Tugas Big Data Architeture and Infrastructure

Pertemuan 5

Hadoop Infrastructure Layer

Nama	**: Lukas Febrian Laufra**
Kelas/Nim	**: TI22J/20220040076**
Dosen Pengajar	**: Pak Ir. Somantri, S.T, M.Kom**
Waktu Pengerjaan	**: 03.52/21 November 2024**

Soal

1. Jelaskan komponen utama dalam infrastruktur Hadoop dan peran masing-masing. Berikan contoh penggunaan kasus untuk setiap komponen tersebut?

Jawaban

1. Berdasarkan informasi yang diberikan, komponen utama dalam infrastruktur Hadoop dan peran masing-masing adalah:

    1. HDFS (Hadoop Distributed File System)

        - Berfungsi untuk menyimpan data dalam jumlah besar secara terdistribusi di seluruh cluster Hadoop.

        - Data disimpan dalam potongan-potongan (blocks) dan diduplikasi di beberapa node untuk fault tolerance.

        - Contoh penggunaan kasus: Menyimpan data log, gambar, video, dan dataset lainnya yang berukuran besar.

    2. MapReduce

        - Merupakan framework pemrosesan data terdistribusi untuk aplikasi yang dapat dieksekusi secara pararel pada cluster Hadoop.

        - MapReduce memecah pekerjaan besar menjadi tugas-tugas kecil yang dapat dijalankan secara paralel pada berbagai node.

        - Contoh penggunaan kasus: Analisis log, pemrosesan gambar/video, pencarian kata kunci, dll.

    3. YARN (Yet Another Resource Negotiator)

- Berfungsi sebagai resource manager dan scheduler untuk aplikasi yang berjalan di atas Hadoop cluster.

- Mengatur alokasi sumber daya (CPU, memori, penyimpanan) untuk aplikasi MapReduce maupun aplikasi lainnya.

- Contoh penggunaan kasus: Menjalankan berbagai jenis aplikasi (batch processing, streaming, interactive) secara bersama-sama dalam satu cluster Hadoop.

4. Ecosystem Hadoop

- Terdiri dari berbagai komponen pendukung seperti Hive, Pig, Spark, Kafka, Zookeeper, dll.

- Menyediakan kemampuan tambahan seperti SQL-on-Hadoop, streaming data, graph processing, dsb.

- Contoh penggunaan kasus: Analisis data terstruktur menggunakan Hive, pemrosesan data streaming menggunakan Kafka dan Spark Streaming.

Dengan komponen-komponen tersebut, Hadoop dapat menyediakan kemampuan untuk menyimpan, memproses, dan menganalisis data dalam jumlah besar secara terdistribusi dan fault toleran. Hal ini sangat berguna untuk berbagai kasus penggunaan big data di berbagai industri.