

Summary of Day 1 – Data Engineering

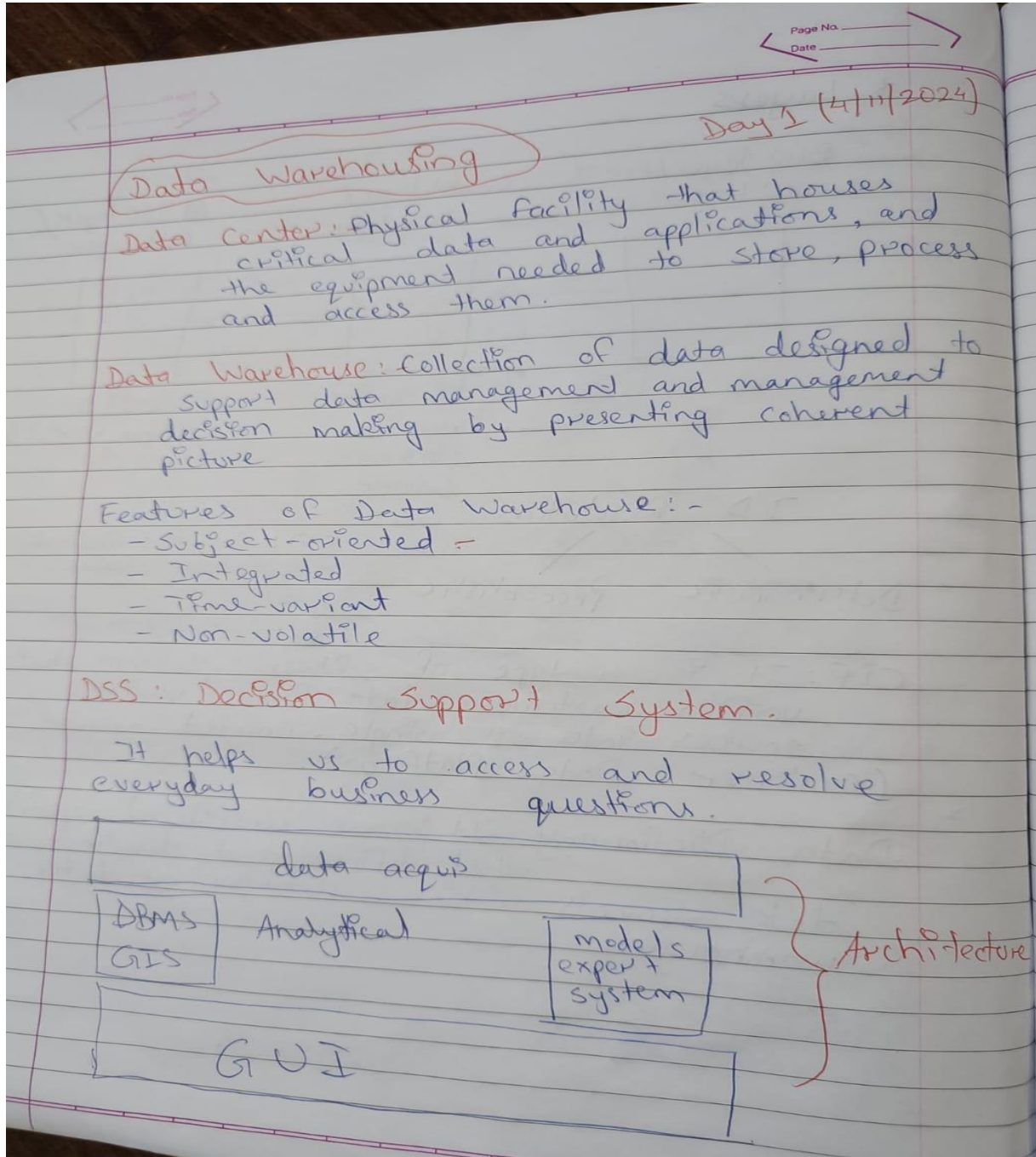
-Sarthak Niranjana Kulkarni (Maverick)

- sarthakkul2311@gmail.com

- (+91) 93256 02791

Day 1 – 4/11/2024 (Monday)

➤ Handwritten Notes:-



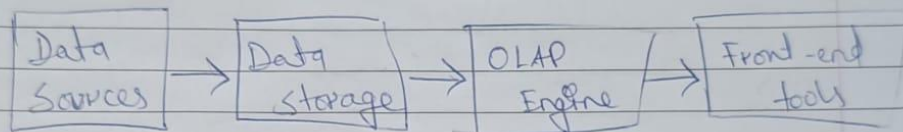
RDBMS = Relational Database Management Systems.
ODS = Operation Data Store.

DSS architecture styles:-

- OLTP (Online Transaction Processing) (RDBMS)
- OLAP (Online Analytical Processing) (Data warehouse)

A typical ETL-based data warehouse uses.

- Staging
- Integration
- Access layers (Data Marts)



Data Mart: It is a subset of the data warehouse that is usually oriented to a specific business line.

It typically contains targeted and limited set of data, structured to support specific analysis, reporting or decision making tasks.

