**SYEDA FATIMA SHAH**

**14th March 2023**

***Task # 1:***

Please familiarize yourself with the basics of Data Engineering which include the following topics:

**- Big Data**

Big data refers to extremely large and complex data sets that are difficult to process and analyze using traditional data processing tools and techniques. The term "big data" typically refers to data sets that are so large and varied that they cannot be managed with traditional databases, data warehouses, or data processing applications.

Big data is characterized by the 3Vs:

1. Volume: The sheer amount of data generated and collected is enormous, ranging from terabytes to petabytes and beyond.
2. Velocity: The speed at which data is generated and processed is very high, as it comes from a variety of sources, such as sensors, social media, and other real-time data feeds.
3. Variety: The data comes in a variety of formats, such as structured, semi-structured, and unstructured data, and may include text, images, audio, and video.

The processing of big data requires specialized tools and techniques such as Hadoop, Spark, and NoSQL databases, which can handle the large-scale and distributed nature of big data. Big data is used in a variety of industries and applications, including healthcare, finance, e-commerce, and social media analysis.

**- Data Lake**

A data lake is a centralized repository that allows you to store all your structured and unstructured data at any scale.

* S3, ADLS, HDFS

**- Database**

* a database is an organized collection of data stored and accessed electronically.
* Usually used for OLTP Systems
* Postgres, MySQL, SQL Server etc

**- Data Warehouse**

A data warehouse is a type of data management system that is designed to enable and support business intelligence (BI) activities, especially analytics