Task 1

Q1) What is big data?

Big data refers to large, complex, and diverse data sets that require advanced technologies and analytical methods to process and extract insights from. The term "big data" typically describes data sets that are too large, too fast, or too varied for traditional data processing tools and techniques to handle. Big data is often characterized by its "three Vs": volume (the sheer amount of data), velocity (the speed at which data is generated and processed), and variety (the range of data types and sources). Other important characteristics of big data include veracity (the reliability and accuracy of data) and value (the potential insights and benefits that can be gained from analyzing the data).

Q2) What is data lake?

A data lake is a centralized repository that allows for the storage and analysis of large volumes of structured and unstructured data. Unlike traditional data warehouses, which store data in a structured manner, data lakes are designed to store raw data in its original form. Data lakes can be used to store a wide variety of data, including social media data, machine-generated data, sensor data, and more. Data is stored in its original format, allowing for greater flexibility in how the data is analyzed and processed.

Q3) What is Database?

A database is a collection of organized and structured data that can be accessed, managed, and updated by computer systems. It is designed to allow for efficient storage, retrieval, and manipulation of data, and can be used to support a wide range of applications and business operations. Databases can be managed using a database management system (DBMS), which provides tools and interfaces for creating, querying, and maintaining the database.

Q4) What is Data warehouse?

A data warehouse is a centralized repository of integrated data that is used to support business intelligence and decision-making activities. It is designed to provide a comprehensive view of an organization's data, by collecting data from multiple sources, cleaning and transforming it, and organizing it into a format that can be easily analyzed and queried. A data warehouse typically contains historical data, rather than real-time data, and is optimized for complex queries and reporting. It can be used to support a wide range of business intelligence and analytics applications, including data mining, trend analysis, and predictive analytics. Data warehouses typically use a process called ETL (extract, transform, load) to integrate data from multiple sources into a unified format. This process involves extracting data from various sources, cleaning and transforming the data to ensure consistency and accuracy, and then loading it into the data warehouse.