# 1. Perbedaan Sistem Distribusi dan Pengolahan Paralel

Sistem Distribusi: Sistem distribusi terdiri dari beberapa komponen yang terpisah secara fisik tetapi saling terhubung melalui jaringan. Setiap komponen dapat dikendalikan secara lokal dan beroperasi secara mandiri. Sebagai contoh, server yang tersebar di berbagai tempat dapat bekerja sama untuk menyediakan layanan atau menyelesaikan tugas.  
  
Sistem Paralel: Sistem paralel berfokus pada penggunaan banyak prosesor dalam satu sistem untuk melakukan tugas yang sama secara bersamaan. Tujuannya adalah untuk meningkatkan kecepatan pemrosesan dengan membagi beban kerja ke banyak prosesor dalam satu mesin.  
  
Perbedaan Kunci:  
- Pengolahan Paralel: Tugas dibagi di dalam satu sistem menggunakan beberapa prosesor.  
- Pengolahan Terdistribusi: Tugas dibagi di antara berbagai komputer yang saling terhubung dalam jaringan.

# 2. Hadoop sebagai Framework dan Platform Software

Hadoop sebagai Framework Software: Hadoop menyediakan struktur dan alat dasar untuk memproses data besar secara terdistribusi. Contohnya, Hadoop menggunakan metode seperti MapReduce dan HDFS untuk penyimpanan data.  
  
Hadoop sebagai Platform Software: Hadoop juga menyediakan lingkungan untuk menjalankan aplikasi big data, seperti Hive dan HBase, yang dapat berjalan di atas sistem Hadoop.  
  
Perbedaan antara Framework dan Platform:  
- Framework: Menyediakan instruksi dan alat untuk membangun aplikasi (contoh: Hadoop MapReduce).  
- Platform: Tempat berbagai aplikasi bekerja dan berinteraksi (contoh: Hadoop yang mendukung aplikasi seperti Hive dan Pig).

# 3. HDFS (Hadoop Distributed File System)

HDFS adalah sistem file terdistribusi yang digunakan untuk menyimpan data dalam blok-blok besar dan mendistribusikannya ke berbagai mesin dalam kluster.  
  
Direktori di Komputer: HDFS memiliki struktur folder dan file yang mirip dengan sistem file konvensional, namun data dibagi dan didistribusikan ke berbagai mesin dalam kluster.  
  
File System: HDFS menyimpan data dalam blok besar yang kemudian disalin ke berbagai mesin untuk memastikan data tetap tersedia meskipun ada kegagalan pada beberapa node.

# 4. Metode MapReduce (MR)

MapReduce adalah metode untuk memproses data besar dalam sistem terdistribusi secara paralel. Proses MapReduce terdiri dari beberapa langkah:  
  
1. Step Map: Data input dibagi menjadi potongan-potongan kecil yang disebut splits. Fungsi map menangani setiap split untuk mengubah data menjadi pasangan kunci-nilai.  
2. Step Shuffle and Sort: Setelah pemrosesan map, data dengan kunci yang sama dikelompokkan dan disusun dalam tahap shuffling.  
3. Step Reduce: Data yang telah dikelompokkan diproses lebih lanjut untuk menghasilkan output akhir yang menggabungkan data yang serupa menjadi hasil yang lebih terstruktur.

# 5. Perbedaan antara YARN dan MapReduce

MapReduce (MR): Merupakan model pemrograman yang memecah tugas menjadi dua bagian utama: Map dan Reduce. MapReduce digunakan untuk mengolah data besar dalam sistem yang terdistribusi secara paralel.  
  
YARN (Yet Another Resource Negotiator): Merupakan manajer sumber daya yang mengelola dan mengatur sumber daya dalam kluster Hadoop, seperti memori dan CPU. YARN bertanggung jawab untuk mengelola sumber daya untuk aplikasi Hadoop yang berjalan, termasuk MapReduce dan aplikasi tambahan seperti Spark.  
  
Perbedaan Kunci:  
- MapReduce: Mekanisme pemrograman yang mengatur bagaimana data diproses.  
- YARN: Pengelola sumber daya yang memastikan aplikasi dapat berjalan dengan memanfaatkan sumber daya komputasi yang tersedia dalam kluster Hadoop.