

## 1.1.1 Data and Databases and Database Users

## \* Data and Information.

## • Data

→ Raw facts

→ Unprocessed facts

→ Data is building block of information

## • Information

→ It is the output of processing raw data.

→ It is produced after processing data.

## • Files

→ A data file is a computer file which stores data to use by a computer application or system.

→ It contains information used as input or written as output by some other software program.

## • Field

→ Used to define and store data.

→ Represents a characteristic of individual/entity.

Eg: name, rollno, empid

## • Record

→ Logical collections of fields that describes a person / thing / place, etc.

Eg: (1, 'Anam', CS(h))

## \* Limitations of File System for data management.

→ Extensive programming is required

→ Ad-hoc queries are not permitted

→ System administration is complex and difficult

→ Difficult to make changes in existing structures

\* Types of Databases and Database applications

- Traditional Database application

- Store numeric and textual information

- Typically used for managing enterprises

- Multimedia Databases

- Store documents, digital images, audio and video streams.

- Geographic Information Systems (GIS)

- Store maps, weather data and satellite images.

- For route-finding, agriculture and natural

- resource management.

- Data Warehouses

- Data Warehouses and online analytical processing systems (OLAP)

- Store historical business information.

- For business analytics and decision support.

- Real-time and Active Databases.

- Store process models, constraints and key

- performance indicators.

- Control industrial and manufacturing processes.

\* Data: Known facts that can be recorded and have an implicit meaning.

Business Intelligence

Business intelligence

Business intelligence

**\* Database**

- Database is a collection of related and self-describing data.
- Database has following properties :
  1. It represents some aspect of real world called mineworld / universe of discourse (UoD).
  2. It is a logically coherent collection of data.
  3. It is designed, built, and populated with data for specific purpose.
  4. A database can be of any size and complexity.

**\* Database Management System (DBMS)**

- A software package/system to facilitate the creation and maintenance of a computerized database.
- DBMS is a general-purpose software system that facilitates the process of defining, constructing, manipulating and sharing databases for applications.

