Construction of Financial Big Data Analysis Platform in Universities under Big Data Technology

Yingmin Zhang[1], Hong Zhang[2]
* Yingmin Zhang: 10827@cqmpc.edu.cn, 176889578@qq.com

[1]Chongqing Medical and Pharmaceutical College, Chongqing 401331, China

[2]University of Shenyang, Shenyang 110044, China

Abstract—The high development of information technology has brought high-speed and convenient services to school teaching, scientific research, learning and life. At the same time, a large number of data recording the daily management and teaching work of universities has been generated, which constitute a huge big data resource with the accumulation of time. At present, the massive data of colleges and universities has not been effectively analyzed and utilized. To this end, this paper uses big data technology and Python for development. Through the establishment of an analysis model, this paper puts forward the construction scheme of a financial big data analysis platform from the aspects of construction ideas, overall application architecture, data management platform, data mining platform, technical solutions, etc.

Keywords-big data; information technology; colleges and universities

1 INTRODUCTION

The rapid development of technology has greatly improved the ability of information exchange, storage and processing. People and things are more closely connected, forming a huge ecosystem of everything interconnection. At the same time, the speed and scale of data generation have been subversive development, and the society has entered the digital age of "data everywhere" [1]. The high development of information technology has brought high-speed and convenient services to school teaching, scientific research, learning and life. At the same time, a large number of data recording the daily management and teaching work of universities has been generated, which constitute a huge big data resource with the accumulation of time. Although these resources are of high value, most colleges and universities have a long time span for information construction. The information construction of the whole school can not be planned uniformly, and big data is in a "data closed" and "data island" state [2]. Massive data has not been effectively analyzed and utilized. Although financial personnel can truly reflect and preliminarily analyze the financial situation and achievements of colleges and universities, their analysis depth, scope and application of results have strong limitations. In view of this, based on practice and theoretical learning, this paper put forward the construction scheme of university financial big data analysis platform, which provides reference for the construction of university financial intelligent finance.

ICICA 2022, December 02-04, Chongqing, People's Republic of China

DOI 10.4108/eai.2-12-2022.2328061

2 CONSTRUCTION GOAL

By building a financial big data analysis platform, the school integrates global data, presents and analyzes the financial operation of colleges and universities, and improve the financial analysis ability of the school. Through intelligent data analysis and all-round data model construction, it provides strong support for the school's business development, policy improvement, decision optimization, risk early warning, etc, at the same time, it helps the school to make scientific decisions and improve the financial management level of the school.

3 CONSTRUCTION IDEAS

Adopt advanced technology architecture to meet the present and the future. Centralized deployment of the analysis system, multi-level and multi-party collaborative application. Build mid-platform support capability and fully apply intelligent tools. Data-driven financial management, risk-oriented work planning. ROADS design concept to improve the user experience.

4 SYSTEM CONSTRUCTION SCHEME

Under the premise of not affecting the operation and security of the original financial business system, the school's financial big data analysis platform collects data from different sources in different systems, departments, regions and businesses, structured and unstructured data of different types through online collection, web crawler acquisition, intelligent tool identification and other methods. The platform conducts big data governance on the collected raw data to form a standard database table structure covering three layers of audit business domain, subject data set and business data table. The platform connects the data of each system horizontally, and realizes the integrated analysis and early warning of industry, finance, tax and capital.

4.1 Overall application framework

The financial big data analysis platform is divided into data management platform layer, data mining platform layer and data application platform layer. The data management platform layer is the data base support environment of the big data intelligent analysis platform, mainly including the source of the original business data of the data center. The data mining platform layer deeply excavates the data of the data management platform. Business logic design and functional design are carried out for different financial departments and leaders, and independent business models and analysis display of each business and unit are realized. The analysis and presentation layer mainly realizes the personalized management of the user interface through the functions of visual large screen customization, content customization, style customization, interface conversion, etc.

