Kho dữ liệu dùng chung và các CSDL chuyên môn

1. Xây dựng kho dữ liệu

Kho dữ liệu (Data Warehouse – DW) là một hệ thống tổng hợp tất cả dữ liệu từ nhiều nguồn khác nhau vào một Kho dữ liệu nhất quán để hỗ trợ phân tích và khai thác dữ liệu. DW cho phép người dùng thực hiện các phân tích chuyên sâu dựa trên khối lượng lớn dữ liệu lịch sử đã lưu trữ.

Kho dữ liệu là một phần quan trọng trong giải pháp kinh doanh thông minh (BI) ngày nay, giúp doanh nghiệp có cái nhìn sâu sắc hơn về hiệu suất hoạt động kinh doanh thông qua việc so sánh dữ liệu được hợp nhất từ nhiều nguồn khác nhau, và được thiết kế để đáp ứng cho các câu hỏi, phân tích dữ liệu có nguồn gốc từ các nguồn giao dịch.

**Kiến trúc Kho dữ liệu**

Nhìn chung, một kho lưu trữ dữ liệu sẽ có kiến trúc gồm 3 tầng dưới đây:

* **Tầng trên cùng**: Hiển thị một số loại giao diện người dùng ngoại vi hoặc công cụ báo cáo, cho phép người dùng cuối cùng tiến hành phân tích chuyên sâu dựa trên dữ liệu kinh doanh của họ.
* **Tầng giữa**: Chứa máy chủ OLAP, giúp truy vấn dữ liệu nhanh chóng. Đặc biệt sẽ gồm 3 loại mô hình OLAP gồm: ROLAP, MOLAP và HOLAP – tùy vào loại hệ thống cơ sở dữ liệu hiện có mà sử dụng loại mô hình phù hợp.
* **Tầng dưới cùng:** Bao gồm một máy chủ Kho dữ liệu, thường là hệ thống cơ sở dữ liệu quan hệ đóng vai trò thu thập, “làm sạch” và chuyển đổi dữ liệu từ nhiều nguồn thông qua một quy trình được gọi là Trích xuất, Chuyển đổi và Tải (ETL).

## Các giai đoạn chung sử dụng Kho dữ liệu

Dưới đây là 4 giai đoạn chung khi sử dụng Kho dữ liệu:

### 1. Cơ sở dữ liệu hoạt động ngoại tuyến

Đây là giai đoạn bắt đầu, tất cả dữ liệu được sao chép từ hệ điều hành sang máy chủ khác. Trong suốt quá trình tải, xử lý và báo cáo dữ liệu sẽ không ảnh hưởng đến hiệu suất của hệ điều hành.

### 2. Kho dữ liệu ngoại tuyến

Tất cả dữ liệu trong kho được cập nhật liên tục từ cơ sở dữ liệu hoạt động – được ánh xạ và chuyển đổi nhằm đáp ứng các nhu cầu sử dụng của người dùng.

### 3. Kho dữ liệu thời gian thực

Kho dữ liệu sẽ cập nhật ngay lập tức khi có bất kỳ giao dịch nào diễn ra trong cơ sở vận hành. Ví dụ là hệ thống đường sắt hoặc đặt vé máy bay.

### 4. Kho dữ liệu tích hợp

Kho dữ liệu cũng cập nhật liên tục khi có bất kỳ giao dịch nào được thực hiện, đồng thời tạo ra các giao dịch sau đó – được chuyển trở lại hệ thống vận hành.

## Các mô hình biểu diễn của Kho dữ liệu

Kho dữ liệu có 3 mô hình biểu diễn chính – mỗi mô hình sẽ có ưu – nhược điểm riêng. Tùy theo quy mô hoạt động và nhu cầu kinh doanh mà bạn có thể lựa chọn phương án phù hợp nhất. Cụ thể:

### 1. Mô hình chòm sao

Phù hợp với những doanh nghiệp có nhu cầu lưu trữ dữ liệu lớn. Các bảng thông số và bảng sự kiện liên kết với nhau tạo thành một đồ thị chi tiết. Mặc dù, doanh nghiệp có thể giảm số lượng bảng thông số cần dùng, nhưng mô hình này cũng ảnh hưởng lớn đến quá trình phân tích và trích xuất dữ liệu.

Ví dụ, doanh nghiệp bạn có hệ thống bán hàng và giao hàng được thiết kế hoàn toàn tách biệt. Người quản lý sẽ lấy thông tin từ 2 nguồn này và ghi vào bảng thông số chung là Địa điểm.

Vì tính chất công việc khá phức tạp – đòi hỏi người dùng phải có kỹ năng, bám sát hướng dẫn, thấu hiểu tổ chức để đọc báo cáo số liệu chính xác.

