**17/07/25-Simple Notes**

**1. An Introduction to Data Warehousing**

-- A **data warehouse** is a Subject oriented, integrated, time variant, non-volatile collection of data in support of management’s system.

-- In simple words, it is a place where we store data for business purposes.

**2. Purpose of Data Warehouse**

* To integrate data from multiple sources
* To support **decision-making** by offering fast access to data
* To provide **historical data analysis**
* To improve **data quality and consistency**

**3. Data Warehouse Architecture**

**A diagram of data warehouse

AI-generated content may be incorrect.**

The architecture has **3 main layers**:

1. **Data Source Layer**: Contains raw data from multiple sources (databases, applications, etc.)
2. **Data Staging Area (ETL)**: Extracts, Transforms, and Loads data
3. **Data Storage Layer**: Cleaned and transformed data is stored here (in fact and dimension tables)

**ETL = Extract, Transform, Load**

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

* An **Operational Data Store** is used for real-time data integration.
* Unlike a data warehouse, it stores current data (not historical).
* It is used when businesses need up-to-date data quickly.

**5. OLTP vs Data Warehouse Applications**

|  |  |  |
| --- | --- | --- |
| **Feature** | **OLTP (Online Transaction Processing)** | **Data Warehouse** |
| Purpose | Day-to-day transactions | Data analysis |
| Data | Current, detailed | Historical, summarized |
| Operations | Read & write | Mostly read |
| Speed | Very fast | Slower |
| Users | Clerks, admins | Analysts, managers |

**6. Data Marts**

* A **Data Mart** is a smaller version of a data warehouse.
* It focuses on a single department or business function (like Sales or HR).
* Data marts are faster and easier to use for specific teams.

**7. Data Marts vs Data Warehouses**

|  |  |
| --- | --- |
| **Data Warehouse** | **Data Mart** |
| Data warehouse is a Centralised system. | While it is a decentralised system. |
| In data warehouse, lightly denormalization takes place. | While in Data mart, highly denormalization takes place. |
| Data warehouse is top-down model. | While it is a bottom-up model. |
| To built a warehouse is difficult. | While to build a mart is easy. |
| In data warehouse, Fact constellation schema is used. | While in this, Star schema and snowflake schema are used. |
| Data Warehouse is flexible. | While it is not flexible. |

**8. Data Warehouse Life Cycle**

The life cycle includes:

1. **Requirement Gathering** – Understand business needs
2. **Design** – Create architecture, data models
3. **Development** – ETL process, schema design
4. **Testing** – Ensure data accuracy and quality
5. **Deployment** – Move to production
6. **Maintenance** – Monitor, optimize, and update