



Nama : Triyana Dewi Fatmawati
NIM : 2241720206
Kelas : TI – 3D
Nomor : 21
Mata Kuliah : Big Data

Tugas 1 – Pengantar Big Data

1. Contoh kasus yang membutuhkan teknologi “big” data.

- Pada perusahaan logistik, seperti JNE dan SiCepat, data pengiriman paket termasuk lokasi pengambilan, tujuan, status pengiriman, dan waktu transit, akan terus direkam. Data ini sangat besar karena melibatkan jutaan paket setiap harinya.
- Pada platform media sosial, seperti Instagram atau TikTok, setiap aktivitas pengguna, seperti unggahan, like, komentar, dan waktu penayangan konten akan selalu dicatat. Data yang disimpan memiliki skala yang sangat besar karena melibatkan jutaan pengguna aktif.
- Pada platform e-commerce seperti Tokopedia atau Shopee, setiap interaksi pengguna, seperti pencarian produk, pembelian, dan ulasan, akan selalu dicatat. Data tersebut memiliki volume yang sangat besar karena mencakup jutaan transaksi setiap harinya.

Ketiga contoh kasus di atas menurut saya membutuhkan teknologi big data untuk menyimpan data karena mengingat data yang dihasilkan dalam jumlah besar. Selain itu, membutuhkan teknologi big data untuk memproses dan menganalisis pola atau tren pengguna.

2. Pengertian tentang Hadoop

Hadoop adalah sebuah framework open-source yang digunakan untuk menyimpan, memproses, dan menganalisis data dengan skala besar. Daripada menggunakan satu komputer besar untuk menyimpan, memproses, dan menganalisis data, Hadoop memungkinkan pengklasteran/pengelompokkan beberapa komputer untuk menganalisis data besar tersebut secara paralel sehingga lebih cepat.

3. Pengertian HDFS

HDFS atau Hadoop Distributed File System adalah sistem file yang digunakan Hadoop untuk membantu dalam menyimpan dan mendistribusikan data ke berbagai node/komputer dalam klaster.

4. Pengertian MapReduce

MapReduce adalah teknologi yang digunakan Hadoop untuk memproses data secara paralel pada berbagai node/komputer dalam klaster. MapReduce membagi tugas pemrosesan data menjadi beberapa bagian ke beberapa node/komputer dalam klaster, kemudian melakukan tugasnya tersebut secara bersamaan/paralel, setelah setiap node/komputer selesai dengan tugasnya maka hasil dari masing-masing node/komputer tersebut digabungkan menjadi satu hasil akhir atau output.