ROLAP: Relational OLAP

MOLAP: Multi-dimensional OLAP

* Three kinds of data warehouse Application

- Information Processing (tables, charts, graphs)
- Analytical Processing (slice, dice, drilling, multidimension)
- Data Mining (Analytical models, knowledge discovery, prediction)

➤ Digital Notes:-

1. Introduction to Data Warehouse:-

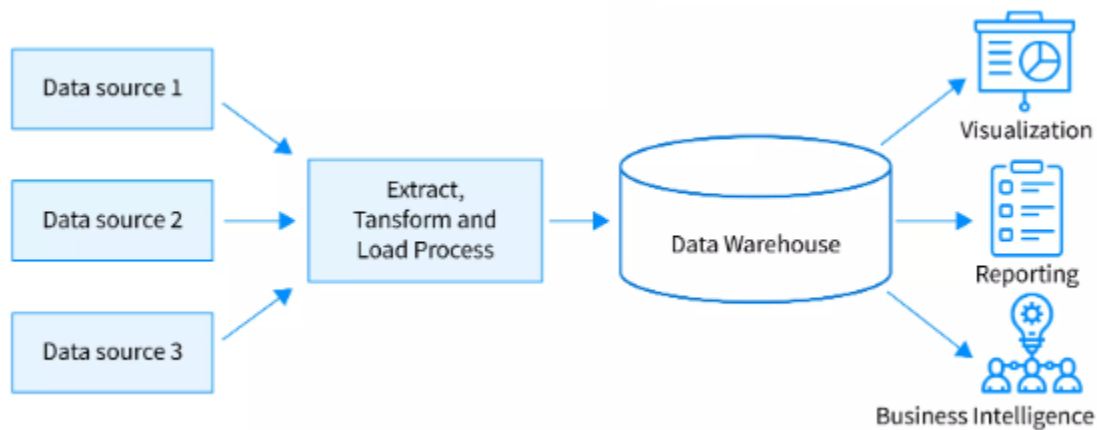
A **data warehouse** is a centralized system that consolidates data from multiple sources, optimized for fast querying, reporting, and analysis. It stores historical data for tracking trends and supports business intelligence by organizing data in a way that's easy to analyze, often using ETL processes to ensure data quality.

- **Central Repository:** Collects data from various sources for unified access.
 - **Historical Data:** Stores past data for trend analysis and reporting.
 - **Optimized for Analysis:** Designed for complex queries and faster data retrieval.
 - **ETL Process:** Uses Extract, Transform, Load to prepare and clean data.
 - **Supports BI:** Enables business intelligence and data-driven decision-making.
-

2. Purpose of Data Warehouse:-

- **Centralized Data Storage:** Combines data from multiple sources into one system for easier access and management.
 - **Optimized for Analysis:** Structures data for fast querying, complex analysis, and reporting, unlike transactional databases.
 - **Supports Historical Analysis:** Stores historical data, enabling trend analysis and tracking changes over time.
 - **Enhances Decision-Making:** Provides accurate, consolidated data that supports informed, data-driven business decisions.
 - **Facilitates Business Intelligence (BI):** Powers BI tools and dashboards, making insights readily available for stakeholders.
-

3. Data Warehouse Architecture: -



4. Operational Data Store:-

An **Operational Data Store (ODS)** is a centralized database that integrates and stores current, real-time data from various operational systems. Unlike a data warehouse, it focuses on recent transactional data, providing a quick, up-to-date view for day-to-day operations and short-term decision-making. The ODS is often used as an intermediary between operational systems and a data warehouse, ensuring data consistency across systems without heavy analytical processing.

5. OLTP vs Warehouse Applications:-

OLTP (Online Transaction Processing) and **Data Warehouse Applications** serve different purposes:

- **OLTP:** Designed for managing real-time, day-to-day transactional data. It supports frequent, quick updates like insert, update, and delete operations in applications (e.g., banking systems, e-commerce). OLTP systems prioritize speed and efficiency in handling many small transactions.
 - **Data Warehouse:** Optimized for analytical processing and historical data storage, supporting complex queries and reporting. Data warehouses focus on data aggregation and trend analysis, making them ideal for business intelligence but not for real-time transaction management.
-

6. Data Marts:-

A **data mart** is a smaller, focused subset of a data warehouse, tailored to meet the specific needs of a particular department or business unit, such as sales, marketing, or finance. Data marts store only relevant data for quick access and analysis, making it easier for users to retrieve specific insights without navigating the entire data warehouse. They can be independent (sourced directly from operational systems) or dependent (sourced from a central data warehouse), and they help streamline reporting and decision-making within specific business areas.

7. Data Marts vs Data Warehouse:-

Data Marts and **Data Warehouses** both support data storage and analysis but differ in scope and purpose:

- **Data Warehouse:** A centralized repository for storing comprehensive, enterprise-wide data from multiple sources. It's designed for complex analysis, historical data storage, and supporting business intelligence across the entire organization.
 - **Data Mart:** A smaller, focused subset of a data warehouse, targeted at a specific department or function (e.g., sales or finance). Data marts provide quick, tailored insights relevant to particular business areas, making them more efficient for specific analysis.
-

8. Data Warehouse Lifecycle:-

The **Data Warehouse Lifecycle** includes steps to build and maintain a data warehouse: gathering requirements, designing data models, developing ETL processes, loading data, testing for quality, deploying to users, and ongoing maintenance. This process ensures reliable, optimized data storage and access for business analysis.

