The Effective Application of Cloud Computing Technology in Large Data User Behavior Engine Design

He Wenmeng

College of Computer Science and Technology，Wuhan University of Science and Technology, Wuhan 430065, China

1204563618@qq.com

Keywords: Cloud computing; Big data; User behavior; Engine design; Computer information technology

Abstract. With the constant development of socialist modernization construction in our country, our country's computer information technology has made effective progress, and makes the world into an information age. People's production and life style experience a series of change. However, with the emergence of information diversity and multi-user mode, traditional computer information technology already can't satisfy people's needs, and the development and application of cloud computing technology come to the stage. This article focuses on analyzing the big data user behavior engine design under cloud computing technology, covering big data flow management, multi-user system design, and other aspects to explore the effect of the system test.

In recent years, China's mobile Internet technology has made full development, which makes the Internet operators in our country face a new development opportunity and began from traffic management to flow management. The law of user behavior is analyzed to explore the real demand of market and the users. In order to fully meet the changing needs of users, operators must constantly develop and launch new products and strengthen the function of computer technology. Cloud computing technology is such kind of engine system which can satisfy the analysis and processing of huge amounts of data.

RESEARCH ON CLOUD COMPUTING SYSTEM OVERALL DESIGN

The overall architecture of cloud computing system

This study mainly uses cloud computing technology's huge amounts of data computing to set up complete mobile Internet data mining analysis system. Realize the analysis for Internet user behavior engine, and according to the user's preferences online habits and behavior, provide users with targeted personalized service and form a unified organic whole of data collection, analysis and service type and marketing strategy to improve enterprise's marketing efficiency. In addition, cloud computing system mainly realizes data collection with the help of FTP server, and makes distributed computing and data batch processing at the system interface. Large data shall be deposited in Hbase database. The system can not only realize mass data storage, but those unstructured data storage. Then through Hive integration layer and summary EIL processing, use MapReduce data analysis model to transmit the processing result into the database. The system's overall structure is shown in Figure 1:
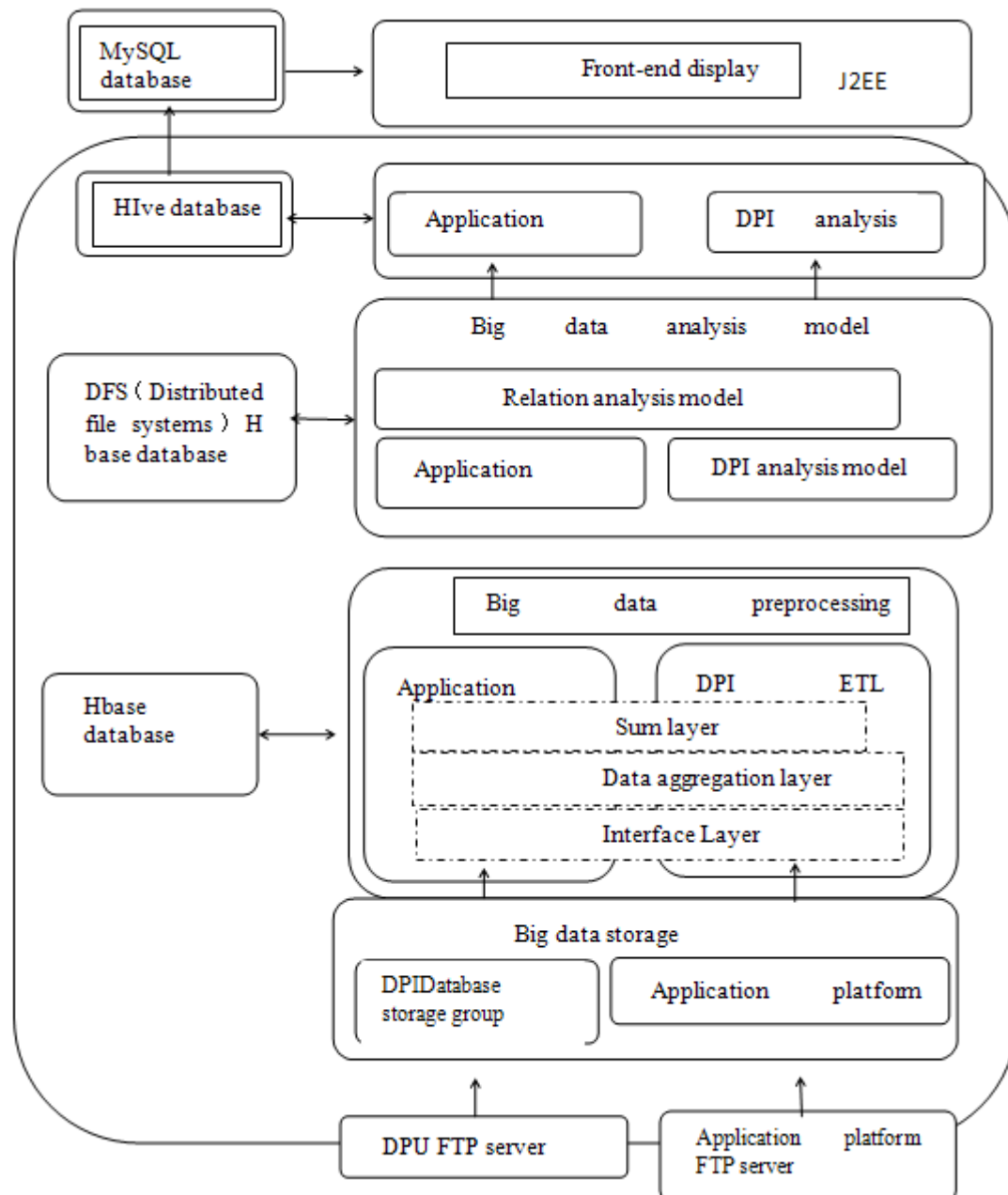
Figure 1 The overall architecture of cloud computing system

