I. Khái niệm

Hệ quản trị cơ sở dữ liệu DBMS: là một phần mềm giúp người dùng tạo, quản lý và thao tác dữ liệu một cách hiệu quả. Nó đóng vai trò trung gian giữa người dùng và cơ sở dữ liệu ( là một tập hợp có tổ chức của dữ liệu được lưu trữ trên hệ thống máy tính) đảm bảo dữ liệu được lưu trữ 1 cách có tổ chức, an toàn và truy xuất nhanh chóng.

II. Cách thiết kế

1. Xác định yêu cầu mục tiêu

* Loại dữ liệu hệ thống sẽ quản lý (cấu trúc, phi cấu trúc)
* Dữ liệu cấu trúc: là dữ liệu được tổ chức theo một định cố định có dạng hàng, cột. Vd: danh sách khách hàng, Hệ thống quản lý kho
* Dữ liệu phi cấu trúc: dữ liệu không tuân theo một mô hình cụ thể hoặc không dễ dàng tổ chức trong các bảng quan hệ. Vd: Hình ảnh, video…
* Hiệu suất mong muốn (tốc độ truy vấn, xử lý giao dịch).
* Khả năng mở rộng và bảo mật của hệ thống.

2. Chọn mô hình dữ liệu

* Quan hệ (Relational Database - RDBMS): Sử dụng bảng (MySQL, PostgreSQL) 🡪 xử lý dữ liệu có quan hệ chặt chẽ
* NoSQL: Key-Value, Document, Column-Family, Graph (MongoDB, Cassandra)
  + hiệu suất cao
* NewSQL: Kết hợp ưu điểm của SQL và NoSQL (Google Spanner, CockroachDB).

3. Thiết kế kiến trúc hệ thống

Các thành phần của 1 DBMS:

* Storage Manager (Quản lý lưu trữ): Quản lý tệp, chỉ mục, caching, nén dữ liệu.
* Query Processor (Bộ xử lý truy vấn): Dịch, tối ưu, và thực thi truy vấn.
* Transaction Manager (Quản lý giao dịch): Đảm bảo ACID (Atomicity, Consistency, Isolation, Durability).
* Concurrency Control (Kiểm soát đồng thời): Hỗ trợ nhiều người dùng truy cập cùng lúc mà không gây lỗi dữ liệu.
* Security & Access Control (Bảo mật và kiểm soát truy cập): Cung cấp cơ chế xác thực, phân quyền.

4. Xây dựng và tối ưu bộ xử lý truy vấn

* Viết trình biên dịch cho SQL hoặc ngôn ngữ truy vấn riêng.
* Thiết kế bộ tối ưu hóa truy vấn để tăng hiệu suất.
* Sử dụng kỹ thuật lập chỉ mục (Indexing) và caching dữ liệu.

5. Quản lý giao dịch và kiểm soát đồng thời

* Triển khai các cơ chế như Two-Phase Locking (2PL) hoặc MVCC (Multi-Version Concurrency Control).
* Xử lý Deadlock detection & prevention.

6. Hỗ trợ sao lưu và phục hồi dữ liệu

* Thiết kế cơ chế Backup & Restore để tránh mất dữ liệu.
* Sử dụng Write-Ahead Logging (WAL) để đảm bảo phục hồi dữ liệu khi có sự cố.

7. Kiểm tra và triển khai

* Kiểm thử khả năng mở rộng (Scalability).
* Kiểm thử bảo mật (Security Testing).
* Cải tiến hiệu suất dựa trên phản hồi thực tế.