

Trường Đại học CNTT – Khoa Hệ thống thông tin

BÀI 4: CƠ SỞ DỮ LIỆU ĐA CHIỀU VÀ CÁC MÔ HÌNH BIỂU DIỄN

Giảng viên: ThS. Nguyễn Thị Kim Phụng
ThS. Hồ Trung Thành

Nội dung

- ▶ Khái niệm
- ▶ Cơ sở dữ liệu đa chiều
- ▶ Bảng chiều
- ▶ Bảng sự kiện
- ▶ Khối
- ▶ Các mô hình lưu trữ dữ liệu đa chiều
(Lướt đồ hình sao, Lướt đồ bông tuyết)

Cơ sở dữ liệu đa chiều

- ▶ Các khái niệm
- ▶ Mô hình chiều
- ▶ Đặc điểm
- ▶ Các mô hình lưu trữ dữ liệu đa chiều

Khái niệm

- ▶ Phù hợp và gần gũi, trực quan hơn với người làm việc với dữ liệu.
- ▶ Cơ sở dữ liệu đa chiều đã phát triển và đi vào công nghệ cơ sở dữ liệu như là một sự lựa chọn cho những ứng dụng phân tích dữ liệu.
- ▶ Ứng dụng dạng này được chấp nhận rộng rãi ngày nay là OLAP.
- ▶ Cơ sở dữ liệu đa chiều hết sức cần thiết như:
 - Tạo điều kiện thuận lợi cho việc truy cập cơ sở dữ liệu linh hoạt.
 - Thực hiện và phân tích với khối lượng dữ liệu lớn, dữ liệu phức tạp.

Khái niệm (2)

- ▶ Theo W.H. Inmon [9]: “Cơ sở dữ liệu đa chiều là hệ thống phần mềm máy tính được thiết kế cho phép có khả năng và phù hợp với việc lưu trữ và truy xuất một khối lượng lớn dữ liệu:
 - ❑ Có quan hệ với nhau.
 - ❑ Lưu trữ, xem xét và phân tích từ nhiều khía cạnh khác nhau. Những khía cạnh đó được gọi là các chiều”.

Khái niệm (3)

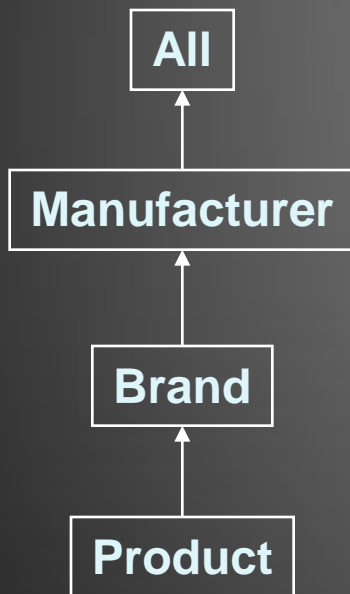
- ▶ Theo Oracle: “Cơ sở dữ liệu đa chiều là một dạng của cơ sở dữ liệu được tối ưu hoá nhằm cho kho dữ liệu và những ứng dụng xử lý phân tích trực tuyến. Cơ sở dữ liệu đa chiều thường xuyên được tạo ra bằng cách sử dụng dữ liệu đầu vào từ những cơ sở dữ liệu quan hệ đã tồn tại”.

Mô hình dữ liệu đa chiều (1)

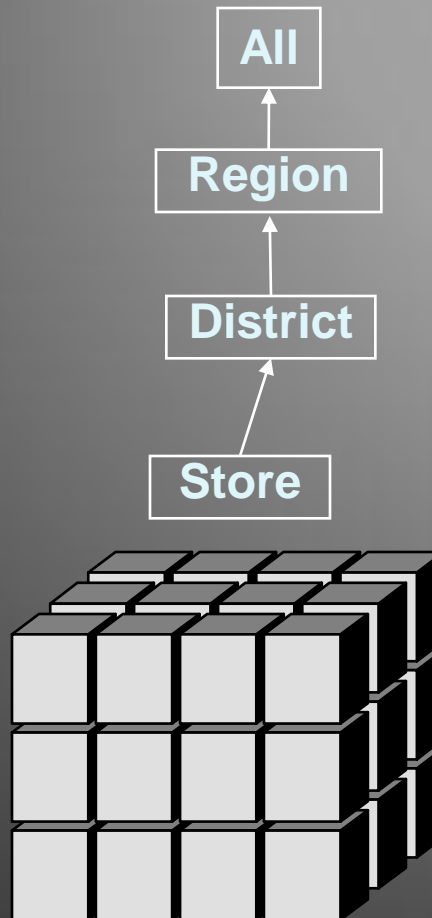
- ▶ Được đề xuất và thiết kế cho một mục đích phân tích dữ liệu
- ▶ Mô hình dữ liệu này không phù hợp cho hệ thống OLTP
- ▶ Mô hình dữ liệu này được thao tác bởi các công cụ OLAP
 - Các công cụ này cung cấp các phương tiện truy vấn mạnh dựa trên thiết kế mô hình dữ liệu đa chiều
 - Ví dụ như: TARGET Analysis, SQL OLAP Server

