Design of Mobile APP User Behavior Analysis Engine Based on Cloud Computing

Qiang Ren*

School of Information Engineering, Xi'an University, No.1 Keji 6th Road, Xi'an

*Corresponding author e-mail: rqemail@qq.com

Abstract. With the rapid development of mobile Internet, operators are facing more and more fierce competition, traffic management is imperative. Accurate marketing based on user behavior analysis is an important means. However, in the era of big data, with the rapid growth of mobile Internet services and the increase in the number of users, the traditional architecture is difficult to adapt to the needs of mass data mining. The purpose of this paper is to make the Internet operators in China face a new development opportunity, start to move towards the road of traffic management to traffic management, and conduct an in-depth analysis of users' behavior rules, to explore the real needs of the market and the majority of users. This paper proposes a cloud-based mobile Internet big data user behavior analysis engine solution, including the design of the overall system architecture, big data warehousing and preprocessing components, big data user behavior analysis model and other key modules, and finally analyzes the system test results.

Keywords: Cloud Computing, Mobile APP Users, Engine Design, User Behavior

1. Introduction

In recent years, mobile APP users in China have started to move from traffic management to traffic management, and conducted in-depth analysis of users' behavior rules to explore the real needs of the market and the majority of users. In order to fully meet the changing needs of users, operators must constantly develop and launch new products and enhance the functions of computer technology. Cloud computing technology is such a powerful engine system that can meet the needs of massive data analysis and processing. And user behavior analysis is somehow to mining of user traffic situation,
then access to the data for effective statistical analysis, in order to get the user access to the basic law, and then to adjust enterprise's network marketing strategy, so that is consistent with the user's site access rules, make more scientific and effective network marketing strategy. In general, operators based on user behavior monitoring, can obtain the effective dynamic data.

Cloud computing is a powerful technology for performing large-scale complex computing. It eliminates the need to maintain expensive computing hardware, dedicated space, and software. There has been a massive increase in the size of data generated through cloud computing, or big data. Noraziah and A introduced several big data processing technologies from the aspects of system and application. First, from the perspective of cloud data management and big data processing mechanism, the key issues of big data are proposed. Next, Noraziah, A will introduce cloud computing of big data and related work. In addition, Noraziah, A also discussed the shift of big data to the cloud [1]. Sen Liu's goal is to focus on the value creation potential of cloud computing in inter-enterprise partnerships. It examines how the flexibility and integration capabilities of a cloud-based It infrastructure contribute to the agility of partnerships to improve enterprise performance. Sen Liu also introduced business life cycle and market turbulence as internal and external environment variables to investigate the different roles of cloud computing in value creation. Design/method/method -- the questionnaire surveyed 184 customer companies of alibaba cloud, China's largest cloud computing service provider. Partial least squares (PLS) method was used to test the theoretical model. Results - cloud infrastructure (CI) flexibility has a positive impact on partner agility, and the impact of CI integration on partner agility is moderated by the business lifecycle and market volatility [2]. With the rapid increase of demand-side monitoring equipment and controllable facilities, more information and communication technology (ICT) resources are needed to support the development of DSM. Unlike traditional computing in power systems, where ict resources are individually tailored for mapping applications, DSM specifically requires scalability and economic efficiency because of the growing number of stakeholders involved in the computing process. Zijian Cao proposes a new cost oriented optimization model for cloud-based ICT infrastructure to allocate cloud computing resources in a flexible and cost-effective manner. The model also takes into account uncertainties such as imprecision of load prediction and unavailability of calculation examples. In order to effectively solve the proposed optimization model, Zijian Cao specially developed an improved priority list algorithm and compared it with the mature simulated annealing algorithm [3-4].

Is proposed in this paper for the mobile end user path analysis model of the App, and through the data in the App page buried points, used to collect user data, Session data, page access data, by combining the depth traversal, improved maximum forward algorithm, output the user access to the maximum forward path during each Session, finally combining ApacheSpark data processing platform for data arrival rate calculation App each function module. This topic research path for mobile
terminal App user behavior model, user behavior path algorithm, to optimize the design of the mobile
terminal App product, has the very big Guiding significance.

