



TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
Viện Toán Ứng dụng và Tin học



Phân tích thiết kế hệ thống

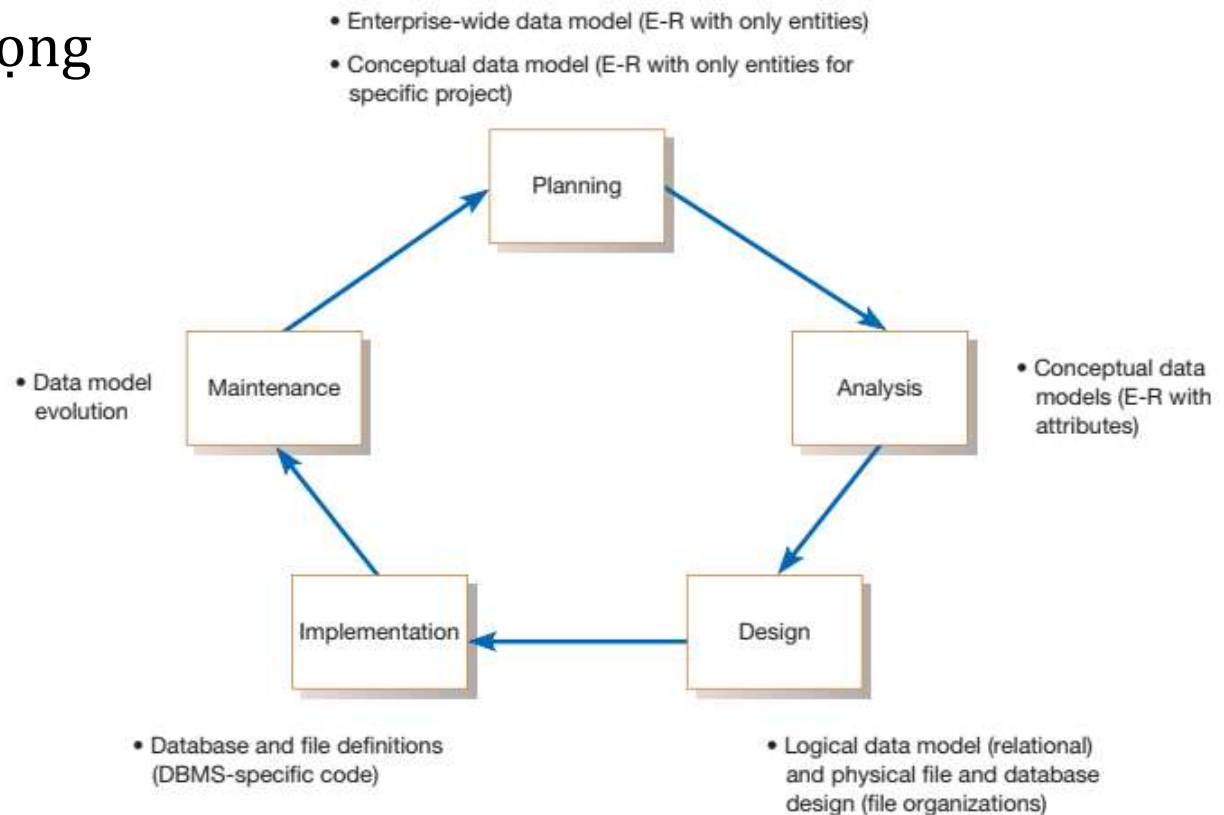
\$7 Thiết kế hệ thống về dữ liệu
- Mô hình dữ liệu quan hệ



Nội dung

- Chuyển từ mô hình E/A sang mô hình quan hệ để thực hiện việc chuẩn hóa
- Bổ sung các ràng buộc toàn vẹn
- Thiết kế các phi chuẩn
- Thiết kế các chỉ mục
- Thiết kế các thủ tục lưu sẵn, trigger

- Dữ liệu: góc nhìn quan trọng
- Mô hình hóa dữ liệu:
 - sơ đồ E-R
 - sơ đồ Class (OOAD, UML)
- Thiết kế: 2 bước
 - Logical design
 - Physical design