Mô hình dữ liệu đa chiều (2)

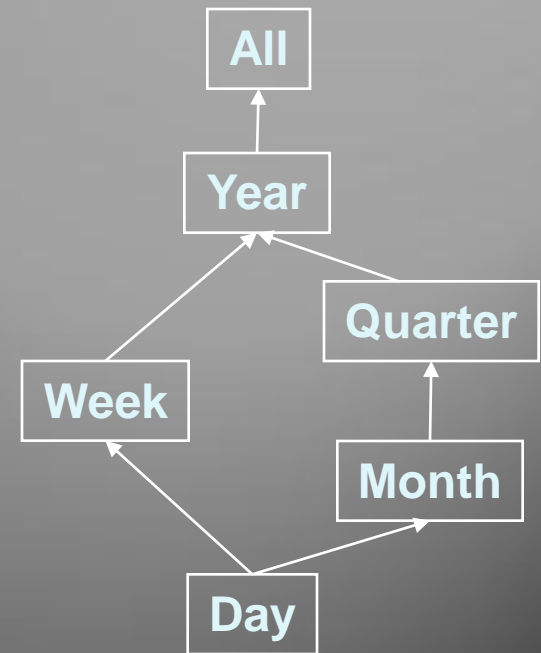
Product



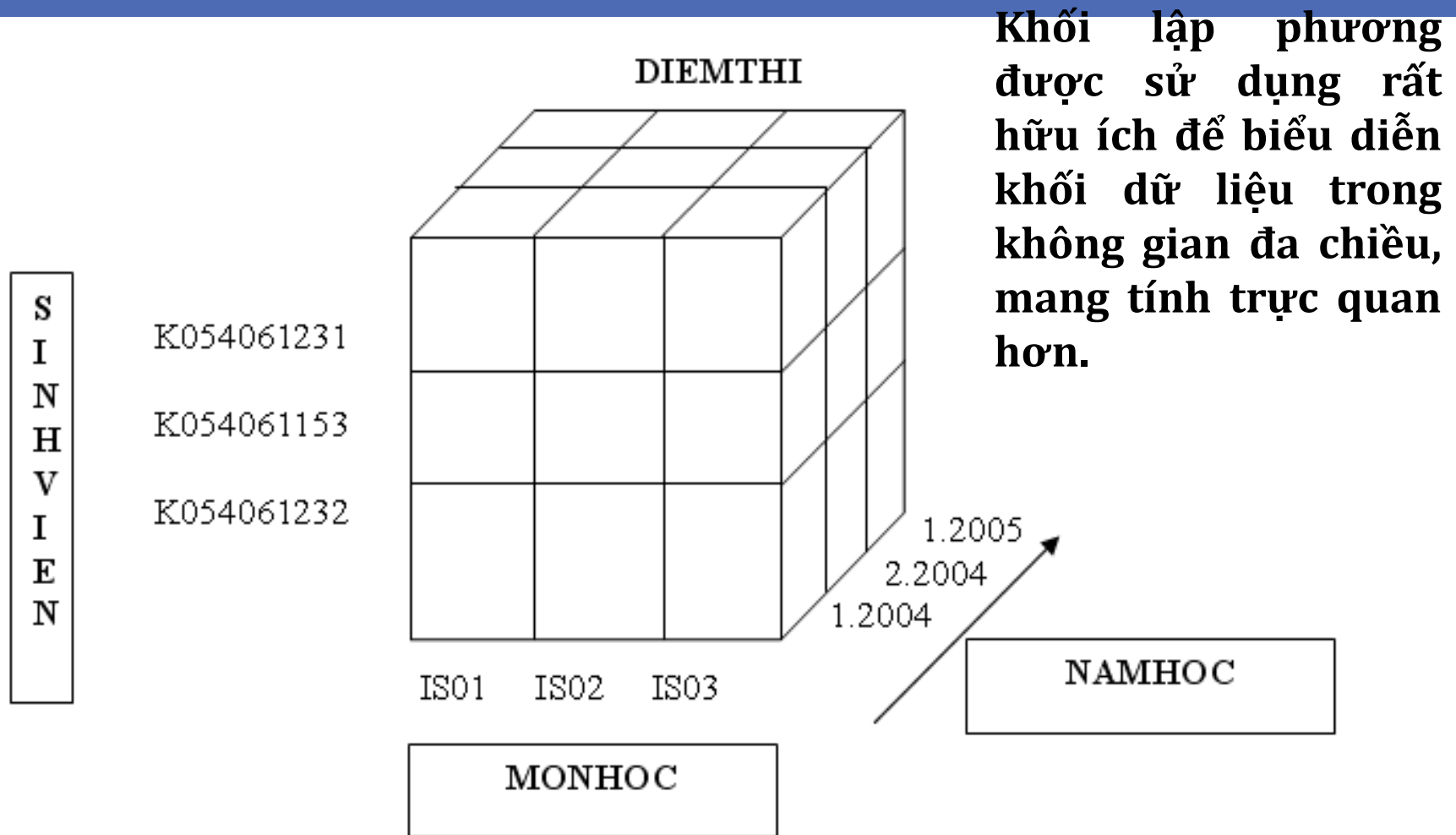
Location



Time



Mô hình dữ liệu đa chiều, ví dụ khác (5)



Dữ liệu được tìm thấy tại điểm giao của các chiều

Mô hình dữ liệu đa chiều (4)

Thể hiện dữ liệu:

SINHVIEN	MONHOC	NAMHOC	DIEMTHI
K054061231	IS01	1.2004	7
K054061231	IS02	2.2004	8
K054061153	IS01	1.2004	9
K054061153	IS02	2.2004	4
K054061153	IS03	1.2005	7
K054061232	IS01	1.2004	6
K054061232	IS02	2.2004	5
K054061231	IS03	1.2005	9
K054061232	IS03	1.2005	6

Đặc điểm

- ▶ Cơ sở dữ liệu đa chiều thường quan tâm đến hai mức:
 - Tổng hợp
 - Chi tiết (dữ liệu nguồn)
- Nên kích thước thường rất lớn.
- ▶ Cơ sở dữ liệu đa chiều thường quan tâm rất nhiều đến yếu tố thời gian, dùng để theo dõi biến động của thực tế theo thời gian
- Do đó chiều thời gian được dùng làm bản lề cho mọi phép phân tích.

Đặc điểm (2)

- ❑ Khái niệm bảng chiều
 - ❑ Bảng chiều chứa dữ liệu miêu tả về một công việc, đối tượng.
- ❑ Bảng chiều thường tương đối nhỏ so với các bảng sự kiện, chứa các thông tin mô tả.

Đặc điểm (3)

- Khái niệm bảng chiều
