

E-Commerce Recommendation System based on MapReduce

Wei Zhao¹, Hong Tao Zhang^{1*}

¹ ZhengZhou university software college, ZhengZhou,450002, China

Received 10 October 2014, www.cmnt.lv

Abstract

According to the Present Situations were that there is an urgent demand for large data analysis in Electronic Commerce, by using cloud computing's advantage in storing and analyzing mass data, the solution of new project which can analysis data was proposed, based on the cloud computing. Firstly, aiming to the advantages of the cloud computing platform, the novel data- analysis architecture was designed, then the business flow chart by using cloud computing analysis on the architecture. Finally the project is validated by practical application and possesses certain reference meaning in E-Commerce Recommendation System based on cloud computing.

Keywords big data; Cloud Computing; architecture; parallel computing; Cloud safety

1 Introduction

The trend is large of data increases in exponential in the fields at present, for example, Internet, E-C. The fields store massive data called big data. Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools, the challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization. Every search in internet, every transaction in website and every input is data, by using computer in selecting and cleaning and analyzing the data, the results which are not only simply and subject conclusions but also guide the more consumption power and which s shown. It is important for E-C website to propose a project which is simple and high expansibility.

At present, some mature software technique, for example, Data Mining and Association Rules and so on which have made a progress in big data analysis of the e-commerce site, but also have showed the poor scalability and the data sharing and have congenital defect in the hardware of the higher requirement [1-3].

Nowadays cloud computing technique is becoming mature as the advantage in storing and analyzing massive data, the technical features of cloud computing is that it can store mass data, read them and abundance analysis, furthermore, the reading operation frequency of the data is much larger than the updating operation frequency of the data, so data management for cloud computing is the reading optimization data management[4]. The new computing and service model in Internet are brought by cloud computing, which is becoming data center or supercomputers by means of distributed computing and virtualization, which is free mode or lease to provide storing data and analyzing data for the developer or enterprise customer[5].

2 MapReduce Programming Model

As a data parallel model, MapReduce is a patented software framework introduced by Google to support distributed computing on large datasets on clusters of computers. Known as a simple parallel processing mode, Mapreduce has many advantages: such as, it is easy to do parallel computation, to distribute data to the processors and to load balance between them, and provides an interface that is independent of the backend technology[6]. MapReduce is designed to describe the process of parallel as Map and Reduce. The user of the MapReduce library expresses the computation as two functions: Map and Reduce[7].

Map, written by the user, takes an input pair and produces a set of intermediate key/value pairs. The MapReduce library groups together all intermediate values associated with the same intermediate key I and passes them to the Reduce function.

The Reduce function, also written by the user, accepts an intermediate key I and a set of values for that key. It merges together these values to form a possibly smaller set of values. Typically just zero or one output value is produced per Reduce invocation. The intermediate values are supplied to the user's reduce function via iterator. This allows us to handle lists of values that are too large to fit in memory.

The Map and Reduce functions are both defined with respect to data structured in (key, value) pairs. MapReduce provides an abstraction that involves the programmer defining a "mapper" and a "reducer", with the following signatures[8]:

Map::(key1) list(key2,value2)

Reduce::(key2,list(value2) list(value2)

* *Corresponding author's* e-mail: iezhaowei@163.com

Hadoop is a free and open-source and distributed programming framework which implement MapReduce ,in fact, use java to implement the algorithm model of the Google MapReduce. The developer write codes by means of Hadoop platform which run on MapReduce cluster and implement efficient parallel processing for massive data. Hadoop are made of the algorithm execution of MapReduce and Distributed File System[9].

The E-Commerce Recommendation System is an intermediary program (or an agent) with a user interface that automatically and intelligently generates a list of information which suits an individual’s needs. So, According to the theory of cloud computing Google's proposed programming model for parallel computing as data analysis and processing, the solution of the E-Commerce Recommendation System was proposed based on the cloud computing, firstly the Source data to analyze which is collected by means of the cloud computing platform, then by the association rule analysis of the Source data, at last recommend product to the customer want to buy. The solution is designed based on the cloud computing platform and association rules mining technique to improves system scalability and the performances of data mining. The solution is validated by practical application and greatly enhances Data Security and improves distributed storage.

