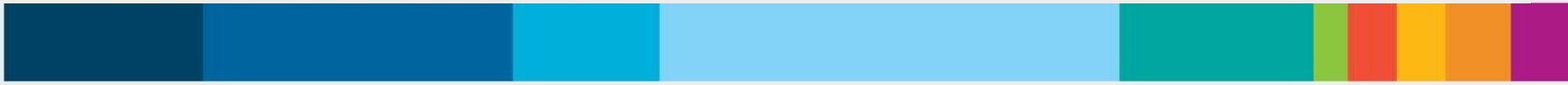
IBM Brighter Blue

# Introduction to Database Management Systems



IBM

# Module Overview

This module will introduce you to the fundamentals of data processing and should take about 60 minutes of your time.

At the end of this module, you should be able to:

- Explain the fundamentals of data processing
- Identify the various data processing models
- Describe the traditional method of data storage
- Identify problems associated with the traditional approach
- Explain the database management system (DBMS)
- Identify the different types of databases
- Examine the detailed system architecture of a database
- Identify the users of a DBMS
- List the advantages of a DBMS

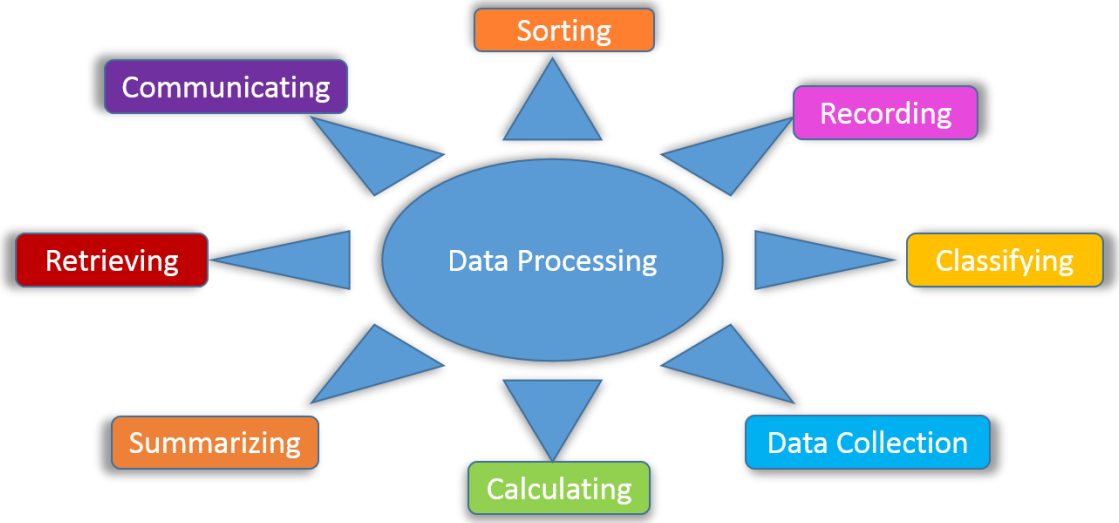
# 01 What is Data Processing?

# What is Data Processing?

**Data processing**

Data processing is any process that uses a computer program to enter data and summarize, analyze, or otherwise convert data into usable information.

Data Processing, in an enterprise, encompasses various activities.



# Data Processing Models

Let us learn about some of the data processing models.

*Click each data processing model to learn more.*


In this model, transactions are collected in a group and processed together.


In this model, transactions are processed as and when they appear.

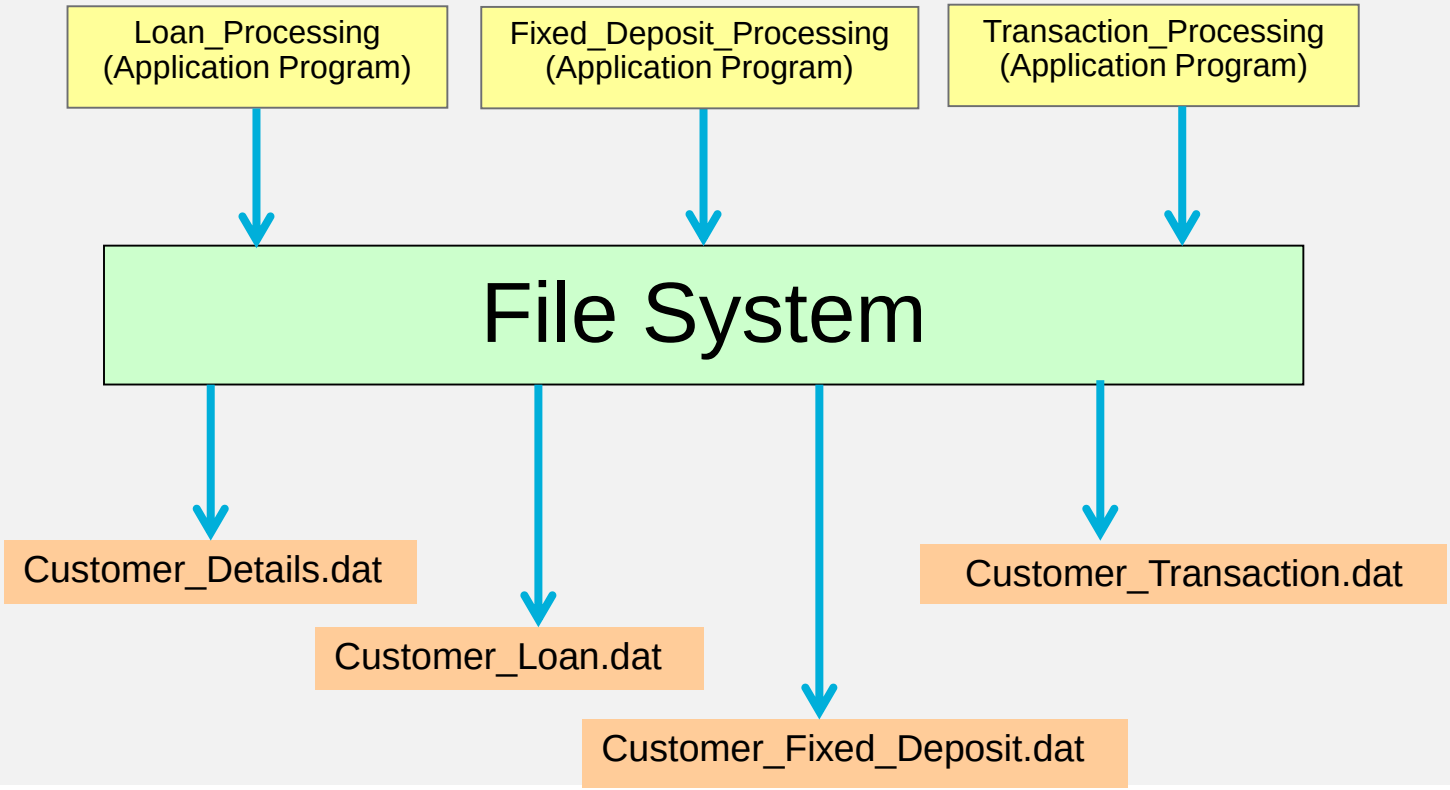

In this model, there is a parallel time relationship with on-going activity. The information that is produced is useful in controlling the current or dynamic activity.

# 02 Traditional Method of Data Storage

# How was Data Stored in the Traditional Method?

Traditionally, data was organized in *file* formats for easy access. Each department in the organization owned its data by creating its own files.

The diagram below depicts a traditional data storage method.



# Storing Customer Data in Files

In the traditional method, the customer data was stored in the form of records in the files. These records consisted of various fields which were delimited by a space, comma, tab, and so on. Special characters used to mark the end of records and end of files.

Here is a sample table that shows the ways of storing customer data in files.

| | | |
|---|---|---|
| 4176 | Aniruddha Sarkar | SBU1 |
| 4181 | Manoj Saha | SBU1 |
| 4183 | Moushumi Dharchoudhury | SBU1 |
| 4203 | Suryanarayana D.V.S.S. | SBU1 |
| 4204 | Vivek Rai | SBU1 |

Predefined length

# Problems Associated with the Traditional Method

Data Redundancy

Difficulty in Accessing Data

Data Integrity

Concurrent Access Anomalies

**Data redundancy**

This means same information is duplicated in several files. This makes data redundant or unnecessary.

Data Inconsistency

Data Isolation

Atomicity

Security Issues

*Click each button to learn more*

# Problems Associated with the Traditional Method

Data Redundancy

Difficulty in Accessing Data

Data Integrity

Concurrent Access Anomalies

## Data inconsistency

This means different copies of the same data are not matching, that is, different versions of the same basic data are existing. This occurs as a result of update operations that do not update the same data stored at different places.
For example: The Address information of a customer is recorded differently in different files.

Data Inconsistency

Data Isolation

Atomicity

Security Issues

*Click each button to learn more*

# Problems Associated with the Traditional Method

Data Redundancy

Difficulty in Accessing Data

Data Integrity

Concurrent Access Anomalies

**Accessing data**

It is difficult to access data as it is not easy to retrieve information using a conventional file processing system. Convenient and efficient information retrieval is almost impossible.

Data Inconsistency

Data Isolation

Atomicity

Security Issues

*Click each button to learn more*

# Problems Associated with the Traditional Method

Data Redundancy

Difficulty in Accessing Data

Data Integrity

Concurrent Access Anomalies

**Data isolation**

This indicates that data is scattered in various files, which may be in different formats. Therefore, writing a new application program to retrieve the data is difficult.

Data Inconsistency

Data Isolation

Atomicity

Security Issues

*Click each button to learn more*

# Problems Associated with the Traditional Method

Data Redundancy

Difficulty in Accessing Data

Data Integrity

Concurrent Access Anomalies

**Data integrity problems**

This implies that the data values may need to satisfy some integrity constraints. For example, the balance field Value must be greater than 5000. You can handle this through a program code in the file processing systems. However, in the database, you can declare the integrity constraints along with the definition itself.

Data Inconsistency

Data Isolation

Atomicity

Security Issues

*Click each button to learn more*

# Problems Associated with the Traditional Method

Data Redundancy

Difficulty in Accessing Data

Data Integrity

Concurrent Access Anomalies

## Atomicity

It is difficult to ensure atomicity in the file processing system. For example: transferring $100 from account A to account B. If a failure occurs during execution there could be a situation where $100 is deducted from Account A and not credited to Account B.

Data Inconsistency

Data Isolation

Atomicity

Security Issues

*Click each button to learn more*

# Problems Associated with the Traditional Method

Data Redundancy

Difficulty in Accessing Data

Data Integrity

Concurrent Access Anomalies

**Concurrent access anomalies**

If multiple users are updating the same data simultaneously, it will result in inconsistent data state. In the file processing system, it is very difficult to handle this using a program code. This results in concurrent access anomalies.

