

## BAB 6

### KESIMPULAN DAN SARAN

#### 6.1 Kesimpulan

Setelah melaksanakan serangkaian penelitian pada karya ilmiah ini, penulis mengambil beberapa kesimpulan, diantaranya adalah:

- Dengan melakukan studi literatur Apache Spark dan Hadoop terutama HDFS, maka dapat dipahami bahwa Apache Spark merupakan alternatif dari MapReduce Hadoop pada sistem terdistribusi. Apache Spark mampu menjalankan berbagai komputasi paralel di memori. Namun tetap mempertahankan model Map dan Reduce yang sudah ada pada Hadoop. Apache Spark dapat berjalan independen dan lingkungan Hadoop. Namun, Spark tidak memiliki sistem penyimpanan. Hadoop menyediakan *file system* dan sekaligus menyediakan *resource* bagi Spark.
- Apache spark memiliki *library* MLlib yang menyimpan berbagai macam implementasi algoritma pengolahan data. Beberapa fungsi diantaranya adalah *statistic, classification, regression, collaborative filtering, clustering, Dimensionality reduction, dan feature extraction*. Dengan demikian, MLlib mampu melakukan analisis big data dengan menyediakan kelas-kelas yang siap dipanggil untuk teknik pengolahan data tertentu.
- Salah satu fungsi MLlib yaitu k-means memiliki kebutuhan yang belum tersedia, maka dilakukan modifikasi pada *source code* MLlib. Hal ini dilakukan dengan menganalisis *source code* Apache Spark MLlib, merancang modifikasi yang akan dilakukan, dan mengimplementasikan pada perangkat lunak demo. Implementasi modifikasi dapat berjalan sesuai fungsinya setelah dilakukan pengujian fungsional.
- Berdasarkan hasil pengujian performa, Apache Spark MLlib dapat berjalan dengan baik dalam menganalisis big data. Hal ini ditunjukkan dengan waktu eksekusi pada pengujian yang masih berada dibawah kelipatan ukuran datanya.
- Pengembangan perangkat lunak untuk mempraolah data diimplementasikan dengan menggunakan API Spark meliputi Spark Core dan Spark MLlib dan berjalan pada *platform* Java dengan bahasa pemrograman Scala. Fungsi-fungsi pada MLlib dipanggil untuk menganalisis data dan menampilkan hasil analisis data pada *graphical user interface* (GUI) yang disediakan perpustakaan Swing Scala.

#### 6.2 Saran

Setelah melaksanakan serangkaian penelitian pada karya ilmiah ini, penulis mengambil beberapa saran, diantaranya adalah:

- Berdasarkan hasil eksperimen yang telah dilakukan, perlu dilakukan pengujian dengan variasi jumlah node cluster yang lebih besar yaitu 10 hingga 20 node.

- Apache Spark MLlib ini dijalankan pada cluster Hadoop agar dapat memanfaatkan *file system* yang dimiliki Hadoop.

## DAFTAR REFERENSI

- [1] Odersky, M., Spoon, L., dan Venners, B. (2016) *Programming in Scala: Updated for Scala 2.12*, 3rd edition. Artima Incorporation, USA.
- [2] Tanenbaum, A. S. dan Steen, M. v. (2006) *Distributed Systems: Principles and Paradigms (2Nd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- [3] Lam, C. (2010) *Hadoop in Action*, 1st edition. Manning Publications Co., Greenwich, CT, USA.
- [4] Holmes, A. (2012) *Hadoop in Practice*. Manning Publications Co., Greenwich, CT, USA.
- [5] Dean, J. dan Ghemawat, S. (2004) Mapreduce: Simplified data processing on large clusters. *Proceedings of the 6th Conference on Symposium on Operating Systems Design & Implementation - Volume 6*, Berkeley, CA, USA OSDI'04, pp. 10–10. USENIX Association.
- [6] Yadav, R. (2017) *Apache Spark 2.X Cookbook: Cloud-ready Recipes for Analytics and Data Science*, 2nd edition. Packt Publishing.
- [7] Abbasi, M. A. (2017) *Learning Apache Spark 2.0*. Packt Publishing.
- [8] Karau, H., Konwinski, A., Wendell, P., dan Zaharia, M. (2015) *Learning Spark: Lightning-Fast Big Data Analytics*, 1st edition. O'Reilly Media, Inc.
- [9] Harrington, P. (2012) *Machine Learning in Action*. Manning Publications Co., Greenwich, CT, USA.
- [10] Shalev-Shwartz, S. dan Ben-David, S. (2014) *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, New York, NY, USA.
- [11] Kaufman, L. dan Rousseeuw, P. (1990) *Finding Groups in Data: an introduction to cluster analysis*. Wiley.
- [12] Alexander, A. (2013) *Scala Cookbook: Recipes for Object-Oriented and Functional Programming*, 1st edition. O'Reilly Media, Inc.