



COMPUTER SCIENCE

12

(MS Access and C)

CHAPTER 1: Data Basics

Topics

- Data & Information
- Data Processing & its Activities
- File, Record and Field
- File Types from Usage Point of View
- File Types from Functional Point of View
- Organize Files on Storage Media
- Problems in File Processing System
- Database
- Database System and its Components
- Objectives of Database
- Database Model
- DBMS & its Objectives

Topics (continued)

- Features of DBMS
- Advantages of DBMS
- Disadvantages of DBMS

Data

- Collection of facts and figures related to an object
- Data may consist of text, numbers, images, sound and videos
- Processed to produce useful information

Object can be person, event or anything about which data is collected

Importance of Data

- Used by managers to perform effective and successful operations of management
- Provides a view of past activities related to the rise and fall of an organization
- Enables the organization to make better decision for future activities

Example

- The data of student may consist of Roll Number, Student Name, and marks of different subjects

Information

- Processed data
 - Organized
 - Meaningful
 - Useful
- Used for making decisions
- Data is used as input for processing and information is the output of this processing



Example

- The data of student can be processed to produce useful information such as:
 - Total Marks
 - Grade
 - Find the number of passed and failed students etc.

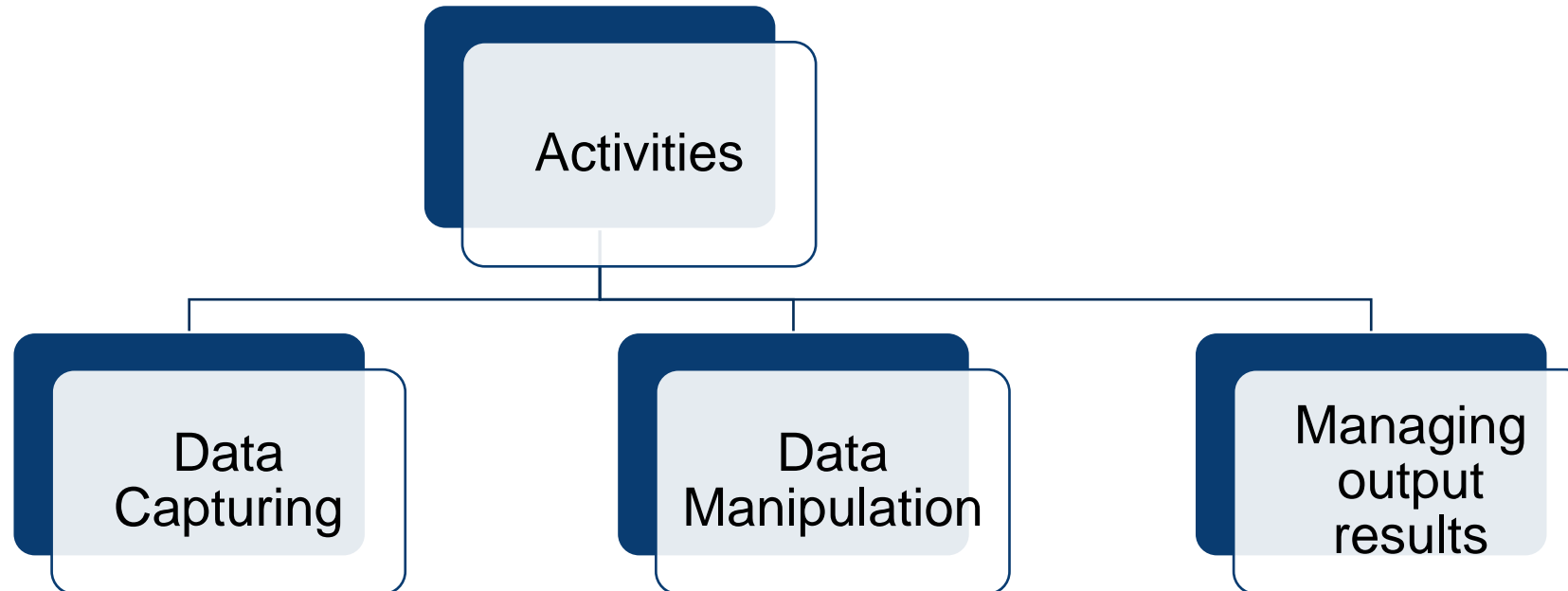
Difference between Data and Information

Data	Information
<ul style="list-style-type: none">• Collection of raw facts and figures	<ul style="list-style-type: none">• Processed form of data
<ul style="list-style-type: none">• Used as input in the computer	<ul style="list-style-type: none">• Output of computer
<ul style="list-style-type: none">• Huge in its volume	<ul style="list-style-type: none">• Short in its volume
<ul style="list-style-type: none">• Difficult or even impossible to reproduce. For example, it is very hard to reproduce the data of census if it is lost	<ul style="list-style-type: none">• Easier to reproduce. For example, number of graduate citizens can be recalculated from the stored data
<ul style="list-style-type: none">• Used rarely	<ul style="list-style-type: none">• Used frequently
<ul style="list-style-type: none">• Does not dependent on information	<ul style="list-style-type: none">• Depends on data

Data Processing or Operations on Data

- Series of actions or operations are performed on data to get required output or result (Information)
- Software is used to process data
- Software convert the data into meaningful information