 - ❑ Bảng chiều chứa các dữ liệu cần thiết cho việc thực hiện các giao tác nghiệp vụ theo một chiều, hay phạm vi nào đó.
 - ❑ Mỗi bảng chiều là một đối tượng của cơ sở dữ liệu, bao gồm các giá trị mà dựa trên đó ta có thể phân loại dữ liệu.

Đặc điểm (4)

- Khái niệm bảng sự kiện
 - ❑ Bảng sự kiện thường rất lớn có thể lên tới hàng triệu dòng
 - ❑ Bảng sự kiện chứa dữ liệu định lượng hoặc sự kiện (là những dữ liệu có độ đo số học).
 - ❑ Bảng sự kiện đóng vai trò của một thực thể kết hợp, kết nối **các thể hiện** của các chiều khác nhau.

Đặc điểm (5)

- Khái niệm thuộc tính, gồm 2 dạng:
 - ❑ **Thuộc tính:** là những thuộc tính của chiều, được dùng để tìm kiếm, truy xuất.
 - ❑ **Thuộc tính phân cấp:** là những thuộc tính được sắp xếp phân tầng tuần tự của phụ thuộc hàm.
- Ví dụ:
 - ❑ Học kỳ → Năm học → Khóa học
 - ❑ Sinh viên → Lớp → Ngành → Bộ môn → Khoa

Đặc điểm (6)

- ▶ Trên cơ sở dữ liệu đa chiều, người dùng có thể phân cấp dữ liệu trên các chiều của chúng.
- ▶ Ví dụ, trên chiều *Thời gian*, người dùng có thể định nghĩa mối quan hệ cấp bậc: *học kỳ* → *năm học* → *khóa học*.
- ▶ Thông qua quan hệ cấp bậc này người dùng có thể xem kết quả học tập ở nhiều cấp như: học kỳ, năm học, khóa học.

