# DATABASE DESIGN AND DEVELOPMENT (IT 2140)

## LECTURE 01- INTRODUCTION TO DBMS, DATABASE DESIGN PROCESS AND DATA-INTENSIVE APPLICATIONS

# LECTURE CONTENT

- Define key database-related terminologies

- Introduction to databases, DBMS and their benefits

- Database design process

- Data-intensive applications

# LEARNING OUTCOMES

- Explain what is a database and a DBMS.

- Identify situations where using a database would be beneficial.

- Explain the database design process.

- Identify the importance of DBMSs in data-intensive applications.

# WHAT IS DATA VS INFORMATION?

- Data refers to raw, unprocessed facts or values that, on their own, may not carry meaning. It can be in the form of numbers, text, images, or symbols.

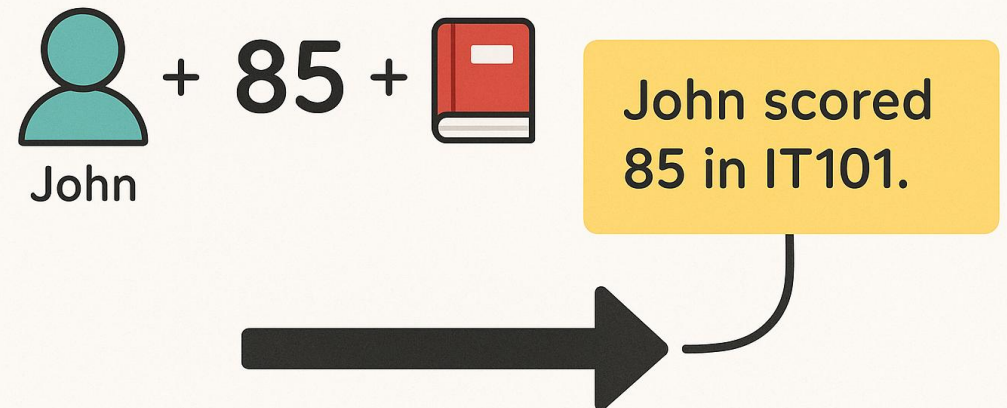- Information is data that has been processed, organized, or structured in a way that makes it meaningful and useful.
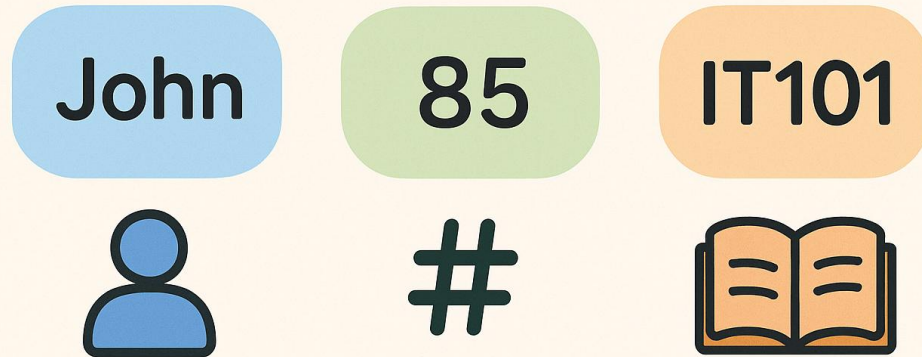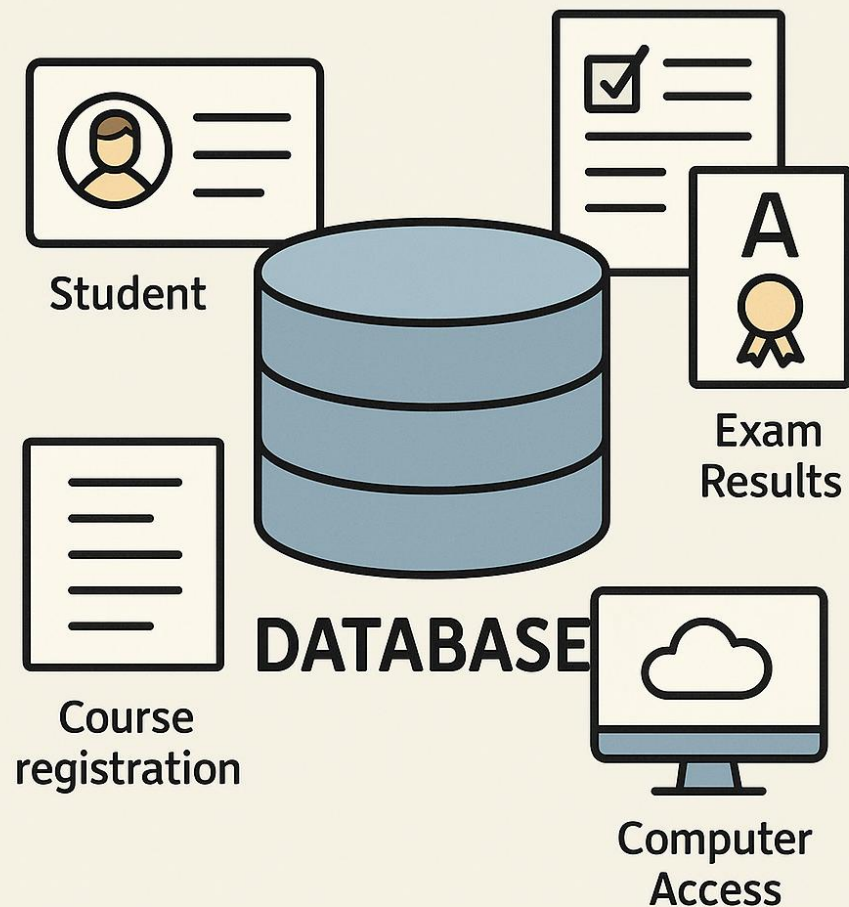
# WHAT IS A DATABASE & A DBMS?

