

Lesson 1: Introduction to Database Systems

**Tozammel Hossain
Assistant Professor
Information Science @UNT**

Announcements

- **Quiz 1 is due this Sun.**
 - Do you think Mon night is a good option as a deadline for the upcoming weeks?
- **Lesson 2 will be release on Mon**

Agenda

- **Data vs Information**
- **Database and Database system**
- **Components of a database**
- **Database design significance**
- **Database management systems**

Data vs Information

- **Three key differences between data and information presented in a tabular form**

| Characteristic | Data | Information |
|----------------|-----------------------------------------|---------------------------------------------|
| Definition | Raw facts and figures, unprocessed. | Processed, organized, and meaningful facts. |
| Context | Lacks context and relevance on its own. | Provides context, insights, and relevance. |
| Purpose | Typically used as a basis for analysis. | Used for decision-making and knowledge. |

Raw Facts to Information



DATA ENTRY

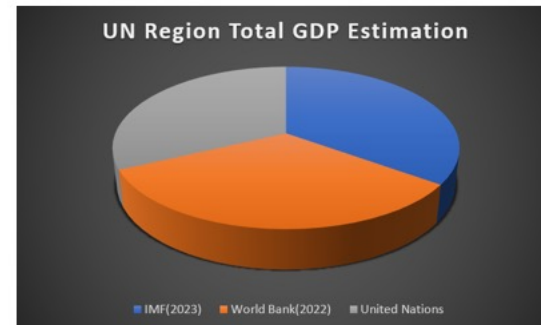
| Country | UN Region | IMF | | World Bank | | United Nations | |
|----------------|-----------|------------|------|------------|------|----------------|------|
| | | Estimate | Year | Estimate | Year | Estimate | Year |
| United States | Americas | 26,854,599 | 2023 | 25,462,700 | 2022 | 23,315,081 | 2021 |
| China | Asia | 19,373,586 | 2023 | 17,963,171 | 2022 | 17,734,131 | 2021 |
| Japan | Asia | 4,409,738 | 2023 | 4,231,141 | 2022 | 4,940,878 | 2021 |
| Germany | Europe | 4,308,854 | 2023 | 4,072,192 | 2022 | 4,259,935 | 2021 |
| India | Asia | 3,736,882 | 2023 | 3,385,090 | 2022 | 3,201,471 | 2021 |
| United Kingdom | Europe | 3,158,938 | 2023 | 3,070,668 | 2022 | 3,131,378 | 2021 |
| France | Europe | 2,923,489 | 2023 | 2,782,905 | 2022 | 2,957,880 | 2021 |
| Italy | Europe | 2,169,745 | 2023 | 2,010,432 | 2022 | 2,107,703 | 2021 |
| Canada | Americas | 2,089,672 | 2023 | 2,139,840 | 2022 | 1,988,336 | 2021 |
| Brazil | Americas | 2,081,235 | 2023 | 1,920,096 | 2022 | 1,608,981 | 2021 |

a) Data Entry

b) Raw Data

| UN Region | Country | IMF(2023) | World Bank(2022) | United Nations |
|-----------|----------------|-----------------|------------------|-----------------|
| Americas | United States | 31025506 | 29522636 | 26912398 |
| | Canada | | | |
| | Brazil | | | |
| Asia | China | 27520206 | 25579402 | 25876480 |
| | Japan | | | |
| | India | | | |
| Europe | Germany | 12561026 | 11936197 | 12456896 |
| | United Kingdom | | | |
| | France | | | |
| | Italy | | | |
| | Total | 71106738 | 67038235 | 65245774 |

