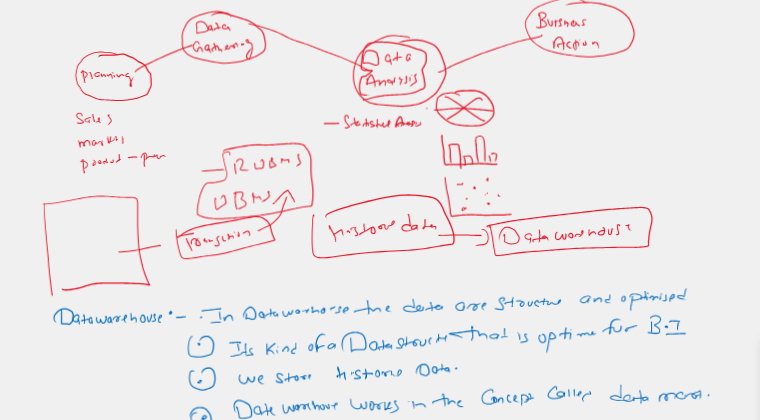
Day 3 –[**Data Warehousing**](https://wbd.ms/share/v2/aHR0cHM6Ly93aGl0ZWJvYXJkLm1pY3Jvc29mdC5jb20vYXBpL3YxLjAvd2hpdGVib2FyZHMvcmVkZWVtLzVkZTA5YTRiZDhkYTRlZWE4ZjZjMGM1MTU2MjVkNWMwX0JCQTcxNzYyLTEyRTAtNDJFMS1CMzI0LTVCMTMxRjQyNEUzRF8zODFjYjMzOS1mZTk1LTQzM2UtYjhjNy00OTMyNDlkZTkxMWI=)Concepts & Importance in Analytics

**📅 03-06-2025**

**📘 1. Data Warehousing**

A **Data Warehouse (DWH)** is a **centralized data storage** system designed for **reporting and analytics**. It integrates data from multiple sources (e.g., sales, marketing) and is optimized for **querying historical data**.

🧠 **Real-world example:** Amazon collects millions of transaction logs daily → loads into DWH for quarterly sales analysis.

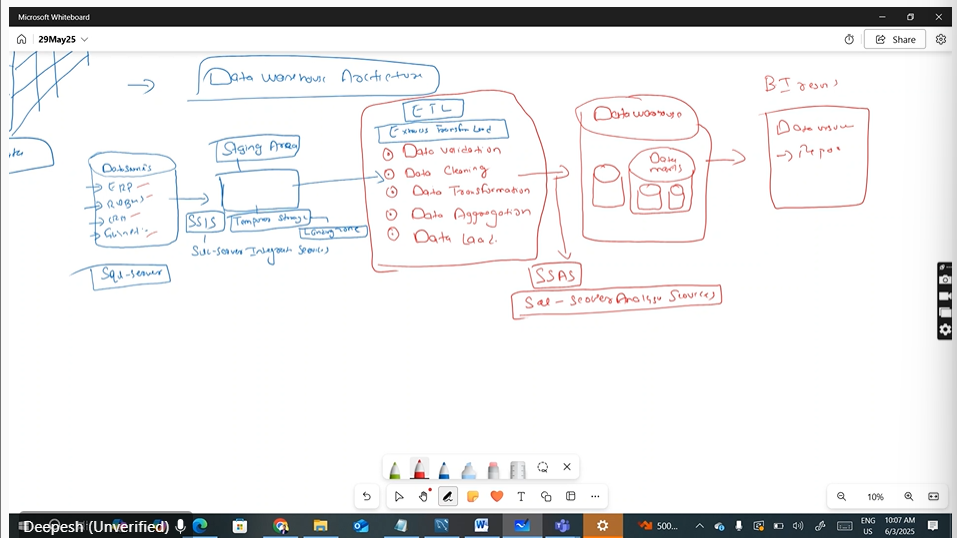
[Data Warehousing](https://www.youtube.com/watch?v=NphMcnU8ymU)

**🎯 2. Purpose of a Data Warehouse**

* Enables **business decision-making** using historic & clean data.
* Supports **data analysis, dashboards, KPIs, ML, AI models.**
* Improves **data consistency, governance, and quality**.

🔗  [Data Warehousing](https://learn.microsoft.com/en-us/azure/architecture/data-guide/relational-data/data-warehousing)

**🏛️ 3. Data Warehouse Architecture**

A typical DWH architecture has **3 layers**:

| **Layer** | **Role** |
| --- | --- |
| **Source Layer** | Pulls raw data (CRM, ERP, APIs) |
| **Staging Layer** | Temporary processing/cleansing |
| **Presentation Layer** | Final structured format for analysis |
|  |  |

Image Reference:  
🔗 [DWH Architecture](https://posts.dwhacademy.com/design_iko.png)

**🏪 4. Operational Data Store (ODS)**

🧾 A **real-time or near real-time** temporary database used before data moves to the DWH.