- A database is a collection of related data.

- A database management systems (DBMS) is a general-purpose software system that facilitates the processes of defining, constructing, manipulating, and sharing databases among various users and applications.

  - What do you understand by defining, constructing, manipulating and sharing databases?

# WHAT IS A DATABASE & A DBMS?
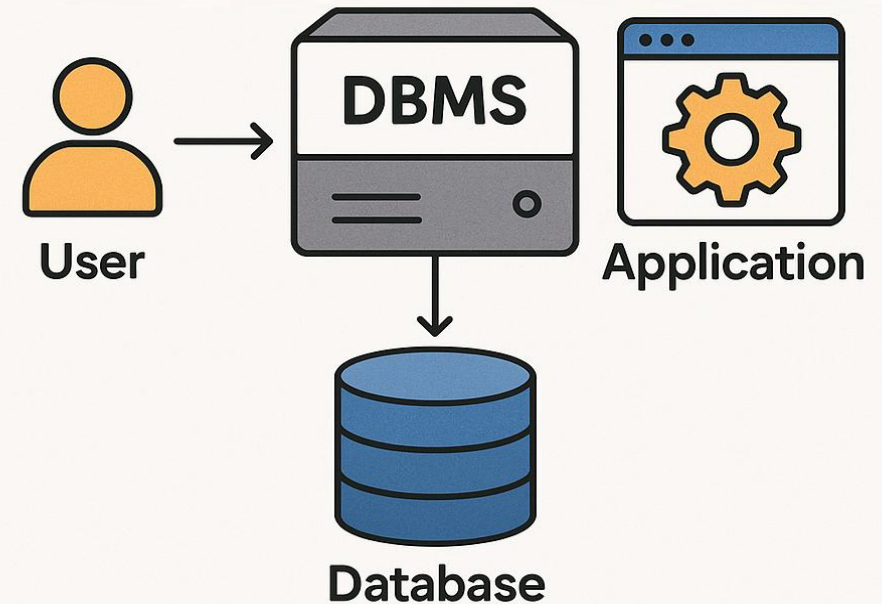
# ❓ QUESTION FOR STUDENTS

Can you identify one real-world place or system where databases are used? For each example,

Explain:

- What kind of data is stored ?

- Who uses the data ?

- Why a database is necessary in that context?

# DATABASE SYSTEM ENVIRONMENT

- Database system environment normally can be considered to have five major parts.
  1. Hardware
  2. Software
  3. People
  4. Procedures
  5. Data

# DATABASE APPROACH VS FILE PROCESSING SYSTEM

- Database system is of self-describing nature

- Database enables insulation between programs and data, and data abstraction

- Databases support of multiple views of the data

- Sharing of data and multiuser transaction processing

- Security and access levels

# ADVANTAGES OF DATABASE APPROACH

- **Data independence** – application programs are independent of the way the data is structured and stored

- **Efficient data access** – make the data retrieval efficient using sophisticated techniques to store and access data

- **Enforcing integrity** – provide capabilities to define and enforce constraints

  - Ex: Data type for a name should be string

- **Restricting unauthorized access** – allows only authorized users to access data

- **Providing backup and recovery** – protects users from the effect of system failures

- **Concurrent access** – simultaneous access of data by different users

# DATABASE DESIGN

Why is Database Design important?

