

CHAPTER-1 INTRODUCTION TO DATABASE SYSTEM

1.1 INTRODUCTION

(DATABASE- database is collection of logically relate data items.)

❖ DEFINATION

- **“DATABASE MANAGMENT SYSTEM is a software to build and manage the data in databases.”**
- PRIMARY GOAL OF DBMS is to provide way to store and retrieve database information that is both convenient and efficient.
- DBMS designed to manage large amount of information.
- DBMS allows insertion, update, deletion and processing of data.
- Ex: oracle, SQL server, MySQL,Microsoft Access etc are famous DBMSs.

❖ DATABASE SYTEM APPLICATION

- Banking: all transactions
- Airlines: reservations, schedules
- Universities: registration, grades
- Sales: customers, products, purchases
- Online retailers: order tracking, customized recommendations
- Manufacturing: production, inventory, orders, supply chain
- Human resources: employee records, salaries, tax deductions

1.2 BASIC CONCEPT AND DEFINATION

❖ DATA

- “DATA is a raw fact that can be stored or recorded”.
- It is not meaningful.
- Data is unorganized that need processed to make it useful.
- It is input processing unit.
- Data is basic row material which taken by certain observation, certain experiment and storing in paper, stored memory, human mind.
- Ex:- marks of student ,account information

❖ INFORMATION

- Information is a processed from of data.
- It is always meaningful.
- “Meaningful data is called information”.
- It is output processing units.
- When ever we organize the data and process is done on it then we get proper information.

➤ Example :-

NO.	BRANCH	STUDENT
1	CE	60
2	EC	50

- Here their data of student branch and no of student study in student.

❖ DATA VS INFORMATION

<u>DATA</u>	<u>INFORMATION</u>									
➤ Data is a raw fact or material	➤ Information is processed from the data									
➤ It may not be meaningful	➤ It is always meaningful									
➤ It is used for analysis and reasoning	➤ The result of analysis and reasoning									
➤ It is input processing unit	➤ It is output processing unit									
➤ It may be accurate but not necessarily passes other characteristics of information	➤ It is relevant, timely, accurate, reliable, concise, and completes									
➤ Ex: marks of the student, account number	<table><tr><td>➤ NO.</td><td>BRANCH</td><td>STUDENT</td></tr><tr><td>1</td><td>CE</td><td>60</td></tr><tr><td>2</td><td>ME</td><td>50</td></tr></table> <p>Information of student, branch and no of student study in college</p>	➤ NO.	BRANCH	STUDENT	1	CE	60	2	ME	50
➤ NO.	BRANCH	STUDENT								
1	CE	60								
2	ME	50								

❖ DATA WAREHOUSE

- Data warehouse contains historical and huge data.
- It is generally used to generate different reports.
- “A subject oriented, integrated, nonvolatile, time-variant data store in support of management decision”.
- Nonvolatile:-“once inserted, data cannot be changed, though it might be deleted”.

❖ METADATA

- The data structure of database is specifying by database schema and it called metadata.
- Metadata is the data about the data.
- Data that describe the properties or characteristics of other data
- Properties include data definitions, data structures, and rules or constraints.
- Metadata describe data but not include that data

❖ SYSTEM CATALOG

- System catalog is a system created database that describes all database objects, data dictionary information and user access information.
- A DBMS needs to maintain data about the relations such as the schema of the relations. This information is called system catalog.
- The types of information that the system must store are these
 - Name of the relations(table).
 - Name of the attributes of each relation
 - Domain and lengths of attributes
 - Integrity constraints (for ex :- key constraints)

❖ DATA ITEMS (FIELDS, ATTRIBUTES, COLUMNS)

- The characteristics of an entity are called fields or attributes or columns.
- Entity can have a number of characteristics like name, address, phone no. etc.
- Entity is a group of similar information or data.

❖ RECORD (ROW, TUPLE)

- Collection of logically related filed or data items
- Multiple fields placed in a horizontal plane are called a record or row or TUPLE.
- An organization will deal with many clients and the same information must be recorded for each client multiple fields placed in a horizontal plane is called a record or row or TUPLE.

❖ Files

- A file is collection of records.
- Files are included into database
- For example if you have 100 records in student database then these 100 records create a file.

❖ DATA DICTIONARY (DATA DIRECTORY)

- Data dictionary, which stores metadata about the structure of database, in particular the schema of the database.
- The data dictionary is a file that contains metadata that is data about data.
- The data structure of database is specifying by database schema and it called metadata.
- The file which contains that metadata is called DATA DICTIONARY (DIRECTORY).
- Example: - In ORACLE special tables **tab** is contain the information about all tables, called table of table. It gives all the name of table in the database so tab contains data about data. And it may refer as data dictionary.
- A DBMS needs to maintain data about the relations such as the schema of the relations. This information is called the data-dictionary or system catalog.

