DATA WAREHOUSE

A data warehouse centralizes and consolidates large amounts of data from multiple sources. The data within a data warehouse is usually derived from a wide range of sources such as application log files and transaction applications. A data warehouse is a type of data management system that is designed to enable and support business intelligence (BI) activities, especially analytics. Data warehouses are solely intended to perform queries and analysis and often contain large amounts of historical data. Because of these capabilities, a data warehouse can be considered an organization's "**single source of truth**."

CHARACTERISTICS OF DATA WAREHOUSE:

- **Subject-oriented:** They can analyse data about a particular subject or functional area (such as sales).
- **Integrated:** Data warehouses create consistency among different data types from disparate sources.
- **Non-volatile:** Once data is in a data warehouse, it's stable and doesn't change.
- **Time-variant:** Data warehouse analysis looks at change over time.

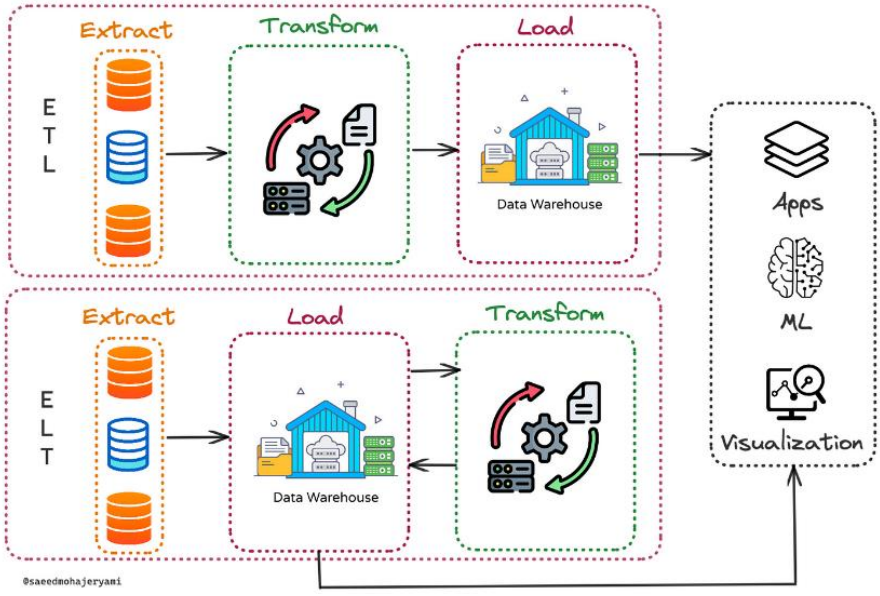
COMPONENTS OF DATA WAREHOUSE:

The main components of a data warehouse include:

1) Data Sources: These are the various operational systems, databases, and external data feeds that provide raw data to be stored in the warehouse.

2) ETL (Extract, Transform, Load) Process: The ETL process is responsible for extracting data from different sources, transforming it into a suitable format, and loading it into the data warehouse.
ELT (Extract, Load, Transform) is more advanced than ETL (Extract, Transform,

Load) primarily because it leverages the processing power of modern cloud data warehouses to handle transformations after data is loaded, rather than before.

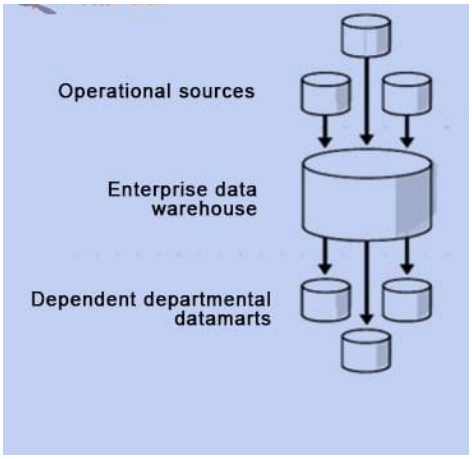


3) Data Warehouse Database: This is the central repository where cleaned and transformed data is stored. It is typically organized in a multidimensional format for efficient querying and reporting.

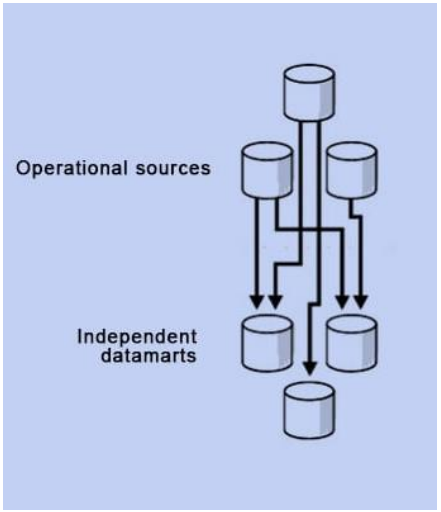
4) Metadata: Metadata describes the structure, source, and usage of data within the warehouse, making it easier for users and systems to understand and work with the data.

5) Data Marts: These are smaller, more focused data repositories derived from the data warehouse, designed to meet the needs of specific business departments or functions.

