

Definition:

A data warehouse is a subject oriented, integrated, non-volatile and time variant collection of data in support of managements decisions.

other definition

The data in data warehouse is? **(SAT IN A)**

- Separate
- Available
- Integrated
- Time stamped
- Subject oriented
- Accessible
- Non-volatile

Defining features:

1) → Subject - Oriented Data

- In operational systems, we store data by individual applications.

BUT

- In data sets, we keep data for that particular application.

Ex: Order processing application.

These data sets provide the data for all functions for entering orders, checking stocks, etc.

∴ These data set contains only the data that is needed for those functions relating to this particular application.

" In data warehouse, data is not stored by operational applications, but by business subjects "

- Operational App → Order processing, Customer loans etc.
- Data warehouse subjects → Sales, Product, Customer, Account, Claims, Policy -

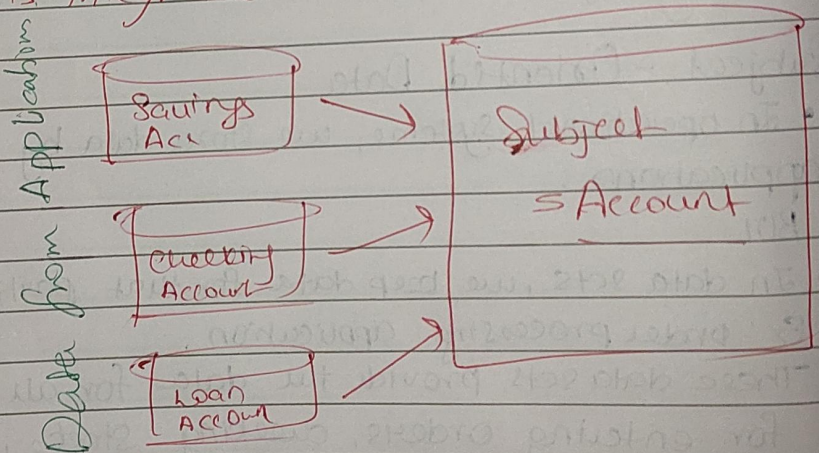


## 2) Integrated

Before the data from various disparate sources can be usefully stored in a data warehouse, you have to remove inconsistencies. You have to standardize the various data elements and make sure of the meanings of data names in each source application.

→ Before moving the data into data warehouse you have to go through a process of transformation, consolidation and integration of source data.

"Data in consistencies are removed, data from diverse operational applications is integrated" Data warehouse Subj



## 3) Time variant data

The time variant nature of the data in data warehouse

- Allows for analysis of past
- Relates info to the present
- Enables to ~~for~~ forecast future

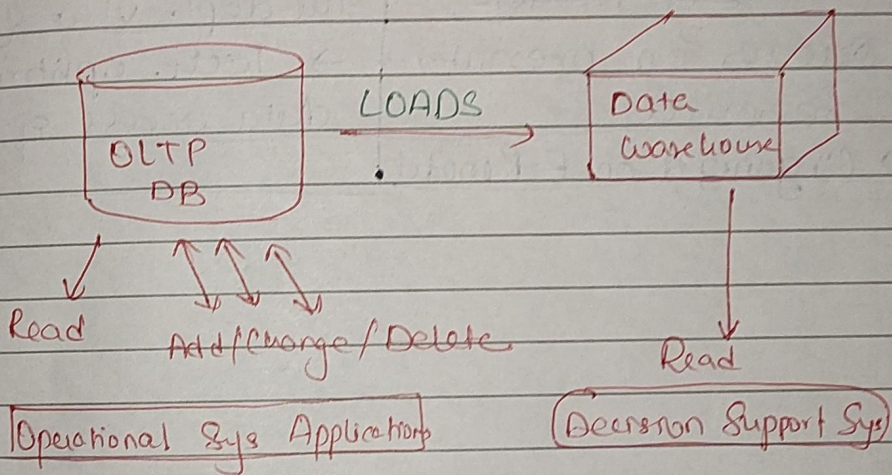


Non volatile Data.

The data in data warehouse is not as volatile as data in operational db.

The data in data warehouse is primarily for query and analysis.

"Usually the data in data warehouse is not updated or deleted".



Data Granularity - level of Detail.

→ In operational sys, data is usually kept the lowest level of detail

→ In data warehouse, user usually starts by looking summary of data.

→ Depending on query one can go into particular level of detail to satisfy the query.

→ Many ~~data~~ data warehouse have at least dual level of granularity.



## Data warehouse VS Data mart

### Data Warehouse

- Corporate/ Enterprise-wide
- Union of all data marts
- Data retrieved from staging area.
- ~~Structure for corporate view~~  
~~Organized on data~~  
E-R model
- Focus on presentation resource
- Organized on E-R model

### Data mart

- Departmental
- A single business process
- Star join (facts & dimensions)
- Structure to suit the dept. view of data
- Tech. optimal for data access & analysis



## **Bottom-Up Approach**

The advantages of this approach are:

- Faster and easier implementation of manageable pieces
- Favorable return on investment and proof of concept
- Less risk of failure
- Inherently incremental; can schedule important data marts first
- Allows project team to learn and grow

The disadvantages are:

- Each data mart has its own narrow view of data
- Permeates redundant data in every data mart
- Perpetuates inconsistent and irreconcilable data
- Proliferates unmanageable interfaces

In this bottom-up approach, you build your departmental data marts one by one. You would set a priority scheme to determine which data marts you must build first. The most severe drawback of this approach is data fragmentation. Each independent data mart will be blind to the overall requirements of the entire organization.

# Top-Down Versus Bottom-Up Approach

## ***Top-Down Approach***

The advantages of this approach are:

- A truly corporate effort, an enterprise view of data
- Inherently architected—not a union of disparate data marts
- Single, central storage of data about the content
- Centralized rules and control
- May see quick results if implemented with iterations

The disadvantages are:

- Takes longer to build even with an iterative method
- High exposure/risk to failure
- Needs high level of cross-functional skills
- High outlay without proof of concept

This is the big-picture approach in which you build the overall, big, enterprise-wide data warehouse. Here you do not have a collection of fragmented islands of information. The data warehouse is large and integrated. This approach, however, would take longer to build and has a high risk of failure. If you do not have experienced professionals on your team, this approach could be dangerous. Also, it will be difficult to sell this approach to senior management and sponsors. They are not likely to see results soon enough.