Khối

- ▶ Một khối có thể có nhiều chiều
 - Nếu có nhiều hơn 3 chiều sẽ được gọi là ‘siêu khối’ (“hypercube”)
 - Về mặt lý thuyết thì số chiều là không hạn chế
 - Thường thì số chiều là từ 4 đến 12
- ▶ Một khối bao gồm nhiều ô dữ liệu
 - Là một liên kết giữa các giá trị của chiều
 - Một ô có thể là rỗng (không có dữ liệu cho liên kết này)
 - Khối thưa: có nhiều ô rỗng
 - Khối dày đặc: có ít ô rỗng

Các mô hình lưu trữ dữ liệu đa chiều

- ▶ Lược đồ hình sao
- ▶ Lược đồ bông tuyết

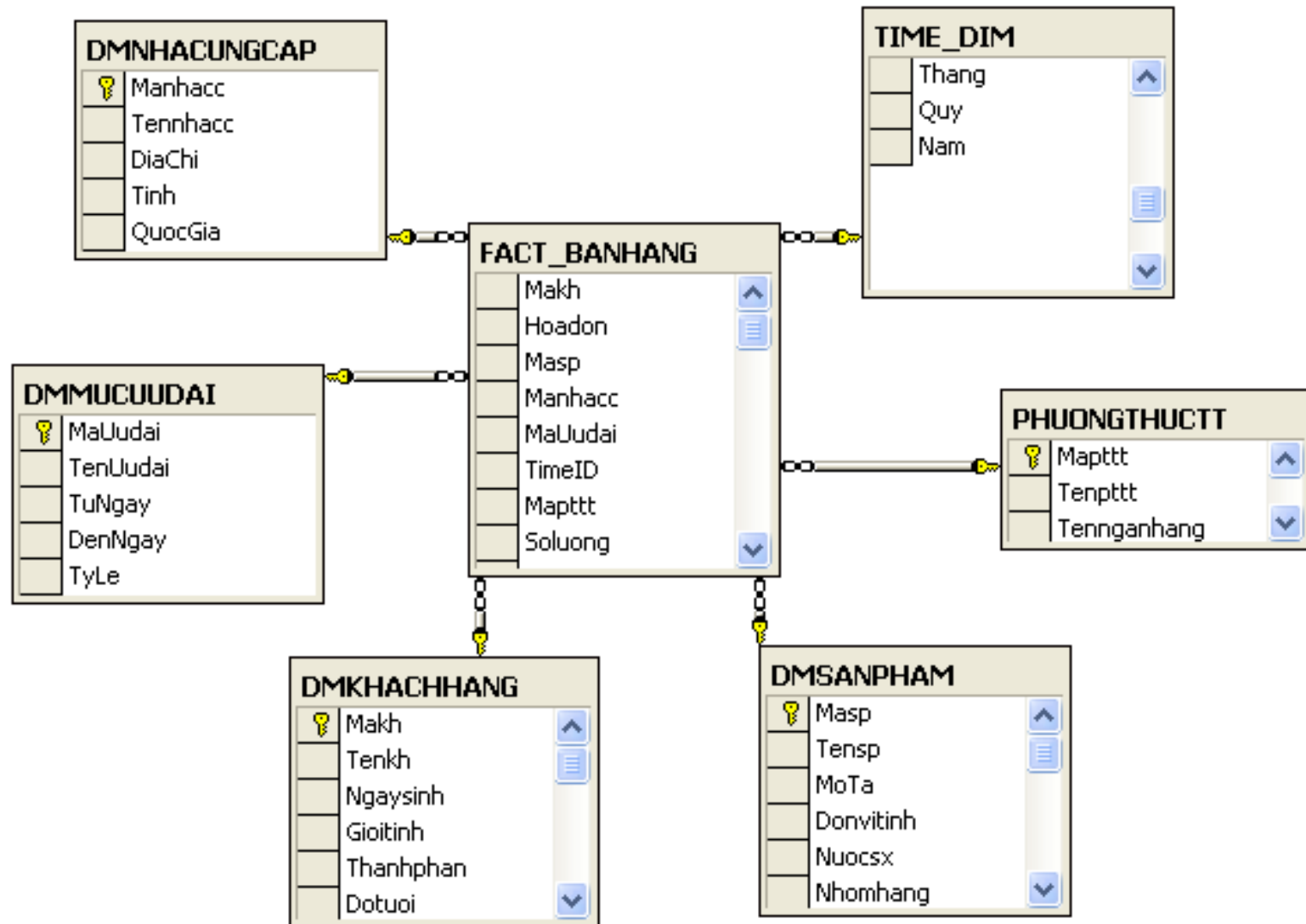
Lược đồ hình sao (2)

- ▶ Cấu trúc của một mô hình chiều cũng có thể được biểu diễn bởi *lược đồ sao*.
- ▶ Trung tâm của lược đồ sao là bảng sự kiện, và bảng sự kiện này là bảng có nhiều kết nối đến các bảng khác.
- ▶ Bảng sự kiện là nơi mà các sự kiện của giao dịch được lưu trữ , còn các bảng khác là bảng chiều.
- ▶ Một MDB có thể có nhiều lược đồ sao với các bảng chiều chồng lấp nhau

