risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 07/06/2016
Aceitação/Acceptance: 14/09/2016

Software Design of Online Learning System based on Hadoop Distributed System

Jiarong Zhao

* **zhaojiarong2016@163.com**

School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China

Pages: 132–141

Abstract: Cloud computing has changed the modern education model. Any educational institutions can share teaching experience through the global online learning resources based on unlimited network resources. By collecting the data generated by the learners in the process of network learning, we construct a learning analysis system based on the distributed Hadoop based on cloud computing technology. Application of cloud computing and big data analysis of online education platform, companies can improve service content, improve the quality of personalized service, while achieving tracking services, precision services, knowledge related services.

Keywords: Software design, Hadoop, Distributed system, Learning system

1. Introduction

With the increasing development of the Internet, people began to "through the information technology and Internet technology for content dissemination and rapid learning", that is, to accept online education (Bhushan, 2014). Moore proposes that online interactive activities through various forms can inspire online learners, effectively reduce the distance between learners and the sense of loneliness (Dong, 2015). Hiltz believes that all kinds of interactive activities are the key factors for students to learn online. Guna believe that learners through interactive media to achieve the level of interaction, a direct impact on the success of online learning. Online instructional interaction hierarchy tower theory model, the data of online education interaction began from the lower to the advanced analysis process. By using the social network analysis method, the interaction model and the structural data of the online teachers and students are analyzed and studied (Essinger,2013). With the help of computer and network technology, the development of online education analysis system, the systematic evaluation of students and teaching resources. Using cloud computing technology to build an online learning support service system, we can see that building a high level of interactive activity system is an important way to promote the development of online education (Fang,2013; Jiménez, 2015). At present the construction of online education interaction analysis system did not involve the analysis of unstructured data, which is currently a key factor restricting the development of online education, so the use of big data technology to realize acquisition, identification, analysis of the data mining and

information value, find hidden aspirations and predict the trend of learning services. In order to guide the improvement of online education services, to achieve the two-way balance between online education services and learners' needs (Huang, 2014; Gonçalves et al., 2015). This paper is based on the demand analysis of the construction of online education, online education interaction model based on big data and cloud computing technology, focuses on the interactive platform, the use of big data technology to dig the interactive data, solve the drawbacks of the current online education interaction, online interaction to achieve the perfect.

Cloud computing has a new generation of online education system of the foundation, it has changed the modern education mode, infinite cyber source based on any educational institutions of learning resources sharing teaching experiences through the global online. Free network open class, the new business model, which attracts the world's attention, many learners are eager to get the world's elite free courses, many learners register or participate in MOOC learning. Through the acquisition of learning data generated in the process of MOOC network computing technology in a Hadoop open source framework to compile and run the application of distributed large-scale data processing to construct a learning analysis system based on cloud, using the system to analyze the status of online learning, help MOOC service providers to improve class rate.

2. Cloud computing and big data

2.1. Big data

The definition of big data is the information and content of unstructured large, from "infinite" activities on the Internet, the general nontraditional sources, such as the web log, click stream, social media, email, sensor, image and video (Ju, 2014). To analyze and make use of the real-time information of large data can provide a great opportunity for related products, and even the political decision-making services. Some online education can benefit from big data, the sentiment analysis, activity analysis, fraud detection can be carried out through data mining, analysis and study of big data binding plays an important role in online education in the future.

