

# What is Data Warehouse ? Definition & Overview

## What is Data Warehouse?

---

Data Warehouse is nothing but relational database management system which is used for Querying the data for the purpose to do some analysis and to take some managerial decisions.

The definition for Data Warehouse (DWH) is collecting / Integrating data from different sources and converting that data into Information format for the purpose of taking managerial decisions.

## Data Warehouse definition?

---

Data Warehouse is nothing but subject oriented, time variant, Integrated, history data and non volatile collection of data to do some analysis and to take some managerial decisions.

According to **Bill Inmon**, Data warehouse is a Subject-Oriented, Integrated, time-variant and non-volatile collection of data.

According to **Ralph Kimball**, Data Warehouse is a transaction data specifically structured for query and analysis.

So what is Data ? Data is a collection of raw material in unorganised format. Now we have to convert that data into Information format. To make decision, we need to collect the data, using that data we get some information and finally we take decision.



**Example:-** In an organisation, we have many departments like Sales dept, Product dept, Hr department and many other. Before releasing any product to the market, CEO collects the data form the Sales department and product department to take some decisions on profits & losses.

In an Organisation, there are several department available and each individual department perform different kind of transactions, all these transactions are saved in Operational data store (ODS).

The main characteristics of ODS is data is volatile and it doesn't maintain any history data. So what is volatile ? Data in volatile means, the data changes in regular interval of time.

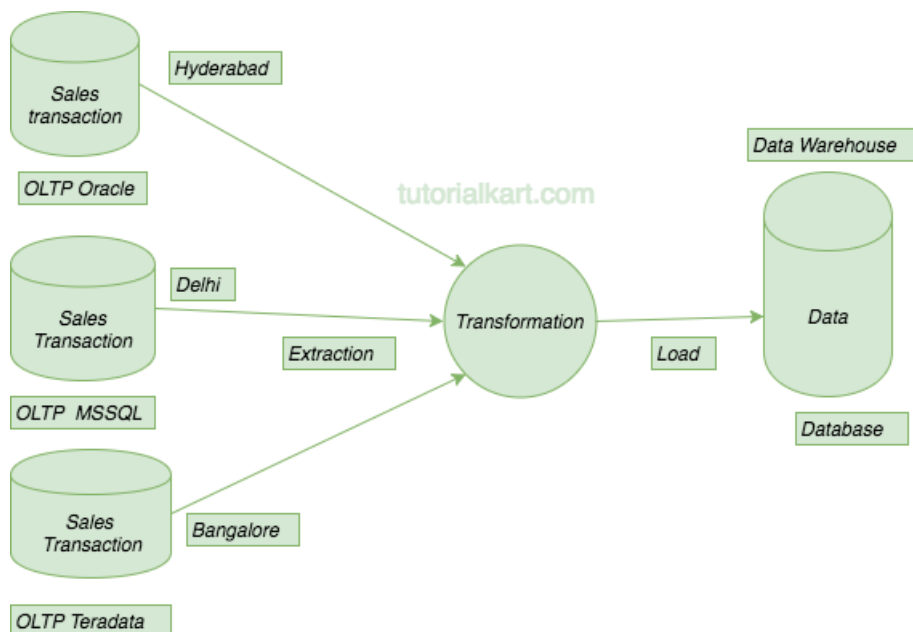
**Example**:- Big Bazaar, CEO needs to take decision about a particular product. So he needs 3 to 5 years of data. But in ODS, it doesn't maintain any history data. So, every organisation should maintain history data to take decisions based on product sales.

## Why Data Warehousing is required?

An enterprise or organisation is a business unit where every department is integrated with various units. For example in Banking, we have different business units like Accounts, Loans, Mutual funds etc. Every business transaction is an operation and these online transaction are stored in Operational database or OLTP (Online Transactional Processing Database).

*Data Warehouse (DW)* centralises data from multiple Operational Databases (OLTP's) because data is scattered in various places and it becomes difficult in gathering data.

- Using Data Warehousing, we can create DWH tables.
- We can get the data from Operational data store (ODS).
- Data is not volatile and DWH maintains history data.
- Data changes in particular interval of time.



## Characteristics of Data Warehouse (DWH).

- **Non volatile** :- Data changes in particular interval of time and also changes in regular interval of time.
- **History data** :- DWH stores the history data and current data.
- **Integrated** :- Integration of data from different types of sources can be made possible using different Data Warehousing tools.
- **Time variant** :- We can load the data in particular intervals of time and also we can retrieve the data in particular interval of time.

- **Subject Oriented** :- We will have many departments in an organisation. Different subject areas are there in an single database.

## Relational Database types.

---

- **Low Range Database** :- Low range Database can manage and organise mega bytes of data. Ex :- Ms-Access, D-Base, Fox PRO, Chipper, Paradox etc.
- **Mid Range Database** :- Mid Range Database can manage and organise Giga bytes of data. Ex : – Oracle, Microsoft SQL Server, IBM DB2 etc.
- **High Range Database** :- High Range Databases can organise and manage Tera-Bytes of data. Ex:- TD, Netezza, Green Plum etc.

## How to choose Database?

---

To choose right Database that required for your organisation, it depends upon the size of the organisation, Transaction frequency per day and No of year's of transaction history to be provided.

## How to build Data Warehouse?

---

To build Data Warehouse, the database must support the following features. Check the following requirements before building DWH.

- A Relational Database is required.
- Database that support enormous storage capacity like billions of Rows and Tera-Bytes of Data.
- DB that supports parallel processing, Distributed file storage (DFS) pattern.
- Database that supports high availability, where data is available to the users even the software and hardware compounds are down.
- We must select Database that have “mature optimisers” and low TCO (Total Cost of Owner ship).

### Data Warehousing

- ◆ [Data Warehouse Tutorial](#)
- ◆ [What is Data Warehousing](#)
- ◆ [Data Warehouse - Characteristics](#)
- ◆ [Data Warehouse - Architecture](#)
- ◆ [Data Warehouse - Data mart](#)
- ◆ [Data Warehouse - Star Schema](#)
- ◆ [Data Warehouse - Snow Flake Schema](#)

◆ Data Warehouse - Confirmed Dimension

◆ Data Warehouse - Junk Dimension

◆ Data Warehouse - Slowly Changing Dimension

◆ Data Warehouse - Degenerate Dimension

◆ Data Warehouse - Role-playing Dimension

◆ Data Warehouse - ETL Concepts