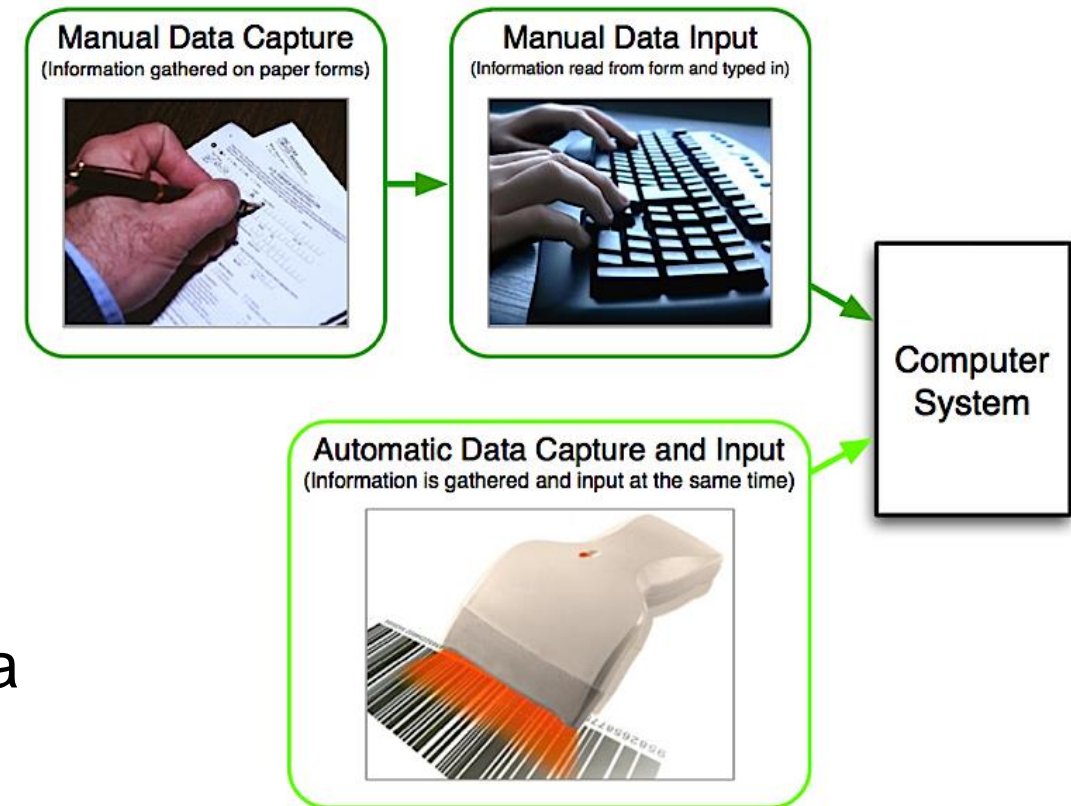
Activities in Data Processing



Activities in Data Processing

1. Data Capturing

- Process of recording the data in some form
- Data is captured before it can be processed
- Data may be recorded on source document
- Source document may include:
 - Photographs, checks, or product label, brochures
- Recording data directly into Computer
- Provide a quick and efficient way to input data
 - Saves time
 - Increases accuracy



Activities in Data Processing

2. Data Manipulation

- Process of applying different operations on data
- **Classifying** - The process of organizing data into classes or groups
 - **Example** - Data of college can be in two group → Data of student and Data of teacher
- **Calculation** - The process of applying arithmetic operations(+, -, ÷ , ×) on data
 - **Example** - The total marks of student are calculated to find the grade
- **Sorting** - The process of arranging data in logical sequence
 - **Example** - Name of students can be sorted according to obtained marks
- **Summarizing** - Process of reducing a large amount of data in more concise form
 - **Example** - Data of students in class can be summarized to show number of passed and failed students

Activities in Data Processing

3. Managing Output Result

Performed on data after the data has been captured and manipulated

- **Storage** - The process of retaining data on storage media such as hard disk for future use
 - **Example** - The student data is stored on the hard disk
- **Calculation** - The process of accessing or fetching the stored data
 - **Example** - Student data can be retrieved from the hard disk any time to prepare result card
- **Communication** - The process of transferring data from one location to another
 - **Example** - The result can be sent to the students via email
- **Reproduction** - The process of copying or duplicating data
 - Data can be reproduced if different users need data at different locations

Field, Record and File

Field

- A combination of one or more characters
- Represents Smallest unit of data
- Name of each field in a record is unique
- Each field contains one specific piece of information
- **Example** - The EmployeeID, Name, HireDate, JobTitle and Phone

Fields
↓

Record →

EmployeeID	0001
Name	Ali Raza
HireDate	05/11/1998
JobTitle	Manager
Phone	2642707

Record

- A collection of related fields used as single unit
- **Example** - An Employee's record includes a set of fields that contains EmployeeID, Name, HireDate, JobTitle and Phone

