

Hadoop

Hadoop adalah library software (pustaka software) yang merupakan framework open source dari bahasa pemrograman Java di bawah lisensi Apache yang digunakan untuk melakukan pemrosesan big data menggunakan model pemrograman sederhana.

Hadoop dirancang untuk melakukan komputasi data dari satu server sampai ribuan server yang terhubung antara satu sama lainnya. Hal ini dapat memberikan kemudahan dari sisi penyimpanan data untuk melakukan analisis data. Selain itu Hadoop dapat memberikan informasi real time untuk mendeteksi kesalahan terkait kegagalan akses dan ketersediaan data pada masing-masing server.

A. Pengertian dan Fungsi Hadoop dari Produk Utamanya

Hadoop merupakan sebuah framework yang terus dikembangkan untuk melakukan pemrosesan big data. Berikut produk utama yang dikembangkan dalam Hadoop.

1. Hadoop Common

Hadoop Common adalah library-library umum yang mendukung library lainnya untuk dapat digunakan. Ini terkait perintah-perintah dasar yang ada pada Hadoop.

2. Hadoop Distributed File System (HDFS™)

Berbeda dengan system file data pada umumnya yaitu FAT32 dan NTFS yang dapat menyimpan 1 file data berkisaran antara 4 GB hingga 16 TB. HDFS adalah format sistem file yang dapat menampung 1 file data yang sangat besar dengan mengecilkan cluster sekelompok host data storage.

3. Hadoop YARN

Hadoop YARN adalah framework yang digunakan untuk mengatur pekerjaan secara terjadwal (*schedule*) dan manajemen cluster data.

4. Hadoop MapReduce

Hadoop MapReduce adalah paradigma pemrosesan data yang mengambil spesifikasi big data untuk menentukan bagaimana data tersebut dijadikan input dan output untuk diterapkan. MapReduce terintegrasi erat dengan HDFS untuk menyimpan data yang diperlukan