System topology and function distribution

System topology mainly refers that a server is taken as Hapdoop platform's master node server, and the others are Hapdoop platform's node servers. In general, node server can be dynamically extended according to the actual need, and the master node server not only assigns tasks and flow to node server, but also monitors the work execution of the node server. More than one node server participates to complete the task, and it can improve data processing efficiency. The master node server's software running status is shown in Figure 2:
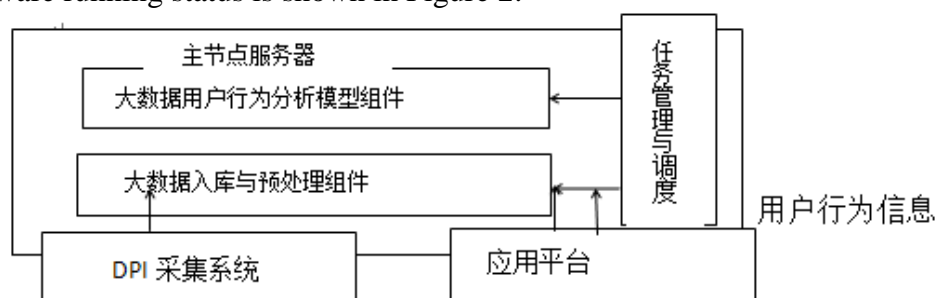

Figure 2 Master node server function structure

1. Task management and scheduling

Task management and scheduling is mainly used to control scheduling, which can not only provide task creation function, but also has task adjusting and delete function. Choose the type of business, set up scientific life-cycle, and define the according logic to effectively control data extraction and organize and run various projects [i]. It can realize data automation processing, and also provide personalized management functions such as suspend, resume.

2. Large data storage and pre-treatment

This system mainly processes DPI users' access to the Internet and application platform user behavior information data processing, which can transmit these data into system user behavior analysis engine, and provide a data reference and basis for data analysis and model mining.