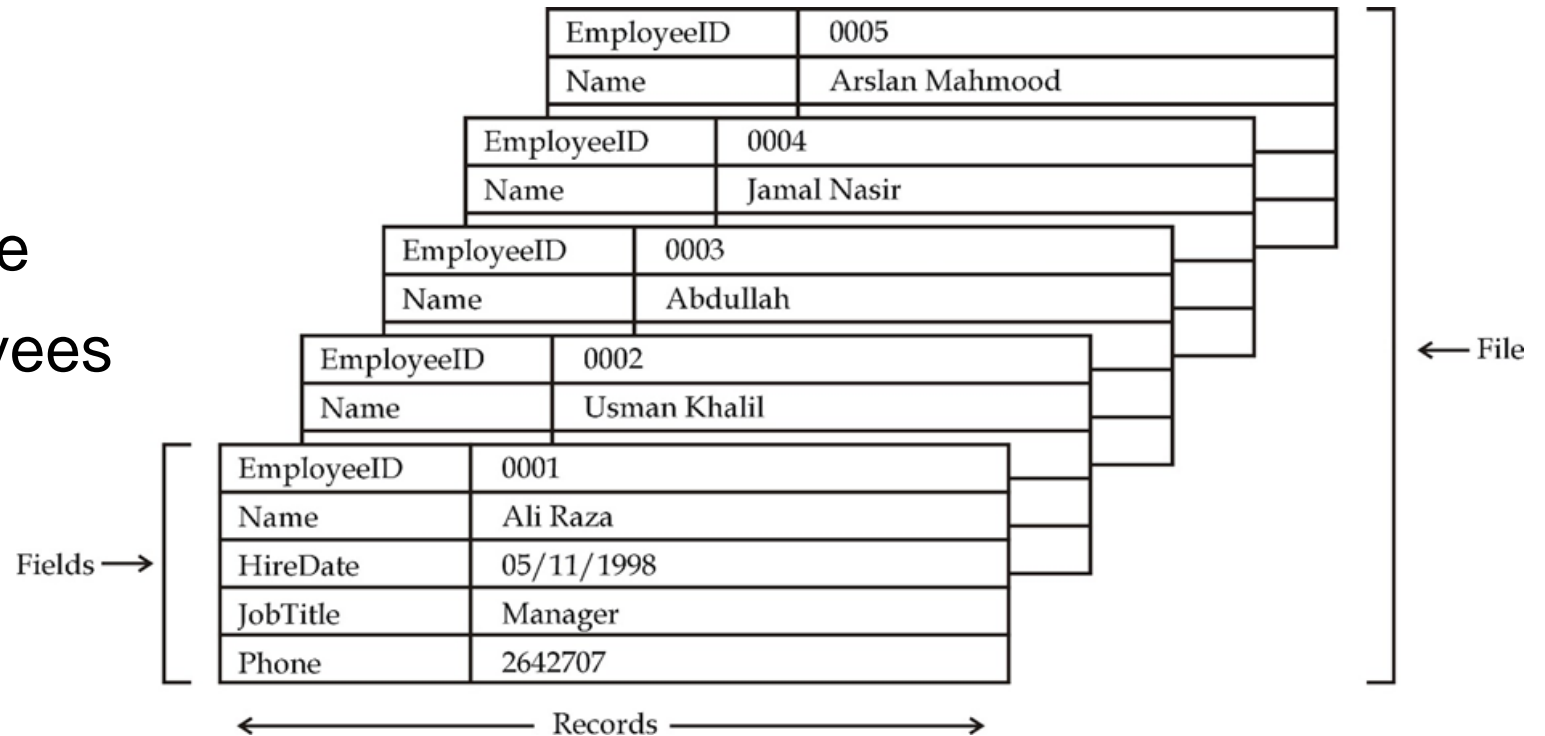
Field, Record and File

File

- A collection of related records used as single unit
- Files are stored on different storage media such as hard disk, USB flash drive or optical disc (CDs and DVDs)

Example

- Employee file may contain the records of hundreds of Employees



File Types from Usage point of View

1. Master File

- Used to store the information that remains constant for a long period of time
- **Example** – A college maintains a master file of all students
- It is updated when any change in its contents is required
- These files are never empty since they are created

2. Transaction File

- Used to store the input data before processing
- It may be temporary file
- The data in transaction files is used to update the master files
- It may exist until the master file is updated
- It may also be used to maintain a permanent record of data about transaction
 - **Example** – A transaction file can be used to store the fee deposited by the student

File Types from Usage point of View

3. Backup File

- Used to take the backup of important data
- Permanent file
- Make additional copy of data
- The data can be recovered from backup files if any data file is lost or damaged
- Backup files are mostly created by using specific software (utility program)

File Types from Functional Point of View

- A file consists of file name and file extension
- Name and extension of file is separated by dot (.)
- The extension of a file is normally assigned by the software in which it is created

1. Program File

- Contains the software instructions
- File extensions: .exe or .com
- Contains instructions that can be directly executed by the computer

File Types from Functional Point of View

2. Data File

- A type of file that contains data
- Data files are created by the software being used
- Different software store data in the data files using different formats
- Data files is generally opened in the same software in which it is created
- It can also be opened in different software that supports the format of that data file

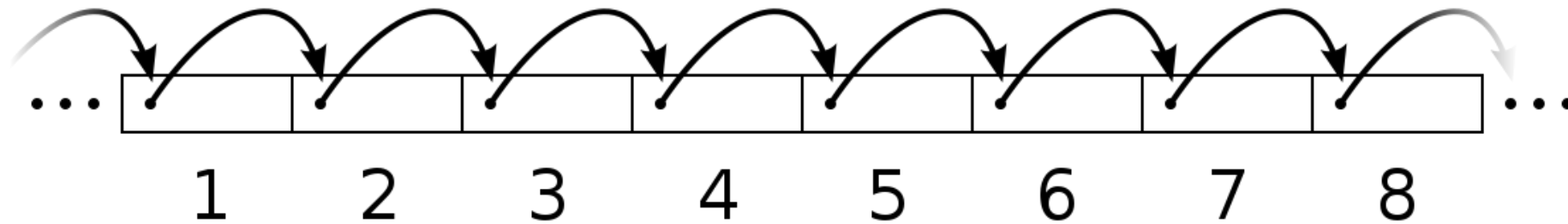
Software	File Type	File Extension
Notepad	Text File	.txt
Word Processor	Document	.doc, .rtf
Spreadsheet	Worksheet	.xls, .wks
Database	Data File	.dat, .dbf, .mdb
Image Processor	Image File	.tif, .jpg, .eps, .gif, .bmp
Audio Software	Audio File	.wav, .mid
Video Software	Video File	.avi, .mpg

File Organization

A technique for physically arranging records of file on secondary storage devices

1. Sequential Files

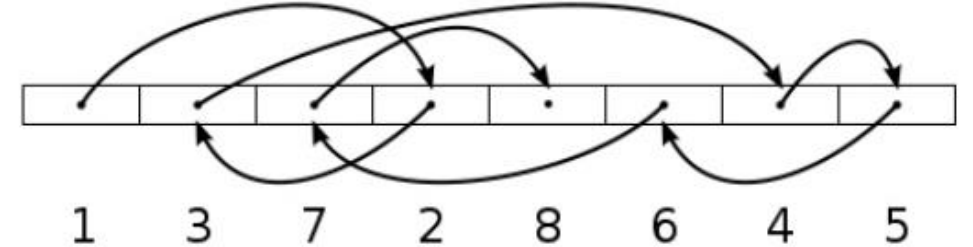
- Records are stored on the storage media in a sequence
- Records can be retrieved only in sequence in which they were stored
- Major disadvantage is very slow access time for a particular record



File Organization

2. Direct or Random Files

- Records are not stored in a particular sequence
- The records are stored at known address or location
- The address or location is calculated against the value of the key field of the record
- Synonym problem → If the same address is calculated to store two or more records
- Faster than sequential file organization for finding a particular record
- Storage media for direct file organization are hard disk , optical discs(CDs, DVDs)



