Database as data store

At the heart of any GIS is the database

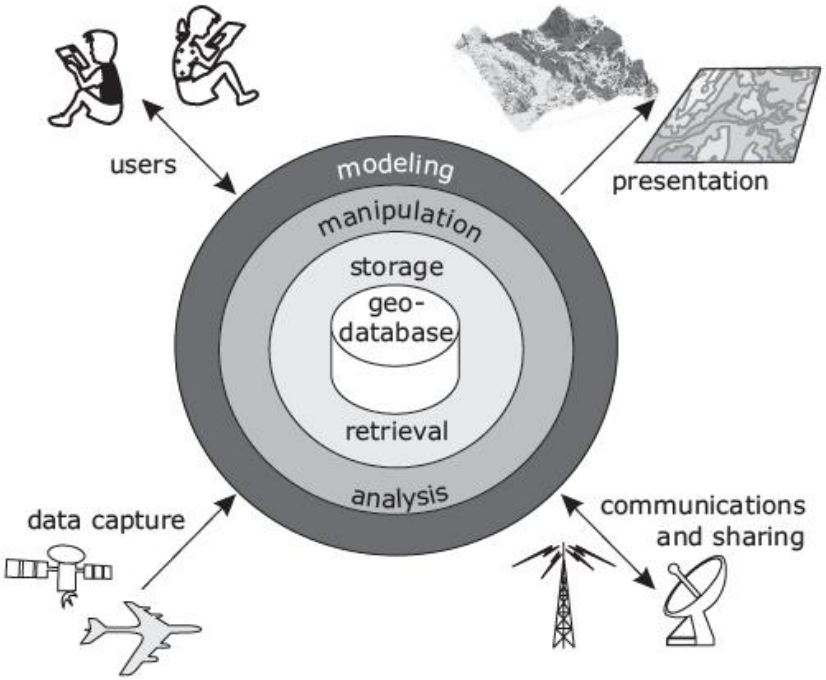


Figure 1.1: Schematic of a GIS

1.1.1 The shape of GIS

- A database is a repository of data that is logically related, but possibly physically distributed over several sites, and required to be accessed by many applications and users.
- A database is created and maintained using a general-purpose piece of software called a database management system (DBMS).

- A database is a repository of data that is logically related, but possibly physically distributed over several sites, and required to be accessed by many applications and users.
- A database is created and maintained using a general-purpose piece of software called a database management system (DBMS).
  PostgreSQL

Useful Database

Reliable:

- A database must be able to offer a continual uninterrupted service when required by users, even if unexpected events occur, such as power failures.
  - For example, a database must be able to contend with the situation where we deposit money into an automatic teller machine (ATM) and the power fails before our balance has been updated.

Useful Database

Reliable:

- A database must be able to offer a continual uninterrupted service when required by users, even if unexpected events occur, such as power failures.
  - For example, a database must be able to contend with the situation where we deposit money into an automatic teller machine (ATM) and the power fails before our balance has been updated.

Correct and consistent:

- Data items in the database should be correct and consistent with each other.
- This has been a problem with older file systems, when data held by one department contradicts data from another. While it is not possible to screen out all incorrect data, it is possible to control the problem to some extent, using integrity checking at input. Clearly, the more we know about the kinds of data that we expect, the more errors can be detected.

Technology proof :

- A database should evolve predictably, gracefully, and incrementally with each new technological development.
- Both hardware and software continue to develop rapidly, as new processors, storage devices, software, and modeling techniques are invented.
- Database users should be insulated from the inner workings

of the database system. For example, when we check our balance at the ATM, we do not really want to know that on the previous night new storage devices were introduced to support the database.

Technology proof :

- A database should evolve predictably, gracefully, and incrementally with each new technological development.
- Both hardware and software continue to develop rapidly, as new processors, storage devices, software, and modeling techniques are invented.
- Database users should be insulated from the inner workings

of the database system. For example, when we check our balance at the ATM, we do not really want to know that on the previous night new storage devices were introduced to support the database.

Secure:

- A database must allow different levels of authorized access and prevent unauthorized access.
- For example, we may be allowed to read our own, but no one else’s, bank balance (read access). At the same time we may not be allowed to change our bank balance (write access).

Data Capture

The process of collecting data from observations of the physical environment is termed data capture

- **The geodatabase** (a database or file structure used primarily to store, query, and manipulate spatial data) is a powerful **data model** for storing and managing your GIS data in one place.