c) Information in Summary Format



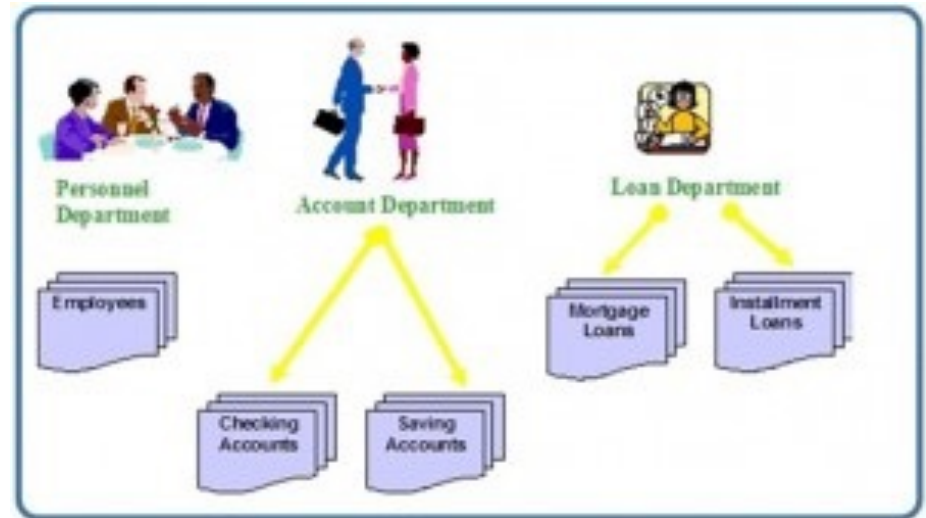
d) Information in Visual Format

Before the Invent of Database (slide 1 of 3)

- **Sumerian prescription tablet (2400 BC)**
- **Business transactions with pen + paper**
- **Product Inventory**
- **Card catalogue**

Before the Invent of Database (slide 2 of 3)

- **Computer data management has come a long way**
- **File-based System**
 - keep information by storing it in permanent files
 - Less expensive



Before the Invent of Database (slide 3 of 3)

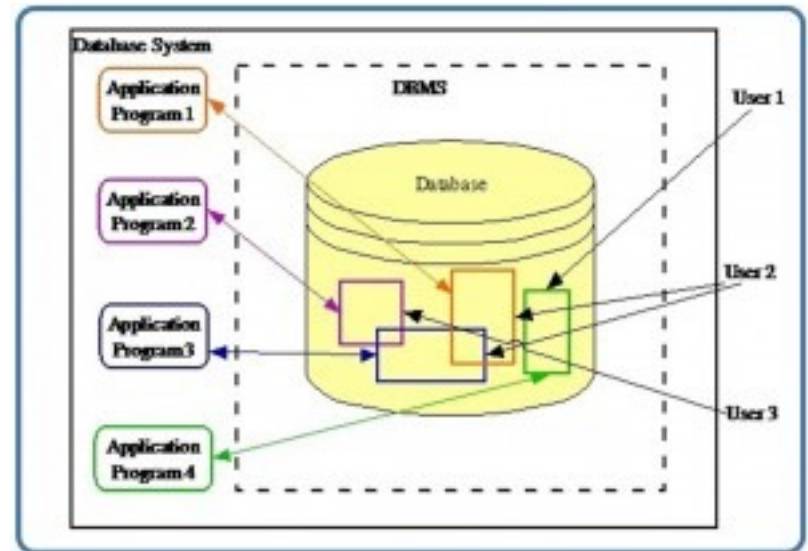
- **Cons**
 - Data redundancy
 - Data isolation
 - Integrity problems
 - Security problems
 - Concurrency access

Why use a database?

- **To remove the cons mentioned**
- **Data organization**
 - Offers structured storage, guaranteeing logical data organization
 - Essential for effective data management and retrieval
 - **Avoids chaos when dealing with large data**
- **Data security**
 - To safeguard sensitive data, databases include robust security mechanisms
 - **E.g., restrict access, encrypt data, and maintain audit trails, reducing the risk of data breaches and ensuring compliance with privacy rules.**