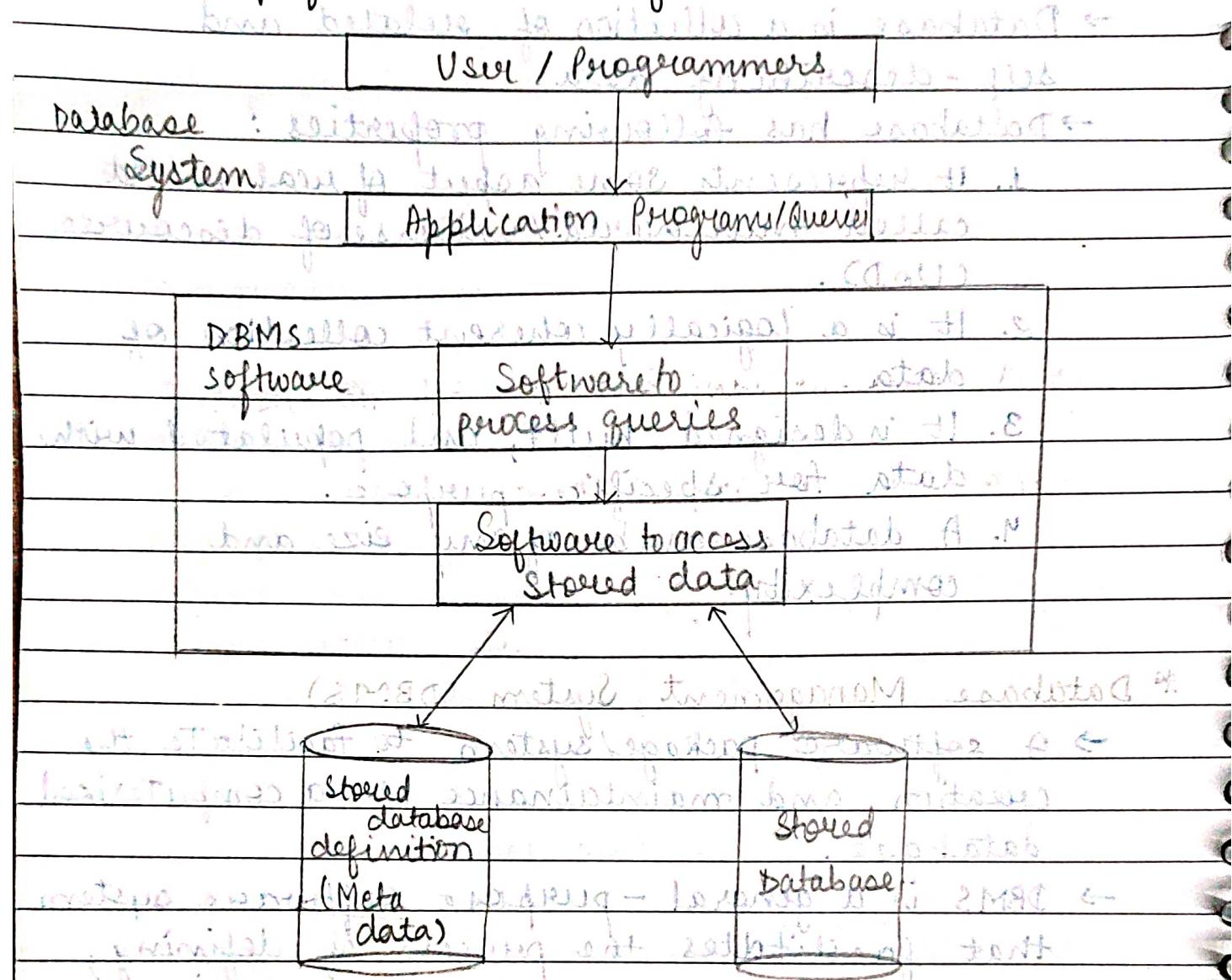
**• Database System**

- DB + DBMS : Here, DB is the h/w as it is stored on the hard disk and DBMS is s/w
- The DBMS software is together with the data itself.

**• Meta-data :** Data about data

- Definition or descriptive information is also stored in the form of a database catalog or dictionary, it is called meta-data.

\* A Simplified database system environment



## \* DBMS Functionality

- Defining a database : It involves specifying the data types, structures and constraints of the data to be stored in the database.
  - Constructing or loading the database : is the process of storing the data on some storage medium that is controlled by the DBMS! (on secondary storage medium) - ptm
  - Manipulating the database includes functions such as querying, up to retrieve specific data, updating the database and generating reports.

- Concurrent Processing and Sharing : allows multiple users and programs to access the database simultaneously.
- Protection or Security measures : includes system protection against hardware or software malfunction and security protection against unauthorized or malicious access.
- "Active" processing to take internal actions on data when want to use DBMS for real-time application.
- Presentation and visualization of data

#### \* Main Characteristics of the Database Approach

- Self-describing nature of a database system  
The information stored in the catalog - is called meta-data and it describes the structure of the primary database. This allows the DBMS software to work with different databases.
- Insulation between Programs and Data  
The structure of data files is stored in the DBMS catalog separately from the access programs.  
This property is called program-data independence.
- Data Abstraction  
A data model is a type of data abstraction that is used to hide storage details and present the user with a conceptual view of the database.  
Data model hides storage and implementation details that are not of interest to database user.
- Support of Multiple Views of the Data.  
Each user may see a different view of the database, which describes only the data of interest to that user. A multiuser DBMS whose users have

have variety of distinct applications must provide facilities for defining multiple views.

- Sharing of data and multuser transaction processing

Allows multiple users to access the database at the same time. It allows concurrent users to retrieve and update the database in on-line time. These app are called OLTP (online transaction processing) applications.

Concurrency control guarantees that each transaction is correctly executed or completely aborted.

DBMS takes more space than ES but retrieval of data is easier.

#### \* Database Users

1. Actors on the scene

- People whose jobs involve day-to-day use of large databases are called actors on the scene.

#### → Database Administrators

- Administering the use of resources is the responsibility of DBA

• Responsible for authorising access to the database, for coordinating and monitoring its use, acquiring software and hardware resources, controlling its use and monitoring efficiency of operation.

- DBA is accountable for security breaches and poor system response time.

→ Database Designers

- Responsible to define the content, the structure, the constraints and the functions of transactions against the database.
- They must communicate with end-users and understand their needs.