- Through the geodatabase, you can import different datasets from various sources and use them in your GIS analysis.

Advantages of using a geodatabase

- **Centralized repository:** All data is stored in the same database, as opposed to in many separate files.

Advantages of using a geodatabase

- **Centralized repository:** All data is stored in the same database, as opposed to in many separate files.
- **Scalable data model:** As your GIS needs increase, you can migrate data from one geodatabase to an upgraded format that allows for more users and editors.

Advantages of using a geodatabase

- **Centralized repository:** All data is stored in the same database, as opposed to in many separate files.
- **Scalable data model:** As your GIS needs increase, you can migrate data from one geodatabase to an upgraded format that allows for more users and editors.
- **Increased data integrity:** You can create spatial and attribute behaviors to facilitate editing, help eliminate data entry errors, and maintain spatial and attribute relationships between your data.

Advantages of using a geodatabase

- **Centralized repository:** All data is stored in the same database, as opposed to in many separate files.
- **Scalable data model:** As your GIS needs increase, you can migrate data from one geodatabase to an upgraded format that allows for more users and editors.
- **Increased data integrity:** You can create spatial and attribute behaviors to facilitate editing, help eliminate data entry errors, and maintain spatial and attribute relationships between your data.
- **Support for imagery:** Mosaic datasets in the geodatabase allow you to manage multiple images as one.

Mosaic datasets

- Large collections of raster data
- It can be a heterogeneous collection of raster datasets (individual images) with multiple formats, sources, data types, resolutions, number of bands, pixel depths, file sizes, and coordinate systems.

Types of geodatabases

- **File geodatabase**

A file geodatabase is a collection of GIS datasets and is stored in a file system folder. File geodatabases work across operating systems and can store individual datasets up to 1 terabyte (TB) in size by default. The size limit can be increased with keywords to 256 TB. File geodatabases support multiple people editing different feature classes or tables at the same time.

Because of their increased capacity and editing capabilities, file geodatabases are the recommended type for ArcGIS Pro users.

Types of geodatabases

- **File geodatabase**

A file geodatabase is a collection of GIS datasets and is stored in a file system folder. File geodatabases work across operating systems and can store individual datasets up to 1 terabyte (TB) in size by default. The size limit can be increased with keywords to 256 TB. File geodatabases support multiple people editing different feature classes or tables at the same time.

Because of their increased capacity and editing capabilities, file geodatabases are the recommended type for ArcGIS Pro users.

- **Enterprise geodatabase**

Enterprise geodatabases support versioning and replication. They require a database management system (DBMS), such as DB2, Informix, Microsoft SQL Server, Oracle, or PostgreSQL.

Typically found in larger organizations, enterprise geodatabases support multiple users viewing and editing the GIS database at the same time. In fact, enterprise geodatabases allow for unlimited editors and database size.

- **Mobile geodatabase**

Mobile geodatabases—A mobile geodatabase is stored in an SQLite database that is entirely contained in a single file and has a `.geodatabase` extension.

Benefits

Mobile geodatabases are stored in an SQLite database, giving them the following advantages:

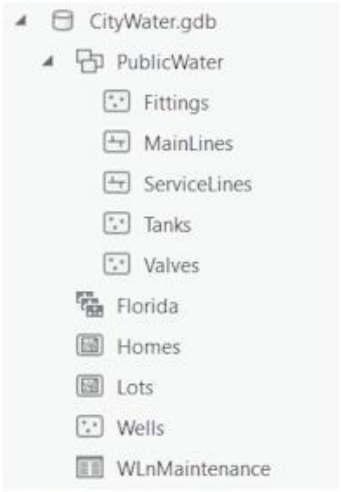
- SQLite is a widely available, stable, and trusted database.
- SQLite is open source in the public domain, so no licensing is required.
- SQLite databases are supported cross-platform and are stored in a single file on disk, making them portable and an efficient data exchange format.
- SQLite is a full-featured relational database allowing for querying and reporting workflows and supports operations such as views and indexes.
- SQLite is interoperable and is ubiquitous in mobile app development.