Data Inconsistency

Data Isolation

Atomicity

Security Issues

*Click each button to learn more*

# Problems Associated with the Traditional Method

Data Redundancy

Difficulty in Accessing Data

Data Integrity

Concurrent Access Anomalies

## Security Issues

The data is maintained in flat files that is easily accessible and hence, not secure. Enforcing security constraints in the file processing system is very difficult as the application programs are added to the system in an ad-hoc manner.

Data Inconsistency

Data Isolation

Atomicity

Security Issues

*Click each button to learn more*

# What is the Solution?

As we just saw the traditional method of data storage is associated with many problems. So, what is the solution?

There is a more effective approach of data storage – the Database approach!

In today's world, where information matters most, Database Management Systems (DBMS) form an integral part of an organization. Large enterprises depend on the DBMS for their business operations in some way or the other.

Let us, therefore, understand the concept of DBMS and how you can efficiently design and implement databases to build robust solutions.

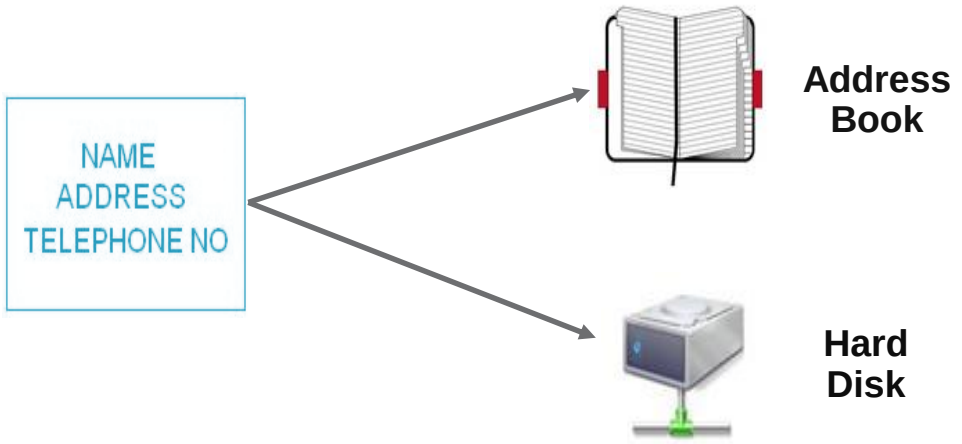
# 03 Database Management Systems

# What is a Database?

**Database**

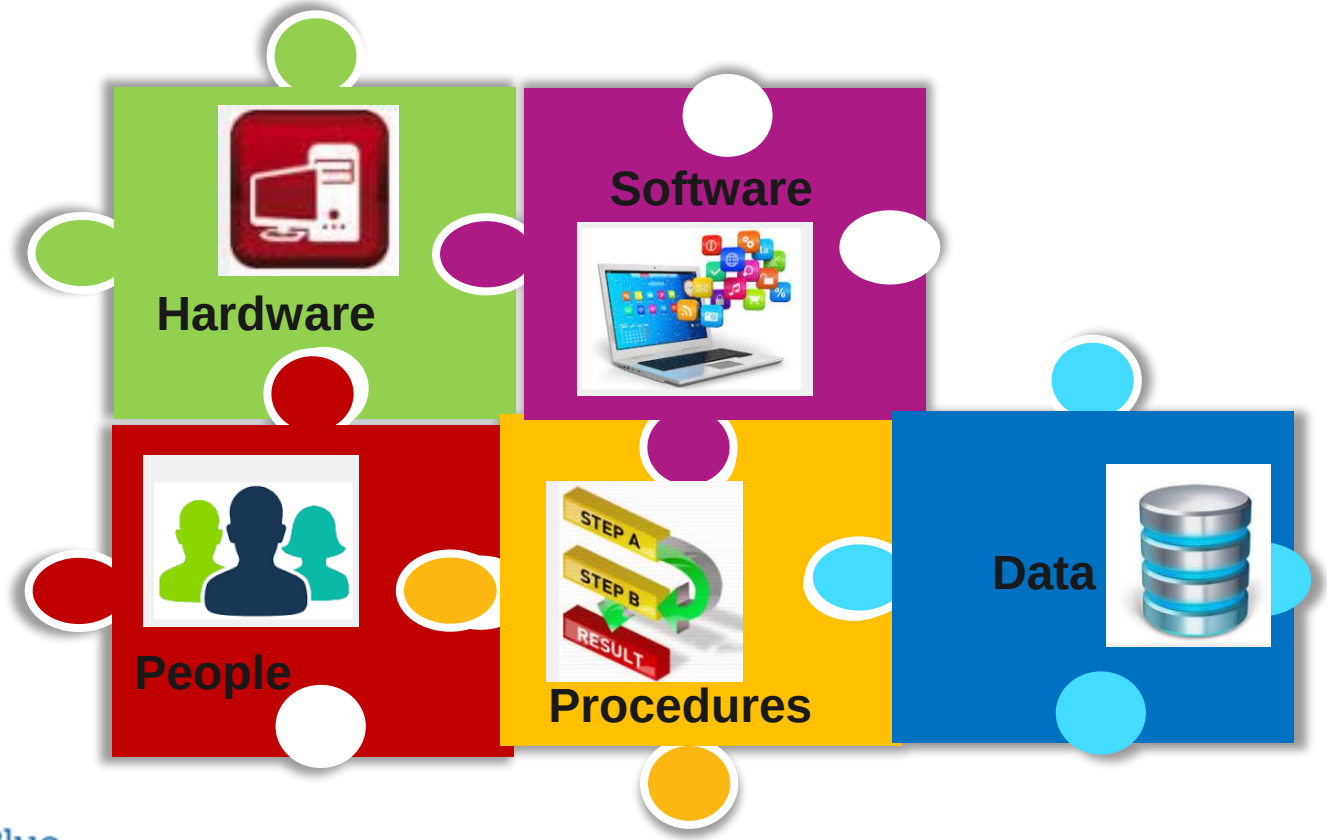
A database is a collection of logically related data. Database is an integral part of an organization.

- Data means known facts, which are meaningful and can be recorded.
- For example: name, address, telephone number form an address book, whether stored manually in an address book, or on a hard disk in the form of a file.

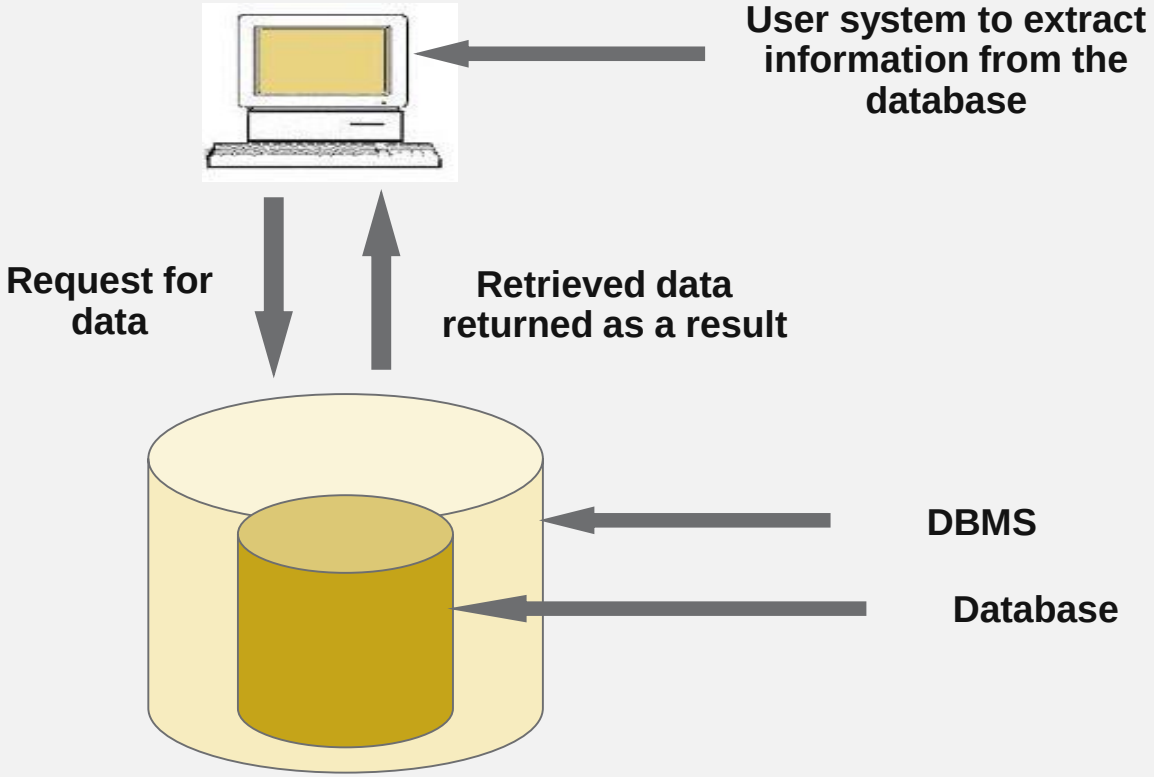


# Important Components of the Database System

The database system refers to an organization of components that define and regulate the collection, storage, management, and use of data within a database environment. In a high level view, the database system is composed of the following five major parts.



