

NAMA : Juan Fakhri

KELAS : TI.19.D5

NIM : 311910361

TUGAS -3 BIG DATA

1. Jelaskan perbedaan dari Distributed dan Parallel Systems, serta Distributed dan Parallel Processing!

Komputasi Paralel digunakan dalam komputasi kinerja tinggi seperti pengembangan superkomputer. Sedangkan **Komputasi terdistribusi** memberikan skalabilitas dan konsistensi data. Google dan Facebook menggunakan komputasi terdistribusi untuk penyimpanan data.

Pemrosesan paralel (parallel processing) adalah penggunaan lebih dari satu CPU untuk menjalankan sebuah program secara simultan. Idealnya, parallel processing membuat program berjalan lebih cepat karena semakin banyak CPU yang digunakan. Tetapi dalam praktek, seringkali sulit membagi program sehingga dapat dieksekusi oleh CPU yang berbeda-beda tanpa berkaitan di antaranya.

2. Jelaskan maksud dari Hadoop disebut sebagai software framework dan software platform! Berikan contohnya, dan jelaskan apa perbedaan antara framework dan platform!

maksud dari **hadoop disebut sebagai software framework** adalah, dikarenakan memungkinkannya pemrosesan data berukuran besar secara terdistribusi dengan melibatkan berkluster-kluster komputer.

Sedangkan, maksud dari **hadoop di sebut sebagai software platform** adalah, dikarenakan menghubungkan beberapa komputer sehingga dapat saling bekerja sama dan sinkron dalam menyimpan dan mengolah data sebagai satu kesatuan.

Perbedaan dari framework dan platform adalah, **framework** hanya mengacu pada perangkat lunak (software) dan mengabaikan perangkat keras (hardware), Jadi jika Anda berkecimpung dalam hardware Anda tidak akan punya framework.

platform itu memungkinkan suatu software untuk running. sedangkan framework tidak, framework hanya fokus pada design.

3. Jelaskan maksud dari HDFS (Hadoop Distributed File System) sebagai “direktori di komputer” dan sebagai “file system”!

HDFS sebagai **direktori di komputer** di mana data Hadoop disimpan. Untuk pertama kalinya, direktori ini akan di “format” agar dapat bekerja sesuai spesifikasi dari Hadoop.

HDFS sebagai **file system**, yang mana artinya tidak sejajar dengan jenis file system dari OS seperti NTFS, FAT32. HDFS ini menumpang diatas filesystem milik OS baik Linux, Mac atau Windows.

4. Jelaskan cara kerja MapReduce (MR) s

MapReduce adalah suatu software framework dan programming model yang digunakan untuk pemrosesan jumlah data yang besar. Program **MapReduce bekerja** dalam dua fase, yaitu Map dan Reduce. Tugas Map berurusan dengan splitting dan mapping dari data sedangkan tugas Reduce melakukan shuffle dan reduce terhadap data. Proses keseluruhan bergerak melalui empat fase eksekusi, yaitu splitting, mapping, shuffling, dan reducing.

5. Jelaskan perbedaan antara MR dan YARN!

Yet Another Resource Negotiator (YARN): Sistem yang mengatur dan memonitor cluster node dan resource usage.

MapReduce: Framework yang membantu program untuk melakukan komputasi data secara paralel.