|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Họ tên** | **MSSV** | **Hình ảnh** | **Ghi chú** |
| 1 | Hà Thị Phương Thảo | 0912430 | Description: 0912430 |  |
| 2 | Nguyễn Thị Thanh Thảo | 0912431 | Description: 0912431 |  |
| 3 | Trương Nguyễn Thủy Tiên | 0912463 | Description: 0912463 |  |
| 4 | Nguyễn Văn Tiến | 0912469 | Description: 0912469 | Nhóm trưởng |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Nội dung công việc** | **Thành viên 1**  **0912430** | **Thành viên 2**  **0912431** | **Thành viên 3**  **0912463** | **Thành viên 4**  **0912469** |
| 1 | Dịch tài liệu |  |  |  |  |
| 2 | Tổng hợp kết quả, viết tài liệu |  |  |  |  |

### Topic #8b – Cơ sở dữ liệu hướng đối tượng

Mục lục

[I. DẪN NHẬP: 5](#_Toc309861610)

[II. CÁC LOẠI HỆ THỐNG CƠ SỞ DỮ LIỆU ĐỐI TƯỢNG:.](#_Toc309861611) 5

[III. Nội dung trình bày: 6](#_Toc309861617)

[1. Minh họa lý do tại sao lại mở rộng mô hình quan hệ: 6](#_Toc309861619)

1.[1. Kiểu dữ liệu mới.](#_Toc309861620) 7

[1.2 Thao tác với các loại dữ liệu mới 8](#_Toc309861619)

[2. Thảo luận về kiểu dữ liệu trừu tượng, định nghĩa và thao tác. 1](#_Toc309861620)0

2.[1. Định nghĩa phướng thức của kiểu dữ liệu trừu tượng 1](#_Toc309861619)2

[3. Cách thức và cấu trúc của kiểu dữ liệu trừu tượng. 1](#_Toc309861620)2

3.[1. Thao tác ở kiểu dữ liệu có cấu trúc 1](#_Toc309861619)2

[4. Xem xét và nhận dạng đối tượng.. 1](#_Toc309861620)2

4.[1. Khái niệm về Equality 1](#_Toc309861619)3

4.[2. Bất tham chiếu đến các loại tham chiếu. 1](#_Toc309861620)4

[5. Sự kế thừa và phân cấp các loại. 1](#_Toc309861619)4

[5.1. Định nghĩa các loại với sự kế thừa. 1](#_Toc309861620)4

5.[2. Tính ràng buộc của những phương thức 1](#_Toc309861619)5

[5.3. Tập các hệ thống phân cấp, loại mở rộng, và truy vấn. 1](#_Toc309861620)5

[6. Xem xét làm thế nào để tận dụng lợi thế của khái niệm hướng đối tượng mới để thiết kế cơ sở dữ liệu ORDBMS 1](#_Toc309861619)6

[6.1. Loại có cấu trúc và loại trừu tượng. 1](#_Toc309861620)6

[6.2. Định danh đối tượng 1](#_Toc309861620)7

[6.2. Mở rộng mô hình thực thể kết hợp (ER) 1](#_Toc309861619)8

[6.3. Sử dụng tập Nested(Lồng nhau). 1](#_Toc309861620)9

[7. Thảo luận một số thách thức mới đặt ra bởi hệ thống quan hệ đối tượng. 1](#_Toc309861619)9

[7.1. Lưu trữ và truy cập phương thức 1](#_Toc309861620)9

7.[2. Chỉ mục trên loại dữ liệu mới 1](#_Toc309861619)9

7.[3. Xử lý truy vấn](#_Toc309861620) 20

7.4[. Tối ưu truy vấn 21](#_Toc309861619)

[8. Thảo luận về ODL và OQL, các tiểu chuẩn cho OODBMSs. 2](#_Toc309861620)2

[8.1. Mô hình dữ liệu ODMG và ODL 2](#_Toc309861619)2

[8.2. OQL 2](#_Toc309861620)2

[9. Trình bày so sánh ngắn gọn của ORDBMSs và OODBMs 3](#_Toc309861619)3

[9.1. So sánh RDBMS và ORDBMS 2](#_Toc309861620)3

9.[2. So sánh OODBMS và ORDBMS 2](#_Toc309861619)3

9.[3. So sánh OODBMS và ORDBMS 2](#_Toc309861620)3

**CƠ SỞ DỮ LIỆU ĐỐI TƯỢNG**

**Nội dung trình bày:**

* Từ phần 25.1 đến 25.5 tạo động cơ và giới thiệu khái niệm về đối tượng. Các khải niệm được thảo luận trong các phần phổ biến cho cả OODBMSs và ORDBMSs, mặc dù cứ pháp tương tự SQL: 1999. Cụ thể:
  + Phần 25.1: Minh họa lý do tại sao lại mở rộng mô hình quan hệ.
  + Phần 25.2: Thảo luận về kiểu dữ liệu trừu tượng, định nghĩa và thao tác.
  + Phần 25.3: Cách thức và cấu trúc của kiểu dữ liệu trừu tượng.
  + Phần 25.4: Xem xét và nhận dạng đối tượng.
  + Phần 25.5: Sự kế thừa và phân cấp các loại.
* Phần 25.6: Xem xét làm thế nào để tận dụng lợi thế của khái niệm hướng đối tượng mới để thiết kế cơ sở dữ liệu ORDBMS.
* Phần 25.7: Thảo luận một số thách thức mới đặt ra bởi hệ thống quan hệ đối tượng.
* Phần 25.8: Thảo luận về ODL và OQL, các tiểu chuẩn cho OODBMSs.
* Phần 25.9: Trình bày so sánh ngắn gọn của ORDBMSs và OODBMs.

1. **DẪN NHẬP:**

* Với mô hình, hệ thống cơ sở dữ liệu quan hệ hỗ trợ rất tốt các kiểu dữ liệu nhỏ, các tập cố định như (integer, dates, strings), Trong đó có cho thấy hỗ trợ đầy đủ cho các lĩnh vực ứng dụng truyền thống như xử lý dữ liệu hành chính. Trong nhiều lĩnh vực ứng dụng, tuy nhiên, các loại dữ liệu phức tạp hơn nhiều cần được xử lý. Thông thường các loại dữ liệu phức tạp được lưu trữ trong một hệ thống file lưu trữ. Ví dụ về các lĩnh vực với các dữ liệu phức tạp bao gồm hỗ trợ thiết kế và mô hình hóa máy tính (CAD / CAM), kho lưu trữ đa phương tiện, và quản lý tài liệu.
* Với một lượng lớn dữ liệu ngày càng tăng nhanh chóng, một số tính năng từ DBMS. Ví dụ: giảm thời gian phát triển ứng dụng, đồng thời kiểm soát và phục hồi dữ liệu, hỗ trợ lập chỉ mục, và khả năng truy vấn ngày càng trở nên cần thiết và hấp dẫn. Để hỗ trợ các ứng dụng như vậy, một DBMS phải hỗ trợ các kiểu dữ liệu phức tạp. Khái niệm hướng đối tượng có ảnh hướng mạnh mẻ đến những nổ lực tăng cường hỗ trợ cơ sở liệu phức tạp và dẫn đến sự phát triển của hệ thống cơ sở dữ liệu đối tượng.