# How does the DBMS work?

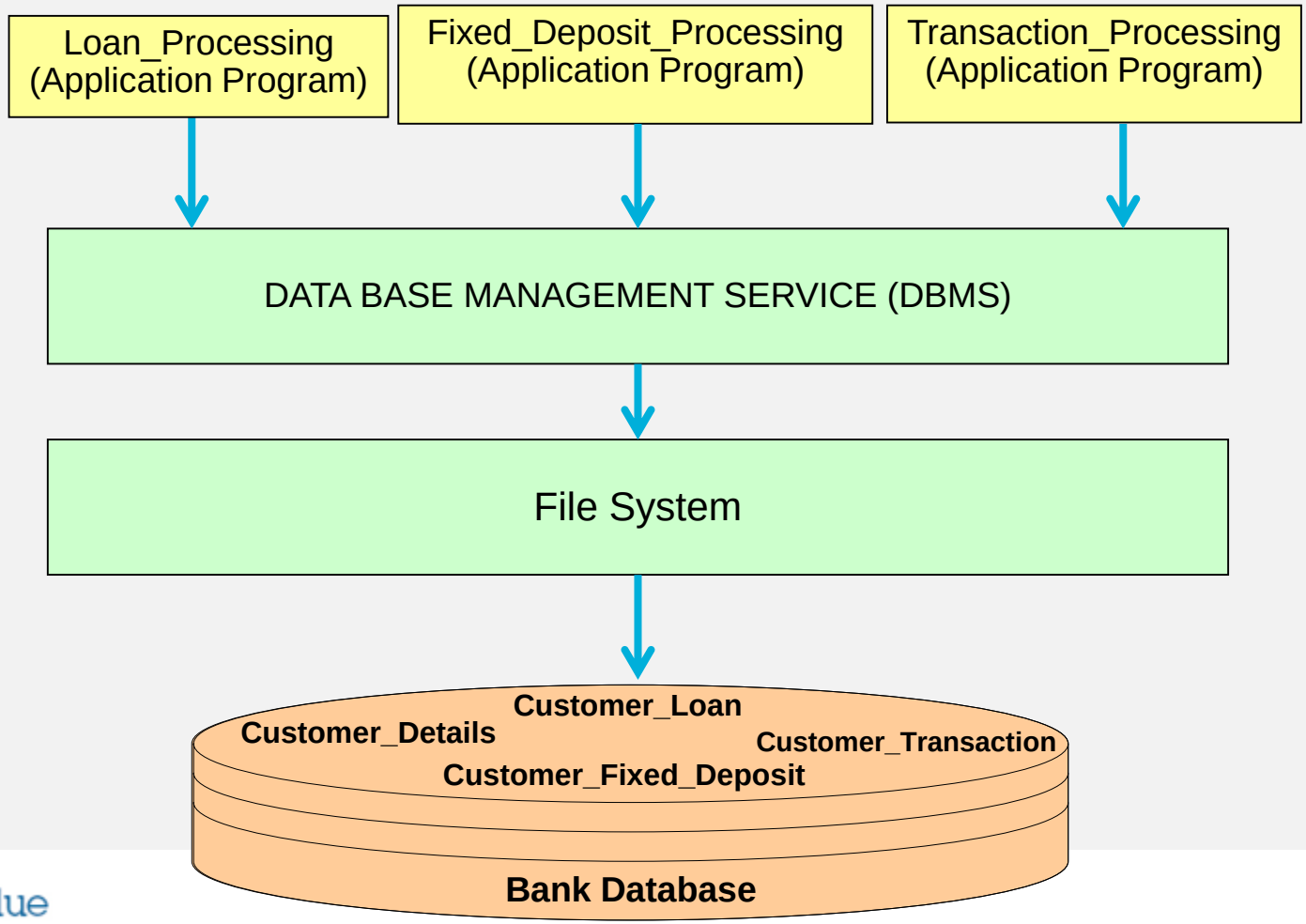



There are three technologies involved to make a DBMS work:

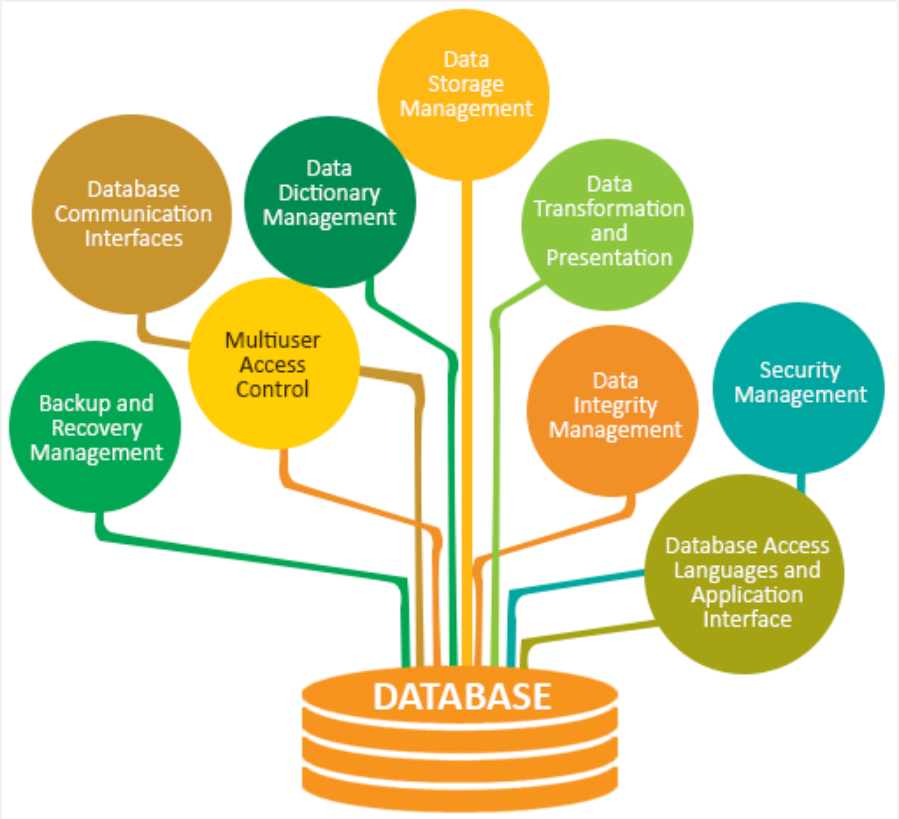
- **Networking**
- **Client Server Architecture**
- **Relational Algebra (forms the mathematical foundation for the data model)**

# Where does the DBMS fit in?

The diagram shows the position of a DBMS. The DBMS acts as a layer of abstraction on top of the **File** system.



# Functions of DBMS

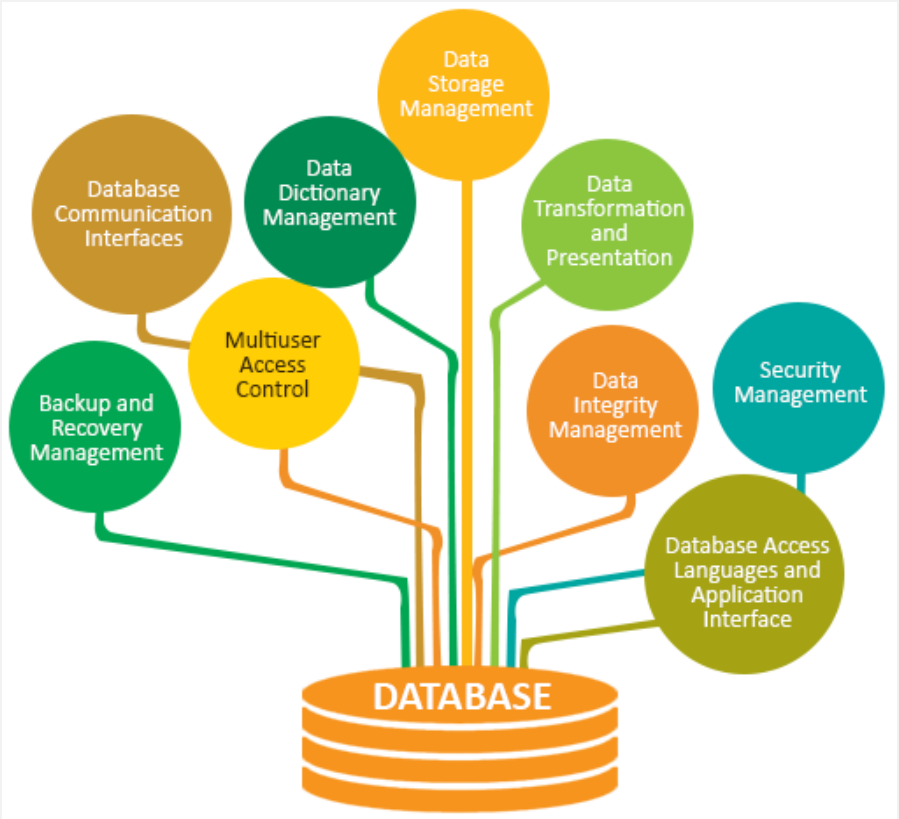



The Database Management System performs various functions.

*Click each function on the tree to learn more.*

**Data Dictionary Management**
The data dictionary stores the definitions of data elements and their relationships. This information is termed as metadata. The metadata includes definition of data, data types, relationship between data, integrity constraints etc. Any changes made in a database structure are automatically reflected in the data dictionary. In short the DBMS provides data abstraction and it removes structural and data dependency from the system.

# Functions of DBMS



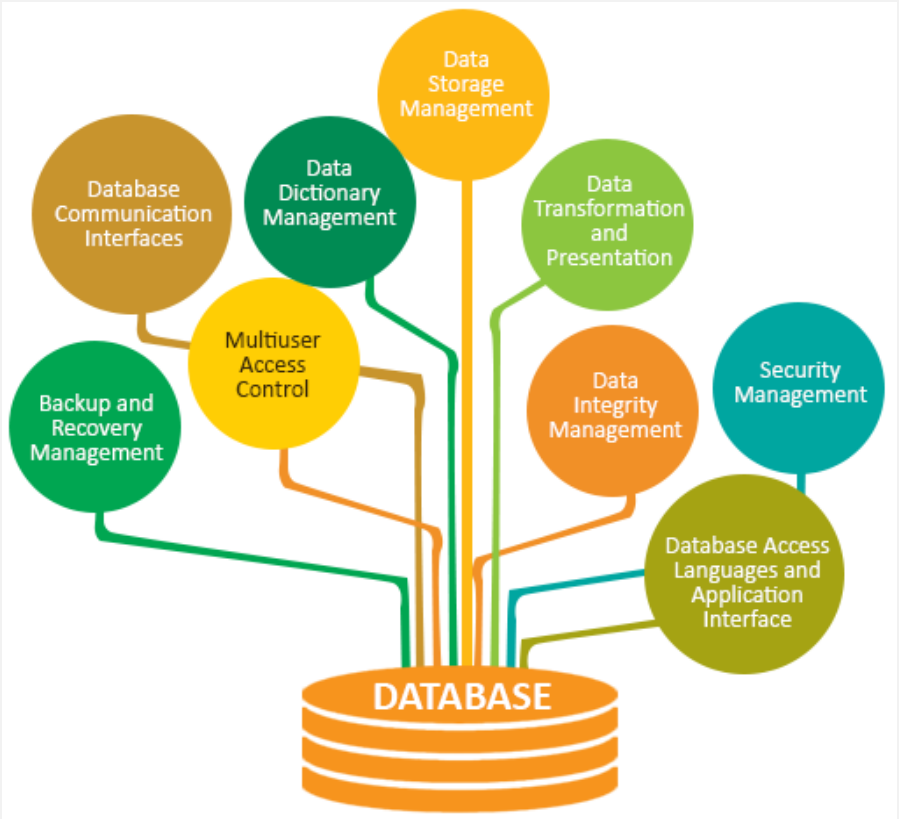

The Database Management System performs various functions.

*Click each function on the tree to learn more.*

**Data Storage Management**

The DBMS creates the complex structures required for data storage. The users are freed from defining, programming, and implementing the complex physical data characteristics.

