fashion.

In many instances, the organization's philosophy of management further contributes to the negative attitude held by these specialists. Organizational management often does not identify with its own personnel. Many managers do not understand the technical environment and have no desire to learn about them. These managers are often in a position where they must accept the use of the systems, but they do not try to understand them.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Chapter VII-2
IT Organizational Styles

Madeline Weiss

For many organizations, continuous improvement and rapid response to market needs are a way of life. Because of these pressures, few organizations are able to provide the long-term security of traditional careers. Whereas mergers and acquisitions were rampant a few years ago, the current trend is toward corporate downsizings and restructurings that lead to the creation of a just-in-time work force or to the outsourcing of services or entire departments. Organizations have higher expectations for quality, productivity, and commitment from their remaining employees. At the same time, employees expect more-satisfying jobs and greater input into decision making.

These trends hold true not only for organizations as a whole but for the IS departments within them. IS must support the organization's use of information technology with agility and innovation. Some IS departments are converting from chargeback systems to staff services for sale, and others can anticipate pressure from their organizations to do so. In effect, with staff services for sale, business units outsource work to IS, which must compete with outside IS service providers. To compete effectively, IS departments must produce high-quality products and services with fewer resources.

STRATEGIES TO ACHIEVE A NEW ORGANIZATIONAL STYLE

Leading-edge IS organizations are responding to these trends by adopting a number of successful strategies, including:

- Flattening the management hierarchy.
- Adopting total quality management.
- Establishing empowered, customer-focused organization cultures and structures.

This chapter addresses these strategies.

FLATTENING THE MANAGEMENT HIERARCHY

Over the past few years, Weiss Associates, Inc., has conducted organization simulations with IS managers and professionals. In each case, participants are divided into three categories — top, middle, and bottom — arbitrarily. Instructions to each group are intentionally vague. Invariably, the same dynamics emerge: Those at the top feel isolated; those in the middle feel like useless go-betweens and messengers, and those at the bottom feel frustrated, powerless, and angry. No matter how open they try to be, the top and middle categories are mistrusted by participants at the bottom, and nothing is ever accomplished. Discussions after the simulation always focus on how bureaucratic chains of command stand in the way of agility and innovation — to say nothing of how they create frustration and anger.

For decades, IS departments have emulated the management style of their organizations, which is based on a management model of bureaucracy that is 50 years old. The model prescribes clear systems of hierarchical relationships with precise direction and close supervision by cascading layers of managers who collect information and prepare reports for higher-level managers. The result is an expensive pyramid of supervisory managers who keep the business humming (not blazing) and reward themselves with generous salaries and high status.

Times are changing. IS professionals are highly educated and can direct their own activities. Most are highly motivated (before the organizational hierarchy discourages them) and want considerable autonomy in their work. Communications are rapid. Organizations cannot wait for bureaucratic hierarchies to make decisions. They need answers from flexible, responsive, committed professionals on the frontline, who are closest to their customers' ever-changing needs.

In these flatter organizations, everyone accepts ownership for delivering quality products and services. Remaining managers focus their energy and time on setting and communicating visions, coaching, mentoring, and providing resources to IS professionals.

ADOPTING TOTAL QUALITY MANAGEMENT

Organizations recognize the need to establish systems and practices that motivate, support, and enable members to consistently design, produce, and deliver quality offerings that meet or exceed customer requirements. Although each organization's approach to establishing a quality environment varies, certain common elements underlie the variations: leadership, information and analysis, planning, human resources use, quality assurance of products and services, quality results, and customer satisfaction. These elements are the seven basic criteria for the Malcolm Baldrige National Quality Award. Leading IS organizations incorporate these seven elements into their culture.

1. *Leadership.* Senior IS managers create, communicate, and manage daily through clear visions of where their organizations are headed in terms of technology leadership, customer service, and IS staff work life.