Mô hình quan hệ

■ Ưu điểm:

- Đơn giản: biểu diễn dưới 1 dạng duy nhất là quan hệ
- Chặt chẽ: Các khái niệm được hình thức hóa cao, cho phép áp dụng các công cụ toán học
- Trừu tượng hóa cao: ở mức quan niệm, độc lập với mức vật lý
- Cung cấp các ngôn ngữ truy cập mức cao

Quan hệ, thuộc tính, lược đồ quan hệ

- Quan hệ

$$R \subseteq D_1 \times D_2 \times \cdots \times D_n$$

- Thuộc tính

- Tên gọi của mỗi cột

- Lược đồ quan hệ

- Khuôn mẫu cho một họ các quan hệ cùng ngữ nghĩa
- Gồm 2 yếu tố:
 - Cấu trúc $R(A_1, A_2, \dots, A_n)$
 - Tập hợp các ràng buộc toàn vẹn, tức là các ràng buộc mà mọi quan hệ trong lược đồ đều phải thỏa mãn

Chiếu, chọn, kết nối

- Phép chiếu: π
- Phép chọn: σ
- Phép kết nối, kết nối tự nhiên: \bowtie
- Phân rã một quan hệ:
 - Tách một quan hệ R thành 2 quan hệ S và T nhỏ hơn mà không mất mát thông tin:
 - S, T là chiếu của R
 - Kết nối của S và T là R

Phụ thuộc hàm

- **Phụ thuộc hàm:** $A \rightarrow B$
- Nói phụ thuộc hàm cho một lược đồ quan hệ hiểu theo nghĩa là phụ thuộc hàm này phải đúng cho mọi quan hệ trong lược đồ quan hệ đó.
- Phụ thuộc hàm **sơ đẳng**: $A \rightarrow B$ là sơ đẳng nếu không tồn tại $A' \subset A$ mà $A' \rightarrow B$
- Phụ thuộc hàm **trực tiếp**: $A \rightarrow B$ trong R là phụ thuộc hàm trực tiếp nếu không tồn tại tập C trong R khác với A, B mà $A \rightarrow C$ và $C \rightarrow B$
- A là **khóa** của R(A,B) nếu $A \rightarrow B$ là phụ thuộc hàm sơ đẳng trong R

Tiên đề Armstrong

- i. Tính phản xạ (reflexivity): Nếu $B \subseteq A$ thì $A \rightarrow B$
- ii. Tính tăng cường (augmentation): Nếu $A \rightarrow B$ thì $A, C \rightarrow B$ với C bất kỳ
- iii. Tính bắc cầu (transitivity): Nếu $A \rightarrow B$ và $B \rightarrow C$ thì $A \rightarrow C$

Một số dẫn xuất:

- iv. Tính phân rã (decomposition): Nếu $A \rightarrow B$ và $D \subseteq B$ thì $A \rightarrow D$
- v. Tính gộp (union): Nếu $A \rightarrow B$ và $A \rightarrow C$ thì $A \rightarrow B, C$
- vi. Tính giả bắc cầu (pseudotransitivity): Nếu $A \rightarrow B$ và $B, D \rightarrow C$ thì $A, D \rightarrow C$

Một số dạng chuẩn

- *Tại sao cần chuẩn hóa ??*
- Chuẩn 1 (**1NF**): miền thuộc tính là các miền đơn
- Chuẩn 2 (**2NF**): là 1NF và phụ thuộc hàm sơ đẳng vào khóa
- Chuẩn 3 (**3NF**): là 2NF và phụ thuộc hàm trực tiếp vào khóa
- Dị thường dữ liệu:
 - Khi cập nhật
 - Khi xóa bỏ
 - Khi bổ sung

Các ràng buộc toàn vẹn

- Ràng buộc toàn vẹn (integrity constraint) là một đặc tả mà một cơ sở dữ liệu phải thỏa mãn để giữ được tính đúng đắn của nó.
- Các ràng buộc toàn vẹn gắn với một mô hình quan hệ thường chia thành 2 loại: các ràng buộc toàn vẹn tĩnh và các ràng buộc toàn vẹn động

