

## BAB 6

### KESIMPULAN DAN SARAN

#### 6.1 Kesimpulan

Pada skripsi ini telah berhasil dibangun perangkat lunak yang terdiri dari beberapa fungsi untuk mengimplementasikan RHadoop. Selain itu juga berhasil dilakukan analisis data studi kasus menggunakan RHadoop. Setelah melaksanakan serangkaian penelitian pada karya ilmiah ini, penulis mengambil beberapa kesimpulan, diantaranya adalah :

- Setelah dilakukan studi literatur tentang analisis statistika berdasarkan variabelnya dapat disimpulkan bahwa analisis statistika dibutuhkan untuk mendapatkan informasi yang berguna dari big data. Pada eksperimen studi kasus telah dilakukan analisis statistika berdasarkan beberapa variabel menggunakan analisis univariat, analisis bivariat, dan analisis multivariat.
- Dengan melakukan studi literatur R dan statistika dapat disimpulkan bahwa R memiliki banyak *package* untuk kebutuhan menganalisis data dalam pendekatan statistik.
- Dengan melakukan studi literatur RHadoop, dapat disimpulkan bahwa RHadoop dapat menjadi alternatif dari mapreduce Hadoop pada sistem terdistribusi. RHadoop menggunakan bahasa R untuk menjalankan mapreduce *job*. R memiliki perintah yang sederhana untuk manipulasi data sehingga pengguna dimudahkan saat membuat kode mapreduce.
- Telah dibangun modul-modul RHadoop untuk analisis big data. RHadoop berjalan diatas Hadoop sehingga pengguna harus mengerti kekurangan dan kelebihan dari Hadoop saat membangun modul-modul RHadoop.
- Berdasarkan hasil pengujian kinerja, RHadoop dapat berjalan dengan baik dalam menganalisis big data. Hal ini ditunjukkan dengan waktu eksekusi pada pengujian yang masih berada dibawah kelipatan ukuran datanya. Terdapat satu fungsi yang kinerjanya tidak baik karena tidak menjamin skalabilitas, yaitu fungsi kuartil. Untuk menghitung kuartil suatu variabel harus diurutkan terlebih dahulu, sedangkan mapreduce tidak efektif dalam mengurutkan nilai karena semua nilai harus masuk kedalam satu reducer. Jika semua nilai harus diurutkan dalam 1 reducer maka mapreduce tidak bekerja secara paralel dan tidak cocok digunakan untuk data yang ukurannya besar.

#### 6.2 Saran

Setelah melaksanakan serangkaian penelitian pada karya ilmiah ini, berikut saran yang dapat penulis berikan :

- Untuk mendapatkan hasil eksekusi waktu yang lebih baik dapat menggunakan perangkat keras dengan spesifikasi yang lebih baik dan menambahkan jumlah node pada cluster Hadoop.
- Fungsi-fungsi RHadoop yang telah dibangun hanya menerima *input file* dalam bentuk csv. Akan lebih baik jika fungsi-fungsi tersebut dapat menerima input dalam bentuk lain seperti json, avro, dll.

- Pada skripsi ini penulis menggunakan *package* `rmr2` dan `rhdfs` pada RHadoop untuk menganalisis big data. Akan lebih baik jika dilakukan implementasi menggunakan *package* lainnya seperti `ravro`, `rhbase`, dan `plyrmr`.