Research and Application of the Electricity Information Collection System Based on Cloud Computing

QIAN Qi[1, a], CHEN Guang[2,b], HAN Yu[3,c], LIU Tao[4,d]

[1]NARI Group Corporation （State Grid Electric Power Research Institute）,Beijing 100192, China;

[2]NARI Group Corporation （State Grid Electric Power Research Institute）,Beijing 100192, China;

[3]NARI Group Corporation （State Grid Electric Power Research Institute）,Beijing 100192, China

[4]NARI Group Corporation （State Grid Electric Power Research Institute）,Beijing 100192, China;

[a]qianqi@sgepri.sgcc.com.cn,[b]chenguang@163.com,[c]hanyu7@sgepri.sgcc.com.cn,[d]liutao7@sgepri.sgcc.com.cn

Key words: Cloud computing; Electricity information acquisition; Distributed computing

Abstract: With the increasing huge amounts of data, there is a high requirement on the electricity information acquisition system. In this paper, a platform for the electricity information collection system is proposed to solve the problems in the current electricity network by employing the cloud computing technology. Then an interactive mode and a cloud computing infrastructure are described in detail and are employed to demonstrate the characteristics of real time and high efficiency in the electricity information collection system.

Introduction

With the fast promotion of grid construction, the scope of data of electricity information collection is larger and larger and the collection frequency is higher and higher, at the same time, the requirement on the capability of real time analysis is more and more strict, the information processing capability of current electricity information collection system will not be capable of completing the collection, storage and analysis function for massive data, cannot meet the unified requirements of State Grid Corporation of China “Cover, collect and control comprehensively” on electricity information collection system.

In order to solve the existing problems in electricity information collection system, this paper puts forward that, to realize the high turnover collection and storage of large scale electricity data, support computing, analysis and treatment of large scale electricity data by taking cloud computing as core technology, and realize the linear improvement of system storage and computing capability through dynamic expansion.

Concept of cloud computing

Cloud computing is the development of distributed computing, parallel computing and network computing is the commercial form of computer scientific technology. It distributes the computer tasks in the resource pool constituted by large number of computers, to make different application systems obtain computing capability, storage space and various software services as required.

Through the application of cloud computing it may construct a new electricity information collection system, which solves the technical bottlenecks in data storage, data computing capability,

data service capability and other aspect caused by big data, platform integration and business expansion.

This paper designed an electricity information collection system structure based on cloud computing, and conducted the research on specific key technologies, information sharing and platform integration etc.

System interactive mode

In order to solve the existing problems of electricity information collection system about massive data and accurately and rapidly analyze the accuracy and timeliness of electricity information collection of massive data, improve the reusability of data synchronism between different data systems and electricity information collection system as well as improve the massive data mining technology, we have developed electricity information collection system cloud computing platform based on Apache Hadoop and other open source modules.

The electricity information collection system cloud computing platform takes SOA as application structure mode, adopting distributed file system for unified data storage, various data collection terminals will transmit the data of structuralization, half structuralization and non-structuralization to distributed file system for unified data management through data collection channels, and construct distributed computing engine taking several servers as basis, so as to solve the performance bottleneck of massive data computing and massive data mining, as shown in Figure 1.

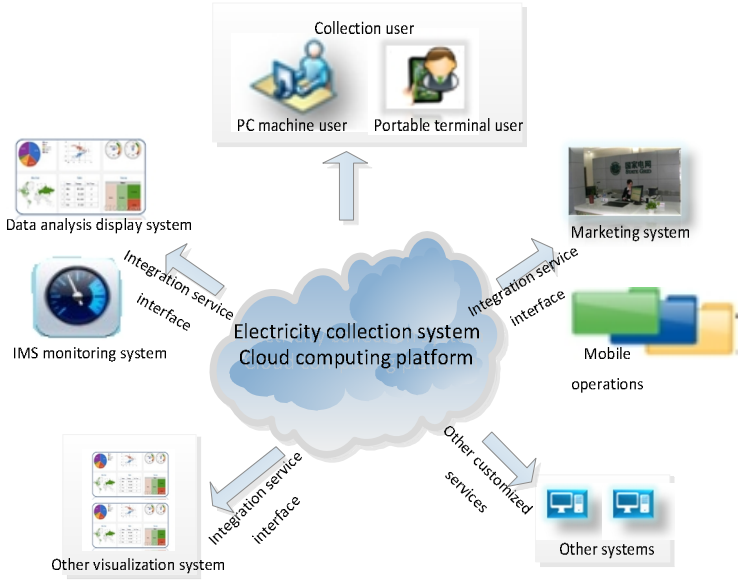

Figure 1 Electricity information collection system cloud computing platform and different systems interactive model