Các ràng buộc toàn vẹn tĩnh

- Ràng buộc trên thuộc tính
 - Tồn tại
 - Miền giá trị
 - Giá trị mặc định
- Các ràng buộc trên bộ
- Các ràng buộc trên quan hệ
 - Khóa
 - Phụ thuộc hàm
 - Dạng chuẩn
 - Bản số
- Các ràng buộc trên nhiều quan hệ
 - Ràng buộc về khóa ngoại
 - Ràng buộc toàn vẹn tham chiếu

Các điều kiện mà CSDL phải thỏa mãn vào bất kỳ thời điểm nào

CATEGORY Table

| CATEGORY_ID | NAME | MODIFICATION_DATE | PARENT_ID |
|-------------|---------------------------|-------------------|-----------|
| 1 | Junkyard Cars | 21-jan-2002 | |
| 2 | Sixties Home Décor | 22-jan-2003 | |
| 3 | Obsolete Computers | 23-jan-2002 | |
| 4 | Outdated Technology Books | 22-jan-2004 | |

Row

Column

Các ràng buộc toàn vẹn động

- *Là các đặc tả một thay đổi trạng thái đúng đắn của CSDL*
- Các ràng buộc định nghĩa bởi các tiên đề, hậu đề
 - Ràng buộc loại tiên đề
 - Ràng buộc loại hậu đề
 - Ràng buộc tiên và hậu đề
- Các ràng buộc định nghĩa bởi luật ứng xử (*chỉ ra các hành động phải thực hiện khi có một thay đổi nào đó trong CSDL*)

