

UNIT 4

Database Management System

(4 Hours) — Operating Systems

4.1 Introduction to DBMS

A Database Management System (DBMS) is software that creates, organizes, stores, and manages data in a database, providing controlled access for users and application programs. Rather than storing data in scattered, independent files, a DBMS keeps data in a structured, centralized form and offers tools to insert, retrieve, update, and delete it reliably and efficiently.

DBMS as an Intermediary Between Users and Data

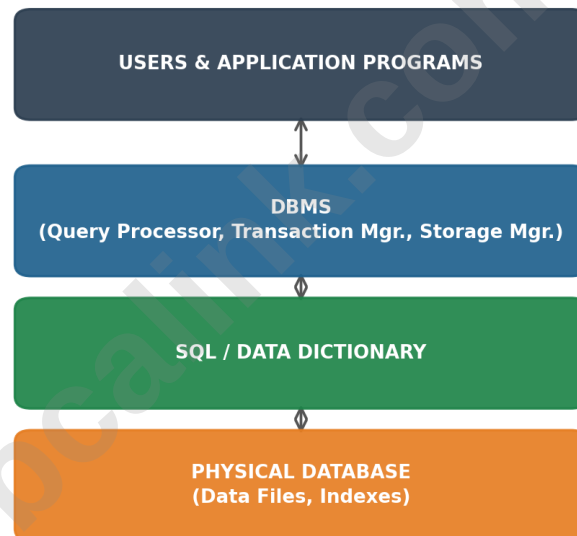


Figure 4.1 — The DBMS acting as an intermediary between users/applications and the physical data.

Compared with traditional file-based systems, a DBMS reduces data redundancy and inconsistency, enforces data integrity through constraints, supports concurrent multi-user access, and provides backup and recovery mechanisms. Popular DBMS products include MySQL, PostgreSQL, Oracle Database, and Microsoft SQL Server.

Example 4.1 — File System vs. DBMS

A college storing student records in separate Excel sheets for each department risks duplicate entries and mismatched updates if a student's address changes in one sheet but not another. A DBMS stores this data once, in a shared, consistent structure, so every department sees the same, up-to-date record.

4.2 Database Models

A database model defines the logical structure of a database — how data is organized and how relationships between data items are represented.

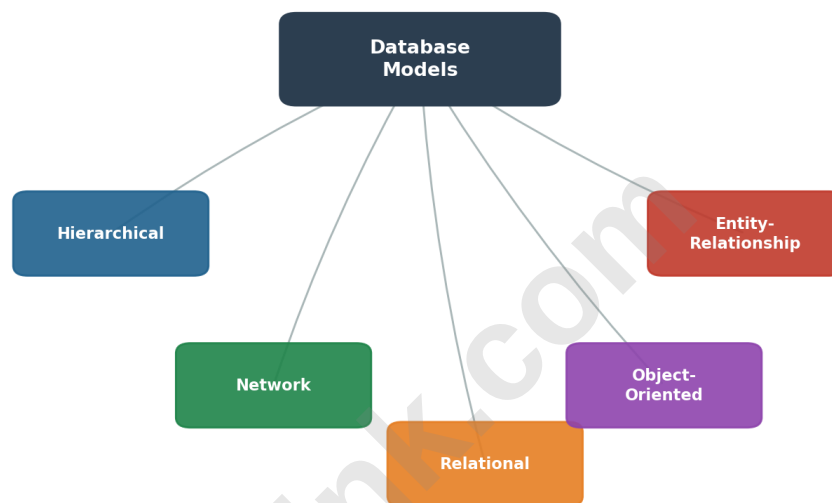


Figure 4.2 — Common database models.

Model	Structure
Hierarchical	Data organized as a tree, with parent-child (one-to-many) links between records.
Network	Similar to hierarchical, but a child can have multiple parents, forming a graph.
Relational	Data organized into tables (relations) of rows and columns, linked via keys.
Object-Oriented	Data represented as objects, combining attributes and behaviour, as in OOP.
Entity-Relationship (ER)	A conceptual model describing entities, attributes, and relationships, used for design.