- The types of information that the system must store are these
 - Name of the relations
 - Name of the attributes of each relation
 - Domain and lengths of attributes
 - Integrity constraints (for ex :- key constraints)
- **Components of data dictionary**
 - **Entities**
 - Entity is the real world object OR an event
 - In other words, any item about which information is stored is called entity.
 - **Example:** Table Name:-**Client_master**
 - **Attributes**
 - The characteristics of an entity are called fields or attributes or columns.
 - Entity (**Client_master**) can have a number of characteristics (Column name) like name, address, phone no. etc.
 - **Relationships**
 - Different entities relate to each other is called relationships
 - Relationships has three types
 - ◆ One – to – one relationships
 - Example:-employee – employee_no
 - ◆ One – to – many relationships
 - Example:-department – employee
 - ◆ Many – to – Many relationships
 - Example:-employee – salary
 - **Key**
 - Key is a single attributes or combination of attributes of any entity set that is used to identify one or more instance of the set.
 - There are various types of keys
 - ◆ Primary Key
 - ◆ Referenced Key
 - ◆ Foreign Key
 - ◆ Super Key
- **Data dictionaries can also classified as Active or Passive**
 - **Active data dictionary** is automatically updated by the DBMS with every database access, thereby keeping its access information up to date.
 - **Passive data dictionary** is not updated automatically and usually requires running a batch process
 - Data dictionary access information is normally used by the DBMS for query optimization purposes.

1.3 Database

- A database is a collection of persistent data that is used by application systems of some given enterprise.
- Persistent means the data is not transient in nature like input and output data.
- In short database is collection of data.
- Database can be define by meaningful data
- An organized collection of logically related data.
- Database consists of the four components
 - Data item
 - Relationships
 - Constraints
 - Schema

1.4 Database system

- Database system is basically computerized record keeping system, whose overall purpose is to maintain information and to make that information available on demand
- **Operations performed on database systems.**
 - **Create** – Add new data to the database.
 - **Read** – Read current database data (often presented in a useful format on a Computer screen or a printed report).
 - **Update** – Update (or modify) current database data.
 - **Delete** – Delete current data from the database.

❖ database system environment

- **Data**
- **Hardware**
- **Software**
- **User**
- **Data**
 - Without Data, database does not mean at all.
 - DBMS stores both; data and metadata.
 - Data can be shared or restricted.
 - **Shared** - The individual piece of data in the database can be shared among several different users.
 - **Restricted** – restricted data can be used only by the authorized users.
- **Hardware**
 - It constitutes physical components that are used to construct the DBMS.

- It includes computer, storage devices, I/O channels etc
- Processor and main memory – they support the execution of database system software.

➤ Software

- Software is a collection of a program.
- It is the layer between the physical database itself and the users called database manager or more usually database management system. It shields the users of database from hardware level details.

➤ Users

- There are the four different types of database-system users, different by the way that they expect to interact with the system.
 - Application programmers
 - Sophisticated users
 - Specialized users
 - Naïve users
- **Application programmers** : responsible for writing database application programs in some programming language such as COBOL, PL/I, C++, JAVA
- **Sophisticated users**: they interact with the system without writing programs. They form their request in the database query processor whose function is to break down **DML** statement into instruction that the storage manager understands.
- **Specialized users**: They are sophisticated users who write specialize database application that do not fit into the traditional data-processing framework.
- **Naïve users**: They are unsophisticated users who interact with the system by invoking one of the permanent application programs that have been written previously.

1.5 Data administrator

- We elaborate on this concept of centralized control.
- The concept implies that there will be some identifiable person in the enterprise who has this central responsibility for the data.
- Given that the data is one of the enterprise's most valuable assets, it is imperative that there should be some person who understands that data, and needs of the enterprise with respect to that data, at a senior management level.
- The data administrator is that person.
- Thus, it is the administrator's job decide what data should be stored in the database in the first place, and establish policies for maintaining and dealing with that data once it has been stored.

- Example such a policy might be one that dictates who can perform what operations on what data in what circumstances-in other words, data security policy.
- Data administration is a manager, not a technician.

1.6 Database administrator

- A person who has such central control over the system is called **database administrator**.
- DBA have responsibility of controlling and protecting the data.
- DBA creating, modifying and maintain data.
- The functions of DBA include:
 - **Schema definition:** The DBA creates the original database schema by executing a set of data definition statements in DDL (data-definition language).
 - **Storage structure and access-method definition:-**DBA decide how the data is to be represented in stored database(Physical database design).DBA defines the storage structure of the database and the access method of data from the database
 - **Schema and physical-organization modification:** The DBA carries out changes to the schema and physical organization to reflect the changing needs of the organization, or physical organization improve performance.
 - **Granting of authorization for data access:** Granting different types of authorization, the database administrator can regulate which parts of the database various users can access. The authorization information is kept in a special system structure that the database system consults whenever someone attempts to access the data in the system.
 - **Routine maintenance :** example of the database administrator's routine maintenance activities are:
 - Periodically backing up the database, either onto tapes or not remote servers, to prevent loss of data in case of disasters as flooding.
 - Ensuring that enough free disk space is available for normal operations, and upgrading disk space required.

