

[CNTT > Lập Trình > CSDL Nâng Cao >](#)

## Những khái niệm và kiến trúc của hệ thống cơ sở dữ liệu

### Mô hình dữ liệu, lược đồ và trường hợp (Data Models, Schemas, Instances)

Một trong những đặc điểm cơ bản của cơ sở dữ liệu là cung cấp một số mức độ trừu tượng hoá dữ liệu bằng cách làm ẩn đi cách thức tổ chức dữ liệu- cái mà hầu hết người dùng không cần biết đến.

Mô hình dữ liệu (Data Model): Là một tập những khái niệm dùng để biểu diễn cấu trúc của cơ sở dữ liệu-cung cấp những điều kiện cần thiết để đạt được mức độ trừu tượng dữ liệu. Cấu trúc dữ liệu bao gồm kiểu dữ liệu (data types) và mối quan hệ giữa các dữ liệu (relationships) và những ràng buộc (constraints) mà cơ sở dữ liệu phải tuân theo.

Hầu hết mô hình dữ liệu đều có một tập hợp các thao tác cơ bản (basic operations) để truy vấn và cập nhật dữ liệu.

Thao tác chung (generic operations): Thêm (insert), Xoá (delete), Sửa (modify), Truy cập (retrieve)

Thao tác do người dùng định nghĩa (user-defined operations)

### Phân loại mô hình dữ liệu

Có rất nhiều mô hình dữ liệu đã được đưa ra, chúng ta có thể phân loại chúng theo những kiểu khái niệm mà họ đã dùng để biểu diễn cấu trúc cơ sở dữ liệu. Mô hình dữ liệu được chia làm 3 loại sau

### Mô hình khái niệm (Conceptual (high-level, semantic) data models)

Cung cấp những khái niệm gần gũi với đa số người sử dụng, mô hình này chỉ ra cái gì được đưa vào để quản lý. Mô hình này là phương tiện để những người phân tích thiết kế giao tiếp với người sử dụng, nhằm thu thập thông tin, xác định đúng đắn và đầy đủ yêu cầu của hệ thống.

Mô hình này sử dụng cấu trúc dữ liệu là: thực thể (entity), thuộc tính (attribute) và mối liên kết (relationship)

### Mô hình dữ liệu vật lý (Physical (low-level, internal) data models)

Cung cấp những khái niệm để biểu diễn chi tiết cách thức dữ liệu được lưu trữ trong máy tính. Mô hình này chỉ ra định dạng bản ghi (record formats), thứ tự sắp xếp các bản ghi (record ordering) và đường dẫn để truy cập dữ liệu (access paths).

### Mô hình dữ liệu thể hiện (Implementation (record-oriented) data models)

Mô tả các dữ liệu bằng cách sử dụng những ký pháp tương ứng với mô hình dữ liệu mà một hệ quản trị cơ sở dữ liệu sử dụng.

Các loại mô hình cơ sở dữ liệu thể hiện

#### Mô hình phân cấp

Mô hình CSDL phân cấp được biểu diễn dưới dạng cây và các đỉnh của cây là các bản ghi. Các bản ghi liên kết với nhau theo mối quan hệ cha-con.

Một cha có nhiều con

Một con chỉ có một cha



Minh họa mô hình cơ sở dữ liệu phân cấp

. Minh họa mô hình cơ sở dữ liệu phân cấp

### Ưu điểm

Thể hiện dễ dàng quan hệ 1-N.

Việc phân chia dữ liệu dễ thể hiện, đảm bảo an toàn dữ liệu

Tính độc lập của chương trình và các dữ liệu được đảm bảo


### Nhược điểm

Không thể hiện được mối quan hệ M-N

Trong một hệ thống phân cấp, dữ liệu được tổ chức như trên dẫn đến khó sửa đổi dữ liệu.

### Mô hình mạng

Cấu trúc cơ bản trong mô hình mạng là những tập hợp và mỗi tập hợp có bản ghi là bản ghi chủ và một số bản ghi thành viên. Mỗi thành viên có thể thuộc về nhiều tập hợp.

 Minh họa mô hình cơ sở dữ liệu mạng

### Ưu điểm

Dễ thể hiện mối liên kết M-N

Kiểu truy cập dữ liệu mềm dẻo hơn kiểu phân cấp

### Nhược điểm

Việc sửa đổi số liệu khó khăn.

Với những lập trình viên, việc thiết kế CSDL khó.

### Mô hình quan hệ

Trong mô hình quan hệ, các dữ liệu được biểu diễn ở dạng các bảng với các dòng và các cột.

