

Define the following terms.

Data

- Data means known facts that can be stored.
- For example age, salary, result etc...

Database

- A Database is a collection of inter-related Data.
- For example consider a collection of student roll no, name, address, contact number for some student's information. Here all the data - roll no, name, address, contact number are inter-related. They all belong to some particular student.
- Any random collection of data is not considered as a database. For example collections of some vehicle number, balance of account, percentage of student do not form a database.

Management

- Manipulation, searching and security of data
- For example searching of product in amazon, viewing result in GTU website etc...

System

- Programs or tools used to manage database
- For example SQL Server Studio, Oracle 11g

DBMS

- A database management system is a collection of inter-related data and a set of programs to manipulate those data.
- Data manipulation involves various operations such as store data, modify data, remove data and retrieve data.
- DBMS = Database + A set of programs

Metadata

- Metadata is data about data. Data such as table name, column name, data type, authorized user, user access privileges for any table is called metadata for that table.

Data dictionary

- Data dictionary is an information repository which contains metadata. It is usually a part of the system catalog.

- Data dictionary contains description of schema, i.e. overall logical structure of the database. This can involve information such as table names, owners, column names, data types, size and constraints.
- A data dictionary contains the following components:
 - ✓ **Entities:** An entity is a thing or object or person in the real world that is distinguishable from all other objects. E.g. book, student, employee, college etc...
 - ✓ **Attribute:** Attribute is a property or characteristic of an entity. E.g. entity is student and attributes of students are enrollmentno, name, address, cpi etc ...
 - ✓ **Relationship:** Relationship is an association (connection) between several entities. E.g. book is issued by student where book and student are entities and issue is relation.
 - ✓ **Key:** A data item or a field which is used to identify a record in a database is referred as key. A primary key is used to uniquely identify a record. For example enrollmentno.

Data warehouse

- Data warehouse is an information repository which stores data. It is designed to facilitate reporting and analysis.

Data Items (Field)

- A field is a character or group of characters (alphabetic or numeric) that have a specific meaning.
- It is also called a data item. It is represented in the database by a value.
- For Example customer id, name, society, and city are all fields for customer Data.

Record

- A record is a collection of logically related fields.
- Here, each field in a record contains a fixed size and fixed data type.
- For example, collection of fields – id, name, society & city – forms a record for customer.
- A record contains values for each field.

Files

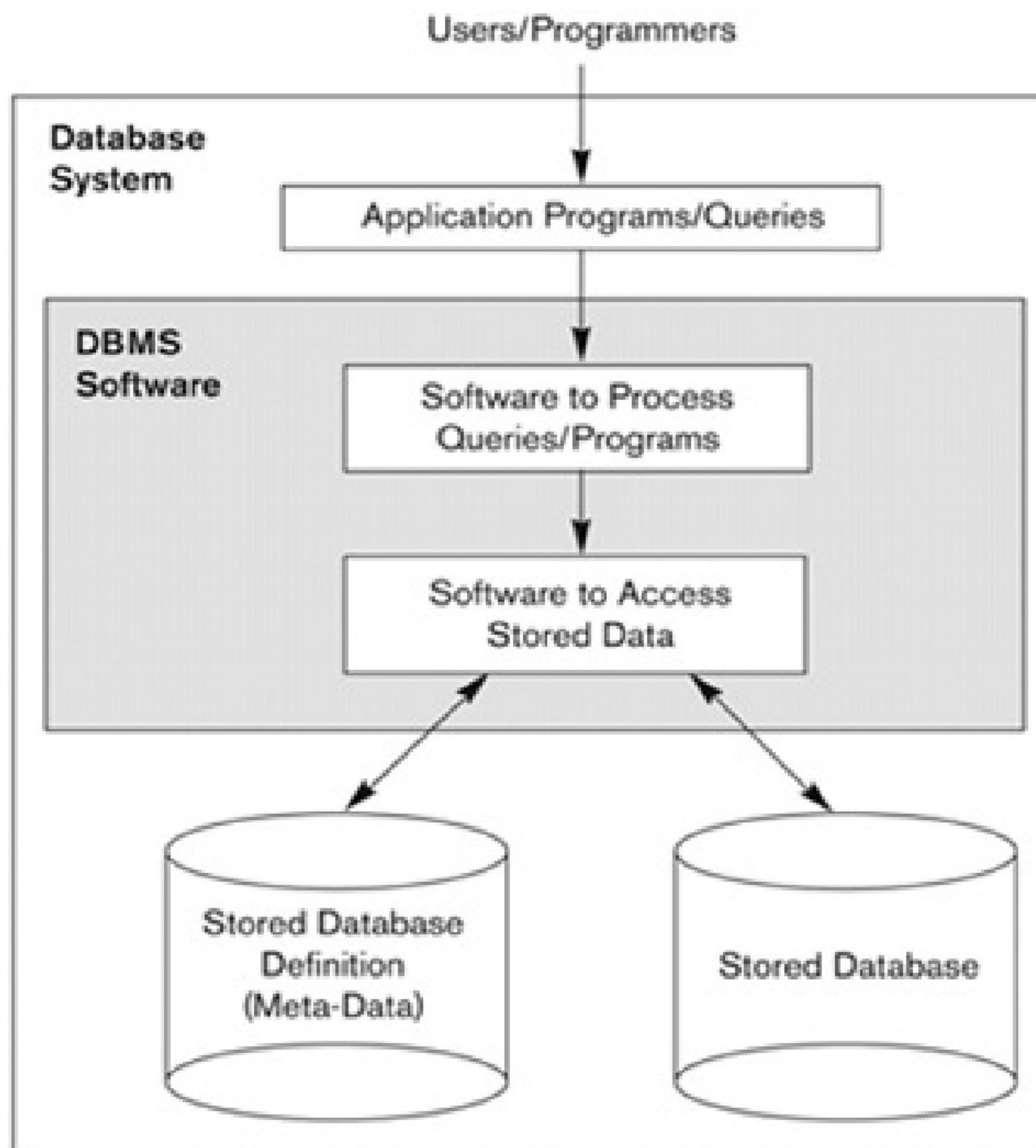
- A file is collection of related records.
- These records are generally arranged in a specific sequence.