2. *Information and Analysis.* The IS department seeks information on customer satisfaction through surveys, interviews, and informal exchanges, analyzing feedback to determine how to serve its customers better.

3. *Planning.* The IS department takes responsibility for understanding customer requirements for information technology and developing plans for meeting and even anticipating those requirements.

4. *Human Resources Use.* IS professionals are encouraged to grow by learning and using new skills and knowledge. There is a recognition that customer relations mirror staff relations because well-respected and satisfied staff members treat their customers well.

5. *Quality Assurance of Products and Services.* IS members have accountability and skills for continuously improving their delivery of quality products and services.

6. *Quality Results.* There are quantitative and qualitative measures of quality based on overall IS quality goals. IS members recognize that customers are the ultimate arbiters of quality.

7. *Customer Satisfaction.* The IS department stays close to its customers to understand changing requirements and ensure that they are met, if not exceeded.

ESTABLISHING EMPOWERED, CUSTOMER-FOCUSED CULTURES AND STRUCTURES

Other forward-thinking IS departments have gone beyond flattening the hierarchy and establishing total quality management processes. By thinking creatively about relationships between people, work, technology, and information, some corporate IS departments are developing work systems capable of achieving significantly higher levels of sustained performance. These work systems frequently include empowered, customer-focused cultures and structures, many of which are characterized by self-managing teams.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter VII-3
Implementing Employee Empowerment

James A. Ward

Total quality management (TQM) is built on the four cornerstones of customer focus, continuous process improvement, management leadership, and employee empowerment. Although empowerment is often the most difficult of the four concepts to implement effectively, organizations that have been successful in empowering their employees invariably see bottom-line improvements.

Empowerment has been defined as that group of practices designed to drive day-to-day decision making to ever lower organizational levels. Such practices include implementing self-managed work teams, highly autonomous improvement teams, workers with ultimate responsibility for their own quality assurance, and highly empowered customer interface personnel. Empowered employees make decisions affecting their work and their customers, often as members of teams, and without constant supervisory review and approval.

Ultimate success in implementing empowerment depends on the sincere commitment of both employees and management. It is the role of IS managers to ensure that their organizations are truly on the road to commitment and not merely forcing compliance or heading toward chaos.

CREATING THE VISION

In an empowered organization, all employees must be singing from the same song book. In other words, management should articulate and communicate organizational goals and strategies in such a way that the public, employees, vendors, and customers understand what to expect from the organization. To attempt empowerment without first creating the common vision is to invite anarchy. As a freed slave is said to have remarked on hearing of the Emancipation Proclamation, “Free [empowered] to do what?” Management cannot abdicate its responsibility to communicate a vision, which employees, in turn, must embrace and own.

Communication in all directions and across all levels of management and employees is the key to creating a common vision. Ideally, customers and vendors should be included in the exchange. Electronic mail, Internet access and such systems as Lotus Notes have given organizations tremendous new capabilities to communicate, give and receive feedback, and carry on dialogues.

PUBLICIZING AND ADHERING TO THE MISSION STATEMENT

An organization should have a published and well-publicized quality mission statement with which organizational behavior is absolutely consistent. Getting management to subscribe to a quality mission statement is the easy part. The difficulties arise in achieving wide dissemination of the mission statement and in actually conducting business in a manner consistent with TQM principles.

Implicit in the TQM methodology is the belief that employees will act with understanding and good faith consistent with objectives that the organization must achieve to meet customer requirements. Unfortunately, in some organizations with supposedly active quality programs, employees are unaware of quality efforts. They regard quality as just another management propaganda program that does not affect them. How often have IS organizations shipped software or implemented a system with known bugs or functionality that does not meet stated requirements in order to meet a published ship date? This is more the rule than the exception. In such organizations, attempts to meet customer requirements are called “phase 2” or “system enhancements.”

Employees who don’t even know that their organization has a quality mission statement, let alone what the statement says, cannot be empowered. In most such cases, the organization’s customers and vendors are similarly unaware of the quality mission. Customers are used to living with what they get and feeling lucky to get anything at all.