🔄 Ideal for **operational reports** that need fresh data (e.g., today’s sales, current order status).

🕹️ *Think of it as a “buffer zone” before final analysis.*

**🔁 5. OLTP vs OLAP (Warehouse Applications)**

| **Feature** | **OLTP (Online Transaction Processing)** | **OLAP (Online Analytical Processing)** |
| --- | --- | --- |
| Focus | Daily operations | Analysis, reporting |
| Example | ATM Withdrawal, Shopping Cart | Quarterly Sales Analysis |
| Data Volume | Small transactions | Large, historical datasets |
| Tables | Normalized | Denormalized |

Limit 100 MB-2GB 100GB-1 TB

Relationship ER Schema – snowflake, Star, etc.

[OLTP vs OLAP](https://www.youtube.com/watch?v=iw-5kFzIdgY)

**🧱 6. Data Marts**

A **data mart** is a subset of the DWH designed for a **specific department or purpose**.

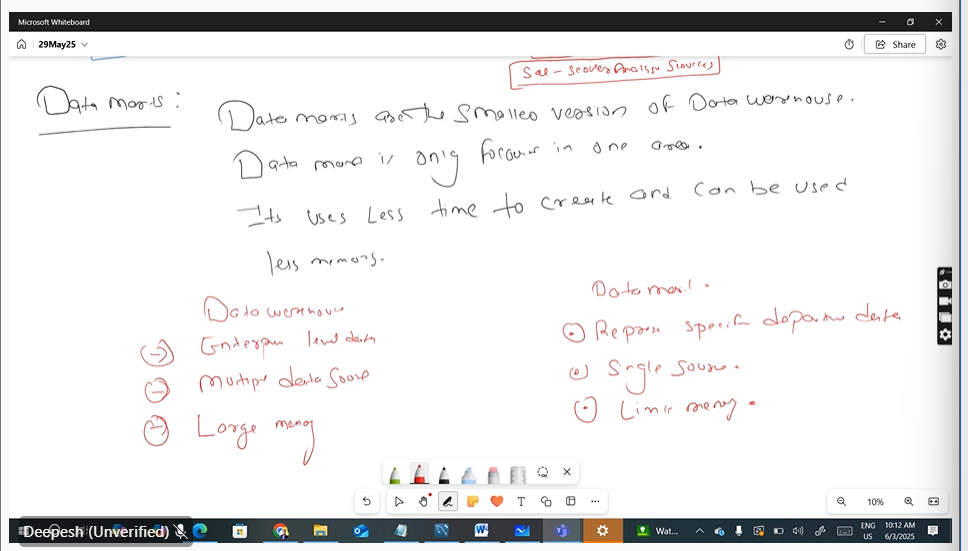
🔹 *Example:* Sales Mart, HR Mart, Finance Mart.

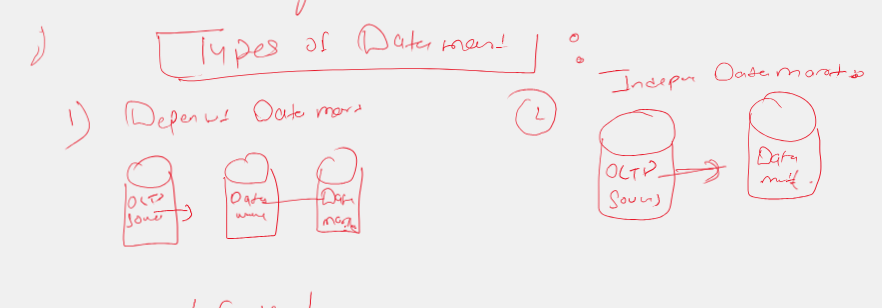
📦 Instead of querying entire DWH, teams access **custom-cut views**.

**⚔️ 7. Data Marts vs Data Warehouses**

| **Feature** | **Data Mart** | **Data Warehouse** |
| --- | --- | --- |
| Scope | Department-specific | Enterprise-wide |
| Size | Smaller | Larger |
| Development Time | Short | Long |
| Data Source | One or few | Many |