The electricity information collection users visit system application through Ethernet or Internet, for other systems, such as: marketing system, mobile operation, data analysis exhibition system, IMS monitor system and other visual systems, we have developed different cloud services, provide integration service open interface and standard Web Service services for access and integration of those systems; at the same time, in order to provide the system expandability and platform integration, the platform also supports the provision of customized standard services for other service systems, so as to realize the perfect integration of more systems and platforms.

System structure

As shown in Figure 2, the electricity information collection system cloud computing platform is composed of cloud service interactive layer, cloud computing service layer and cloud distributed data service layer.

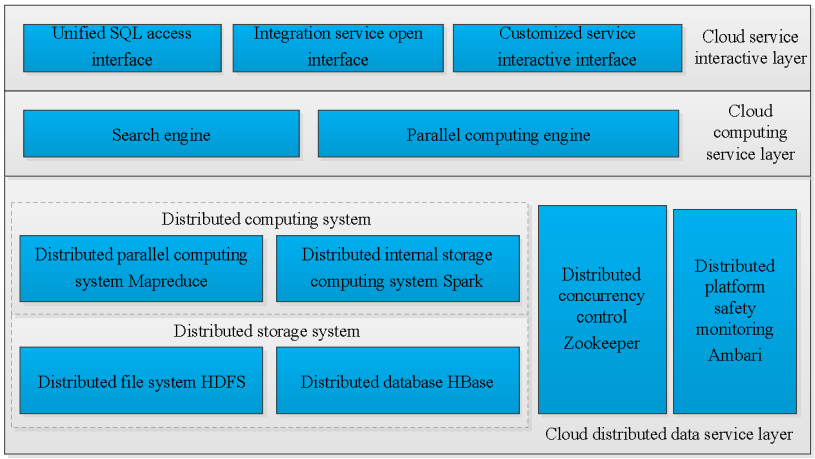

Figure 2 Electricity information collection system cloud computing platform system structure design drawing

The cloud service interactive layer mainly provides the collection users or other systems the interactive access function for the platform, users visit the conversion interface through unified SQL, providing the function of unified data SQL access of relational database and distributed NoSQL database for users; the integration service open interface provides the function to integrate the system with other third party ETL, data visualization, data unified analysis and modeling tools, improving the usability of system, in addition, it also provides the function to dock with other external application systems, so as to provide more data supporting services; customized service interactive interface provides customized services for other systems through sealing the collection data by customized data collection mechanism.

The cloud computing service layer provides supports for cloud service interactive layer, main providing the search engine and parallel computing work engine. Among which, the search engine realizes the document storage, index and search services. The parallel computing work engine conducts various electric data applications and analysis on the basis of distributed computing system MapReduce or Spark, which can provide the electrical power system application high performance parallel computing capability and commonly used parallel algorithm development service interface.

The cloud distributed data service layer includes distributed storage system、distributed concurrent computing system、distributed concurrent control as well as distributed platform safety monitor platform.

The distributed storage system includes distributed file system HDFS and distributed database HBase. The distributed file system adopts distributed redundancy storage method to store data, store and manage enterprise level non-structural data, and store the massive data in relation to voltage, current and power on the platform. The data fragment treatment for data will be conducted after collected, and the data fragment will be stored at least three joints, so as to improve the high availability, high reliability and economy of system. In processing and analyzing the massive data set, HDFS can reach high data turnover at the cost of high delay, but in actual application, such as under the extreme high demand of timeliness in voltage detection, HDFS is not sufficient to process these applications of higher timeliness, while the distributed database HBase is a better choice. HBase is a column-oriented distributed data storage system of high reliability which is expandable,

providing random and real time read/write functions on big data set. In the system design, for analysis of applications of high timeliness, we will abstract the data in HDFS by fragments and store the data about analysis results to HBase.

The distributed computing system includes distributed parallel computing system Map Reduce and distributed memory computing system Spark. The platform adopts MapReduce parallel mode to automatically divide tasks into several sub-tasks, and conduct statistics, computing and analysis for electricity information collection in large scale computing nodes. Spark is a fast and common used distributed computing engine for big data set processing and data analysis, which can conduct parallel analysis and processing by MapReduce, while for real time task, compared to MapReduce, Spark is based on memory computing and the computing speed is higher.

