



Introduction to Databases



Data and Information

- Data are **raw facts**
 - Building block of information.
 - Information is obtained from processed data
 - Data has to be interpreted in order to obtain information
- Data are a representation of facts, concepts, or instructions
- Data and Information are used interchangeably
- The data in DBMS can be broadly classified into two types
 - collection of information needed by the organization and
 - “**metadata**” which is the information about the database.



What is a Database?

- “A set of information held in a computer”
 - Oxford English Dictionary
- “One or more large structured sets of **persistent** data, usually associated with software to update and query the data”
 - Free On-Line Dictionary of Computing
- “A collection of data arranged for ease and speed of search and retrieval”
 - Dictionary.com



What is a Database (cont.)

- A database is a **well-organized** collection of data that are related in a meaningful way
- A database contains a **model** of something
- The database contains all the data needed by the organization
- several users can access the data in an organization still the integrity of the data should be maintained.
- A database is **integrated** when same information is not recorded in two places



Database Users

- End users
 - Use the database system to achieve some goal
- Application developers
 - Write software to allow end users to interface with the database system
 - Uses a DBMS
- Database Administrator (DBA)
 - Designs & manages the database system
 - Top-level database expert for an organization
- Database systems programmer
 - Writes the database software itself

Database Management System

- A Database System consists of
 - Data (the database)
 - Software
 - Hardware
 - Users
- A DBMS consists of collection of **interrelated** data and a set of programs to **access** that data.
- A software that is used to maintain and use a database.
- A DBMS is a software system designed to
 - **store**,
 - **manage**, and
 - **facilitate access** to databases.

DBMS (cont.)

- A DBMS consists of:
 - A collection of interrelated and persistent data.
 - This part of DBMS is referred to as database (DB).
 - A set of application programs used to access, update, and manage data.
 - This part constitutes data management system
- A DBMS is general-purpose software
 - i.e., not application specific.
 - The same DBMS (e.g., Oracle, SQLServer, etc.) can be used in different systems

DBMS (cont)

- A DBMS takes care of storing and accessing data,
 - leaving only application specific tasks to application programs.
- DBMS allows user to
 - input data,
 - share the data,
 - edit the data,
 - manipulate the data, and
 - display the data in the database.
 - allows more than one user to share the data;



Objectives of DBMS

- The main objectives of database management system are
 - data availability,
 - data integrity,
 - data security, and
 - data independence

1 Data Availability

- Data availability refers to the fact that the data are made available to wide variety of users
 - so that the users can easily access the data.



Objectives of DBMS (cont)

■ 2 Data Integrity

- refers to the correctness of the data in the database.
- the data in the database should be a reliable data.

■ 3 Data Security

- only authorized users can access the data.
- Data security can be enforced by passwords.
- If two separate users are accessing a particular data at the same time, the DBMS must not allow them to make conflicting changes.