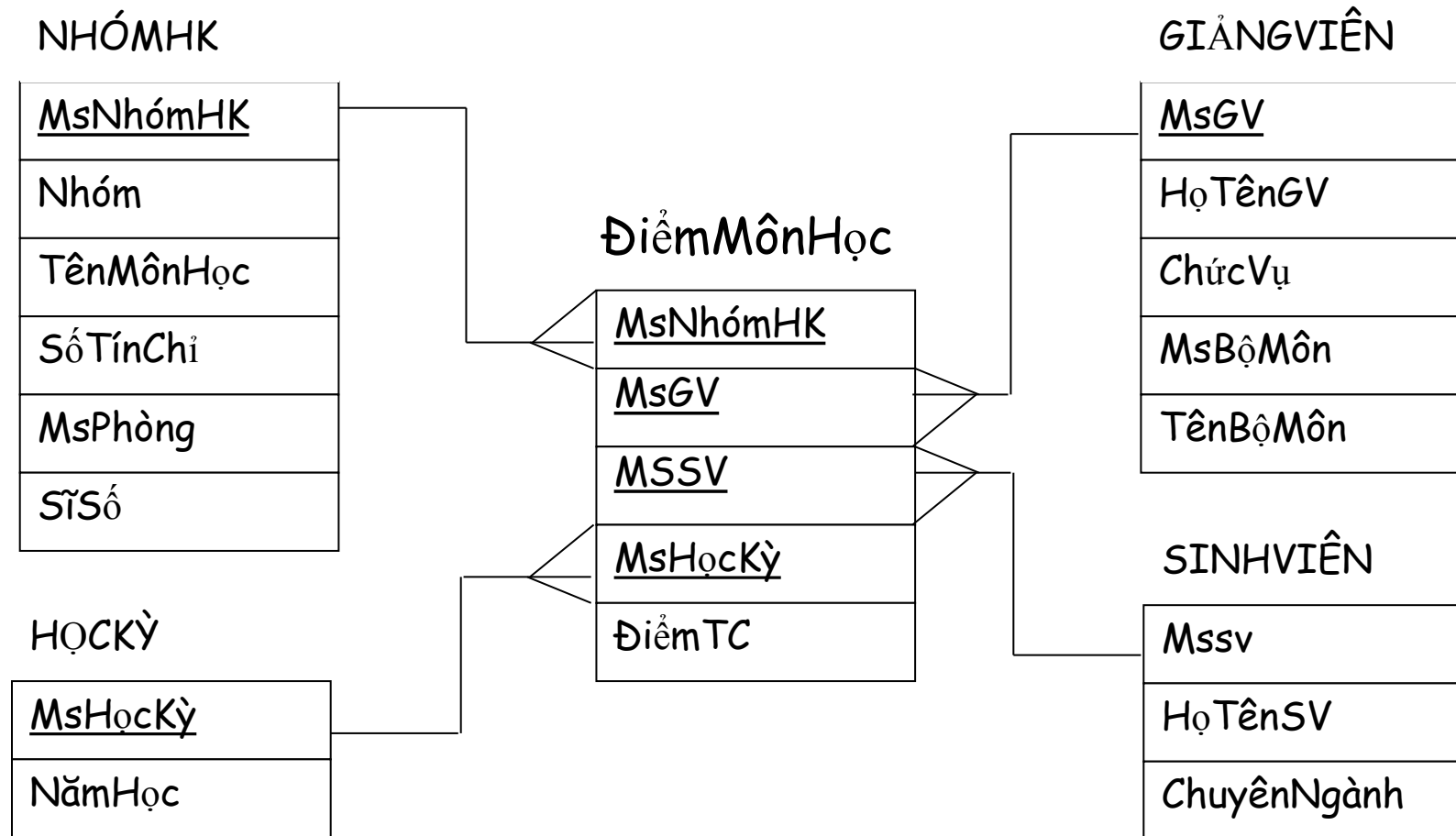
Lược đồ hình sao (3)

- ▶ Phân đoạn dữ liệu
- ▶ Dữ liệu không chuẩn hóa
- ▶ Một bảng dữ kiện và các chiều bao quanh
- ▶ Các câu hỏi nhằm vào bảng dữ kiện và được cấu trúc bởi các bảng chiều
- ▶ Mô hình nổi bật của KDL

Ví dụ 1: lượt đồ sao quản lý bán hàng



Ví dụ 2: lượt đồ sao quản lý điểm sinh viên



Ưu điểm của lược đồ hình sao

- ▶ Các sự kiện và chiều được mô tả rõ ràng, dễ hiểu.
- ▶ Các bảng chiều là dữ liệu tĩnh, tương đối, dữ liệu được nạp (bằng thao tác) vào bảng sự kiện.
- ▶ Lược đồ hình sao còn giúp cải tiến hiệu suất cho các câu truy vấn kho dữ liệu.