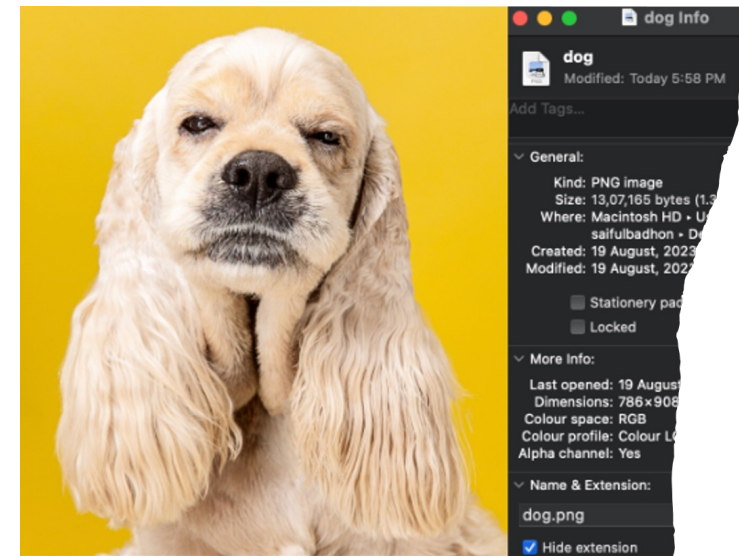
What is a Database? (slide 1 of 2)

- **A database is a shared collection of related data**
 - used to support the activities of a particular organization.
- **A database can be viewed as a repository of data that is defined once and then accessed by various users**



What is a Database? (slide 2 of 2)

- **A database manages two types of data**
 - Data from end users
 - **Unprocessed, fundamental information that is significant to end users**
 - Example of end users?
 - Metadata
 - **Information on how end-user data is compiled and handled, as well as information on the data's attributes and connections**



A Brief History of Modern Databases

- **IMS by IBM**
 - Information Management System
 - A tree-like hierarchical model
 - E.g. NASA used this system for a lunar lander
- **Network model by Charles Bachman**
 - A graph-like model
- **Relational model by Ted Codd at IBM**
 - Collection of tables; 1970
 - C J Date supported the idea
- **Michael Stonebraker at UC-Berkeley**
 - Ingress relational model (freely available)
- **System R by IBM in 1975**
 - Used SQL by Chamberlin and Boyce
- **Larry Ellison formed Oracle in 1977**
 - Invested \$2000
 - Miner and Oates
- **DB2 by IBM in 1983**

Properties of a Database (slide 1 of 2)

- **A representation of some aspect of the real world or a collection of data elements (facts) representing real-world information**
- **Logical, coherent, and internally consistent**
- **Designed, built, and populated with data for a specific purpose**

Properties of a Database (slide 2 of 2)

- **Each data item is stored in a field**
- **A combination of fields makes up a table**
 - E.g., each field in an employee table contains data about an individual employee.
- **A database can contain many tables**

Applications of Databases

- **Business Operations**
- **E-commerce**
- **Healthcare**
- **Banking**
- **Education**
- **Manufacturing**
- **Social Media**
- **Government**
- **Scientific Research**
- **Libraries**
- **Telecommunications**
- **Online Content**
- **Retail**
- **Transportation**
- **Travel**

Database types based on location

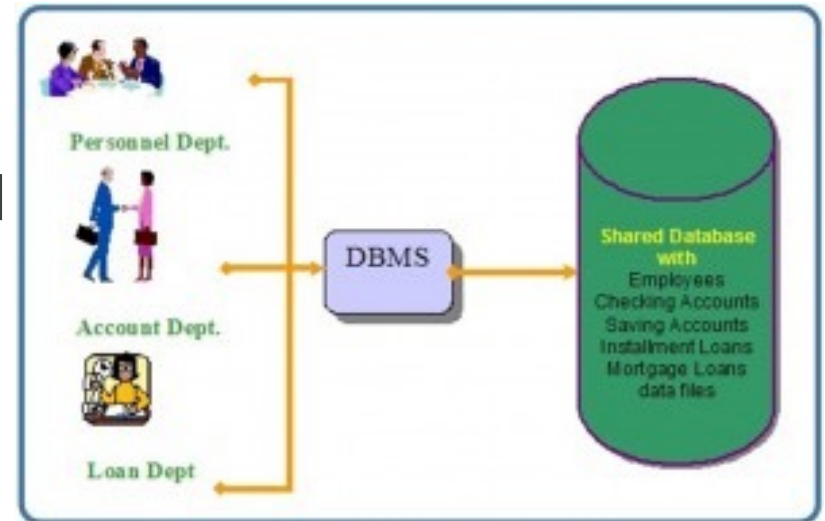
- **Centralized database**
 - Information is kept in one place through a central database
 - Prone to a single point of failure
- **Distributed database**
 - data dispersed over several sites
 - Require better maintenance
- **Cloud databases**
 - created and managed using cloud data services that provide performance information particular to each database