File Organization

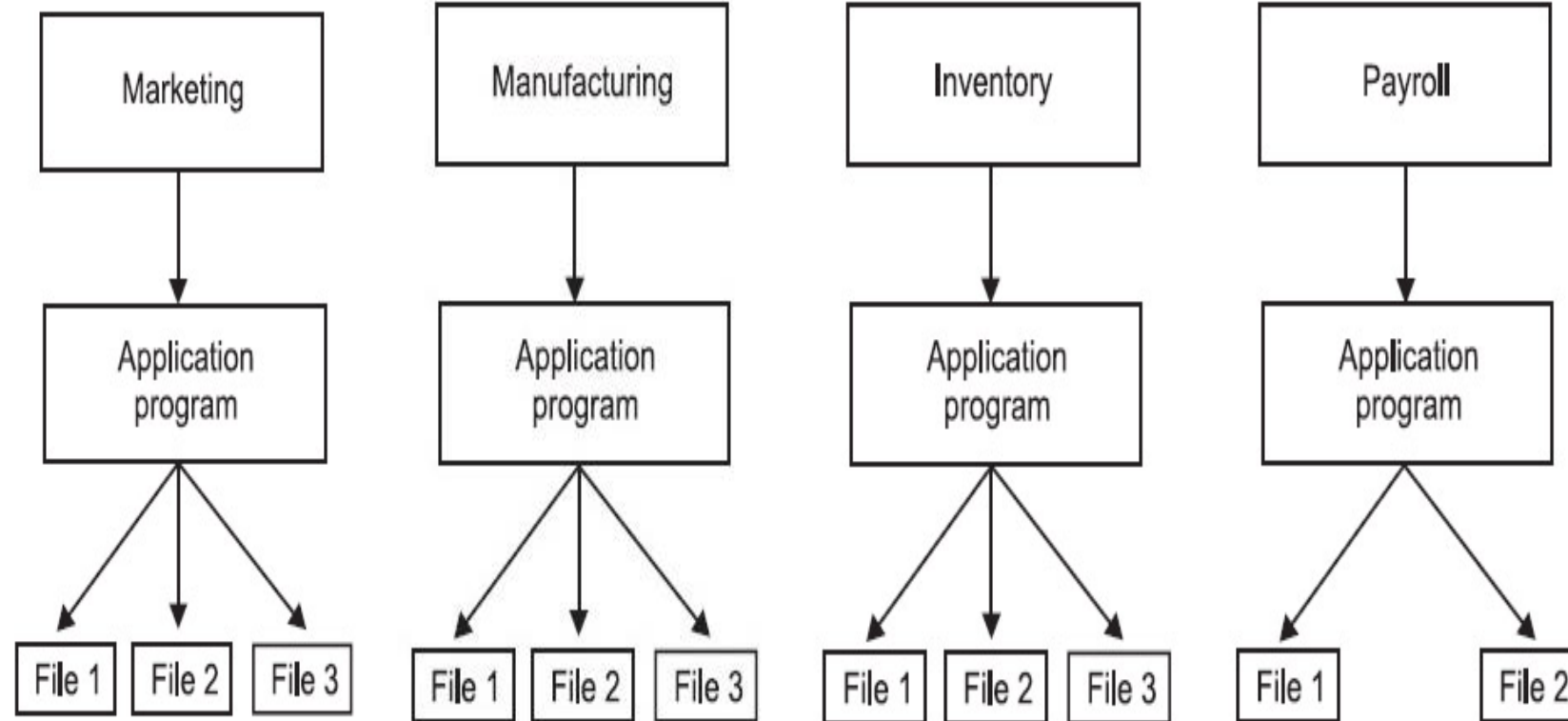
3. Indexed Sequential Files

- Records are stored in ascending or descending order based on value called key
- An index value is generated for each key and mapped with the record
- Index refers to the location or address on a disk where a record is stored
- The index is stored in a file called **index file**
- Index file contains the value of :
 - Each key field
 - Disk address of record with corresponding key field
- Index file is updated whenever a record is added or deleted from the file
- **Main advantage**
 - Allows for both random and sequential processing
- **Main disadvantages**
 - Extra space is required to store indexes
 - Extra time necessary to access and maintain indexes

File Processing System

- This system is used by different organization to store and manage data
- Each department has its own set of data files and application program
- Each program defines and manages its own data
- Every Process generate its separate files and does not communicate with each other

Example



Problems in File Processing System



Example

Suppose that two files are used in a college. The **Students** file contains the data such as **RollNo**, **Name**, **Address**, **Phone** and other details of the students. The **Library** file contains the same data of the students who borrow books from library along with the information about the book. The data of one student appears in two files. It causes wastage of storage and creates many problems.

Problems in File Processing System



Data Inconsistency

Two files may contain different data about the same thing.

Example

The address of a student must be updated in all files if any change occurs. It is possible that it is changed in **Students** file but not in **Library** file. The data becomes inconsistent in this situation.

Problems in File Processing System



Program Dependence

The application program has to be changed if the format of file is changed.

Example

If there is a change in the length of postal code, it requires change in the program. The changes may be costly to implement.

Problems in File Processing System



Lack of Flexibility

Combined reports are very difficult to display as data is scattered in different files.

Example

Suppose a students report is required in the college. The data will be collected from various files to prepare the report. It requires a lot of time and effort to write program for such types of reports in file processing system.

Problems in File Processing System



Example

RollNo and **Marks** of the students should be numeric value. It is very difficult to apply these constraints on files in file processing system.

Problems in File Processing System



Lack of Data Security

It is not possible to define different access levels for different users.

Example

A data entry operator should only be allowed to enter data. The chairman of the organization should be able to access or delete the data completely. Such types of security options are not available in file management system.

Problems in File Processing System



Limited Data Sharing

The file processing system provides a limited data sharing

Each application has its own data files.

It is very difficult for one application program to access the data from a file that is created by other application program.

The file processing system also provides very limited data sharing among different users.