Ưu điểm của lược đồ hình sao (2)

- ▶ Lược đồ hình sao rất dễ sử dụng, trực quan, thể hiện khung nhìn đa chiều của dữ liệu dùng ngữ nghĩa của cơ sở dữ liệu quan hệ.
- ▶ Khoá của bảng sự kiện được tạo bởi những khoá của bảng chiều.

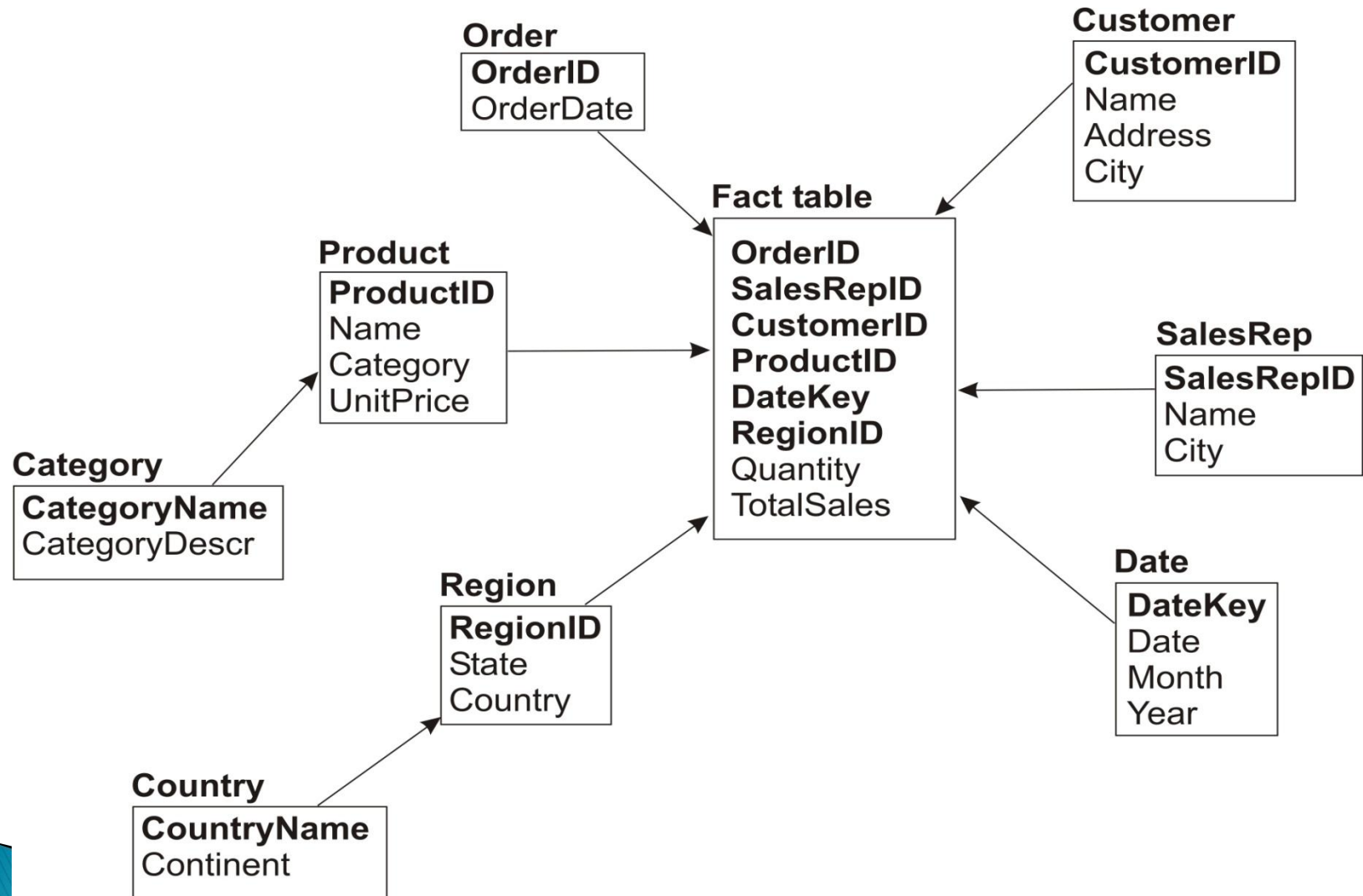
Lược đồ hình bông tuyết

- ▶ Căn cứ vào lược đồ hình sao
- ▶ Bảng sự kiện giống như lược đồ hình sao
- ▶ Các bảng chiều được chuẩn hoá (dạng chuẩn 3)
- ▶ Các chiều được cấu trúc rõ ràng