The preceding paragraphs highlight a major weakness in the way IS people cooperate, manage, and organize themselves. Although the goals of developers, customers, project managers, testers, maintainers, operators, and management usually conflict, empowered employees in cross-functional teams can work through these conflicts. IS employees are eager to do so, in many cases far more so than their managers, who often exercise power by encouraging and fomenting conflict. Intense focus on the customer should be the essential guiding principle.

For an organization to be truly successful in the pursuit of quality, all employees, customers, and vendors must be brought into the process. Commitment and mutual trust must be established.

Previous	Table of Contents	Next

Products | **Contact Us** | **About Us** | **Privacy** | **Ad Info** | **Home**

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Chapter VII-4
IT Career Planning

Layne C. Bradley

Given the rapid change and turmoil the information systems industry is experiencing during the 1990s, the idea of career planning might seem to be almost irrelevant. In the face of downsizing and outsourcing, merely keeping a job could be considered the extent of current career planning. In fact, the idea of a career now has a much different connotation than in the past.

Many data center managers might wonder whether they actually have a career anymore. Some have found themselves outsourced, downsized, or reengineered out of a job. Others who have so far escaped the sweeping management changes taking place in many corporations wonder how much longer they will be able to practice their chosen profession. This chapter provides data center managers with a better understanding of the changes that are taking place today, why they are taking place, and how to deal with them.

CAREER STRATEGIES

Data center managers must understand that simply being a highly skilled technician is no longer sufficient. Although technology is still important, it is rapidly becoming a commodity. Companies generally are no longer interested in technology for technology’s sake, being leading-edge companies, or experimenting with new technologies merely to uncover potential applications. Rather, most companies are interested in using technology and systems only if they add value to the business. Adding value usually means increasing revenue, reducing expenses, reducing personnel, gaining a competitive edge, or moving new products to market faster. Thus, the data center manager must become recognized as someone capable of playing a key role in determining how new technology can add value to the company.

Data center managers must think not only in technical terms but also in business terms. While data center managers clearly will continue to have responsibility for managing the technical environment of the data center on a daily basis, they should not make that the sole focus of their jobs. All data center tasks should be approached from the perspective of how they can improve overall business operations. Improved efficiency, reduced cost, and higher productivity should be the primary focus for data center managers.

When dealing with corporate management, data center managers must think and talk in business terms, not technical ones. To be perceived as a business professional, the data center manager must learn to speak the language of business. In other words, it is essential to use such terms as cost-benefit analysis and return on investment when proposing new projects or plans for the data center. Corporate management now includes such positions as the Vice President of Information Systems or the Chief Information Officer. These positions are gaining increased visibility and stature in many corporations. In many cases, however, they are also being filled with nontechnical executives from operational areas other than information systems. Thus, the data center manager could easily be reporting to an executive with little or no technical background. As a result, data center managers must be able to present their ideas from a business, rather than a technical, perspective.

To keep up with the rapid changes in technology, data center managers must expand their skills. For example, if the data center manager's skill set involves COBOL, DB2, and a centralized mainframe environment, it is time to broaden that skill set — quickly. The data center manager must become skilled in such technologies as LANs, client/server architecture, imaging, object-oriented programming, and wireless communications. Although mainframes will continue to be the hub of the data center for some time to come, their role is clearly changing and the data center manager must understand that changing role.

Data center managers also must improve their general business skills. At one time, it was unusual for a data center manager to have an MBA. Today, such a level of business acumen is almost mandatory. To have a successful IS career, the data center manager must become part of the business management team. IS is rapidly being absorbed by the business functions, and only those IS players with strong technical and business skills will achieve leadership roles in the company.