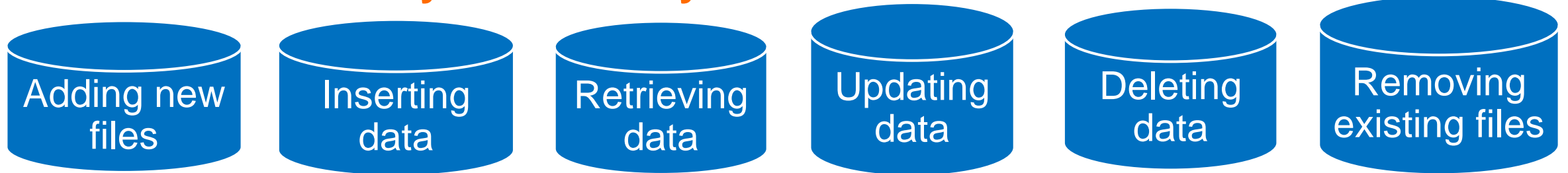
Database

- Collection of logically related data sets or files
- Each data set or file may contain different kind of information

Example – Client Database of Bank have different files:

- Saving Accounts
- Current Accounts
- Automobile Loan
- Personal Loan
- Clients Information etc.

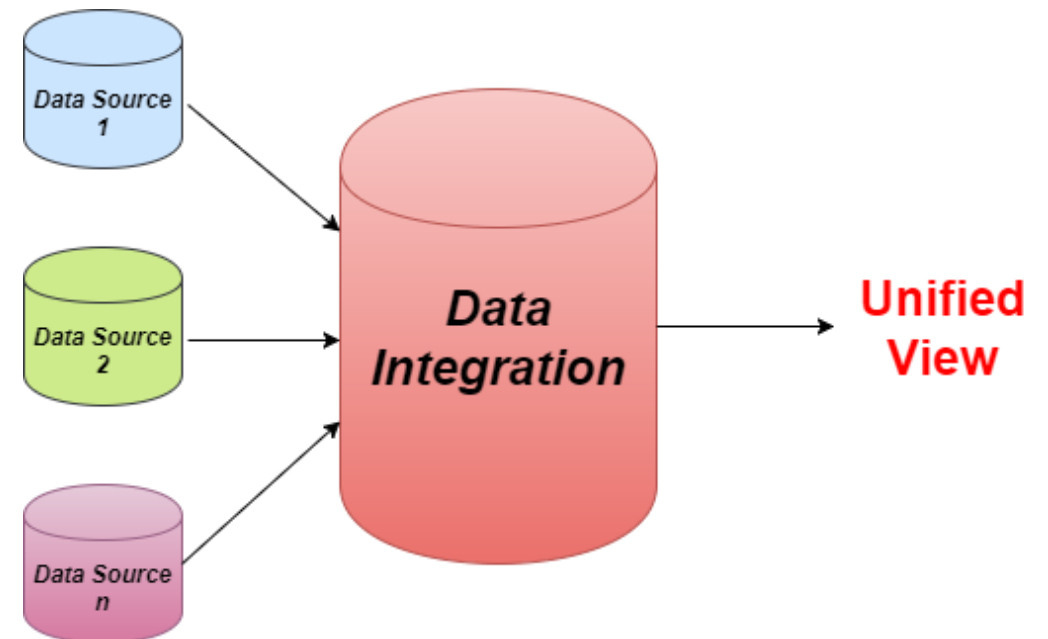
Facilities Provided by database System



Objective of Database

1. Data Integration

- An efficient approach to utilizing data
- Data integration involves combining data located in different computers and providing users with a unified view of them
- Logically, data is centralized



Objective of Database

2. Data Integrity

- Reliability & accuracy of data
- Rules are designed to keep data consistent and correct
- Enforcing data integrity ensures the quality of data
- **Example**
 - Same Employee ID is not assigned to multiple employees

3. Data Independence

- Data and application programs are separate from each other
- User can change data storage structure without changing the application program
- The user can also modify programs without affecting data

Database System

- Collection of data as well as program required to manage that data
- Computerized record-keeping system

Purpose

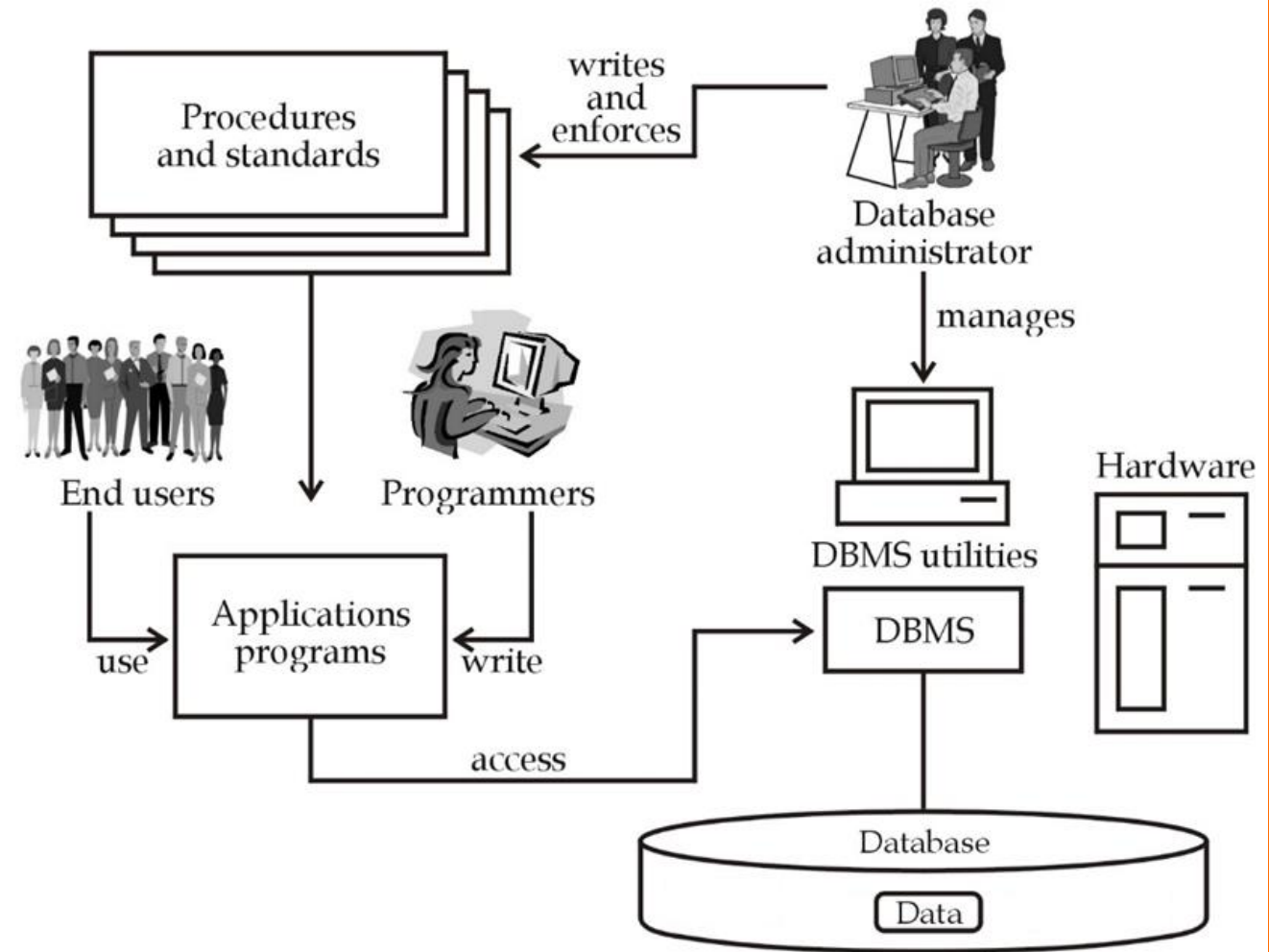
- Maintain data
- Provide it to the user when it is required

