**Madiha Aimon Tappal**

[**madihaaimon@gmail.com**](mailto:madihaaimon@gmail.com)

**Data Engineering Batch – 1**

**Day – 1 Assignment**

**What is a Data Warehouse?**

* **Data Warehouse (DW)** is a Subject oriented, integrated, time variant, non-volatile collection of data in support of management’s system.
* A **Data Warehouse**is separate from DBMS, it stores a huge amount of data, which is typically collected from multiple heterogeneous sources like files, DBMS, etc. The goal is to produce statistical results that may help in decision-making. For example, a college might want to see quick different results, like how the placement of CS students has improved over the last 10 years, in terms of salaries, counts, etc.
* 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.
* Data Warehousing can be applied anywhere where we have a huge amount of data and we want to see statistical results that help in decision making.
* **E.g. Social Media Websites, Banking**

**Need for Data Warehouse**

* An ordinary Database can store MBs to GBs of data and that too for a specific purpose. For storing data of TB size, the storage shifted to the Data Warehouse.
* Besides this, a transactional database doesn’t offer itself to analytics. To effectively perform analytics, an organization keeps a central Data Warehouse to closely study its business by organizing, understanding, and using its historical data for making strategic decisions and analysing trends.
* Features of **Data Warehouse**: -
  + Subject-oriented
  + Integrated
  + Time-variant
  + Non-volatile
* Subject-oriented:
  + Data are organized according to the subject instead of application.
  + It mainly focuses on modelling and analysis of data for decision makers, not on daily operations or transaction processing.
* Integrated:
  + Constructed by integrating multiple, heterogeneous data sources like relational databases, flat files, on-line transaction records.
  + Ensure consistency in naming conventions, encoding structures, attribute measures, etc. among different data sources.
* Time-variant:
  + The time horizon for the data warehouse is significantly longer than that of operational systems. i.e. provide information from a historical perspective (e.g., past 5-10 years).

* Non-volatile:
  + No updates are allowed. Once the data entered into the data warehouse, they are never removed.
  + The data in warehouse represent company’s history, the operational data representing near term history are always added to it.

### **Benefits of Data Warehouse**

* **Better business analytics:**Data warehouse plays an important role in every business to store and analysis of all the past data and records of the company. which can further increase the understanding or analysis of data for the company.
* **Faster Queries: The data** warehouse is designed to handle large queries that’s why it runs queries faster than the database.
* **Improved data Quality:**In the data warehouse the data you gathered from different sources is being stored and analysed it does not interfere with or add data by itself so your quality of data is maintained and if you get any issue regarding data quality then the data warehouse team will solve this.
* **Historical Insight:**The warehouse stores all your historical data which contains details about the business so that one can analyse it at any time and extract insights from it.

**Advantages of Data Warehousing**

* **Intelligent Decision-Making:**With centralized data in warehouses, decisions may be made more quickly and intelligently.
* **Business Intelligence:**Provides strong operational insights through business intelligence.
* **Historical Analysis:** Predictions and trend analysis are made easier by storing past data.
* **Data Quality:**Guarantees data quality and consistency for trustworthy reporting.
* **Scalability:** Capable of managing massive data volumes and expanding to meet changing requirements.
* **Effective Queries:**Fast and effective data retrieval is made possible by an optimized structure.
* **Cost reductions:** Data warehousing can result in cost savings over time by reducing data management procedures and increasing overall efficiency, even when there are setup costs initially.
* **Data security:** Data warehouses employ security protocols to safeguard confidential information, guaranteeing that only authorized personnel are granted access to certain data.

**Disadvantages of Data Warehousing**

* **Cost:**Building a data warehouse can be expensive, requiring significant investments in hardware, software, and personnel.
* **Complexity:**Data warehousing can be complex, and businesses may need to hire specialized personnel to manage the system.
* **Time-consuming:**Building a data warehouse can take a significant amount of time, requiring businesses to be patient and committed to the process.
* **Data integration challenges:**Data from different sources can be challenging to integrate, requiring significant effort to ensure consistency and accuracy.
* **Data security:**Data warehousing can pose data security risks, and businesses must take measures to protect sensitive data from unauthorized access or breaches.