

TOPIC NAME : Data Warehouse

DAY :

TIME :

DATE : 10 / 02 / 2024

Why do we need a data warehouse?

There are two purposes of how and why we use data in the company.

- Operational Purposes

- Receive orders
 - React to complaints
 - Fill up stocks
- (OLTP)

- Analytical Decision Making

- What's best category
 - How many sales compared to last month.
 - How can we improve
- (OLAP)

OLTP: (Online Transaction Processing)

- One record at a time
- Data input (New data comes)
- Don't keep any long history of data

OLAP: (Online Analytical Processing)

- Can get a thousands of records at a time
- Fast query performance
- Store long historical data
- Used for data analysis and reporting

DWH is there to address those analytical data needs.

TOPIC NAME : _____

DAY : _____

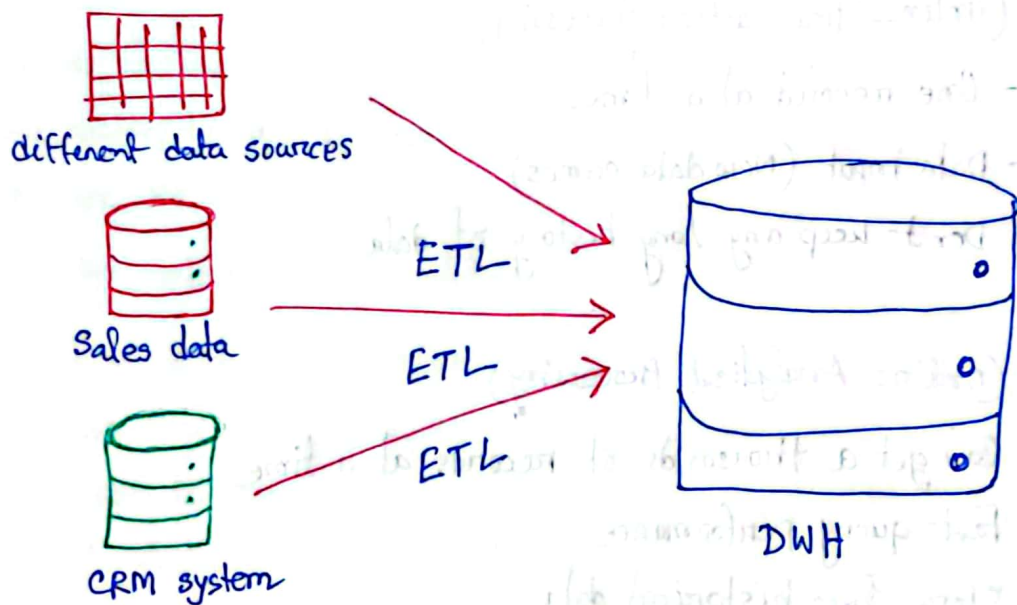
TIME : _____ DATE : / /

What is a data warehouse?

A database that is used and optimized for analytical purpose.

Important things about data warehouse

- User friendly.
- Fast query performance
- Enabling data analysis



From different format of data sources, different structures of data, DWH bring all the data that is important and store that at a centralized location.

The process using which DWH do that, is called ETL process.

TOPIC NAME : _____

DAY : _____

TIME : _____

DATE : / /

ETL

Extract

Transform

Load

Extract: We extract data from different sources

Transform: We transform the data according to our desired format, validate and verify data, then aggregate together to be in a same platform. (Consistent and repeated process)

Load: We load the transformed data in a centralized location which is the data warehouse. Then the data is ready for analysis.

Goals of DWH:

- ① Centralized and consistent location for data
- ② Data must be accessible fast (Query performance)
- ③ User-friendly (easy to understand)
- ④ Must load data consistently and repeatedly (ETL)
- ⑤ Reporting and Data Visualization built on top.

What is BI? And why we need DWH for it?

In BI we gather raw data and transform them in such a way that we can turn into meaningful insights. We do that by data visualization, reportings etc. We do that to understand a company better and take important decision for the company.

TOPIC NAME : _____

DAY : _____

TIME : _____ DATE : / /

DWH is a very important component for BI as we use the DWH for the storage purpose. It also act as the centralized location for structured and transformed data using which we can directly do our analysis.

Data Lake VS Data Warehouse?

Data lake and DWH Both work as a centralized data storage.

① In data lake we store raw data,
where in DWH, we store processed data.

② In data lake the volume of technology data is so large that we use different technology (Big data)

But in DWH, we use database to store data.

③ In data lake, the data is unstructured,
In DWH, data is structured.

④ In data lake, data are not ready for analysis or anything.
But in DWH, data is ready to use.

⑤ The users of Data Lake → Data Scientist
The user of DWH → Business users & IT

We can make datawarehouse from some part of the Data lake.