Components of Database System

1. Data

- Main purpose of database system is:

- To store data
- To maintain data
- To Process data



Components of Database System

2. Hardware

- Physical components of computer
- Used to perform different tasks such as input, output, storage and processing
- **Example of Hardware components**
 - Secondary storage
 - I/O devices
 - Processors
 - Main memory

Components of Database System

3. Software

- Collection of programs used by the computer within database system
- **DBMS**
 - Use to create and manage a database in database system
- **Application Program**
 - Used to access and process the data stored in database
- **Operating System**
 - Manage all hardware components
 - Enables all other software to run on the computer

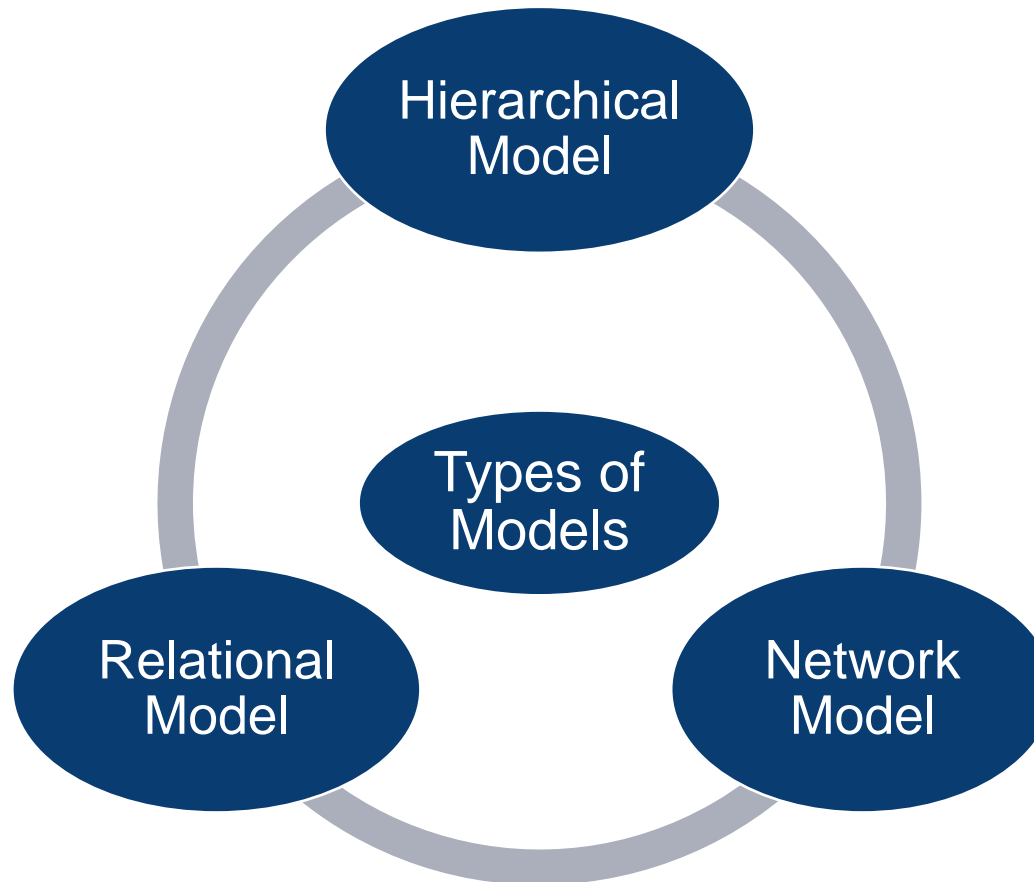
Components of Database System

4. Personnel

- People related to the database system
- **Database Administrator (DBA)**
 - Person who is responsible to manage the whole database system
- **Application Programmer**
 - Person who writes the application program to access data from database
- **End Users**
 - Persons who perform different operations on database
 - Access DBMS through Application program