2. Proposed method

2.1. Overall design scheme of the system based on cloud computing

The mobile Internet user behavior analysis engine designed in this paper realizes distributed and
concurrent large-scale computing power through cloud computing technology, and builds an
end-to-end big data mining and analysis system for the mobile Internet, thus realizing preference
analysis. The online behavior of DPI and application platform users. In the whole process from data
collection and analysis to service provision and marketing implementation, personalized
recommendation services are provided [5]. The system consists of one server as the main node server
of Hadoop platform and Hbase, and another server as the slave node server of Hadoop platform and
Hbase. The primary node server is mainly responsible for assigning tasks and traffic to the slave node
server and monitoring the execution status of the slave node server. Multiple slave servers perform
specific tasks under the control of the primary node server [6].

2.2. Design of big data user behavior engine components

The analysis of big data user behavior is mainly realized through Hado knocking platform, which has
a variety of functional structures and components. The model evaluation system mainly verifies the
created model, compares the calculated results of the model with the actual data, outputs the verified
indicators, and makes the evaluation related to the model. More business data correlation model by the
user's Internet behavior and the behavior of the business platform game, both for effective analysis,
and the Internet to DPI users behavior is analyzed, the correlation of preference may discover the laws
of user behaviour on the Internet, to enter these laws system, promote the realization of enterprise's
marketing; Personalized recommendation is the use of computer filtering technology to extract
effective information, and the product content and customer behavior for comprehensive analysis, to
provide users with personalized recommendation services; In addition, through the measurement of
DPI access preference, users' Internet behaviors can be classified, and the types of web pages visited
by users can be analyzed to obtain users' behavior rules and preferences [7].

Mainly through changing the position of the user to determine, through analyzing the user's
location, you can learn the user's social networking sites and networks, whereas interest graph to a
certain extent, to establish a mutual interest relationship between users. Construction of these models
are all the graphs with different levels of application, and in the process of using, mainly through the
map and reduce these two functions. The first completed by the map to the input of data and
calculation, and then in the form of <key, value> output, then the output values using Hadoop collection and transmission. The reduce function is handed to the reduce function, and then the reduce function is processed accordingly, which is also output in the form of <key, value>, and the calculation of user preference [8]. <key, value> Generally, the algorithm is divided into three steps. First, Hbase is used to read and analyze the user's behavior data and combine the obtained data. Key represents the user's ID and value represents the user's Internet behavior and browsing content. Second, the reduce function is used to calculate the behavior information. In this case, the key represents the user ID and the ID of browsing content, while the value represents the content preference. Finally, the list of recommended contents is determined through the calculation results. The user's analysis model is distributed to each specific project, forming an integrated operation mode and improving the operation efficiency [9-10].

3. Experiments

3.1. Experimental environment

In order to verify the design scheme of mobile Internet big data user behavior analysis engine based on cloud computing proposed in this paper. In the laboratory, a four-node Hadoop distributed platform was built using PC server, one of which was the primary node and three secondary nodes. The big data store was used as a test solution and compared to a standalone operating environment. Five sets of system operations with different data sizes were tested.

3.2. Experimental data

The platform mainly USES is bulk storage way, every day for a data warehouse, the size of the file is commonly GB level. Usually for large data files use Had00p platform more into the Treasury, but through the practice has proved that graphs distributed processing Hbase warehousing efficiency generally is not high. Therefore, can be optimized on this system, on the basis of the distribution graphs processing, and then a batch library processing. In the process of Hadoop system, use TextInputF0Hnat more, its more than shown in the map is a single record file '5. . Therefore, NLineInputFormat class can be used to realize mass storage in the graphs, with the support of such a system, each shard will leave N rows, and then through the proper parameter configuration, implement every time can read N rows, batch warehousing related operations on the map, boc, can to a certain extent, improve the efficiency of data analysis.

4. Discussion

4.1. Mobile app user management and scheduling analysis
Experiment is mainly used to control task scheduling, can not only provide the task creation function, but also has the task of adjusting and delete function, as long as to choose the type of business, set up the science the lifecycle, definition of processing logic corresponding to the data extraction, organize, and run the project for effective control, can realize the automation of data processing, and also provide personalized management functions such as suspend, resume.

(1) Storage and preprocessing of big data

This system mainly deals with the Internet access of DPI users and the relevant user behavior data in the application platform. It can transfer these data to the user behavior analysis engine of the system, providing a data reference and basis for data analysis and model mining.

