

# Giới thiệu

Phạm Thị Ngọc Diễm  
[ptndiem@cit.ctu.edu.vn](mailto:ptndiem@cit.ctu.edu.vn)

Khoa CNTT & TT - Trường Đại học Cần Thơ

# Nội dung

- Sự phát triển của CSDL
- Ưu nhược điểm của CSDL quan hệ
- Nhu cầu về các mô hình CSDL mới

# Sự phát triển của CSDL



- 1960s: CSDL phân cấp (IMS), CSDL mạng (CODASYL)
- 1970s: CSDL quan hệ
  - Mô hình đơn giản
  - Toàn vẹn dữ liệu
  - An toàn dữ liệu
  - Tính nguyên tố của các giao dịch
  - Độ tin cậy của dữ liệu
- 1986 : Sự xuất hiện CSDL hướng đối tượng
- 1993 : Sự ra đời của chuẩn ODMG cho CSDL HĐT
- 1998 : Chuẩn UML cho thiết kế những ứng dụng HĐT
- 1999 : Chuẩn SQL3 cho CSDL quan hệ - đối tượng
- 2010s: CSDL NoSQL

# Ưu điểm của CSDL quan hệ

- Mô hình dữ liệu đơn giản và dễ hiểu đối với người sử dụng,
- Mô hình dựa trên cơ sở lý thuyết vững chắc, nó cho phép định nghĩa:
  - Các phương pháp thiết kế các lược đồ CSDL(lý thuyết chuẩn hóa) và
  - Ngôn ngữ SQL ( ngôn ngữ truy vấn chuẩn),
- Phù hợp với những ứng dụng quản lý cổ điển đòi hỏi kiểm tra RBTV chặt chẽ như quản lý ngân hàng
- ...

# Nhược điểm của CSDL quan hệ



- Biểu diễn các thực thể của thế giới thực không phong phú
- Dữ liệu và xử lý tách rời nhau
- Quá tải về ngữ nghĩa:
  - quan hệ (relation) vừa có thể được tạo ra từ thực thể (entity) ,
  - vừa có thể được tạo ra từ liên kết (relationship) trong mô hình thực thể- liên kết (ERD: Entity-Relationship Diagram).

# Nhược điểm của CSDL quan hệ



- Cấu trúc dữ liệu đồng nhất: theo chiều dọc và chiều ngang:
  - Các bộ có các thuộc tính giống nhau
  - Các giá trị trong cùng thuộc tính phải cùng miền giá trị
  - Giao tại một dòng và một cột phải là một trị nguyên tố (không chia nhỏ ra được nữa)
  - Nhiều HQT CSDL cho phép lưu trữ và hiển thị BLOB (Binary Large Object) nhưng không cho sử dụng hoặc hiển thị một phần của nó. BLOB không có tính đệ qui và không biểu diễn được khía cạnh hành xử (behaviour aspect) của đối tượng.  
=> quá hạn chế đối với các đối tượng trong thế giới thực, vốn có cấu trúc phức

# Nhược điểm của CSDL quan hệ



- Các thao tác (operation) bị hạn chế:
  - Chỉ có các thao tác trên tập hợp và trên các bộ
  - Không cho phép thêm mới phép toán
  - Không tương ứng giữa ngôn ngữ SQL và ngôn ngữ lập trình (impedance):
    - SQL: ngôn ngữ khai báo, sử dụng cùng lúc nhiều dòng dữ liệu
    - Ngôn ngữ lập trình: ngôn ngữ thủ tục, chỉ có thể sử dụng 1 dòng dữ liệu/ thời điểm
    - Hai loại ngôn ngữ này có cách biểu diễn dữ liệu khác nhau => trình ứng dụng phải chuyển đổi qua lại 2 cách biểu diễn này => không hiệu quả

# Nhược điểm của CSDL quan hệ



- Truy cập nghèo nàn: CSDL quan hệ dùng các truy cập kết hợp (associative access: dựa trên sự kết hợp các tân từ để chọn) mà ít dùng các truy cập thông thường (navigational access: dựa trên sự chuyển động giữa các bộ)  
=> Không sử dụng con trỏ (pointer)
- Các vấn đề về tính cạnh tranh, sử dụng thay đổi lược đồ trong RDBMS
  - Kiểm soát cạnh tranh không còn phù hợp với các giao dịch kéo dài
  - Khó thay đổi lược đồ: quản trị phải can thiệp để thay đổi lược đồ và ứng dụng phải thay đổi cho phù hợp



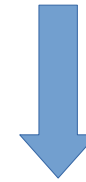
# Nhu cầu về các mô hình CSDL mới



- Nhu cầu về các ứng dụng với các dữ liệu có cấu trúc phức tạp, dữ liệu đa phương tiện,...
- Các công ty đa quốc gia hay công ty có nhiều chi nhánh đặt nhiều nơi => nhu cầu phân tán dữ liệu
- Sự gia tăng nhanh chóng thông tin (ví dụ hơn 14 lần 1990-2000),
- Số lượng giao dịch tăng nhanh (gấp hơn 10 lần trong vòng 5 năm).

=> Cần các mô hình CSDL có khả năng đáp ứng nhanh trên các dữ liệu lớn

**facebook  
data**  
**500+ Terabytes Per Day**  
**2012**



**Facebook generates 4  
petabytes of data per day**  
(<https://kinsta.com/blog/facebook-statistics/>,  
2020)

petabyte (PB) = 1.024 TB = 1 triệu GB

1,5 PB = 10 tỷ ảnh trên Facebook

20 PB = Lượng dữ liệu được Google xử lý hàng ngày trong năm 2008.