Data Models

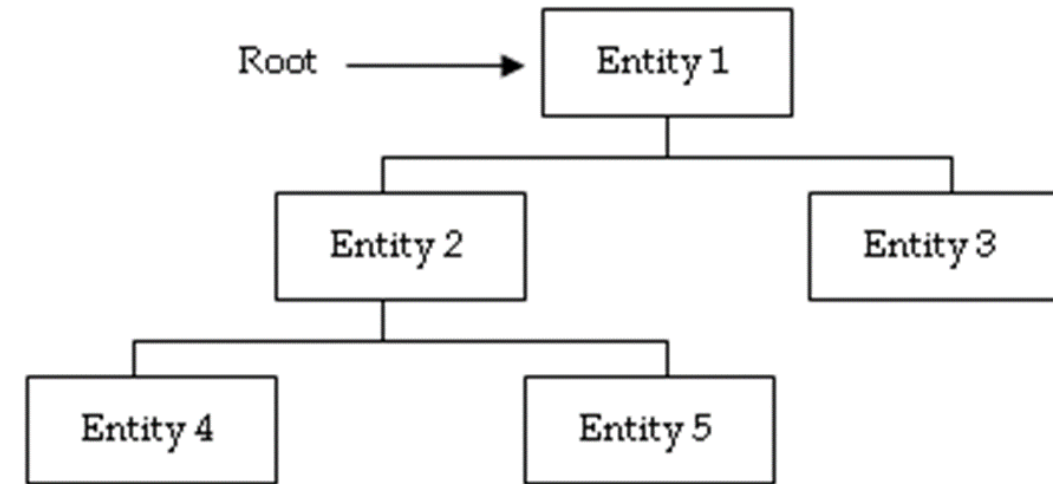
- A set of rules and standards that define how the database organizes data



Data Models

1. Hierarchical Model

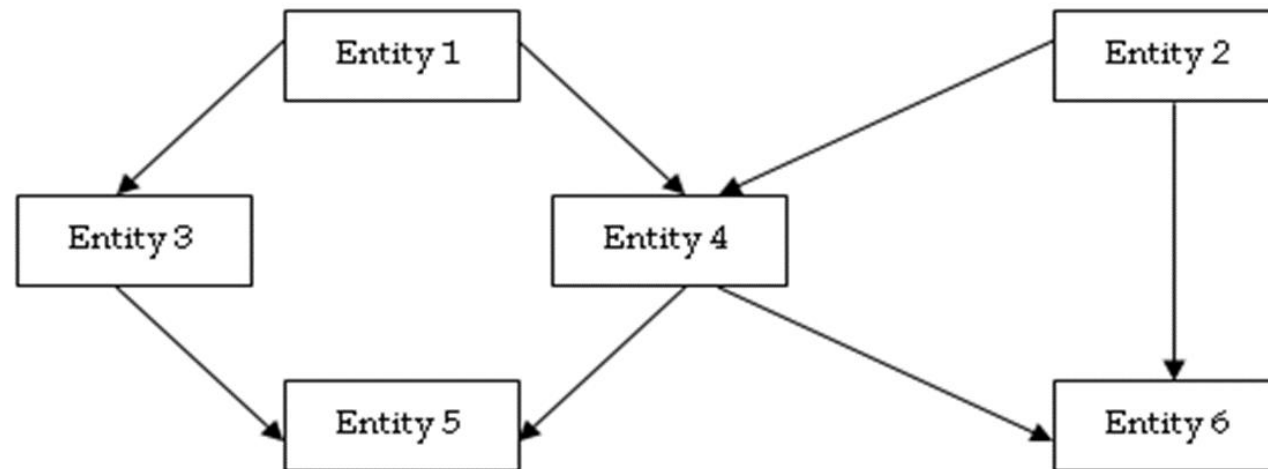
- Records are arranged in a hierarchy like an organizational chart
- Each record type is called a **node** or **segment**
- Node represents a particular entity
- Topmost node is **root**
- Use Parent /Child relationship
 - Each parent node can have many child nodes
 - Each child node may have only one parent node
 - One-to-many relationship between data entities
- Kind of structure → **Inverted tree**



Data Models

2. Network Model

- Similar to hierarchical model but one difference
 - A child node may have any number of parent nodes
- Child nodes represented by arrows
- Complex diagram to represent a database
- Provides more flexibility than hierarchical model



Data Models

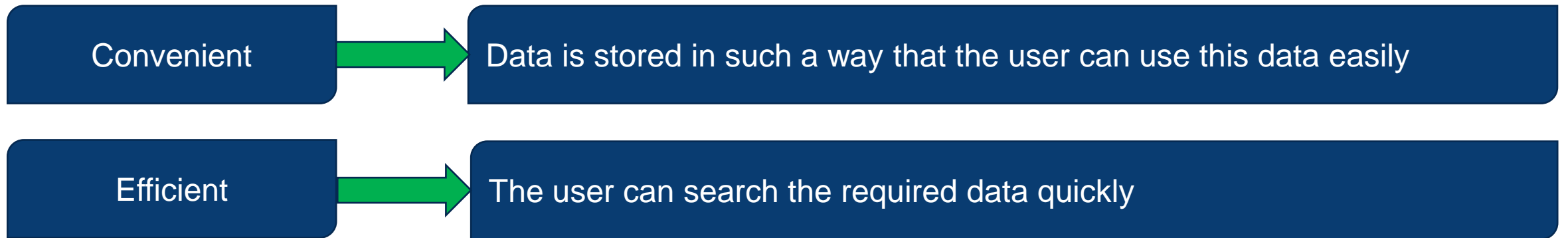
3. Relational Model

- Most commonly used database model
- More flexible than hierarchical and network database model
- Consists of a collection of simple relations or tables
- Relation represents a particular entity to store information about entity
- Relationships are based on the data of the entities
- Relationship between entities is represented by diagram



DBMS

- Collection of programs used to create, maintain, and access databases in convenient and efficient manner



- DBMS uses Database Manager
 - Software
 - Controls the overall structure of database
- Some Relational DBMS vendors/products like:
 - Access & SQL Server/MS , MySQL, Oracle ..etc