Trong mô hình quan hệ không có một cấu trúc vật lý nào của dữ liệu mô tả sự kết nối giữa các bảng. Thay vào đó, sự kết nối giữa các bảng được mô tả logic bằng các giá trị được lưu trữ trong các dòng của bảng. Chẳng hạn trong hình dưới đây, thuộc tính ProCode(Mã tỉnh) được lưu trong cả 2 bảng PROVINCE và bảng STUDENT, giá trị chung này cho phép người dùng liên kết được 2 bảng.

 Minh họa mô hình cơ sở dữ liệu quan hệ

### Mô hình hướng đối tượng

Trong mô hình hướng đối tượng, các thuộc tính dữ liệu và các thao tác trên các dữ liệu này được bao gói trong một cấu trúc gọi là đối tượng.

Đối tượng có thể chứa các dữ liệu phức hợp như văn bản, hình ảnh, tiếng nói và hình ảnh động. Một đối tượng có thể yêu cầu hoặc xử lý dữ liệu từ một đối tượng khác bằng việc gửi đi một thông báo đến đối tượng đó. Mô hình hướng đối tượng biểu diễn một sơ đồ mới để lưu trữ và thao tác dữ liệu. Từ một đối tượng có thể sinh ra một đối tượng khác.

 Minh họa mô hình cơ sở dữ liệu hướng đối tượng

## Lược đồ (Schema) , minh họa (instances), và trạng thái (State)

Lược đồ cơ sở dữ liệu (Database Schema): là biểu diễn của cơ sở dữ liệu, bao gồm cấu trúc cơ sở dữ liệu và những ràng buộc trên dữ liệu.

Sơ đồ của lược đồ cơ sở dữ liệu (Schema Diagram): Là lược đồ cơ sở dữ liệu được biểu diễn thông qua sơ đồ.

 Lược đồ cơ sở dữ liệu UNIVERSITY

Minh học cơ sở dữ liệu (Database Instance): Là dữ liệu thực sự được lưu trữ trong cơ sở dữ liệu ở thời điểm hiện tại. Database Instance cũng được gọi là trạng thái của cơ sở dữ liệu (database state)



Cơ sở dữ liệu UNIVERSITY

Như vậy, Database Schema rất hiếm khi thay đổi, còn Database State thay đổi bất kỳ khi nào có sự cập nhập dữ liệu.

## Lược đồ kiến trúc của hệ quản trị cơ sở dữ liệu (DBMS Architecture) và sự độc lập dữ liệu (Data Independence)

Như chúng ta đã biết, các tính chất quan trọng nhất của cơ sở dữ liệu là: (1) Đảm bảo sự độc lập giữa chương trình ứng dụng và dữ liệu. (2) Hỗ trợ nhiều khung nhìn cho các đối tượng người dùng khác nhau. (3) Sử dụng danh mục để lưu trữ biểu diễn dữ liệu (schema). Trong phần này, chúng ta sẽ tìm hiểu kiến trúc của hệ quản trị cơ sở dữ liệu, gọi là Lược đồ kiến trúc 3 mức (three – schema architecture). Sau đó chúng ta sẽ tìm hiểu về khái niệm độc lập dữ liệu.

### Lược đồ kiến trúc 3 mức của HQTCSĐL

Mục đích của việc phân thành 3 mức trong kiến trúc của mô hình HQTCSĐL là để tách biệt các ứng dụng của người sử dụng với cơ sở dữ liệu vật lý. Trong kiến trúc này, lược đồ có thể được định nghĩa ở 3 mức sau

Lược đồ mức trong (Internal schema) ở Mức trong (Internal level) để biểu diễn chi tiết cấu trúc lưu trữ dữ liệu và cách thức truy cập dữ liệu. Lược đồ mức trong sử dụng mô hình dữ liệu vật lý (physical data model).

Lược đồ khái niệm (Conceptual schema) ở Mức khái niệm (Conceptual level) để biểu diễn cấu trúc và các ràng buộc trong toàn bộ cơ sở dữ liệu phục vụ cho việc giao tiếp với người sử dụng. Lược đồ khái niệm ẩn đi cách thức tổ chức vật lý của dữ liệu, chỉ tập trung vào việc biểu diễn các thực thể, các kiểu dữ liệu, mối quan hệ giữa các thực thể, các thao tác của người sử dụng và các ràng buộc giữa các dữ liệu. Mô hình dữ liệu mức khái niệm (Conceptual data model) hoặc Mô hình dữ liệu thể hiện (Implementation data model) có thể được sử dụng ở mức này.

Lược đồ mức ngoài (External Level) ở Mức ngoài (External level hoặc View level) để biểu diễn hàng loạt những khung nhìn của người sử dụng (user views). Mô hình dữ liệu mức cao (High-level data model) hoặc Mô hình dữ liệu thể hiện (Implementation data model) có thể được sử dụng ở mức này.