1. **CÁC LOẠI HỆ THỐNG CƠ SỞ DỮ LIỆU ĐỐI TƯỢNG:**

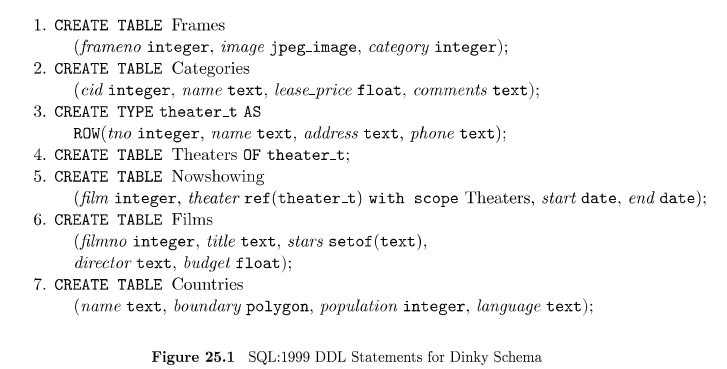
* Hệ thống cơ sở dữ liệu đối tượng có 2 hướng phát triển khác nhau:
  + Hệ thống cơ sở dữ liệu hướng đối đối tượng: hệ thống cơ sở dữ liệu hướng đối tượng là những lĩnh vực nơi các đối tượng phức tạp đóng một vai trò trung tâm. Cách tiếp cận này bị ảnh hưởng nhiều bởi các ngôn ngữ lập trình hướng đối tượng và có thể được hiểu như là một cố gắng để thêm chức năng  DBMS một môi trường ngôn ngữ lập trình.
  + Hệ thống cơ sở dữ liệu quan hệ đối tượng:  Nó có thể được coi là một nỗ lực để mở rộng hệ thống cơ sở dữ liệu quan hệ với các chức năng cần thiết để hỗ trợ một class rộng lớn hơn của các ứng dụng, và bằng nhiều cách, cung cấp một cầu nối giữa các mô hình quan hệ và hướng đối tượng.

🡪Trong seminar này sẽ tập trung vào cơ sở dữ liệu quan hệ đối tượng và nhấn mạnh làm thế nào nó có thể được xem như là một sự phát triển của RDMSs, chứ không phải là một mô hình hoàn toàn khác nhau.

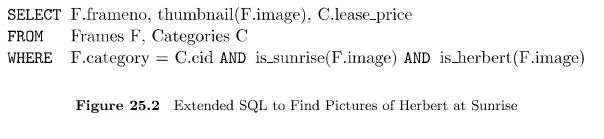
* SQL: tiêu chuẩn năm 1999 là dựa trên mô hình ORDBMS, chứ không phải là mô hình OODBMS. Tiêu chuẩn này bao gồm hỗ trợ cho rất nhiều các tính năng kiểu dữ liệu phức tạp được thảo luận trong chương này. Bài seminar này tập trung vào việc phát triển các khái niệm cơ bản, chứ không phải là trình bày SQL: 1999, một số trong những tính năng mà chúng tôi thảo luận là không bao gồm trong SQL: 1999. Chúng tôi đã cố gắng để phù hợp với SQL: năm 1999 với ký hiệu, mặc dù chúng tôi có đôi khi tách ra một chút cho rõ ràng cho rõ ràng. Điều quan trọng là nhận ra rằng các khái niệm chính được thảo luận phổ biến cho cả hai ORDBMSs và OODBMSs, và chúng tôi thảo luận làm thế nào họ được hỗ trợ trong tiêu chuẩn ODL / OQL được đề xuất cho OODBMSs tại Mục VIII.
* Các nhà cung cấp RDBMS bao gồm cả IBM, Informix, Oracle, có thêm chức năng ORDBMS (mức độ khác nhau) trong các sản phẩm của họ, và điều quan trọng là nhận ra làm thế nào các cơ quan hiện có của kiến thức về thiết kế và thực hiện các cơ sở dữ liệu quan hệ có thể được thừa hưởng để giải quyết các ORDBMS phần mở rộng. Nó cũng rất quan trọng để hiểu những thách thức và cơ hội mà các phần mở rộng trình bày cho người sử dụng cơ sở dữ liệu, thiết kế, và người thực hiện.

1. **Nội dung trình bày:**
2. **Minh họa lý do tại sao lại mở rộng mô hình quan hệ:**

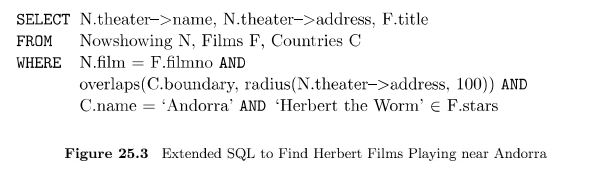
* Như một ví dụ cụ thể về sự cần thiết cho các hệ thống quan hệ đối tượng, chúng tôi tập trung vào một vấn đề kinh doanh xử lý dữ liệu mới khó hơn và (theo quan điểm của tác giả ) giải trí nhiều hơn so với đồng đô la và xu sổ sách kế toán của thập kỷ trước.
* Ngày nay, các công ty trong các ngành công nghiệp như giải trí trong kinh doanh bán dữ liệu số, tài sản cơ bản của công ty của họ không phải là sản phẩm hữu hình, nhưng thay vì các đồ tạo tác phần mềm như video và âm thanh.
* Chúng tôi xem xét Công ty tưởng tượng về Giải trí dinky, một tập đoàn lớn Hollywood có tài sản chính là một bộ sưu tập của các nhân vật hoạt hình, đặc biệt là Herbert âu yếm và quốc tế yêu quý Worm.Dinky có một số bộ phim Herbert Worm, nhiều trong số đó đang được thể hiện trong các rạp chiếu phim trên toàn thế giới tại bất kỳ thời điểm nào. Dinky cũng làm cho một thỏa thuận tốt về tiền bạc cấp giấy phép Herbert của hình ảnh, giọng nói, và đoạn băng video cho các mục đích khác nhau, con số hành động, trò chơi video, xác nhận sản phẩm, và như vậy. Cơ sở dữ liệu của Dinky được sử dụng để quản lý việc bán hàng và hồ sơ cho thuê cho các sản phẩm khác nhau liên quan đến Herbert, cũng như video và âm thanh tạo nên nhiều bộ phim của Herbert.
  1. **Kiểu dữ liệu mới**
* Một vấn đề cơ bản phải đối mặt khi thiết kế cơ sở dữ liệu của Dinky là cần hỗ trợ các loại dữ liệu phong phú hơn đáng kể hơn là có sẳn trong một DBMS:
  + **Người dùng định nghĩa kiểu dữ liệu trừu tượng (ADTs):** Những dữ liệu của Dinky bao gồm hình ảnh, giọng nói, video của Herbert và chúng phải được lưu trữ trong cơ sở dữ liệu. Hơn nữa chúng ta cần các chức năng đặc biệt để thao tác với các đối tượng này. Vi dụ như chúng ta có thể viết các chức năng tạo ra bản nén của một hình ảnh hoặc hình ảnh có độ phân giải thấp hơn.
  + **Kiểu dữ liệu có cấu trúc:** Trong các ứng dụng quả thực chúng ta cần xây dựng loại mới dựa vào các hàm dựng như sets, tuples, arrays, sequences.
  + **Tính kế thừa:** Khi số lượng các loại dữ liệu phát triển, điều quan trọng là phải nhận ra sự phổ biến giữa các loại và tận dụng lợi thế của nó. Ví dụ nén hình ảnh và hình ảnh có độ phân giải thấp là 2 loại, nhưng ở mức độ nào đó, chỉ cần 1 loại là đủ. Do đó, mong muốn kế thừa một số tính năng của các đối tượng hình ảnh khi xác định các đối tượng hình ảnh nén và các đối tượng có độ phân giải thấp.
* Làm thế nào chúng ta có thể giải quyết những vấn đề này trong một RDBMS? Chúng ta có thể lưu trữ hình ảnh, video, và như vậy là BLOB trong hệ thống hiện tại quan hệ, một nhị phân lớn đối tượng (BLOB) là chỉ một dòng dài byte, và việc của DBMS hỗ trợ lưu trữ và lấy BLOB trong cách một ví dụ mà người dùng không phải lo lắng về kích thước của BLOB, một BLOB có thể chiều dài một vài trang, không giống như một thuộc tính truyền thống. Tất cả các quá trình gia công, chế biến thêm BLOB đã được thực hiện bằng chương trình ứng dụng của người sử dụng, trong ngôn ngữ máy chủ lưu trữ trong đó các mã SQL được nhúng vào. Giải pháp này không hiệu quả bởi vì chúng ta buộc phải lấy tất cả các BLOB trong một bộ sưu tập ngay cả trong số họ có thể được lọc ra câu trả lời bằng cách áp dụng chức năng người dùng định nghĩa (trong DBMS). Nó là không thỏa đáng từ một quan điểm nhất quán dữ liệu, bởi vì ngữ nghĩa của dữ liệu là phụ thuộc nhiều vào mã ứng dụng ngôn ngữ sở tại và không thể được thi hành bởi DBMS
* Đối với các loại cấu trúc và thừa kế, chỉ đơn giản là không có hỗ trợ trong mô hình quan hệ. Chúng ta buộc phải ánh xạ dữ liệu với cấu trúc phức tạp như vậy vào một bộ sưu tập các bảng phẳng. (Chúng ta đã thấy ví dụ của ánh xạ như vậy khi chúng tôi thảo luận về các bản dịch từ sơ đồ ER, thừa kế cho mối quan hệ tại Chương 2.)
* Ứng dụng này rõ ràng đòi hỏi các tính năng không có sẵn trong mô hình quan hệ. Là một minh họa các tính năng, hình 25,1 hiện nay SQL: 1999 DDL báo cáo cho phần của giản đồ ORDBMS Dinky của mà sẽ được sử dụng trong ví dụ tiếp theo Mặc dù DDL là một hệ thống quan hệ truyền thống, nó có một số khác biệt quan trọng mà làm nổi bật việc dữ liệu mới mô hình khả năng của một ORDBMS.Nhìn lướt qua các câu lệnh DDL là đủ cho bây giờ, chúng tôi sẽ nghiên cứu chi tiết trong phần tiếp theo, sau khi trình bày một số các khái niệm cơ bản mà ứng dụng mẫu của chúng tôi cho thấy là cần thiết trong một DBMS thế hệ tiếp theo
  1. **Thao tác với các loại dữ liệu mới**
* Như vậy đến thời điểm này, chúng ta đã mô tả các loại dữ liệu phải được lưu trữ trong cơ sở dữ liệu dinky. Ở đây ta không nói bất cứ điều gì về cách sử dụng các loại mới trong các truy vấn, vì vậy chúng ta hãy nghiên cứu hai truy vấn dữ liệu của dinky cần phải hỗ trợ. Cú pháp của các truy vấn không quan trọng, nó là đủ để hiểu những gì họ thể hiện.Chúng tôi sẽ trở về đặc biệt của cú pháp truy vấn như chúng ta tiến hành