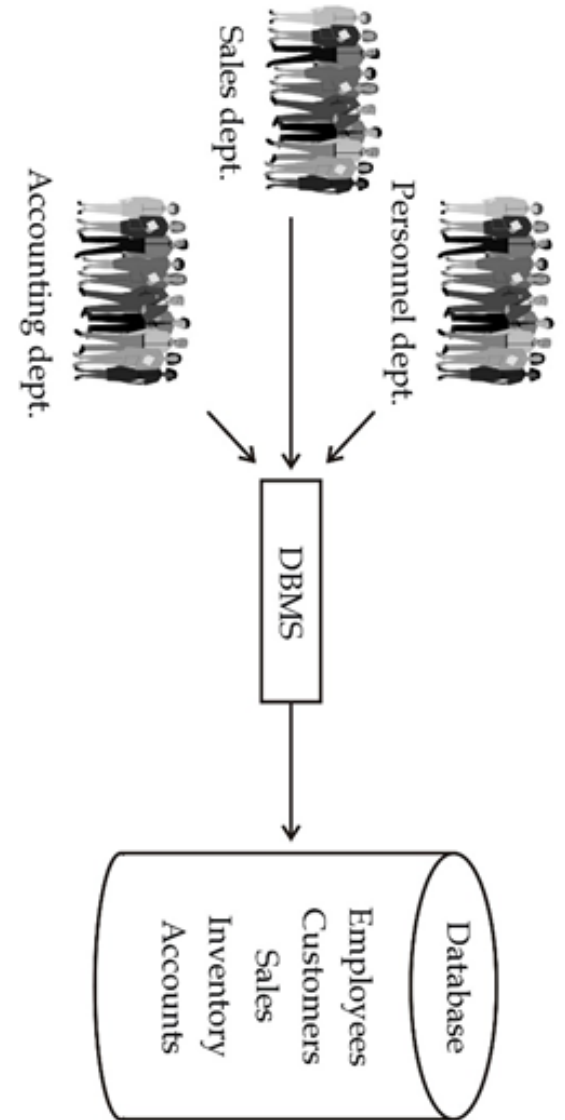
Objectives of DBMS

1. Shareability

- Data is shared by different people at the same time
- Data is stored at central place
- Different users can share the data from different locations
- Reduce storage space and provides data consistency

2. Availability

- Users must be able to access data and DBMS easily
- Data should be available when it is required



Objectives of DBMS

3. Evolvability

- Provide the facility to change the database due to
 - Increase in user requirements or change in the technology

4. Database Integrity

- Accuracy of Data
 - Quality of data entered determines the quality of generated information

Features of DBMS

Data Dictionary

Utilities

Query Language

Report Generator

Access Security

Backup and Recover

Features of DBMS

1. Data Dictionary / Repository

- Contains data definitions for a database:
 - Data Definition is the process of describing the properties of data to be included in a database table
 - During data definition, each field is assigned:
 - Name (must be unique within the table)
 - Data type (such as Text, Number, Currency, Date/Time)
 - Properties (field size, format of the field, allowable range, if field is required, etc.)
 - Finished specifications for a table become the table structure
- Ensures that data is according to the data definition rules
- Used for data access authorization (Password, etc) for database users

Features of DBMS

2. Utilities

- Programs used to maintain database
- Some of these programs are also used for backup and recovery of data

3. Query Language

- A **query** is a request for specific data from the database
- A **query language** consists of simple, English-like statements that allow users to specify the data to display, print, store, update, or delete

Structured Query Language (SQL) is a popular query language that allows users to manage, update, and retrieve data

Features of DBMS

4. Report Generator / Report Writer

- Program that is used to generate reports
- Retrieves data from database and displays it to the user in different formats
- Produce useful and attractive reports by using report generator

5. Access Security

- Protection of database from unauthorized access
- DBMS provides several procedures to maintain data security
 - Allowing access to the database through the use of username and passwords
- Different users have different levels of access rights to database
 - A data entry operator should only be allowed to enter data
 - The chairman of the organization should be able to access or delete the data completely

Features of DBMS

6. Backup and Recover

- DBMS Provides the facility of backup and recover
- Backup facility is used to store an additional copy of data
- Data can be recovered from backup file
 - if the original data file is lost or damaged

Advantages of DBMS

1. Data Independence

- Data and application programs are separate from each other
- User can change data storage structure without changing the application program
- The user can also modify programs without affecting data

2. Redundancy Control

- Redundancy means duplication of data in multiple files → causes wastage of storage
- The data in database appears only once and is not duplicated
- The same data is used at all required places

Advantages of DBMS

3. Consistency Constraints

- Allows user to design complex data structures
- Enables users to view and access data in different ways

4. Data Security

- Protection of database from unauthorized access
- DBMS provides several procedures to maintain data security
 - Allowing access to the database through the use of username and passwords
- Different users have different levels of access rights to database
 - A data entry operator should only be allowed to enter data
 - The chairman of the organization should be able to access or delete the data completely

Advantages of DBMS

5. Backup and Recover

- DBMS Provides the facility of backup and recover
- Backup facility is used to store an additional copy of data
- Data can be recovered from backup file
 - if the original data file is lost or damaged.

6. Advanced Capabilities

- Provides advance capabilities
 - Online access - Access data through Internet

Disadvantages of DBMS

1. High Cost of DBMS

- Expensive to purchase database management software e.g. Oracle

2. Higher Hardware Cost

- DBMS Software requires powerful hardware to work properly and efficiently
 - Requires large size memory and high speed processor

3. Appointing Technical Staff

- DBMS is a complex system
- Technical staff such as database administrator and application programmer is required to manage DBMS
- Pay good salaries to the technical staff increases cost

Disadvantages of DBMS

4. Cost of Staff Training

- DBMS is a complex system
- Requires trained user to use it properly
- Use training is required in all fields:
 - Programming
 - Application development
 - Database administrator
- Spend a lot amount for staff training

Disadvantages of DBMS

5. Problem in Wrong Database Environment

- The problems may occur if wrong type of database environment is selected
- The database system may also need to change due to change in requirements
- The change can be costly due to conversion and testing
- Difficult and time-consuming process
- Require a lot of cost to implement the changes

6. Need of Data Dictionary

- Useful tool but expensive
 - Requires installation costs as well as hardware requirements

Thank you!