

BAB I

PENDAHULUAN

1.1 Latar belakang

Seiring perkembangan teknologi informasi yang cepat memberikan konsekuensi pertumbuhan dan peningkatan jumlah data, di dalam sebuah riset dari Vcould News mencatat pada tahun 2015, pertumbuhan data statistik per hari mencapai 2,5 quintillion (10 pangkat 18) byte atau sekitar Giga Byte (GB) perdetik. Pada tahun 2018 pertumbuhan data ini diproyeksikan mencapai 50 ribu GB perdetik [1]. Dengan angka sebesar ini membuat suatu perusahaan atau organisasi mencari cara untuk menyimpan serta mengolah data yang banyak tersebut. Data yang banyak dapat diolah serta dianalisis sehingga melahirkan sebuah *value*. Nilai yang merupakan hasil analisis tersebut dapat dijadikan sebuah pertimbangan dalam pengambilan keputusan disuatu perusahaan atau organisasi.

Pertumbuhan data yang besar menyebabkan teknologi *big data* yang ada saat ini menjadi solusi untuk menyimpan serta mengolah data yang banyak tersebut, maka dibutuhkan juga suatu sistem yang mampu memanajemen data-data tersebut hingga menghasilkan suatu *value*. *Project open source* bernama Hadoop hadir sebagai salah satu sistem yang mampu memanajemen data secara terdistribusi.

Yahoo! dan Facebook termasuk kedalam salah satu perusahaan besar yang menggunakan Hadoop untuk mengelola data mereka. Vice President Infrastructure Engineering Facebook Jay Parikh berkata bahwa facebook mengelola *cluster* terbesar di dunia, *cluster* terbesar yang melampaui 4.000 mesin dan menyimpan lebih dari ratusan juta gigabyte[3]. Hadoop menyediakan infrastruktur umum untuk facebook dengan efisiensi dan kehandalan.

Pada tahun 2005, Doug Cutting dan Mike Cafarella menciptakan Hadoop saat bekerja pada perusahaan Yahoo!. Hadoop adalah inspirasi yang didapatkan dari mainan gajah kecil berwarna kuning milik anak Doug Couting. Hadoop versi 0.1.0 akhirnya rilis pada bulan April 2006, sampai versi terakhir Hadoop yang rilis pada Maret 2017 adalah Apache Hadoop 2.8. Pada versi terbaru ini, layanan yang diberikan Hadoop juga termasuk untuk HDFS (Hadoop Distributed File System), Yarn (Yet Another Resource Negotiator) dan MapReduce [2].

Hadoop sendiri menawarkan teknologinya yaitu HDFS (Hadoop Distributed File System) dimana data-data yang ada akan di distribusikan dalam bentuk *block-block* data untuk disimpan

dalam setiap media penyimpanan (*node*) dalam sebuah *cluster*. Penggunaan Hadoop di era *cloud computing* seperti saat ini sangatlah dibutuhkan karena hadoop sendiri menyediakan *file system* yang dapat menyimpan data dalam ukuran besar secara *scalable*. MapReduce juga termasuk kedalam project Apache Hadoop yang dimana HDFS dan MapReduce yang berguna untuk menyelesaikan permasalahan data besar berbasis *Java* dan *open source*.

Hadoop *file system* juga mengklaim dirinya memiliki fitur toleransi tinggi, ketersediaan tinggi, keandalan data, replikasi, *scalability*, dan *distributed storage*. Melalui tugas akhir ini, akan di implementasikan serta di analisis kinerja HDFS yang sudah dipaparkan dalam ruang lingkup yang sederhana serta mudah dipahami.

1.2 Perumusan Masalah

Dari paparan latar belakang mengenai beberapa fitur yang telah dibahas di atas, maka masalah yang dapat dirumuskan sebagai berikut:

“Bagaimana mengimplementasi dan menganalisa kinerja HDFS sebagai infrastruktur pembangunan *Big Data*”

Untuk menjawab rumusan tersebut, maka pada tugas akhir ini akan menguji beberapa hal, sebagai berikut:

1. Langkah-langkah apa saja yang dilakukan dalam implementasi HDFS sebagai infrastruktur pembangunan *Big Data*?
2. Faktor-faktor apa saja yang menentukan kinerja dari rancangan sistem HDFS yang diteliti?
3. Bagaimanakah kinerja HDFS sebagai infrastruktur pembangunan *Big Data*?

1.3 Tujuan dan Manfaat Penelitian

Penyusunan tugas akhir ini memiliki tujuan untuk:

1. Faktor-faktor yang digunakan untuk menentukan kinerja HDFS.
2. Melakukan penilaian atau pengukuran unjuk kerja HDFS.

Penyusunan tugas akhir ini memiliki manfaat untuk:

1. Memahami langkah-langkah yang dilakukan.
2. Mengetahui faktor-faktor yang menentukan kinerja HDFS.

3. Menghasilkan suatu karya tulis yang bisa menjadi rujukan didalam mengimplementasikan HDFS sebagai infrastruktur pembangunan *Big Data*.

1.4 Batasan Masalah

Pada penelitian kali ini, penulis membatasi penelitian pada hal-hal sebagai berikut:

1. Lingkungan penelitian adalah lingkungan Virtualisasi, dimana sistem yang dipasang perangkat Hadoop adalah *Virtual Machine*.
2. Sistem HDFS ini diimplementasikan dalam konsep sistem terdistribusi dengan jumlah *node* adalah 3.
3. Penilaian kinerja peneliti menggunakan peralatan atau *tools* yang sudah ada atau tidak membuatnya sendiri.
4. *Platform* Sitem Operasi yang digunakan Homogen (Linux).
5. Faktor yang dianalisa untuk menentukan kinerja HDFS adalah waktu eksekusi faktor I/O atau *read & write* menggunakan aplikasi *benchmark* Hadoop NNBench & TestDFSIO, pengaruh *block size* pada HDFS, dan pengujian *availability* atau ketersediaan data.

1.5 Sistematika Penulisan

Sistematika penulisan Tugas Akhir ini adalah sebagai berikut:

1. BAB I PENDAHULUAN, merupakan bab pembuka yang memberikan gambaran umum mengenai pelaksanaan Tugas Akhir. Bab ini terdiri dari latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah, dan sistematika penulisan
2. BAB II KAJIAN LITERATUR, bab ini akan mengkaji lebih dalam mengenai HDFS dalam pembangunan infrastruktur *Big Data*, tinjauan pustaka dan menjelaskan teori-teori yang digunakan dalam menyusun tugas akhir ini.
3. BAB III METODOLOGI PENELITIAN, bab ini akan menjelaskan tentang tahapan penelitian, mulai dari tahapan yang sudah dilakukan sampai tahapan yang akan dilakukan. Selain itu rancangan penelitian ini berisi bagaimana metode penyelesaian masalah yang ada, jenis penelitian yang dilakukan, pengumpulan data, dan rancangan *design* penelitian.
4. BAB IV ANALISIS DAN PERANCANGAN, bab ini berisi tentang rancangan sistem dan pengujian HDFS sebagai infrastruktur pembangunan *Big Data*.

5. BAB V IMPLEMENTASI DAN PENGUJIAN, bab ini berisi implementasi dari HDFS dan pengujian HDFS sebagai infrastruktur pembangunan *Big Data*.
6. BAB VI KESIMPULAN DAN SARAN, bab ini berisi kesimpulan dari Tugas Akhir yaitu inti dari jawaban pada rumusan masalah, dan saran untuk peneliti selanjutnya yang meneliti HDFS sebagai infrastruktur pembangunan *Big Data*.



STT - NF