* Thách thức đầu tiên  đến từ công ty Clog cung cấp ngũ cốc ăn sáng. Clog sản xuất ngũ cốc được gọi là Delirios, và muốn thuê một hình ảnh của Herbert Worm ở phía trước của mặt trời mọc, để kết hợp trong thiết kế hộp Delirios. Một truy vấn để trình bày một bộ sưu tập các hình ảnh có thời điểm này có và giá cho thuê có thể được thể hiện trong SQL-giống như cú pháp như trong hình25,2. Dinky có một phương pháp số được viết bằng một ngôn ngữ bắt buộc như Java và đăng ký với hệ thống cơ sở dữ liệu. Những phương pháp này có thể được sử dụng trong các truy vấn trong cùng xây dựng trong các phương pháp, chẳng hạn như =, +, -, được sử dụng trong một ngôn ngữ quan hệ như SQL. Các phương pháp hình ảnh thu nhỏ trong mệnh đề Select sản xuất một phiên bản nhỏ của hình ảnh đầu vào đầy đủ kích thước của nó.Các phương pháp is\_sunrise là một chức năng boolean phân tích một hình ảnh và trả về true nếu ảnh có chứa một mặt trời mọc;phương pháp is\_herbert trả về true nếu ảnh có chứa một hình ảnh của Herbert.Truy vấn sản xuất mã số khung, thumhnail hình ảnh, và giá cho tất cả các khung có chứa Herbert và mặt trời mọc.



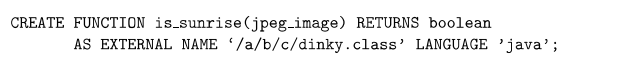
* Thách thức thứ hai đến từ giám đốc điều hành dinky.Họ biết Delirios là cực kỳ phổ biến ở đất nước nhỏ bé của Andorra, vì vậy họ muốn đảm bảo rằng một số bộ phim Herbert đang chơi tại nhà hát gần Andorra khi cung cấp ngũ cốc việc kệ hàng. Để kiểm tra về hiện trạng của vấn đề, việc giám đốc điều hành muốn tìm tên của tất cả các rạp chiếu phim cho thấy có bộ phim Herbert trong vòng 100km Andorra. Hình 25,3 cho có truy vấn này trong một cú pháp SQL:



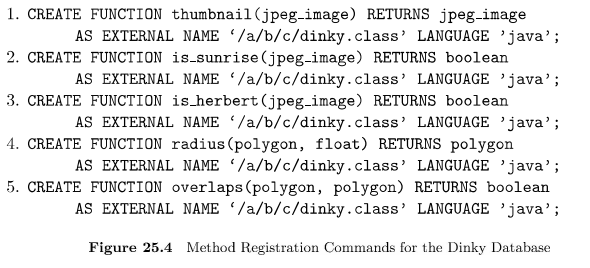
* Các thuộc tính rạp chiếu phim của bảng Nowshowing là một tham chiếu đến một đối tượng trong bảng khác, trong đó có các thuộc tính name, address, và location. Tham chiếu đối tượng này cho phép việc ký hiệu N.theater-> name và N.theater-> address, trong đó đề cập các thuộc tính của đối tượng theater\_t tham chiếu có trong hàng Nowshowing N. Thuộc tính có ngôi sao của bảng bộ phim là một tập hợp các tên mỗi ngôi sao của bộ phim. Phương pháp radius trả về một vòng có trung tâm ở số đầu tiên của nó với radius tương đương với số thứ hai của có. Phương pháp overlaps kiểm tra chồng chéo lên nhau không gian. Vì vậy, Nowshowing và Films là sự tham gia của các điều khoản kết bằng, trong khi Nowshowing và các nước tham gia củaviệc khoản chồng chéo không có. Các lựa chọn "Andorra và phim ảnh có chứa ‘Hertbert Worm’ hoàn chỉnh với câu truy vấn.
* Hai câu truy vấn trên tương tự như truy vấn SQL-92 nhưng có một số tính năng không bình thường:
  + Phương thức do người dùng định nghĩa: Người dùng tự định nghĩa các kiểu dữ liệu trừu tượng để thực thi cho các phương thức của họ ví dụ is\_heabert.
  + Các toán tử có cầu trúc: Cùng với các loại cấu trúc có sẵn trong mô hình dữ liệu, ORDBMSs cung cấp việc phương pháp tự nhiên cho các loại dữ liệu đó. Ví dụ, các thiết lập của chuẩn của các loại…
  + Toán tử liên kết: tính tham chiếu trong loại được thế hiện bằng toán tử -> (arrow).

