

Distributed Big Data Modeling Research

Min Tu¹ and Fangqiang Zhong²

¹Jiangxi Police College, Nanchang, Jiangxi, 330100, China

²Nanchang University, Nanchang, Jiangxi, 330100, China

Abstract—This paper focuses on how to better build a model from resource aggregation to analytics mining in mainstream cloud computing and big data. Distributed Big Data Analysis Modeling Research, provide high reliability, high security and high efficiency services in analysis and judgement for various intelligence analysis and intelligent decision-making.

Keywords--big data; mining; cloud computing; modeling

I. INTRODUCTION

In the information era, A lot of traces produced by people's various activities and brought so much information resources. How to quickly and effectively extract potential and valuable data from a variety of information resources to provide assistance for various types of analysis, control, service, and decision-making is the most effective way to improve government capacity at this moment. Cloud computing and big data technologies have the ability to support the storage and analysis of massive data. Big data project's core is the efficient and highly compatible Data Governance between cloud and applications, data and applications, clouds and clouds.

At present, the key technologies for secure collection, aggregation and management of massive and heterogeneous information in the field of cross-domain security exchange are still in their infancy in the whole country. The exchange platform in physical isolation directly connect with the database server and cannot well cope with massive data processing and high concurrency. data engine has been introduced in foreign projects as middleware and in the field of distributed big data service bus, foreign ESB service bus has been used as mainstream technology, and foreign ESB projects such as Mule have joined with Load balancing in a way, there is no mature case in China; in the aspect of big data analysis and mining, the combination of distributed processing technology and traditional model in analytical mining is a relatively advanced method. It is mature for large Internet projects in our country. some underlying technologies have been used such as MapReduce, Spark, and the use of Storm which has higher real-time support rarely used in China.

Based on mainstream cloud computing and big data platform, The Research of distributed big data analysis modeling is of great significance to the information era. There are so many functions should be included such as adapting to the complicated requirements of data mining, having the capabilities of cross-network and cross-cloud platform, providing support based on distributed big data analysis modeling for Intelligence and intelligent decision-making, providing new solutions for applications of big data, various risks and challenges.

Based on the cloud computing and big data technology system, "data + model + service" this scheme was adopted by us to research distributed big data analysis modeling.

II. OVERALL DESIGN

Our idea of overall design is based on main-stream cloud computing and big data infrastructure, and builds a distributed big data analysis modeling that supports mainstream cloud and big data platforms and meet those needs: high-efficiency, high-security, high-reliability. finally everyone can quickly mine the required information in the massive data.

The logical structure system of design is divided into: data resource layer and service resource layer.

Data resource layer

The data layer is based on cloud computing resources and big data resources, and provides basic technical support for data resource applications and services, data resource standardization and management. Data storage and organization, data processing and analysis, and data applications and services based on development, advanced, mature, and compatible principles.

According to the requirements of building database in mining big data resources, realize the integration and standardization of data resources, decompose and correlate the time and space elements of data resources, and establish a library of events based on the elements of geography and time. Using the technology of big data, it is possible to quickly locate the relevant personnel, vehicles, trajet-ories in a manner named event stream and you can set the factors of temporal and spatial, and truly display the complete life cycle trajectory for the entity of data from generation to end.

Through the collection and integration of data resources, relying on the big data storage of the cloud computing platform, deepen the planning and management of mining resource data, and deepen the management of data reso-urces.

Through the collection and integration of data resources, relying on the big data storage of the cloud computing platform. The Data resource can be integrated and managed deeply.

Service resource layer

The center of System Sharing Service, which provides support for big data service bus and Distributed mining service built by visual analysis mining modeling system, establishes service standards, analyzes and mines service resource management system, and provides distributed mining service based on Big data for various applications by unified standards.

This layer is the core, researching reliable big data service bus platform, establishing standardized analysis and mining service resource management, improving service resource standards, deepening service resource management, providing analysis and mining big data service development accomplished by visual analysis and analysis modeling system.