Without careful planning you may create a database that…

- contains unnecessary data which occupies the storage space

- omits data required to create critical reports

- takes a considerable time to respond to user questions

- produces results that are incorrect or inconsistent

- is unable to accommodate changes in the user's requirements

A poorly designed database will require more time in the long-term…!

# DATABASE DESIGN PROCESS

- There are six main phases of the process to develop a database
  - Requirement collection and analysis
  - Conceptual database design
  - Logical database design
  - Schema refinement
  - Physical database design
  - Security design

# STEP 1: REQUIREMENT COLLECTION AND ANALYSIS

The goals of the requirements analysis are:

- to determine the **data requirements** of the database in terms of objects

- to identify and describe the information about these objects

- to identify the relationships among these objects

- to identify different transactions that will be performed on the database

- to identify performance, integrity, security or administrative constraints to be imposed on the database

- to identify design and implementation constraints if any (ex: specific technologies, hardware and software, programming languages, policies, standards, etc…)

# STEP 1: REQUIREMENT COLLECTION AND ANALYSIS

Information needed for the requirements analysis can be gathered in several ways:

- **Review of existing documents** –

Can become familiar with the organization/ activity you need to model by reviewing the existing documentation. Ex: existing forms and reports, written guidelines, job descriptions, personal narratives, etc…

- **Interviews with end users** –

Can organize individual/group meetings with the end users. Can use a blackboard, flip chart, or overhead transparencies to record information gathered from the interviews.

- **Review of existing automated systems** –

Can review the system design specifications and documentation, if the organization already has an automated system

# STEP 2: CONCEPTUAL DATABASE DESIGN

- The information gathered in the requirements analysis phase is used to create a:

- high-level description of the data in a conceptual data model or *Semantic Data Model.*

Eg. ER Model

# STEP 3: LOGICAL DATABASE DESIGN

- In this step, we determine the DBMS to implement the database & also the data model.

- We utilize the conceptual schema created in the previous step and convert it into a schema of a particular data model* (e.g. Relational Data Model )

- *we will cover this in the next two lectures.

# STEP 4: SCHEMA REFINEMENT

- The schema created by the logical database design phase is further refined for potential problems such as redundancies.

# STEP 5: PHYSICAL DATABASE DESIGN

- In this step, performance criteria are taken into consideration and further enhancements to the schema & creation of indexes are considered.

# STEP 6: SECURITY DESIGN

- Different user groups and their roles are identified.
- Appropriate levels of access are then provided to the data ensuring that users have access to only the necessary data.

