

LAPORAN TUGAS MATA KULIAH  
BIG DATA  
PERTEMUAN 1



NAMA : MUHAMAD IQBAL MUHTARAM  
NIM : 2241720265  
KELAS : TI 3D

PROGRAM STUDI D-IV TEKNIK INFORMATIKA  
JURUSAN TEKNOLOGI INFORMASI  
POLITEKNIK NEGERI MALANG

Jl. Soekarno Hatta No.9, Jatimulyo, Kec. Lowokwaru, Kota Malang, Jawa Timur 65141

## **Contoh Kasus yang Membutuhkan Big Data**

Big Data adalah kumpulan data dalam jumlah sangat besar dan beragam, yang sulit diolah dengan cara biasa. Contoh kasus yang membutuhkan Big Data bisa dilihat pada perusahaan e-commerce seperti Tokopedia atau Shopee. Mereka mengumpulkan data transaksi, riwayat pencarian, hingga ulasan pelanggan dalam jumlah yang sangat besar. Data tersebut kemudian dianalisis untuk membuat rekomendasi produk, mendeteksi penipuan, atau meningkatkan kualitas layanan. Karena datanya sangat banyak dan terus bertambah, teknologi Big Data dibutuhkan agar analisis bisa dilakukan lebih cepat dan efisien.

## **Pengertian Hadoop**

Hadoop adalah sebuah kerangka kerja (framework) open-source yang digunakan untuk menyimpan dan memproses data berukuran besar. Dengan Hadoop, data dibagi menjadi beberapa bagian kecil dan disimpan di banyak komputer (cluster). Tujuannya adalah agar data dapat diproses secara paralel (bersamaan), sehingga pemrosesan data besar menjadi lebih cepat. Hadoop dikembangkan oleh Apache Software Foundation dan terdiri dari beberapa komponen utama, antara lain HDFS dan MapReduce.

## **Pengertian HDFS (Hadoop Distributed File System)**

HDFS adalah sistem penyimpanan terdistribusi dalam Hadoop. Cara kerjanya adalah dengan memecah data menjadi blok-blok kecil, lalu menyimpannya di beberapa komputer yang tergabung dalam cluster. Kelebihan HDFS antara lain:

- Toleransi Kerusakan: Jika satu komputer rusak, data tetap aman karena disalin atau direplikasi ke komputer lain.
- Skalabilitas: Kapasitas penyimpanan bisa ditambah dengan menambah jumlah komputer dalam cluster.
- Kecepatan Akses Tinggi: Karena data tersebar, pembacaan dan penulisan data bisa dilakukan secara paralel.

## **Pengertian MapReduce**

MapReduce adalah model pemrograman di Hadoop untuk memproses data dalam jumlah besar. Prosesnya dibagi menjadi dua tahap:

Map: Data dibagi menjadi potongan kecil (split), lalu setiap potongan diproses di komputer yang berbeda. Hasilnya berupa pasangan kunci-nilai (key-value).

Reduce: Hasil dari tahap Map digabungkan sesuai kunci, lalu dirangkum menjadi output akhir.

Dengan konsep ini, banyak komputer bisa bekerja bersama untuk memproses data secara paralel, sehingga analisis data yang besar bisa diselesaikan dengan lebih efisien.