3 The data-analysis architecture figure of the e-commerce recommendation system based on cloud computing

The whole architecture of the e-commerce recommendation system based on cloud computing is divided into client, SaaS, PaaS and IaaS.

Client It provides terminal various forms,for example, PC, Brower and Mobile

Terminal. The commodity purchase program was out-putted by means of the client.

SaaS It composed of data collecting, access control, data sieving and commodity purchase program. It provides the development interface of system customization and interface of the software application services, then collects the massive source data which data mining.

PaaS It composed of identity management, access control, data processing, association rules and cloud storage services. It converts data which SaaS collected into normative data which has sieved. It is the key of the e-commerce recommendation system based on cloud computing.

IaaS It provides virtualization platform based on cloud computing and distributed group environment, provides run-time infrastructure for Pass. It composed of basic services and infrastructure. Infrastructure composed of host computer, network, relational database and distributed storage. basic services composed of parallel processing, multi-user, distributed cache and virtualization.

The E-Commerce Recommendation System work by collecting data from account, commodities, and transactions, examples of implicit data collection include the following: observing the items that a user views in an on-line store, keeping a record of the items that a user purchases online, obtaining a list of items that a user has listened to or watched on his/her computer, et al. The data are stored by disk array,,and mined in MapReduce that is a programming model and an associated implementation for processing and generating large data sets.

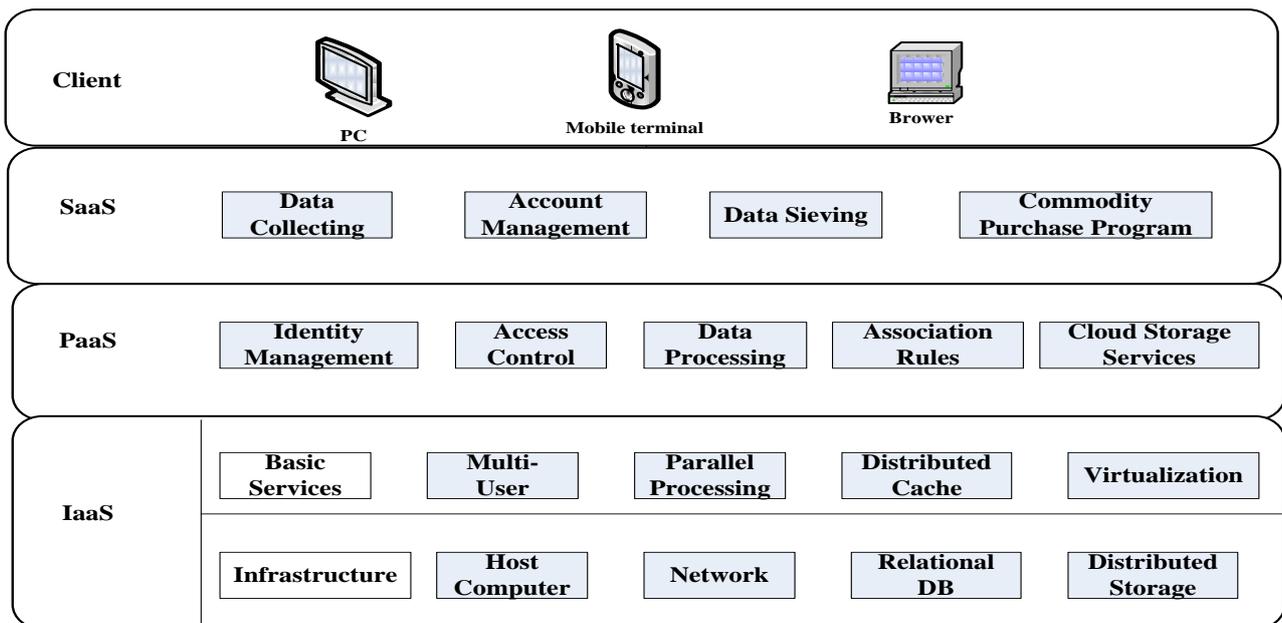


FIGURE 1 the data- analysis architecture figure of the e-commerce recommendation system based on cloud computing

