***Day 13 – 20 October 2023***

Database:

A database is a structured collection of data, typically stored electronically in a computer system. A database management system (DBMS) is a software application that interacts with end users, applications, and the database itself to capture and analyze the data.

Relational databases: Relational databases store data in tables, but not all databases do. Non-relational databases can store data in other formats, such as documents, graphs, and key-value pairs.

MySQL, Oracle Database, PostgreSQL, SQL Lite

Non-relational databases: Non-relational databases do not use tables to store data. Instead, they use a variety of other data models, such as document databases, graph databases, and key-value stores.

MangoDB, Cassandra, Redis

Database Management System (DBMS): A DBMS is a software application that is used to create, manage, and access databases. It is not limited to relational databases. Non-relational databases also use DBMSs.

CRUD operations: CRUD stands for Create, Read, Update, and Delete. These are the four basic operations that can be performed on data in a database.

Schema: A database schema is a definition of the structure of a database. It specifies the tables, columns, and relationships between them.

Major data types

* String: Text data, such as names, addresses, and product descriptions.
* Numeric: Numerical data, such as integers, floats, and decimals.
* Date and time: Data representing dates and times.

Example: A customer table might have a string column for customer name, a numeric column for customer ID, and a date and time column for customer registration date.

Primary key and foreign key

Primary key: A unique identifier for each record in a table. Primary keys cannot be null.

Foreign key: A column in one table that references the primary key of another table. Foreign keys are used to create relationships between tables.

Example: A customer table might have a primary key column for customer ID. An orders table might have a foreign key column that references the customer ID column in the customer table. This would allow us to track which orders belong to which customers.

Unique key: A column in a table that must contain unique values, but can be null.

Example: A customer table might have a unique key column for customer email addresses. This would prevent customers from having duplicate email addresses.

Types of relationships

1. One-to-one: One record in one table can be associated with only one record in another table.
2. One-to-many: One record in one table can be associated with multiple records in another table.
3. Many-to-many: Multiple records in one table can be associated with multiple records in another table.

Example: A customer can have many orders (one-to-many relationship). A product can be ordered by many customers (one-to-many relationship). A customer can order many products, and a product can be ordered by many customers (many-to-many relationship).

A junction table is a table that is used to create a many-to-many relationship between two tables. It contains the primary keys of both tables, and each record in the junction table represents a relationship between two records in the other two tables.

For example, if you have a database of customers and products, and you want to track which customers have ordered which products, you could create a junction table called orders. The orders table would have two columns: customer\_id and product\_id. Each record in the orders table would represent an order that a customer has placed for a product.

Junction tables are useful because they prevent duplication of data. If you did not use a junction table, you would have to store the customer ID and product ID for each order in both the customers table and the products table. This would be inefficient and could lead to data errors.

Differences between OLTP and OLAP

OLTP - Databases OLAP - Data warehouses

OLTP (Online Transaction Processing) databases are designed for handling large volumes of transactions in real time. OLAP (Online Analytical Processing) databases are designed for analytical workloads, such as running reports and performing data mining.

OLTP databases typically have the following characteristics:

* High concurrency: OLTP databases must be able to handle many users accessing the database at the same time.
* High throughput: OLTP databases must be able to process many transactions per second.
* Low latency: OLTP transactions must be processed quickly, with minimal delay.

OLAP databases typically have the following characteristics:

* High performance: OLAP databases must be able to run complex queries on large datasets quickly.
* Scalability: OLAP databases must be able to scale to handle large datasets.
* Data security: OLAP databases must protect sensitive data from unauthorized access.