Ánh xạ giữa các mức này là cần thiết. Những chương trình làm việc với dữ liệu ở mức ngoài và được HQTCSĐL ánh xạ tới dữ liệu vật lý ở mức trong để thực hiện.



Lược đồ kiến trúc 3 mức của HQTCSĐL

### Độc lập dữ liệu

Kiến trúc 3 mức của HQTCSĐL có thể được sử dụng để giải thích khái niệm về độc lập dữ liệu. Độc lập dữ liệu là khả năng thay đổi lược đồ ở một mức nào đó của hệ thống cơ sở dữ liệu mà không cần phải thay đổi lược đồ ở mức cao hơn. Chúng ta có thể định nghĩa 2 kiểu của độc lập dữ liệu:

Độc lập dữ liệu logic (Logical data independence): cho phép thay đổi lược đồ khái niệm mà không cần phải thay đổi lược đồ mức ngoài hoặc những chương trình ứng dụng. Chúng ta có thể thay đổi lược đồ khái niệm để mở rộng (thêm các trường dữ liệu, các bản ghi) hoặc thu nhỏ cơ sở dữ liệu (xóa các trường dữ liệu, các bản ghi).

Độc lập dữ liệu vật lý (Physical data independence): cho phép thay đổi lược đồ mức trong mà không cần thay đổi lược đồ khái niệm. Có khi chúng ta cần thay đổi lược đồ mức trong vì các file vật lý đôi khi cần tổ chức lại để tăng hiệu quả thực hiện. Nếu kiểu dữ liệu không thay đổi thì chúng ta không cần thay đổi lại lược đồ khái niệm.

### Ngôn ngữ của HQTCSĐL

Vì HQTCSĐL phục vụ có nhiều đối tượng người sử dụng khác nhau nên nó phải hỗ trợ ngôn ngữ để người sử dụng tương tác với nó. Trong phần này chúng ta sẽ tìm hiểu một số những ngôn ngữ được HQTCSĐL hỗ trợ.

Ngôn ngữ định nghĩa dữ liệu (Data Definition Language - DDL): Là ngôn ngữ được các nhà quản trị cơ sở dữ liệu (DBA) và các nhà thiết kế cơ sở dữ liệu (database designers) dùng để xây dựng lược đồ khái niệm của cơ sở dữ liệu. Trong nhiều HQTCSĐL, DDL cũng được sử dụng để định nghĩa lược đồ mức trong và mức ngoài (các khung nhìn). Một số HQTCSĐL chia thành 2 ngôn ngữ: Ngôn ngữ định nghĩa lưu trữ (storage definition language – SDL) và Ngôn ngữ định nghĩa khung nhìn (view definition language -VDL) được dùng để định nghĩa lược đồ mức trong và mức ngoài.

Ngôn ngữ thực hiện dữ liệu (Data Manipulation Language -DML): Là ngôn ngữ được sử dụng để thao tác với dữ liệu bao gồm việc truy cập đến bản ghi và cập nhật dữ liệu cho bản ghi (thêm, sửa, xóa). Các lệnh DML có thể được nhúng trong ngôn ngữ lập trình hoặc thực hiện độc lập (ngôn ngữ truy vấn).

## Các tính năng của HQTCSĐL

Nạp dữ liệu đang lưu trữ ở các tệp tin vào cơ sở dữ liệu (Conversion Tool).

Cung cấp các thao tác truy xuất

Đảm bảo tính độc lập dữ liệu

Cung cấp thủ tục sao lưu và phục hồi (backup và recovery)

Cung cấp các thủ tục điều khiển đồng thời (Do tính truy xuất đồng thời và cạnh tranh)

Cung cấp các thủ tục kiểm soát bản quyền, kiểm tra tính đúng đắn của dữ liệu (để đảm bảo tính an toàn và toàn vẹn dữ liệu)

## Phân loại HQTCSĐL

Người ta phân loại HQTCSĐL dựa trên một số tiêu chí

Dựa trên mô hình dữ liệu được HQTCSĐL sử dụng: Mô hình quan hệ (Relational Data Model), Mô hình hướng đối tượng (Object Data Model), Mô hình mạng (Network Data Model), mô hình phân cấp (Hierarchical Data Model), mô hình quan hệ thực thể (Entity – Relationship Data Model)...

Tiêu chí khác

Số người sử dụng HQTCSĐL: Đơn người dùng (Single-user), Đa người dùng (multi-user). Hầu hết các HQTCSĐL hiện nay đều là HQTCSĐL đa người dùng.

Vị trí HQTCSĐL: tập trung (sử dụng một máy đơn) hay phân tán (sử dụng nhiều máy tính).

Giá của HQTCSĐL: 100\$- 300\$; 10000\$- 100000\$