International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-3 Issue-2, July 2013

The Design and Test of a Private Cloud Storage System, Part I

Zhaorui Wang

***Abstract*—Currently, cloud computing in the development stage, and technology is not mature. The core technology tends to be open source software. For security reasons or for mistrust of public cloud, some companies or organizations prefer to build their own private cloud, while reluctant to use public cloud services. In this context, this paper presents a general architecture, which well meets the needs of enterprises or institutions. This private cloud storage system has a highly scalable, high stability and high concurrent processing capability and low cost advantages. The key part of this architecture chooses ParaStor and MongoDB, and performance was tested when they work together and benchmarking tool YCSB was used to do the test. The test results indirectly illustrate the feasibility of the program.**

***Index Terms*—cloud computing, private cloud and YCSB.**

I. INTRODUCTION

Open source cloud computing, such as OpenStack [1], and Eucalyptus [2], presents the trend of core and allows anyone to build and deliver their own cloud computing services and to build a firewall within the "private cloud" (Private Cloud). It offers various departments within the organization or enterprise shared resources. But these open source software has high entry barriers, the popularity of the domestic usage is not high. Due to reasons, such as language barriers, sometimes nowhere to find help, security reasons or mistrust of private cloud, some companies or organizations prefer to build their own private cloud. In this context, this paper, we propose a private cloud storage system design. The recommended software is the industry's more commonly used open source software. To test our proposed system, we used data from a wireless sensor network (WSN). WSN usually have low-power sensors while cloud usually has super computation power [3][4]. Therefore, WSN can be combined with cloud to perform as an integrated system. For example, cloud can be used to handle the data from many sensors. WSN has many application and operation scenarios [5]-[12]. For example, sensors can be used to gather date for detection [11] and for estimation [8][9]. Usually, if sensors are used for long-term monitoring, the data can be huge and to handle the data, cloud computing may be needed. That's why the proposed system is tested using data from WSNs. Tools and software used in this framework can be found in [13]-[19]. For more research about cloud computing, detection techniques and WSNs, readers can refer to [20]-[48].

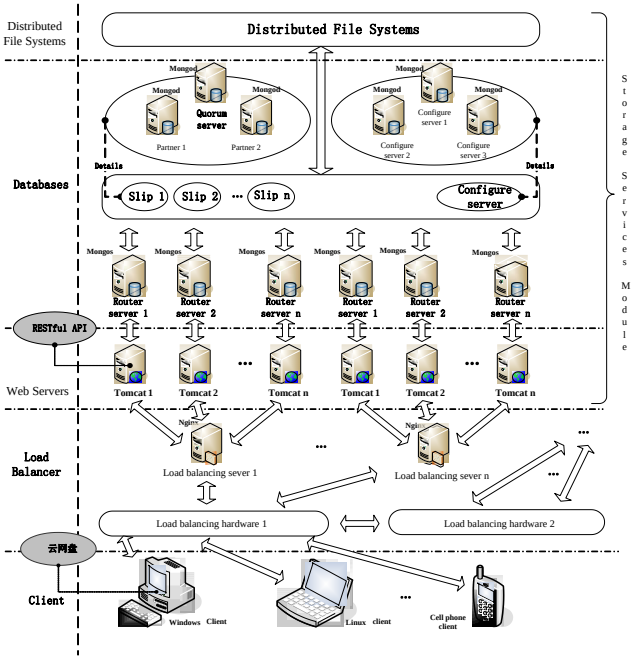

Figure 1. The private Cloud storage system design

II. PROGRAM FRAMEWORK

Figure 1 shows, private cloud storage system consists of five major components: distributed file systems, databases, Web servers, load balancer and Client. Distributed file system and database part is also known as Storage Services Module. The whole design uses a redundant design.

The design of the distributed file system uses of open source software Lustre [13] or HDFS [14] to build. Part of the database is using the popular NoSQL database, MongoDB [15], and configured the cluster approach. Web server is used Tomcat [16], which can run the REST architectural style API program. The system provides developers API. Load balancer using Nginx [17], and then one needs to decide whether to introduce load balancing hardware and with dual hot standby. In client part, one can develop his own cloud network disk, cloud notes software, etc.. Software used to build the entire system, except for the client part, is open source software. The design has the following advantages:

1) multi-level scalability. As a result of a distributed file system, storage capacity can be dynamically extended laterally. Database uses the replica set technology, so you can dynamically expand horizontally. Web server and load balancer can expand as needed.

2)

Manuscript published on 30 July 2013.
***Correspondence Author(s)**
Dr. Zhaorui Wang, National Astronomical Observatories, Chinese Academy of Sciences, Beijing, China.