Differentiate the Data and Information.

Data	Information
Data means known facts, that can be recorded and have implicit meaning	Information means processed or organized data.
Examples: <ul style="list-style-type: none"> • Student no: 7001 • Student name: Ram • City name: Rajkot 	Examples: <ul style="list-style-type: none"> • Percentage : 82.20% (Derived from marks of all subjects) • Run rate in cricket match: 6.0

<ul style="list-style-type: none"> • Account No: A01 • Balance: 5000 	run/over (Derived from total runs and over)
Data are raw materials used to derive information.	Information is a product derived from Data.
Data is comparatively less useful.	Information is comparatively more useful.

Explain Database System Environment.



Hardware

- All the physical devices of a computer system are referred as hardware. A computer system can have number of different hardware such as processor, memory, hard disk, monitor, keyboard, mouse, printer, scanner etc.
- From the database point of view important hardware can be divided into two category:
 1. The processor and main memory:
 - ✓ Supports the execution of the database software.
 2. The secondary storage devices:
 - ✓ Used to store data of system permanently
 - ✓ These includes hard disks, magnetic tapes, compact disks etc

Software

- Software provides the interface between users and database stored in physical devices. Application programs, DBMS software and OS generate the software component here.
- Application program are developed using programming language like C, C++, Java etc.
- These application programs use the functionalities of the DBMS software to perform various operations on the database.

- The OS manages all hardware of the computer.

Users

- Any person who interacts with a database in any form is considered as a database user.
- There are four different database users.

1) Database Administrator

- ✓ Has central control over the database system – including data and programs.
- ✓ DBA is responsible for proper functioning of the database system.

2) Database designer

- ✓ Identify the data to be stored in database and designs structure of the database for an organization.

3) Application programmers

- ✓ These users write application programs to interact with the database.
- ✓ Application programs can be written in some programming language such as C++, JAVA.
- ✓ Such programs access the database by issuing the request, typically a SQL statement to DBMS.

4) End Users

- ✓ Interact with the system by using pre-developed application programs.
- ✓ End users do not know about the working, database design, the access mechanism etc.
- ✓ They just use the system to get their task done.

Differentiate the DA and DBA.