■ 4 Data Independence

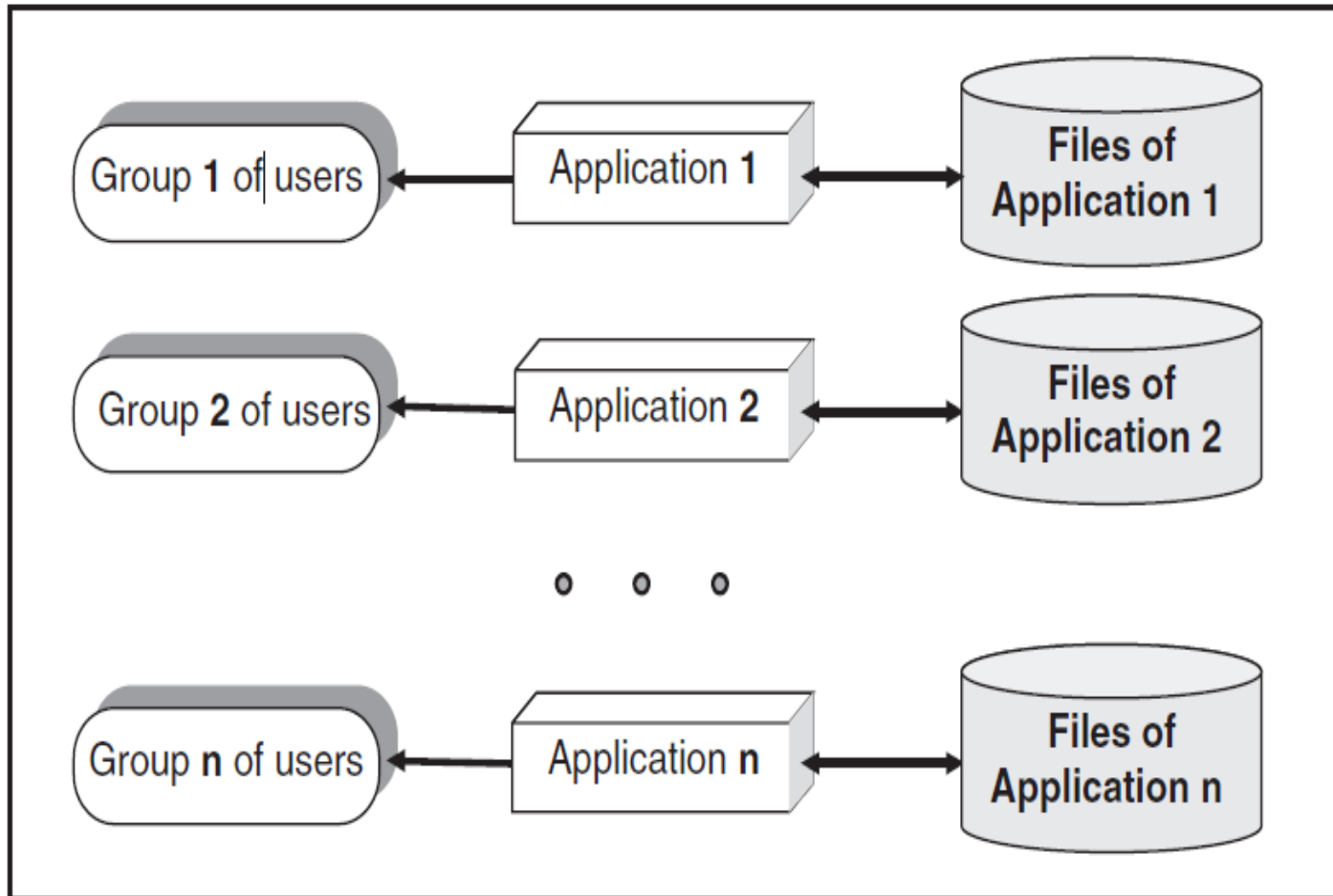
- DBMS provides an “abstract view” of how the data is stored in the database.
- The system hides certain details of how the data are stored and maintained



File-Based System

- Data is stored in distinct files provided by OS
- Application programs are responsible for program logic and data maintenance
- End users can access data via the application programs only
- Each program defines and manages its own data.
- Each file is created and accessed by a separate application program

File-Based System(cont)



Limitations of File-Based System

1 Duplication of Data

- Duplication of data means same data being stored more than once.
- also termed as data redundancy.
- Data redundancy is a problem due to the decentralized data.
- Duplication of data leads to wastage of storage space.
- Has a direct impact on cost
- Duplication of data can lead to loss of data integrity;
- the data are no longer consistent.

Limitations of File-Based System (cont)

2 Data Dependence

- In file-based system there is program-data dependence.
- Data dependence means the application program depends on a specific structure of the data.
- If some modifications have to be made in the data, then the application program has to be rewritten.
- If the application program is independent of the storage structure of the data, then it is termed as data independence.
- Data independence is generally preferred as it is more flexible.

Limitations of File-Based System (cont)

3 Incompatible File Formats

- The structure of the file depends on the application programming language.
- For example, the structure of the file generated by FORTRAN program may be different from the structure of a file generated by “C” program.
- The incompatibility of such files makes them difficult to process jointly.