Eg. Bank

Customer   – read access

Teller         – read/update (limited)
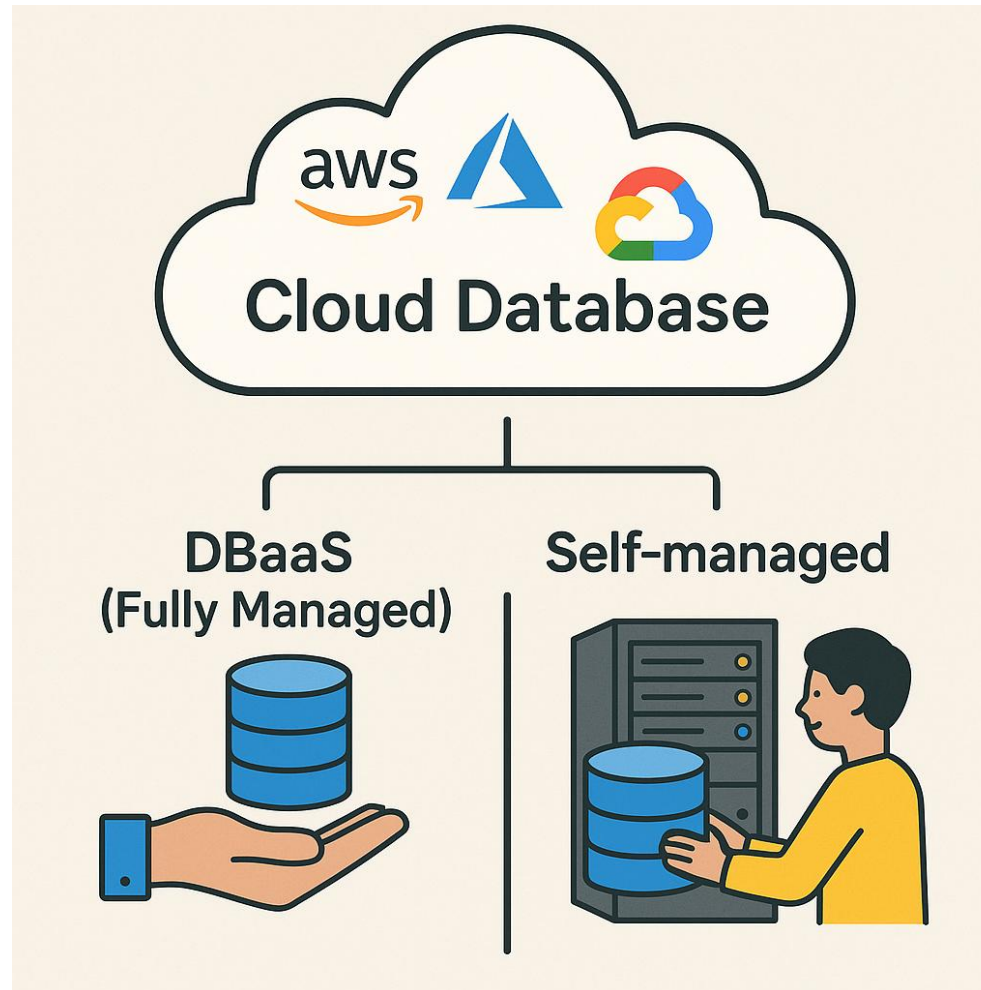
Manager     – read/update

# DATABASE HOSTING

- Database hosting defines where and how a database is deployed, stored, and accessed.

- Selecting the appropriate hosting type depends on:
  - Scalability
  - Cost
  - Security
  - Performance

# TYPES OF DATABASE HOSTING

- On-Premise Hosting

- Cloud Hosting

# WHAT IS CLOUD DATABASE?

# EXAMPLES OF CLOUD DATABASE

| Cloud Database | Examples |
|---|---|
| **Relational Cloud Databases** | AWS RDS, Azure SQL Database, Google Cloud SQL |
| **NoSQL Cloud Databases** | MongoDB Atlas, AWS DynamoDB, Google Firestore, Azure Cosmos DB |

# THINK–PAIR–SHARE ACTIVITY

- Cloud Databases – Advantages and Challenges

# DATA-INTENSIVE APPLICATIONS

What are Data-Intensive Applications?

- Focus on handling large volumes of data rather than performing complex computations.
- Involve extensive data storage, retrieval, and processing.
- Require fast and consistent access to data.
- Must be scalable to handle growing data and user loads.
- Common in systems with real-time access and multiple concurrent users.
- Depend heavily on reliable databases and DBMSs for performance and stability.

# EXAMPLES OF DATA-INTENSIVE APPLICATIONS

| Domain | Example Application | Types of Data Handled |
|---|---|---|
| Social Media | Facebook, Instagram | Posts, images, comments, likes, user profiles |
| E-Commerce | Amazon, eBay | Product listings, transactions, user carts, reviews |
| Banking | Online banking apps | Customer info, transactions, loan history, statements |
| Education | Learning Management Systems (LMS), like Moodle | Grades, courses, materials, attendance |
| Search Engines | Google, Bing | Crawled pages, user queries, ranking metadata |

# WHY IS A DBMS CRITICAL FOR DATA-INTENSIVE APPLICATIONS?

| DBMS Feature | How It Helps |
|---|---|
| Data Storage & Retrieval | Efficient storage of huge volumes of structured data |
| Indexing | Fast lookups and search operations |
| Transactions & Concurrency | Handles simultaneous users and operations safely |
| Security & Access Control | Ensures authorized access to sensitive data |
| Backup & Recovery | Prevents data loss in case of failure |
| Query Optimization | Speeds up processing of complex queries |
| Scalability (esp. in Cloud) | Supports growth in data, users, and workloads dynamically |

# HOMEWORK QUESTION

- Identify and describe five (5) modern technologies used in data-intensive applications.
- For each technology, include:
  - The name of the technology
  - Its purpose or function
  - One real-world example where it is used

# WHAT YOU HAVE TO DO BY NEXT WEEK

- Try out the self-test questions on the course web.

- Try out tutorial and bring the answers to the class.