

DAY 1 ASSIGNMENT

1. Introduction to Data Warehousing

- **Definition:**
 - Centralized storage system for data.
 - Stores data from multiple sources for analysis.
 - Provides a unified view of business information.
 - Supports decision-making processes.
- **Features:**
 - **Subject-Oriented:** Focuses on data specific to subjects like sales, inventory.
 - **Integrated:** Combines data from different sources with consistency.
 - **Time-Variant:** Stores historical data.
 - **Non-Volatile:** Data is stable and not altered after entry.
- **Purpose:**
 - Provides historical data for analysis.
 - Facilitates business forecasting and planning.
 - Supports data-driven decision-making.

2. Purpose of Data Warehouse

- **Decision Support:**
 - Enables quick and informed decisions.
 - Helps assess business performance and strategy.
 - Improves response time to market changes.
- **Centralized Data Source:**
 - Integrates data from different departments.
 - Offers a consolidated view of the organization.
 - Reduces data redundancy across systems.
- **Enhanced Data Quality:**
 - Ensures consistency in data definitions and formats.
 - Reduces discrepancies in reports.
 - Maintains data accuracy across various functions.

3. Data Warehouse Architecture

- **Layers:**

- Staging Layer: Extracts and transforms data.
- Data Integration Layer: Cleanses and merges data.
- Access Layer: Makes data available for analysis.
- **Components:**
 - **Data Sources:** Operational systems, external sources.
 - **ETL Processes:** Extraction, transformation, loading.
 - **Metadata:** Describes data warehouse structure.
- **Types of Warehouses:**
 - **Enterprise Data Warehouse (EDW):** Centralized storage for entire organization.
 - **Operational Data Store (ODS):** Temporary storage for real-time data.
 - **Data Mart:** Subset focused on specific functions or departments.
- **ETL Tools:**
 - Automate data extraction, transformation, and loading.
 - Ensure data quality and consistency.
 - Support scheduling and job management.
- **OLAP Engines:**
 - Facilitates fast, multi-dimensional data queries.
 - Supports slicing, dicing, and pivoting data.
 - Enables interactive data exploration.

4. Operational Data Store (ODS)

- **Definition:** Interim area for data storage before data warehouse.
 - Used for short-term data processing.
 - Provides operational reporting capabilities.
- **Purpose:**
 - Offers real-time access to recent data.
 - Supports high-frequency data updates.
 - Bridges operational databases and data warehouse.
- **Characteristics:**
 - Stores current data, unlike historical data in warehouses.
 - Supports quick response for daily business needs.
 - Integrates data from multiple operational systems.

5. OLTP Vs. Data Warehouse Applications

- **OLTP (Online Transaction Processing):**

- Optimized for transaction processing.
- Supports daily business operations.
- Focuses on short-term data storage.
- **OLAP (Online Analytical Processing):**
 - Used for data analysis and reporting.
 - Optimized for querying large datasets.
 - Supports business intelligence and decision support.
- **Comparison:**
 - OLTP handles high-volume transactions.
 - OLAP supports complex, multi-dimensional queries.
 - OLTP is real-time; OLAP is historical.
- **Example Applications:**
 - **OLTP:** ATM transactions, order entry.
 - **OLAP:** Sales forecasting, trend analysis.

6. Data Marts

- **Definition:** Subset of a data warehouse.
 - Oriented toward a specific business area.
 - Provides focused, quick access to relevant data.
- **Types:**
 - **Dependent Data Marts:** Draws data from centralized warehouse.
 - **Independent Data Marts:** Operate independently from warehouse.
 - **Hybrid Data Marts:** Combination of both types.
- **Uses:**
 - Tailors data for specific departments or functions.
 - Speeds up access and analysis.
 - Supports targeted business intelligence.
- **Advantages:**
 - Reduces load on main data warehouse.
 - Customizable to department needs.
 - Enhances analysis for specific business lines.

7. Data Warehouse Life Cycle

- **Planning:**
 - Define goals and scope of the warehouse.

- Identify key stakeholders and requirements.
- Develop a project plan and timeline.
- **Data Modeling:**
 - Design logical and physical data models.
 - Define relationships between entities.
 - Create schemas (e.g., star, snowflake).
- **ETL Process:**
 - Extract data from various sources.
 - Transform data to fit warehouse schema.
 - Load cleansed data into the warehouse.
- **Testing:**
 - Verify data accuracy and completeness.
 - Test ETL processes and data retrieval.
 - Ensure performance meets requirements.
- **Deployment:**
 - Make the data warehouse available to users.
 - Provide training and documentation.
 - Implement ongoing maintenance and support.

OLAP Types in Data Warehousing

- **ROLAP (Relational OLAP):**
 - Uses relational databases for dynamic, multi-dimensional analysis.
 - Processes data directly in the database with SQL queries.
 - Works with large volumes of data, suitable for detailed and complex queries.
- **MOLAP (Multidimensional OLAP):**
 - Organizes data in multi-dimensional cubes, enabling “slice and dice” analysis.
 - Pre-calculates and stores summarized data, offering fast retrieval.
 - Efficient for high-level analysis across dimensions like time, geography, and product lines.