Data center managers must understand all aspects of the company. In the past, IS professionals often believed they could practice their particular skills anywhere. At one time, that view was reasonably accurate and not likely to create any career problems. However, given the current idea that systems must add value to the business, this approach is no longer acceptable. The current corporate view is that systems must add value, and data center managers can be certain their actions are in fact adding value only if they thoroughly understand the company's business.

Finally, data center managers must revise their concept of a career. It is becoming less likely that any individual will have a long-term career with a single company. Rather, a career will probably involve working for several companies, perhaps in different industries. Thus, data center managers must take responsibility for managing their own careers. They must remain knowledgeable of new technology, expand and maintain their business skills, and develop reputations as managers who can clearly combine business and technology skills to produce value for their employer. A data center manager who can do these things well will be in demand.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management
(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter VII-5
Developing a Personal Business Plan

Stewart L. Stokes, Jr.

Today's IS professionals are striving to shape their futures in a paradox. On the one hand, the needs of rapidly changing enterprises and the challenges of new technologies offer unlimited opportunities. On the other hand, enterprises are relentlessly shrinking their headcounts, and new technologies often empower users before they benefit IS professionals. Added to this contradictory situation is the realization that consultants and contractors frequently have the opportunity to learn new technologies before the IS staff does.

It is no surprise, therefore, that job satisfaction is declining for senior IS executives, IS managers, and IS staff alike. * IS professionals at all levels are finding themselves grappling with the following questions:

*R. Garner, "Job Satisfaction: IS Adrift," *Computerworld* (May 27, 1996), pp. 88–91.

- Where am I going in my career? Am I pursuing a career or merely holding a series of jobs linked together by circumstances?
- Am I meeting my career goals or spending my time fulfilling someone else's goals?
- Am I able to express my personal values in my career decisions, or have I put my personal values on hold in order to survive?

- Am I an active participant in my own future or merely an interested bystander?

To answer these questions and survive in today's paradoxical world, IS professionals need to incorporate proactive career planning into their lives and develop personal business plans that they can maintain in the face of the ever-changing marketplace for IS talent. This column's blueprint of personal business planning discusses the following issues:

- The need for personal business plans
- The ways in which a personal business plan differs from a résumé
- The contents of a personal business plan
- A suggested structure for a personal business plan

NEED FOR PERSONAL BUSINESS PLANNING

An experienced IS manager caught in his company's latest wave of rightsizing called me to network and discuss how to approach the next chapter in his career. He was receiving outplacement assistance that focused on helping him polish his résumé, create a list of contacts, and rehearse for interviews.

The IS manager was pleased that his réumé was circulating and that he was learning how to answer, and ask, the right questions during interviews. What he believed he was lacking, though, was long-term direction. He'd always assumed his stable employment history would be a strength, but some interviewers now viewed it with concern, if not as an outright weakness. A couple of them wondered if he might not have gotten too comfortable and lost his learning edge, as they put it.

Although I knew that becoming comfortable was not his style, I also knew that his latest job did not require that he deal directly with new technology. I also knew that his managerial experience had been in hierarchical, command-and-control organizations. He had little experience leading and managing in the newer, team-based and less-structured environments. As we discussed his job history, he concluded that he might, in fact, have lost his learning edge, and although he was not yet a dinosaur, he could be on the road to career extinction.

We discussed his career direction, including his personal vision, core competencies, and how he might see himself as a value-adding product. It became clear that because of his history of stable employment, he had little career vision. He knew what he liked to do (which he described using vague terms such as "coordinate, oversee," and "facilitate"), but he admitted he was having trouble summarizing his strengths and describing them in a convincing manner.

We talked about the need to focus on his long-term employability while finding another job in the short run. Employability is the security that comes not from an employer but from oneself; built on an ever-expanding base of knowledge, skills, contacts, results, and reputation, it enables us to focus on opportunities. These opportunities might not appear as specific jobs but as circumstances that might be shaped into a job or a combination of jobs.

