# GIS DATA STORAGE AND MANAGEMENT

**Biruk G.**

**March 2020**

# Introduction to DBMS

- **Databases and database systems have become an essential component of everyday life in modern society.**
- **In the course of a day, most of us encounter several activities that involve some interaction with a database.**

# Some examples

- **In the past few years, advances in technology have been leading to exciting new applications of database systems.**
  - ***Multimedia databases***
    - **can now store pictures, video clips, and sound messages.**
  - ***Data warehouses and on-line analytical processing (OLAP)*** **systems**
    - **are used in many companies to extract and analyze useful information from very large databases for decision making.**
  - ***Real-time and active database technology***
    - **is used in controlling industrial and manufacturing processes.**
  - ***Geographic information systems (GIS)***
    - **can store and analyze maps, weather data, and satellite images.**

# What is database and DBMS?

- **Databases are organized collections of interrelated data.**
- **Are managed by a software known as database management systems (DBMS) and shared by multiple applications**.

# WHY DBMS?

- A DBMS supports the storage and manipulation of *very large data sets.*

Some data sets are so big that storing them in text files or spreadsheet files becomes too awkward for use in practice. The result may be that finding simple facts takes minutes, and performing simple calculations perhaps even hours.

- A DBMS can be instructed to guard over some levels of *data correctness.*

For instance, an important aspect of data correctness is data entry checking: making sure that the data that is entered into the database is sensible data that does not contain obvious errors. Since we know in what study area we work, we know the range of possible geographic coordinates, so we can make the DBMS check them.

- A DBMS supports the *concurrent use* of the same data set by many users.

# WHY DBMS?

- A DBMS provides a high-level, *declarative query language*.

The most important use of the language is the definition of queries. A *query* is a computer program that extracts data from the database that meet the conditions indicated in the query. We provide a few examples below.

- A DBMS supports the use of a *data model*. A data model is a language with which one can define a database structure and manipulate the data stored in it.

- A DBMS includes *data backup* and *recovery* functions to ensure data availability at al times.

# GIS AND DBMS

- **Database management systems are an integrated and crucial components of most successful GIS.**
- **DBMSs are used to store, manipulate and retrieve data from a database.**
- **A key element in creating spatial database is database design using a variety of data modeling techniques.**

# SPATIAL DATABASE

- **Definition**
  - **A spatial database is a collection of spatially referenced data that acts as a model of reality**
    - **a database is a model of reality in the sense that the database represents a selected set or approximation of phenomena**
    - **these selected phenomena are deemed important enough to represent in digital form**
    - **the digital representation might be for some past, present or future time period (or contain some combination of several time periods in an organized fashion)**

# SPATIAL DATABASE

- **Fundamental database elements**
  - **Elements of reality modeled in a GIS database have two identities:**
    - **Entity - the element in reality**
    - **Object - the element as it is represented in the database**

# Spatial database

- **Entity**
  - **An entity is "a phenomenon of interest in reality that is not further subdivided into phenomena of the same kind"**
    - **e.g. a city could be considered an entity and subdivided into component parts but these parts would not be called cities, they would be districts, neighborhoods or the like**
    - **e.g. a forest could be subdivided into smaller forests**

# SPATIAL DATABASE

- **Object**
  - **An object is "a digital representation of all or part of an entity"**
  - **The method of digital representation of a phenomenon varies according to scale, purpose and other factors**
    - **e.g. a city could be represented geographically as a point if the area under consideration were continental in scale**
    - **the same city could be geographically represented as an area if we are dealing with a geographic database for a state or a county**

# GIS DATABASE REQUIREMENTS

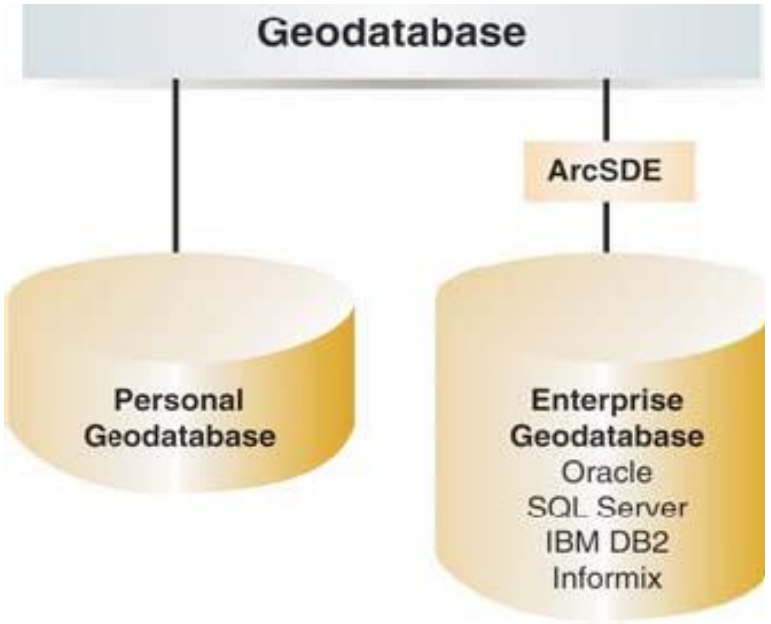
- **Scale to *large sizes* (multiple terabytes)**
- **Scale to *large number of users* (hundreds to thousands)**
- **Provide *advanced GIS data models* and behavior**
- **Maintain *spatial data integrity***
- **Deliver *fast data retrieval***
- **Support *long transactions* and GIS work flows**
- **Support *multiple uses and applications***
- **Proven to *work through real case* studies**

