

Design of Cloud Computing Architecture for DIOT

Liang Chen^{1, a}, Jing Qiu^{2, b}, Xueping Gu^{1, c}

¹ School of Electrical and Electronic Engineering, North China Electric Power University, Baoding, 071003, China

² School of Computer Science, North China Electric Power University, Baoding, 071003, China

^aemail: raymond05@yahoo.cn, ^bemail: 750923115@qq.com, ^cemail: gxp_ncepu@126.com

Keywords: Cloud Computing; Domestic Internet of Things; Data Storage; Data Analysis and Mining

Abstract. Considering the problems of the current domestic internet of things intelligent terminals—single processing mode, non-dynamic generated mode and low level humanized service mode, a cloud computing architecture for the domestic internet of things is proposed in this paper. The message management mechanism of the proposed architecture is based on MINA framework and JMS. And its data storage module is based on HBase, achieving data segmentation. Besides, MapReduce technology is applied to realize data analysis and data mining. Accordingly, a complete cloud computing framework for the domestic internet of things is established. At last, the performances of data storage, data analysis and data mining of proposed architecture are tested and the results prove its validity and efficiency.

Introduction

In the 21st century, with the development of smart grid, the technology of the Internet of things(IOT) will further strengthen the smart grid, such as intelligent interaction between power grid and users[1-3]. At present, most research of the domestic internet of things(DIOT) focus on sensing layer at the fore-end, however, the value promotion and efficiency of IOT sensory information depends on the integration processing and system control of sensory information at the back-end, etc., the cloud computing services background of IOT long chain[4]. The sensory ability and control ability of current DIOT for the home environment isn't so excellent, operating in a fixed mode without dynamically generated humanized service according to different users. So a back-end processing layer is demanded to comprehensively analyze and mine the environment data collected form DIOT. And this requires an efficient communication management strategy, a storage platform, which can store vast amounts of data, and a powerful cloud computing platform.

Traditional data analysis is based on the relational data model, its parallel operation is realized through expensive hardware, such as massively parallel processing (MPP) system and/or symmetric multiprocessing (SMP) system[5], which has a poor scalability and high cost and can't adapt to the analysis of vast amounts of unstructured data. Based on Hadoop, MapReduce parallel analysis supports column oriented data model. It's flexible and extensible, and can be realized by the Mahout algorithm framework[6].

A large number of DIOT intelligent terminals and cloud computing background always keep in a long connection state, and the security, reliability and real-time performance should be assured. So classification method is applied in the solution of messages management. The messages of IOT environment is divided into two parts: asynchronous messages and synchronization messages. They are respectively realized by JMS and MINA framework.

Design of Cloud Computing Architecture for DIOT

Considering the requirements of DIOT, the cloud computing architecture is proposed and shown in Fig. 1, which mainly includes system maintenance module, security management module,

function layer, business logic layer, data processing layer, data storage layer, data access layer.

(1)System maintenance module

The system maintenance module is mainly responsible for system audit log 、exception handling and data backup.

(2)Security management module

The security management module mainly includes data encryption and access control, ensuring data security of the cloud system.

(3)Function layer

The function layer is the major interface to provide services, including knowledge download module, knowledge update module and so on.

(4)Business logic layer

Business logic layer includes task scheduling control module, domain knowledge base and workflow control module. According to the different processing request, task scheduling control module calls different business processes or method modules; workflow control module which monitors and manages the state of the business can return specific information parameters to the task scheduling control module.

(5)Data processing layer

Data processing layer consists of parallel data processing module, user mode evaluation module, data access module. Parallel data processing module for data analysis and data mining provides parallel algorithms, mainly includes the MapReduce parallel computing algorithm library; user mode evaluation module—the results in line with user requirements are stored in the domain knowledge base and can play a supporting role in the field of data-processing procedures; data access module according to different needs can read or write data from data storage layer.

(6)Data storage layer

Data storage layer includes core database, business auxiliary database, file system and so on. HBase is applied in core database, and it's function is storing system metadata; Oracle database is applied in business auxiliary database to assist business fast processing; HDFS distributed file system is applied in file system.

(7)Data access layer

Data access layer mainly includes the synchronous message module and asynchronous message module, the asynchronous message part is based on JMS to solve load problems caused by the frequent communication between intelligent terminals and the cloud. And the synchronization message processing part is based on MINA framework to ensure the real-time processing of the message.

