

Kho dữ liệu



Mục tiêu:

- Hiểu được khái niệm và mục đích của kho dữ liệu
- So sánh OLTP với OLAP
- Một số ví dụ minh họa

1

Nội dung



- Đặt vấn đề
- Khái niệm kho dữ liệu
- Mục đích của kho dữ liệu
- OLTP với OLAP
- Tóm tắt

2

Đặt vấn đề



- CSDL có vai trò quan trọng trong các hệ thống thông tin quản lý:
 - Xử lý nghiệp vụ
 - Xử lý giao dịch
- Một CSDL được thiết kế cho nhu cầu lưu trữ và xử lý thông tin của tổ chức, doanh nghiệp, gồm:
 - Một tập hợp dữ liệu có cấu trúc
 - Một bản thông tin mô tả dữ liệu có cấu trúc đó (metadata)

Kho dữ liệu – TS. Phan Anh Phong

3

3

Đặt vấn đề



- Từ những năm 80 của thế kỷ XX doanh nghiệp đã nhận ra sự cần thiết sử dụng dữ liệu lịch sử để phân tích nhằm hỗ trợ ra quyết định, tạo lợi thế cạnh tranh
- Thuật ngữ “kho dữ liệu” được sử dụng vào năm 1988 trong bài báo kỹ thuật của IBM: *An architecture for a business and information system*

Kho dữ liệu – TS. Phan Anh Phong

4

4

Kho dữ liệu



- Kho dữ liệu là dữ liệu lịch sử được trích ra định kỳ từ các nguồn dữ liệu khác nhau (chủ yếu từ CSDL tác nghiệp) và được chuyển đổi tới một CSDL có thiết kế đặc biệt để xử lý thông tin, xử lý phân tích
- Theo W. H. Inmon thì kho dữ liệu là tập hợp dữ liệu hướng chủ thể, tích hợp, thay đổi theo thời gian và có tính ổn định, với mục đích hỗ trợ ra quyết định

Kho dữ liệu – TS. Phan Anh Phong

5

5

Kho dữ liệu



- Hướng chủ đề
 - Được tổ chức xung quanh chủ đề chính của doanh nghiệp, chẳng hạn như Bán hàng, Khách hàng, Nhà cung cấp, Sản phẩm...
- Tích hợp, thay đổi theo thời gian, ổn định
 - Thu thập dữ liệu ổn định
 - Làm sạch
 - Tải dữ liệu
 - Phát sinh

Kho dữ liệu – TS. Phan Anh Phong

6

6

Kho dữ liệu



- Có nhiều công ty lớn đầu tư vào công nghệ kho dữ liệu và công cụ liên quan:
 - Microsoft
 - Oracle
 - IBM Infosphere
 - Amazon Redshift
 - ...

Kho dữ liệu – TS. Phan Anh Phong

7

7

Kho dữ liệu



- Các công cụ
 - Tích hợp dữ liệu
 - Truy vấn
 - Báo cáo
 - Phân tích
 - Khai phá dữ liệu
 - Quản trị kho dữ liệu
- Kho dữ liệu và Khai phá dữ liệu?

Kho dữ liệu – TS. Phan Anh Phong

8

8

Mục đích của kho dữ liệu



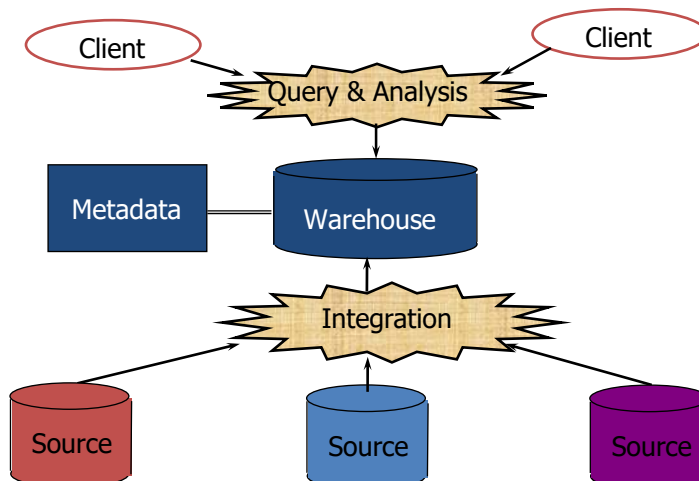
- Xử lý thông tin - Dữ liệu được xử lý bằng các truy vấn, phân tích thống kê cơ bản, báo cáo bằng cách sử dụng bảng biểu, biểu đồ, đồ thị
- Xử lý phân tích - Dữ liệu có thể được phân tích bằng các thao tác OLAP cơ bản, như roll up, drill down, drill up...
- Khai phá dữ liệu - Tìm các mô hình và các mối liên kết ẩn, thực hiện phân loại và dự đoán.

Kho dữ liệu – TS. Phan Anh Phong

9

9

Kiến trúc của kho dữ liệu



Kho dữ liệu – TS. Phan Anh Phong

10

10

OLAP là gì?



- Kho dữ liệu cho phép người dùng ở mức quản lý, ra quyết định thực hiện các phép phân tích tương tác với data bằng hệ thống xử lý phân tích trực tuyến (online analytical processing – OLAP).
- Ngoài ra kho dữ liệu cũng được dùng cho báo cáo, data mining và phân tích thống kê.
- Database và kho dữ liệu, do đó chỉ khác nhau về mặt khái niệm, một cơ sở dữ liệu nếu dùng riêng cho các mục đích trên cũng được coi là kho dữ liệu.

OLAP là gì?



- Xử lý phân tích trực tuyến (On-Line Analytical Processing – OLAP)
 - Xử lý phân tích dữ liệu (thời gian thực) để xác định xu hướng, hỗ trợ quyết định chiến lược kinh doanh
 - Xây dựng báo cáo về tài chính, bán hàng, tiếp thị, quản trị, dự báo...
- So sánh OLAP với OLTP (On-Line Transaction Processing)

OLTP và OLAP



OLTP

- Phần lớn là cập nhật
- Truy vấn đơn giản
- Kích thước dữ liệu MB-GB
- Dữ liệu thô
- Người dùng đa dạng, nhiều, truy cập đồng thời
- Lượng giao dịch cao
- Định hướng xử lý

OLAP

- Phần lớn là đọc dữ liệu
- Truy vấn dài, phức tạp
- Kích thước dữ liệu GB-TB
- Dữ liệu tổng hợp
- Người dùng: lãnh đạo, người quản lý, chuyên gia
- Lượng giao dịch trung bình, thấp
- Định hướng chủ đề

Kho dữ liệu – TS. Phan Anh Phong

13

13

OLAP



- Dữ liệu trong Kho dữ liệu được tổ chức dưới dạng các khối dữ liệu đa chiều (Multi Dimensional Cube) và OLAP dùng để phân tích trên các khối dữ liệu đó.
- OLAP cho phép người dùng phân tích dữ liệu qua việc cắt lát (slice) dữ liệu theo nhiều khía cạnh khác nhau, khoan xuống (drill down) mức chi tiết hơn hay là cuộn lên (roll up) mức tổng hợp hơn của dữ liệu Ứng dụng của OLAP

Kho dữ liệu – TS. Phan Anh Phong

14

14

ROLAP và MOLAP



- ROLAP:
Relational On-Line Analytical Processing
- MOLAP:
Multi-Dimensional On-Line Analytical Processing

15

Các ví dụ minh họa



- Kho dữ liệu về bán hàng
 - xây dựng báo cáo về tài chính, bán hàng, tiếp thị, quản trị, dự báo
 - ...
- Kho dữ liệu về khách hàng

16