# Functions of DBMS



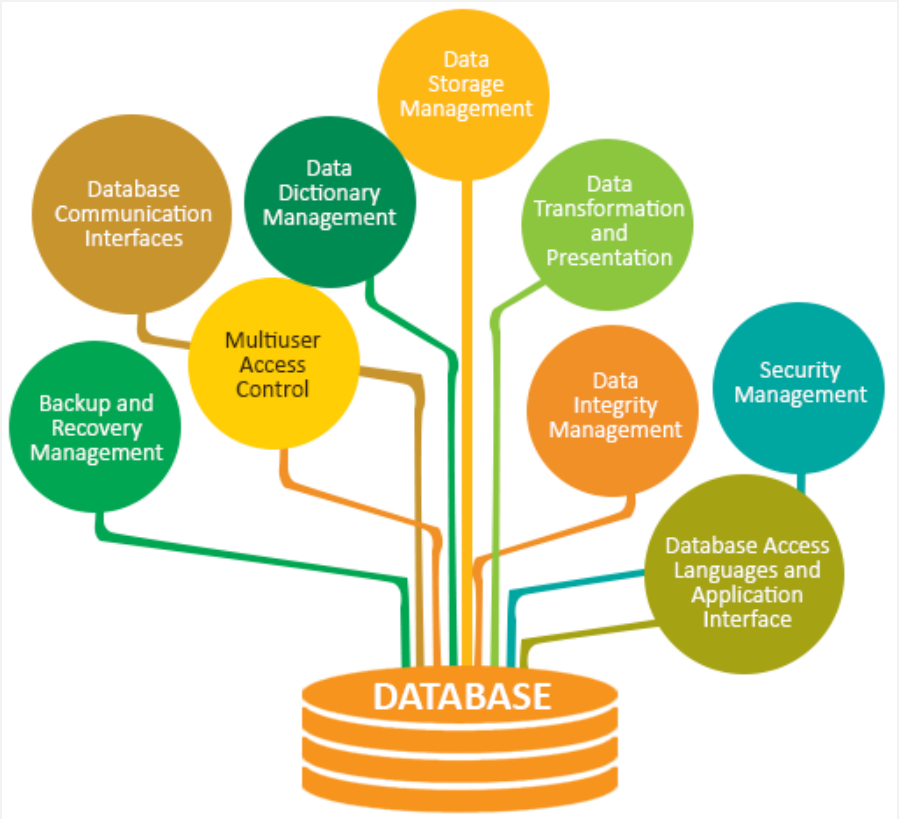

The Database Management System performs various functions.

*Click each function on the tree to learn more.*

**Data Transformation and Presentation**

DBMS supports data independence. Hence, the DBMS translate logical request into commands that physically locate and retrieve the requested data. The DBMS formats the physically retrieved data according to the logical data format specifications.

# Functions of DBMS



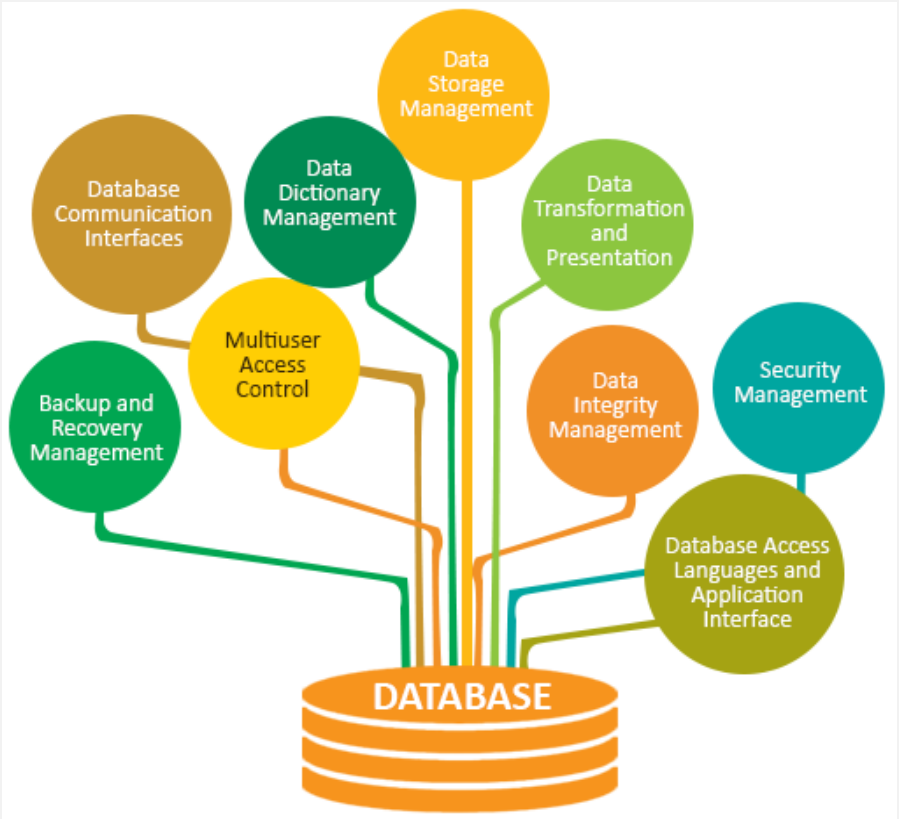

The Database Management System performs various functions.

*Click each function on the tree to learn more.*

**Security Management**

The DBMS creates a security system that enforces user security and data privacy within the database. Security rules determine the access rights of the users. Read or write access given to the user is specified using access rights.

# Functions of DBMS



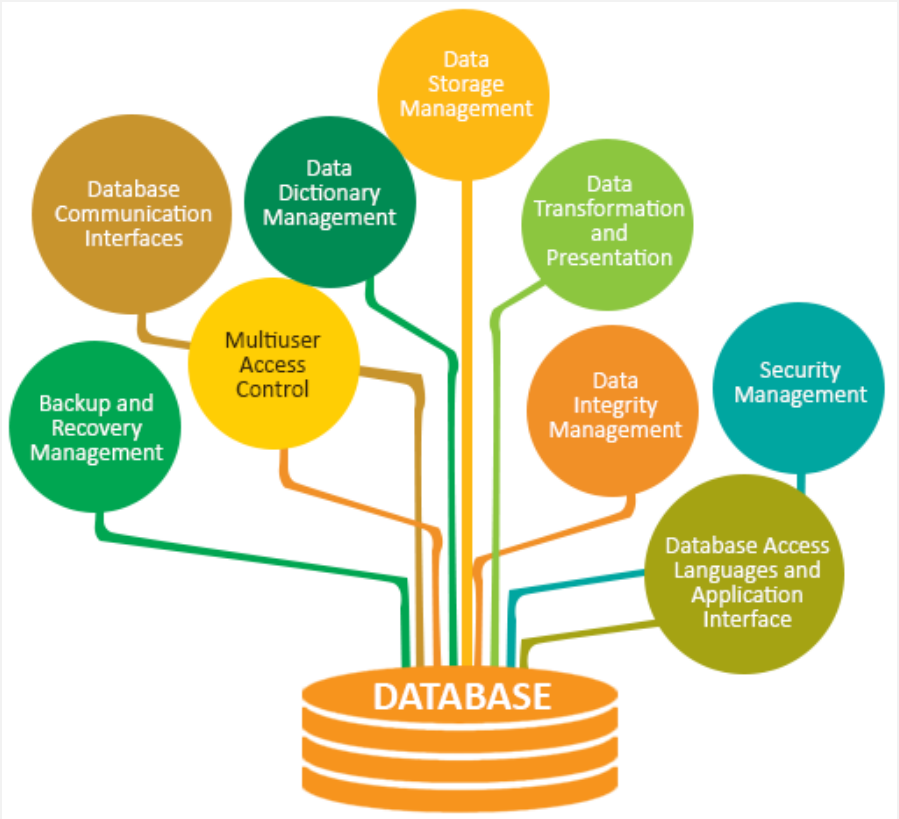

The Database Management System performs various functions.

*Click each function on the tree to learn more.*

**Multiuser Access Control**

The DBMS ensures that multiple users can access the database concurrently without compromising the integrity of the database. Hence, the database ensures data integrity and data consistency.

# Functions of DBMS



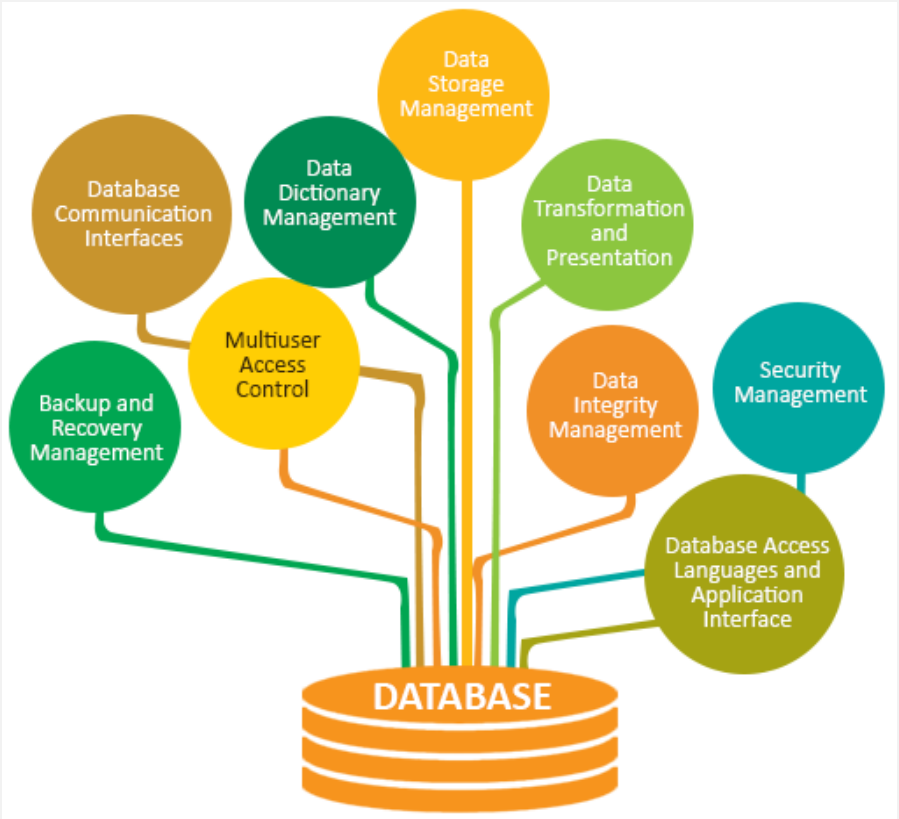

The Database Management System performs various functions.

*Click each function on the tree to learn more.*

**Backup and Recovery Management**

The DBMS provides backup and data recovery procedures to ensure data safety and integrity. The DBMS system provides special utilities which allow the DBA to perform routine and special backup and restore procedures. Recovery Management deals with the recovery of the database after a failure.