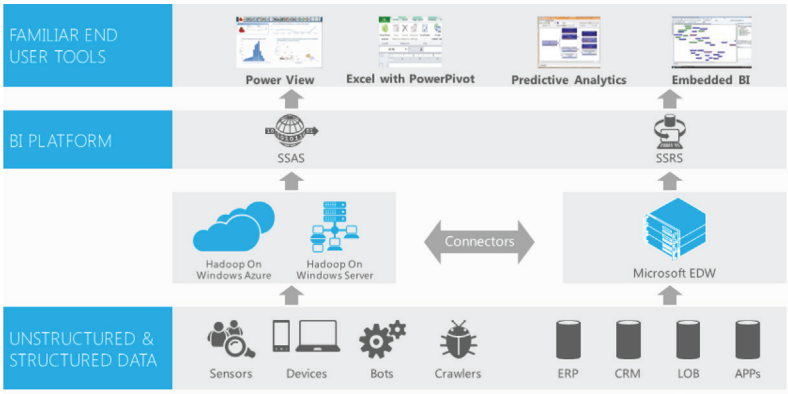

Figure 1 – Hadoop distributed processing

Hadoop can perform distributed processing of massive data; it is the core of HDFS and Reduce Map. A number of data nodes and a node name composed of HDFS, mapping of the relevant documents by the client access management, file or directory management, data block and corresponding data nodes are responsible for the management of the name node, each node as a set of a data node, which is responsible for storage management on the node (Qian, 2014). Map Reduce the task process is divided into Map stage and Reduce stage, Map will change the user input data to key/value pairs form through the user-defined mapping process into a set of intermediate key value pairs set. Reduce is on the middle of the intermediate temporary keys as input for processing, and the results of the final output.

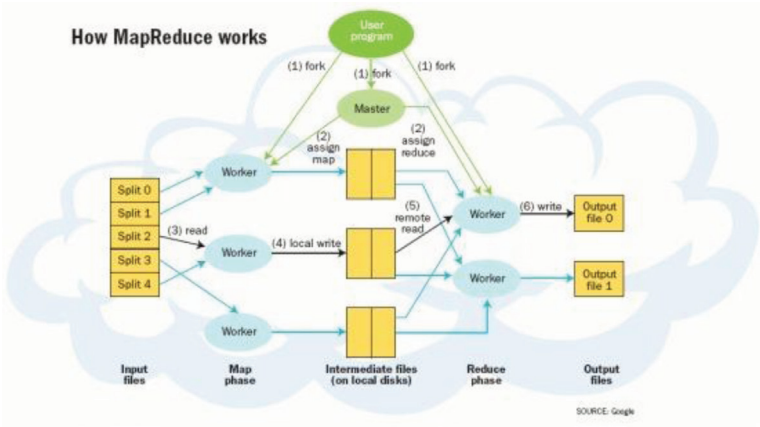

Figure 2 – Map Reduce

2.2. Cloud computing

Cloud computing is the development of distributed processing, parallel processing and grid computing, or the business concepts in computer science. The basic principle of cloud computing is distributed in a large number of distributed computers through the calculation, rather than a local computer or a remote server, enterprise data center operation and the Internet will be more similar. This enterprise can switch resources to need the application, according to demand access to the computer and storage system. This is a revolutionary move, for example, this is from a single generator mode to the old power plant centralized power supply mode (Qiong, 2013). It means that computing power can be used as a commodity circulation, like gas and electricity, convenient use, low cost. The biggest difference is that it is transmitted through the Internet. In the domestic and foreign well-known IT enterprises are to develop and promote cloud computing, such as Google launched Google Apps services, to postpone the Amazon Elastic Compute Cloud (EC2) services, IBM launched the “blue cloud” plan, China launched the Cloud Computing Oriented “sea of clouds” operating system, preliminary cloud computing products have been put into formal the use of cloud computing, but is still in the stage of development, the related technology is not mature enough.

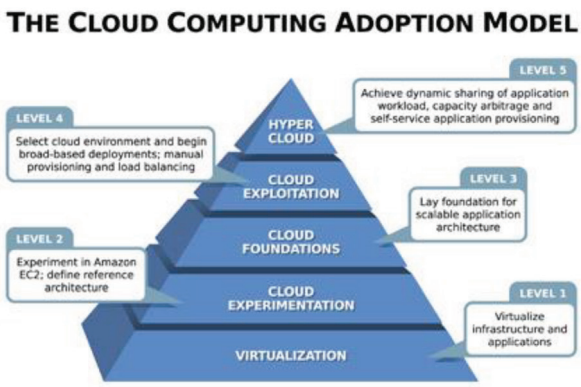

Figure 3 – Cloud computing

Cloud computing needs to process and analyze the distribution and massive data. Therefore, the data management technology must be able to manage a large number of Data. Cloud computing data are massive, heterogeneous, non deterministic characteristics, need to adopt effective data management technology to analyze and process the data and information, build a distributed data storage system is highly available and scalable, the cloud computing system in data management technology is mainly Google GFS, BigTable, MapReduce data management technology and the Amazon Dynam. Cloud computing features in data is mainly manifested in the following aspects:

1. ***Massive nature:*** In recent years, with the rise of the Internet of things applications, the sensor is used to collect data. With the expansion of the scale of this application and applied in more and more fields, the amount of data will show a trend of explosive growth. How to effectively improve the technology and method of existing or proposed new technology and method to efficiently manage and deal with these data will be the key to extract information from the data and further integration, reasoning and decision making.
2. ***Heterogeneous:*** The application of a variety of computing in the cloud, different industries in different areas in the stage of data acquisition equipment, means and methods are different, the data in the data form and data structure is also different. There are different types of sensors, such as carbon dioxide concentration sensor, temperature sensor, humidity sensor, sensor in different categories the capture, transfer the information content and format of information is different. The above factors lead to data access, analysis and treatment of a variety of ways. Multi-source data lead to data with different points, different categories have different data formats, resulting in structured data, semi-structured data and unstructured data coexist.
3. ***Non deterministic:*** The uncertainty of data in cloud computing has obvious characteristics, mainly including the uncertainty of the data itself, uncertainty

and query semantic matching analysis of uncertainty, and so on. In order to obtain accurate information object, so that people more comprehensive expression and reasoning.

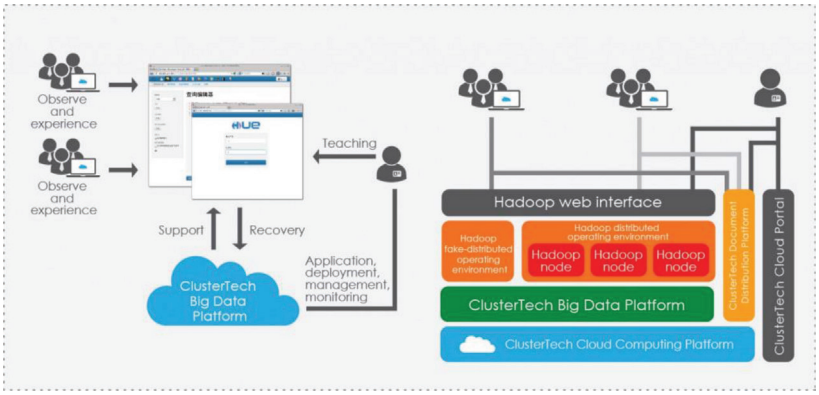

Figure 4 – Cloud computing platform

3. Online education platform based on Cloud Computing

3.1. MOOC online education platform

MOOC is a massive open online courses, namely: Massive, refers to the learning of the course registration and access number; Open refers to the learning needs of learners, can on-line learning; Online refers to the learning time and place is not restricted, open 24 hours, through the network industry, discussion, interaction and evaluation of relevant knowledge; Course, MOOC curriculum design is similar to the university courses, but MOOC courses will be re decomposition, re recycling, so as to adapt to the online education.

MOOC as a new teaching model, it will be the curriculum re decomposition, re - recycling, so that it can adapt to online education. It mainly has the following several aspects of the characteristics.

1. ***The change of teaching idea:*** The "Teacher centered" is the traditional current education philosophy, MOOC is the "student centered", flipping the classroom, the teacher just the tutor, network classroom teacher-student interaction is the communication between places, by providing appropriate teaching materials to guide learners, learners can according to individual needs, learning and rhythm control the way, will work hard, diligent, active and personalized involved, and get the desired knowledge.
2. ***Change of teaching scale:*** The traditional university courses generally only tens to hundreds of learners, and a course Mu hundreds of thousands of people, is a course with a maximum of 240000 students participate in MOOC, completely broke the class size limit, breaking the traditional concept of teaching mode in class.

3. ***Change in learning:*** Traditional learners in accordance with the syllabus of the course to learn, lack of flexibility. MOOC learners can according to their own time, the course of the master of the situation, the use of time to learn, self control of learning progress, the completion of the corresponding curriculum.
4. ***Open change:*** The traditional university is in the classroom, laboratory, library and other places to learn, and the corresponding course between each university is not open to the outside world. MOOC learners with a computer or a mobile terminal, as long as the Internet, you can learn all kinds of high-quality courses, and communicate with learners from around the world, these MOOC curriculum resources are open to all.
5. ***Changes in the way of learning:*** The traditional education in 45 minutes as a unit of classroom teaching, MOOC video course was cut into 5-15 minutes of micro course, learners can use online resources and some additional learning materials, and found in the course of teaching through many small problems, and on the platform directly put forward their own ideas and doubts. There will be teachers and students to discuss or provide answers, learning interest and initiative learning will be greatly improved.