High reliability. Distributed File System, MongoDB clusters, Web servers, load balancers allow partial node failure, which does not affect the normal operation of the entire system.

3) High concurrent processing capabilities. With the introduction of load balancing and load balancing part only perform distribution requests, thereby improving the system's parallel processing capabilities.

4) Low cost. Since the program are recommended to use open source software and the hardware used is not special, you can effectively utilize existing equipment to build low cost system. General business or organization can accept it, and capacity expansion means a horizontal expansion and will not cause hardware waste related to vertical expansion.

III. Performance Testing

The whole design is mainly determined by the performance of the bottom design of the distributed file system and database two parts. But under normal circumstances, the database data is stored in a distributed file system, so the following test is done on the distributed file system and database performance when they work together. Testing tools use Yahoo's cloud service benchmarking tool YCSB [8], distributed file system uses the company's domestic dawn ParaStor [9] distributed parallel file systems, and databases uses MongoDB.

The Purpose of testing is to understand the performance indicators when MongoDB and ParaStor work together in the current hardware and software, namely:

1) understand the number of concurrent and throughput the MongoDB can support.

2)understand whether ParaStor work with MongoDB normally, and whether it can meet the performance requirements.

YCSB stands for Yahoo! Cloud Serving Benchmark, which was developed by Yahoo as a cloud service performance benchmarking tool. If YCSB performance testing is used, the test results can be compared using the horizontal alignment, which facilitates exchange of a higher degree of recognition. YCSB easily be extended, can easily add support for new systems, and can be set to any read and write ratio, any number of threads, and has flexible definition of workloads.

A. Test environment and programs

Test environment

1)MongoDB version：v2.2.3（64bits）

2)YCSB version：v0.1.4（Based on the latest source compiler）

3)ParaStor Client-driven version：1.5.3.1.18370.3-Built on Mon Mar 4 08:34:40 CST 2013

4) Server Configuration: Using Dawn 6000 YG partition as a server blade (Inter X5650 dual-CPU, 48G DDR3 memory, 146G SAS HDD, Gigabit Ethernet card, mount ParaStor parallel file system, operating system is 64-bit SLES 11 SP1)

5) Client Configuration: Using Dawn 6000 YG partition virtual machine as a test client (quad-core CPU, 24G memory, 32G system disk, 1000G data disk, Gigabit Ethernet card, the operating system is 64-bit SLES 11 SP1)

Testing program

Testing program template: MongoDB on the data stored in the *** for Workload a, Workload b, Workload c and Workload f four kinds of load testing # # # insert records, read, update, use 8 different threads (1,5, 10,20,30,50,75,100) test running time were recorded, operands, throughput, average delay;

Test scenario template were filled ParaStor and *** at the local hard disk, # # # were filled at 100,000 and one million, four test solutions obtained. Workload a, Workload b, Workload c, Workload f compared with four typical data load mode, where:

1) Workload a: 50% + 50% read the update operation

2) Workload b: 95% + 5% read the update operation

3) Workload c: 100% read operations (read-only access)

4) Workload f: 50% + 50% read operations read - modify - write operations

B. Preparations

1) Download and install MongoDB

From MongoDB's official website to download precompiled binary packages 64 (http://fastdl.mongodb.org/linux/mongodb-linux-x86_64-2.2.3.tgz), Extract the appropriate directory, complete the installation. Run the following command to start MongoDB:

mongod --dbpath /home-yg/data/db --logpath /home-yg/logs/mongod.log --fork --rest.

2) YCSB compile and install

The source library from YCSB (https://github.com/brianfrankcooper/YCSB)

Download the latest source code (YCSB-master.zip), unzip, get YCSB-master directory in the directory, enter "mvn clean package" command, and perform the compilation. After compilation, you can go to distribution / target directory to find prepackaged binary packages ycsb-0.1.4.tar.gz, extract it to the appropriate directory, and complete the installation. You can also use pre-compiled binary packages.

IV. Conclusions

Program testing is a complex process. Due to space issues, program testing steps and results will be discussed in the second part of the paper.

