

BAB 6

KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari awal hingga akhir penelitian beserta saran untuk penelitian selanjutnya.

6.1 Kesimpulan

Kesimpulan yang dapat ditarik dari awal penelitian ini sampai selesai adalah sebagai berikut:

- Pada penelitian ini, telah dipelajari algoritma *Hierarchical Agglomerative Clustering*.
- Pada penelitian ini, telah diimplementasikan algoritma *Hierarchical Agglomerative Clustering* pada lingkungan Spark dengan menggunakan *transformation* dan *actions*. Fungsi *map()*, *groupByKey()*, *flatMap()* dapat digunakan untuk menggantikan fungsi *map()* dan *reduce()* pada MapReduce.
- Pada penelitian ini, telah dilakukan pengukuran kinerja Spark dengan membandingkan waktu eksekusi Spark dan Hadoop.
- Pada penelitian ini, telah dilakukan eksperimen perbandingan performa antara perangkat lunak Spark dan Hadoop. Dari hasil pengujian dapat disimpulkan bahwa perangkat lunak Spark memiliki performa yang lebih baik asalkan diatur dan dikonfigurasi dengan benar. Waktu eksekusi Spark lebih cepat dibanding Hadoop karena Spark menyimpan data pada memori, sebaliknya Hadoop banyak melakukan proses I/O kepada disk yang membuat Hadoop lambat. Proses *shuffling* dan *sorting* menghambat dan meningkatkan waktu eksekusi Hadoop ketika jumlah partisi ditingkatkan. Waktu eksekusi Hadoop akan meningkat seiring meningkatnya jumlah partisi. Sebaliknya, waktu eksekusi Spark menurun ketika jumlah partisi ditingkatkan. Dengan meningkatkan jumlah partisi pada Spark, data akan lebih terdistribusi. Hal ini akan mengurangi waktu yang dibutuhkan untuk mengirim data dari satu komputer ke komputer lain dan mengurangi waktu komunikasi.
- Pada penelitian ini, telah dibangun perangkat lunak untuk melihat hasil reduksi data dan menjalankan proses reduksi data.

6.2 Saran

Saran untuk penelitian selanjutnya adalah sebagai berikut:

- Pada penelitian ini, Spark dijalankan pada Hadoop *YARN*. Oleh karena itu, penulis berharap agar penelitian selanjutnya dapat menguji performa perangkat lunak pada Spark *cluster* atau *cluster* lainnya.
- Pada penelitian ini, pengujian yang dilakukan masih terbatas dengan 10 *worker* dan ukuran data sampai 20GB. Untuk penelitian selanjutnya, penulis berharap agar pengujian yang dilakukan dapat menggunakan jumlah *worker* dan data yang lebih besar.

DAFTAR REFERENSI

- [1] Moertini, V. S., Suarjana, G. W., Venica, L., dan Karya, G. (2018) Big data reduction technique using parallel hierarchical agglomerative clustering. *IAENG International Journal of Computer Science*, **45**, 188 – 205.
- [2] Ishwarappa dan J, A. (2015) A brief introduction on big data 5vs characteristics and hadoop technology. *Procedia Computer Science*, **48**, 319 – 324.
- [3] Jain, A. K. dan Dubes, R. C. (1988) *Algorithms for Clustering Data*. Pearson College Div, New Jersey.
- [4] Holmes, A. (2012) *Hadoop in Practice*. Manning, New York.
- [5] White, T. (2015) *Hadoop The Definitive Guide*, 4th edition. O'Reilly Media, Sebastopol.
- [6] Lam, C. (2010) *Hadoop in Action*. Manning Publications, New York.
- [7] Karau, H., Konwinski, A., Wndell, P., dan Zaharia, M. (2015) *Learning Spark*, 1th edition. O'Reilly Media, Sebastopol.