

**Mata kuliah : Big Data**  
**Materi : 01. Pengantar Big Data**  
**Nama : Sukma Bagus Wahasdwika**  
**NIM : 2241720223**  
**No. Presensi : 20**  
**Kelas : TI – 3D**

I. Contoh kasus yang membutuhkan teknologi “big” data

a. E-commerce

E-commerce memerlukan Big Data yang berguna untuk mengetahui serta memahami perilaku pengguna, memberikan pengalaman belanja, dan mengoptimalkan rantai pasokan. Dalam hal ini, data yang dianalisis mencakup dari kegiatan/aktivitas yang dilakukan pengguna seperti transaksi, riwayat pencarian, ulasan pengguna dengan tujuan platform penyedia e-commerce dapat memberikan rekomendasi produk yang tepat dan akurat kepada pengguna.

Salah satu aplikasi misalnya, Penentuan Harga Dinamis yang digunakan oleh “Tokopedia” dan “Shopee” menggunakan “Big Data Analytics” untuk menyesuaikan harga produk berdasarkan permintaan pasar dan harga pesaing secara real-time.

b. Analisis Tren Media Sosial

Penggunaan media sosial di zaman saat ini sudah menjadi hal yang utama bagi masyarakat untuk mendapatkan informasi terbaru. Big Data pada kasus ini memberikan analisis sentimen, tren viral, dan sebagainya. Misalnya, pada Analisis Sentimen aplikasi seperti Google Cloud Natural Language API menggunakan NLP(Natural Language Processing) untuk menganalisis opini publik terhadap suatu produk berdasarkan ulasan dan review di media sosial, ini diterapkan pada beberapa perusahaan e-commerce.

II. Pengertian tentang Hadoop

Hadoop merupakan sebuah framework open-source yang berfungsi sebagai penyimpanan, memproses, dan menganalisis data dalam jumlah besar (Big Data) sesuai alur. Hadoop mampu melakukan proses data secara paralel menggunakan cluster komputer, sehingga dapat efisien dan cepat daripada pemrosesan data tradisional. Hadoop dikembangkan oleh Apache Software Foundation serta digunakan oleh beberapa perusahaan besar seperti Facebook, Twitter, Amazon, dan Google untuk menangani Big Data.

III. Pengertian HDFS (Hadoop Distributed File System)

HDFS merupakan sistem penyimpanan yang tersalurkan dalam ekosistem Hadoop. Tujuannya dirancang untuk menyimpan “big data” agar dapat diakses dengan cepat dengan membagi data besar menjadi blok-blok kecil dan disalurkan ke beberapa komputer dalam cluster sehingga data dapat diproses secara paralel.

2 komponen utama cara kerja HDFS (Hadoop Distributed File System):

1. NameNode (Master Node)

Bertugas sebagai tanggung jawab untuk mengelola metadata atau informasi mengenai letak posisi setiap blok data disimpan.

2. DataNode (Slave Node)

Tugasnya adalah menyimpan blok data yang dikirim HDFS, dari bagian file disimpan di setiap DataNode dan bekerja sama membentuk agar file utuh, jika satu DataNode gagal maka sistem otomatis menggunakan salinan data dari node lain.

IV. Penegertian MapReduce

MapReduce merupakan model pemrograman dari Hadoop yang berfungsi untuk memproses data besar secara paralel di beberapa komputer dalam cluster.

MapReduce memiliki 2 tahap utama dari proses komputasi yang dibagi:

1. Map : Memisahkan data besar menjadi bagian kecil dan diproses secara paralel di beberapa node dalam cluster.
2. Reduce : Hasil dari Map tersebut digabungkan untuk mendapatkan hasil akhir berdasarkan key/kunci. Kemudian untuk menentukan hasilnya setiap key dihitung total kemunculannya. Misalnya, dari data Map pada tahap reduce akan menjumlahkan kata yang sama.