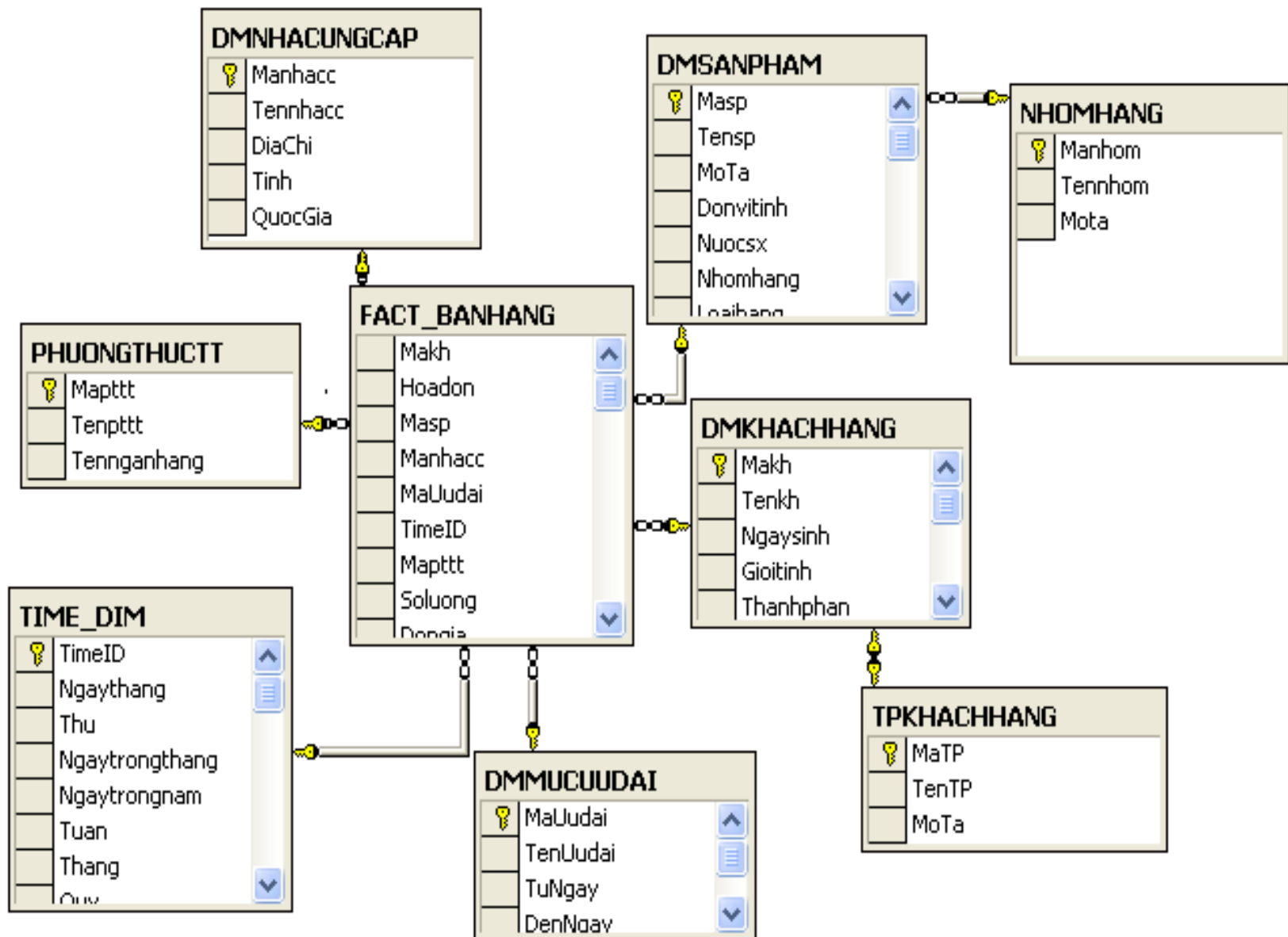
Lược đồ hình bông tuyết (2)

- ▶ Nếu một bảng chiều được chia ra thành chiều chính và chiều phụ thì cấu trúc thu được của kết quả là một lược đồ bông tuyết hoặc một cấu trúc sao mở rộng.

