Design and Implementation of Financial Information Management Information System Based on Data Warehouse Technology

Linlin Yu

zhenxinwn@163.com

Shandong Institute of Commerce and Technology, Jinan, Shandong, China, 250103

Abstract. Based on the combination of data warehouse technology and financial information management, this paper is based on Hadoop and spark big data processing ecosystem framework, combined with HDFS distributed storage, MapReduce computing and spark streaming data interface, and completed the packaging and publishing of the system in SSH development environment in Java language environment. The system is presented in the form of Web, which is convenient for managers to complete all kinds of inquiries and risk assessment through simple and convenient operations, and provides comprehensive application solutions for the problems of complexity, concealment, difficulty and risk in the financial management process in the data age. It not only improves the ability of managers to deal with financial fraud, but also helps to standardize the behavior of enterprise information disclosure, and further makes a useful attempt for the digitalization and information construction of financial management.

Keywords: data warehouse; information management system; SSH; finance

1 Introduction

Internet finance in China has emerged with vitality that can't be ignored. At the same time, it has brought new challenges to internal management and operational risk management. Internet finance has changed the collection method of internal management evidence and the processing method of internal management information. The related financial products of Internet financial institutions, the risks it faces as a new industry, and the internal management problems brought about by it, put forward some suggestions for the internal management measures of Internet financial enterprises. Traditional management has the shortcomings of heavy tasks and strong temporality, lack of systematic management, and difficulty in effective early warning and control of risks, which makes the prevention of economic risks within enterprises lag behind. [1]

However, most financial systems are based on the data warehouse technology in big data technology, so as to realize further data analysis and data mining. In view of the above situation, the author thinks that an Internet financial information management system should be designed. Considering the large and diverse data volume of Internet finance, the system should be developed in combination with data warehouse technology. In this paper, how to apply the management system under the framework of Hadoop and spark is studied. By combing the business processes related to the management of Internet financial institutions and using the

ICBBEM 2023, May 19-21, Hangzhou, People's Republic of China

DOI 10.4108/eai.19-5-2023.2334262

industry-advanced SSH network application development framework of spring+springmvc+hibernate, a management data collection platform is provided, and management model tools and management ideas are integrated into the system design functions, thus helping managers to analyze and manage online management data. [2]

2 Key technologies

2.1 Data warehouse technology

The data warehouse is a widely used technology for decision-making by management departments of large enterprises. Data warehouse usually carries all the data of the whole financial system, and hive in Hadoop big data ecosystem is usually used to build data warehouse.

The basic data warehouse architecture can be divided into four parts, namely front-end tools, OLAP data warehouse servers, data computing tools and data sources. The data source of this system is the financial data of each functional business system collected by the enterprise's internal ERP system. The data calculation module can clean and process the collected data for further reorganization. The core of data warehouse is data logical layering. ODS layer is used to save original data, DWD data warehouse detail layer is used for behavior data and log data, DWS data warehouse summary layer, and ADS data application layer is used to display data. [3]

2.2 Spark

Spark is another big data framework launched by Apache Foundation, which is different from Hadoop. Spark is a processing framework that highlights the performance of speed, ease of use and complex analysis. At the same time, its data sources are more diversified, besides text data and chart data, there are real-time streaming data. Spark provides a data framework that can uniformly manage and process data with different attributes.

Spark SQL is a structured data processing tool based on Spark platform, Spark streaming is an API interface that facilitates the development of streaming data processing programs, MLlib is a distributed machine learning algorithm library based on Spark platform, and GraphX is a computing framework for parallel graphs. Spark ecosystem has been involved in many fields, such as machine learning, data mining, database, information retrieval, etc. At the same time, it has also been widely used in natural language processing and speech recognition. [4]

2.3 Development environment

In this paper, the author briefly introduces the related technologies used in platform development. The financial information management information system uses the big data server cluster which combines spark and Hadoop ecology to process data, and uses JavaWeb technology to develop the corresponding application platform to present the functions and effects of the system. [5]

The operating system of the server cluster is Linux centos7, and the server is established by Apache tomcat. According to the data volume of small and medium-sized financial institutions, the server cluster chooses to build four nodes. Choose a master server as the namenode of Hadoop cluster and also the master of spark cluster. The remaining three servers are slave

nodes, as datanode and slave. The three servers are numbered 1, 2 and 3 in turn. The IP address of the response is assigned to each server node, and the nodes are connected through the ping command. And install the components of Hadoop cluster and spark cluster in the above nodes according to the needs of each node, such as flume, hbase, hive and Spark streaming. After all functional components are installed, the data processing cluster can be initially set up. When configuring the spark cluster, first find the spark-env.sh.template file of /conf in the installation directory of Spark, and copy it to the parent directory. Then use a text editor to edit spark and configure related environment variables. It includes the configuration of parameters such as JAVA_HOME that specifies the Java installation path, SPARK_MASTER_IP memory size that allocates the IP address of the node, scala installation path and Hadoop file directory. [6]

