**Date: 17-07-2025**

**Data Warehousing**

### **1. Introduction to Data Warehousing:**

* A **Data Warehouse** is a centralized system used to store and manage large volumes of data from multiple sources.
* Unlike regular databases used in day-to-day operations, a data warehouse is designed for **analysis, reporting, and decision-making**.
* It stores **historical, cleaned, and organized data** that supports Business Intelligence (BI) tools.
* E.g.: Think of a data warehouse like a digital library that collects books (data) from different publishers (sources) into one well-organized place where readers (analysts) can find insights easily.

**2.** **Purpose of a Data Warehouse:**

A data warehouse is built to help businesses use their data in smarter ways. Here's why it's important:

* **Bring Data Together**: It collects information from different places like sales, finance, and marketing into one system.
* **Look at the Past**: It helps you see how things have changed over time, like sales trends over months or years.
* **Make Better Decisions**: Managers can use this data to make smart choices based on facts, not guesses.
* **Easy Reporting**: It powers reports and dashboards that show clear summaries of business performance.
* **Faster Answers**: It can run complex questions (queries) much faster than normal systems.

### 

### **3. Data Warehouse Architecture:**

Data Warehouse Architecture shows how data moves and is stored in the system. It usually has four simple parts:

#### **a. Data Source Layer**

* This is where the data comes from. It can be from different systems like apps for sales, customer info, or files.

#### **b. Data Staging Layer**

* This is like a cleaning room. Here, data is collected, cleaned, and changed to a useful format before being stored.

#### **c. Data Storage Layer**

* This is the main storage area. Cleaned data is stored here in an organized way, so it's easy to find and use later.

#### **d. Presentation Layer**

* This is the part users see. It includes tools and charts that help people analyze the data and make decisions.

### **4. Operational Data Store (ODS)**

An Operational Data Store (ODS) is a place where fresh and current data is collected from different sources before being moved into the main data warehouse. It helps make quick decisions using up-to-date information.

#### **In Simple Terms:**

* It holds **current and fresh data**.
* It's good for **real-time reports** (like daily sales).
* It acts like a **waiting room** for data before it goes to the warehouse.

| **Feature** | **ODS (Operational Data Store)** | **Data Warehouse** |
| --- | --- | --- |
| Data Freshness | Real-time or very recent | Older, historical data |
| Use Case | Daily updates, quick actions | Long-term planning |
| Data History | Very short-term | Stores many years of data |

### 

### **5. OLTP vs Data Warehouse Applications (OLAP)**

**Definition:**

* **OLTP (Online Transaction Processing):** Systems that manage day-to-day business transactions like order entry, banking, or inventory.
* **OLAP (Online Analytical Processing):** Systems used to analyze and understand historical data for business insights and decision-making.

**Explanation:**

* OLTP is fast and handles lots of small changes like **adding or updating records**. It's used in apps like banking or retail checkout.
* OLAP is slower but **powerful for asking big questions** like "How were sales last year?" or "What are our top 10 products?"

**Summary Table:**

| **Feature** | **OLTP (Online Transaction Processing)** | **OLAP (Online Analytical Processing)** |
| --- | --- | --- |
| Purpose | Run daily operations | Analyze business data |
| Data Type | Current and detailed | Historical and summarized |
| Operation Type | Insert, Update, Delete | Read, Aggregate |
| Query Complexity | Simple and fast | Complex and slower |
| Example | ATM | Sales trends |
| Schema Design | Highly normalized (3NF) | De-normalized (Star/Snowflake) |

### **6. Data Marts:**

A **Data Mart** is a smaller, more focused version of a data warehouse. It contains data specific to a department or business function such as Sales, HR, or Finance.

#### **Types of Data Marts:**

* **Dependent**: These get their data from the **main data warehouse**. Think of them as mini versions made for specific departments.
* **Independent**: These get data **directly from other systems**, not from the main warehouse. They're built separately.
* **Hybrid**: These use a **mix of both**—some data comes from the warehouse and some from outside sources.

#### **Benefits:**

* Faster access for specific users
* Easy to maintain
* Cost-effective for small teams

### **7. Data Marts vs Data Warehouses:**

### 

| **Feature** | **Data Mart** | **Data Warehouse** |
| --- | --- | --- |
| Scope | Department-specific | Organization-wide |
| Size | Smaller | Large |
| Development | Faster | Slower |
| Maintenance | Simple | Complex |
| Data Source | Few specific sources | Multiple integrated sources |

### **8. Data Warehouse Life Cycle (Step-by-Step)**

A well-structured process is followed to build and manage a data warehouse. Here's a detailed life cycle:

#### **1. Requirement Gathering**

* Understand the business needs.
* Identify users, KPIs, and data sources.

#### **2. Data Modeling**

* Design logical and physical models.
* Choose appropriate schema (Star, Snowflake).

#### **3. ETL Design and Development**

* Extract data from sources.
* Transform data (cleaning, mapping).
* Load it into the data warehouse.

#### **4. Data Warehouse Development**

* Build the actual storage and indexing structures.
* Set up user access, security, and backups.

#### **5. Testing**

* Validate data correctness.
* Test performance and query outputs.

#### **6. Deployment**

* Move the warehouse into a live environment.
* Integrate with reporting tools.

#### **7. Maintenance and Monitoring**

* Regular updates, performance checks.
* Add new data sources and features as needed.