**VAISHALI BOKADIYA**

**DAY 1 ASSIGNMENT**

**DATA ENGINEERING & DATA WAREHOUSE**

**DATA ENGINEERING:**

Data Engineering is designing, building and scaling systems that enable data collection, processing and storage for analytical use.

**WHY DATA ENGINEERING?**

Now a days, a huge amount of data is being generated daily, through various sources such as IOT devices, websites etc. To transform this data into usable form, data engineering comes into action. It transforms raw data into a form that can be used for analysis.

**DATA CLASSIFICATION:**

* Raw data
  + Unprocessed data
  + No schema applied
* Processed data
  + Processed and organized data
  + Schema applied, stored in tables
* Cooked data
  + Processed data that can be summarized

**BIG DATA PROPERTIES:**

* Volume: Amount of data
* Velocity: Incoming data rate
* Variety: Different format data
* Veracity: Reliability of data

**DATA PROCESSING METHODS:**

* Batch processing: Processing of data that has been stored over a period of time.
* Stream processing: Processing of data as the data comes in.

**DATA STORAGE:**

* Relational Database (SQL)
* Document Store (NoSQL )

**DATA WAREHOUSE:**

Subject oriented, integrated, time variant, non-volatile collection of data that supports management decision making by presenting a coherent picture of business conditions at a single point of time.

**FEATURES OF DATA WAREHOUSE:**

* Subject oriented: Data is organized according to the subject. It is mainly focused on modelling and analysis of data.
* Integrated: It integrates data from various heterogeneous sources (formats) and maintains consistency among different sources.
* Time variant: Provides historical information of a longer period of time, example: past 8-10 years.
* Non Volatile: Data in the warehouse can neither be updated nor deleted.

**DECISION SUPPORT SYSTEM:**

Decision making system enables to make quick decisions. It processes useful information from raw data, documents, personal knowledge or business models.

* Structured components: Directly helps in decision making.
* Unstructured components: Requires human interaction to be processed.

**DSS ARCHITECTURAL STYLES:**

* Online Transaction Processing (OLTP): Used by traditional operational systems (RDBMS)
* Online Analytical Processing (OLAP): Used by data warehouse

**ONLINE TRANSACTION PROCESSING (OLTP):**

OLTPrefers to a class of systems that facilitate and manage transaction oriented applications, typically for data entry and retrieval transaction processing.

**BENEFITS OF OLTP:**

* Simplicity: OLTP systems make things simpler.
* Efficiency: OLTP systems provides accurate data.
* Data integrity: OLTP systems maintain data integrity
* Fast query processing: OLTP systems provide fast query processing in multiple access environment.

**PITFALLS OF OLTP:**

* Resultant data is not suitable for data analysis.
* OLTP requires instant update.
* For single transaction, we need to query multiple tables by using joins.