4 Key Technology Research and Implementation

4.1 THE MAIN E-R DIAGRAM AND ANALYSIS

The new system puts forwards a new access control mode based on user, user group, role, screen and group-screen to

make access control more flexible. The order is designed by general master-slave table, the po_head and th are bound together by po_no, the po_head and the basic_user are bound together by user_id, then e po_detail and product are bound together by product_id.

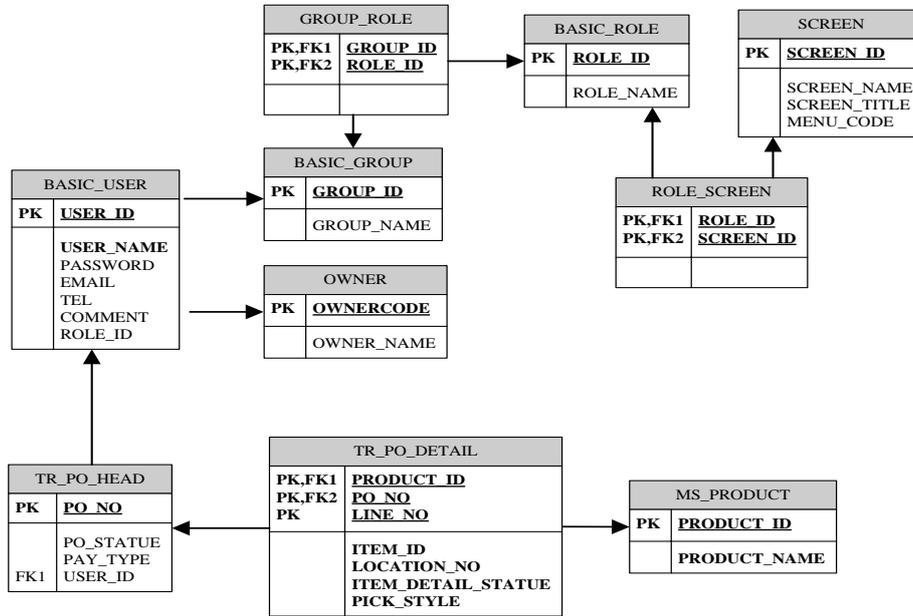


FIGURE 2 The main E-R diagram of the e-commerce recommendation system based on cloud computing

4.2 THE BUSINESS FLOW CHART OF THE E-COMMERCE RECOMMENDATION SYSTEM BASED ON CLOUD COMPUTING

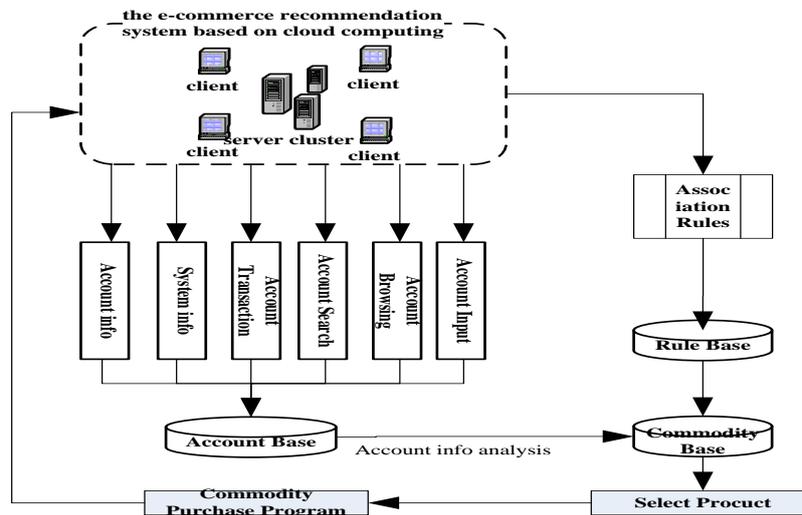


FIGURE 3 The business flow chart of the e-commerce recommendation system based on cloud computing

