

LAPORAN PERTEMUAN 1
BIG DATA



MOCHAMMAD ZAKARO AL FAJRI
2241720175
TI – 3D

D-IV TEKNIK INFORMATIKA
POLITEKNIK NEGERI MALANG
2025/2026

Daftar Isi

Daftar Isi	2
Pengertian Big Data Dan Contoh Kasusnya	3
Pengertian Hadoop.....	3
Pengertian HDFS.....	4
Pengertian MapReduce.....	4
Daftar Pustaka.....	4

Pengertian Big Data Dan Contoh Kasusnya

Big Data adalah istilah yang diberikan pada kumpulan data yang berukuran sangat besar dan kompleks, sehingga tidak memungkinkan untuk diproses menggunakan perangkat pengelola database konvensional ataupun aplikasi pemroses data lainnya. Dalam Gartner IT Glossary, Big Data didefinisikan sebagai berikut : *Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.* Menurut Cloudera (salah satu perusahaan terkemuka di bidang Big Data) mendefinisikan: “Suatu data dapat dikatakan sebagai big data apabila terlalu besar untuk bisa disimpan dan diolah dalam satu mesin (komputer).”

Big data menjadi solusi penting dalam menangani volume data yang sangat besar, terutama di era digital. Salah satu contoh kasus yang sering dibahas adalah analisis perilaku konsumen di e-commerce. Perusahaan seperti Amazon dan Alibaba menggunakan big data untuk memahami pola pembelian, merekomendasikan produk, dan meningkatkan kepuasan pelanggan. Tanpa teknologi big data, proses analisis data dalam volume besar akan menjadi tidak efisien dan memakan waktu. Selain itu, big data juga digunakan dalam bidang Kesehatan. Salah satu penggunaan big data dalam bidang Kesehatan yaitu pada pengembangan obat-obatan. Perusahaan farmasi menggunakan big data untuk menganalisis keefektifan obat yang diuji berdasarkan data uji klinis dan membantu mempercepat proses pengembangan obat. Tidak berhenti disitu, big data juga digunakan dalam bidang keuangan yaitu dalam memprediksi pasar saham dan manajemen risiko, bidang energi dan lingkungan yaitu pemantauan konsumsi energi dan manajemen limbah, dan di bidang Pendidikan yaitu dalam prediksi keberhasilan akademik dan peningkatan efisiensi administrasi. Big Data merupakan teknologi yang *future-proof* dan memiliki potensi besar untuk terus berkembang serta mengubah kehidupan manusia secara signifikan.

Pengertian Hadoop

Hadoop merupakan *framework open-source* yang diciptakan dan dirancang untuk menyimpan dan melakukan pemrosesan data berskala besar secara terdistribusi. Hadoop dikembangkan oleh *Apache Software Foundation* sebagai solusi dalam menangani data skala terabyte bahkan petabyte. Hadoop memiliki 4 komponen utama yaitu HDFS atau *Hadoop Distributed File System* yang menjadi sistem penyimpanan terdistribusi dengan sistem kerja berupa membagi data menjadi blok kecil dan menyimpannya di berbagai node dalam kluster, *MapReduce* yang menjadi sistem pemrosesan data dan hasilnya berupa kumpulan data besar, *Yarn* atau *Yet Another Resource Negotiator* yang menjadi pengatur dalam manajemen, mendistribusikan, dan menjadwalkan tugas pemrosesan data, dan yang terakhir *Common* yang berisikan kumpulan utilitas dan pustaka yang mendukung modul-modul lainnya dalam melakukan pemrosesan data.

Untuk cara kerja dari Hadoop sendiri dapat dibagi menjadi 4 proses besar berupa pemecahan data, distribusi tugas, pemrosesan data dan pengumpulan hasil. Berikut ini merupakan penjelasan singkatnya.

1. **Pemecahan Data :** memecah data dalam blok-blok kecil, dengan HDFS yang bertugas untuk menyimpannya. Setiap blok akan direplikasi ke beberapa *node*, untuk memastikan keandalan dan toleransi kesalahan.

2. **Distribusi Tugas** : YARN akan mengatur sumber daya dan mendistribusikan tugas pemrosesan ke berbagai node dalam kluster. Fungsinya adalah memastikan pemanfaatan sumber daya yang optimal dan pemrosesan data yang efisien.
3. **Pemrosesan Data** : memproses data melalui MapReduce, yang membaginya ke dalam dua fase utama. Pertama adalah fase Map, yang memecah data menjadi pasangan kunci-nilai di berbagai node. Hasil dari fase Map tersebut kemudian akan berlanjut ke fase Reduce, yang bertugas menggabungkan data untuk mendapatkan hasil akhir.
4. **Pengumpulan Hasil** : HDFS akan menyimpan hasil akhir dari fase Reduce, atau mengirimkannya ke aplikasi lain untuk analisis lebih lanjut.

Pengertian HDFS

HDFS atau *Hadoop Distributed File System* adalah sistem penyimpanan data yang digunakan untuk menyimpan file dengan ukuran yang sangat besar. HDFS menggunakan metode pemecahan data dan file menjadi blok blok kecil kemudian menyimpannya di berbagai node dalam cluster.

Pengertian MapReduce

MapReduce merupakan model pemrograman yang dibuat untuk melakukan pemrosesan dan menghasilkan kumpulan data besar. MapReduce memiliki 2 proses kerja utama yaitu :

1. **Map** : pada proses ini, dilakukan pemecahan data menjadi berpasangan . Kemudian data tersebut akan diproses secara paralel yang memungkinkan analisis data dalam jumlah besar dilakukan lebih cepat dan efisien.
2. **Reduce** : pada proses ini, dilakukan menggabungkan dan peringkasan data menjadi output akhir.

Daftar Pustaka

Apache Software Foundation. (t.t.). *Dokumentasi Apache Hadoop*. Diakses pada 15 Februari 2025, dari <https://hadoop.apache.org>

Apache Software Foundation. (t.t.). *Dokumentasi Apache HDFS*. Diakses pada 15 Februari 2025, dari <https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html>

Apache Software Foundation. (t.t.). *Dokumentasi Apache MapReduce*. Diakses pada 15 Februari 2025, dari <https://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html>