ASSIGNMENT – 1

Trainee Name: Swathi Baskaran

1. **An Introduction to Data Warehousing**

   It is a set of process of subject-oriented, integrated, time-variant and non-volatile collection of data which is helpful for management decision making. Collecting, storing and managing large volumes of data in a centralized repository to support business intelligence and decision-making activities takes place. This reads the transactional data, performs the analytics using ETL tools and finally stores the analytical data. This is mainly used for reporting and analytics purpose.

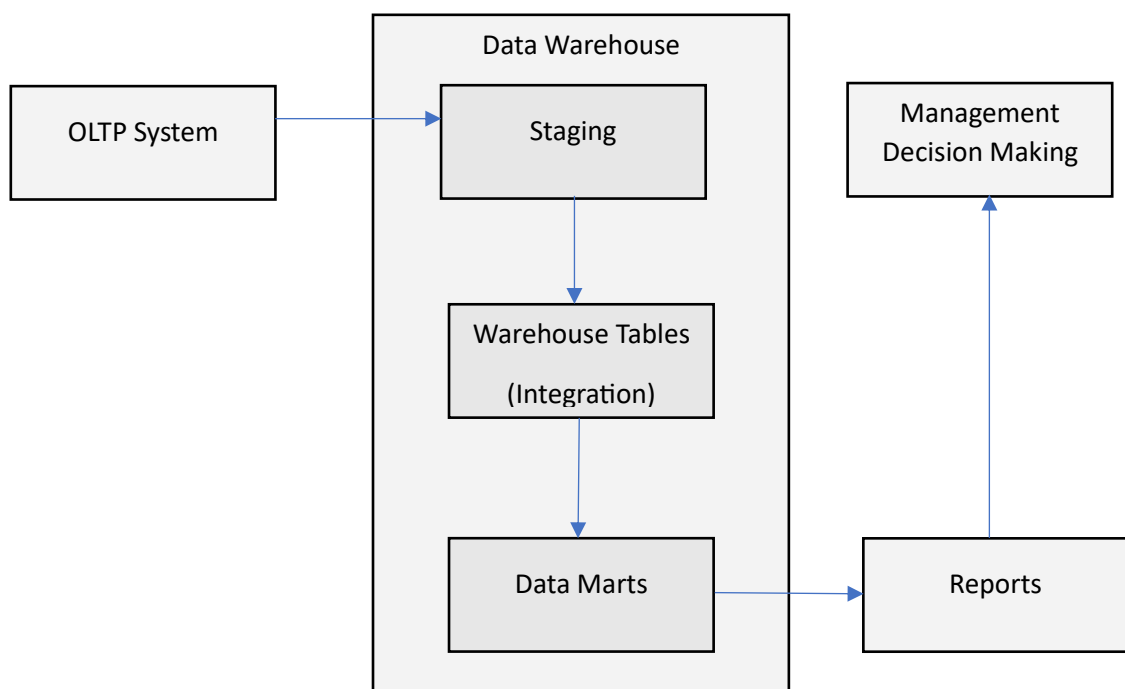2. **Purpose of Data Warehouse**

   Limitations of OLTP:

   - OLTP requires instant update.
   - Data that we get from OLTP system is not suitable for data analysis.
   - For simple calculations, we need to query multiple tables using multiple joins.

   Purpose of Warehouse:

   To overcome the above limitations and to perform data analysis and help make business decisions, data warehousing is essential.

3. **Data Warehouse Architecture**

4. **Operational Data Store**

An Operational Data Store (ODS) in a data warehouse acts as a bridge between transactional systems and analytical systems, providing a near real-time, consolidated view of operational data. It's designed to support quick decision-making and operational reporting, unlike a data warehouse which is used for historical analysis.

5. **OLTP vs Warehouse Applications**

OLTP	Warehouse Applications
Contains limited time period data	Contains historical data
Volatile in nature	Non-volatile in nature
Used for simple efficient purpose	Used for complex analytical purpose
Used for capturing transactional data	Used for deriving analytical data

6. **Data Marts**

- It is a subset of data warehouse.
- It contains analytical data corresponding to a single specific business line item/subject area.
- Data marts consist of facts and dimension tables
- Data marts act as source for data visualisation (Reporting).

7. **Data Marts vs Data Warehouses**

Data Marts	Data Warehouses
It is a subset of data warehouse.	It is a superset of data mart.
There could be multiple data marts, one each for a subject area/line of business.	There will be only 1 data warehouse environment.

8. Data Warehouse Lifecycle