

BÀI 5: PHÂN TÍCH VÀ THIẾT KẾ KHO DỮ LIỆU

Phân tích hệ thống

- Phương pháp luận xây dựng kho dữ liệu
- Xác định yêu cầu
- Giai đoạn mô tả
- Giai đoạn xây dựng kho dữ liệu
- Giai đoạn nhập và quản trị dữ liệu

Phương pháp luận xây dựng kho dữ liệu

- Chu trình xây dựng kho dữ liệu gồm nhiều pha. Chi tiết từng pha có thể chia làm các giai đoạn sau:
 - Xác định yêu cầu
 - Giai đoạn mô tả
 - Giai đoạn xây dựng kho
 - Giai đoạn nhập và quản trị dữ liệu

Xác định yêu cầu

- Thu thập mẫu báo cáo hiện đang sử dụng, nhu cầu khai thác thông tin của lãnh đạo.
- Xác định các nguồn dữ liệu.
- Thực hiện truy vấn dữ liệu nhanh nhất.
- Thể hiện dữ liệu dưới dạng khung nhìn khác nhau.
- Theo dõi sự thay đổi giá trị trong một thời gian dài.
- Dễ dàng sử dụng với người sử dụng đầu cuối.

Giai đoạn mô tả

- Xác định, thiết kế các bảng chiều.
- Xác định, thiết kế các bảng sự kiện.
- Xác định, thiết kế các bảng tổng hợp.
- Xác định, thiết kế các bảng trung gian.
- Xác định lược đồ hình sao.
- Ánh xạ việc trích lọc, chuyển đổi và tải dữ liệu vào.

Giai đoạn xây dựng kho

- Cấu hình kho vật lý.
- Tạo các kết nối.
- Tạo các chỉ mục.
- Phân vùng.
- Triển khai (trích lọc, ánh xạ, v.v..).
- Nâng cấp và bảo trì kho.

Giai đoạn nhập và quản trị dữ liệu

- Ở pha này, cơ sở hạ tầng của kho dữ liệu đã được hoàn chỉnh. Việc quản trị dữ liệu, đặt lịch cho các công việc quản trị có thể được tiến hành thông qua công cụ độc lập. Các công việc được sắp xếp trong SSIS

Yêu cầu chức năng hệ thống

- **Chức năng tích hợp dữ liệu, chuyển số liệu vào kho dữ liệu**
 - Tích hợp và chuyển dữ liệu từ nguồn dữ liệu sinh hiện đang được quản lý tại đơn vị bằng phần mềm: FoxPro, Excel, SQL Server,...
 - Chuyển số liệu vào kho tự động hoặc theo chu kỳ định trước, hoặc theo yêu cầu của người dùng.

Yêu cầu chức năng hệ thống (2)

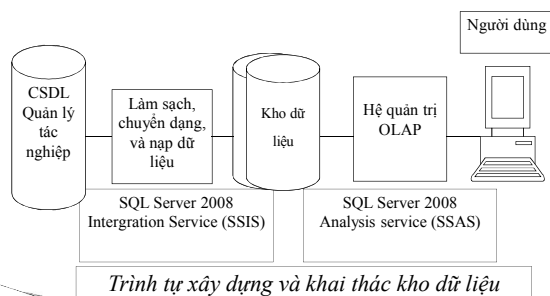
- **Chức năng phân tích số liệu hỗ trợ ra quyết định**
 - Cho phép phân tích số liệu theo từng chủ đề theo kho dữ liệu theo chủ đề
 - Cho phép phân tích số liệu ở nhiều mức độ khác nhau cũng như từ mức chi tiết đến mức tổng hợp.

Yêu cầu chức năng hệ thống (3)

► Chức năng biểu diễn thông tin, tạo báo cáo và thống kê

- Cho phép khai thác và biểu diễn thông tin dữ liệu đã phân tích bằng các công cụ hỗ trợ như: Excel, hoặc bằng dịch vụ SQL Server như SSAS, ...
- In báo cáo, thống kê kết quả theo từng chủ đề.