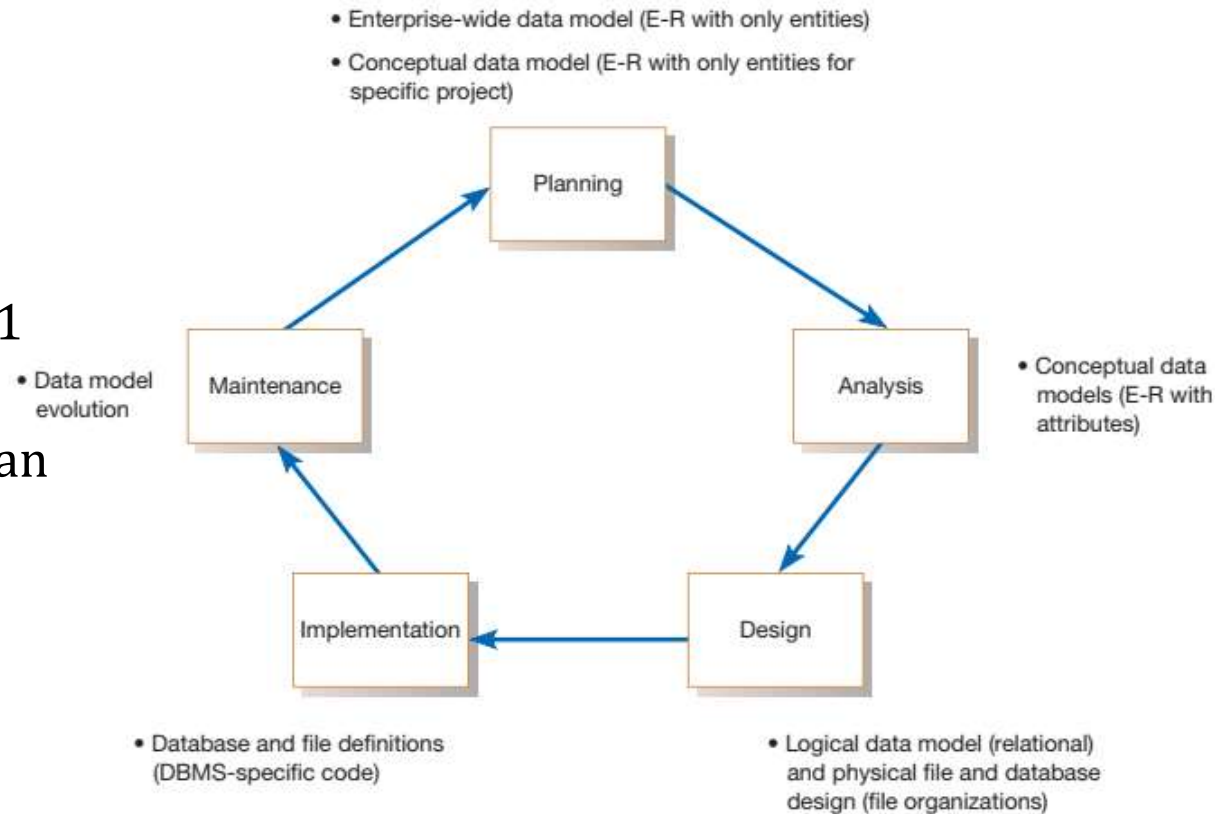
Biểu diễn các ràng buộc

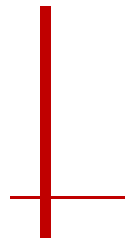
- Dạng bảng

| TT | Tên trường | Diễn giải | PK | FK | NULL | Default | Ghi chú |
|----|-------------|---------------------------------|----|----|------|---------|---------|
| 1 | <u>MaSV</u> | Mã sinh viên | X | | | | |
| 2 | TenSV | Tên sinh viên | | | | | |
| 3 | NgaySinh | Ngày sinh | | | | | |
| 4 | DiaChi | Địa chỉ | | | | | |
| 5 | DienThoai | Điện thoại | | | | | |
| 6 | Active | 1; nếu dừng học; 0 nếu đang học | | | | | |
| 7 | GhiChu | Ghi chú | | | | | |

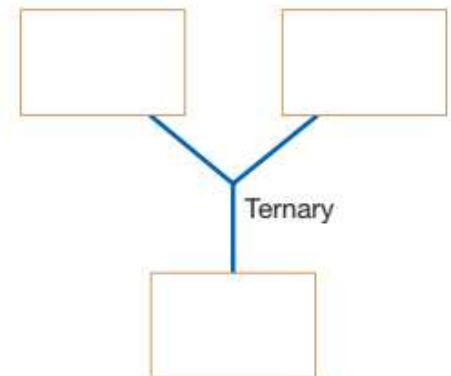
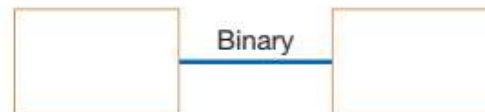
■ Các tác vụ cơ bản

- logical model cho từng form/report
- tổng hợp dl chuẩn hóa thành 1 mô hình
- xd các yêu cầu dữ liệu liên quan lĩnh vực nghiệp vụ
- so sánh và hoàn thiện db logical model





Relationship Degrees



Relationship Cardinality



Mandatory One



Mandatory Many



Optional One



Optional Many

vd

- Xd logical model cho từng form/report

- hóa đơn
- báo cáo hàng tồn

- Tổng hợp

| <u>PRODUCT ID</u> | <u>BACKLOG QUANTITY</u> |
|-------------------|-----------------------------|
| B381 | 0 |
| B975 | 0 |
| B985 | 6 |
| E125 | 30 |
| ⋮ | |
| M128 | 2 |
| ⋮ | |

PRODUCT(Product_ID)
LINE ITEM(Product_ID,Order_Number,Order_Quantity)
ORDER(Order_Number,Order_Date)
SHIPMENT(Product_ID,Invoice_Number,Ship_Quantity)
INVOICE(Invoice_Number,Invoice_Date,Order_Number)

