**1. Kho lưu trữ dữ liệu (data warehouse):**

* Là một hệ thống tập trung dữ liệu từ **nhiều nguồn khác nhau** trong tổ chức, bao gồm:
  + Hệ thống kế toán
  + Hệ thống quản lý quan hệ khách hàng (CRM)
  + Hệ thống điểm bán hàng (POS)
  + Hệ thống quản lý kho hàng
  + v.v.
* Dữ liệu được lưu trữ trong kho dữ liệu có thể là **cấu trúc** (ví dụ: dữ liệu khách hàng) hoặc **phi cấu trúc** (ví dụ: email, ảnh, video).
* Mục đích của kho dữ liệu là cung cấp **cái nhìn tổng quan** về hoạt động kinh doanh và hỗ trợ **phân tích dữ liệu** cho các mục đích khác nhau, bao gồm:
  + Báo cáo và phân tích kinh doanh
  + Hỗ trợ ra quyết định
  + Phân tích dự đoán
  + Khai phá dữ liệu
* **Ví dụ:**
  + Một công ty bán lẻ có thể sử dụng kho dữ liệu để lưu trữ dữ liệu bán hàng từ các cửa hàng khác nhau. Dữ liệu này có thể được sử dụng để phân tích xu hướng bán hàng, hiệu quả của các chiến dịch marketing, v.v.

**2. Xử lý phân tích trực tuyến (OLAP):**

* Là một loại **truy vấn dữ liệu** tập trung vào việc phân tích và tổng hợp dữ liệu theo nhiều chiều khác nhau.
* OLAP thường được sử dụng để **phân tích dữ liệu lịch sử** trong kho dữ liệu để **trả lời các câu hỏi kinh doanh** như:
  + Doanh số bán hàng của sản phẩm nào cao nhất trong quý trước?
  + Khu vực nào có tỷ lệ khách hàng quay lại cao nhất?
  + Mức độ hiệu quả của các chiến dịch marketing khác nhau?
* OLAP sử dụng các kỹ thuật **tổng hợp dữ liệu** như lập bảng tổng hợp, tính toán trung bình, tỷ lệ phần trăm, v.v. để giúp người dùng **dễ dàng hiểu và phân tích dữ liệu**.
* **Ví dụ:**
  + Một nhà phân tích kinh doanh có thể sử dụng OLAP để phân tích dữ liệu bán hàng theo sản phẩm, khu vực, thời gian, v.v. để xác định các xu hướng và đưa ra các khuyến nghị kinh doanh.

**3. Dữ liệu lớn (big data):**

* Là thuật ngữ dùng để chỉ **lượng dữ liệu khổng lồ** được thu thập với tốc độ ngày càng tăng, khối lượng ngày càng lớn và đa dạng về định dạng hơn bao giờ hết.
* Dữ liệu lớn có thể bao gồm:
  + Dữ liệu giao dịch
  + Dữ liệu cảm biến
  + Dữ liệu mạng xã hội
  + Dữ liệu log
  + v.v.
* Dữ liệu lớn mang đến **nhiều cơ hội** cho doanh nghiệp, nhưng cũng đặt ra **nhiều thách thức** về việc lưu trữ, xử lý và phân tích dữ liệu.
* **Ví dụ:**
  + Một công ty sản xuất có thể sử dụng dữ liệu cảm biến từ các nhà máy của mình để theo dõi hiệu quả hoạt động và dự đoán các vấn đề tiềm ẩn.

**4. Dữ liệu thời gian thực (real-time data):**

* Là dữ liệu được **truyền trực tiếp** từ nguồn gốc, không bị trễ.
* Dữ liệu thời gian thực được sử dụng cho các ứng dụng cần **phản hồi nhanh chóng**, chẳng hạn như:
  + Giám sát hệ thống
  + Phát hiện gian lận
  + Giao dịch tài chính
  + v.v.
* Dữ liệu thời gian thực có thể được kết hợp với dữ liệu lịch sử trong kho dữ liệu để cung cấp **cái nhìn toàn diện** về hoạt động kinh doanh.
* **Ví dụ:**
  + Một công ty chứng khoán có thể sử dụng dữ liệu thời gian thực để theo dõi giá cổ phiếu và đưa ra các quyết định giao dịch.

**Lưu ý:**

* Các khái niệm này có liên quan mật thiết với nhau. Kho dữ liệu có thể được sử dụng để lưu trữ và xử lý cả dữ liệu lớn và dữ liệu thời gian thực. OLAP có thể được sử dụng để phân tích cả dữ liệu lịch sử và dữ liệu thời gian thực.

**Đặc điểm:**

**1. Tập trung vào việc cung cấp câu trả lời cho các truy vấn phức tạp:**

* Khác với các hệ thống cơ sở dữ liệu truyền thống tập trung vào xử lý giao dịch (OLTP), kho lưu trữ dữ liệu được thiết kế để **phân tích dữ liệu** và **trả lời các câu hỏi kinh doanh phức tạp**.
* Các truy vấn trong kho dữ liệu thường liên quan đến **nhiều bảng dữ liệu** và **nhiều bước xử lý**.
* Kho lưu trữ dữ liệu sử dụng các kỹ thuật **tối ưu hóa** để thực hiện các truy vấn phức tạp một cách hiệu quả.