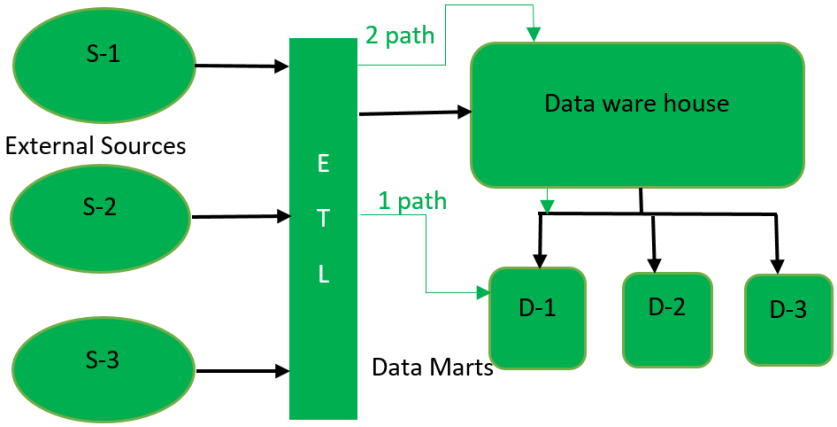
1. Dependent Data Mart - Dependent Data Mart is created by extracting the data from central repository, Datawarehouse.



2. Independent Data Mart - Independent Data Mart is created directly from external sources instead of data warehouse. This model of data mart is used by small organizations and is cost effective comparatively.



3. Hybrid Data Mart - This type of Data Mart is created by extracting data from operational source or from data warehouse. 1Path reflects accessing data directly from external sources and 2Path reflects dependent data model of data mart.



6) OLAP (Online Analytical Processing) Tools: OLAP tools allow users to analyse data in multiple dimensions, providing deeper insights and supporting complex analytical queries.

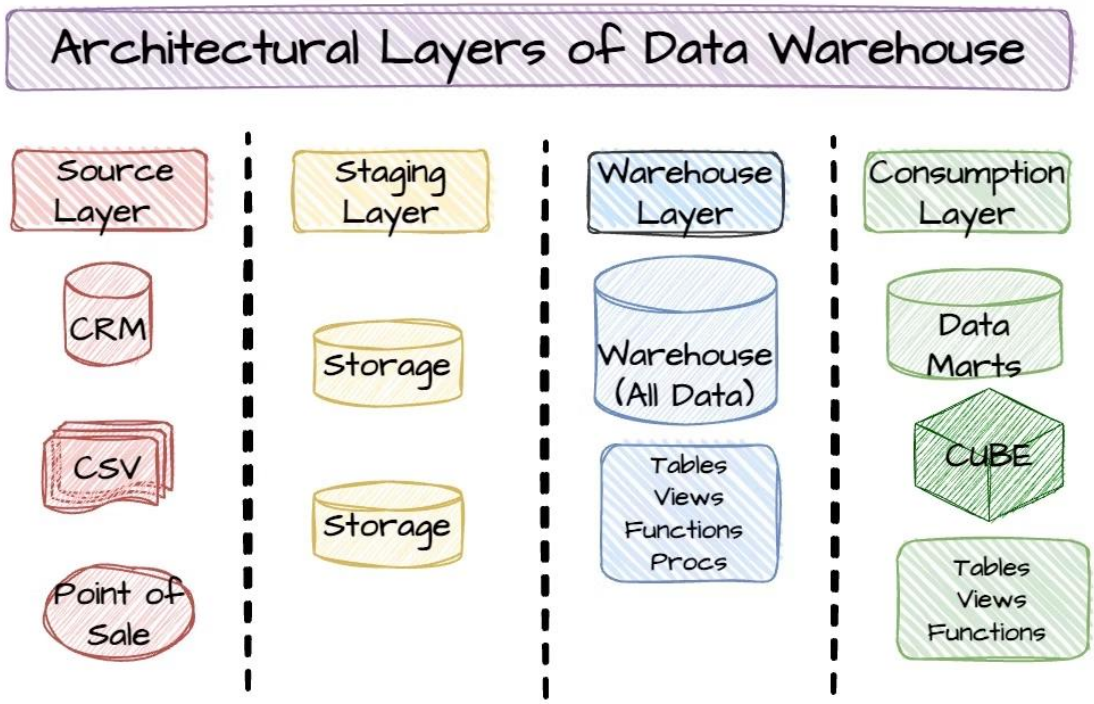
7) End-User Access Tools: These are reporting and analysis tools, such as dashboards or BI tools, that enable business users to query the data warehouse and generate reports.

