

## 1. Pendahuluan

### 1.1 Latar Belakang

Pengguna media sosial di Indonesia semakin bertambah seiring dengan banyaknya pengguna *smartphone* (telepon pintar). Fungsi media sosial saat ini bukan hanya sekedar untuk berkomunikasi, berbagi informasi dan eksistensi. Masyarakat pun dapat melakukan jajak pendapat (*polling*) di media sosial. Metode *polling* menjadi metode yang paling efektif dalam mendapatkan informasi yang cepat mengenai masalah atau isu yang berkembang di masyarakat. *Polling* digunakan untuk mendapatkan informasi tentang suatu fenomena, dalam hal ini yang ingin didapat dari *polling* adalah sikap, pandangan, dan keyakinan masyarakat terhadap isu-isu yang berkembang [1]. Karena itu dapat juga dikatakan bahwa *polling* adalah penerapan praktis dari metode survei, pemakaian metode survei untuk mengukur pendapat publik seperti isu-isu politik. Media Sosial merupakan *platform* komunikasi yang memiliki kemampuan untuk mengetahui opini publik terhadap suatu isu secara cepat [1].

Biasanya untuk melakukan sebuah *polling* kita butuh setidaknya data yang cukup besar, untuk mendapatkan hasil yang maksimal. Data yang besar ini sering disebut juga dengan, istilah *Big Data*. *Big Data* merupakan kumpulan data yang besar, sangat variatif, dan mungkin tidak terstruktur [2]. *Big Data* sangatlah besar sehingga sulit untuk prosedur konvensional di dalam menganalisa *big data*. Dengan data yang besar, dapat menjadikan analisa terhadap suatu fenomena lebih sempurna, dan jika berhasil menganalisa sebuah data tersebut akan membantu di dalam pengambilan sebuah keputusan dengan lebih baik [2]. Yang diharapkan juga data hasil analisa bisa efisien mewakili suara yang didapat, kost lebih murah, dan kecepatan di dalam pemrosesan. Sehingga diperlukan algoritma khusus sehingga informasi yang mendalam mudah didapatkan dan membantu di dalam pengambilan keputusan yang lebih baik.

Hadoop merupakan salah satu sistem terdistribusi yang diperuntukan untuk memproses data berukuran besar [3]. *Hadoop* merupakan sebuah *framework* untuk penyimpanan dan pemrosesan data skala besar yang terdiri dari beberapa modul, yaitu *Hadoop Common*, *Hadoop Distributed File System* (HDFS), *Hadoop YARN*, dan *Hadoop MapReduce* [4].

Tugas Akhir ini mengimplementasikan metode *MapReduce* pada *Framework Hadoop* yang digunakan untuk memproses data *polling* supaya lebih efisien dalam segi waktu proses data. Data yang akan diproses berasal dari *web* ([happypoll.id](http://happypoll.id)), dimana untuk responden yang akan melakukan *polling* diharuskan untuk *login* terlebih dahulu. Hal tersebut untuk mendapatkan informasi, diantaranya alamat, *username*, *email*, dan informasi akun *twitter* responden lainnya. Pengolahan data memanfaatkan komputasi *MapReduce* dengan program *wordcount*, yang mana dilakukan beberapa kali pengujian. Yaitu pengolahan data *polling* dengan program *wordcount* dengan *MapReduce* pada *Hadoop* dan program *wordcount* tanpa *MapReduce* atau program *wordcount* biasa dengan bahasa java. Hal ini dilakukan untuk mengetahui seberapa efektif jika *Hadoop* digunakan untuk proses data *polling* berbasis media sosial pada Tugas Akhir ini.

### 1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, terdapat beberapa permasalahan yang dibahas dalam Tugas Akhir ini, yaitu sebagai berikut:

1. Bagaimana merancang sistem untuk pengolahan data *polling* berbasis media sosial dengan memanfaatkan *MapReduce* pada *Hadoop*?
2. Bagaimana program implementasi program *wordcount* dengan *MapReduce* pada *Hadoop* untuk proses data *polling*?
3. Bagaimana kecepatan proses data *wordcount* dengan *MapReduce* dan *wordcount* tanpa *MapReduce*?
4. Apakah *Block* dapat mempengaruhi proses data pada *Hadoop* yang di-*setting* dengan *Multi Node Cluster*?

### 1.3 Batasan Masalah

Batasan masalah dalam Tugas Akhir ini adalah:

1. Data yang digunakan merupakan data hasil *polling* di *web happypoll.id*, dimana *voters* harus *login* terlebih dahulu melalui *twitter*.
2. Data *polling* berbentuk file *.text* yang disesuaikan.
3. Pengolahan data memanfaatkan komputasi *MapReduce* pada *Framework Hadoop* dengan program *wordcount*.
4. Dilakukan pengujian untuk membandingkan program *wordcount* dengan *MapReduce* pada *Hadoop* dan program *wordcount* tanpa *Hadoop*.
5. Dilakukan pengujian pada *Hadoop* dengan ukuran *block* yang diperbesar.

### 1.4 Tujuan

Adapun tujuan dari Tugas Akhir ini, yaitu sebagai berikut:

1. Membuat sistem pengolahan data *polling* berbasis media sosial, dengan memanfaatkan *MapReduce* pada *framework Hadoop*.
2. Menjalankan aplikasi *wordcount* yang berjalan dengan komputasi *MapReduce* pada *Framework Hadoop* yang dikonfigurasi secara *MultiNode Cluster*.
3. Melakukan pengujian program *wordcount* dengan *MapReduce Hadoop* dan tanpa *MapReduce Hadoop*, untuk mengetahui metode mana yang lebih unggul dari segi kecepatan proses data.
4. Melakukan pengujian pada program *wordcount* dengan *MapReduce Hadoop*, dengan memperbesar ukuran *block* untuk mengetahui pengaruh dari *block*.

### 1.5 Organisasi Penulisan

Penelitian TA ini disusun dengan struktur sebagai berikut: Bagian pertama berupa pendahuluan yang berisi mengenai latar belakang, perumusan masalah, batasan masalah, tujuan, dan organisasi penulisan. Bagian kedua berisi mengenai studi terkait, referensi-referensi yang berkaitan dengan TA penulis. Bagian ketiga berisi tentang penjelasan sistem yang dibangun. Bagian keempat mengenai analisi dan evaluasi sistem yang dibangun. Bagian ke lima berisi kesimpulan dan yang terakhir daftar pustaka.