DATA ENGINEERING TRAINING

What is Db, Datawarehouse, Datamart

Data Eng — Data warehouse Concepts

↓

Here input is

(data)

↓

Source.

→ Data warehouse is Subject oriented, integrated, time variant, non-volatile collection of data in support of management's System

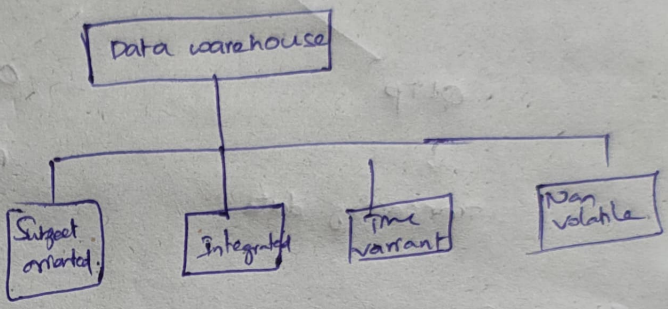
Feature of data warehouse

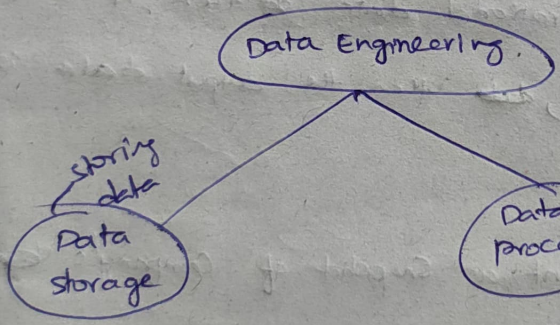
→ Subject - oriented

→ Integrated

→ Time - variant

→ Non - volatile





Processing techniques

1) DSQL, 2) pyspark

↓

Python + Spark

↓

Data processing technique running on Datawarehouse Solutions

SQL

Python

Pyspark

Azure

Analytics

DSS - Decision Support System is very much needed in organizations

structured and unstructured Components are there for DSS

Structured means - when you see that data you can know how to do

unstructured - u need to get help like human interactions to performs the data operations

DSS Architectural Styles :-

→ OLTP (Online transaction processing) — Simplicity, efficiency, data integrity.

- used by traditional operational systems (RDBMS)

→ OLAP (online Analytical processing)

- use by Data warehouse

DSS — Decision support System

OLTP → transacting data

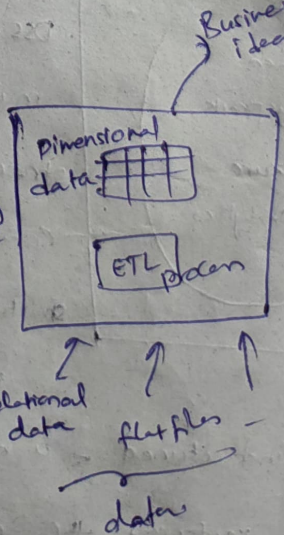
OLAP → measuring data

→ Data is first as operational data then if it had performed some ETL operations, then the output data is Analytical data.

→ Business data represent a snapshot of Company's situation.

ETL :- Extract, Transform — Load

Staging
integration
access layers (Data marts)



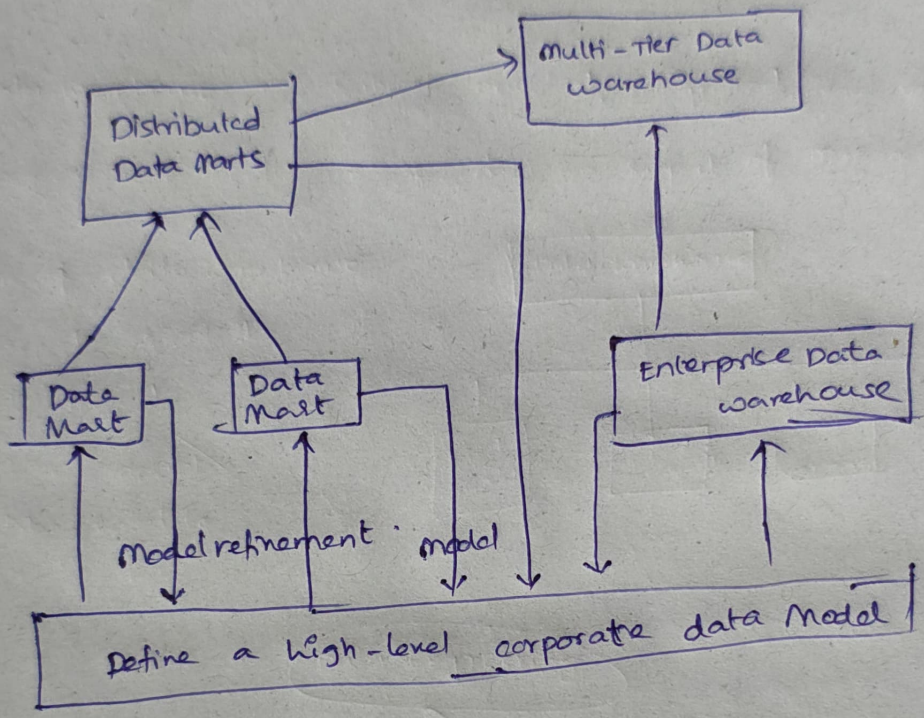
→ The data that arrived at data warehouse are first passed to operational data store (ODS)

→ operational systems are used for DSS

Data Marts :-

→ The data in the data warehouse is stored in the form of data marts.

→ data mart is subset of data warehouse that is usually oriented to specify business teams.



⇒ OLAP data stored in star schema - which is a combination of dimensions and fact tables.

Three kinds of data coarehouse applications :-

1. Information processing - Querying
2. Analytical processing - ETL
3. Data-Mining - Knowledge discovery from hidden pattern.