(2) Behavior analysis of big data users

All converge to the system of the mobile Internet user behavior data analysis, mainly through the graphs user behavior analysis model of resources, to the user's online habits, preferences, and even the user's social relations for effective analysis, and provides a comprehensive business services, to users recommend specific content. In general, from the node server and the master node server not much differences in the structure of the software, the only difference from the node server does not have the deployment task scheduling and management functions. Taking data analysis of sogou as an example, the market share of sogou search has reached 15.68%, making it the third largest Internet search company in China. The data analysis format of user behavior is shown in Table 1.

| School name | Fields that                      |
|-------------|----------------------------------|
| QueryTime   | Access time                      |
| UserID      | The user ID                      |
| QueryTerm   | The query words                  |
| Rank        | The ranking of the URL in the returned result |
| SequenceN0  | The serial number the user clicks |
| URL         | The URL the user clicks on       |

4.2 Analysis of operation results

As shown in Figure 1, it can be seen from the test results that the performance of the library in the
single machine batch mode is very low, with about 5 sets of data of 3 GB, 6 GB, 9GB, 12 GB and 15GB. The system running time is respectively 32 minutes, 60 minutes, 95 minutes, 121 minutes and 151 minutes. The running time of the system increases almost linearly with the increase of data volume. For mass data, it is difficult to complete the operation of data warehouse in a reasonable time. Using cloud computing distributed batch storage can greatly reduce system processing time. For the 5 data sizes of 3GB, 6GB, 9GB, 12GB and 15GB, the system running time was 14min, 29min, 35min, 47min and 55min, respectively. It can be seen from the comparison of the test results of the two methods that with the increase of data scale, the processing time gap between cloud computing distributed batch storage and independent batch storage becomes larger and larger, and the advantages become more and more obvious. In this way, the system assigns data-intensive operations of the big data user behavior analysis model to each computing node. Therefore, the performance of the model is greatly improved.

![Diagram](image)

**Figure 1.** Performance comparison between single and distributed batch incoming test

### 5. Conclusions

This paper aims at the challenges brought to operators by the era of mobile Internet big data. This paper proposes a design scheme of mobile Internet big data user behavior analysis engine based on cloud computing. Verified by experiments. This scheme can effectively deal with the "short, flat and
fast" features of new mobile Internet services and new products, and the data scale is rapidly expanded, which can complete the task of big data processing and analysis in effective time. Mobile Internet the advent of the era of big data, the operation of the enterprise is confronted with great challenges, the study on the big data under the background of user behavior analysis engine puts forward the design scheme of actualizing the supplier for the development of new business. It is able to produce effective tracking to the customer, and the operation cost is low, has great application value in the enterprise's marketing, is worthy of popularization and application.

References

[1] Noraziah, A, Fakherldin, Mohammed Adam Ibrahim, Adam, Khalid. Big Data Processing in Cloud Computing Environments[J]. Advanced Science Letters, 2017, 23(11):11092-11095.

[2] Sen Liu, Yang Yang, Wen Guang Qu. The business value of cloud computing: The partnering agility perspective[J]. Industrial Management & Data Systems, 2016, 116(6):1160-1177.

[3] Zijian Cao, Jin Lin, Can Wan. Optimal Cloud Computing Resource Allocation for Demand Side Management in Smart Grid[J]. IEEE Transactions on Smart Grid, 2017, 8(4):1943-1955.

[4] Yahya Al-Dhuraibi, Fawaz Paraiso, Nabil Djarallah. Elasticity in Cloud Computing: State of the Art and Research Challenges[J]. IEEE Transactions on Services Computing, 2017, PP(99):1-1.

[5] Yumin Wang, Jiangbo Li, Harry Haoxiang Wang. Cluster and cloud computing framework for scientific metrology in flow control[J]. Cluster Computing, 2019, 22(1):1-10.

[6] Hamid Madni, Abd Latiff Muhammad Shafie, Coulibaly Yahaya. Resource Scheduling for Infrastructure as a Service (IaaS) in Cloud Computing: Challenges and Opportunities[J]. Journal of Network & Computer Applications, 2016, 68(C):173-200.

[7] Jiaxi Liu, Zhibo Wu, Jin Wu. A Weibull distribution accrual failure detector for cloud computing[J]. Plos One, 2017, 12(3):e0173666.

[8] Najme Mansouri. Adaptive data replication strategy in cloud computing for performance improvement[J]. Frontiers of Computer Science (print), 2016, 10(5):925-935.

[9] Jian Shen, Member, IEEE. Anonymous and Traceable Group Data Sharing in Cloud Computing[J]. IEEE Transactions on Information Forensics & Security, 2018, 13(4):912-925.

[10] Humphrey M. Sabi, Faith-Michael E. Uzoka, Kehbuma Langmia. Conceptualizing a model for
adoption of cloud computing in education[J]. International Journal of Information Management, 2016, 36(2):183-191.