DA (Data Administrator)	DBA (Database Administrator)
The data administrator is a person in the organization who controls the data of the database.	The database administrator is a person in the organization who controls the design and the use of the database.
DA determines what data to be stored in database based on requirements of the organization.	DBA provides necessary technical support for implementing a database.
DA is involved more in the requirements gathering, analysis of data.	DBA is involved more in the design, development, testing and operational phases.
DA is a manager or some senior level person in an organization who understands organizational requirements with respect to data.	DBA is a technical person who have the knowledge of database technology.
DA does not need to be a technical person.	DBA does not need to be a business person.
DA is a business focused person, but, he/she should understand more about the database technology.	DBA is a technically focused person, but, he/she should understand more about the business to administrate the databases effectively.

Describe functions & responsibility of DBA.

- Database Administrator is a person in the organization who controls the design and the use of database.

Schema definition

- DBA defines the logical schema of the database. A schema refers to the overall logical structure of the database.
- According to this schema, database will be developed to store required data for an organization.

Storage structure and access method definition

- DBA decides how the data is to be represented in the database.
- Based on this, storage structure of the database and access methods of data is defined.

Defining security and integrity constraints

- DBA decides the security and integrity constraints.
- DDL provides facilities to specifying such constraints.

Granting of authorization for data access

- The DBA decides which user needs access to which part of the database.
- According to this, various types of authorizations are granted to different users. This is required to prevent unauthorized access of a database.

Assisting Application programmers

- DBA provide assistance to application programmers to develop application programs.

Monitoring performance

- The DBA monitors performance of the system.
- The DBA ensure that better performance is maintained by making change in physical or logical schema if required.

Backup and Recovery

- Database should not be lost or damaged.
- The task of DBA is to backing up the database on magnetic tapes or remote servers.
- In case of failures, such as flood or virus attack, Database is recovered from this backup.

Explain Advantages of File Oriented System.

- File oriented systems possess some advantages compared to paper based system.

Back-up

- It is possible to take faster and automatic back-up of database stored in file of computer-based systems.

Compactness

- It is possible to store data compactly. For example to store all the words of English dictionary only few kilobytes of memory is required in computer-based systems.

Data Retrieval

- Computer-based systems provides enhanced data retrieval techniques to retrieve data stored in file in easy and efficient way

Editing

- It is easy to edit any information stored in computers in form of files. Specific application programs or editing software can be used for this purpose.

Remote Access

- In computer-based systems it is possible to access data remotely. So to access data it is not necessary for a user to remain present at location where these data are kept.

Sharing

- Data stored in files systems can be shared among multiple users at a same time.

Explain disadvantages of file management system over DBMS.

Data Redundancy

- It is possible that the same information may be duplicated in different files. This leads to data redundancy. Data redundancy results in memory wastage.
- For example, consider that some customers have both kinds of accounts - saving and current. In this case, data about customers - name, address, e-mail, contact number - will be duplicated in both files, file for saving accounts and file for current accounts. This leads to requirement of higher storage space. In other words, same information will be stored in two different locations (files). And, it wastes memory.

Data Inconsistency

- Due to data redundancy, it is possible that data may not be in consistent state.
- For example, consider that an address of some customer changes. And, that customer has both kinds of accounts. Now, it is possible that this changed address is updated in only one file, leaving address in other file as it is. As a result of this, same customer will have two different addresses in two different files, making data inconsistent.

Difficulty in Accessing Data

- Accessing data is not convenient and efficient in file processing system.
- For example, suppose, there is a program to find information about all customers. But, what if there is a need to find out all customers from some particular city. In this case, there are two choices here: One, find out all customers using available program, and then extract the needed customers manually. Second, develop new program to get required information. Both options are not satisfactory.
- For each and every different kind of data access, separate programs are required. This is neither convenient nor efficient.

Limited Data Sharing

- Data are scattered in various files. Also, different files may have different formats. And these files may be stored in different folders (directories) may be of different computers

of different departments. So, due to this data isolation, it is difficult to share data among different applications.

Integrity Problems

- Data integrity means that the data contained in the database is both correct and consistent. For this purpose, the data stored in database must satisfy certain types of constraints (rules).
- For example, a balance for any account must not be less than zero. Such constraints are enforced in the system by adding appropriate code in application programs. But, when new constraints are added, such as balance should not be less than 0, application programs need to be changed. But, it is not an easy task to change programs whenever required.