The relational model, introduced by E. F. Codd, is by far the most widely used today because of its simplicity and strong mathematical foundation in set theory.

4.3 SQL (Structured Query Language)

SQL is the standard language used to create, query, and manage relational databases. Its commands are grouped into four categories:

Category	Purpose	Example Commands
DDL — Data Definition Language	Defines and modifies database structure	CREATE, ALTER, DROP
DML — Data Manipulation Language	Manipulates the data stored in tables	SELECT, INSERT, UPDATE, DELETE
DCL — Data Control Language	Controls access and permissions	GRANT, REVOKE
TCL — Transaction Control Language	Manages transactions	COMMIT, ROLLBACK, SAVEPOINT

Example 4.2 — A Simple SQL Query

```
SELECT name, marks FROM students WHERE marks >= 75 ORDER BY marks DESC;
```

This retrieves the name and marks of every student scoring 75 or above, sorted from highest to lowest.

4.4 Database Design and Data Security

Good database design begins with an ER diagram that captures entities and their relationships, which is then translated into relational tables and refined through normalization — a step-by-step process (1NF, 2NF, 3NF, and beyond) that removes redundant data and avoids update, insert, and delete anomalies.

Data security protects the database from unauthorized access, corruption, and loss. Common measures include:

- Authentication & Authorization — verifying user identity and restricting actions via roles and privileges.
- Encryption — protecting data at rest and in transit from unauthorized reading.

- Access Control — SQL's GRANT/REVOKE commands limit which users can read or modify which data.
- Backup & Recovery — regular backups and transaction logs allow restoration after failure.

Note: Normalization improves data integrity but can increase the number of joins needed in queries — real-world designs often balance normalization with performance.

4.5 Data Warehouse

A data warehouse is a large, centralized repository that consolidates data from multiple operational sources for the purpose of reporting and analysis, rather than day-to-day transaction processing. It is typically described by four characteristics: subject-oriented, integrated, time-variant, and non-volatile.

Characteristic	Meaning
Subject-Oriented	Organized around major subjects (e.g., sales, customers) rather than applications.
Integrated	Combines data from multiple, inconsistent sources into a unified format.
Time-Variant	Stores historical data over long periods to support trend analysis.
Non-Volatile	Once loaded, data is not updated in place — only read and analyzed.

Data warehouses support On-Line Analytical Processing (OLAP), which allows analysts to view data from multiple dimensions (e.g., sales by region, by month, by product).

4.6 Data Mining

Data mining is the process of discovering hidden patterns, correlations, and trends within large volumes of data, often stored in a data warehouse, using statistical and machine-learning techniques.

- Classification — assigning data items into predefined categories (e.g., spam vs. not spam).
- Clustering — grouping similar data items together without predefined labels.
- Association Rule Mining — finding relationships between items (e.g., market-basket analysis).

- Prediction/Regression — estimating future or unknown numeric values from existing data.

Illustration — Market-Basket Analysis

A supermarket applies association rule mining to till receipts and discovers that customers who buy bread frequently also buy butter. The store uses this insight to place the two items near each other, increasing sales.

4.7 Database Administrator (DBA)

A Database Administrator is the person responsible for the overall management, performance, integrity, and security of a database system.

- Installing, configuring, and upgrading DBMS software.
- Designing schemas and managing storage structures.
- Managing user accounts, roles, and access permissions.
- Monitoring performance and tuning queries.
- Planning and testing backup and disaster-recovery procedures.

Key Takeaways — Unit 4

- ✓ A DBMS centralizes data storage and access, reducing redundancy and inconsistency.
- ✓ The relational model, using tables linked by keys, is the most widely used database model.
- ✓ SQL commands fall into four categories: DDL, DML, DCL, and TCL.
- ✓ Good design combines ER modeling and normalization with practical security measures.
- ✓ Data warehouses support historical analysis (OLAP); data mining discovers hidden patterns within that data.
- ✓ The DBA keeps the database available, secure, and performing well.

— End of Unit 4 —