HIGHEST-VOLUME CUSTOMER

ENTER PRODUCT ID.: M128
START DATE: 11/01/2017
END DATE: 12/31/2017

CUSTOMER ID.: 1256
NAME: Commonwealth Builder
VOLUME: 30

CUSTOMER(Customer_ID,Name)
ORDER(Order_Number,Customer_ID,Order_Date)
PRODUCT(Product_ID)
LINE ITEM(Order_Number,Product_ID,Order_Quantity)

vd

- Xd logical model cho từng form/report


- hóa đơn
- báo cáo hàng tồn

- Tổng hợp

- Tổng hợp quan hệ
- XD mô hình khái niệm

CUSTOMER(Customer_ID,Name)
ORDER(Order_Number,Customer_ID,Order_Date)
PRODUCT(Product_ID)
LINE ITEM(Order_Number,Product_ID,Order_Quantity)

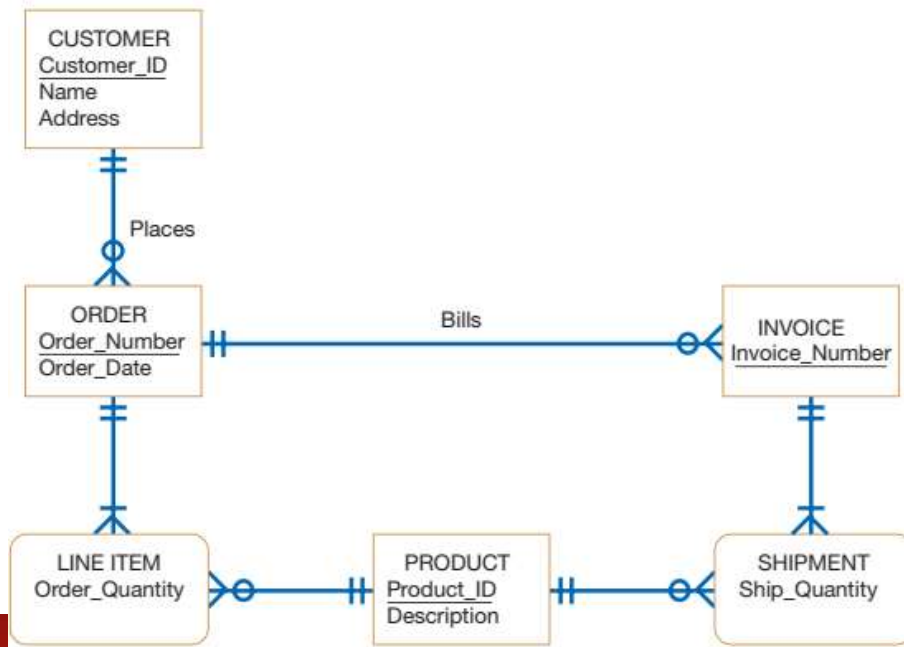
PRODUCT(Product_ID)
LINE ITEM(Product_ID,Order_Number,Order_Quantity)
ORDER(Order_Number,Order_Date)
SHIPMENT(Product_ID,Invoice_Number,Ship_Quantity)
INVOICE(Invoice_Number,Invoice_Date,Order_Number)



CUSTOMER(Customer_ID,Name)
PRODUCT(Product_ID)
INVOICE(Invoice_Number,Invoice_Date,Order_Number)
ORDER(Order_Number,Customer_ID,Order_Date)
LINE ITEM(Order_Number,Product_ID,Order_Quantity)
SHIPMENT(Product_ID,Invoice_Number,Ship_Quantity)

vd

- Tổng hợp
 - Tổng hợp quan hệ
 - XD mô hình khái niệm



CUSTOMER(Customer_ID,Name)
PRODUCT(Product_ID)
INVOICE(Invoice_Number,Invoice_Date,Order_Number)
ORDER(Order_Number,Customer_ID,Order_Date)
LINE ITEM(Order_Number,Product_ID,Order_Quantity)
SHIPMENT(Product_ID,Invoice_Number,Ship_Quantity)



CUSTOMER(Customer_ID,Name,Address)
PRODUCT(Product_ID,Description)
ORDER(Order_Number,Customer_ID,Order_Date)
LINE ITEM(Order_Number,Product_ID,Order_Quantity)
INVOICE(Invoice_Number,Order_Number)
SHIPMENT(Invoice_Number,Product_ID,Ship_Quantity)