3. Big data user behavior analysis

All converge of mobile Internet user behavior data in the system is analyzed, which is mainly with the help of MapReduce user behavior analysis model resources. It can make effective analysis for user's online habits, preferences, and even user's social relations and provide a comprehensive business services, and recommend specific content to users. In general, sub-node server has little difference from the master node server in software structure, and the only difference is that sub-node server does not have function deployment and management scheduling [ii]. Sogou company data analysis is taken for example. According to the third party Internet data statistics institution CNZZ released September 2015 Chinese Internet users search engine usage statistics report, Sogou search market share has reached 15.68% and becomes China's third largest Internet search company. User behavior data analysis format is shown in Table 1:

Table 1 Cloud computing technology of data formats

Field name	Field explanation
Query Time	Query Time
Use ID	User ID
Query Term	Query term
Rank	The URL ranking in the returned results
Sequence NO	User click serial number
URL	Users click URL

THE BASIC DESIGN OF LARGE DATA STORAGE COMPONENT

Usually, the data of mobile Internet user behavior analysis engine is mainly gain through two channels: DPI data and application platform data, and the two data sources are obviously different. First of all, application platform data is concentrated and in one access behavior table, there is a complete show, and everyday performance is a file, and file size is classified by GB; DPI data is in many scattered small files, and these files are generally no more than 10 MB. One of its most important features is that source file is particularly fast, and averagely every 2 min, we can receive multiple new files. About 1 TB amount of data can be collected in 1 day, so it is of high efficiency. According to the characteristics of these two kinds of data sources, different design schemes are respectively adopted.

Application platform data storage

The platform mainly adopts bulk storage way. Every day there is data storage, and the size of the file is commonly GB level. Usually large data files use Hadoop platform to get stored, but through the practice, it has proved that `MapReduce` distributed processing Hbase storage efficiency is not generally high; therefore, this system can be optimized. On the basis of `MapReduce` distribution processing, practice bulk storage processing. In Hadoop system application process, use TextInputFormat more. It is often shown as file single record in map [iii]. Therefore, NLineInputFormat can be used to realize mass storage in `MapReduce`. With the support of such system, each shard will leave N rows, and then through proper parameter configuration, implement every time N rows reading record. Implement batch storage related operations in map, which to a certain extent improves the efficiency of data analysis.

DPI data storage design

DPI data storage mainly aims at large number of small data files. Although these documents contain small data volume, data sources frequency is particularly high. First of all, use SequenceFiles software to package small file compression, and at the beginning of file package, take measures to reduce the number of small files. The reading, however, is subject to certain restrictions. Both Hadoop shell software and Map reading software are difficult to realize flexible access to data. Adopt HAR software for small file archiving. Although it can to a certain extent reduce the number of small files, still it is hard to achieve flexible data read. This largely is the result of low HAP software read performance. Although Hadoop append system can supplement these small files to the same file, these small files sizes are differ in thousands ways and the data source frequency is extensive with peak and trough, therefore, the control of these data is relatively complicated. In addition, there are Flume, FLumeNG and Scribbe system. These systems can process the data of middle layer, effectively reduce the number of small files, but the two transmissions and compress file function are not strong and have certain defects. Thus, these four kinds of ways cannot realize effective processing of data files. Therefore, to make the scientific analysis of the characteristics of DPI data, CombineFile InputFornt should be effectively inherited. Set data fragmentation size in order to realize CreateRecordReader, and this way can also promote DPI large data storage.

BIG USER BEHAVIOR DATA ENGINE COMPONENT DESIGN

Big data user behavior analysis is mainly through Hadoop platform to achieve. It has many kinds of functional structure and components, and the relationship between the components is shown in Figure 3:
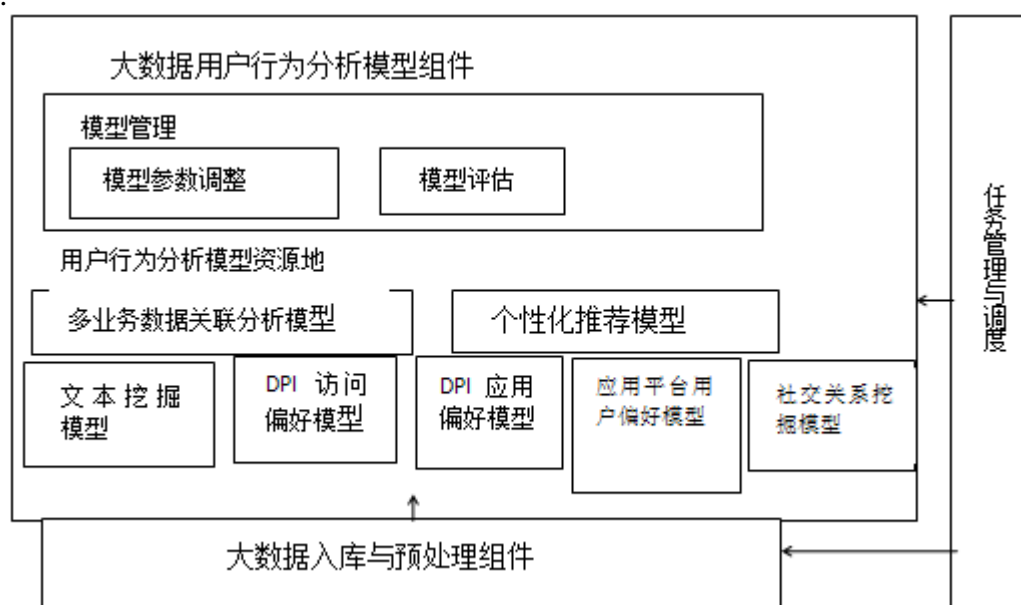