The business flow chart of the e-commerce recommendation system based on cloud computing as Figure3. Firstly, the source data and the information about account landing were collected by the e-commerce recommendation system based on cloud computing, and were formatted and saved the account base, then were analyzed by all of the host computers based on cloud computing and server

clusters, including the association rule mining. Then commodity information were general analyzed and becoming commodity base which provided commodity recommendation service for customers. At last, according to the requirement of the customers, commodity was selected and the solution was recommended for customers.

4. 3 THE SEQUENCE DIAGRAM OF THE E-COMMERCE RECOMMENDATION SYSTEMS BASED ON CLOUD COMPUTING

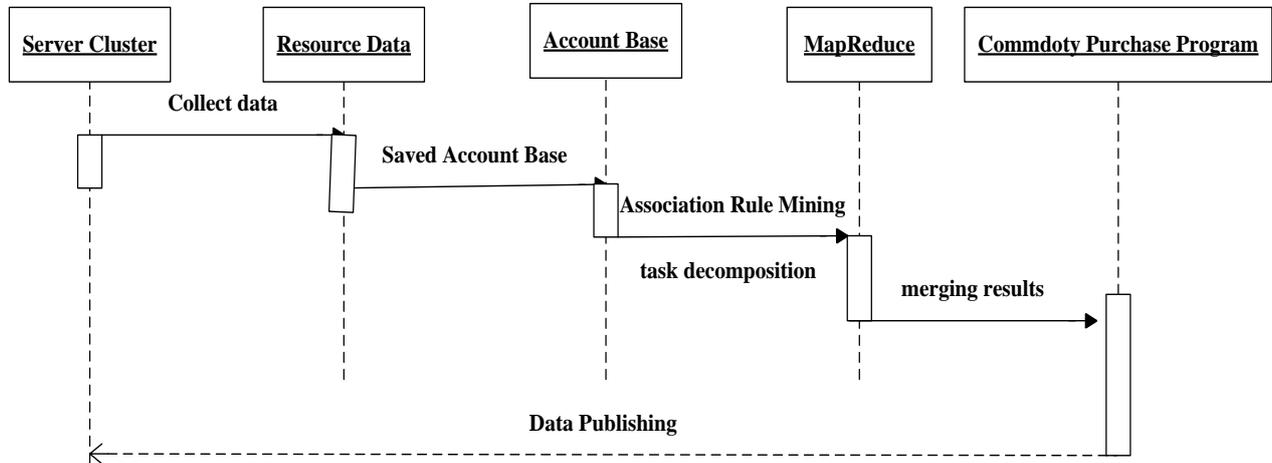


FIGURE 4 the Sequence Diagram of the e-commerce recommendation systems based on Cloud computing

1. The source data and the information about account landing were collected by server cluster based on cloud computing.
2. The data was saved account base on cloud model.
3. The data for Mining Association Rules according to cooperate with the host computer and server cluster, task was decomposed and intermediate result was complicated using MapReduce model.
4. At last, the merging results were introduced into the destination database.
5. The destination database published data to the server cluster. where superscripts '+' and '-' label the asymptotic behavior in terms of d-dimensional waves, d is the atomic structure dimension.

