

Data Warehouse

Name: PRIYESHWAR M D

Mail: privesh2664@gmail.com

An introduction to Data Warehousing:

Data warehousing • Data Warehouse (DW) is a Subject oriented, integrated, time variant, non-volatile collection of data in support of management's system.

It is a collection of data designed to support management decision making by presenting a coherent picture of business conditions at a single point of time.

Subject-oriented: Data are organized according to the subject instead of application. It mainly focuses on modeling and analysis of data for decision makers, not on daily operations or transaction processing.

Integrated: Constructed by integrating multiple, heterogeneous data sources like relational databases, flat files, on-line transaction records. Ensure consistency in naming conventions, encoding structures, attribute measures, etc. among different data sources.

Time-variant: The time horizon for the data warehouse is significantly longer than that of operational systems. i.e. provide information from a historical perspective (e.g., past 5-10 years).

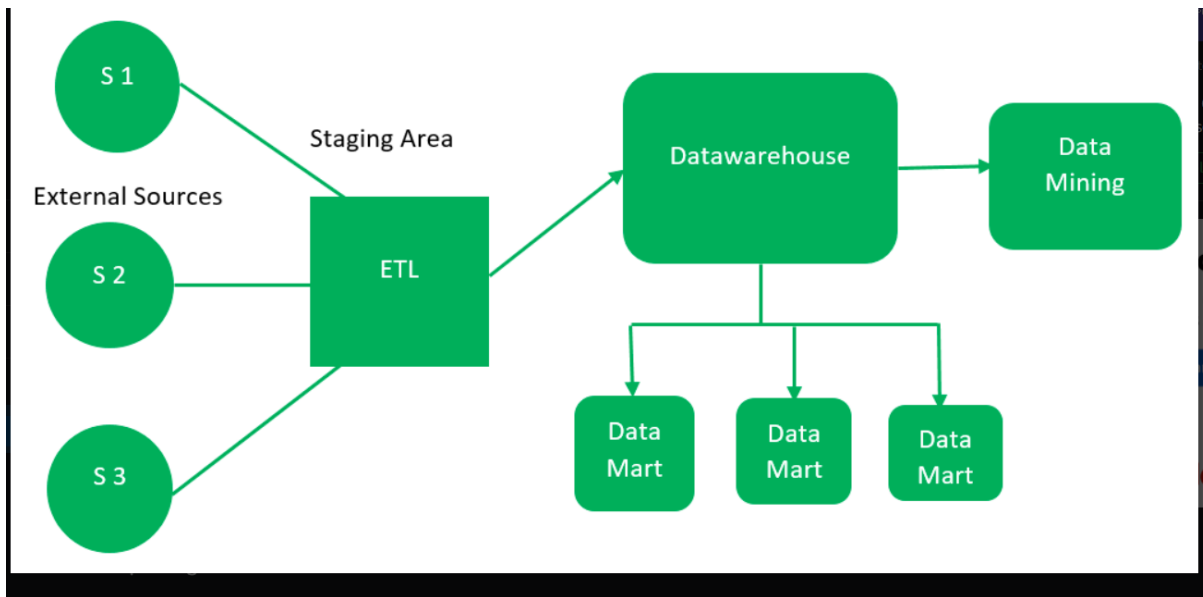
Non-volatile: No updates are allowed. Once the data entered into the data warehouse, they are never removed. The data in warehouse represent company's history, the operational data representing near term history are always added to it.

Purpose of Data Warehouse

- **Data Consolidation:** Integrates data from multiple heterogeneous sources.
- **Historical Analysis:** Supports time-based analysis and trend tracking.
- **Business Intelligence:** Enables reporting, dashboards, KPIs, and data mining.
- **Improved Performance:** Optimized for analytical queries, not transactions.
- **Data Quality & Consistency:** Cleansed, transformed, and standardized data.

Data Warehouse Architecture

1. Source Layer (where raw data comes from)
2. Staging Layer (data is cleaned here)
3. Data Storage Layer - Data WareHouse (organized and stored in warehouse)
4. Presentation Layer (final form for users or analysts).



Operational Data Store

An ODS is a type of database that serves as an intermediary between operational systems and the data warehouse. It contains current, up-to-date data and is used for real-time reporting and operational decisions.

- Short-term storage
- Supports frequent updates
- Not optimized for deep historical analysis

OLTP Vs Warehouse Applications

OLTP (Online Transaction Processing) is used for day-to-day tasks like ATM transactions, online orders, etc. It handles many small operations quickly.

OLAP (Online Analytical Processing) is used for analysis, reports, and insights. It works on large chunks of historical data. This is one of the methods used in the warehouse application. It helps doing multi dimensional analysis and report generation.

Feature	OLTP	Data Warehouse (OLAP)
Main Function	Transactional processing	Analytical processing
Data	Real-time, current	Historical, aggregated
Query Type	Simple, fast (CRUD)	Complex, slow (aggregations, joins)
Schema Design	Highly normalized	Denormalized (star/snowflake)
Use Cases	Day-to-day operations	Reporting, forecasting, decision-making

Data Marts

A Data Mart is a subset of a data warehouse focused on a specific business function or department (e.g., Sales, HR).

- Easier and faster to implement
- Can be dependent (from a data warehouse) or independent (from source systems)
- Enables focused analytics for a specific user group

Data marts Vs Data Warehouses

Feature	Data Mart	Data Warehouse
Scope	Departmental	Enterprise-wide
Size	Smaller	Larger
Complexity	Lower	Higher
Source	Can be from DW or direct	Multiple operational systems
Time to Deploy	Shorter	Longer

Data Warehouse Life cycle

The DW Life Cycle includes all the stages from planning to using the data warehouse:

1. Planning - why and what data is needed.
2. Data Collection - getting data from sources.
3. Data Cleaning & ETL - preparing and transforming the data.
4. Storage - storing it properly.
5. Access & Analysis - using it for reports and business decisions.
6. Maintenance - updating and improving regularly.

