**DATA WAREHOUSING**

**INTRODUCTION TO DATA WAREHOUSING**

A Data Warehouse (DW) is a *subject-oriented*, *integrated*, *time-variant*, and *non-volatile* collection of data that supports decision-making in an organization.

As per W.H. Inmon, often referred to as the father of data warehousing:

“**A data warehouse is a subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management’s decision-making process.**”

**PURPOSE OF DATA WAREHOUSE**

**** To provide a centralized repository of integrated data from multiple sources.

 Helps in strategic decision-making by providing a coherent picture of business conditions at a specific point in time.

 Enables historical data analysis for trend predictions.

 Supports Business Intelligence (BI) tools and reporting systems.

**DATA WAREHOUSE ARCHITECTURE**

Data warehouse architecture typically consists of the following layers:

* Data Source Layer: Internal databases, flat files, external data.
* ETL Layer: Extract, Transform, Load operations to clean and prepare data.
* Data Storage Layer: Central data warehouse repository.
* Data Mart Layer: Subsets of the data warehouse, focused on specific business lines.
* Presentation Layer: Used by reporting and analysis tools.

Architectural types:

* **Single-tier**: Minimizes data redundancy but not often used in practice.
* **Two-tier**: Separates data warehouse from user interface.
* **Three-tier**: Most common, separates data warehouse, OLAP engine, and presentation.

**OPERATIONAL DATA STORE**

 A type of database that serves as an **intermediate staging area** for operational data before it enters the data warehouse.

 Used for **short-term decisions** and operational reporting.

 Unlike a data warehouse, ODS contains **current** (not historical) data.

**OLTP Vs WAREHOUSE APPLICATIONS**

|  |  |  |
| --- | --- | --- |
| FEATURE | OLTP | Data Warehouse(OLAP) |
| Data Type | Current,detailed | Historical,summarized |
| Operations | Insert,Update,Delete | Select (Read-only) |
| Speed | High for inserts/updates | Optimized for complex queries. |
| Data Volume | Low to medium | Very high |
| Purpose | Day-to-day transactions | Strategic decision making |

**DATA MARTS**

 A Data Mart is a subset of a Data Warehouse, designed for a specific department or function like sales, finance, or marketing.

 It is subject-oriented and focuses on individual business areas.

 Can be:

* Dependent: Sourced from a central DW.
* Independent: Built directly from operational systems.

**DATA MARTS Vs DATA WAREHOUSES**

|  |  |  |
| --- | --- | --- |
| FEATURE | DATA MART | DATA WAREHOUSE |
| Scope | Specific business functions | Entire Organization |
| Data Volume | Lower | High |
| Implementation | Quick, department-wise | Time consuming and complex |
| Source | Can be independent or from DW | Integrated from multiple sources |

**DATA WAREHOUSE LIFE CYCLE**

**** Requirement Analysis: Understand business needs.

 Data Modeling: Design logical & physical models.

 ETL Development: Extract, clean, transform, and load data.

 Data Warehouse Implementation: Build and test the warehouse.

 Data Marts Creation: Based on specific business lines.

 Front-End Development: Reporting and visualization tools.

 Deployment & Maintenance: Monitor performance and update.