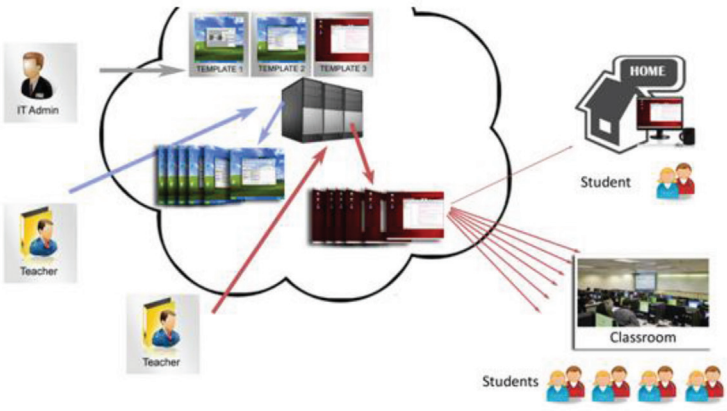

Figure 5 – Cloud computing education platform

3.2. The basic hardware configuration

In this paper, we use a minimal set to illustrate the specific process of cloud environment construction. The Hadoop cluster needs at least 3 servers, each server is configured with 4GB memory, 360GB disk space, and ubuntu12.10 operating system. In HDFS Hadoop environment, the server node is divided into the name node and data node, the name of the node is only one, data nodes can have more than one. For MapReduce cloud computing, server nodes can be divided into JobTracker and TaskTracker JobTracker, which is only one, TaskTracker can have multiple. Among them, 1 servers deployed as a named node and the MapReduce HDFS of the JobTracker; the other 2 servers deployed for the data node and the MapReduce HDFS TaskTracker Java JDKl.60 and above.

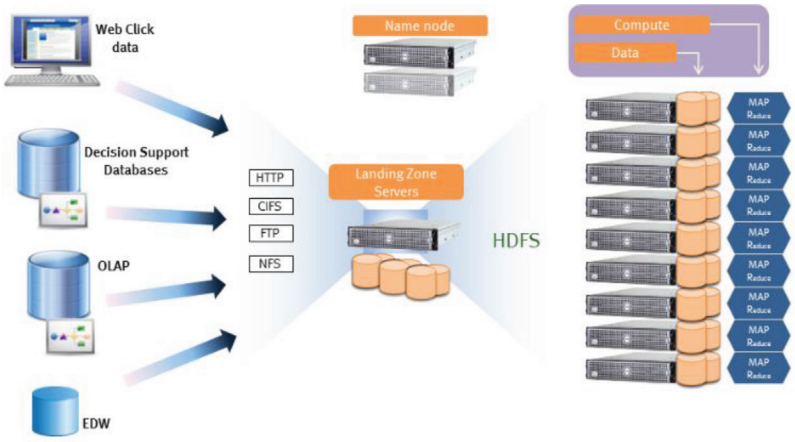

Figure 6 – Hadoop deployments

3.3. Cloud platform environment settings

Hadoop cluster is the name of the node through the SSH (SecureShell) to start and stop the process of the guardian of the node. Configure SSH to use public key authentication without password to use no password authentication of public key, all in the cloud computing no longer need to enter the password to execute commands between nodes, so the SSH authentication configuration of this step is very important.

Hadoop requires that all computing nodes on the deployment directory structure is exactly the same, and there is an account of the same user name. Cloud platform of the 3 machines have a CCEA account, the main directory is / home / ccea.

The configuration defines the required in Hadoop conf / directory Hadoop site.xml, its value will override the default Hadoop value in default.xml, can be completed according to the actual needs of the custom. Cloud cloud education network platform in the Hadoopsite.xml settings are as follows:

< property>
< name > hadoop. tmp. dir < / name >
< value > / root/hadoop/data</ value >
< description >a base for other temporary
Directories < /description >

< property >
< name > fs. Default. name < / name >
< value > HDFS: / / node01:9000< /value >
< description > a base for other temporary
Directories < /description >

4. Online education interactive platform application advantage

The combination of cloud computing and big data plays an important role in the development of online education, not only to promote the development of modern education, but also to enhance the core competitiveness of online education, and to maintain the healthy development of online education.

