**DAY 1 NOTES**

**Name : Podutur Lahari - DE126**

**Date : 04-11-2024**

**Date Warehousing :** Data Warehouse (DW) is a Subject oriented,integrated, time variant, non-volatile collection of data in support of management’s system.

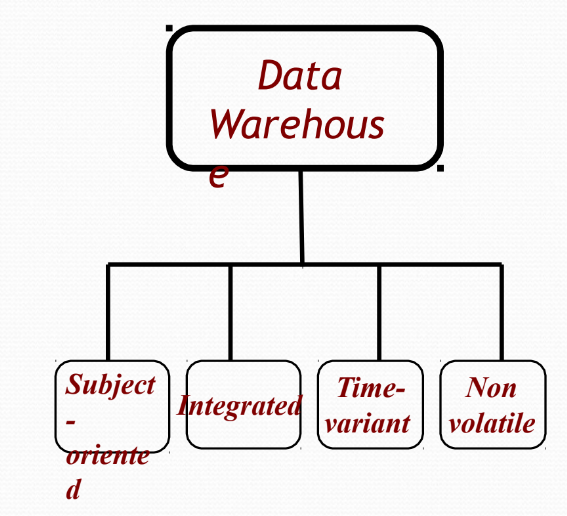
It is a collection of data designed to support management decision making by presenting a coherent picture of business conditions at a single point of time.

Features of Data Warehouse:-

❖ Subject-oriented

❖ Integrated

❖ Time-variant

❖ Nonvolatile

* Subject-oriented:

1. Data are organized according to the subject instead of Application.
2. It mainly focuses on modeling and analysis of data for decision makers, not on daily operations or transaction processing.

* Integrated:

1. Constructed by integrating multiple, heterogeneous data sources like relational databases, flat files, on-line transaction records.
2. Ensure consistency in naming conventions, encoding structures, attribute measures, etc. among different data sources.

**Types of Data**

1. Structured Data:

• Format: Tables with rows and columns.

• Examples: Relational databases with primary keys (PK) and foreign keys (FK).

2. Semi-Structured Data:

• Format: JSON, XML.

3. Unstructured Data:

• Format: Images, audio, video, files.

• Databases: Non-relational databases.

**DSS architectural styles**

➢ OLTP (Online Transaction Processing) -used by traditional operational systems (RDBMS).

➢ OLAP (Online Analytical Processing) -used by Data Warehouse.

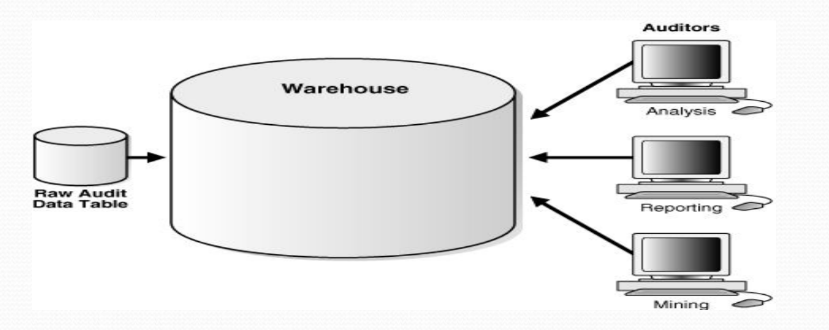
**OLTP**

● OLTP is a methodology to provide end users with access to large amounts of data

● It works in an intuitive and rapid manner to assist with deductions based on investigative reasoning.

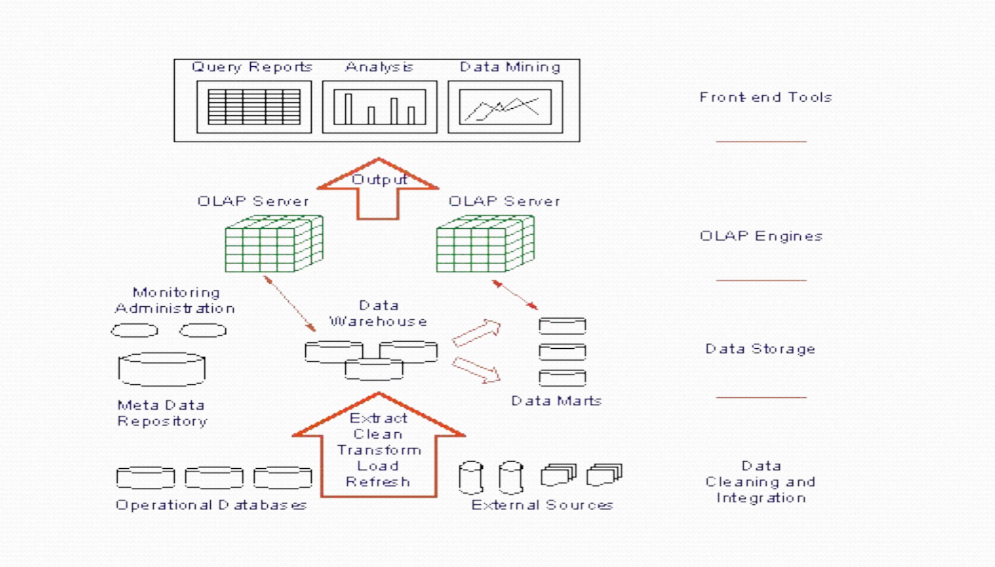
● OLTP refers to a class of systems that facilitate and manage transaction-oriented applications,

typically for data entry and retrieval transaction processing.



**OLAP**

● OLAP is an approach to answer multi-dimensional analytical queries which also encompasses relational reporting and data mining.

● An OLAP cube is an array of data that is understood in terms of its 0 or more dimensions which enables the users to gain insight into their data in a fast, interactive, easy-to-use manner.

**OLTP vs OLAP**

• OLTP (Online Transactional Processing):

* Focus: Fresh data from servers/apps/IoT devices.
* Role: Data Engineer.
* Size: Small (e.g., 600GB).
* Operations: High write (80%), low read (20%)

• OLAP (Online Analytical Processing):

* Focus: Historical data.
* Role: Data Analyst, Data Scientist.
* Size: Large.
* Operations: High read (80%), low write (20%)

**ROLAP**

● It is a form of OLAP that performs dynamic multi- dimensional analysis of data stored in a relational

database rather than in a multi-dimensional database (which is usually considered the OLAP standard).

● Data processing may take place within the database system, a mid-tier server, or the client.

● In two-tier architecture, the user submits a Structured Query Language (SQL) query to the database and receives back the requested data.

**MOLAP (Multi-dimensional OLAP)**

● It is a form of OLAP that helps the user to “slice and dice” information, providing multi-dimensional

analysis of data by putting data in a cube structure.

● Most MOLAP products use a multi-cube approach in which a series of small, dense, pre-calculated cubes make a hypercube.

**Data Marts**

* Purpose: Categorize data for quicker access.
* Example: Movies sorted into Telugu, Hindi, English folders.