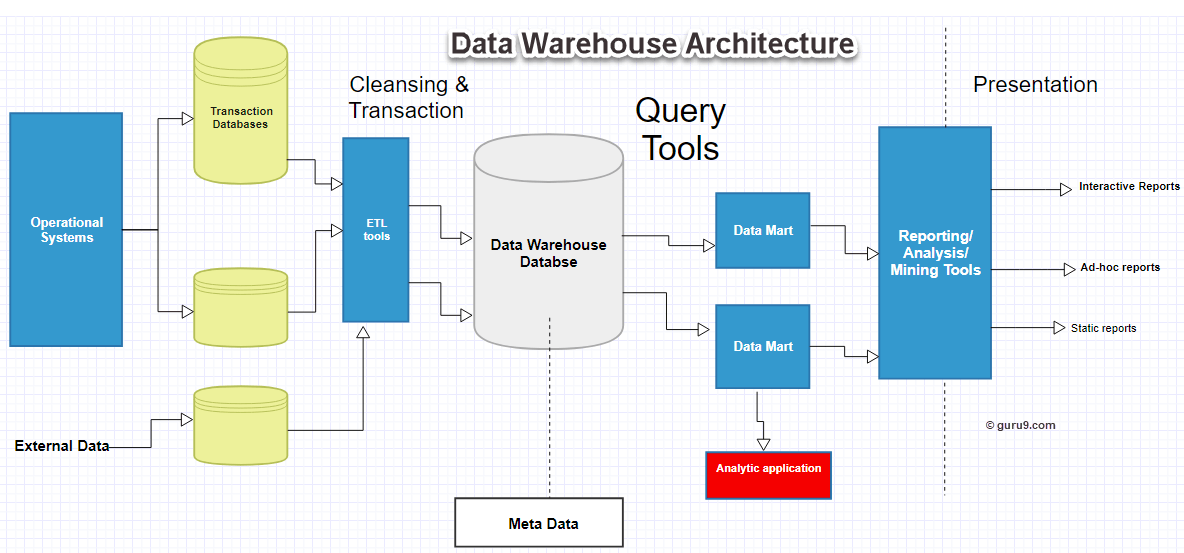
# **Các thành phần Datawarehouse**

Chúng ta sẽ tìm hiểu về các Thành phần Datawarehouse và Kiến trúc của Data Warehouse với Sơ đồ như hình dưới đây:



Trong đó, chủ yếu có năm Thành phần Data Warehouse:

1. **Data Warehouse Database**

Cơ sở dữ liệu trung tâm là nền tảng của môi trường Data Warehouse. Cơ sở dữ liệu này được thực hiện trên công nghệ RDBMS.

1. **ETL Tools**

ETL – Extract, Transform and Load

Các công cụ ETL được sử dụng để thực hiện tất cả các chuyển đổi, tóm tắt và tất cả các thay đổi cần thiết để chuyển đổi dữ liệu thành một định dạng thống nhất trong Data Warehouse.

Chức năng của chúng bao gồm:

* Ẩn danh dữ liệu theo quy định của pháp luật.
* Loại bỏ dữ liệu không mong muốn trong cơ sở dữ liệu hoạt động khi tải vào Data Warehouse.
* Tìm kiếm và thay thế các tên và định nghĩa phổ biến cho dữ liệu đến từ các nguồn khác nhau.
* Tính toán tóm tắt và dữ liệu dẫn xuất
* Trong trường hợp thiếu dữ liệu, điền chúng bằng các giá trị mặc định.
* Loại bỏ trùng lặp dữ liệu lặp lại đến từ nhiều nguồn dữ liệu.

Các Công cụ ETL này phải đối phó với các thách thức về tính không đồng nhất của Cơ sở dữ liệu & Dữ liệu.

1. **Metadata**

Metadata là dữ liệu về dữ liệu xác định Data Warehouse. Nó được sử dụng để xây dựng, duy trì và quản lý Data Warehouse.

Trong Kiến trúc Data Warehouse, metadata đóng một vai trò quan trọng vì nó chỉ định nguồn, cách sử dụng, giá trị và tính năng của dữ liệu Data Warehouse. Nó cũng xác định cách dữ liệu có thể được thay đổi và xử lý. Nó được kết nối chặt chẽ với Data Warehouse.

Metadata giúp trả lời các câu hỏi sau:

* Data Warehouse chứa những bảng, thuộc tính và khóa nào?
* Dữ liệu đến từ đâu?
* Dữ liệu được tải lại bao nhiêu lần?
* Sự biến đổi nào đã được áp dụng làm sạch?

Metadata có thể được phân thành các loại sau:

1. Technical Meta Data: Loại metadata này chứa thông tin về kho được sử dụng bởi các nhà thiết kế và quản trị Data Warehouse.
2. Business Meta Data: Loại metadata này chứa thông tin chi tiết giúp người dùng cuối hiểu được thông tin được lưu trữ trong Data Warehouse một cách dễ dàng.
3. **Công cụ truy vấn**

Các công cụ truy vấn cho phép người dùng tương tác với hệ thống Data Warehouse.

Các công cụ này được chia thành bốn loại khác nhau:

1. Công cụ truy vấn và báo cáo

2. Công cụ phát triển ứng dụng

3. Công cụ khai thác dữ liệu

4. Công cụ OLAP

1. **Công cụ truy vấn và báo cáo:**

Các công cụ truy vấn và báo cáo có thể được chia thành:

* Công cụ báo cáo
* Công cụ truy vấn được quản lý

**Công cụ báo cáo:**

Các công cụ báo cáo có thể được chia thành công cụ báo cáo sản xuất và trình viết báo cáo trên máy tính để bàn.

1. Người viết báo cáo: Loại công cụ báo cáo này là những công cụ được thiết kế cho người dùng cuối để phân tích.

2. Báo cáo sản xuất: Loại công cụ này cho phép các tổ chức tạo ra các báo cáo hoạt động thường xuyên. Nó cũng hỗ trợ các công việc hàng loạt khối lượng lớn như in ấn và tính toán.

**Công cụ truy vấn được quản lý:**

Loại công cụ truy cập này giúp người dùng cuối giải quyết các lỗi trong cơ sở dữ liệu và SQL và cấu trúc cơ sở dữ liệu.

1. **Công cụ phát triển ứng dụng:**

Các báo cáo tùy chỉnh được phát triển bằng cách sử dụng các công cụ phát triển Ứng dụng.

1. **Công cụ khai thác dữ liệu:**

Các công cụ khai thác dữ liệu được sử dụng để làm cho quá trình khai thác dữ liệu được tự động hoá.

1. **Công cụ OLAP:**

Các công cụ này dựa trên các khái niệm về cơ sở dữ liệu đa chiều. Nó cho phép người dùng phân tích dữ liệu bằng cách sử dụng các khung nhìn đa chiều kỹ lưỡng và phức tạp.

1. **Kiến trúc Data warehouse Bus**

Data Warehouse Bus xác định luồng dữ liệu (data flow) trong kho của bạn. Luồng dữ liệu trong Data Warehouse có thể được phân loại thành luồng vào (Inflow), luồng lên (Upflow), luồng xuống (Downflow), luồng ra (Outflow) và luồng meta (Meta flow).

**Nguồn:** [**https://www.guru99.com/data-warehouse-architecture.html**](https://www.guru99.com/data-warehouse-architecture.html)