The distributed concurrency control ZooKeeper is the distributed coordinative service for Hadoop, used for configuring and supporting distributed dispatch, realizing the synchronization service, configuration maintenance and naming service etc. The distributed platform safety monitoring Ambari is used for solving the Hadoop ecologic system deployment, including the module deployment and configuration, deployment process tracking and multi-machine deployment problems etc.

Cloud computing based electricity information collection system, construct the distributed computing engines based on Hadoop. Spark taking several servers as foundation, which can solve the bottlenecks in massive data computing, massive data mining. Through utilizing the distributed batch processing function for massive data of MapReduce, realize the fast analysis and processing for mass historic data, adopt Spark iterative memory computing engine, reduce the data transmission and repeat computing, make up the defect of MapReduce in complex computing, improve the distributed processing capability of system in complex computing, greatly enhance the important role of electricity information collection system in intelligent power grid.

System application and realization

Cloud computing based electricity information collection system has been deployed and applied in certain provincial electric company. After the optimization this time, servers for the project is increased from the original 3 to current 20 units, the model of which is DELL M610. The original three servers will take Oracle database as main storage platform, conduct data mining and data extracting through using Oracle database programming, construct twenty servers as Hadoopcloud computing environment after optimization, among which, the main node CPU is 4×8 cores, 64G memory, 4×300GB hard disk, for other data nodes CPU is 2×6 cores, 32G memory, 2×300GB hard disk, the operation environment of Hadoop is 2.6.0.

In order to verify the advantages of cloud computing in electricity information collection system, we selected the power curve function of measuring points as test, the number of data at test is 0.12 billion pieces, the data about power at measuring point are obtained every 15 min, and the number of power data at each measuring point per day after analysis and statistics are 96 pieces. Before optimization, it requires 90 min to analyze the data about power curve at measuring point, after the system optimization, it only requires 5 min by cloud computing platform. The comparative results of test are as shown in Figure 3.

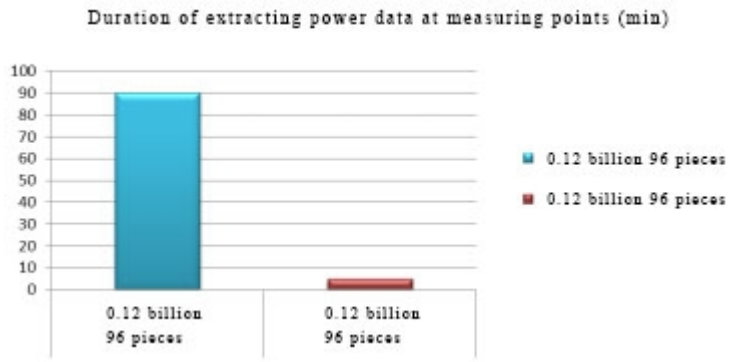

Figure 3 Comparison chart for duration of extracting power data at measuring points

Above all, we can learn that, it saves 94%wwhen adopting cloud computing for analysis, which verifies the advantages of cloud computing platform in electricity information collection system.

Conclusions

This paper realized the distributed storage and concurrence computing of electricity information collection system and improved the computing speed of system data analysis as well as the expansion capability and storage capability of system data storage through establishing a cloud computing based electricity information collection system structure.

References

[1] WANG Zhongwen Cloud Computing Based electricity Data Collection and Management Terminal System IT China Science and Technology Information 2013:16.

[2] HONG Jianguang Research and Application of Key Technology for Performance Improvement of Cloud Computing Based Electricity Information Collection System ELECTRIC POWER ICT 2014, Vol.12 Issue 3.

[3] FAN Bangkui Research on Cloud Computing Structure for Application in Smart Power Grids ELECTRIC POWER ICT 2014, Vol. 12, Issue 01

[4] SONG Zhenwei Application of Cloud Real Time Database in Electricity Information Collection System China Electric Power Education 2014, Issue 09

[5] CAO Yang GAO Zhiyuan Application of Cloud Computing Mode in Power Dispatching System China Power 2012, Issue 06

[6] HE Ming, ZHENG Xiang Disccusion on Development and Application of Cloud Computing Technology Telecommunications Science 201502:56-57

[7] Yun Yang, Lie Wu,etc. The structure of intelligent grid based on cloud computing and risk analysis [J]. Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2012:123-126