Thiết kế vật lý Physical design

Thiết kế trường dữ liệu

- Kiểu dữ liệu
 - Kích thước lưu trữ => min
 - Biểu diễn tất cả domain value
 - Tăng tính toàn vẹn
 - Hỗ trợ các tác vụ xl dữ liệu của field
- Calculated field
- Coding & compression
- Kiểm soát tính toàn vẹn
 - default value
 - range
 - ref integrity (FK)
 - null value

TABLE 9-2 Commonly Used Data Types in Oracle 10g

| Data Type | Description |
|-----------|---|
| VARCHAR2 | Variable-length character data with a maximum length of 4000 characters; you must enter a maximum field length (e.g., VARCHAR2(30) for a maximum length of 30 characters). A value less than 30 characters consumes only the required space. |
| CHAR | Fixed-length character data with a maximum length of 255 characters; length is 1 character (e.g., CHAR(5) for a field with a fixed length of 5 characters, capable of holding a value from 0 to 5 characters long). |
| LONG | Capable of storing up to two gigabytes of one variable-length character field (e.g., to hold a medical instruction or a customer comment). |
| NUMBER | Positive and negative numbers in the range 10^{-130} to 10^{126} ; can specify precision (total number of digits to the left and right of the decimal point) and the scale (the number of digits to the right of the decimal point). NUMBER(5) specifies an integer field with a maximum of 5 digits; NUMBER(5, 2) specifies a field with no more than five digits and two digits to the right of the decimal point). |
| DATE | Any date from January 1, 4712 BC to December 31, 4712 AD; displays the century, year, month, day, hour, minute, and second. |
| BLOB | Binary large object, capable of storing up to four gigabytes of binary data (e.g., a photograph or sound clip). |

Thiết kế bảng vật lý

- Không phải luôn tương ứng 1-1 giữa logical (E-R) và physical (DB table)
- Các yếu tố vật lý:
 - Lưu trữ
 - Truy cập
- Partitioned table

lưu trữ liên tục/cạnh nhau

Thiết kế phi chuẩn

- Thiết kế là sự thỏa hiệp giữa logic thuần và điều kiện vật lý
- VD thỏa hiệp giữa tối ưu hóa kích thước dữ liệu lưu trữ và tốc độ xử lý
- Nguyên tắc đảm bảo tính sẵn sàng: dư thừa => phi chuẩn

CUSTOMER

| <u>Customer_ID</u> | Name | Region | Annual_Sales |
|--------------------|--------|----------|--------------|
| 1256 | Rogers | Atlantic | 10,000 |
| 1323 | Temple | Pacific | 20,000 |
| 1455 | Gates | South | 15,000 |
| 1626 | Hope | Pacific | 22,000 |
| 2433 | Bates | South | 14,000 |
| 2566 | Bailey | Atlantic | 12,000 |

A_CUSTOMER

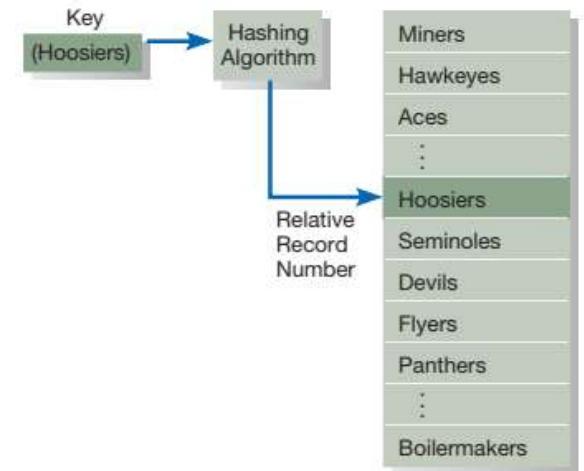
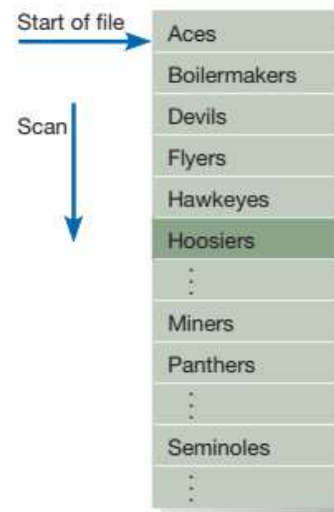
| <u>Customer_ID</u> | Name | Region | Annual_Sales |
|--------------------|--------|----------|--------------|
| 1256 | Rogers | Atlantic | 10,000 |
| 2566 | Bailey | Atlantic | 12,000 |

