Day 30 – SQL Theory Introduction

---

1. What is Data?

Data refers to raw facts, figures, or information that can be stored, processed, and analyzed. It can exist in many forms:

- **Structured**: Names, dates of birth, account balances, tables, Excel sheets.
- **Unstructured**: Images, videos, audio files, PDF documents, XML files, sensor logs, social media posts.

---

2. What is a Database?

A **database** is an organized collection of data, stored electronically, that allows easy **storage, retrieval, modification, and analysis**. Think of it as a **container** that holds your information.

- Example: Social media platforms store your posts, images, and messages in their databases.

---

3. Types of Databases

Databases can be classified into:

A. Structured Databases

- Store data in tables (rows & columns).
- Example: **Relational Databases (RDBMS)** like MySQL, PostgreSQL, Oracle.
- Used when the data follows a fixed schema.

B. Unstructured Databases

- Handle data without a fixed schema (images, videos, text).
- Example: **NoSQL Databases** like MongoDB, Cassandra, HBase.

C. Vector Databases

- Store data as vectors for AI and Machine Learning applications.
- Used in Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG).
- Examples: Pinecone, ChromaDB, Milvus, FAISS.

---

4. SQL vs NoSQL vs Vector Databases

Feature	SQL (Relational DB)	NoSQL (Non-relational DB)	Vector DB
Data Format	Tables (rows & columns)	Key-value, documents, graphs	Vectors (numeric embeddings)
Examples	MySQL, PostgreSQL	MongoDB, Cassandra	Pinecone, Milvus
Use Case	Banking, Insurance, Healthcare	Streaming, Social Media	AI, Chatbots, Search
Structure	Fixed schema	Flexible schema	Embedding-based

---

5. Types of NoSQL Databases

- **Key-Value Stores** → Redis, DynamoDB
- **Document Stores** → MongoDB, CouchDB
- **Graph Databases** → Neo4j, Amazon Neptune
- **Wide Column Stores** → Cassandra, HBase

---

6. Database Deployment

- **Server**: A collection of databases stored in on-premises servers.
- **Cloud**: Databases stored in remote **data centers** (e.g., AWS, Azure, GCP).
- **Application Server vs Production Server**:
  - Development happens on application server.
  - Production server runs the live project.

---

7. What is DBMS?

Database Management System (DBMS) is software that interacts with databases to store, retrieve, update, and manage data.

- Examples: MySQL, PostgreSQL, Oracle, SQL Server, MongoDB, Redis.
- **CRUD Operations** → Create, Read, Update, Delete.

---

8. Evolution of Databases

1. **Flat File DB** – Data stored in plain text or spreadsheets.
2. **Hierarchical DB** – Tree-like parent-child structure.
3. **Network DB** – Multiple relationships between records.
4. **Relational DB (SQL)** – Data stored in related tables.
5. **NoSQL DB** – Flexible schema for unstructured data.
6. **Vector DB** – Data stored as numerical embeddings for AI.

---

9. Relational Database Concepts

- **Table**: Collection of rows & columns.
- **Row (Record)**: Single entry in a table.
- **Column (Field)**: Attribute with a specific data type.
- **Primary Key**: Unique identifier for a record.
- **Foreign Key**: Field linking two tables.

---

10. Why SQL is Important for Data Professionals?

- Enables **data analysis** for structured datasets.
- Essential for **Data Analysts, Data Engineers, and ML Engineers**.
- Works well with Python libraries like Pandas, NumPy, and visualization tools like Matplotlib/Seaborn.

---

11. Real-World Applications of SQL

1. **Banking & Finance**
    - Storing customer profiles, account balances, and transaction history.
    - *Example:* Retrieving the last 10 transactions of a user for fraud detection.
2. **Healthcare**
    - Managing patient records, lab reports, and prescriptions.
    - *Example:* Fetching all patients with a particular diagnosis in the last 6 months.
3. **E-Commerce**
    - Tracking products, customer orders, and payments.
    - *Example:* Generating monthly sales reports by category.
4. **Social Media**
    - Managing user profiles, posts, likes, and comments.
    - *Example:* Displaying top 10 trending hashtags in the last 24 hours.
5. **Education**
    - Maintaining student records, courses, and grades.
    - *Example:* Listing all students who scored above 90% in Mathematics.