

Nama : Restu Wibisono

NPM : 2340506061

#### Pertanyaan

1. Sebutkan dan jelaskan yang dimaksud dengan 10 V's of Big Data!
2. Bagaimana langkah-langkah yang dilakukan dalam menyiapkan Big Data Stack?
3. Sebutkan beberapa platform untuk mengelola Big Data Stack dan jelaskan karakteristik masing-masing! (minimal 3 platform)

#### Jawab

1. 10V's Big Data
  - a. **Volume:** Jumlah data yang besar, biasanya dalam skala terabyte atau bahkan petabyte.
  - b. **Velocity:** Kecepatan di mana data dihasilkan dan diproses. Big Data harus mampu menangani aliran data yang terus-menerus dan real-time.
  - c. **Variety:** Variasi dari berbagai data yang ada, mulai dari data terstruktur, semi-terstruktur, dan tidak terstruktur.
  - d. **Veracity:** Mengacu kepada keakuratan data.
  - e. **Value:** Nilai dari suatu data seberapa jauh akan bermanfaat, data yang menghasilkan wawasan berguna dan keputusan yang tepat.
  - f. **Variability:** Perubahan data serta pola yang tidak konsisten.
  - g. **Visualization:** Bagaimana data itu ditampilkan secara visual.
  - h. **Vulnerability:** Potensi dan masalah keamanan yang berkaitan dengan data. Perlu memiliki perlindungan yang kuat.
  - i. **Verifiability:** Kemampuan untuk memastikan dan memverifikasi data yang digunakan benar.
  - j. **Viability:** Mengukur data bisa digunakan secara praktis dan efisien dengan tujuan yang sesuai.
2. Langkah-langkah menyiapkan big data
  - a. Identifikasi kebutuhan dan tujuan
  - b. Pengumpulan data
  - c. Penyimpanan data
  - d. Pembersihan data

- e. Pengolahan data
- f. Analisis data
- g. Visualisasi data
- h. Keamanan dan privasi
- i. Pemeliharaan dan monitoring
- j. Peningkatan dan skalabilitas

3. Platform untuk mengolah big data dan karakteristiknya

a. Apache Hadoop

Apache Hadoop adalah sistem terintegrasi yang menggunakan HDFS untuk penyimpanan dan MapReduce untuk pemrosesan data serta dikenal karena skalabilitasnya, Hadoop juga dapat menangani data besar dengan menambah node dan mendukung berbagai alat tambahan seperti HBase dan Hive untuk fleksibilitas analisis.

b. Apache Spark

Apache Spark ini unggul dalam kecepatan pemrosesan data dengan metode in-memory-nya, yang lebih cepat jika dibandingkan MapReduce. Spark mudah digunakan dengan API untuk Java, Scala, Python, dan R, serta mendukung batch, stream, dan pemrosesan interaktif.

c. Amazon Web Service (AWS) Big Data Stack

Amazon Web Services (AWS) Big Data Stack memiliki terintegrasi seperti dengan layanan Amazon S3 untuk penyimpanan, Amazon EMR untuk pemrosesan, dan Amazon Redshift untuk data warehousing.