P_CUSTOMER

| <u>Customer_ID</u> | Name | Region | Annual_Sales |
|--------------------|--------|---------|--------------|
| 1323 | Temple | Pacific | 20,000 |
| 1626 | Hope | Pacific | 22,000 |

Thiết kế các chỉ mục

- *Chỉ mục giúp tăng tốc độ tìm kiếm*
- Hiểu biết về chỉ mục và các truy vấn giúp thiết kế các chỉ mục tốt
- Nguyên tắc:
tìm kiếm trên thuộc tính gì thì lập chỉ mục trên thuộc tính đó



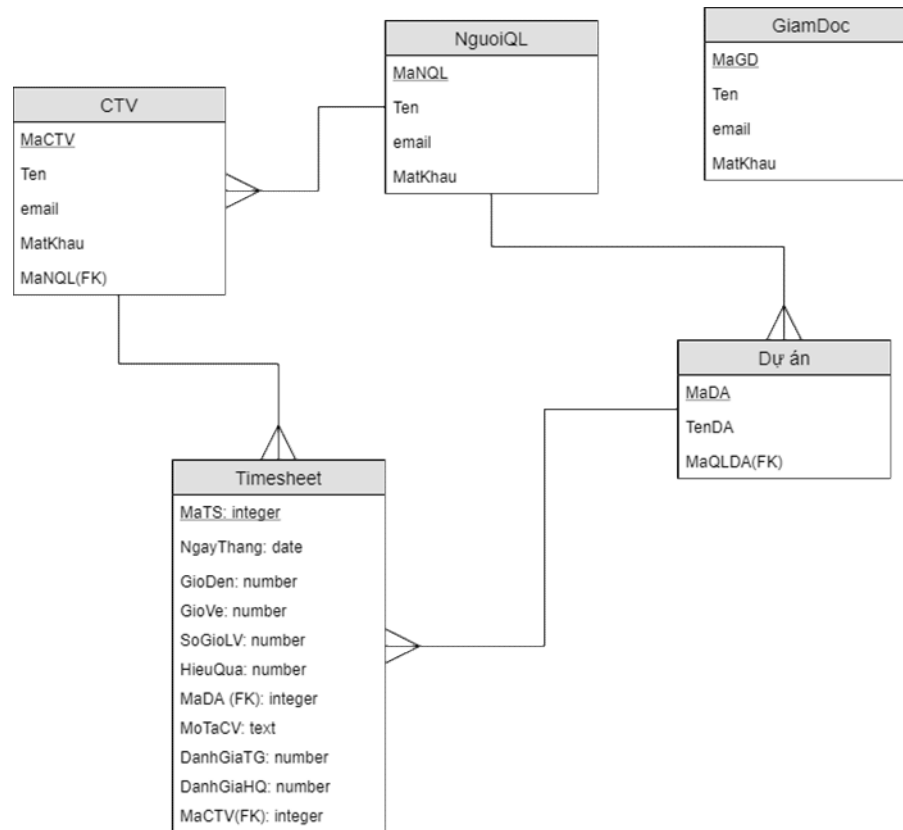
Thiết kế các thủ tục lưu sẵn, trigger

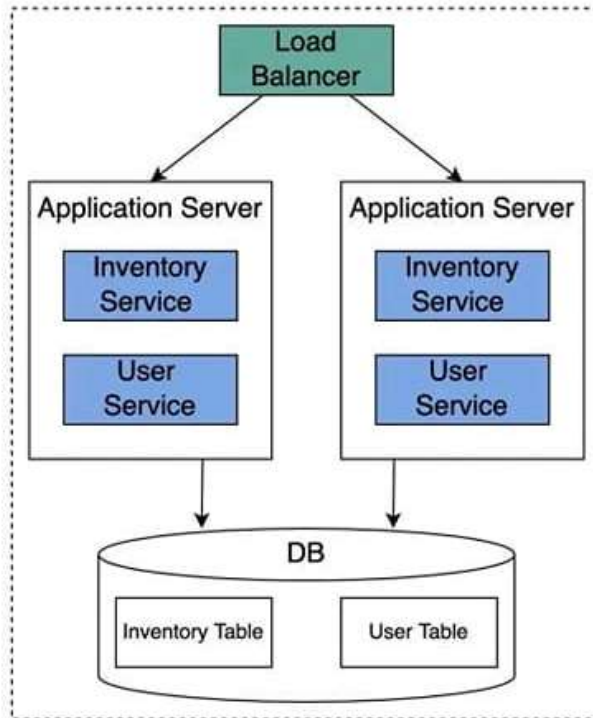
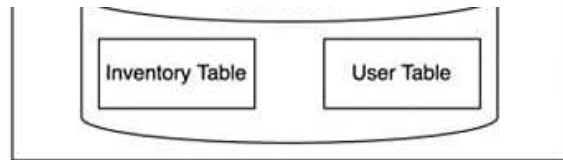