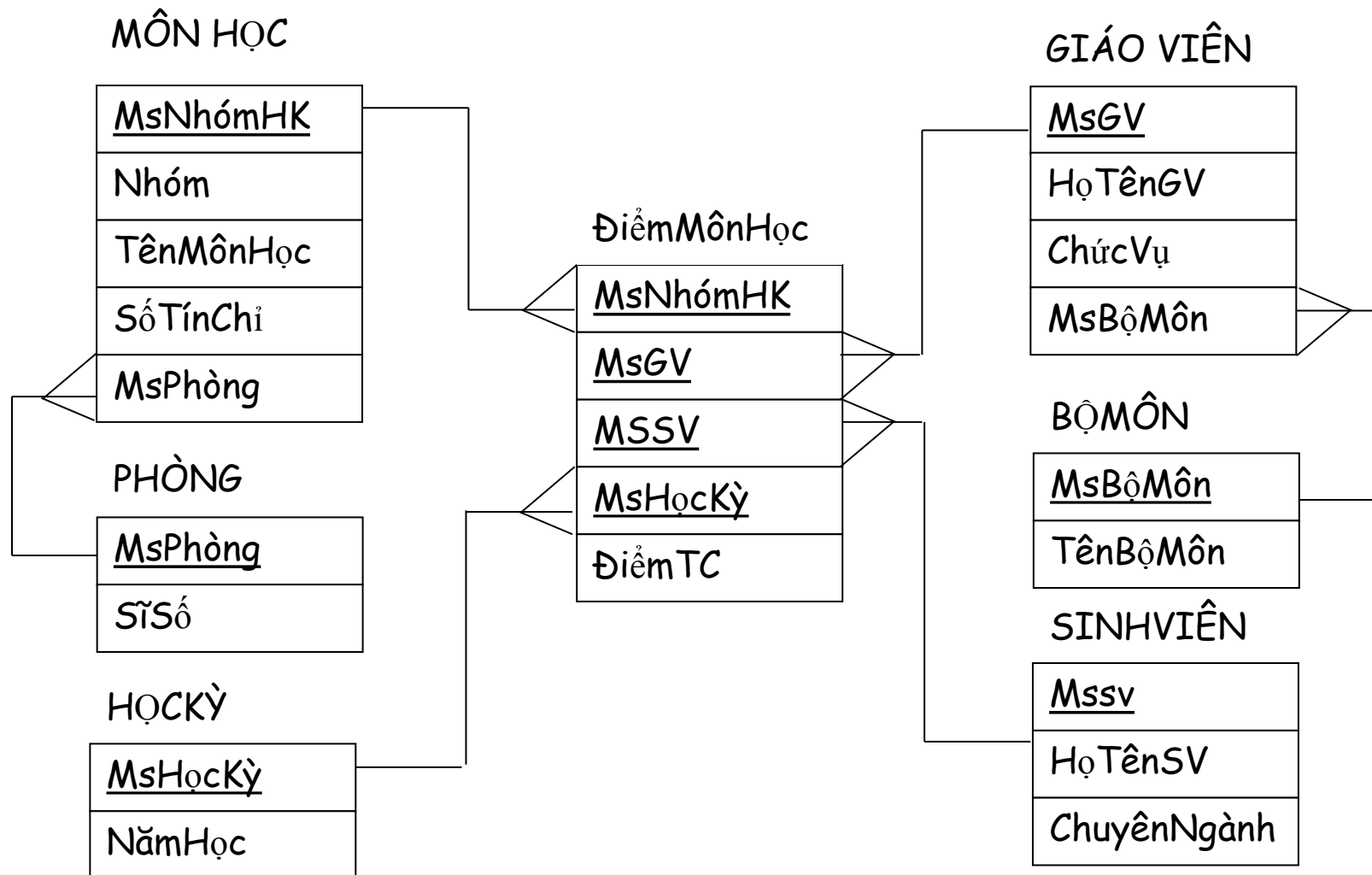
Ví dụ 1 về lược đồ bùng tuyết



Ví dụ 2 về lược đồ bùng tuyết



Ví dụ 3 về lược đồ bông tuyết



Ưu điểm của lược đồ bông tuyết

- ▶ Lược đồ bông tuyết (Snowflake schema) là một cải tiến của lược đồ hình sao
- ▶ Trong đó một số chiều được phân cấp để thể hiện rõ ràng dạng chuẩn của bảng chiều.
- ▶ Tuy nhiên cấu trúc phi dạng chuẩn của các bảng chiều trong lược đồ hình sao có thể thích hợp hơn cho việc duyệt các chiều.

Tóm tắt

- ▶ Qua bài này chúng ta đã khái quát được mô hình cơ sở dữ liệu đa chiều, cách thức lưu trữ dữ liệu đa chiều trong kho dữ liệu
- ▶ Trong bài sau chúng ta sẽ đi tìm hiểu kỹ thuật OLAP trong việc phân tích dữ liệu trên kho dữ liệu hỗ trợ ra quyết định