# Geodatabase – Introduction

- **The Geodatabase has three key aspects:**
  - **It is a comprehensive information model and a transaction model for GIS**
  - **It is the common application logic used in ArcGIS for accessing and working with all geographic data files and formats**
  - **It is a physical instance of a collection of datasets stored in a file system or DBMS**

# Geodatabase – Introduction

- **Geodatabases work across a range of DBMS architectures and file systems, come in many sizes, and have varying numbers of users.**
- **They can scale from small, single-user databases built on files up to larger work group, department, and enterprise databases accessed by many users.**
- **Two types of Geodatabase architectures are available: personal Geodatabases and multi-user Geodatabases.**



# Geodatabase - Personal

- **Personal Geodatabases use the Microsoft Jet Engine database file structure to persist GIS data in smaller databases.**
- **Personal Geodatabases are much like file-based folders and hold databases up to 2 GB in size.**
- **Microsoft Access is used to work with attribute tables in personal Geodatabases.**

# Geodatabases - Multi-user

- **Readily scale to extremely large sizes and numbers of users.**
- **Through many large Geodatabase implementations, it has been found that DBMSs are efficient at moving the type of large binary objects required for GIS data in and out of tables.**
- **In addition, GIS database sizes and the number of supported users can be much larger than GIS file bases.**

| Geodatabase type | DBMS | Notes |
| --- | --- | --- |
| Personal geodatabase | Microsoft Jet Engine (Access) | · Single-user editing<br>· 2 GB size limit<br>· No versioning support |
| Multiuser, versioned geodatabase | · Oracle<br>· Oracle with Spatial or Locator<br>· IBM DB2<br>· IBM Informix<br>· Microsoft SQL Server | · Requires ArcSDE Gateway<br>· Multiuser editing<br>· Version-based work flows<br>· Database size and number of users up to RDBMS limits |

# Geodatabase - Key database concepts

- **The Geodatabase architecture is based on a series of simple yet essential database concepts.**
- **The DBMS provides a simple, formal data model for storing and working with information in tables.**
- **Users tend to think of the DBMS as inherently open because the simplicity and flexibility of the generic relational data model enable it to support a broad array of applications.**

# Geodatabase - Key database concepts

- **Key DBMS concepts include:**
  - **Data is organized into tables.**
  - **Tables contain rows.**
  - **All rows in a table have the same columns.**
  - **Each column has a type, such as integer, decimal number, character, date, and so on.**
  - **Relationships are used to associate rows from one table with rows in another table. This is based on a common column in each table, the primary key or the foreign key.**
  - **Relational integrity rules exist for table-based dataset**

Feature class table

| Shape | ID | PIN | Area | Addr | Code |
|---|---|---|---|---|---|
| | 1 | 334-1626-001 | 7,342 | 341 Cherry Ct. | SFR |
| | 2 | 334-1626-002 | 8,020 | 343 Cherry Ct. | UND |
| | 3 | 334-1626-003 | 10,031 | 345 Cherry Ct. | SFR |
| | 4 | 334-1626-004 | 9,254 | 347 Cherry Ct. | SFR |
| | 5 | 334-1626-005 | 8,856 | 348 Cherry Ct. | UND |
| | 6 | 334-1626-006 | 9,975 | 346 Cherry Ct. | SFR |
| | 7 | 334-1626-007 | 8,230 | 344 Cherry Ct. | SFR |
| | 8 | 334-1626-008 | 8,645 | 342 Cherry Ct. | SFR |

Related ownership table

| PIN | Owner | Acq.Date | Assessed | TaxStat |
|---|---|---|---|---|
| 334-1626-001 | G. Hall | 1995/10/20 | $115,500.00 | 02 |
| 334-1626-002 | H. L Holmes | 1993/10/06 | $24,375.00 | 01 |
| 334-1626-003 | W. Rodgers | 1980/09/24 | $175,500.00 | 02 |
| 334-1626-004 | J. Williamson | 1974/09/20 | $135,750.00 | 02 |
| 334-1626-005 | P. Goodman | 1966/06/06 | $30,350.00 | 02 |
| 334-1626-006 | K. Staley | 1942/10/24 | $120,750.00 | 02 |
| 334-1626-007 | J. Dormandy | 1996/01/27 | $110,650.00 | 01 |
| 334-1626-008 | S. Gooley | 2000/05/31 | $145,750.00 | 02 |

# At the heart of any GIS there is always a strong database support


THANK YOU