More detailed comparison of geodatabases

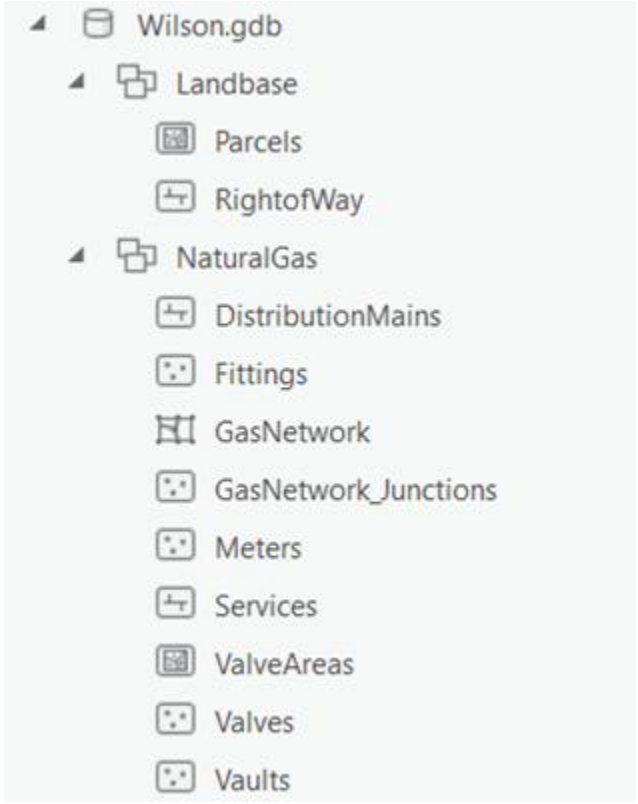
There are two types of geodatabases, as detailed in the following table. Each type has unique specifications and capabilities to meet specific organizational workflows.

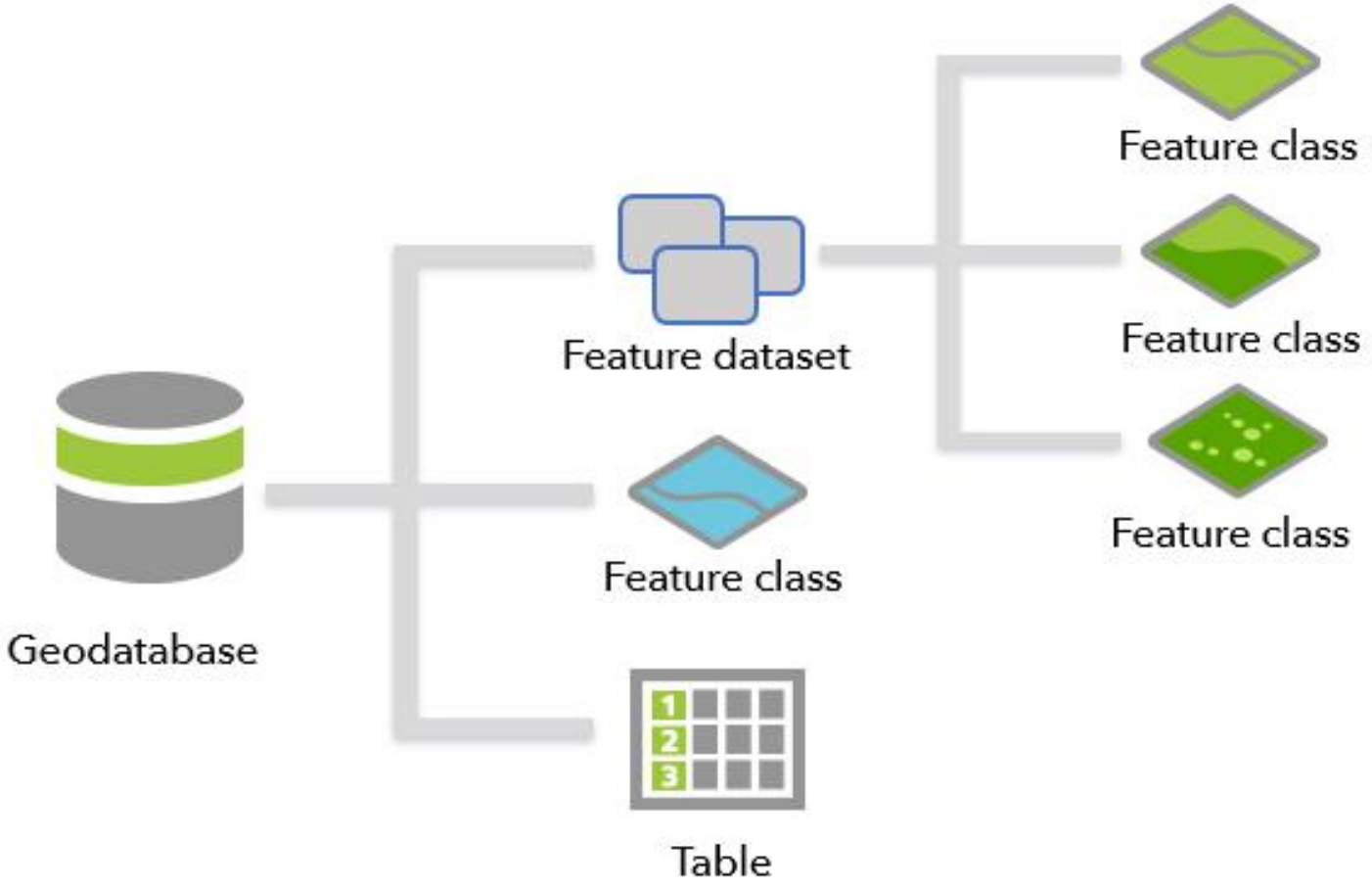
Differences between types of geodatabases		
	File	Enterprise
Purpose	Geodatabase replication	Large organization
Storage mechanism or platform	SQL Server Express (.mdf)	Oracle, DB2, SQL Server, Informix, PostgreSQL
Storage limit	1 TB (256 TB with keywords)	Unlimited
Connections	Three viewers, one editor (same functionality as multiuser geodatabases)	Unlimited
Additional benefits	Only requires a Standard ArcGIS Pro license	Use of DBMS tools, memory management, security

