**Data Engineering**

**Week 1**

**Day 01: 14/02/23**

**Big Data**

Big data refers to extremely large and complex data sets that traditional data processing methods are unable to handle efficiently. Big data is characterized by its volume, velocity, and variety, and often requires specialized tools and techniques to store, process, and analyze. Examples of big data include social media data, sensor data, machine data, and scientific data.

**Data Lake**

A data lake is a central repository that stores all types of raw data, both structured and unstructured, at any scale. Unlike traditional data warehouses, data lakes don't require a predefined schema or data model, which makes them more flexible and scalable. Data lakes are designed to store and analyze large volumes of data, and they often leverage technologies such as Hadoop, Spark, and NoSQL databases. An example of a data lake is the data lake used by Netflix, which stores data from various sources, such as customer data, streaming data, and content metadata, and is used to provide personalized recommendations and improve the user experience.

**Database**

A data lake is a central repository that stores all types of raw data, both structured and unstructured, at any scale. Unlike traditional data warehouses, data lakes don't require a predefined schema or data model, which makes them more flexible and scalable. Data lakes are designed to store and analyze large volumes of data, and they often leverage technologies such as Hadoop, Spark, and NoSQL databases. An example of a data lake is the data lake used by Netflix, which stores data from various sources, such as customer data, streaming data, and content metadata, and is used to provide personalized recommendations and improve the user experience.

**Data Warehouse**

A data warehouse is a centralized repository that stores data from different sources and transforms it into a common format for efficient analysis and reporting. Data warehouses are designed to support business intelligence (BI) activities such as data mining, trend analysis, and decision-making. They often use a structured schema, such as a star or snowflake schema, to organize data and facilitate querying and analysis. An example of a data warehouse is the data warehouse used by Amazon, which stores data from various sources, such as customer data, transaction data, and inventory data, and is used to analyze customer behavior, optimize pricing, and manage inventory.