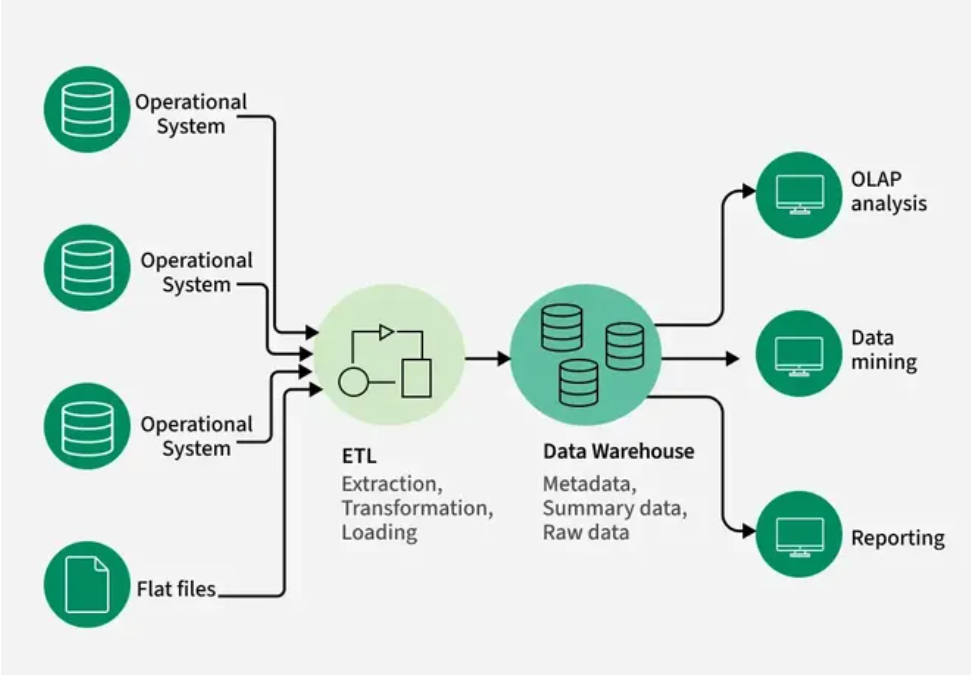
15.05.2025 ASSIGNMENT 1

**Data Warehousing**

A data warehouse is a centralized system used for storing and managing large volumes of data from various sources. It is designed to help businesses analyze historical data and make informed decisions. Data from different operational systems is collected, cleaned, and stored in a structured way, enabling efficient querying and reporting.

* Goal is to produce statistical results that may help in decision-making.
* Ensures fast data retrieval even with the vast datasets.



**Needs of data warehousing:**

**1. Centralized Data Storage:** Combines data from various sources into a single repository for easier access and management.

**2. Improved Decision-Making:** Supports business intelligence by providing accurate and timely information for strategic planning.

**3. Historical Data Analysis:** Maintains historical data to identify trends and patterns over time.

**4. Faster Query Performance:** Optimized for running complex queries and generating reports quickly, even with large datasets.

**5. Data Integration:** Merges data from different departments or systems to provide a unified and comprehensive view.

**How Datawarehouse works?**

A Data Warehouse works as a central repository where information arrives from one or more data sources. Data flows into a data warehouse from the transactional system and other relational databases.

Data may be:

1. Structured
2. Semi-structured
3. Unstructured data

The data is processed, transformed, and ingested so that users can access the processed data in the Data Warehouse through Business Intelligence tools, SQL clients, and spreadsheets. A data warehouse merges information coming from different sources into one comprehensive database.

By merging all of this information in one place, an organization can analyze its customers more holistically. This helps to ensure that it has considered all the information available. Data warehousing makes data mining possible. Data mining is looking for patterns in the data that may lead to higher sales and profits.

**components of a data warehouse**

**1. Data Source:**

Raw data originates from various operational systems such as databases, CRM, ERP, flat files, and external sources.

**2. Data Staging Area (ETL):**

A workspace where data is Extracted, Transformed, and Loaded. It cleanses, integrates, and prepares data before loading it into the warehouse.