1. **Thảo luận về kiểu dữ liệu trừu tượng, định nghĩa và thao tác**

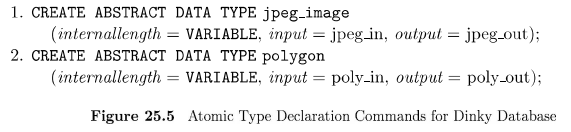
* Xem xét bảng Frames - 25,1. Nó có một cột hình ảnh của loại jpeg\_image, dùng để lưu trữ một hình ảnh nén đại diện cho một Frame duy nhất của một bộ phim. Các loại jpeg\_image không nằm trong nhân DBMS được xây dựng trong các loại và được xác định bởi một người dùng cho việc ứng dụng dinky để lưu trữ dữ liệu hình ảnh nhất nén bằng cách sử dụng tiêu chuẩn JPEG. Một ví dụ khác, bảng Countries xác định trong dòng 6 của hình 25,1 có một ranh giới cột kiểu của đa giác, trong đó có đại diện của các hình dạng của các Countries vạch ra trên bản đồ thế giới.
* Cho phép người dùng định nghĩa kiểu dữ liệu tùy ý mới là một tính năng quan trọng của ORDBMSs. DBMS cho phép người dùng lưu trữ và lấy đối tượng jpeg\_image loại, giống như một đối tượng của bất kỳ loại khác, chẳng hạn như số nguyên. Các kiểu dữ liệu nguyên tử thường cần phải có các loại hoạt động cụ thể quy định bởi người sử dụng tạo ra chúng. Ví dụ, người ta có thể xác định các hoạt động trên một kiểu dữ liệu hình ảnh và các cộng sự của nó phương pháp được gọi là một kiểu dữ liệu trừu tượng, hoặc ADT. Truyền thống SQL đi các với xây dựng trong ADTs, chẳng hạn như số nguyên (với các phương pháp số học có liên quan), hoặc các chuỗi (với sự bình đẳng, so sánh, và phương thức tương tự). Hệ thống quan hệ đối tượng bao gồm các ADTs và cũng cho phép người dùng xác định ADTs của riêng mình.
* Nhãn "abstract" được áp dụng cho các loại dữ liệu vì hệ thống cơ sở dữ liệu không cần phải biết làm thế nào dữ liệu của ADT được lưu trữ cũng như phương thức làm việc của ADT. Nó chỉ đơn thuần là cần phải biết những phương thức có sẵn và các loại đầu vào và ouput cho các phương thức. Cất giấu bên trong ADT được gọi là đóng gói. Lưu ý rằng ngay cả trong một hệ thống quan hệ, các loại nguyên tử chẳng hạn như số nguyên có liên quan đến phương pháp được đóng gói vào ADTs. Trong trường hợp các số nguyên, phương pháp tiêu chuẩn cho ADT là các nhà khai thác số liệu thông thường và so sánh. Để đánh giá các nhà điều hành bổ liệu vào số nguyên, hệ thống cơ sở dữ liệu không cần phải hiểu những quy định của - luật bổ sung - nó chỉ cần biết làm thế nào để gọi phương thức đoạn mã khai thác và loại dữ liệu để mong đợi trong trả về.
* Trong một hệ thống quan hệ đối tượng, đơn giản hóa do sự đóng gói là rất quan trọng bởi vì nó che giấu bất kỳ sự phân biệt nội dung giữa các loại dữ liệu và cho phép một ORDBMS được thực hiện mà không cần dự đoán các loại và phương thức mà người dùng có thể muốn thêm. Ví dụ, hình ảnh bổ sung thêm các số nguyên và phủ có nhất được xử lý đồng nhất của hệ thống, với sự phân biệt đáng kể duy nhất được rằng mã khác nhau trong gọi cho hai hoạt động và đối tượng khác nhau đã nhập được dự kiến sẽ được trả lại từ mã đó.
  1. **Định nghĩa phướng thức của kiểu dữ liệu trừu tượng**
* Ở mức tối thiểu, đối với từng loại nguyên tử mới, một người sử dụng phải xác định các phương pháp cho phép các DBMS để đọc và để tạo ra các đối tượng của loại hình này và để tính toán lượng lưu trữ cần thiết để giữ đối tượng.Người dùng tạo ra một loại mới nguyên tử phải đăng ký các phương thức sau đây với các DBMS:
  + **Size:** Trả về số byte cần thiết dùng để lưu trữ đối tượng này hoặc một giá trị động đặc biệt, nếu như nó có kích thước động.
  + **Import:** Là phương thức cho phép tạo ra một đối tượng mới dữ liệu đầu vào.
  + **Export:** Ánh xạ các đối tượng để thích hợp cho việc in ấn, hoặc sử dụng trong các chương trình ứng dụng.
* Để đăng ký một phương pháp mới cho một loại nguyên tử, người sử dụng phải viết mã cho phương pháp này và sau đó thông báo cho hệ thống cơ sở dữ liệu về phương pháp.Các mã được viết phụ thuộc vào các ngôn ngữ được hỗ trợ bởi các DBMS, và có thể là hệ điều hành trong câu hỏi.Ví dụ, ORDBMSS có thể xử lý mã Java trong hệ điều hành Linux.Trong trường hợp này, các phương pháp mã phải được viết bằng Java và biên dịch vào một tập tin bytecode Java được lưu trữ trong một hệ thống tập tin Linux.Sau đó, một phong cách phương pháp lệnh SQL đăng ký được trao cho ORDBMS để nó nhận ra phương pháp mới:



* Các khai báo này là những điểm đặc trưng của phương thức: các loại ADT liên quan, kiểu trả về, vị trị của đoạn mã này. Khi phương thức này được đăng ký, DBMS sử dụng một máy ảo Java để thực thi mã:



* Những khai báo định nghĩa kiểu dữ liệu nguyên tử do người dùng định nghĩa để sử dụng trong lược đồ của Dinky



1. **Cách thức và cấu trúc của kiểu dữ liệu trừu tượng**

* Các loại nguyên tử và các loại do người dùng định nghĩa có thể kết hợp để mô tả cấu trúc phức tạp hơn bằng cách sử dụng cac loại hàm dựng. Ví dụ dòng thứ 6 của hình 25.1 xác định cột ngôi sao của loại setof(text), mối mục trong cột đó là một tập cac chuỗi văn bản, đại diện cho các ngôi sao trong bộ phim. Các cú pháp setof là một ví dụ của một hàm dựng. Các hàm dựng phổ biến khác bao gồm:
  + **ROW**(n­1 t1, …, nn tn): Một loại trình bày cho một dòng, hoặc một bộ của n cột dữ liệu với n1,…,nn thuộc loại t1,…,tn tương ứng.
  + **Listof(base):** Một kiểu đại diện cho một list các phần tử cơ sở.
  + **ARRAY(basse):** Một kiểu đại diện cho một mảng các phần tử cơ sở.
  + **setof(base):** Một kiểu đại diện cho một tập (set) các phần tử cơ sở **.** Một set thì không có 2 phần tử trùng nhau.
  + **bagof(base):** Một kiểu đại diện cho một bag bao gồm các phần tử cơ sở hoặc các set.
* Đánh giá đầy đủ khả năng của các hàm dựng, ví dụ ARRAY (ROW (age: integer, sal: integer)). Các loại được xác định bằng cách sử dụng các hàm dựng của các kiểu được gọi là kiểu có cấu trúc. Ngoài ra có thể sử dụng listof, ARRAY, bagof, hoặc setof như là các hàm dựng sau cùng đôi khí được gọi là tập hợp các loại, hoặc dữ liệu với số lượng lớn.
  1. **Thao tác ở kiểu dữ liệu có cấu trúc**
* DBMS cung cấp phương thức được xây dựng cho các loại hỗ trợ thông qua hàm dựng loại. Những phương pháp này tương tự như trong hoạt động xây dựng như phép cộng và phép nhân đối với các loại nguyên tử chẳng hạn như số nguyên.Trong phần này chúng tôi trình bày các phương pháp để xây dựng loại hình khác nhau và minh họa các truy vấn SQL có thể tạo ra và thao tác giá trị với các loại cấu trúc.
* **ROWS:** Với mỗi item có loại là ROW(n1 t1, …, nn tn) phương pháp lấy các field cho phép chúng ta truy cập đến một nk sử dụng ký hiệu i.nk truyền thống các hàm dựng được lồng trong một kiểu định nghĩa, dấu (.) có thể được lồng vào nhau để truy cập vào các field của hàng lồng nhau.
* Ký hiệu các chấm lồng nhau này thường được gọi là một biểu thức đường dẫn bởi vì nó mô tả một đường thông qua các cấu trúc lông nhau.
* **Sets và multisets:** Một tập đối tượng có thể so sánh được với nhau sử dụng các toán tử truyền thống của tập (các toán tử tập hợp). Một phần tử trong tập setof(foo) có thể so sánh với các tập khác sử dụng phương thức thuộc tập phương thức. Hai tập đốio tượng (các phần tử có cùng kiểu) có thể tổ hợp để tạo nên đối tượng mới sử dựng (giao, hợp, hiệu).
* **Lists:** Danh sách bình thường sẽ bao gồm head, trả về yếu tố đầu tiên là tail, trả về danh sách thu được bằng cách bỏ các yếu tố đầu tiên, thêm vào trước, trong đó có một phần tử và chèn nó vào như là các yếu tố đầu tiên trong một list và thêm vào một danh sách khác.
* **Arrays:** Mảng loại hỗ trợ một phương thức ‘mảng chỉ số’ cho phép truy cập vào các mảng tại một thiết lập cụ thể. Có có dấu ngoặc vuông phía sau dùng để truy cập đến phần tử nào.