# Functions of DBMS



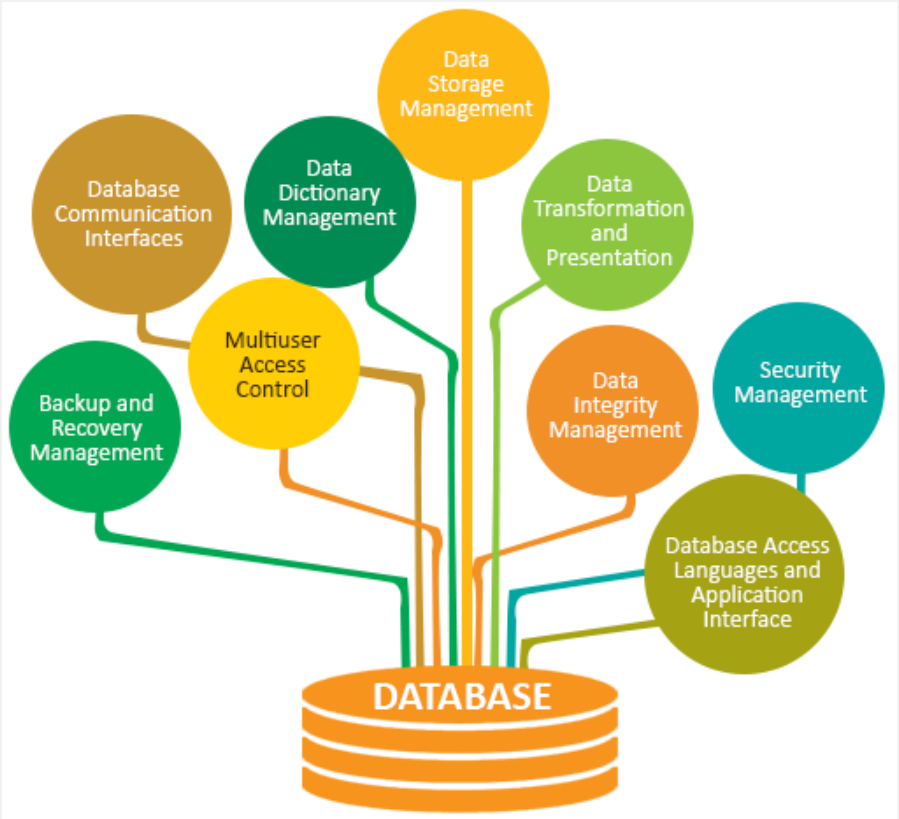

The Database Management System performs various functions.

*Click each function on the tree to learn more.*

**Data Integrity Management**

The DBMS promotes and enforce integrity rules to eliminate data integrity problems, thus minimizing the data redundancy and maximizing data consistency.

# Functions of DBMS



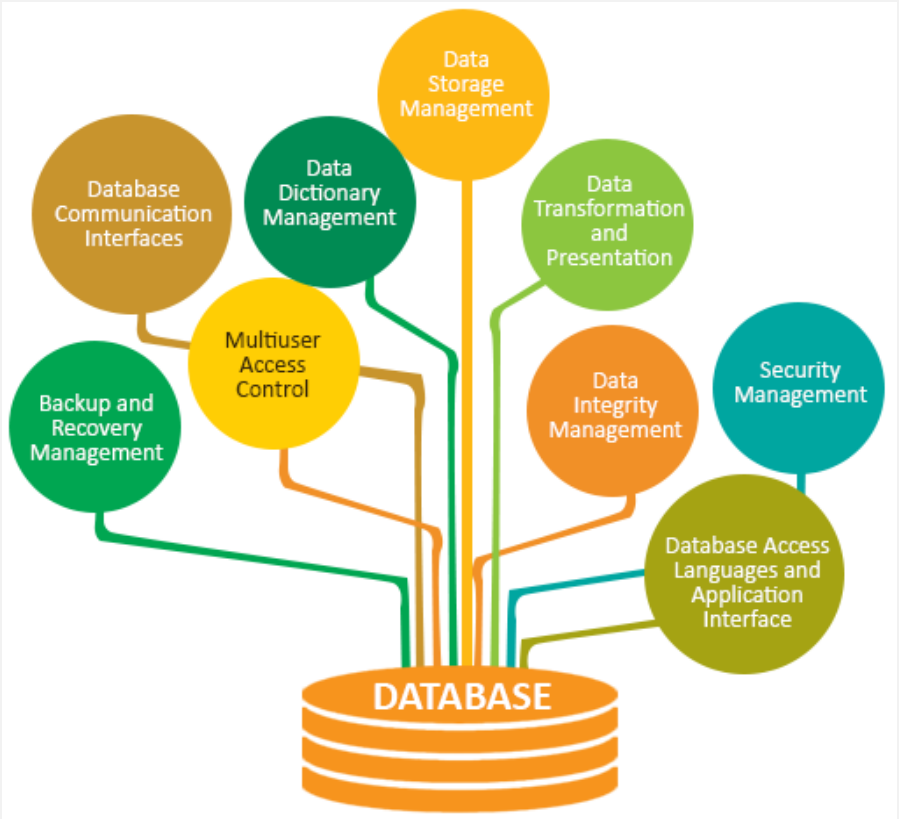

The Database Management System performs various functions.

*Click each function on the tree to learn more.*

**Database Access Languages and Application Interface**

The DBMS provides data access via query language. A query language is a non-procedural language, that is, the user only needs to specify what must be done without specifying how it is to be done. The DBMS's query language contains two components: a data definition language (DDL) and a data manipulation language (DML). The DBMS also provide data access to programmers via programming languages.

# Functions of DBMS




The Database Management System performs various functions.

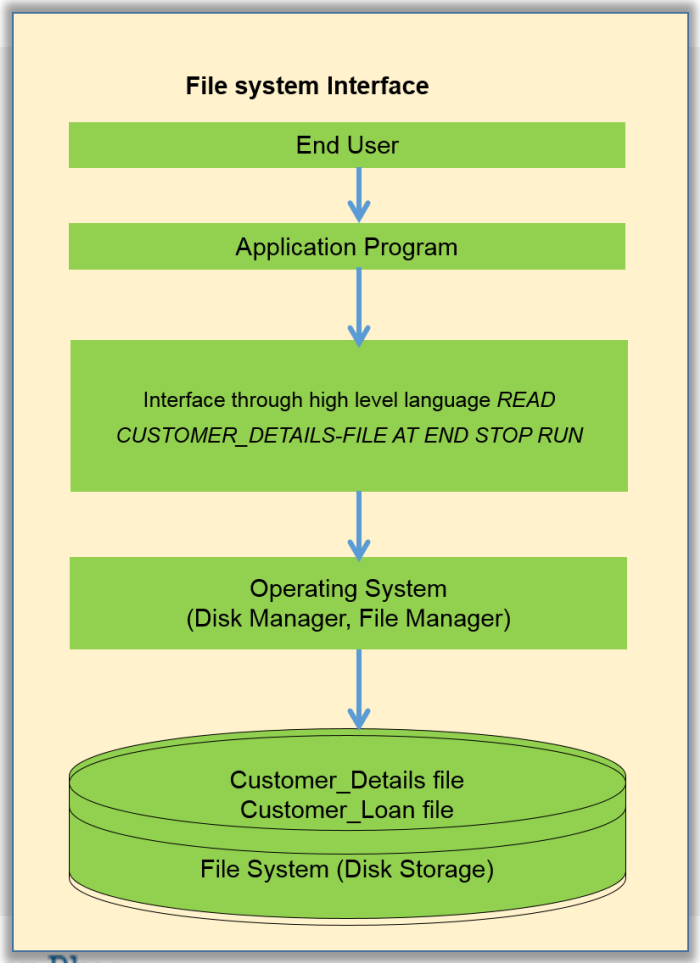
*Click each function on the tree to learn more.*

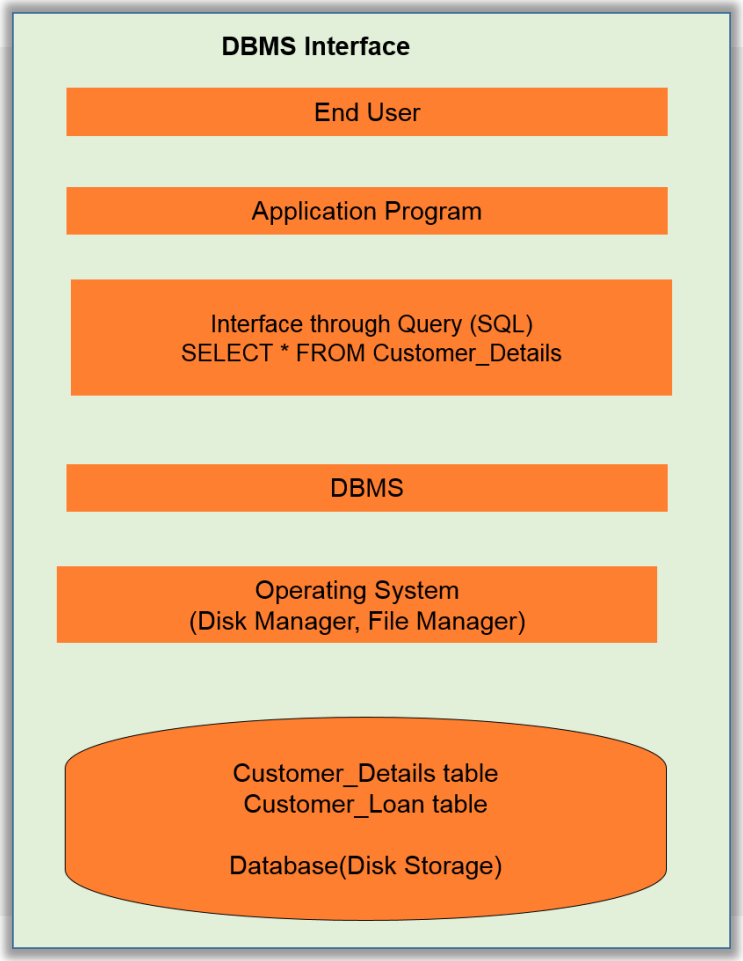
**Database Communication Interfaces**

Different users may access the database through a network environment. So the DBMS provides communication functions to access the database through computer network environment.

# File versus DBMS Operations

Let us now understand the difference between the traditional file operations versus the DBMS operations.



# Spot Quiz

**01** In which type of data processing model, transactions are processed as and when they appear?

A. Batch processing

B. Online processing

C. Real-time processing

D. Transaction processing

# Spot Quiz

**03**

Which of the following is not true with regard to the traditional method of data storage?

