

Module 3 - Introduction to Data Warehousing and Multi-Dimensional Modelling.

* Operational Support System vs Decisional Support Sys.

O.S.S

D.S.S

① In terms of purpose, O.S.S automates routine tasks and processes.

① In terms of purpose, D.S.S provides data and tools for decision making and analysis.

② In terms of data, O.S.S has structural and transactional data.

② In terms of data, D.S.S has historical and predictive data.

③ The concurrent transaction volume in O.S.S is very high.

③ concurrent transaction levels in D.S.S is medium or low.

④ Regarded as update transactions.

④ Regarded as query transactions. (Read-only)

⑤ O.S.S supports transactions which are happening in real-time

⑤ D.S.S supports transactions which have happened a long time ago to analyze and make decisions on them.

* Data Warehousing

- Data Warehousing is a process of collecting, organizing and storing large amounts of data from various sources in a centralized repository.
- The data warehouse is designed to support business intelligence activities such as data analysis, reporting, and decision-making.

* Features of Data Warehouse

- ① Subject-oriented - A data warehouse is designed around specific subjects or topics related to an organization's business such as sales, customer data and financial data.
- ② Integrated - Data in a Data Warehouse is integrated from multiple sources, including internal systems and external data sources. Helps data stay accurate.
- ③ Time-variant - Data in a data warehouse is stored over time allowing analyst to track changes in data and identify trends over time.
- ④ Non-Volatile - Data in a data warehouse is read-only, so data cannot be updated or changed directly.

* The Need for Data Warehousing

→ There are many reasons why organisations need data warehousing.

- ① Centralized data storage - organizations generally collect data from various sources, including transactional systems, websites, social media, articles etc., making it easier to access and analyze data.
- ② Improved data quality - inconsistent data is a major issue, but since data comes from various sources, the accuracy and quality of data is improved significantly.
- ③ Analytics and reports - Data warehouses are designed for analytics and reporting, with tools and techniques optimized for data analysis and visualization.
- ④ Faster decision-making - Data warehouses provide timely access to the data they need. This is important for organizations working in competitive markets.
- ⑤ Historical analysis - Data warehouses store large amounts of data for a long time, allowing organizations to analyze change in trend patterns over many years.

* Meta data

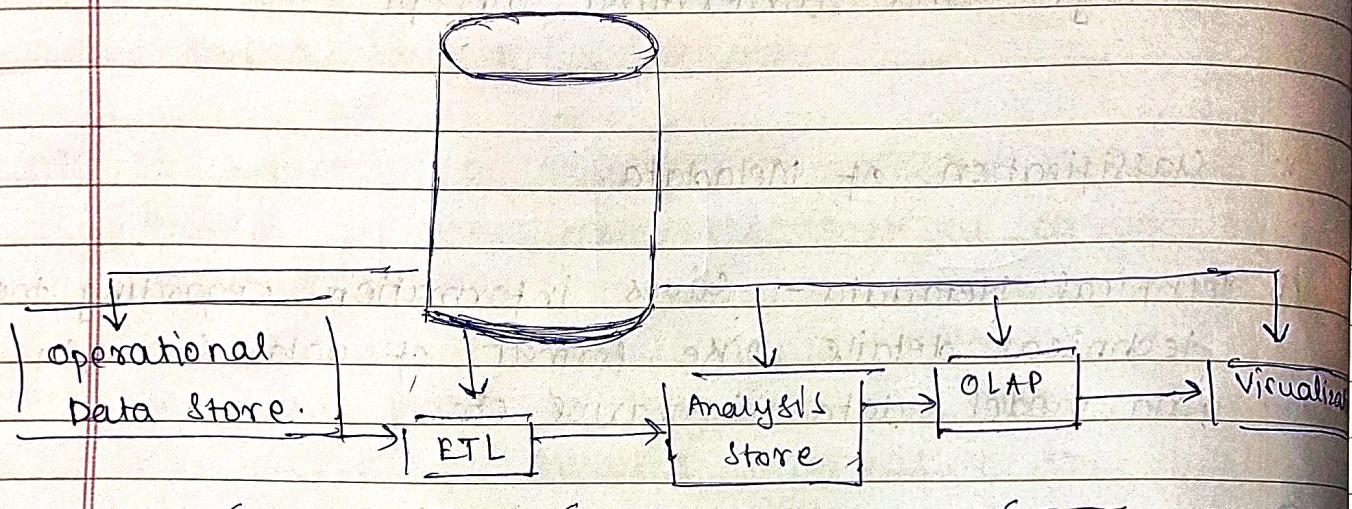
- Meta-Data refers to the info about the data stored in the warehouse.
- It provides a description of data which is stored in the database, such as origin, format, structure.
- Metadata plays a key role in data warehousing by providing valuable information. This helps in
 - ① Understanding the data - Helps user to analyse and report by knowing the origin, format and structure.
 - ② Managing data quality - Metadata also includes information regarding the accuracy, precision and completeness of the information.
 - ③ Data Integration - Metadata helps integrate data from multiple sources ensures data is consistent across the organisation.
 - ④ Data governance - Metadata plays an imp. role in data governance by providing information regarding data ownership, data usage and data security.

④ Data Lineage - Metadata provides info. about the history of data helps users trace the lineage and understand concept.

* Classification of Metadata

- ① Technical Metadata - Gives information regarding the technical details like format of data, storage, data model, data dictionaries etc.
- ② Operational Metadata - Operational metadata gives information regarding data integration, data transformation, data loading and data quality.
- ③ Business Metadata - Contains business concepts represented by data.
- ④ Usage Metadata - Contains info regarding query execution, report generation and data visualisation. used to analyse data.
- ⑤ Data Lineage Metadata - Described the history of the data stored, transformation, date it came into the warehouse etc.