Các thành phần trong hệ thống



Thiết kế kho dữ liệu

- Phương pháp thiết kế kho dữ liệu
 - Từ trên xuống (Top down)
 - Từ dưới lên (Bottom up)
- UML trong thiết kế kho dữ liệu
- Các mức trong thiết kế kho dữ liệu
 - Ý niệm
 - Luận lý
 - Vật lý

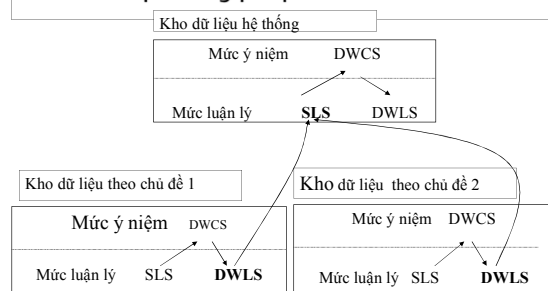
Phương pháp thiết kế kho dữ liệu

- ▶ Ngày nay, trong chiến lược phân tích và thiết kế hệ thống thông thường ta thấy có 2 phương pháp phân tích, thiết kế
 - Từ trên xuống (Top down)
 - Từ dưới lên (Bottom up)
- ▶ Mỗi phương pháp có những ưu và khuyết điểm riêng

Phương pháp thiết kế từ dưới lên

- ▶ Phương pháp thiết kế từ dưới lên
 - Trước tiên những kho dữ liệu theo chủ đề được xây dựng trước
 - Sau đó những kho dữ liệu theo chủ đề này được tích hợp để đi đến mục tiêu xây dựng kho dữ liệu.
- ▶ Với phương pháp này các kho dữ liệu theo chủ đề là nguồn dữ liệu cho kho dữ liệu.

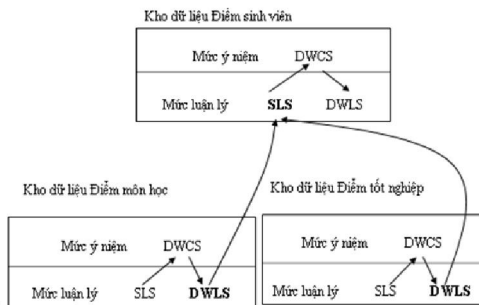
Sơ đồ theo phương pháp thiết kế từ dưới lên



Ví dụ 1

- Sử dụng phương pháp thiết kế từ dưới lên nhằm vào việc tìm kiếm dữ liệu theo từng chủ đề điểm sinh viên ở mức cụ thể hơn như:
 - Kho dữ liệu điểm môn học
 - Kho dữ liệu điểm tốt nghiệp
- Nhằm vào việc hỗ trợ ra quyết định tốt hơn

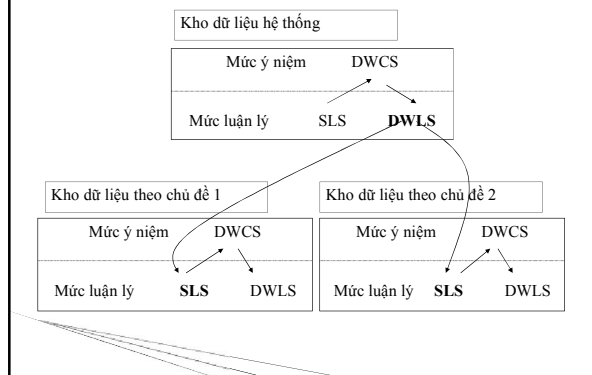
Ví dụ 1 (2)



Phương pháp thiết kế từ trên xuống

- Ngược lại với phương pháp thiết kế từ dưới lên, phương pháp thiết kế từ trên xuống
 - Trước tiên sẽ xây dựng kho dữ liệu hệ thống trước
 - Sau đó phân rã kho dữ liệu này thành các kho dữ liệu theo chủ đề phù hợp với yêu cầu từ phía người sử dụng.
- Khi đó kho dữ liệu lại chính là dữ liệu nguồn cho các kho dữ liệu theo chủ đề.

Sơ đồ theo phương pháp thiết kế từ trên xuống



Ví dụ 2

- Sử dụng phương pháp thiết kế từ trên xuống nhằm vào việc tìm kiếm dữ liệu theo từng chủ đề điểm sinh viên ở mức cụ thể hơn như:
 - Kho dữ liệu điểm môn học
 - Kho dữ liệu điểm tốt nghiệp
- Nhằm vào việc hỗ trợ ra quyết định tốt hơn

Ví dụ 2 (2)