The workflow of the cloud computing architecture is shown in Fig. 2. The requested action of various data acquisition equipment and users are collectively referred to collectors. Different approaches are provided for different types of messages. The specific communication processes are as follows:

(1)Transmission of synchronous messages:

Step 1: Messages will be sent to synchronization message server by the intelligent terminal;

Step 2: User control messages and environmental exception messages will be distinguished;

Step 3: Messages will be written into the cloud by synchronous message server;

Step 4: Based on the message level, data analysis methods will be called to match the message, and finally, a smart decision message will be made.

Step 5: Decision messages will be sent back to the corresponding intelligent terminal by synchronous message server;

(2)Transmission of Asynchronous messages:

Step 1: Messages will be sent to the asynchronous message server by the intelligent terminal

Step 2: After pretreatment, Messages will be written into the cloud by synchronous message server.

The cloud server regularly does some data analysis and data mining to update the knowledge, in order to ensure the accuracy of message decision-making.

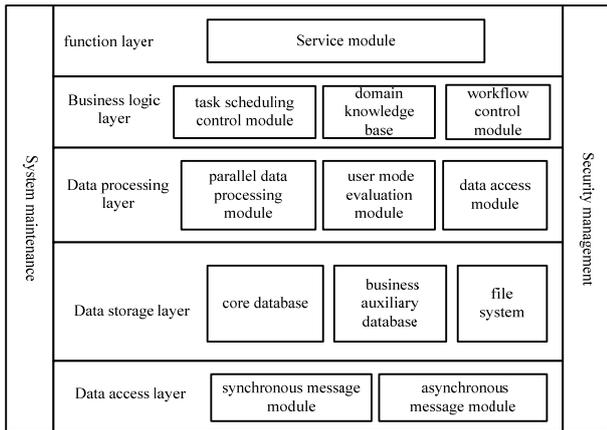


Fig.1. The core schema of cloud computing architecture for DIOT.

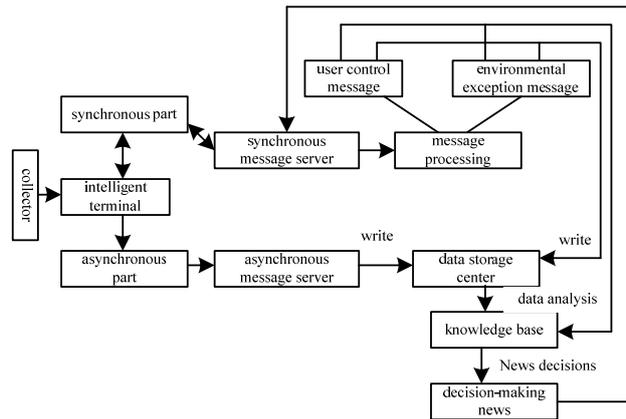


Fig.2. The workflow of cloud computing architecture for DIOT.

Technical Explanation

The Hadoop Distributed File System(HDFS) is applied to realize the system data storage, HDFS is highly fault-tolerant and is designed to be deployed on low-cost hardware. HDFS provides high throughput access to application data and is suitable for applications that have large data sets. Besides, all data storage modules are based on The Hadoop Database (HBase), for example, the HomeTable, which is shown in Table1.

Each HomeTable is a sparse distributed multidimensional orderly figure indexing by the line key, the row key and a timestamp. Each row of the table are dynamic partitions, each partition is called Tablet which is the smallest unit to adjust the data distribution and load balance .In order to improve the efficiency of access and read, the column group of HomeTable is made up of multiple related columns. The efficiency and scalability benefit from using the Tablet and column groups.

Tab.1. HomeTable

Tablet	Rowkey	Dev_Kind	Dev_Id		State		Command		
			RFID	UPnP IP	Dev	Power	Search	Control	Link
Tablet1	510501	air- conditioning	*****	*****	on	low	none	start	link to 510504
Tablet2	510502	water heater	*****	*****	on	low	exist	none	none
	510503	washer	*****	*****	on	low	none	start	none
Tablet3	510504	door and window	*****	*****	close	low	none	none	link to 510501

In order to ensure the data security of DIOT and performance requirements of the system, data encryption and signature technology are applied to encrypt sensitive data storage to protect data integrity and confidentiality, including AES, SHA-256 algorithm.