1. **Xem xét và nhận dạng đối tượng.**

* Trong HQT CSDL hướng đối tượng, đối tượng dữ liệu có thể dc gán 1 giá trị object identifier (oid), xác định duy nhất trong dữ liệu trong suốt thời gian sống của nó. DBMS chịu trách nhiệm tạo ra giá trị này và phải đảm bảo nó luôn xd dc 1 đối tượng duy nhất trong suốt thời gian sống. ở 1 số hqt tất cả các bộ dc lưu trong bất cứ bảng nào thì cũng đều là đối tượng và he qt sẽ tự đống đánh cho chúng 1 giá trị oid. ; 1 số hqt khách, ng dùng có thể chỉ ra các bảng nào có oid.
* 1 oid của đối tượng dc dùng để chỉ tới đối tượng đó mà có dữ liệu dc lưu ở nơi khác. Nó như 1 giá con trỏ tham chiếu có kiểu là loại mà nó tham chiếu đến. ref(base) là 1 hàm dựng kiểu là con trỏ trỏ đến object kiểu base. Ref có thể dc chèn vào trong hàm dựng khác.
* So sánh url với oid:

|  |  |
| --- | --- |
| **Oid** | **URL** |
| Xác định duy nhất 1 đối tượng trong suốt thời gian | Tài nguyên web dc trỏ bởi 1 URL có thể thay đổi theo thời gian |
| Đơn giản, nó không mang ý nghĩa vật lý (không chứa thông tin vật lý của đối tượng mà nó trỏ đến) - dẫn đến nếu thay đổi địa chỉ lưu trữ của đối tượng thì ta khộng phải sửa oid này. | Chứa cả địa chỉ network và tên file system, khi tài nguyên mà url đó trỏ đến thay đổi địa chỉ lưu hoặc địa chỉ network, đường dẫn url đến tài nguyên đó bị sai hoặc phải yêu cầu cơ chế forwarding. |
| Oid được phát sinh tự động bởi DBMS | Người dùng tạo ra |

* 1. **Khái niệm về Equality**
* Sự khác biệt giửa các loại tham chiếu với các loại cấu trúc tham chiếu tự do (reference-free structured types) làm nảy sinh 1 vấn đề khác: định nghĩa đẳng thức. 2 object có cùng loại được xem là deep equal khi và chỉ khi:
* Các Đối tượng là kiểu nguyên tử và có cùng giá trị.
* Các đối tượng là kiểu tham chiếu và
* Đối tượng là loại cấu trúc, toán tử deep equal đúng với tất cả thành phần tương ứng của 2 đối tượng.
* 2 đối tượng có cùng loại tham chiếu được định nghĩa là shallow equal nếu cả 2 chỉ đến cùng đối tượng. định nghĩa shallow equality có thể dc mở rộng cho đối tượng có loại tuỳ ý bằng cách lấy định nghĩa của deep equality và thay deep equals thành shallow equals.
* 2 đối tượng thoả shallow equal thì thoả deep equal, ngược lại không đúng
  1. **Bất tham chiếu đến các loại tham chiếu**
* Loại tham chiếu ref(foo) không giống như loại foo nó trỏ đến. để truy cập foo dc tham chiếu, có phương thức deref() cung cấp song song với hàm dựng loại ref. giống như trong ng lập trình def() = \*a. ví dụ cần lấy 1 bộ trong bảng Nowhowing, lấy giá trị name của theater: Noshowing.deref(theater).name. vì tham chiếu đến các loại bộ thì giống nhau, nên 1 số hqt cung cập 1 toán tử mũi tên kết hợp với hậu tố của toán tử dereference operator . đó là: Nowshowing.theater🡪name.

1. **Sự kế thừa và phân cấp các loại.**
   1. **Định nghĩa các loại với sự kế thừa**

* Trong csdl Dinky, ta mô hình hoá rạp chiếu phim thành kiểu theater\_t. nhưng Dinky muốn kinh doanh thêm loại rạp chiếu-café, rapchiếu-café sẽ cung cấp pizza và các laoị thức ăn uống khi xem phim. Vì vậy trong db cần hỗ trợ lưu trữ các thông tin này, theater-café giống như theater nhưng có thêm các loại thuộc tính thực đơn. T dùng kế thừa, đoạn lệnh DDL sau diễn tả điều đó:
* Create type theatercafe\_t under theater\_t (menu text):
* Trong lệnh này tạo ra theatcafe và tất cả các thuộc tính và phương thức của theater\_t đều dc theatercafe dùng lại nhưng điểu ngược lại thì k. ta nói theatercafe\_t kế thừa các thuộc tính của theater\_t.
* Lưu ý cơ chế kế thừa k phải là macro làm ngắn gọn lệnh create. Nó tạo ra quan hệ rõ ràng trong dữ liệu giữa subtype (theatercafe\_t) và supertype (theater\_t): 1 object của subtype ccũng dc xét là 1 object của supertype.
* The substitution Principle: cho 1 supertype A và 1 subtype B, luôn luôn có thể thay thế 1 đối tượng loại B thành 1 biểu diễn hợp lý được viết bởi object laoị A, ngoại trừ việc tạo ra type errors.
* Note: kế thừa cũng có thể dc dùng cho kiểu nguyên tử, cho cả kiểu row.
  1. **Tính ràng buộc của những phương thức**
* Ta có thể viết lại phương thức của supertype trong subtype. Viết 1 phương thức mới với phương thức cũ gọi là overloading tên phương thức.
* Do việc overloading này, hệ thống phải hiểu phương thức nào dc dành cho 1 biểu diễn cụ thể. Vd, khi sys cần gọi pthức display() trong object kiểu jpeg\_image\_t, nó dùng phương thức display riêng biệt. khi cần vọi display cảu image\_t (k phải loại subtyped) thì n1o gọi pthuc display chuẩn. quá trình (process) quyết định phương thức nào dc gọi lên dc gọi là liên kết phương thức với đối tượng (binding to the object). Binding dc thực hiện khi 1 biểu thức dc phân tách (early binding), 1 số trường hợp
* Early binding : Khi quá trình binding được thực thi trước khi chương trình chạy(by compiler & linker). Vấn đề chính của early binding là compiler không biết chính xác method nào để gọi khi chỉ có một tham khảo tới đối tượng Instrument. C compilers chỉ có một lời gọi method là early binding.
* Late binding : ra đời để giải quyết vấn đề của early binding, nó sẽ binding ở thời điểm chạy, nó có cơ chế xác định đối tượng nào ngay tại thời điểm gọi và gọi method của đối tượng đó.
* Late binding là cái core của Tính đa hình (Polymorphism). Nhờ late binding mà tính đa hình trở thành công cụ mạnh mẽ trong Lập trình hướng đối tượng.
  1. **Tập các hệ thống phân cấp, loại mở rộng, và truy vấn**
* Có 1 chút khác biệt giữa kế thừa trong topic này với ngôn ngữ lập trình hướng đối tượng
* Tuy nhiên vì ORDBMS cung ấp ngôn gnữ truy vấn trên tập dữ liệu dạng bảng, nên các cơ chế từ các ngôn gnữ lập trình dc tăng cường trong CSDL đối tượng để xử lý với bảng và truy vấn tốt. một cách riêng, trong ob-relational sys ta ca thể định nghĩa 1 bảng chứa cá các đối tượng có kiểu cụ thể, như bảng theater trong lược đồ Dinky. Cho 1 subtype như theater-café, nếu mốn tạo bảng theater-café khác để lưu thong tin về theater cafes. Nhưng khi viết 1 truy vấn trên bảng theaters, đôi khi muốn 1 câu truy vấn tượng tự trong bảng theate-cafes, sau hết, nếu ta chiếu các cột, 1 ví dụ của bảng theater-café có thể dc quan tam như 1 ví dụ của bảng.
* Không phải yêu cầu ng dùng chỉ rõ 1 câu query riêng lẽ cho mỗi bảng, ta có thể thông báo cho hệ thống là 1 bảng mới thuộc subtype dc xem như 1 phần của bảng supertype:
* Create table theater\_cafes of type theater\_cafe\_t under theaters;
* Truy vấn này nói cho hệ thống biết các lệnh query trên bảng theaters phải dc chạy trong các bộ trên cả bảng theaters và theater\_cafes. Nếu subtype definition gọi phương thức overlaoding, thì late-binding dc dùng để chắc ràng phước thức đó dc gọi cho các bộ hợp lý.
* Nhìn chung, mệnh để clause dc dùng để tạo 1 cây tuỳ ý của bảng gọi 1 tập phân cấp. lệnh query trên 1 bảng riêng biệt T trong phân cấp dc chạy trên tất cả bộ trong T và các node con của nó. Đôi khi, 1 ng dùng có thể muốn truy vấn chạy chỉ trên T, và k trên các con, thì thêm cú pháp only.
* Một số hệ thống tự động tạo các bảng đạc biệt cho mỗi kiểu, mà chứa tham chiếu đến các instance của type mà tồn tại trong db. Các bảng này dc gọi là type extents và cho phép truy vấn trên tất cả đối tượng của type cho, k kể đến nơi các đối tượng cư trú trong db. Type extents hình thành 1 collection hierarchy mà song song với kiểu phân cấp.

