**To introduce DWH concepts and its importance in Analytics**

**DAY 1 --- SUMMARY**

An introduction to Data Warehousing:

Data warehouse is used for storing data and is a Subject oriented, integrated, time variant, non-volatile collection of data in support of management’s system.

Purpose of Data Warehouse:

It is a collection of data designed to support management decision making by presenting a coherent picture of business conditions at a single point of time.

Features of Data Warehouse:

1. Subject-oriented: Data is grouped by main topics (like customers or sales) to make it easier to study specific areas of the business.
2. Integrated: Data from different places is combined in one place, in a consistent way, so it all matches up.
3. Time-variant: Data is saved with dates, so you can look at how things change over time.
4. Non-volatile: Data doesn’t change or get deleted once it’s in the warehouse, so you can always rely on it staying the same.

Data Warehouse Architecture :

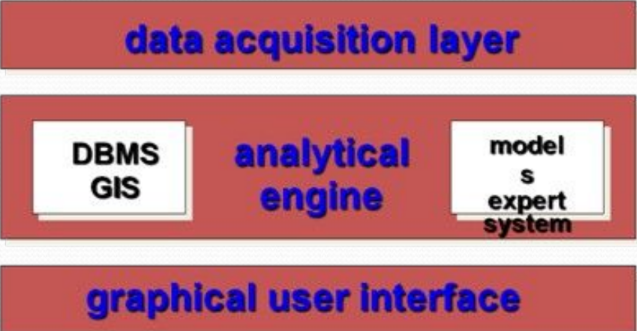
A diagram of data processing

Description automatically generated

What is DSS:

Decision Support Systems helps us to assess and resolve everyday business questions. It works by compiling useful information from a combination of raw data, documents, personal knowledge, or business models.

DSS Architecture:



DSS architectural styles:

1. OLTP (Online Transaction Processing) – used by RDBMS
2. OLAP (Online Analytical Processing) – used by data warehouse

OLTP Vs Warehouse Applications:

* OLTP (Online Transaction Processing): Used for daily business tasks where fast, small transactions happen constantly (like processing sales or banking transactions). It stores detailed, up-to-date data in a way that keeps everything accurate and fast for real-time actions.
* Data Warehouse: Used for analyzing large amounts of data over time to help make business decisions. It keeps a lot of historical data and is organized to make complex searches and reports easier, even if they take a bit longer to run.

Operational Data Store:

It is a central database that consolidates data from multiple systems to support business operations.

Data Marts:

The data mart is a subset of the data warehouse that is usually oriented to a specific business line or team.

Data marts Vs Data Warehouses:

1. Data Warehouse: A large storage of data from across the whole organization, covering many topics and used for broad analysis and decision-making.
2. Data Mart: A smaller, focused part of a data warehouse, created for a specific team or purpose (like sales or marketing) to make it easier for them to find the data they need.

Data Warehouse Life cycle:

1. Requirements Gathering: Figure out what data the business needs and how it will use it to make decisions.
2. Data Modeling: Design how data will be organized in the warehouse, setting up tables and relationships so data is easy to access and analyze.
3. ETL (Extract, Transform, Load):

* Extract: Collect data from different sources.
* Transform: Clean and format the data so it’s consistent.
* Load: Store the cleaned data in the warehouse.

1. Data Loading and Storage: Set up the data warehouse to store large amounts of data reliably and securely.
2. Data Access and Reporting: Create tools for users to access the data, run reports, and generate insights.
3. Maintenance and Optimization: Regularly update the data, monitor performance, and make improvements as needed to keep everything running smoothly.