References

1. OpenStack. Available: http://www.openstack.org/.
2. Eucalyptus. Available: http://www.eucalyptus.com/.
3. K. Ahmed, and M. Gregory, "Integrating Wireless Sensor Networks with Cloud Computing," 2011 Seventh International Conference on Mobile Ad-hoc and Sensor Networks (MSN), pp.364-366, Dec. 2011.
4. W. Kurschl and W. Beer, “Combining cloud computing and wireless sensor networks”, in Proceedings of the 11th ACM International Conference on Information Integration and Web-based Applications & Services, New York, NY, USA, pp. 512-518.
5. Z. X. Luo and T. C. Jannett, “Optimal threshold for locating targets within a surveillance region using a binary sensor network,” in *Proc. of the International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering*, Dec., 2009.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

6. Z. X. Luo and T. C. Jannett, "Energy-based target localization in Multi-hop wireless sensor networks," in *Proc. of the 2012 IEEE Radio and Wireless Symposium*, Santa Clara, CA, Jan. 2012.
7. H. Chen and P. K. Varshney, "Nonparametric quantizers for distributed estimation," IEEE Trans. Signal Process., vol 58, no 7, pp. 3777-3787, July 2010.
8. Z. X. Luo and T. C. Jannett, "Modeling sensor position uncertainty for robust target localization in wireless sensor networks," in *Proc. of the 2012 IEEE Radio and Wireless Symposium*, Santa Clara, CA, Jan. 2012.
9. H. Chen and P. K. Varshney, "Performance Limit for Distributed Estimation Systems with Identical One-Bit Quantizers," IEEE Transaction on Signal Processing, Vol. 58, No. 1, pp. 466-471, Jan. 2010.
10. Z. X. Luo and T. C. Jannett, "Performance comparison between maximum likelihood and heuristic weighted average estimation methods for energy-based target localization in wireless sensor networks," in *Proc. of the 2012 IEEE SoutheastCon Conference*, Orlando, FL, Mar. 2012.
11. Q. Cheng, P. K. Varshney, J. H. Michels, and C. M. Belcastro, "Distributed fault detection with correlated decision fusion," IEEE Trans. Aerosp. Electron. Syst., Volume 45, No. 4, pp.1448 – 1465, October 2009.
12. Z. X. Luo and T. C. Jannett, "A multi-objective method to balance energy consumption and performance for energy-based target localization in wireless sensor networks," in *Proc. of the 2012 IEEE SoutheastCon Conference*, Orlando, FL, Mar. 2012. Lustre. Available: http://lustre.org/.
13. HDFS. Available: http://hadoop.apache.org/.
14. MongoDB Available:http://www.mongodb.org/.
15. Tomcat. Available: http://tomcat.apache.org/.
16. Nginx. Available: http://nginx.org/.
17. Yahoo! Cloud Serving Benchmark. Available:
18. http://research.yahoo.com/Web_Information_Management/YCSB.
19. ParaStor. Available: http://www.sugon.com/product/detail/productid/37.html.
20. Z. X. Luo, "A censoring and quantization scheme for energy-based target localization in wireless sensor networks," *Journal of Engineering and Technology*, vol.2, no.2, Aug. 2012.
21. K. C. Ho, "Bias reduction for an explicit solution of source localization using TDOA," *IEEE Trans. Signal Processing*, vol. 60, pp. 2101-2114, May 2012.
22. Z. X. Luo, "A coding and decoding scheme for energy-based target localization in wireless sensor networks," *International Journal of Soft Computing and Engineering*, vol.2, no. 4, Sept. 2012.
23. Y. Li, K. C. Ho, and M. Popescu, "A microphone array system for automatic fall detection," IEEE Trans. Biomedical Engineering, vol. 59, pp. 1291-1301, May 2012.
24. K. C. Ho and R. Rabipour, "A design and use case for inhibiting the adaptation of echo canceller without using external control," Contribution C816R1, *ITU WP1/SG16 Standard Meeting*, Geneva, Switzerland, May 2012.
25. Z. X. Luo, "Anti-attack and channel aware target localization in wireless sensor networks deployed in hostile environments," *International Journal of Engineering and Advanced Technology*, vol. 1, no. 6, Aug. 2012.
26. Y. Shang, W. Zeng, K. C. Ho, D. Wang, Q. Wang, Y. Wang, T. Zhuang, A. Lobzhanidze, and L. Rui, "NEST: networked smartphones for target localization," in Proc. IEEE CCNC 2012, Las Vegas, Jan. 2012, pp. 732-736.
27. Z. X. Luo, "Robust energy-based target localization in wireless sensor networks in the presence of byzantine attacks," *International Journal of Innovative Technology and exploring Engineering*, vol. 1, no.3, Aug. 2012.
28. L. Yang and K. C. Ho, "Alleviating sensor position error in source localization using calibration emitters at inaccurate locations," IEEE Trans. Signal Processing, vol. 58, pp. 67-83, Jan. 2010.
