**Name: Muhammad Usman**

**Data: 14/03/2023**

**Data Engineering Basics**

**Database: (Main Goals)**

* Focus on Operational Data
* Highly Normalized Data model
* Built to Manage CRUD (Create, Read, Update, Delete) operations quickly.
* Queries are generally simple and very Fast.
* Don’t always track historical information.
* Focuses mainly on Transactional Data.
* My SQL is primary database being used by various social media platforms for storing data.

**Data Warehouse: (Main Goals)**

* Centralizes and integrates business operations data.
* Most data models are denormalized.
* Built to manage analytics and aggregations quickly.
* Queries are more complex and often take longer.
* Track Historical information.
* Used by analysts, data scientists and other business users to built dashboards, reporting and other analytical purposes.
* Integrated, Time variant, non-volatile collection of Data in support of management’s decision making process.

**Big Data: (Main Goals)**

* Collecting and storing large amounts of data: Big Data is all about dealing with large and complex data sets that can't be managed with traditional data processing tools. The first goal of Big Data is to collect, store, and manage large amounts of data from different sources, such as social media, sensors, and machines.
* Analyzing data for insights: Once the data is collected, the next goal is to analyze it to gain insights and knowledge. The analysis can help identify patterns, trends, and correlations that can be used to make informed decisions and improve business outcomes.
* Enhancing decision-making: One of the key goals of Big Data is to use data to make better decisions. By analyzing data and gaining insights, businesses can make informed decisions that can lead to increased efficiency, productivity, and profitability.
* Improving customer experiences: Big Data can help businesses understand their customers better and provide personalized experiences. By analyzing customer data, businesses can identify customer needs, preferences, and behaviors, and use that information to tailor their products and services to meet those needs.

**Data Lake: (Main Goals)**

* Centralizing and storing large amounts of data: The primary goal of a Data Lake is to centralize and store large amounts of data from different sources, such as databases, social media, and IoT devices. The data is stored in its raw and unstructured form, without being transformed or processed.
* Allowing flexible and scalable data storage: Data Lakes are designed to handle massive amounts of data that can be easily scaled up or down based on business needs. They provide a flexible and scalable storage environment that can store data of different types and sizes.
* Enabling easy data access and exploration: Data Lakes make it easy to access and explore data by providing a single repository for all types of data. The data can be accessed by different teams and individuals within the organization, without the need for complex data extraction processes.
* Supporting diverse data analytics: Data Lakes support diverse types of data analytics, including descriptive, diagnostic, predictive, and prescriptive analytics. They enable data scientists and analysts to perform advanced analytics on large and complex data sets, without the need for data transformation or modeling.
* Facilitating data integration and sharing: Data Lakes enable data integration and sharing across different systems and applications within the organization. They provide a common platform for sharing data across different teams and departments, improving data collaboration and efficiency.
* Supporting data governance and security: Data Lakes provide a framework for data governance and security. They allow organizations to implement data policies and controls to ensure data quality, privacy, and security.
* **Top of Form**