1. To achieve personalized precision services

Online education platform services to the core of resources to the user as the core of the personalized service change, to provide personalized services for different users. Study on online education platform users of cloud computing and big data applications to strengthen, and based on the analysis results, to improve services, enhance the quality of personalized service, complete tracking service for users, service, knowledge service platform precise association. Analysis of the data source is the same, because of different needs, the results will be different, the services provided are different; the same needs, for different individuals, analysis of different data sources, different results, provide the services are not the same. Try to gain insight into online learners and online instructors in order to drive the development of online education.

2. The trend of educational development and the change of hot spots

Through the big data and cloud computing technology, changes the situation of the passive update teaching resources, according to the needs of online learners to actively update resources, enhance the role of online education platform. Not only to provide online learners with the needs of learning resources, but also for online teachers to provide more quality of teaching resources to provide research basis. Interactive online education platform through collecting and sorting, analysis, mining depth and summary of user data, analysis of the changes of development in the field of education and focus on the macro view, faster to insight into the latest learners' interest, progress and related content, update the online platform for learning resources, and ensure the real-time leading and learning content.

3. Improve the management ability of the interactive platform for online education

Application of cloud computing and big data online education interactive platform for specific application data needs, to achieve rapid, timely and effective response. To adapt to changes in demand and growth, the platform has high performance expansion and expansion stability, efficient processing of various types of data, and on the basis of making full use of their own data value, realize the transformation of data assets from cost center to profit center. Through the integration of data assets, standardization of data assets, the formation of flexible and extensible and easy to update, control, isolation, environmental protection and efficient analysis of data management platform, implementation support standards development, user service, diversified development of a variety of application support mode, form the basis of data and application data are loosely coupled, heterogeneous level two data management level. At the same time, so that online education institutions can control their own data assets, and comprehensively enhance the platform's ability to manage the data information.

5. Conclusion

Information technology can promote the education of revolutionary change, cloud education advocates information architecture and interactive electronic classroom + Mobile Internet terminal + cloud services platform mode, the architecture has been generally recognized by the industry. This paper studies the key technologies of cloud platform and introduces the "cloud education network environment and the system deployment process, the practice and development direction of the application are discussed, in order to promote the practice of education cloud and application demonstration, explore the new mode of digital school. With the rapid promotion of the MOOC project in the world, study and analysis of automatic identification system of Hadoop and Map Reduce applications to the unfinished "learners also can continue to study on the future can be combined with related sites in depth data mining, and through the user interface for MOOC service providers can use the analysis system to adjust teaching strategies, and improve the class rate.

References

Bhushan, N., Li, J., Malladi, D. (2014). Network densification: the dominant theme for wireless evolution into 5G. *IEEE Communications Magazine*, *52*(2), 82–89.

Dong, X., Thanou, D., Frossard, P. (2015). Laplacian matrix learning for smooth graph signal representation. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3736–3740.

Essinger, S., Zhu, X., Schnee, M., Liu, J., Shen, X., Chen, L., & Lu, J. (2013). *U.S. Patent No. 8,457,013*. Washington, DC: U.S. Patent and Trademark Office, 45–50.

Fang, Y., Gang, H., & Chen, Y. (2013). Construction of Information Security System for the Campus Card System. *International Information Institute (Tokyo). Information*, *16*(2), 1015.

Gonçalves, M. J. A., Rocha, Á., & Cota, M. P. (2015). Interoperability Framework for Competences and Learning Outcomes. *Journal of Universal Computer Science*, 21(8), 1042–1060.

Gutiérrez, J., Villa-Medina, J. (2014). Automated irrigation system using a wireless sensor network and GPRS module. *IEEE transactions on instrumentation and measurement*,*63*(1), 166–176.

Huang, K. (2014). Applied-Information Technology in Virtualized Cloud Storage on Campus. In *Applied Mechanics and Materials*, 685, 571–574.

Jiménez, M., Vicente, E. (2015). Safeguard selection for risk management in information systems: a fuzzy approach. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (15), 83–100.

Ju, H., & Zhang, R. (2014). User cooperation in wireless powered communication networks. *2014 IEEE Global Communications Conference*, 1430–1435.

Qian, Y., & Shi, Q. (2014). Study on the Application of Data Mining Based on Campus Card Platform. In *Advanced Materials Research*, 846, 977–980.

Qiong, Z. (2013). Thinking of Application of College Campus One-Card in University Financial Management. *Science Mosaic*, *4*, 12–15.

Wang, C. X., Haider, F., Gao, X. (2014). Cellular architecture and key technologies for 5G wireless communication networks. *IEEE Communications Magazine*, *52*(2), 122–130.