Components



This CityWater geodatabase contains various components.




Feature class
Feature dataset
Feature class
Feature class
Feature class
Geodatabase
Table

Feature datasets and feature class

- **A feature class is a collection of geographic features with the same geometry type, attributes, and spatial reference.**
- **A feature dataset is composed of feature class**

Geodatabase Schema

- ➔ The **schema** of the geodatabase (its structure or design) defines **the physical structure** along with **rules, relationships, and properties of each dataset** within the geodatabase.
- ➔ **The geodatabase schema is a way to model GIS data**, both spatial and nonspatial, in an organizational container that represents features in the real world.
- ➔ You can also copy templates into your geodatabase to meet specific project goals.
- ➔ The individual feature classes can contain specific fields that are relevant to your organization and analysis.

Database Solutions

- **Lightweight**
  - for one user or a small number of users
  - a lightweight database is a good place to start
  - they have no servers and are very nimble
    - SQLite

Database Solutions

- **Centralized**
  - If you expect tens, hundreds, or thousands of users and applications to use a database simultaneously, lightweight databases are not going to cut it.
  - Runs on a server
  - Handles high volume of traffic
    - MySQL
    - Microsoft SQL Server
    - Oracle
    - **PostgreSQL**
    - Teradata
    - IBM DB2
    - MariaDB

Centralized Solution

- ★ You can install some of these solutions on any computer and turn that computer into a server.
- ★ You can then connect users' computers (also known as clients) to the server so they can access the data.

RDBMS (R = Relational)

- An RDBMS is simply a type of database that holds one or more tables that may have relationships to each other.
- These tables can have relationships to each other, such as an ORDER table that refers to a CUSTOMER table for customer information
- When we go through the ORDER table, we can use the CUSTOMER_ID to look up the customer information in the CUSTOMER table.
- This is the fundamental idea behind a “relational database,” where tables may have fields that point to information in other tables.
- VLOOKUP in Excel is an example of this concept

Database Normalization

- Why Separate Tables?
- Normalization, which is separating the different types of data into their own tables rather than putting them in one table.
- If we had all information in a single table, it would be redundant, bloated, and very difficult to maintain

	NAME	REGION	STREET ADDRESS	CITY	STATE	ZIP	ORDER ID	ORDER DATE	SHIP DATE	ORDER QTY	SHIPPED
1	LITE Industrial	Southwest	729 Ravine Way	Irving	TX	75014	1	2015-04-15	2015-04-18	450	false
2	Re-Barre Construction	Southwest	9043 Windy Dr	Irving	TX	75032	2	2015-04-18	2015-04-21	600	false
3	Re-Barre Construction	Southwest	9043 Windy Dr	Irving	TX	75032	3	2015-04-20	2015-04-23	300	false
4	Marsh Lane Metal Works	Southeast	9143 Marsh Ln	Avondale	LA	79782	4	2015-04-18	2015-04-22	375	false
5	Re-Barre Construction	Southwest	9043 Windy Dr	Irving	TX	75032	5	2015-04-17	2015-04-20	500	false

Figure 2-3. A table that is not normalized