Database types based on data arrangements

- **Raw unstructured database**
 - Data is kept in its original format
 - Hard to retrieve information
- **Structured database**
 - Data is preprocessed
 - Data is stored following a schema
 - Each entity is well defined
- **Semi-structured database**
 - Data has undergone some processing

Database Management System (DBMS)

- **DBMS is a collection of programs that**
 - enables users to create and maintain databases
 - controls all access to them.
- **The primary goal of a DBMS:**
 - provide an environment that is both convenient and efficient for users to retrieve and store information.



Why use a DBMS?

- **Managing information means taking care of it so that it works for us**
- **The information is no longer subject to accidental disorganization**
- **More accessible and integrated with the rest of our work**
- **Allows us to become strategic users of the data we have**

DBMS Components (slide 1 of 3)

- **Hardware**

- The actual parts of the system, including as computers, printers, storage units, networks, and other devices, are referred to as hardware

- **Software**

- Operating system software, which also enables all other software, including DBMS, to operate on computers, manages all physical components
- DBMS software is used to administer the database within the database system
- Application programs and utility programs are required to access and alter data in the DBMS

DBMS Components (slide 2 of 3)

- **People**

- All users of the database system are included in this component
- Based on their primary job duties, there are five user types:
 - system administrators
 - database administrators
 - database designers
 - system analysts and programmers
 - end users

DBMS Components (slide 3 of 3)

- **Procedures**

- The rules and regulations that control the development and use of the database system are known as procedures
- A critical component of the system
 - Often disregarded
- Enforces the standards by which business is conducted both within the company and with customers
- Ensures a systematic way to monitor and audit the data that enters the database and the information that is generated from those data

- **Data**

- The collection of information kept in a database is referred to as data
- Determining which data to add to the database and how to organize that data is an essential component of the database designer's job

Characteristics and Benefits (slide 1 of 11)

- **Self-describing nature of a database system**
 - not only contains data but also metadata
 - **metadata which defines and describes the data and relationships between tables in the database**
 - separation of data and metadata
 - a key difference with the traditional file-based system
 - the data definition is part of the file-based system

Characteristics and Benefits (slide 2 of 11)

- **Insulation between program and data**
 - File-based system:
 - **the structure of the data files is defined in the application programs**
 - **Issues:**
 - if a user wants to change the structure of a file, all the programs that access that file might need to be changed as well
 - Database
 - **the data structure is stored in the system catalogue and not in the programs**
 - one change is all that is needed to change the structure of a file

Characteristics and Benefits (slide 3 of 11)

- **Support for multiple views of data**
 - A database supports multiple views of data
 - A view is a subset of the database
 - Could be defined and dedicated for particular users of the system
 - Multiple users in the system might have different views of the system
 - Each view might contain only the data of interest to a user or group of users

Characteristics and Benefits (slide 4 of 11)

- **Sharing of data and multiuser system**
 - allow many users to access the same database at the same time
 - this access is achieved through features called concurrency control strategies
 - These strategies ensure that the data accessed are always correct and that data integrity is maintained
 - The design of modern multiuser database systems is a great improvement from those in the past which restricted usage to one person at a time

Characteristics and Benefits (slide 5 of 11)

- **Control of data redundancy**
 - each data item is stored in only one place in the database
 - in some cases, data redundancy still exists to improve system performance
 - Such redundancy is controlled by application programming

Characteristics and Benefits (slide 6 of 11)