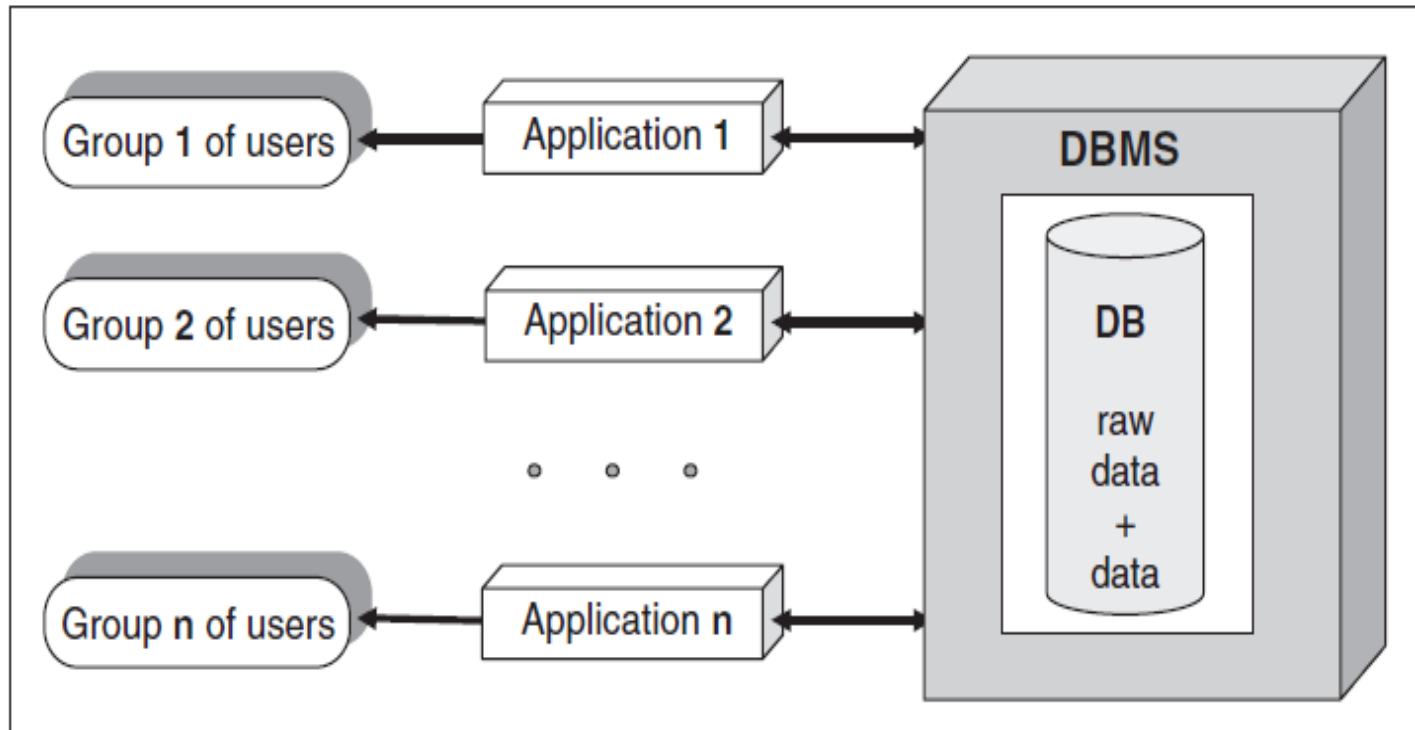
Limitations of File-Based System (cont)

4 Separation and Isolation of Data

- In file-based approach, data are isolated in separate files.
- Hence it is difficult to manage related data.
- The application programmer must synchronize the processing of two files to ensure that the correct data are extracted.
- Data distributed in various files may be in different formats hence it is difficult to share data among different application (Data Isolation).

The DBMS Approach

- DBMS is software that provides a set of functions for defining, accessing, and manipulating data.
- The same data are being shared by different application programs; as a result, data





Advantages of DBMS

- Some of the advantages of database management system:
 - Centralized data management.
 - Data Independence.
 - System Integration.

1 Centralized Data Management

- In DBMS all files are integrated into one system
- reducing redundancies and
- making data management more efficient



Advantages of DBMS (cont.)

2 Data Independence

- Data independence means that programs are isolated from changes in the way the data are structured and stored.
- the database management system provides the interface between the application programs and the data.
- applications need not worry about how the data are physically structured and stored.
- Applications should work with a logical data model and declarative query language.

Advantages of DBMS (cont.)

3 Data Inconsistency

- Data inconsistency means different copies of the same data will have different values.

- Other Advantages of a DBMS
 - Data independence
 - Efficient data access
 - Data integrity & security
 - Data administration
 - Concurrent access, crash recovery
 - Reduced application development time