1. **Xem xét làm thế nào để tận dụng lợi thế của khái niệm hướng đối tượng mới để thiết kế cơ sở dữ liệu ORDBMS**

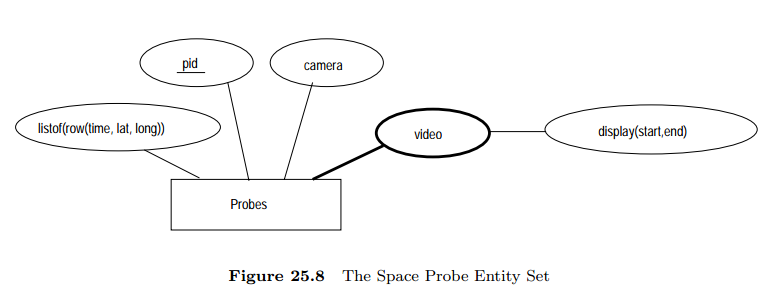
* Sự đa dạng về trạng thái của loại dữ liệu trong ORDBMS đã mang lại cho các nhà thiết kế cơ sở dữ liệu nhiều cơ hội cho việc thiết kế hiệu quả.
  1. **Loại có cấu trúc và loại trừu tượng**
* Ví dụ đầu tiên của chúng ta liên quan đến tàu thăm dò vũ trụ (probe), với mỗi probe sẽ có một loại camera xác định, sẽ ghi lại một đoạn video.
* Mỗi video sẽ được thu lại trong một khoảng thời gian cụ thể và ở vị trí với vĩ độ và kinh độ xác định.
* Do đó thông tin liên quan đến một tàu thăm dò sẽ có 3 phần chủ yếu: probe\_id, video stream, <time, location>
* **Thiết kế CSDL với RDBMS:**
  + Với RDBMS, chúng ta phải lưu video stream với kiểu dữ liệu BLOB và mỗi vị trí tuần tự chúng ta lưu thành một bộ trong bảng.
    - Probe (**pid: integer, time: timestamp, latitude: real, longitude: real**, camera: string, video: BLOB)
    - Vơi một bảng Probes, và có một dòng cho một probe.
    - Và chúng ta có 2 phụ thuộc hàm
      * PTLN🡪 CV
      * P🡪CV
    - Và với 2 phụ thuộc hàm này cấu trúc của chúng ta không đạt dạng chuẩn BCNF, vì vậy tách ra làm 2 bảng như sau:
    - Probe\_Loc (**pid: integer, time: timestamp, latitude: real, longitude: real**)
    - Probe (**pid: integer,** camera: string, video: BLOB)
  + Thiết kế này thì tốt khi được thực hiện trên RDBMS. Tuy nhiên, có nhiều hạn chế:
    - Đối với câu truy vấn “Với probe 10, hiển thị đoạn video được thu từ 1:10pm đến 1:15pm vào ngày 10, 1996” 🡪 Chúng ta phải truy vấn toàn bộ các đối tượng video của probe 10 được thu trong hàng giờ để ra được kết quả trên.
    - Sự thật mà mỗi probe có một dòng các location không rõ nghĩa, và thông tin nối tiếp liên quan tới probe bị phân tán qua các bộ.
    - Chúng ta bắt buộc phải chia thông tin liên tiếp của một probe ra. Điều này không hiệu quả đối với các truy vấn có liên quan đến tất cả thông tin của probe. Ví dụ “Với mỗi probe, hãy in thời gian sớm nhất được thu lại và loại camera”.
* **Thiết kế CSDL với ORDBMS:**
  + Một ORDBMS hỗ trợ nhiều giải pháp tốt hơn.
    - Thứ nhất, chúng ta có thể lưu trữ video dưới dạng một đối tượng ADT và viết các phương thức để chụp hay xử lý đặc biệt như chúng ta mong muốn.
    - Thứ hai, bởi vì chúng ta có thể lưu trữ với các kiểu dữ liệu có cấu trúc như là list, chúng ta sẽ dùng list để lưu trữ vị trí của probe trên một bộ đơn cùng với đoạn video stream.
    - Với ORDBMS ta có một quan hệ duy nhất như sau:
      * Probes\_AllInfo (pid: integer, locseq: location\_seq, camera: string, video: mpeg\_stream)
      * Và khi viết theo cách này chúng ta có thể có các câu truy vấn sau:

**SELECT** display (P.video, 1:10pm, May 10 1996, 1:15pm, May 10 1996)

**FROM** Probes\_Allinfo P

**WHERE** P.id = 1

* 1. **Định danh đối tượng**
* Chúng ta sẽ thảo luận hệ quả của việc sử dụng reference hoặc oid. Việc sử dụng oid đặc biệt có ý nghĩa khi các đối tượng có kích thước lơn, bởi vì các đối tượng đó là một cấu trúc dữ liệu hoặc một đối tượng lớn như một hình ảnh.
* Giải thích các thao tác thường sử dụng:
  + Xóa: các đối tượng tham chiếu bị ảnh hưởng, thường nó sẽ chuyển thành giá trị null
  + Cập nhật: Thay đổi trên tất cả các đối tượng tham chiếu -> không cần phải cập nhật thủ công cho từng đối tượng như đối tượng không tham chiếu
  + Chia sẻ với sao chép: một đối tượng có thể bị tham chiếu bởi nhiều đối tượng khác, do đó một sự cập nhật sẽ được phản chiếu ở nhiều nơi. Trong khi đó, đối tượng cơ sở dữ liệu quan hệ đòi hỏi phải cập nhật ở từng bản copy
  + Chi phí lưu trữ: lưu trữ các bản copy của đối tượng sẽ tốn nhiều không gian hơn lưu một giá trị và tham chiếu đến nó. Điều này có thể ảnh hưởng đến việc sử dụng đĩa và quản lí bộ đệm (đặc biệt khi nhiều bản copy được truy vấn cùng lúc)
  + Phân nhóm: Những phần nhỏ trong đối tượng có cấu trúc thường lưu trữ gần nhau trên đĩa. Do đó các đối tượng nếu có tham chiếu quá xa sẽ khiến đĩa truy xuất lâu hơn, vì vậy đối tượng có cấu trúc sẽ hiệu quả hơn các loại tham chiếu nếu chúng thường được truy vấn theo tổng số
* Nhưng cần lưu ý, trong thiết kế cơ sở dữ liệu, sự lựa chọn giữa kiểu cấu trúc và kiểu tham chiếu cần phải xem xét dựa trên chi phí lưu trữ, vấn đề phân nhóm, và ảnh hưởng của việc cập nhật
* **Dùng định danh đối tượng hay Foreign Keys:**
  + Một vấn đề nữa đặt ra là nên sử dụng định danh đối tượng (Oid) hay khóa ngoại.
  + Oid: một oid có thể trỏ đến một đối tượng được lưu ở bất cứ đâu trong cơ sở dữ liệu.
  + Khóa ngoại: là một ràng buộc trỏ đến đối tượng trong một quan hệ tham chiếu cụ thể.
  + Do đặc điểm trên mà DBMS dễ dàng cung cấp nhiều hỗ trợ cho ràng buộc tham chiếu hơn là con trỏ Oid. Chú ý khi xóa.
  + Chúng ta nên cẩn trọng khi sử dụng oid, và nên sử dụng khóa ngoại thay thế bất cứ khi nào có thể.
  1. **Mở rộng mô hình thực thể kết hợp (ER)**
* Với mô hình ER mô tả ở chương 2 không đầy đủ cho thiết kế ORDBMS. Chúng ta sẽ sử dụng mô hình ER mở rộng và hỗ trợ các giá trị có cấu trúc. Phân biệt giữa một thực thể và oid của tối tượng , và mô hình cho phép mô tả các phương thức của đối tượng:

****

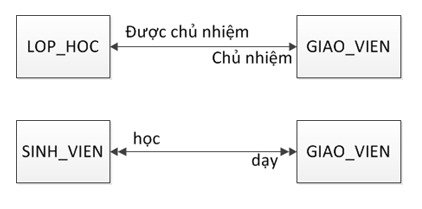
* Cuối cùng, chúng ta cần phải ghi chú các mở rộng có ý nghĩa của mô hình ER là yêu cầu hỗ trợ từ nhà thiết kế nằm từ những tập lồng nhau. Ví dụ như về cách lưu trữ vị trí của một phi thuyền.
  1. **Sử dụng tập Nested(Lồng nhau)**
* **Tập hợp lồng nhau cung cấp sức mạnh mô hình hóa lớn, tuy nhien cũng tạo ra nhiều trở ngại trong quyết định thiết kế:**
  + Probes1 (pid: integer, locseq: location\_seq)
  + Câu truy vấn: Với mỗi probe, hãy xuất ra thời gian sớm nhất được ghi lại và loại camera
  + Probes2 (pid: integer, time:timestamp, latitude: real, longitude: real)
  + Câu truy vấn: Tìm thời điểm sớm nhất được ghi ở latitude =5, và longitude =10
  + Sự chọn lựa trong thiết kế lược đồ phải phụ thuộc vào lượng workload mong muốn
  + Tập hợp lồng nhau thích hợp ở trường hợp cụ thể do đó khi sử dụng cần cẩn trọng, tránh lạm dụng

1. **Thảo luận một số thách thức mới đặt ra bởi hệ thống quan hệ đối tượng.**

* Cơ sở dữ liệu hướng đối tượng ra đời là đã giải quyết được một số khó khăn nhưng đồng thời cũng mang theo các thách thức:
  1. **Lưu trữ và truy cập phương thức**
* Dữ liệu người dùng tự định nghĩa ADT khá lớn. Trong một số trường hợp, nó sẽ lớn hơn page, ví dụ như BLOB. Khi ADT lớn hơn page memory, sư dụng các disk-base pointer được duy trì từ các bộ đến đối tượng mà nó chứa.
* Solution: các con trỏ dựa trên đĩa (disk-based pointer) được duy trì từ các bộ đến đối tượng mà nó chứa
* Đối tượng cấu trúc không giống ADT, có kích thước biến thiên trong vòng đời của cơ sở dữ liệu. Ví dụ: đối với thuộc tính star, qua nhiều năm khi các diễn viên nhỏ trở nên nổi tiếng, Dinky muốn quảng bá họ đã từng tham gia bộ film đó thì sẽ đòi hỏi thểm họ vào thuộc tính star. Bởi vì các thuộc tính này có thể tăng tùy ý, do đó đòi hỏi cơ chế sắp xếp đĩa linh hoạt.
* Đối với loại dữ liệu mảng, thông thường sẽ được lưu trữ liên tiếp trên đĩa theo kiểu từng dòng. Tuy nhiên đối với các truy vấn chỉ truy xuất một phần của mảng, điều này sẽ đòi hỏi chi phí I/O cao. Vì vậy để giảm chi phí mảng thường được chia nhỏ ra thành những khối nhỏ (chunk) và các chunk không cần lưu theo từng dòng hay từng cột
  1. **Chỉ mục trên loại dữ liệu mới**
* Cấu trúc chỉ mục RDBMS chuẩn chỉ hỗ trợ điều kiện bằng (cây B+ và hash index) và điều kiện phạm vi (cây B+).
* Nhiều cấu trúc chỉ mục đã được đưa ra cho từng loại ứng dụng như: vẽ bản đồ, nghiên cứu di truyền, lưu trữ đa phương tiên, web search … nhưng một ORDBMS không thể cài đặt tất cả các chỉ mục được sáng tạo ra. Do đó tập cấu trúc chỉ mục ORDBMS phải linh hoạt đối với người dùng
* Giải pháp thứ nhất: xây dựng một interface phương thức truy vấn cho phép người dùng cài đặt các cấu trúc chỉ mục bên ngoài DBMS. Trong đó chỉ mục và dữ liệu có thể được lưu trong file system và DBMS chỉ đưa ra các yêu cầu lặp đi lặp lại là open, next, close đến đoạn mã chỉ mục bên ngoài của người dùng. Tuy nhiên phần dữ liệu ở chỉ mục bên ngoài không được bảo vệ bởi DBMS cho tính đồng thời và phục hồi.
* Giải pháp thứ hai: xây dựng một template cấu trúc chỉ mục chung có thể đáp ứng tổng quát hầu hết các cấu trúc chỉ mục mà người dùng tạo ra. Bởi vì cấu trúc này sẽ được cài đặt trong DBMS nên nó có thể khắc phục vấn đề đồng thời và phục hồi. => Generalised Search Tree (GiST) là một cấu trúc như vậy. Nó là một template dựa trên cây B+, cho phép hầu hết các cấu trúc chỉ mục có thể cài đặt với vài dòng ADT code do người dùng định nghĩa
  1. **Xử lý truy vấn**
* Hầu hết các ORDBMS cho phép người dùng định nghĩa hàm kết hợp mới. Để định nghĩa một hàm kết hợp, người dùng phải cài đặt 3 phương thức sau:
  + Initialize: khởi tạo trạng thái bên trong cho sự kết hợp
  + Iterate: cập nhật trạng thái cho mỗi bộ nhìn thấy
  + Terminate: tính toán kết quả kết hợp dựa trên trạng thái cuối cùng và sau đó dọn dẹp
* Ví dụ: tìm giá trị cao thứ hai của một thuộc tính:
  + Initialize: xác định lưu trữ cho hai giá trị cao nhất
  + Iterate: so sánh giá trị của bộ hiện tại với hai giá trị cao nhất và cập nhật hai giá trị cao nhất lại nếu cần
  + Terminate: xóa lưu trữ của hai giá trị và trả về bản copy của giá trị cao thứ 2
  1. **Tối ưu truy vấn**
* **Method caching**
  + Đối với phương thức được sử dụng nhiều lần, chúng ta có thể duy trì cache của phương thức input và những output trùng khớp như một bảng trong cơ sở dữ liệu. Sau đó để tìm kiếm giá trị của phương thức với một input cụ thể, chúng ta có thể kết bộ input với cache table. Đối với một truy vấn đơn, có thể tránh gọi phương thức hai lần trên cùng giá trị của một cột bằng cách sắp xếp bảng trên cột đó hoặc sử dụng lược đồ dựa trên hash giống như đã dùng trong chương 12.7. Chúng ta có thể kết hợp cả hai phương pháp trên với nhau
* **Pointer Swizzling**
  + Trong một vài ứng dụng, đối tượng đưa vào bộ nhớ và truy cập thường xuyên thông qua oid. Một số hệ thống duy trì một bảng oid trong bộ nhớ. Khi đối tượng O được đưa vào bộ nhớ, chúng kiểm tra oid trong đối tượng O và thay thế oid của đối tượng trong bộ nhớ bằng con trỏ đến các đối tượng đó. Kỹ thuật này được gọi là pointer swizzling. Tuy nhiên khi một đối tượng page out, thì tham chiếu trong bộ nhớ đến nó cũng không còn tác dụng nữa và sẽ bị thay thế với oid của nó
  1. **Tinh chỉnh truy vấn**
* **Khi chỉ mục mới được thêm vào hệ thống, bộ tinh chỉnh phải được thông báo về sự có mặt của nó và chi phí truy vấn.:**
  + Bộ tinh chỉnh phải biết được
  + Điều kiện gì trong mệnh đề WHERE trùng khớp với chỉ mục
  + Chi phí nạp một bộ đối với chỉ mục là bao nhiêu
  + Dựa vào các thông tin đó mà bộ tinh chỉnh có thể sử dụng bất kì cấu trúc chỉ mục nào trong việc xây dựng kế hoạch thực thi.
* **Hệ số giảm và tính toán chi phí cho phương thức ADT**
  + Việc tính toán hệ số giảm đối với điều kiện do người dùng định nghĩa là một vấn đề khó và vẫn còn đang được nghiên cứu.
  + Cách tiếp cận thông thường hiện nay là để người dùng tự tính toán, người dùng một khi có thể tạo ra một phương thức thì cũng có tạo một hàm bổ trọ việc tính toán hệ số giảm của phương thức đó. Nếu như hàm tính toán đó không được đăng kí thì bộ tinh chỉnh sử dụng một số tùy ý như 1/10.
  + Người dùng có thể tính toán bằng chi phí I/O trong hệ thống. Ngoài ra ORDBMS có thể thực thi phương thức trên nhiều đối tượng có kích thước khác nhau để cố gắng tính toán chi phí một cách tự động, nhưng cách này vẫn chưa được xem xét chi tiết.
* **Tinh chỉnh các phép chọn chi phí cao**
  + Trong cơ sở dữ liệu quan hệ, phép chọn được xem như một thao tác không tốn thời gian. Ví dụ emp.salary không cần I/O và chỉ tốn một vài chu kỳ CPU, nhưng đối với is\_herbert (Frames.image) thì khá đắt đỏ vì nó phải nạp một lượng lớn đối tượng từ đĩa vào và xử lí chúng trong bộ bằng nhiều cách phức tạp.
  + Do đó mà bộ tinh chỉnh DBMS phải xem xét cẩn thận để sắp xếp thứ tự các điều kiện chọn.
  + Ví dụ Frames.frameno < 100 is\_herbert (Frames.image)
  + Một cách tổng quát, thứ tự tốt nhất trong phép chọn là một hàm của chi phí và hệ số giảm, rank = (hệ số giảm -1 )/ cost. Nếu rank trong một truy vấn nhiều bảng lớn thì nên hoãn phép chọn lại sau khi đã thực hiện phép kết. Điều này trái ngược với quan niệm thực hiện phép chọn càng sớm càng tốt trong cơ sở dữ liệu quan hệ