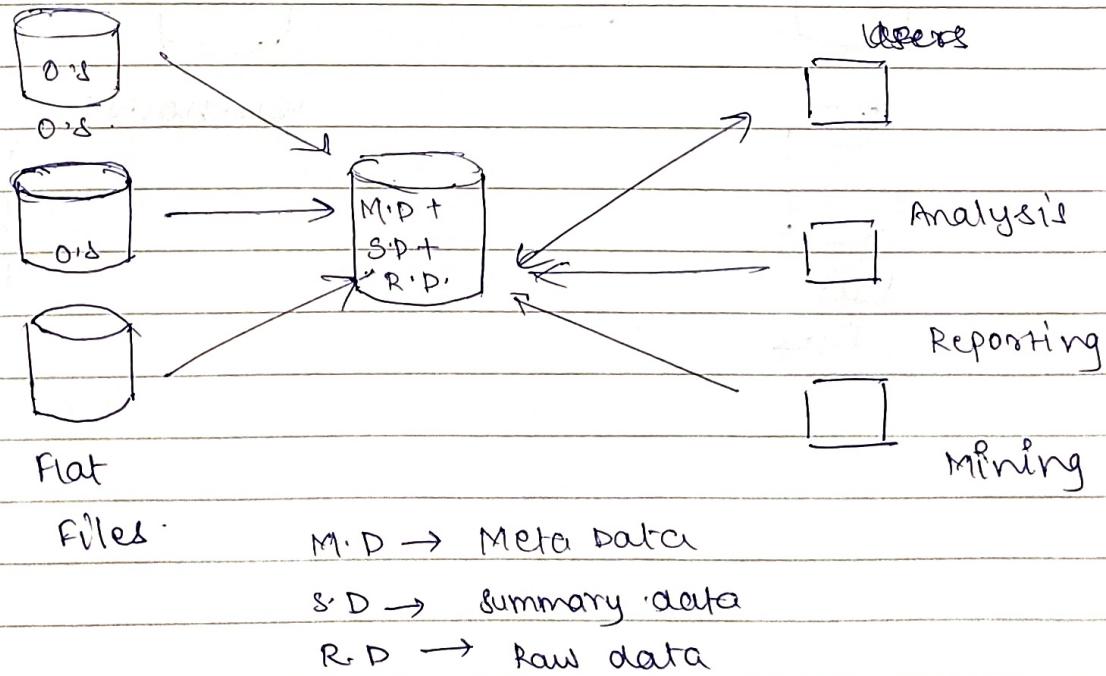
* Information Flow Mechanism and Architecture



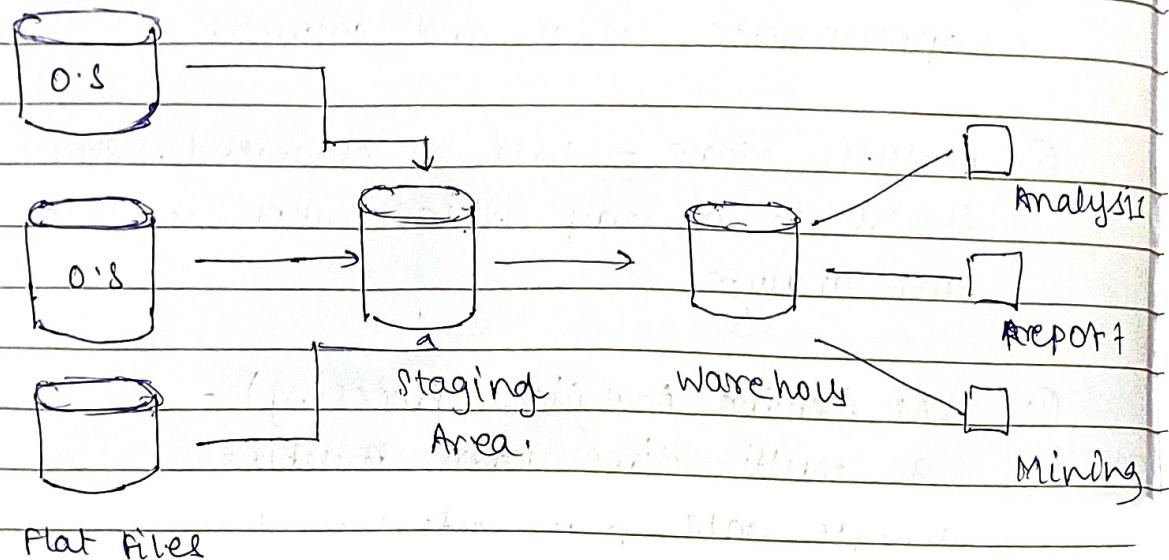
- ① Operational data store (ODS) - ODS is a database which stores the latest and real time data from various operating systems.
- ② ETL - contains 3 steps
 - a) Extract - Data is extracted from ODS and other source systems, such as CRM, ERP etc.
 - b) Transform - The extracted data is then transformed into a format that is suitable for the data warehouse.

- c) Load - transformed data is then loaded onto the data warehouse. This involves putting data into appropriate tables and columns.
- ③ Analysis store - data is organised, structured and stored in a way which makes it easy to query and analyse.
- ④ OLAP (Online Analytical Processing) - organised data in a multi-dimensional structure: this allows users analyse data from different perspectives. OLAP provides tools to analyse trend patterns and anomalies.
- ⑤ Visualisation - creates reports, patterns, trends, insights, dashboard etc using stored data within the data warehouse.

* Architecture (Basic)



* Architecture (with Staging Area)



* Architecture (staging Area and Data Marts)