29. Z. X. Luo, "A new direct search method for distributed estimation in wireless sensor networks," *International Journal of Innovative Technology and Exploring Engineering*, vol. 1, no. 4, Sept. 2012.
30. T. Glenn, J. N. Wilson, and K. C. Ho, "A multimodal matching pursuits dissimilarity measure applied to landmine/clutter discrimination," in Proc. IEEE Int. Geoscience and Remote Sensing Symp. IGARSS, Honolulu, July 2010.
31. Z. X. Luo, "Parameter estimation in wireless sensor networks based on decisions transmitted over Rayleigh fading channels," *International Journal of Soft Computing and Engineering*, vol. 2, no. 6, Jan. 2013.
32. M. Popescu, K. E. Stone, T. C. Havens, J. M. Keller, and K. C. Ho, "Anomaly detection in forward-looking infrared imaging using one-class classifiers," in Proc. SPIE Conf. Detection and Remediation Technologies for Mines and Minelike Targets XV, Orlando, Apr. 2010.
33. Z. X. Luo, "Distributed estimation and detection in wireless sensor networks," *International Journal of Inventive Engineering and Sciences*, vol. 1, no. 3, Feb. 2013.
34. T. C. Havens, C. J. Spain, K. C. Ho, J. M. Keller, Tuan T. Ton, D. C. Wong, and M. Soumekh, "Improved detection and false alarm rejection using FLGPR and color imagery in a forward-looking system," in Proc. SPIE Conf. Detection and Remediation Technologies for Mines and Minelike Targets XV, Orlando, Apr. 2010.
35. George Kousiouris, Tommaso Cucinotta, Theodora Varvarigou, "The Effects of Scheduling, Workload Type and Consolidation Scenarios on Virtual Machine Performance and their Prediction through Optimized Artificial Neural Networks," The Journal of Systems and Software (2011),Volume 84, Issue 8, August 2011, pp. 1270-1291, Elsevier, doi:10.1016/j.jss.2011.04.013.
36. Ko, Ryan K. L. Ko; Kirchberg, Markus; Lee, Bu Sung (2011). "From System-Centric Logging to Data-Centric Logging - Accountability, Trust and Security in Cloud Computing". Proceedings of the 1st Defence, Science and Research Conference 2011 - Symposium on Cyber Terrorism, IEEE Computer Society, 3–4 August 2011, Singapore.
37. Ko, Ryan K. L.; Jagadpramana, Peter; Mowbray, Miranda; Pearson, Siani; Kirchberg, Markus; Liang, Qianhui; Lee, Bu Sung (2011). "TrustCloud: A Framework for Accountability and Trust in Cloud Computing". Proceedings of the 2nd IEEE Cloud Forum for Practitioners (IEEE ICFP 2011), Washington DC, USA, July 7–8, 2011.
38. Z. X. Luo, "Parameter estimation in wireless sensor networks with normally distributed sensor gains," *International Journal of Soft Computing and Engineering*, vol. 2, no. 6, Jan. 2013.
39. Daniel Nurmi, Rich Wolski, Chris Grzegorczyk, Graziano Obertelli, Sunil Soman, Lamia Youseff, Dmitrii Zagorodnov, "The Eucalyptus Open-source Cloud-computing System", Computer Science Department, University of California - Santa Barbara, Santa Barbara, California 93106.
40. Z. X. Luo, "Overview of applications of wireless sensor networks," *International Journal of Innovative Technology and Exploring Engineering*, vol. 1, no. 4, Sept. 2012
41. Mike P. Papazoglou, "Service -Oriented Computing: Concepts, Characteristics and Directions", Tilburg University, INFOLAB.
42. Z. X. Luo, "Distributed estimation in wireless sensor networks with heterogeneous sensors," *International Journal of Innovative Technology and Exploring Engineering*, vol. 1, no.4, Sept. 2012
43. Rajiv Ranjan, Rajkumar Buyya, "Decentralized Overlay for Federation of Enterprise Clouds", Grid Computing and Distributed Systems (GRIDS) Laboratory, Department of Computer Science and Software Engineering, The University of Melbourne, Australia.
44. Z. X. Luo, P. S. Min, and S. J. Liu, "Target localization in wireless sensor networks for industrial control with selected sensors," *International Journal of Distributed Sensor Networks*, 2013.
45. Marianne C. Murphy, Marty McClelland, "Computer Lab to Go: A "Cloud" Computing Implementation", Proc ISECON 2008, v25.
46. Z. X. Luo, "Survey of networking techniques for wireless multimedia sensor networks," *International Journal of Recent Technology and Engineering*, vol. 2, no. 2, May 2013.
47. Z. X. Luo, "Survey of applications of pupil detection techniques in image and video processing," *International Journal of Recent Technology and Engineering*, vol. 2, no. 2, May 2013.
48. Z. X. Luo, "Survey of corner detection techniques in image processing," *International Journal of Recent Technology and Engineering*, vol. 2, no. 2, May 2013.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)