### 2. Mô hình ngôi sao

Cấu tạo của mô hình này là một bảng sự kiện trung tâm, hứa thông tin bán hàng như: số lượng đơn bán ra, số tiền thu được và doanh số trung bình hàng tháng. Xung quanh là các bảng thông số khác là chi nhánh, mặt hàng, thời gian và địa điểm. Từng chủ đề được tạo sẽ hiển thị dưới dạng 1 hoặc nhiều ngôi sao.

Ưu điểm nổi bật của mô hình ngôi sao là tính năng truy vấn đơn giản và thời gian kiểm tra báo cao nhanh chóng. Tuy nhiên, số liệu lưu trữ trong mô hình này dễ bị trùng lặp và dư thừa thông số, gây lãng phí khó.

### 3. Mô hình bông tuyết

Là một biến thể khác, nhằm mục đích khắc phục các nhược điểm của mô hình ngôi sao. Mô hình bông tuyết giúp chuẩn hóa bảng thông số nhờ tính năng thêm vào nhiều thông số nhỏ hơn. Từng chủ đề sẽ được phân tích bằng 1 hoặc nhiều “bông tuyết”.

Mỗi “bông tuyết” đều được làm rõ và tránh trùng tình trạng lặp thông số, nhưng điều này lại làm mất đi ưu điểm của mô hình ngôi sao. Hơn nữa, vì phải liên kết đa chiều nên tốc độ xử lý dữ liệu sẽ mất nhiều thời gian hơn.

## Lợi ích khi sử dụng Kho dữ liệu

Kho dữ liệu mang lại cho doanh nghiệp nhiều lợi ích tuyệt vời sau:

### 1. Quản lý tập trung

Khi chuyển đổi số, doanh nghiệp phải xử lý lượng lớn dữ liệu từ nội bộ đến bến ngoài. Chính vì thế, Kho dữ liệu Data Warehouse trở thành công cụ quản lý thông tin nhất quán và khoa học, giúp người dùng tối ưu thời gian tìm kiếm, phân tích và truy vấn dữ liệu.

### 2. Nhất quán lưu trữ dữ liệu

Doanh nghiệp có thể loại bỏ các dữ liệu thừa và nhất quán hoạt động thu thập dữ liệu vào kho. Điều này giúp người dùng có thể tìm kiếm dữ liệu chính xác, giảm thiểu sai sót và nhầm lẫn thông tin. Bên cạnh đó, tính nhất quán cũng giúp cho quá trình đào tạo nhân viên mới diễn ra dễ dàng và nhanh hơn.

Tất cả dữ liệu sẽ được lưu trữ một cách nhất quán theo từng chủ đề nhằm đảm bảo khả năng truy vấn nhanh chóng và chính xác hơn

### 3. Tăng khả năng cạnh tranh

Nhà lãnh đạo và các nhà quản lý có thể phân tích chuyên sâu để tìm hiểu về những thách thức mà doanh nghiệp đang đối mặt thông qua Kho dữ liệu. Từ đó, ban lãnh đạo có thể đưa ra giải pháp khắc phục kịp thời, nâng cao chất lượng sản phẩm, định hướng đúng phân khúc khách hàng và tăng khả năng cạnh tranh trên thị trường.

## Kho dữ liệu hoạt động như thế nào?

Khi dữ liệu được thêm vào sẽ không thay đổi hoặc sửa đổi vì Kho dữ liệu vận hành các phân tích về dữ liệu lịch sử và tập trung thay đổi các dữ liệu theo thời gian thực. Tất cả dữ liệu trong kho phải được lưu trữ một cách an toàn, đáng tin cậy, dễ quản lý và truy xuất ngay khi cần.

Khi tạo một Kho dữ liệu, có một số bước nhất định mà bạn cần ghi nhớ:

* Bước 1: Trích xuất dữ liệu từ nhiều điểm nguồn.
* Bước 2: Sau khi dữ liệu được biên soạn sẽ trải qua quá trình “làm sạch”: xử lý để tìm lỗi và sửa hoặc loại bỏ (nếu cần thiết).
* Bước 3: Dữ liệu được chuyển đổi từ định dạng cơ sở dữ liệu sang định dạng kho.
* Bước 4: Sau khi lưu trữ, dữ liệu sẽ được sắp xếp và nhất quán theo từng chủ đề để người dùng dễ dàng điều phối và sử dụng.

Nhiều doanh nghiệp sử dụng Kho dữ liệu để nghiên cứu và khai thác thông tin nhằm cải thiện quy trình kinh doanh tốt hơn. Một hệ thống lưu trữ dữ liệu bài bản sẽ giúp các phòng ban trong công ty có thể truy cập và phân tích dữ liệu của nhau hiệu quả hơn.

1. Xây dựng CSDL chuyên môn