DATA WAREHOUSE ARCHITECTURE:

The architecture of a data warehouse is determined by the organization's specific needs. Common architectures include

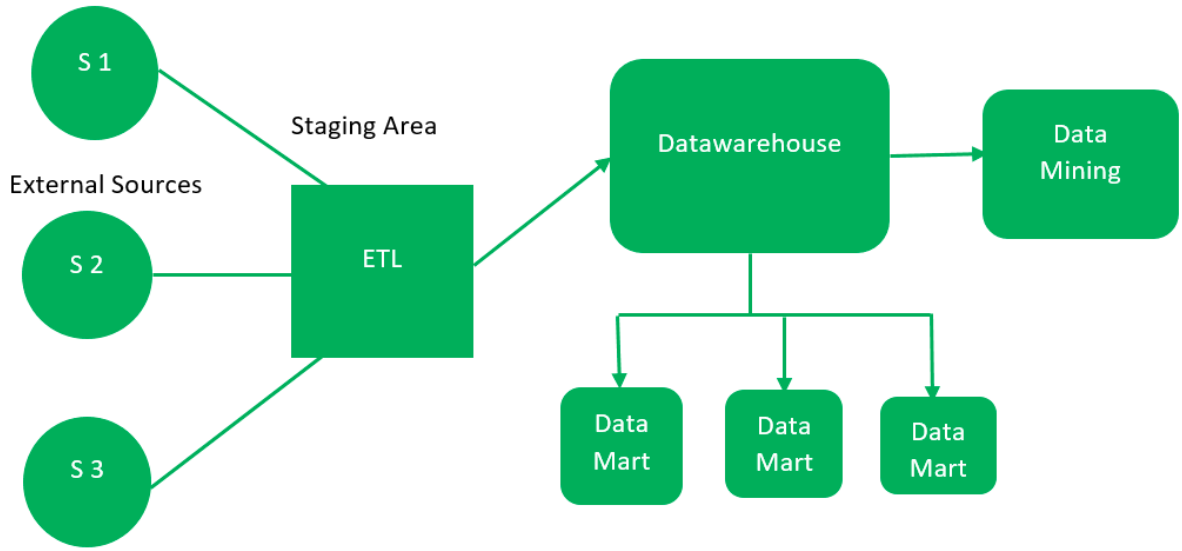
- **Simple:** The repository is fed by data sources on one end and accessed by end users for analysis, reporting, and mining on the other end.
- **Simple with a staging area:** Operational data must be cleaned and processed before being put in the warehouse. Data warehouses add a staging area for data before it enters the warehouse, to simplify data preparation.

- **Hub and spoke.** Adding data marts between the central repository and end users allows an organization to customize its data warehouse to serve various lines of business. When the data is ready for use, it is moved to the appropriate data mart.
- **Sandboxes.** Sandboxes are private, secure, safe areas that allow companies to quickly and informally explore new datasets or ways of analysing data without having to conform to or comply with the formal rules and protocol of the data warehouse.

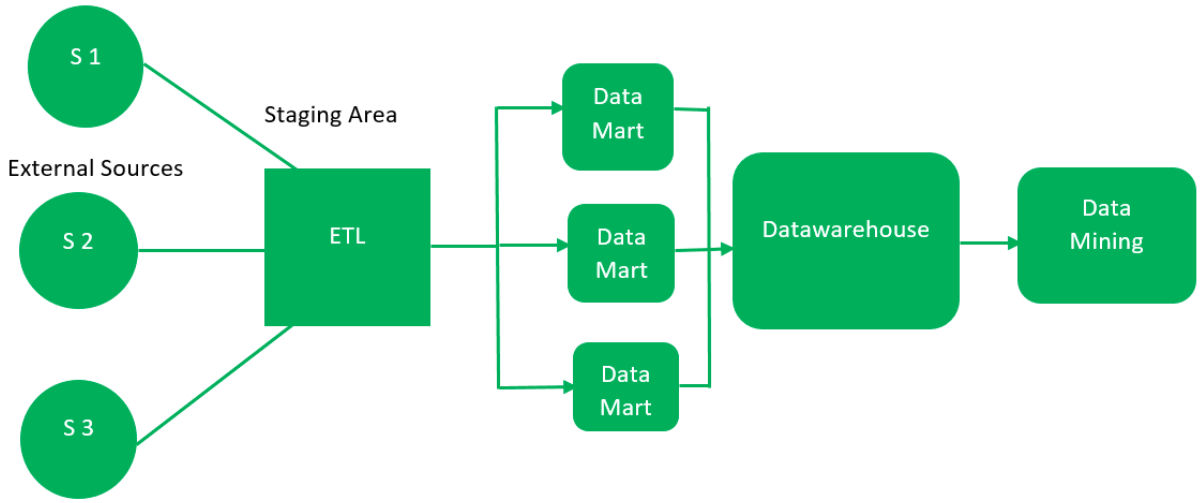


There are two common approaches to constructing a data warehouse:

- **Top-Down Approach:** This method starts with designing the overall data warehouse architecture first and then creating individual data mart.



- **Bottom-Up Approach:** In this method, data marts are built first to meet specific business needs, and later integrated into a central data warehouse.



REAL-WORLD EXAMPLE:

Seimens, a global leader in industrial manufacturing, implemented a company-wide data warehouse called "**One Siemens.**" This system integrates data from various business units and manufacturing facilities worldwide.