1. **Thảo luận về ODL và OQL, các tiểu chuẩn cho OODBMSs.**

* Trong chương này chúng ta định nghĩa một OODBMS như một ngôn ngữ lập trình với sự hỗ trợ của các đối tượng lâu dài. Trong khi định nghĩa này phản ánh nguồn gốc của OODBMSs chính xác, và đến một mức độ nhất định tập trung thực hiện OODBMSs, thực tế là OODBMS hỗ trợ một tập ngôn ngữ truy vấn truy trên dữ liệu. Và một tiêu chuẩn được phát triển bởi ODMG và được gọi là OQL.
* OQL thì tương tự SQL, với cú pháp SELECT-FROM-WHERE (và GROUP BY, HAVING, ORDER BY cũng được hỗ trợ) và dựa trên sự mở rộng của SQL:1999.
* Ngoài ra dựa trên DDL (ngôn ngữ định nghĩa đối tượng) cho OODBMSs (Object Data Language, or ODL) thì ODL sẽ được thêm vào OODBMS, để định nghĩa đối tượng.
  1. **Mô hình dữ liệu ODMG và ODL**
* Mô hình dữ liệu ODMG là cơ bản cho OODBMS, giống như mô hình CSDL quan hệ là cơ bản cho RDBMS. Một CSDL chứa các tập của đối tượng, mà tương tự như các thực thể trong mô hình ER. Mỗi đối tượng đều có một oid duy nhất, và CSDL chứa các tập đối tượng này với các thuộc tính, mỗi tập như vậy được gọi là class.
* Các thuộc tính của class có thì theo quy định sử dụng ODL và gồm 3 loại sau: thuộc tính, mối quan hệ, và phương thức. Các định nghĩa của các loại này tương đương các định nghĩa bên lập trình hướng đối tượng.
* Mối kết hợp: có nhiều loại một-một, một-nhiều, nhiều-nhiều, ví dụ:

****

* Phương thức là các hoạt động của đối tượng, không giống với store procedure trong mô hình quan hệ.
* **ODL:**
  + Hỗ trợ các khái niệm ngữ nghĩa ngữ nghĩa của OODBMS.
  + Không phụ thuộc vào ngôn ngữ lập trình.
  + Dùng để tạo các đặc điểm của đối tượng.
  + Dùng để tạo các đặc điểm của đối tượng.
  + Không dùng để xử lý cơ sở dữ liệu
  1. **OQL**
* Ngôn ngữ truy vấn OQL của ODMG được cố tình thiết kế để có cú pháp tương tự SQL, để làm cho nó dễ dàng cho người sử dụng quen thuộc với SQL để tìm hiểu ODL. Cùng xem qua ODL bằng các ví dụ:
  + **SELECT *mname*: M.movieName, *tname*: T.theaterName**

**FROM Movies M, M.shownAt T**

**WHERE T.numshowing() > 1**

* **SELECT T.ticketPrice,*avgNum*: AVG(SELECT P.T.numshowing() FROM *partition* P)**

**FROM Theaters T**

**GROUP BY T.ticketPrice**

1. **Trình bày so sánh ngắn gọn của ORDBMSs và OODBMs**
   1. **So sánh RDBMS và ORDBMS**

* RDBMS không hỗ trợ phần mở rộng của ORDBMS.
* RDBMS dễ sử dụng hơn vì nó có ít tính năng hơn.
* RDBMS ít linh hoạt hơn ORDBMS.
  1. **So sánh OODBMS và ORDBMS: Sự giống nhau**
* Hỗ trợ các kiểu có cấu trúc, kiểu đối tượng, ADTS và kế thừa.
* Hỗ trợ ngôn ngữ truy vấn để thao tác trên kiểu collection:
  + ORDBMS: phần mở rộng của SQL.
  + OODBMS: ODL/OQL.
  + ORDBMS thêm các tính năng của OODBMS vào RDBMS.
  + OODBMS phát triển ngôn ngữ truy vấn dựa trên ngôn ngữ truy vấn quan hệ.
  1. **So sánh OODBMS và ORDBMS: Sự khác nhau**
* OODBMS thêm ngôn ngữ lập trình vào DBMS, trong khi ORDBMS thêm các kiểu dữ liệu phong phú vào RDBMS.
* OODBMS có mục tiêu là tích hợp liền mạch với ngôn ngữ lập trình, trong khi đây không phải là mục tiêu quan trọng của ORDBMS.\
* OQL không được hỗ trợ hiệu quả trong hầu hết OODBMS, ngược lại chúng là trung tâm của ORDBMS.
* OODBMS đã được tối ưu hóa để hỗ trợ cho các ứng dụng HĐT và các ngôn ngữ OO.
* ORDBMSs được hỗ trợ bở hầu hết các HQTCSDL hiện nay.