→ End Users

- They access/use the database for querying, updating and generating reports.

Categories of end-user

→ Casual end users.

→ Naïve or parametric end users.

→ Sophisticated end users.

→ Standalone users.

→ System Analysts and Application Programmers.

- System analysts study the needs of end users and communicate to the database designer.

- Application programmers implement the specification as programs.

- These analysts and programmers are referred to as Software Engineers.

## 2. Workers behind the Scene.

- Those who work to maintain the database system environment but are not actively interested in the contents of database.

→ DBMS designers and implementers

- Designs and implements the DBMS module and interfaces as a software package.

→ Tool Developers.

- Tool developers design and implement tools.
- Tools are optional packages that are often purchased separately, these facilitate database modeling and design.

→ Operators and maintenance personnel.

- Responsible for the actual running and maintenance of the hardware and S/W environment.

### \* Advantages of using DBMS Approach

• Controlling Redundancy

→ Redundancy causes the duplication of efforts and storage space is wasted.

→ Therefore, DBMS controls redundancy in data storage and minimize efforts in development and maintenance.

• Restricting Unauthorized Access

→ A DBMS should provide a security and authorization subsystem, which the DBA uses to create accounts and to specify account restrictions.

• Providing Persistent Storage for Program Objects

→ Databases can be used to provide persistent storage for program objects and data structures.

impediment mismatch problem — Refer book.

- Providing Storage Structure for efficient query processing
  - Database systems must provide capabilities for efficiently executing queries and updates.
- Providing Backup and Recovery.
  - DBMS must provide facilities for recovering from hardware or software failures.
  - The backup and recovery subsystem of the DBMS is responsible for recovery.
- Providing Multiple User Interfaces.
  - DBMS should provide a variety of user interfaces.
- Representing complex Relationships among Data.
  - DBMS must have the capability to represent a variety of complex relationships among the data.
- Enforcing Integrity Constraints
  - We can specify that every section record must be related to a course record Referential integrity
  - Key or uniqueness constraint specifies uniqueness on data item values.
- Permitting Inference and Actions using Rules
  - Deductive database systems provide capabilities for defining deductions. Inferring new information from the stored database facts.
  - Triggers: Rules activated by updates to the table
  - Stored procedures: More involved procedures to enforce rules.

Date .....

- Additional Implications of using the Database Approach

- Potential for enforcing standards.

Crucial for the success of database applications.

Standard refers to data items names, display formats, screen, report structures, meta-data.

- Reduced application development time.

Reduced incremental time to add each new application.

- Flexibility to change data structures.

Database structure may evolve as new requirements are defined.

- Availability of up-to-date information.

Very important for online transaction system such as airline, hotel, etc.

- Economies of scale.

Wasteful overlap of resources and personnel can be avoided by consolidating data and applications across departments.

- \* Software packages using db as backend.

- CRM : Customer records management

- ERP : Enterprise resource planning.

Include modules for handling different departments and their functional requirements for any organization.

- IR : Information Retrieval

Manage library related articles.

Support keyword searching.

Ranking & classification of documents.

Date .....

- \* When not to use DBMS
  - High initial investment and possible need for additional hardware.
  - Overhead for providing generality, security, recovery, etc.
  - Vendor dependence
  - Maintaining currency.
  - If the db are simple and not expected to change.
  - If access to data by multiple users is not required.
  - If db system is not able to handle the complexity of data.
  - Real time operations
  - If db users need special operations not supported by DBMS.

#### \* Classification of DBMS.

Criteria used :

1. Data Model
  - i) Relational
  - ii) Object oriented
  - iii) Object relational, etc.
2. Number of users
  - i) Single user
  - ii) Multi users
3. Number of sites over which DB is distributed
  - i) Centralized
  - ii) Distributed
4. Distributed DBMS
  - i) Homogeneous
  - ii) Heterogeneous

Date .....

5. Expected type and extent of its use for analysis

i) Operational DB management within organization

ii) Data warehouse

6. Structure of data mining and modeling

i) Structured

ii) Unstructured or semistructured data

between two data mining and data mining soft w

types of data mining and data mining soft w

and data mining soft w