III. INTEGRATION OF HETEROGENEOUS BIG DATA

We can build a big data resource management platform, truly realize the standardization and visual management of data resources. it is a Hadoop/Hbase/HDFS Distributed organic integrated big data resource center(instance: based on Hadoop big data platform).it includes three points. Firstly, efficient, secure, and reliable data integration for various heterogeneous data. Secondly, standardized data construction based on data characteristics. Thirdly, forming relational structured data, distributed semi-structured data, and unstructured data.

Big data resource aggregation

Massive and heterogeneous data can be integrated and connect to the big data platform of Hadoop through many adapted components of data. and complete the access of data, data standardization and datastores by integrating various data such as structured data, binary data, quantity data, cross-net data and the result of big data analysis.

Standardization of Big data

The standardization of data is mainly for the specification of related elements that come from underlying data. It includes two parts: data resource and data exchange and share, in which data resources include metadata, data elements, data dictionary and data directory, etc. Data exchange and share includes standards related to the transaction of data and the open sharing of data.

The core value of big data lies in the Relational Analysis of data from many different sources. Therefore, relevant standards including the transaction of data and the open sharing of data are particularly important. Data transactions and open sharing can be carried out efficiently only if the participants follow a common standard.

IV. RESEARCH ON VISUAL ANALYSIS MINING MODELING

Visual Analysis Mining Modeling is a built-in mining model component. It realizes the customization of visual analysis and mining modeling through flexible programming with data resources. And in the underlying, we can transform the instructions of analysis model into corresponding code based on Big data by analyzing the modeling engine.

Visual analysis mining modeling is mainly divided into three business models.

Off-line analysis mining model: it is Mainly used for analysis of fixed offline data of big data. The input of parameters and conditions are relatively fixed in this scenario. By customizing the offline analysis mining model and setting the timing task. The result of calculation and analysis is pushed to the specified data source. In this mode, the processing of the result data can be automatically pushed by customizing Message Service or encapsulated as a Request Service and it can be invoking by Big data application.

Static real-time analysis mining model: it is mainly applicable to fixed and pre-defined analysis models. The input parameters of the model are generally inputted with the users of big data applications according to the requirements, so the triggering of the model needs to provide the request service to trigger. We can get the result by synchronous or asynchronous according to the business requirement and the actual time of calculating.

Dynamic real-time analytics mining model: This business model is mainly applicable to the scenario where the visual analytics mining modeling is encapsulated into a specific application. The mode has fixed the relevant data and model components according to the requirements of the specific application. Finally user directly design the model according to the requirements of business. After the design is completed, the results can be calculated immediately. The results can be displayed with graph through Echarts(Chart Tool Of Data Visualization) in real time.

V. CONCLUSION

This paper focuses on how to better build a model from resource aggregation to analytics mining in mainstream cloud computing and big data. and provide high reliability, high security and high efficiency services in analysis and judgement for various intelligence analysis and intelligent decision-making by using distributed big data analysis modeling. finally everyone can quickly mine the required information in the massive data.

ACKNOWLEDGMENT

Fund Project: 2017 Science and Technology Research Project of Jiangxi Provincial Department of Education "Distributed Big Data Modeling and Data Mining Practical Application Platform" (Project No.: 171055)

Jiangxi Key R&D Program: Research on Electronic Data Trajectory Analysis Technology under Cloud Computing (No. 20161BBE50044)

REFERENCES

- [1] Lu Bing, cloud computing-based border defense information resource management service platform architecture design [J], computer age, 2017 (10)
- [2] Liu Xiao, Architecture of distributed key-value system in big data environment [J], China Financial Computer, 2015(6)
- [3] Fan Jing, Xiong Lirong, Xu Cong, Research and Application of Data Integration in Distributed Enterprise Service Bus Platform [J], Computer Science, 2014(2)
- [4] Liu Guocheng, Wang Yuetang, Research on Security Auditing Issues of Cloud E-commerce[J], Lanzhou Academic Journal, 2017(5)
- [5] The development trend of big data and the standards at home and abroad[EB/OL].<http://tea.baike.com/article16340553.html>, 2018-10-22
- [6] Current status and development direction of big data standardization at home and abroad[EB/OL].http://www.sohu.com/a/260246819_100016410, 2018-10-18
- [7] Several major areas of big data standardization at home and abroad[EB/OL].<http://www.hizcn.com/dsj-yx/post20181011083809.html> 2018/10/11