4. 4 THE SECURITY OF THE E-COMMERCE RECOMMENDATION SYSTEMS BASED ON CLOUD COMPUTING

In this system accounted information, purchased product information and commodity purchase program of the customers were saved and processed related to cloud computing. If the key data was lost or stolen, the unfavorable influence will be caused. In order to ensure the security of the account information, system safety and security were designed as follows: In cases of large and complicated formulas you can use 1-column text formatting as in cases of large tables and figures.

1. Infrastructure layer provide network security for the e-commerce recommendation system through the firewall, anti-virus systems, security gateways, VPN, routers, switches and other network security devices.
2. The e-commerce recommendation system based on cloud computing can tolerate node failure problem and

- implement its high availability and will not happen downtime through redundancy between multiple inexpensive servers. Therefore, cloud computing platform itself is safe.
3. The system is based on user authentication and authorization access control methods, in distributed environments users are configured one or more roles, users and administrator are real-time identity monitored, authority certified and certificate checked for preventing unauthorized access inter-user[10].
4. In order to ensure the security of cloud computing's the account data, the data of transmission is encrypted in the system, the account data of collected upload to the environment of cloud computing after using the user key encryption, then real-time decryption when using, to prevent the decryption of data stored on physical media, the data of encryption use the current popular and mature data encryption algorithm, such as symmetric encryption, public key encryption and so on[11].
5. Snapshot, backup and tolerance protect the important data in cloud computing platform, This is the kind of the current creative ways to storage data, That is the data stored at the same time in multiple locations, if the one copy of the data is broken, users can read from another location, the entire process is transparent to the users[11].

4. 5 THE SYSTEM DATA STORED

The system data is extracted through each workstation of cloud computing environment need to convert the source data, and then stored in the Microsoft cloud computing platform. Data management and data maintenance by means of the Microsoft cloud computing platform. The

hardware of the system data storage is disk array, the software platform is SQL Azure of the Microsoft, which provides service set of WEB and allows users to create, query and SQL SERVER database using the network in the cloud. The location of SQL SERVER is transparent for users in cloud. SQL Azure supports a large number of data types, including all of data types of SQL Server 2008. SQL Azure is still relational database, which supports TSQL to create, operate and manage cloud database and supports stored procedure. The stored procedure and data type of SQL Azure have a startling likeness to traditional SQL Server. So the developer can easy to develop their application system in local, and deploy on cloud platform. To make rational use of the third party software platform as data storage center is not only reducing costs which customers have brought and managed database server and resource commitment but also accelerating the speed of system development and make the system more reliability and robustness.

4. 6 CLOUD APPLICATION DEVELOPMENT AND PUBLISHING

The development of the new system has used Microsoft's popular VS2010 and cloud application development kit-Windows Azure SDK which it is provided. The cloud application is developed, debugged, deployed and managed by Windows Azure SDK, then cloud application program is efficient developed by means of ASP. NET Components. The system development language is ASP. NET, which can support object oriented programming and has powerful and complete class library support and good program structure [12]. In the VS2010, cloud application development is divided into creating cloud service, configuration cloud service, generation cloud services, running and debugging cloud service and publishing cloud service. In the five steps, the above four steps is the procedure of cloud application development, the last step is the procedure of cloud application deployment. Supply to see the procedure of the application development, there is not much difference between the development of cloud applications and the development of ASP. Net, it is easy to develop for ASP. Net developers which is accustomed to develop ASP. Net.

References

- [1] Zhao Wei,Zhou Bing. Design and implementation of message-oriented middleware based on ERP System [J]. Computer Engineering, 2008, 34(10): 98-100.
- [2] Li Xiao-dong,Yang Yang,Guo Wen-cai. Data share-and-exchange platform based on ESB [J]. Computer Engineering, 2006, 32(21): 217-219.
- [3] Ea Li Chang-he,Zhao Jie,Zhang Ya-ling,et al. Research and implementation of secure data exchange between heterogeneous databases [J]. Computer Engineering 2007, 33(2):88-93.
- [4] Chen Quan,Deng Qian-ni. Cloud computing and its key techniques [J]. Journal of Computer Applications,2009,29(9):2562-2567.