I pointed out that successful enterprises are built on realistic and flexible business plans that include a vision, mission, goals, and objectives; strategies to achieve them; and strengths to support the strategies. In similar fashion, the caller was also an enterprise ("Me, Inc." he laughed), and he, too, needed to take a more disciplined approach to career management.

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Previous | Table of Contents | Next

Chapter VII-6
Developing People Skills: A New IT Priority

Greg Scileppi

In the past few years, hiring managers have been actively seeking a certain breed of IT experts—those with a mixture of technical and interpersonal skills. The current move toward client/server environments is bringing IT staff, including data center personnel, network administrators, LAN specialists, and others—in closer contact with end users. Consequently, possessing strong people skills is one of the leading ingredients for success as a computer professional today.

According to a nationwide survey developed by RHI Consulting, nearly three out of four executives believe that IT professionals with effective interpersonal and communications skills are scarce. To fill this void and become the type of employee or consultant that managers need, data center professionals should consider taking stock of their skills, making adjustments when needed. They should not, however, approach the task of self-evaluation in a haphazard manner. Effective change, especially if the goal is to achieve long-term results, requires careful thought and persistent effort.

TAKING STOCK: UNDERSTANDING BUSINESS ISSUES

One of the best places to start a personal evaluation is with corporate knowledge. In the past, data center professionals may have been able to argue that understanding their company's business had only an indirect impact, at best, on their job responsibilities. Today, however, with the growing importance placed on information, the onus clearly rests on individual data center staff members and the technology they recommend and support to provide the company's working foundation. If they do not know the business well, their effectiveness in formulating solutions is severely limited. They may also err when dealing with colleagues and management, leading to diminished professional credibility within the organization.

When evaluating the strength of business knowledge, the individual should consider the following "big picture" questions:

- How much does he or she really know about the employer's business?
  - — What types of products or services does the company sell?
  - — Who are its target customers?
  - — How does the company stack up against the competition?
  - — Is the company's industry relatively flat or constantly evolving.?
- What role does each department, or division, play in its success?
- How do these groups typically interact?

Increasing the Knowledge of the Business

There are some immediate steps data center professionals can take to learn more about a company's products and goals. For example, one of the most valuable publications a public company has to offer is its annual report. Some firms make sure that every employee receives a report, especially if employee stock options are offered. If the company is a privately held organization, an annual report probably is not available. In that case, data center staff should ask for a copy of the company's backgrounder, a document that describes the company's goals, accomplishments, key products, and other related information. If nothing else is available, the company's marketing materials give a helpful overview.

At the very least, data center professionals should know their employers' mission statements. If they do not understand the strategy, they can seek help from; their interest in the organization is almost always be viewed positively. When data center managers are able to relate their companies' missions to their jobs, they are more likely to perform at their best, because that understanding gives greater perspective to the work and its implications. Managers also find it invaluable when checking that they are operating in a manner that supports their companies' goals.

BUILDING BRIDGES THROUGH COMMUNICATION

Once data center managers have increased their company knowledge and feel comfortable supporting the company mission, they are prepared to break through the stereotype that IT professionals don't spend much time interacting with others. In today's client/server environment, strong interpersonal skills in addition to technical acumen are more important than ever.

Although the data center speaks with people every day, the truth is that they may need to develop some new skills and techniques to improve their effectiveness. Contrary to what many believe, being a “people” person is not easy, and it does not come naturally to everyone. Like programming languages, networking topologies, or any other computer-based discipline, working with people is a skill that can be learned, practiced, and improved.

Previous	Table of Contents	Next

To access the contents, click the chapter and section titles.

Handbook of Enterprise Operations Management

(Publisher: CRC Press LLC)
Author(s): John Wyzalek, Consulting Editor
ISBN: 0849399661
Publication Date: 08/01/98

Search this book:

Table of Contents

Index

B

C

D

E

F

G

H

J

K

L

M

N

O

Q

R

S

T

U

V

W

Table of Contents