Atomicity Problems

- Any operation on database must be atomic. This means, it must happen in its entirety or not at all.
- For example, a fund transfer from one account to another must happen in its entirety. But, computer systems failure due to system crash or virus attack. If a system failure occurs during the execution of fund transfer operation, it may possible that amount to be transferred, say, Rs. 500, is debited from one account, but is not credited to another account.
- This leaves database in inconsistent state. But, it is difficult to ensure atomicity in a file processing system.

Concurrent Access Anomalies

- Multiple users are allowed to access data simultaneously (concurrently). This is for the better performance and faster response.
- Consider an operation to debit (withdrawal) an account. The program reads the old balance, calculates the new balance, and writes new balance back to database. Suppose an account has a balance of Rs. 5000. Now, a concurrent withdrawal of Rs. 1000 and Rs. 2000 may leave the balance Rs. 4000 or Rs. 3000 depending upon their completion time rather than the correct value of Rs. 2000.
- Here, concurrent data access should be allowed under some supervision. But, due to lack of co-ordination among different application programs, this is not possible in file processing systems.

Security Problems

- Database should be accessible to users in a limited way. Each user should be allowed to access data related to his application only.
- For example, a customer can check balance only for his/her own account. He/She should not have access to information about other accounts.
- It is difficult to enforce such kind of security constraints in file systems.

Explain advantages of DBMS over file management system.

OR

Explain purpose of database system.

Minimal Data Redundancy (Duplication)

- Due to centralized database, it is possible to avoid unnecessary duplication of information.
- For example all the information about bank customer can be kept centralized. Both accounts saving as well as current can be sharing this information. This prevents unnecessary duplication of customer information that has both types of accounts.
- This leads to reduced data redundancy.
- It prevents memory wastage.
- It also reduced extra processing time to get required data.

Improved Data Consistency

- Data inconsistency occurs due to data redundancy.
- For example, consider that customer information is maintained separately for saving accounts and current accounts. Now, it is possible that this changed address is updated in only one file, leaving address in other file as it is.
- As a result of this, same customer will have two different addresses in two different files, making data inconsistent.
- With reduced data redundancy such type of data inconsistency can be eliminated.
- This results in improved data consistency.

Efficient Data Access

- DBMS utilizes a variety of techniques to retrieve data.
- Required data can be retrieved by providing appropriate query to the DBMS.
- For example information about all customers or from some particular city can be retrieved easily by providing appropriate query statement (SELECT statement).
- Thus, Data can be accessed in convenient and efficient manner.

Improved Data Sharing

- As database is maintained centrally all authorized user and application program can share database easily.

Improved Data integrity

- Data in database must be correct and consistent.
- So, data stored in database must satisfy certain types of constraints (rules).
- For example balance in an account should not be a negative value.
- DBMS provides different ways to implement such type of constraints.
- This improves data integrity in a database.

Guaranteed Atomicity

- Any operation on database must be atomic. This means, operation must be executed either 100% or 0%.

- It is the responsibility of the DBMS software to ensure such kind of atomicity. If any operation fails due to some problem such as system crash then effects the partially executed operation can be undone.

Improved Concurrent Access

- Multiple users are allowed to access data simultaneously. This is for the better performance and faster response.
- As database is maintained centrally data can be shared among multiple users. Also concurrent access to centralized data can be allowed under some supervision.
- This result in better performance of system and faster response.

Improved Data security

- Database should be accessible to user in a limited way.
- For example, a customer can check balance only for his/her own account. He/She should not have access to information about other accounts.
- DBMS provides way to control the access to data for different user according to their requirement.
- It prevents unauthorized access to data.
- Thus, security can be improved.

Explain disadvantages of DBMS.

High Cost

- High initial cost is required to implement database and application programs.
- It incurs high initial cost behind DBMS software. Also hardware needs to be upgraded according to requirements of the DBMS software.
- Conversion from older file oriented system to database system is also costly in terms of money as well as time.

Specialized Manpower

- Database system requires specialized skilled manpower to design and develop database and to provide database administration services.
- Also due to rapid changes in database technology this manpower needs to be trained and retrained on regular basis.

Increased complexity

- It is necessary to understand the requirements of the organization and different kinds of users to design and implement efficient database.
- With increase in requirements and functionalities complexity of the database increase.

Need of Explicit Backup and Recovery

- As entire database is maintained centrally it is must to provide efficient backup and recovery facility. Else any kind of damage to this database can affect the operation of the entire system