**2. Có thể xử lý dữ liệu lớn:**

* Kho lưu trữ dữ liệu có thể lưu trữ và xử lý **lượng dữ liệu khổng lồ**, bao gồm cả dữ liệu lịch sử và dữ liệu thời gian thực.
* Khả năng xử lý dữ liệu lớn của kho dữ liệu giúp doanh nghiệp có thể **phân tích toàn bộ dữ liệu** của họ, không chỉ một phần nhỏ.
* Việc phân tích toàn bộ dữ liệu có thể giúp doanh nghiệp **nhận ra các xu hướng** và **mối tương quan** mà họ có thể bỏ lỡ nếu chỉ phân tích một phần nhỏ dữ liệu.

**3. Cung cấp thông tin cho các quyết định kinh doanh quan trọng:**

* Kho lưu trữ dữ liệu cung cấp thông tin chi tiết về hoạt động kinh doanh của doanh nghiệp.
* Thông tin này có thể được sử dụng để **đưa ra các quyết định kinh doanh quan trọng**, chẳng hạn như:
  + Mở rộng thị trường sang khu vực nào?
  + Phát triển sản phẩm mới nào?
  + Nâng cao hiệu quả hoạt động như thế nào?
* Việc sử dụng dữ liệu để đưa ra quyết định giúp doanh nghiệp **tăng hiệu quả hoạt động** và **cạnh tranh tốt hơn**.

**Ví dụ:**

* Một công ty bán lẻ có thể sử dụng kho lưu trữ dữ liệu để phân tích xu hướng bán hàng và đưa ra quyết định về việc mở rộng cửa hàng mới.
* Một công ty sản xuất có thể sử dụng kho lưu trữ dữ liệu để phân tích hiệu quả hoạt động của các nhà máy và đưa ra quyết định về việc cải thiện quy trình sản xuất.

**Lưu ý:**

* Các đặc điểm này của kho lưu trữ dữ liệu giúp doanh nghiệp **tận dụng tối đa dữ liệu** của họ và **đưa ra các quyết định kinh doanh sáng suốt**.

**Thành phần:**

**1. Dữ liệu từ nhiều nguồn khác nhau:**

* Kho lưu trữ dữ liệu thu thập dữ liệu từ **nhiều nguồn khác nhau** trong tổ chức, bao gồm:
  + Hệ thống kế toán
  + Hệ thống quản lý quan hệ khách hàng (CRM)
  + Hệ thống điểm bán hàng (POS)
  + Hệ thống quản lý kho hàng
  + v.v.
* Dữ liệu có thể được **trích xuất** từ các hệ thống nguồn bằng nhiều cách khác nhau, bao gồm:
  + Sao chép toàn bộ dữ liệu
  + Sao chép dữ liệu thay đổi
  + Sao chép dữ liệu theo thời gian thực
* Dữ liệu từ các nguồn khác nhau được **hợp nhất** trong kho dữ liệu để tạo ra một cái nhìn thống nhất về hoạt động kinh doanh.

**2. Các công cụ và dịch vụ để xử lý và phân tích dữ liệu:**

* Kho lưu trữ dữ liệu sử dụng nhiều công cụ và dịch vụ để **xử lý** và **phân tích dữ liệu**, bao gồm:
  + Hệ thống quản lý cơ sở dữ liệu (DBMS)
  + Công cụ ETL (Extract, Transform, Load)
  + Công cụ phân tích dữ liệu
  + Công cụ khai phá dữ liệu
* Hệ thống DBMS được sử dụng để **lưu trữ** và **quản lý** dữ liệu trong kho dữ liệu.
* Công cụ ETL được sử dụng để **trích xuất**, **chuyển đổi** và **tải** dữ liệu từ các hệ thống nguồn vào kho dữ liệu.
* Công cụ phân tích dữ liệu được sử dụng để **phân tích** dữ liệu và **trả lời các câu hỏi kinh doanh**.
* Công cụ khai phá dữ liệu được sử dụng để **tìm kiếm các mẫu** và **mối tương quan** trong dữ liệu.

**3. Giao diện để truy cập và trực quan hóa dữ liệu:**

* Kho lưu trữ dữ liệu cung cấp giao diện để người dùng có thể **truy cập** và **trực quan hóa dữ liệu**.
* Giao diện này có thể bao gồm:
  + Báo cáo
  + Biểu đồ
  + Bảng điều khiển
* Giao diện trực quan giúp người dùng **dễ dàng hiểu** và **phân tích dữ liệu**.

**Ví dụ:**

* Một công ty bán lẻ có thể sử dụng một công cụ ETL để trích xuất dữ liệu bán hàng từ hệ thống POS của mình và tải dữ liệu vào kho dữ liệu.
* Một nhà phân tích kinh doanh có thể sử dụng một công cụ phân tích dữ liệu để phân tích dữ liệu bán hàng và tạo báo cáo về xu hướng bán hàng.

**Lưu ý:**

* Các thành phần này của kho lưu trữ dữ liệu giúp doanh nghiệp **thu thập**, **xử lý**, **phân tích** và **trực quan hóa dữ liệu** một cách hiệu quả.

**5. Ví dụ:**

* Azure Data Factory: Tự động hóa di chuyển và chuyển đổi dữ liệu.
* Azure Data Lake Storage: Lưu trữ dữ liệu phi cấu trúc với dung lượng lớn.
* Azure Databricks: Phân tích dữ liệu dạng hồ dữ liệu và học máy.
* Power BI: Phân tích kinh doanh بصريا (b بصريا - بصري - basri = trực quan).