SIMON DIEDONG DOMBO UNIVERSITY OF BUSINESS AND INTEGRATED DEVELOPMENT STUDIES

(SDD UBIDS)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
COURSE: EMERGING TECHNOLOGY
GROUP WORK (5)

STUDENTS NAMES INDEX NUMBERS

OSMAN ADAMS	UC0832323
IDDRISU MOHAMMED KHALIFA	UG0809723
IDDRISU SALIFU JIMBA	UG0809423
SIITA OSMAN	UG0809423
LATIF ZUMATA	UG0707523
ABDUDU ABDUL HAMID	UG0789823
SHABAN KWAO MANOH	UG0706723

1. List and discuss the characteristics of big data?

1. Volume: One of the defining features of big data is its sheer volume. Big data sets are typically massive in size, ranging from terabytes to petabytes. For example, according to a research study published in the International Journal of Computer Applications, it is estimated that the global volume of data will reach 175 zettabytes by 2025. This abundance of data presents both opportunities and challenges for computer scientists in terms of storage, management, and analysis.

2. Velocity: Big data is generated at a high velocity from various sources such as social media, sensors, and Internet of Things (IoT) devices. The speed at which data is generated and collected requires efficient real-time processing and analysis techniques. For instance, a case study from a scholarly journal on big data analytics in healthcare revealed that the velocity of data from patient monitoring devices necessitated the development of streaming analytics tools to enable timely decision-making by healthcare professionals.

3. **Variety:** Big data comes in various forms, including structured, unstructured, and semi-structured data. Structured data is organized and easily searchable, while unstructured data, such as text, images, and videos, presents challenges in terms of analysis. For example, a study published in the Journal of Big Data highlighted the need for specialized algorithms and machine learning techniques to process and derive insights from diverse data types.

4. Veracity: Veracity refers to the trustworthiness and reliability of the data. Big data sets often contain noisy, incomplete, or inconsistent data, which can impact the accuracy of analysis and decision-making. To address this challenge, computer scientists are exploring techniques such as data cleansing, data validation, and anomaly detection to ensure the quality and reliability of big data sets.

5. Value: The ultimate goal of analysing big data is to extract actionable insights and value from the data. This could involve identifying trends, patterns, and correlations that can inform business strategies, scientific research, or policy decisions. For instance, a research paper from an academic journal highlighted how big data analytics in e-commerce enabled companies to personalize customer experiences, improve marketing campaigns, and drive revenue growth.

2. Describe the big data life cycle. Which step you think most useful and why?

Big data life cycle refers to the process of managing large volumes of data throughout its lifespan, from collection and storage to analysis and utilization. The big data life cycle typically consists of six stages: data generation, data acquisition, data storage, data processing, data analysis, and data visualization. Each stage plays a crucial role in making sense of the enormous amount of data available and turning it into valuable insight.

The first stage, data generation, involves the creation of data through various sources such as social media, sensors, and transaction records. For example, in Ghana, data generated from mobile phone usage, social media interactions, and e-commerce transactions contribute to the vast amount of big data available.

The next stage, data acquisition, focuses on collecting and ingesting the data into storage systems. This can include data warehouses, data lakes, or cloud-based storage solutions. Once the data is stored, the processing stage involves cleaning, transforming, and preparing the data for analysis. This is a critical step, as big data is often unstructured and needs to be organized before any meaningful insights can be derived. Following this, data analysis involves using various tools and techniques to uncover patterns, trends, and correlations within the data. This stage can include advanced analytics, machine learning, and data mining to extract valuable information.

Finally, the data visualization stage presents the insights in a visual format, such as charts, graphs, and dashboards, to make it easier for stakeholders to understand and use the findings. This is especially important in making data-driven decisions and communicating findings effectively. In terms of usefulness, all stages of the big data life cycle are essential for deriving value from large datasets. However, the data analysis stage is particularly crucial, as it is where the actual insights and actionable information are uncovered. Without effective data analysis, the data collected and stored remains just that – data. It is the analysis that turns data into knowledge and enables informed decision-making.

3. List and describe each technology or tool used in the big data life cycle

1. Data collection: This involves gathering large volumes of data from various sources. Tools such as Apache Kafka and Flume are used to collect and ingest data from sources such as social media, sensors, and logs.

2. Data storage: Once collected, the data needs to be stored efficiently. Technologies like Hadoop Distributed File System (HDFS) and NoSQL databases such as Apache Cassandra and MongoDB are used for scalable and distributed storage of big data.

3. Data processing: Tools like Apache Spark and Apache Flink are used for processing and analysing large datasets in parallel, enabling real-time processing and complex analytics.

4. Data visualization: Communicating insights from big data is crucial. Tools like Tableau and Power BI help in creating interactive and easy-to-understand visualizations of the analysed data.

These technology tools play a critical role in enabling the big data cycle to extract valuable insights from large and complex datasets, which is increasingly important in today's data-driven world.

4. Discuss the three method of computing over a large dataset

- **Batch Processing:** This method involves processing data in large batches, where data is collected over a period of time and processed together. It is suitable for scenarios where real-time processing is not required, and it allows for efficient processing of large volumes of data.

- **Stream Processing:** Stream processing deals with real-time data streams, where data is processed as it arrives. It is useful for scenarios where immediate analysis or response is needed, such as monitoring social media feeds or analysing sensor data in real-time.

- **Interactive Processing:** This method allows users to interactively explore and analyse data. It provides a more flexible and interactive approach to data analysis, where users.