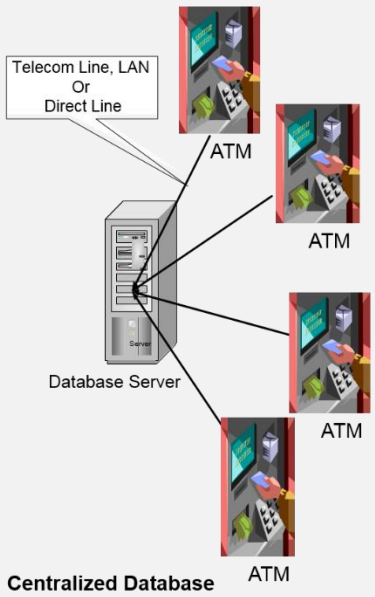
| | |
|---|---|
| A | Data redundancy |
| B | Better flexibility |
| C | Enforces integrity constraints |
| D | Data isolation |

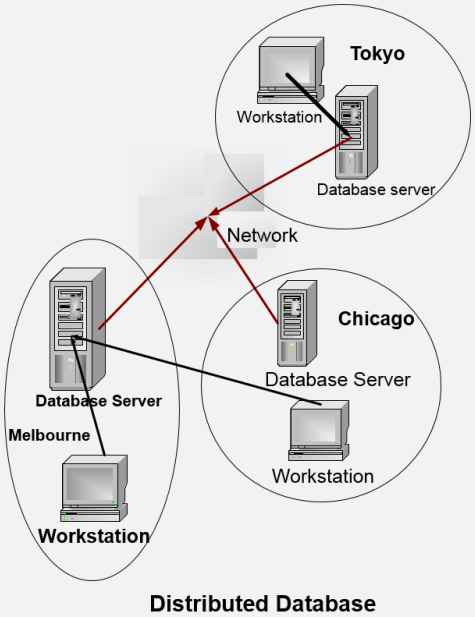
# 04 Types of Databases

# Types of Databases

The databases may be of different types. However, there are two generic database architectures.

*Click each database type to learn more*.

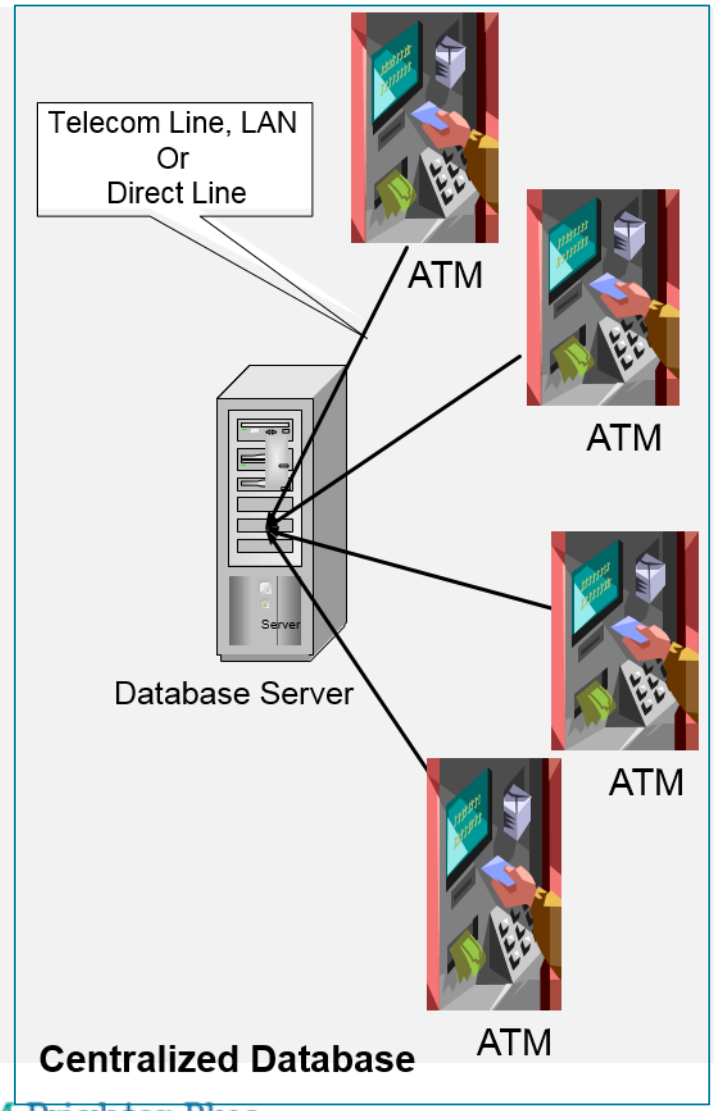



**Centralized Database**

**Distributed Database**

# Centralized Database




A **Centralized** database is a single central database accessed by multiple users.

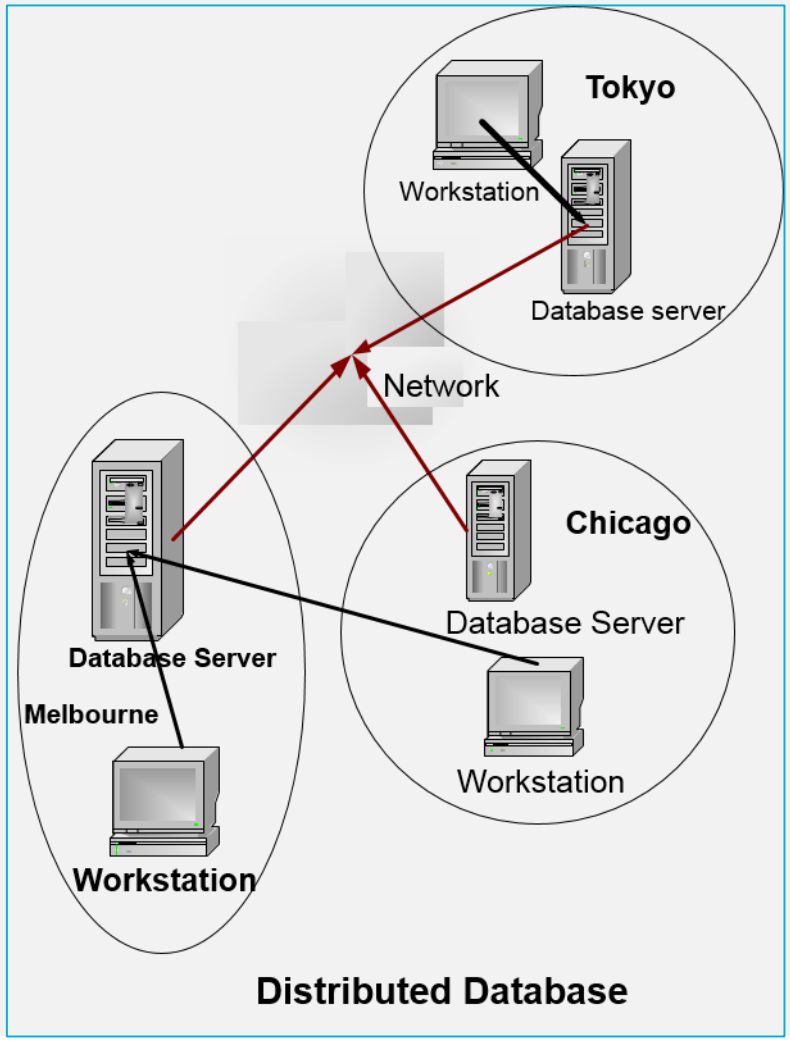
The features of the Centralized database are:

- ✓ Data stored at a single site
- ✓ Greater control over accessing and updating data
- ✓ Vulnerable to failure as there is dependency on the availability of resources at the central site.

Let us understand this type of database with an example.

The account information of customers is stored in a particular branch office of a bank. This information must be shared across all Automated Teller Machines (ATM) so that customers can withdraw money from their accounts. Instead of storing the customer information in every ATM machine, it can be stored at a common place (the branch office of the bank) and shared over a network.

# Distributed Database



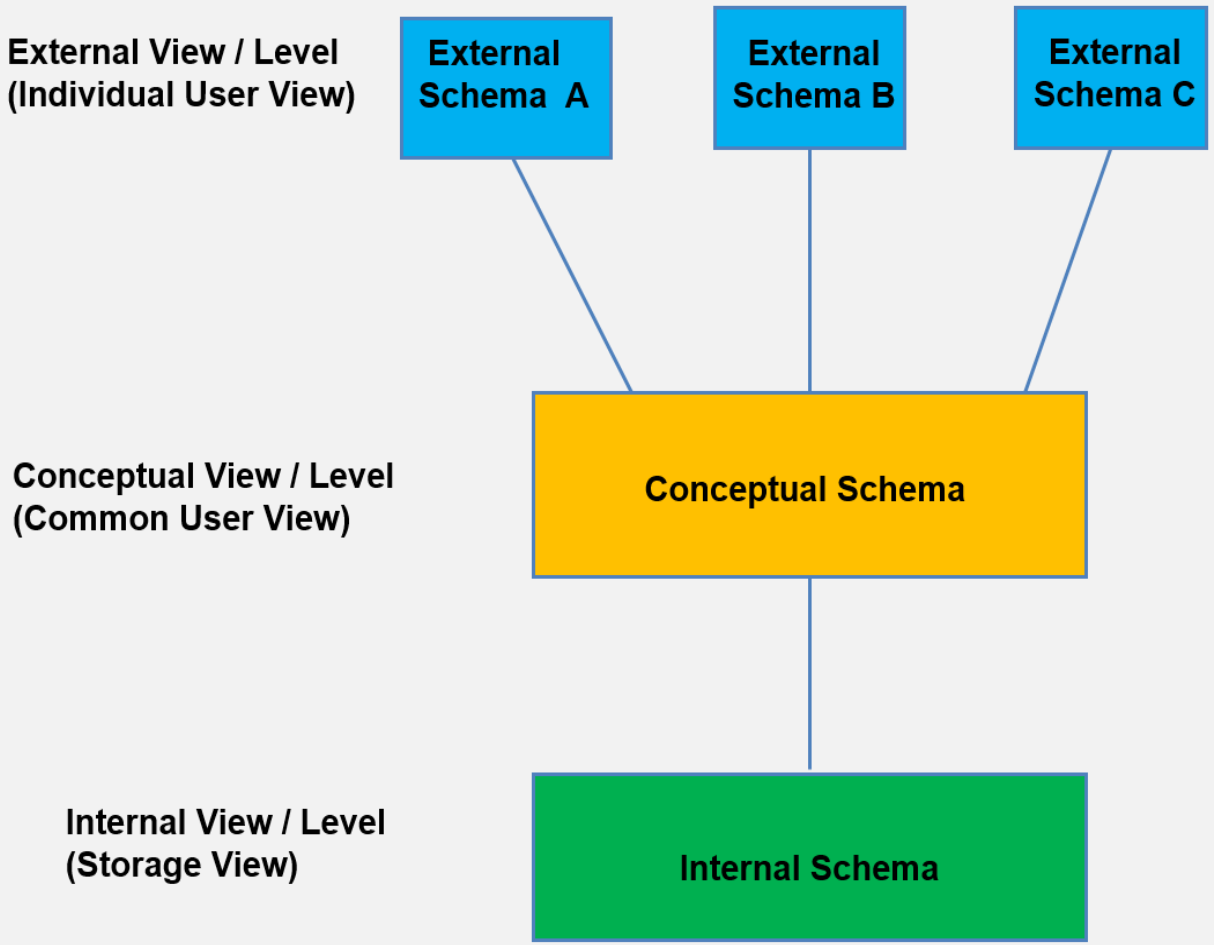

A **Distributed** database is stored on several computers – ranging from personal computers to mainframe systems.
In a distributed system, the computers communicate with one another through various communication media, such as high speed networks or telephone lines.

