**Data Warehouse**

A structured, non-volatile single source of truth for a company. It is a centralized place that stores large volumes of Structured Data from Multiple Sources.

In Information Systems Theory the Single Source of Truth is the practice of structuring all the best quality data in one place.

eg, One file and its different versions. We need to name these versions differently.

Data warehouse helps find out the single source of truth for avoid naming the different versions of the same file. Companies use this Data Warehouse to store their valuable data, like employee data, sales data, customer data. The data is used for reporting and analysis purposes.

**Features:**

1. **Subject Oriented:**

For a company in Data Warehouse Sales Data is more meaningful than Competitors Data. Company likely to store Data related to specific interest.

1. **Integrated:**

Every person or team can have access to the Data and naming the file as per their preferences as per the previously mentioned example. Stores Data form multiple sources through ETL Process.

1. **Time Variant:**

Data Warehouse contains Historical Data for Analysis and Reporting purposes.

1. **Non-Volatile:**

Once the Data is stored in the Data Warehouse, it can’t be changed or deleted.

1. **Summarized:**

Data is used for Data Analytics. Often used some functionalities (like, aggregation, filter, segmentation etc.) to facilitate analysis and reporting.

**Stages (AI Ladder):**

Collect ----> Organize & Store ----> Analyze ----> Consume/ Infuse

Different Data Sources ----> Data Lake ----> Data Warehouse ----> Data Mart ----> Data for Analysis and Consume Purposes.

**Importance and Benefits**:

Data warehouses enhance performance by consolidating data, enabling comprehensive analysis, and supporting business intelligence efforts for better strategic planning and management.

**Users:**

1. Business Analyst.
2. Data Scientist.
3. Data Engineer.

**Tools:**

1. BI Tools.
2. PAML (Predictive Analytics Machine Learning).

**Depletion Places:**

1. **Organization Premises:**
2. For Commodity Hardware through MPP (Massively Parallel Processing) architecture or SMP (Symmetric Multi-Processing) architecture.
3. For Purpose Build Appliance Format.
4. **Cloud Based:**

Data warehouse is delivered as a managed SAAS offering from the data providers who stored their data in a cloud format.

1. **Hybrid Mode: (On Prem + Cloud)**

Premises and Cloud Data are both used for flexibility and recovery purposes.

**Applications:**

1. For Dashboard Preparation.
2. For Application Building.
3. For Automation Purposes.

After applying the data in the real world to solve some specific problems, new possible solutions and features come forward which will be again feed into the Data Lake section where we do some necessary filtration to store the data into the Data Warehouse again. In this way the Cycle continues.

**Data Lake**

An ingestion process is required to put the data in the data warehouses. Because it is impossible to analyze with the actual data from the source, a copy of the data (Meta Data) stored in the Data Warehouse after filtration can be used properly for purposeful analysis. It ensures flexibility.

**Data Lake** is where structured, unstructured, semi-structured, raw data are stored, and after ETL processing, the raw data is transformed into high quality data, stored in the Data Warehouse.

It is a step where we can store huge amounts of data in a cost-effective way. But there are some problems regarding Data Governance and Data Quality.

Also, Data Warehouse is good at Data Governance, but it is highly cost effective because some applications always need fresh data to do their work. It takes time to sort, clean and load data. Developers take the best quality data from Data Lake and Data Warehouse and combine it through new technology and store it in a place called Data Lakehouse.

Here the data is flexible, cost effective, structured and well performed. Here Data Governance takes into place to give quality structure. Now this data is used for BI Visualizations and AI & ML purposes as per business requirements.