The core goal of cloud computing architecture for DIOT is to provide DIOT smart terminal with well performed and highly reliable background data analyses and processing methods. Thus, to help the smart terminal make messages decisions, machine learning algorithms as well as conventional data analysis algorithms are needed in data processing part of DIOT.

The framework of Mahout algorithm includes classification algorithm, clustering algorithms, collaborative filtering recommendation algorithm, dimension-reduction algorithm and ect. Because the algorithm itself is based on Hadoop, many have been realized in MapReduce mode, greatly improving the performance of the algorithm in data processing.

The K-means clustering algorithm in Mahout for data analysis processing. Although the K-means is simple and efficient, there are certain problems. Firstly, K value is artificial and it is difficult to give a reasonable value of K in the case of data unknown. And if the isolated point was selected, it would heavily affect the impact of clustering. Usually Canopy algorithm is coordinated, to determine the number of clusters and the initial cluster centers.

Test results

Test environment: The Hadoop cluster test platform contains 15 machines, and they are of the same configuration: 2G memory, 160G hard drive space, Ubuntu 10.4 operating system, Hadoop 0.20.2, HBase 0.90.5.

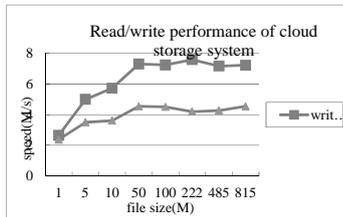


Fig.3. Read/write performance of cloud storage system.

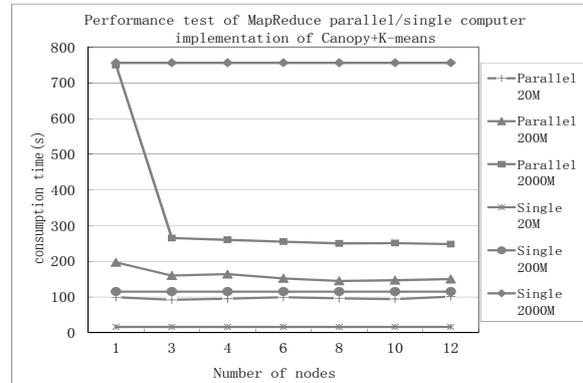


Fig.4. The performance of MapReduce parallel/single computer implementation of Canopy+K-means.

The test results show that: when dealing with files less than 50M, the read/write speed of this system varies greatly, and for files more than 50M, the speed can be maintained in a stable value: the write speed is 7.30 M/s, and the read speed is 4.42 M/s. When dealing with small data sets, single computer implementation is superior to parallel implementation; when dealing with large data sets, MapReduce cluster parallel computing has an obvious performance advantages. It also can satisfy the requirements of DIOT on the efficiency of data analysis and data mining.

Conclusion

Considering the problems of weak processing ability, low level intelligent and humanized service of DIOT, a complete cloud computing framework for DIOT is established this paper. The message management mechanism is designed based on MINA framework and JMS. And data storage module of DIOT is designed based on HBase, achieving data segmentation. Besides, MapReduce technology is applied to realize data analysis and data mining. The performance test results of data storage, data analysis and mining verify the validity of proposed cloud computing architecture. The proposed cloud computing architecture provides a feasible and innovative technical scheme for the efficient and intelligent data processing of DIOT.

References

- [1] ANGI H FARH. The path of the smart grid [J]. IEEE Power and Energy, 2009 (8) 18-28.
- [2] Hongbo Zhu, Longxiang Yang. Investigation of technical thought and application strategy for the internet of things[J]. Journal on Communications, 2010:31(11) :2-4.
- [3] Weihong Wang. Evolution of the Internet of Things and the Related Value Chain[J]. Telecom Engineering Technics and Standardization, 2009(12) 31-36.
- [4] Yu Ming. Disk array storage technology analysis[J]. Advanced Television Engineering, 2006(7) 45-50.
- [5] Jiangong Duan. Differentia and Unification of NAS and SAN[J]. Application Research Of Computers, 2009(12) 17-22.
- [6] Jie Cui, Taoshen Li. Design and Development of the Mass Data Storage Platform Based on Hadoop[J]. Journal of Computer Research and Development, 2012(1)33-37.