**3. Data Storage (Data Warehouse Repository):**

The core storage component where structured and historical data is kept for querying and analysis.

**4. Data Mart:**

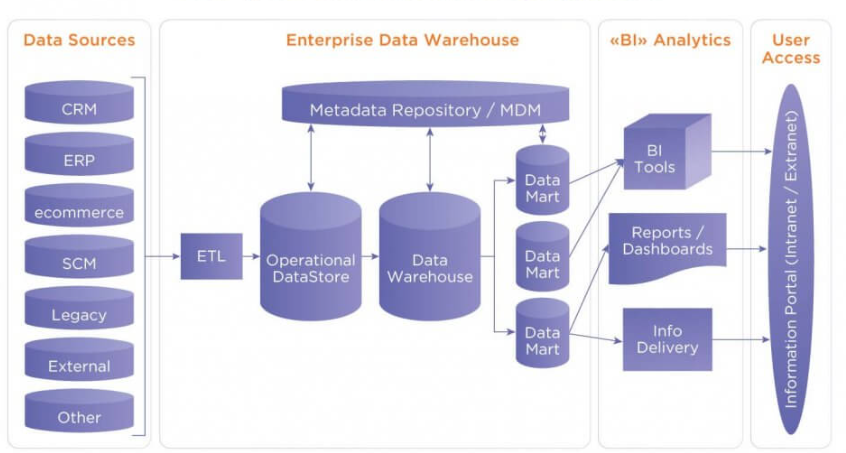
A subset of the data warehouse, tailored for specific business lines or departments (e.g., sales, finance). It provides focused and faster access to relevant data.

**5. Metadata:**

Provides information about the data—such as definitions, data types, source systems, and data lineage—helping users understand and navigate the warehouse.

**6. Query and Reporting Tools (Access Layer):**

Interfaces and tools that allow users to access, analyze, visualize, and generate reports from the data warehouse and data marts.



**characteristics of data warehousing:**

**1. Subject-Oriented:** Organized around key subjects like customers, products, or sales rather than daily operations.

**2. Integrated:** Data from multiple sources is combined into a consistent format.

**3. Time-Variant:** Stores historical data to analyze trends over different time periods.

**4. Non-Volatile:** Once data is entered, it is stable and not frequently changed or deleted.

**5. Supports Decision-Making:** Designed primarily to help with complex queries, reporting, and business analysis.

**Types of Data Warehouses**

1. **Enterprise Data Warehouse (EDW)**: A centralized warehouse that stores data from across the organization for analysis and reporting.
2. **Operational Data Store (ODS)**: Stores real-time operational data used for day-to-day operations, not for deep analytics.
3. **Data Mart**: A [subset](https://www.geeksforgeeks.org/data-marts-storage-component-of-hdfs/) of a data warehouse, focusing on a specific business area or department.
4. **Cloud Data Warehouse**: A data warehouse hosted in the cloud, offering scalability and flexibility.
5. **Big Data Warehouse**: Designed to store vast amounts of unstructured and structured data for big data analysis.
6. **Virtual Data Warehouse**: Provides access to data from multiple sources without physically storing it.
7. **Hybrid Data Warehouse**: Combines on-premises and cloud-based storage to offer flexibility.
8. **Real-time Data Warehouse**: Designed to handle real-time data streaming and analysis for immediate insights.

**Uses of Data Warehouse:**

Here, are most common sectors where Data warehouse is used:

**Airline:**

In the Airline system, it is used for operation purpose like crew assignment, analyses of route profitability, frequent flyer program promotions, etc.

**Banking:**

It is widely used in the banking sector to manage the resources available on desk effectively. Few banks also used for the market research, performance analysis of the product and operations.

**Healthcare:**

Healthcare sector also used Data warehouse to strategize and predict outcomes, generate patient’s treatment reports, share data with tie-in insurance companies, medical aid services, etc.

**Public sector:**

In the public sector, data warehouse is used for intelligence gathering. It helps government agencies to maintain and analyze tax records, health policy records, for every individual.