- **Data sharing**

- The integration of all the data, for an organization, within a database system has many advantages
 - allows for data sharing among employees and others who have access to the system
- gives users the ability to generate more information from a given amount of data than would be possible without the integration

Characteristics and Benefits (slide 7 of 11)

- **Enforcement of integrity constraints**
 - A database constraint is a restriction or rule that dictates what can be entered or edited in a table
 - E.g., a postal code using a certain format or adding a valid city in the City field
 - There are many types of database constraints
 - Data type
 - Data uniqueness

Characteristics and Benefits (slide 8 of 11)

- **Restriction of unauthorized access**
 - Not all users of a database system will have the same accessing privileges
 - For example, one user might have read-only access (i.e., the ability to read a file but not make changes), while another might have read and write privileges
 - For this reason, a database management system should provide a security subsystem to create and control different types of user accounts and restrict unauthorized access

Characteristics and Benefits (slide 9 of 11)

- **Data independence**

- Another advantage of a database management system is how it allows for data independence
 - the system data descriptions or data describing data (metadata) are separated from the application programs
- This is possible because changes to the data structure are handled by the database management system and are not embedded in the program itself

Characteristics and Benefits (slide 10 of 11)

- **Transaction processing**
 - This feature ensures that data remains consistent and valid during transaction processing even if several users update the same information
- **Provision for multiple views of data**
 - By its very nature, a DBMS permits many users to have access to its database either individually or simultaneously. It is not important for users to be aware of how and where the data they access is stored

Characteristics and Benefits (slide 11 of 11)

- **Backup and recovery facilities**
 - Backup and recovery are methods that allow you to protect your data from loss
 - The database system provides a separate process, from that of a network backup, for backing up and recovering data
 - If a computer system fails in the middle of a complex update process, the recovery subsystem is responsible for making sure that the database is restored to its original state

Database architecture

- **1-tier architecture**
 - Database server+ an application on a single machine
- **2-tier architecture**
 - Database server + some applications on different machines
 - Direct communication between them
- **3-tier architecture**
 - Database server + an application server + multiple applications
 - E.g., web-based applications (web server)

Key Terms (slide 1 of 7)

- **concurrency:**
 - the ability of the database to allow multiple users access to the same record without adversely affecting transaction processing
- **data element:**
 - a single fact or piece of information
- **data inconsistency:**
 - a situation where various copies of the same data are conflicting

Key Terms (slide 2 of 7)

- **data isolation:**
 - a property that determines when and how changes made by one operation become visible to other concurrent users and systems
- **data integrity:**
 - refers to the maintenance and assurance that the data in a database are correct and consistent
- **data redundancy:**
 - a situation that occurs in a database when a field needs to be updated in more than one table

Key Terms (slide 3 of 7)

- **database approach:**
 - allows the management of large amounts of organizational information
- **database:**
 - a shared collection of related data used to support the activities of a particular organization
- **database management software:**
 - a powerful software tool that allows you to store, manipulate and retrieve data in a variety of ways

Key Terms (slide 4 of 7)

- **file-based system:**
 - an application program designed to manipulate data files
- **database management system (DBMS):**
 - a collection of programs that enables users to create and maintain databases and control all access to them
- **table:**
 - a combination of fields

Key Terms (slide 5 of 7)

- **concurrency control strategies:**
 - features of a database that allow several users access to the same data item at the same time
- **data type:**
 - determines the sort of data permitted in a field, for example numbers only
- **data uniqueness:**
 - ensures that no duplicates are entered

Key Terms (slide 6 of 7)

- **database constraint:**
 - a restriction that determines what is allowed to be entered or edited in a table
- **metadata:**
 - defines and describes the data and relationships between tables in the database
- **read and write privileges:**
 - the ability to both read and modify a file

Key Terms (slide 7 of 7)

- **read-only access:**
 - the ability to read a file but not make changes
- **self-describing:**
 - a database system is referred to as self-describing because it not only contains the database itself, but also metadata which defines and describes the data and relationships between tables in the database
- **view:**
 - a subset of the database