

Nodes, Compute Nodes, Index Maintenance Nodes and disk arrays are easy to expand or shrink, and the original architecture of system will not be changed. Special nodes are used for Maintaining and querying indices which are time-consuming, which remarkably improves the parallelism of system and lower the data-processing time.

Two running models (the Direct Storage Access Model and the Storage Space Mapping Model) are proposed on the basis of the architecture of this platform. These two models have different hardware requirements. The performance and feature of platform under different models also are different. The Direct Storage Access Model needs to access data from storage when executing tasks, so it is more suitable for the computing-intensive applications, while the Storage Space Mapping Model does not need to access data from storage, it, therefore, is more suitable for the I/O-intensive applications.

In future work, we will build a prototype of this platform, and test the performance of two running models in this prototype. At the same time, we will further to find the application fields which are suitable for this platform, and adjust and optimize platform according to different application circumstances.

5. References

- [1] ORACLE, <http://www.oracle.com/index.html>.
- [2] Sanjay Ghemawat , Howard Gobioff , Shun-Tak Leung, The Google file system, Proceedings of the nineteenth ACM symposium on Operating systems principles, October 19-22, 2003, Bolton Landing, NY, USA.
- [3] Fay Chang , Jeffrey Dean , Sanjay Ghemawat , Wilson C. Hsieh , Deborah A. Wallach , Mike Burrows , Tushar Chandra , Andrew Fikes , Robert E. Gruber, Bigtable: A Distributed Storage System for Structured Data, ACM Transactions on Computer Systems (TOCS), v.26 n.2, p.1-26, June 2008.
- [4] Jeffrey Dean , Sanjay Ghemawat, MapReduce: a flexible data processing tool, Communications of the ACM, v.53 n.1, January 2010.
- [5] Hadoop, <http://hadoop.apache.org/>, 2012.11.18.
- [6] Cassandra. <http://incubator.apache.org/cassandra/>.
- [7] A.Thusoo, J. Sarma, N. Jain, Z. Shao, P. Chakka, S. Anthony, H. Liu, P. Wyckoff and R. Murthy, "Hive – A Warehousing Solution Over a MapReduce Framework," VLDB, Lyon, France, August 2009, pp. 1626-1629.
- [8] B.Cooper, R. Ramakrishnan, U.Srivastava, A. Silberstein, P. Bohannon, H. Jacobsen, N. Puz, D. Weaver and R. Yerneni, "PNUTS: Yahoo!'s Hosted Data Serving Platform," VLDB, Auckland, New Zealand, August 2008, pp. 1277-1288.
- [9] Yingjie Shi , Xiaofeng Meng , Jing Zhao , Xiangmei Hu , Bingbing Liu , Haiping Wang, Benchmarking cloud-based data management systems, Proceedings of the second international workshop on Cloud data management, October 30-30, 2010, Toronto, ON, Canada.
- [10] WANG Shan, WANG Hui-Ju, QIN Xiong-Pai, ZHOU Xuan, "Architecting Big Data: Challenges, Studies and Forecasts," CHINESE JOURNAL OF COMPUTERS, Vol. 34, No.10, pp. 1741-1752, Oct. 2011.
- [11] TDWI Checklist Report: Big Data Analytics, <http://tdwi.org/research/2010/08/Big-Data-Analytics.aspx>.