🧠 *Think of DWH as a university and Data Marts as individual departments.*





**📈 8. Data Warehouse Life Cycle**

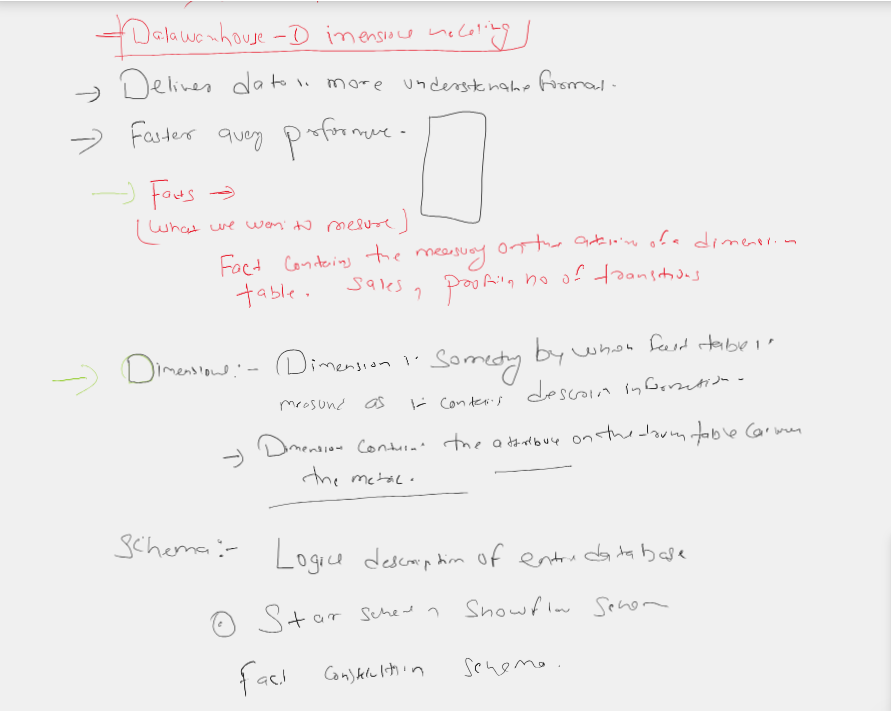
1. **Requirement Analysis** – Define business goals
2. **Data Modeling** – Star/Snowflake schema
3. **ETL Development** – Extract, Transform, Load
4. **Testing & QA**
5. **Deployment**
6. **Monitoring & Maintenance**

📘 [Lifecycle](https://data-flair.training/blogs/data-warehouse-lifecycle/)

**📐 Dimensionality in Data Warehousing**

**Dimensionality** refers to the structure of data in a **dimensional model** used in Data Warehouses, primarily consisting of **Facts** and **Dimensions**.

It helps in **organizing and analyzing** data across different perspectives (time, geography, product, etc.)

**🧱 Key Components of Dimensional Modeling**

| **Component** | **Description** |
| --- | --- |
| **Fact Table** | Contains **measurable, quantitative data** (e.g., sales, profit, revenue) |
| **Dimension Table** | Contains **descriptive attributes** (e.g., product name, customer name, date) that **provide context** to facts |

**🎮 Bonus Quiz:**

Try this interactive data warehouse quiz:  
🔗 [Data Warehousing Quiz](https://www.proprofs.com/quiz-school/story.php?title=data-warehouse-quiz)

**💎 Summary**

| **Concept** | **Summary** |
| --- | --- |
| **DWH** | Storage for analytics-ready data |
| **Purpose** | Business intelligence, KPIs |
| **Architecture** | 3-layer: Source → Staging → Presentation |
| **ODS** | Temporary real-time store |
| **OLTP vs OLAP** | Transactional vs Analytical |
| **Data Marts** | Department-level mini-DWH |
| **DWH vs Mart** | Enterprise-wide vs Specific |
| **Lifecycle** | Plan → Design → ETL → Report |

**MongoDB Overview**

**📘 1. Overview of NoSQL Databases**

| **Topic** | **Details** |
| --- | --- |
| **NoSQL** | **“Not Only SQL” – databases designed to handle unstructured, semi-structured, and massive-scale data** |
| **Types** | **Document-based (MongoDB), Key-Value (Redis), Column-based (Cassandra), Graph-based (Neo4j)** |
| **Flexible Schema** | **No predefined structure; JSON-like** |

**📊 Use case: Storing social media posts, logs, IoT data, and sensor streams.**

**🌐 2. Key Features of MongoDB**

**✅ Stores data as BSON (Binary JSON)  
✅ Flexible schema  
✅ Horizontal scaling with sharding  
✅ Full JavaScript query interface**

**📘 Docs:** [**MongoDB Features**](https://www.mongodb.com/why-use-mongodb)

**🧑‍💻 Basic MongoDB Commands**

| **Command** | **Purpose** |
| --- | --- |
| mongosh | Starts MongoDB shell |
| show dbs | Lists all databases |
| use students | Switches to students DB |
| db.collection.insertOne() | Inserts 1 document |
| db.collection.insertMany() | Inserts multiple documents |
| db.collection.find() | Retrieves documents |
| db.collection.updateOne() | Updates a document |