Distributed databases are:
- ✓ Geographically separated and managed
- ✓ Separately administered
- ✓ Slow in interconnection

Let us understand this type of database with an example.

Consider the bank system. The bank's head office is located at Chicago and the branch offices are at Melbourne and Tokyo. The bank database is distributed across the branch offices. The branch offices are connected through a network.

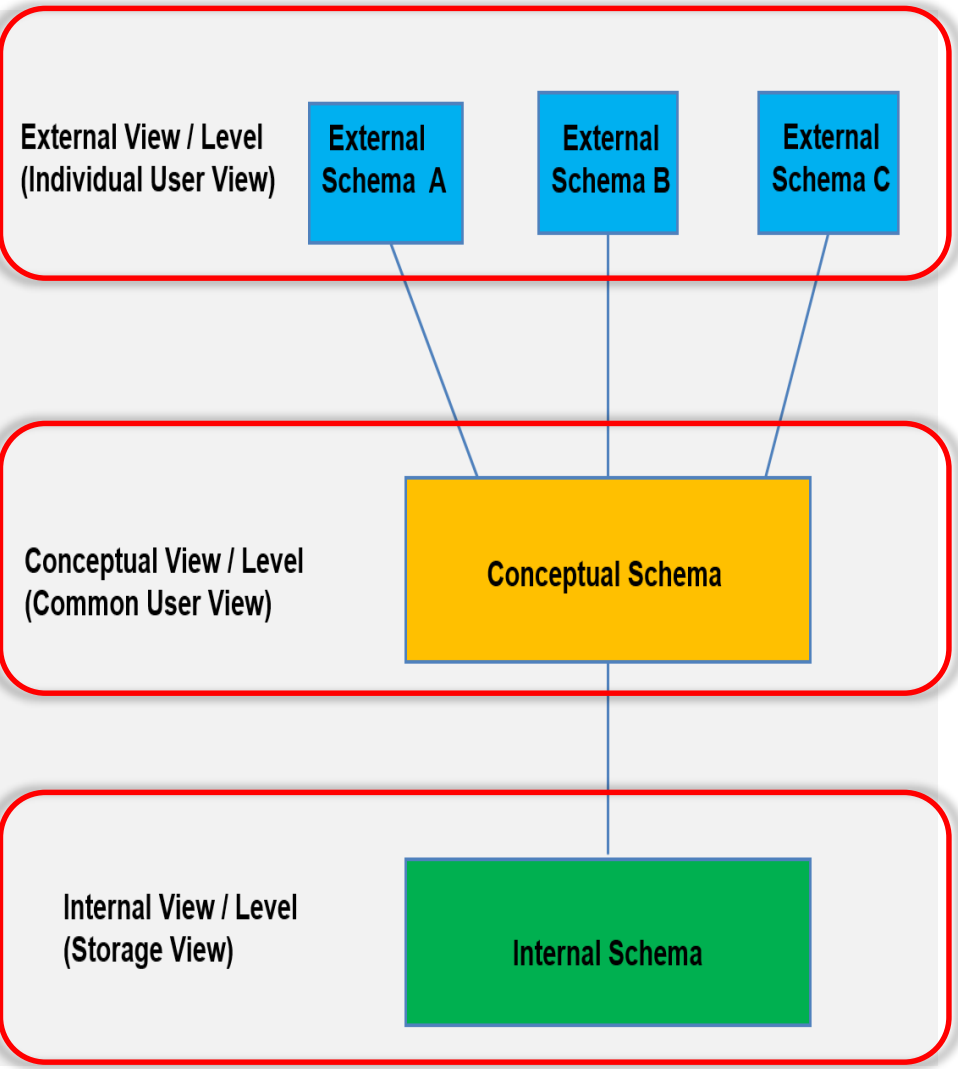
# Three-layer Architecture




A DBMS provides three views of the database data:

- **External level**
- **Conceptual level**
- **Internal level**

# Three-layer Architecture (cont'd)

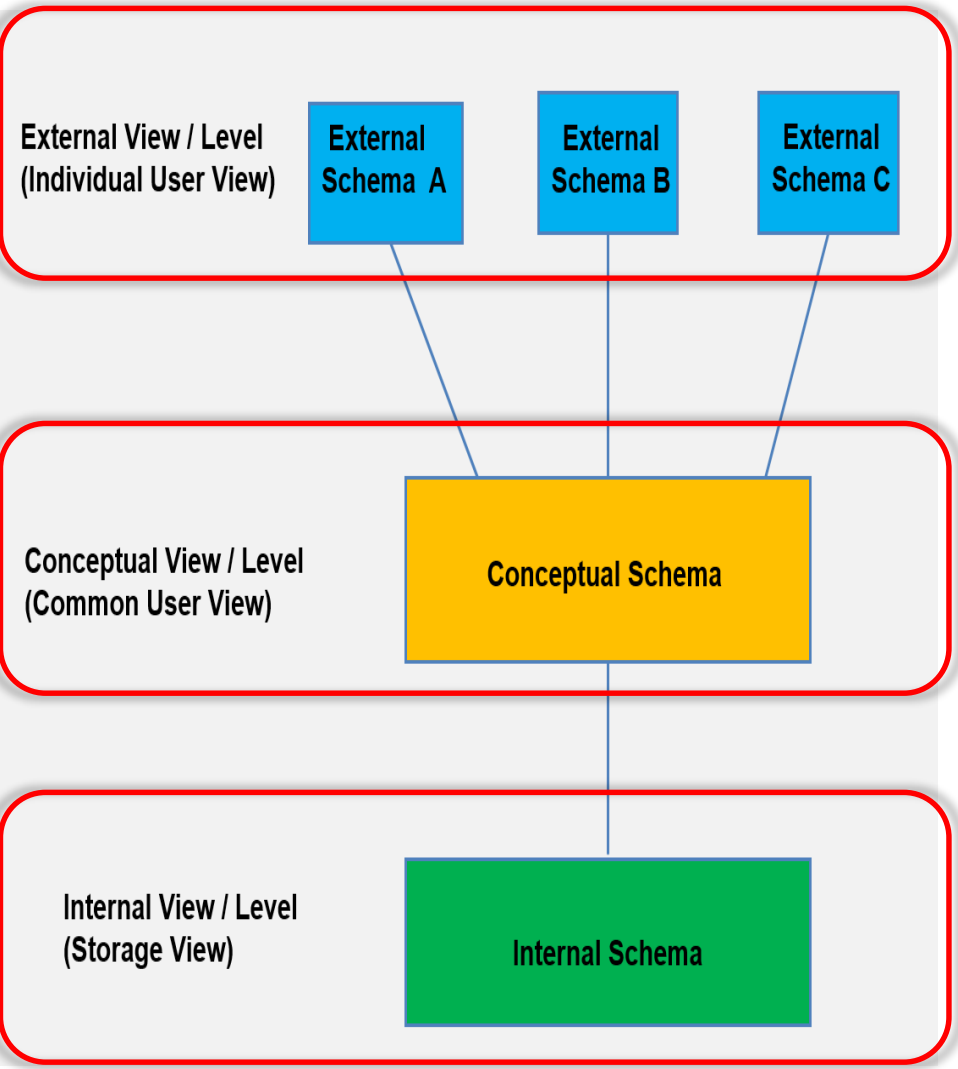



Let us learn about each of these three layers of the database architecture.

*Click each layer to know more.*

The **external** level defines how each group of end-users sees the organization of data in the database. A single database can have any number of views at the external level.

# Three-layer Architecture (cont'd)

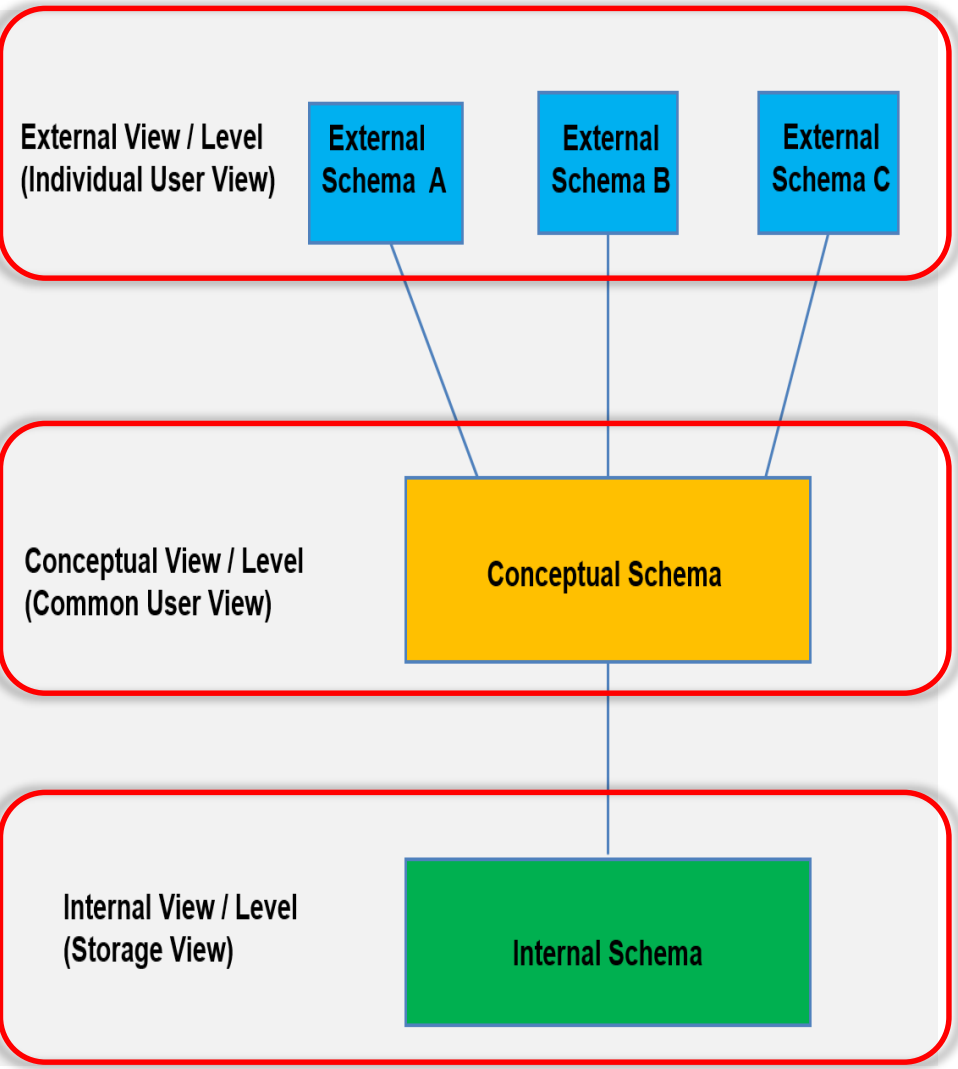



Let us learn about each of these three layers of the database architecture.

