

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Saat ini adalah era dimana teknologi sangat berkembang dengan cepat dan pesat. Seiring dengan berkembangnya teknologi yang ada, maka ukuran data yang diolah juga akan semakin besar.

Data yang memiliki ukuran yang semakin besar tentunya membutuhkan tempat penyimpanan yang sangat besar dan sistem pengelolaan yang tepat agar mudah dalam mengolahnya. Ada beberapa *platform* yang dapat digunakan untuk menyimpan dan mengolah data yang besar antara lain 1010data, Actian, Amazon Web Services (AWS), Cloudera, IBM SmartCloud, dan lain-lain. Untuk menganalisis data berukuran besar tersebut *big data* [1] menjadi kunci dasar persaingan, mendasari gelombang baru pertumbuhan produktivitas, inovasi, dan surplus konsumen. Untuk membangun sistem *super computer* yang berguna untuk menyimpan dan mengolah data yang berukuran besar (*big data*) secara baik dan cepat membutuhkan biaya yang tidak murah. Untuk mengatasi hal ini, maka *platforms* yang digunakan untuk menyimpan dan mengolah *big data* menggunakan sebuah sistem yang disebut *parallel computing*. [2]

*Parallel computing* adalah penggunaan beberapa komputer yang saling terhubung untuk mengolah data dalam ukuran yang besar. Salah satu *platform* yang masih sering digunakan sampai saat ini untuk mengolah data yang berukuran besar (*big data*) secara terdistribusi dan dapat berjalan diatas *cluster* adalah Hadoop. [2]

Penyimpanan HDFS (*Hadoop Distributed File System*) adalah *metadata*, merupakan struktur direktori HDFS dan *file* dalam bentuk *tree*. Hal ini juga mencakup berbagai atribut direktori dan *file* (kepemilikan, perizinan, kuota, dan faktor replikasi).

## 1.2 Rumusan Masalah

Berikut masalah–masalah yang mungkin akan terjadi dan perlu dijawab. Masalah–masalah yang ada sebagai berikut:

1. Bagaimana karakteristik penerimaan data dalam *client server*?
2. Apakah pengaruh ukuran *file* dibawah 500 MB terhadap waktu saat perpindahan data?
3. Bagaimana keterkaitan konfigurasi terhadap *block data* yang dikirim ke *slave*?

## 1.3 Tujuan

Tujuan yang dapat dari rumusan masalah adalah sebagai berikut:

1. Mengamati karakteristik penerimaan data dalam *client server*.
2. Mengamati pengaruh ukuran *file* dibawah 500MB terhadap waktu saat perpindahan data.
3. Mengamati dan mengukur *block data* yang dikirim ke *slave* berdasar konfigurasi pada *master*.

## 1.4 Ruang Lingkup Penelitian

Ruang lingkup dari penelitian ini terbatas pada hal-hal berikut ini:

1. Membahas karakteristik pada Hadoop dalam pengiriman data dari *master server* ke *client server*, mengamati keadaan *client server* apabila *master server* dimatikan secara tiba-tiba, mengamati apa yang terjadi bila ada penambahan *client server* baru.
2. Perancangan yang dibuat adalah dengan menjalankan Hadoop secara *multi node (cluster)*.
3. Physical machine yang digunakan terdiri dari lima komputer yang terdiri dari satu komputer sebagai *master server* dan empat komputer sebagai *client server* memiliki spesifikasi CPU Intel i5-2320 3.00ghz, RAM 4 GB dengan sistem operasi Windows 7.
4. *Virtual machine* yang digunakan memiliki spesifikasi CPU Intel i5-2320 3.00ghz, RAM 1GB dengan sistem operasi OS Linux Ubuntu 14.03.3 server amd64 yang berjalan diatas Oracle VM Virtual Box v5.0.2 pada setiap masing-masing komputernya.

5. Percobaan dilakukan pada satu komputer bertindak sebagai *master server* dan empat komputer sebagai *client server*.
6. Ukuran *file* yang digunakan untuk pengiriman data dibatasi hanya menggunakan tiga ukuran *file* yang berbeda yaitu 127MB, 235MB, 314MB semua dengan *file* format mp4.
7. *Block size* yang digunakan dibatasi hanya menggunakan ukuran 128 MB.
8. Jumlah replikasi pada penyebaran per blocknya dibatasi hanya menggunakan tiga replikasi.

### **1.5 Metodologi Penelitian**

Pada penelitian ini dilakukan berawal dengan mencari informasi mengenai Hadoop. Segala informasi mengenai Hadoop yang didapat melalui internet, buku atau *e-book*. Melakukan penginstalan dan konfigurasi Hadoop dengan beberapa server, menginstal *traffic monitoring* agar dapat dilakukan monitoring jaringan antara *server* atau *node* yang ada. Lalu dilakukan beberapa percobaan dan dibuat laporan mengenai hasil percobaan.

### **1.6 Sistematika Laporan**

Pada proses penyusunan laporan ini, sistematika penulisan yang akan digunakan adalah sebagai berikut :

## **BAB I – PENDAHULUAN**

Pada bab ini berisi latar belakang, rumusan masalah, tujuan, ruang lingkup penelitian, dan metodologi penelitian ini.

## **BAB II – DASAR TEORI**

Pada bab ini berisi penjelasan dari teori–teori yang didapat dalam sumber–sumber yang didapat yang menunjang percobaan yang akan dilakukan.

## **BAB III – ANALISA DAN PEMODELAN**

Pada bab ini berisi penjelasan permodelan rancangan dari penelitian beserta detail mengenai skenario yang akan dilakukan.

#### **BAB IV – HASIL IMPLEMENTASI**

Pada bab ini berisi settingan atau konfigurasi yang dibutuhkan pada sistem Hadoop.

#### **BAB V – PENGUJIAN**

Pada bab ini berisi penjelasan hasil pengujian dan analisis dari skenario rancangan penelitian yang terdapat pada bab 3.

#### **BAB VI – SIMPULAN DAN SARAN**

Pada bab ini berisi simpulan dan saran penelitian dari sistem Hadoop serta dari keseluruhan hasil penelitian yang dapat digunakan pada penelitian selanjutnya.