Difference between DA and DBA

<u>DA</u>	<u>DBA</u>
DA is a person who understands the data and needs of the enterprise.	DBA is a technical person responsible to implementing DA's decisions.
He is not necessarily IT professionals	He must be IT professionals
DA doesn't design database. He just decides what data should be in database and various constraints.	DBA creates actual database which supports all constraints decided by DA.
DA decides which services must be available but not how to implement them.	DBA implements all the services decided by DA.
He has central responsibility of data	He has central responsibility of database.

1.8 File oriented system versus database system

- One way to keep the information on a computer is to store it in operating system files. The system has a number of application programs that manipulate files, including
 - A program to debit or credit an account
 - A program to add anew account
 - A program to find the balance of account
 - A program to generate monthly statements
- File processing system is a system used to store and manage data that involves each department or area within an organization having its own set of files, often creating data redundancy and data isolation.
- File-processing system is supported by conventional operating system.
- The system stores permanent records in various files.
- Before database management system came among, organizations usually stored information in such systems.

File Management Systems

<u>Advantages</u>	<u>Disadvantages</u>
Simpler to use	Typically does not support multi-user access
Less expensive	Limited to smaller databases
Fits the needs of many small businesses and home users	Limited functionality (i.e. no support for complicated transactions, recovery, etc.)
Popular FMS's are packaged along with the operating systems of personal computers (i.e. Microsoft Card file and Microsoft Works)	Decentralization of data
Good for database solutions for hand held devices such as Palm Pilot	Redundancy and Integrity issues

DISADVANTAGE OF FILE PROCESSING (ORIENTED) SYSTEM
OR
ADVANTAGE OF DATABASE SYSTEM
OR
DATABASE APPROACH

- Data redundancy and inconsistency
 - Difficulty in accessing data
 - Data isolation
 - Integrity problems
 - Atomicity Problems
 - Concurrent Access anomalies
 - Security problem
-
- **Data redundancy and inconsistency:** The same information may be duplicated in several files which is called redundancy. This redundancy leads to higher storage space and access cost. In addition it leads to data inconsistency because various copies of data may no longer agree.
 - **Difficulty in accessing data :** In the older system if new request is anticipated at that time there will be no application program on hand to meet the request at that time only two possibilities are there either information may be extracted manually or write necessary application program.
 - **Data isolation:** Data is scattered in various files. And files may be in different formats then it is difficult to write new application programs to retrieve data.

- **Integrity problems:** Integrity constraints are added to the system by adding appropriate code in the various application programs so when new constraints are to be added it is difficult to change the program to enforce them.
- **Atomicity Problems:** A computer system is subject to failure but in May application it is crucial to ensure that once a failure has occurred and has been detected, the data is restored to the consistent state that existed before the failure.
- **Concurrent Access anomalies:** In multi-user systems multiple user the data simultaneously so the probability of data being in consistent state is more.
- **Security problem:** Every user of the database must not be able to access all the data.

➤ **Disadvantage of DBMS**

- Problem associated with Centralization
- Security might be compromised (without good controls)
- Integrity might be compromised (without good controls)
- Additional hardware might be required
- Performance overhead might be significant
- Successful operation is crucial (the enterprise might be highly vulnerable to failure)
- The system is likely to be complex (though such complexity should be concealed from the user)
- Cost of Software/Hardware and Migration
- Complexity of backup and recovery

Database Management Systems

Database Management Systems provide the following advantages and disadvantages:

<u>Advantages</u>	<u>Disadvantages</u>
Greater flexibility	Difficult to learn
Good for larger databases	Packaged separately from the operating system (i.e. Oracle, Microsoft Access, Lotus/IBM Approach, Borland Paradox, Claris FileMaker Pro)
Greater processing power	Slower processing speeds
Fits the needs of many medium to large-sized organizations	Requires skilled administrators
Storage for all relevant data	Expensive
Provides user views relevant to tasks performed	
Ensures data integrity by managing transactions (ACID test = atomicity, consistency, isolation, durability)	
Supports simultaneous access	
Enforces design criteria in relation to data format and structure	
Provides backup and recovery controls	
Advanced security	

❖ **Database System Environment**

- Computer-aided software engineering (CASE) tools
- Repository
- Database management systems (DBMS)
- Database
- Application programs
- User interface
- Data administrators
- System developers
- End users