5 Experimental Result Analyses

According to the presented ideas, we developed the e-commerce commendation testing system based on cloud computing. Test Environment of system is composed of ten host computers, servers, routers, security gateway, switch and so on. Every computer' s CPU is Intel Celeron D 35, memory 2G and WindowsXP. The testing system uses VS2010 and, SQL Server2010.

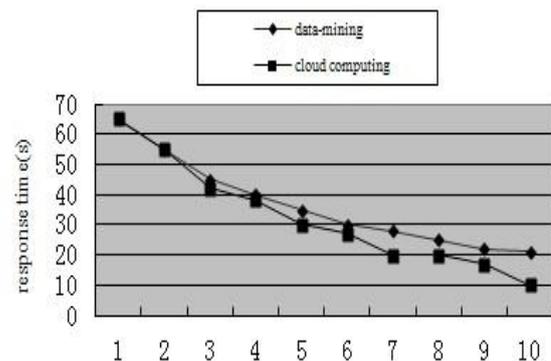


FIGURE5 The performance test of the e-c recommendation system based on Cloud computing

Figure 5 is performance test data of the e-commerce recommendation system base on cloud computing, x-axis is client number, y-axis is the response time of the system, the experimental results show that the performance of the e-commerce recommendation system base on cloud computing has significant improvement gains over Traditional data mining.

6 Conclusion

Cloud computing technology is now applied to more and more aspects, parallel computing of the e-commerce recommendation system is researched based on clouding computing environment in this papers, the programme combines the technical advantages of the platform of cloud computing and association rules in data mining and analysis, is designed and tested in the papers, the results show the programme is feasible, possess certain reference meaning in e-commerce recommendation system based on cloud computing.

- [5] Rittinghouse J,Ransome J. Cloud computing: implementation,management,and security [M] . Boca Raton, FL: CRCPress,2010.
- [6] J. H. C. Yeung, C. C. Tsang, K. H. Tsoi, B. Kwan, C. Cheung, A. P. C. Chan,P. H. W. Leong. Map-reduce as a Programming Model for Custom Computing Machines. Proceedings of the 16th IEEE Symposium on Field-Programmable Custom Computing Machines, FCCM'08.
- [7] J. Dean and S. Ghemawat. Mapreduce: Simplified data processing on large clusters, ACM Commun, vol. 51, Jan. 2008, pp. 107-113.
- [8] T. Elsayed, J. Lin, and Douglas W. Oard. Pairwise Document Similarity in Large Collections with MapReduce. Proceedings – 32nd

Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009.
 [9] Robert L G,Gu Y H,Michael Sabala,etc. Computer and storageclouds using wide area high performance network[J]. Future Generation Computer Systems,2009,25(2):179-183.
 [10]Ning S,Federica P,Elisa B. A privacy-preserving approach to

policybased content dissemination[R]. . CERIAS Tech Report,2009
 [11]Zhang Shui-Ping, LI Ji-zhen,Research and implementation of data center security system based on cloud computing[J]. Computer Engineering and Design.
 [12]Shao Liang-shang,Liu hao-zeng,The practicing course of ASP. NET3. 5(C#)[M]. Beijing:Tsinghua University Press

Authors	
	<p>Wei Zhao, born in 1976, Shangdan County, Gansu province, China</p> <p>Current position, grades: lecturer in Zhengzhou University University studies: postgraduate degree in computer science and application from Zhengzhou University in 2007</p> <p>Scientific interest: distributed data processing, web design and development, Data mining technology, Mass data processing technology.</p>
	<p>Hong Tao Zhang, born in 1977 Xian city, Shanxi province, China</p> <p>Current position, grades: lecturer in Zhengzhou University University studies: postgraduate degree in computer science and application from Xi'an University of Technology University in 2007</p> <p>Scientific interest: distributed data processing, web design and development.</p>