*Click each layer to know more.*

The conceptual level unifies the various external views into a coherent global view. It provides the synthesis of all the external views. It is out of the scope of the various database end-users, and is rather of interest to database application developers and database administrators.

# Three-layer Architecture (cont'd)




Let us learn about each of these three layers of the database architecture.

*Click each layer to know more.*

The internal level (or physical level) is the internal organization of data inside a DBMS. It is concerned with cost, performance, scalability and other operational matters. It deals with storage layout of the data, using storage structures such as indexes to enhance performance. Occasionally it stores data of individual views (materialized views), computed from generic data, if performance justification exists for such redundancy. It balances all the external views' performance requirements, possibly conflicting, in an attempt to optimize overall performance across all activities.

# Three-layer Architecture Relates to Concept of Data Independence

While there is typically only one conceptual or logical, and physical or internal view of the data, there can be any number of different **external** views.

The three-level database architecture relates to the **concept of data independence**. The idea is that changes made at a certain level do not affect the view at a higher level. For example, changes in the internal level do not affect application programs written using conceptual level interfaces, which reduces the impact of making physical changes to improve performance.

The **conceptual** view provides a level of indirection between internal and external. On the one hand, it provides a common view of the database, independent of different external view structures. On the other hand, it abstracts away details of how the data is stored or managed.

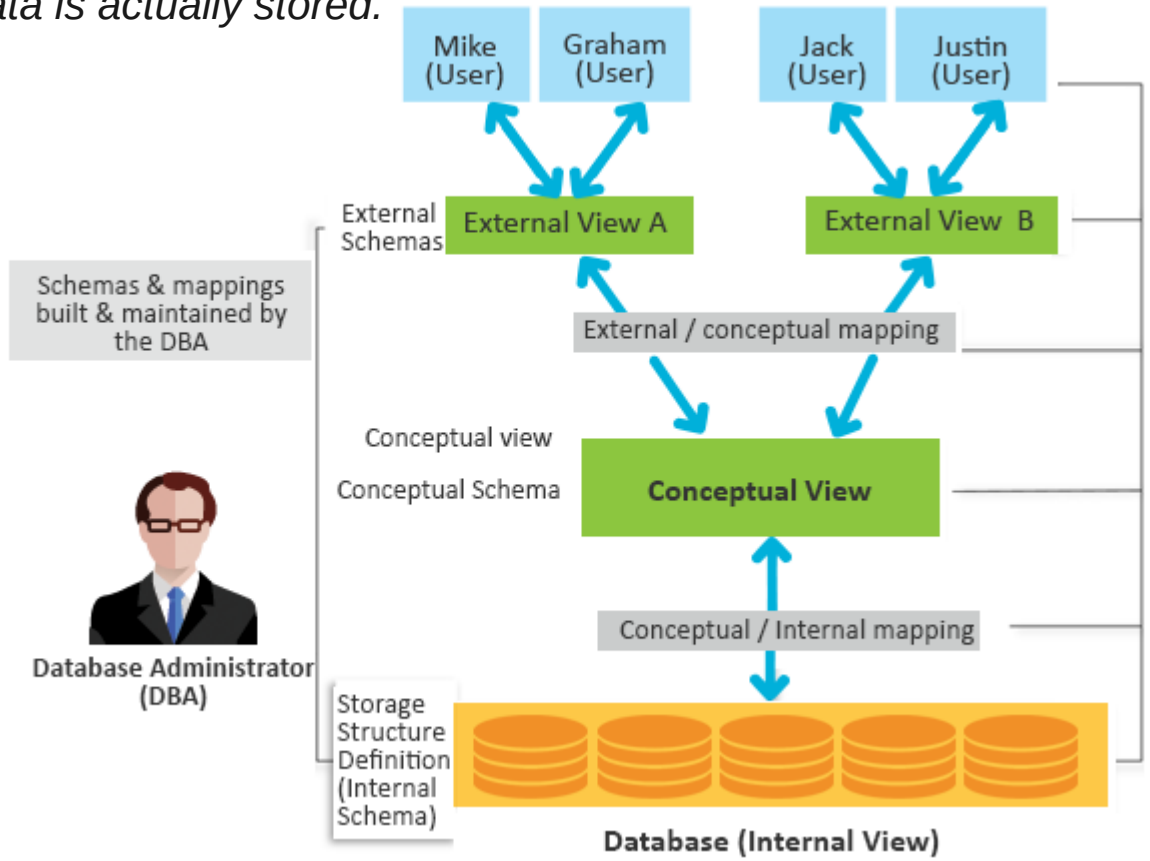
In principle, every level and every external view can be presented by a different data model. In practice, usually a given DBMS uses the same data model for both the external and the conceptual levels.

The **internal** level, which is hidden inside the DBMS and depends on its implementation, requires a different level of detail and uses its own types of data structure types.

# Detailed System Architecture

The diagram depicts the three levels of the DBMS architecture.

- *The External view is how the Customer, Jack views it.*
- *The Conceptual view is how the DBA views it.*
- *The Internal view is how the data is actually stored.*



# 05 Users and Advantages of DBMS

# Users of a DBMS

Let us find out who the users of a DBMS are.

The **Database Administrator** (DBA) is a key person and takes care of most administrative tasks, such as:

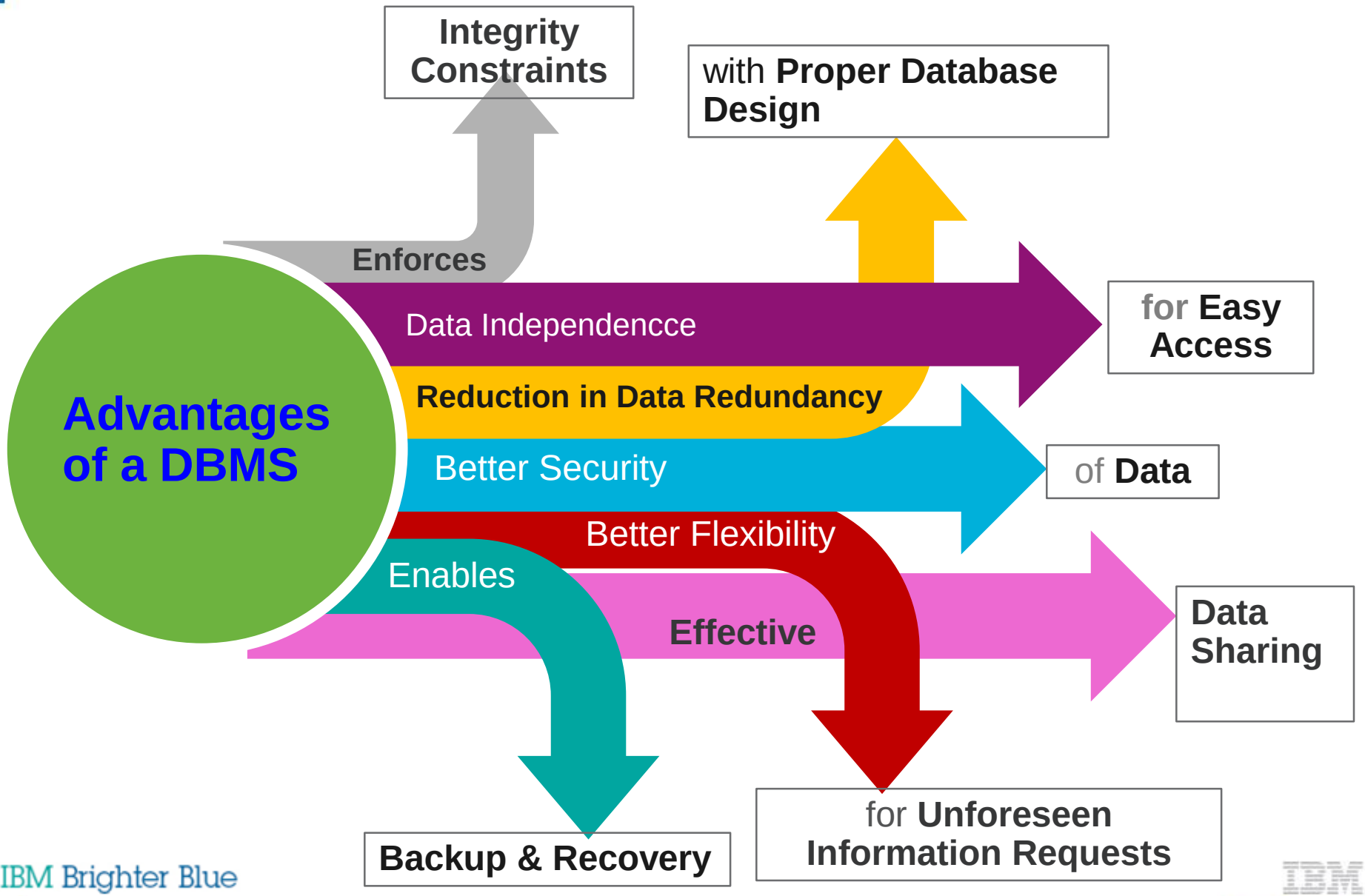
- ✓ Managing information contents
- ✓ Liaison with users
- ✓ Enforcing security and integrity rules
- ✓ Strategizing backup & recovery
- ✓ Monitoring performance

**Database designers** take care of the designing of the database elements.

**Application programmers** make use of the various database elements and write programs to retrieve data from them.

**End users** use the DBMS.

# Advantages of a DBMS



# Spot Quiz

**01**

In the three-layer DBMS architecture, which level defines how each group of end-users sees the organization of data in the database?

| | |
|---|---|
| A | Internal level |
| B | External Level |
| C | Conceptual Level |
| D | Physical level |

# Spot Quiz

01

Which of the following is not true about DBMS?

A. Data Dependency

B. Enforces Integrity Constraints

C. Reduction in Data Redundancy

D. Flexibility

# Module summary

Now that you have completed this module on **Introduction to Database Management System**, you should be able to:

- Explain the fundamentals of data processing
- Identify the various data processing models
- Describe the traditional method of data storage
- Identify problems associated with the traditional approach
- Explain the DBMS
- Identify the different types of databases
- Examine the detailed system architecture of a database
- Identify the users of a DBMS
- List the advantages of a DBMS