- Nguyên tắc: chức năng của hệ thống càng được đặt sâu (các tầng thấp) càng tốt
- Thủ tục (ràng buộc/logic nghiệp vụ) được đặt càng gần dữ liệu càng tốt

Mở rộng

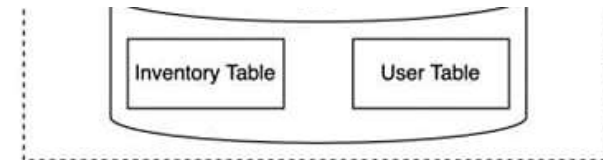
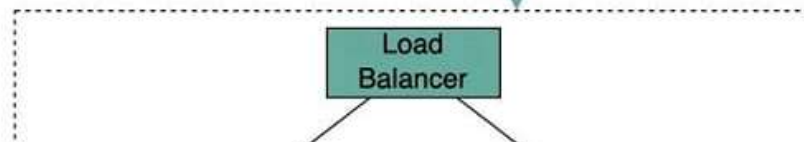
- Các loại hình CSDL
 - SQL
 - NoSQL: document DB, graph DB, timeseries DB, ...
 - New SQL
- Lưu trữ và xử lý các dữ liệu bán cấu trúc, phi cấu trúc

Thí dụ mô hình dữ liệu

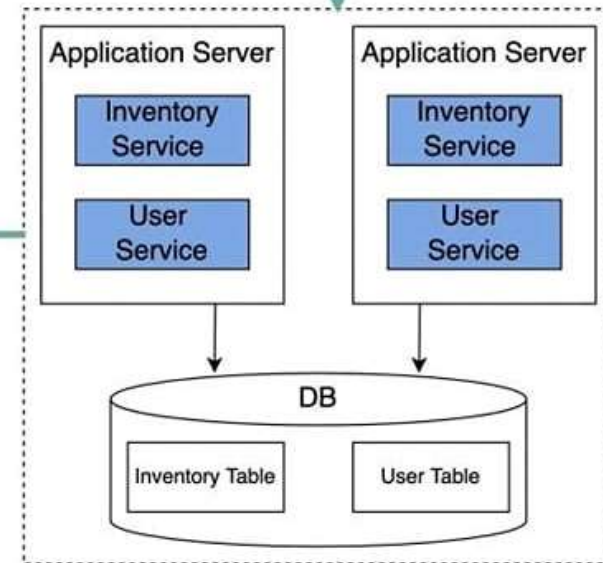




④ Separate DB read/write



② Split services into clusters



③

Add load balancer