The JavaWeb application of this system uses the development and compilation tool IDEA, the development environment is JDK 1.8, the front-end development language is CSS3+HTML5+Javascript, and the back-end development language is Java. The system development is based on MVC pattern, with vue.js as the front end and SSH framework of spring+springmvc+hibernate as the back end. And MySQL 7.8 is selected as the relational database of the system. Through the introduction of the above key technical theories, the overall environment for the development of financial information management information system, the configuration of related software and tools are determined, and the technical feasibility of the overall project is also clarified. [7]

3 Development process

The data warehouse of financial information management information system is divided into source data, data management part, data storage part and application part. The source data format is two types of financial statement data in ERP system. The data management part is processed by ETL by the data collection and distribution system, and the components of this part extract, transform and load the source data. After processing, the source data will still be sent back to the data collection and distribution system for redistribution. In order to handle data interaction, ETL system will provide a unified transaction request interface, and it is also the only interface between ETL subsystem and financial data management system. [8]

4 Functional implementation

The data source of this system is the financial data of various functional business systems collected by enterprise's internal ERP system, which will be stored in MySQL database. These data will be extracted and used through the data warehouse created by HIVE. For example, the data in the payment information step in ERP system can be extracted by SQL code, as shown below. The extracted data will be cleaned and transmitted to kafka component in spark for integration. [9]

```
select
  'DATA _DATE',
  'TABLE_NATURE',
```

```
    'BANK_ORG_COD',
    'BANK_ORG_NAME',
    'PAY_NANE',
    'PAY_ATTR',
    'PAY_ ORG_ID_NUM',
    'ACC_OF_BEN',
    .....
    'COLL_DAY',
    'CLEAR_DORM',
    'PAY_CCY',
    'TOL_AMO_REC',
    'ORI_DEC_NUM',
    'BANK_BUSINESS_NUM',
  FROM
    PISA_TPC_PAYCOMN
```

Because the source data has the characteristics of time series, it is necessary to use time series decision tree algorithm to classify the data. The initial input data set D of the time series decision tree and the number of candidate sequence pairs m, if D has $x, y(x) = c_j$, the node is a leaf node, and it is continuously selected x_1, x_2 from D and input into the candidate sequence pair set s until the number of candidate sequence pairs is m. The information gain and gain rate of each candidate data pair are calculated in turn, and the data with the largest gain rate is divided into sub-nodes, and then the decision tree is constructed recursively. The calculation formula of information gain rate is shown in Formula 1. [10]

$$\text{InfoGainRatio}(D, s) = \frac{Gain(D,S)}{H_S(D)} \tag{1}$$

5 Conclusion

This paper mainly introduces and analyzes the design scheme of applying data warehouse technology to financial information management system, hoping to improve the competitiveness of financial institutions in the market through data warehouse technology. However, due to the limited time and ability of the author, the system designed and implemented in this paper still has room for further improvement. At present, most of the data extracted come from reports, and the form is single, so it is necessary to expand the range of data source formats in the future.

References

[1] Song Jiamei. Applying Data Warehouse Technology to Establish Financial Information Management System. Electronic Technology & Software Engineering.2018.01.

[2] Luo Wei, Liu Gongzong. Research Status of Data Warehouse Based on Big Data. China New Technologies and Products.2020.09.

[3] Lei Bowen. Design and Implementation of Real-time Data Warehouse Based on Big Data. China University of Geosciences (Beijing).2021.05.

[4] Zhang Yanli, Wu Huaibei. Import and Application of Hive Data Warehouse in Hadoop Big Data Environment. Computer Programming Skills & Maintenance.2012.12.

[5] Tong Yixiao. Design and Implementation of Report System for Financial Supervision Data. Guizhou University.2021.08.

[6] Xia Guangwei, Wang Weijian, Zhou Yumei. Research on Metadata Based on Financial Data Warehouse. Journal of Changsha Social Work College.2021.09.

[7] Li Yan, Yang Chunming, He Yuan, Liu Bo. Research and Exploration on the Construction of Graph Data Warehouse in Financial Industry. Financial Technology Time.2022.08.

[8] Yang Rui, Zhang Rongrong. Design and Implementation of Financial Data Warehouse Model Based on Greenplum. Modern Information Technology.2020.11.

[9] Miao Yue. Research and Application of Financial Leasing Data Warehouse Based on Spark. Jiangsu University of Science and Technology.2021.03.

[10] Ren Zhongsheng. Data Mining Technology Based on Data Warehouse. Digital Technology and Application.2021.09.