Figure 3 Big data user behavior analysis model components

First, model parameter adjustment is mainly to complete the variable set of model algorithm, and effectively adjust the parameters and set space model according to the actual situation; Model evaluation system is mainly to check the created model and compare model calculation results with actual data, then output check indexes, and make model related evaluation; Multiple business data correlation model makes effective analysis for both by the user's Internet behavior and game business platform behavior, and DPI users' Internet behavior preference correlation is analyzed to discover the laws of user behavior on the Internet, and transmit these laws to system to promote the realization of enterprise's marketing; Personalized recommendation is to use computer filtering technology to extract effective information, and carry on the comprehensive analysis of product content and customer behavior. Provide users with personalized recommendation service; through DPI access preferences, classify user behavior on the Internet, and analyze the user's access page

types so as to get the user's behaviors and preferences. Social relationship mining model not only contains the user's social graph, but also has the user's interest graph, which is mainly determined through the change of user's location. Through analyzing the user's location, you can learn the user's social networking sites and networks, whereas interest graph, to a certain extent, can establish a mutual interest relationship between users. Construction of these models is of different levels' MapReduce application, and in the process of using, it is mainly through these two functions: map and reduce to implement. First the map completes input data and calculation, then output in the form of < key, value >. Then use Hadoop to collect Hadoop passed number value and output to reduce function, and then by the corresponding processing, output in the form of < key, value > as well. Usually its algorithm is mainly divided into three steps. (1) The calculation of user preferences degree. First of all, use Hbase to read and analyze the user's behavior data, and then combine the obtained data. Key is user ID. The user's Internet behavior, and browse content are expressed as value; And then use reduce function for calculation. Key here refers to user ID and browse content ID. Value is preference degree of the content. Calculate users' content attribute preference. Still map function transmits it to reduce, and uses reduce to calculate content attribute preferences of the user. Finally, determine the recommended content list by the results of calculation. Form the integrated operation mode from the users' analysis model distribution to each specific project, and enhance the operation efficiency.

CONCLUSION

The advent of mobile Internet big data era makes the operation enterprises confronted with great challenges. The study puts forward the design scheme for user behavior analysis engine under big data background, which will actualize the suppliers' new business development. It can form efficient tracking of the user with low operation cost, and has great application value in enterprise marketing. It is worthy of popularization and application.

REFERENCE

[1] Tao Caixia, Xie Xiaojun, Chen Kang, Guo Lirong, Liu Chun. Mobile Internet big data user behavior analysis engine design based on cloud computing [J]. Journal of Telecommuter Science, 2013, 01:27-31.

[2] Liu Lu. Mobile Internet big data user behavior analysis engine design based on cloud computing [J]. Journal of Electronics, 2014, 174:174-175.

[3] Xie Xiaodi. Cloud computing distributed data management and analysis tools study under the environment of big data [J]. Journal of Science and Technology, 2015 (19) : 51-51.

[4] Li Xiaofei. Big data system processing research based on cloud computing technology [J]. Journal of Changchun College of Engineering: Natural Science Edition, 2014, 15 (01).