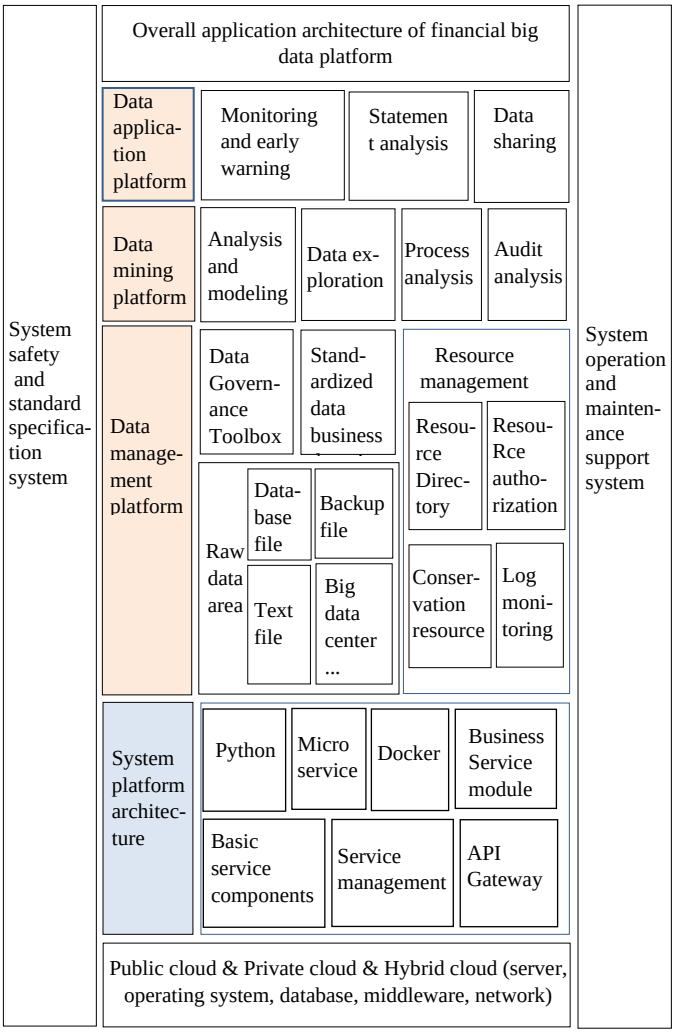

Figure 1. Overall framework of the system

4.2 Data management platform

First, the platform realizes the connection and unified management of the raw data sources of multiple business systems to be collected. Data was collected through file data import, multi-source heterogeneous data migration, network data acquisition, OCR image recognition, etc. The second, the platform converts the collected unstructured data and semi-structured data into structured data, and cleans them, etc. The platform conducts standardized processing of structured data, and establishes a financial analysis database based on the three-tier system of business domain, subject data set and standard data table. And then, the platform can conduct unified maintenance and user authorization of the data according to the business domain, subject set and data table; and form the school financial data assets directory.

4.3 Data mining platform

The platform connects accounting, capital, budget, contract, purchase, sales, inventory, production, cost and other financial business system data, realizes the integrated query analysis and intelligent application of industry and finance, and meets the normal supervision.

1)Data query: Through data governance, the platform can connect the financial and business data of colleges and universities, and realize the horizontal and vertical query and analysis of multiple departments, multiple accounts and multiple years. Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.

2)The data analysis.

a) Comprehensive financial analysis: Aiming at the application scenarios of big data financial analysis, the platform uses ratio analysis models and other models from different dimensions to establish analysis indicators for different business topics, such as budget execution rate index, performance index, solvency index and growth ability index.

b) Predictive data analysis: First, income was predicted using multiple linear regression models based on multiple factors. The multiple regression model is represented by the following formula:

$$Y = k_0 + k_1 x_1 + k_2 x_2 + k_3 x_3 + \ldots \tag{1}$$

In the formula, X_1, X_2, X_3, ... are different feature variables; K_1, K_2, K_3, ... are the coefficients before these feature variables; K_0 is a constant term.

Then we evaluated the model against R-Squared [3]. R-squared is calculated as follows:

$$R_{sq} = 1 - \frac{RSS}{TSS} \tag{2}$$

and

$$\cdot RSS = \sum (y_i - y_{fitted})^2 \tag{3}$$

and

$$TSS = \sum (y_i - y_{mean})^2 \cdot \tag{4}$$

In formula, Y_i is the actual value, Y_{fitted} is the predicted value and Y_{mean} is the mean of all scatter. For a linear regression model with a high degree of fit, the actual value should fall on the fitting curve as possible, namely that the RSS is as small as possible and the R-squared is as large as possible. When RSS tends to 0, it indicates that the actual value basically falls on the fitting curve, the fitting degree of which is very high, when R-squared tends to 1.

3)Data visualization analysis: The platform is equipped with visual modeling and analysis functions, and provides tables, pie charts, area charts and other basic chart design functions. The platform can intelligently map according to the characteristics of data to improve the efficiency of graphic modeling and analysis.

4.4 Risk warning system

According to the wizard prompt, the risk warning system can flexibly build the warning indicator model and send the warning results.

5 TECHNICAL PROPOSAL

5.1 Developed in Python

Python, as a high-level scripting language, is widely used in the field of artificial intelligence because it integrates interpretive, compilable and interactive languages. At the same time, Python has become one of the most commonly used programming languages in the computer field due to its open source, low threshold, strong portability and rich resource base [4]. As a cross-platform computer programming language, Python has become an industry standard for data analysis and processing.

- At present, big data analysis is characterized by complex architecture and high threshold of operation, maintenance and development. With its advantages of glue language, Python greatly reduces the operation and development threshold of big data architecture.

- Based on the support of rich visual controls on the big data analysis platform, it assists in data display and analysis. Combined with the big data visual display technology of the Web, it provides a variety of data display components, such as relationship charts and pie charts. Users can intuitively display relevant business data online in a unified way.

- Python can conduct in-depth analysis of data. It is the preferred programming language in the Python artificial intelligence era, supporting statistical analysis, machine learning, artificial intelligence and other applications.

5.2 The UML Unified Modeling Language

UML (Unified Modeling Language) is a well defined, easy to express, powerful and universally applicable modeling language [5]. UML is best used for data modeling, business modeling, object modeling, and component modeling.

5.3 Component technology

Component technology is the best and fastest developing software reuse technology so far. It completely solves the problems of reusability, poor adaptability and long cycle in software development.

5.4 Data stream processing technology

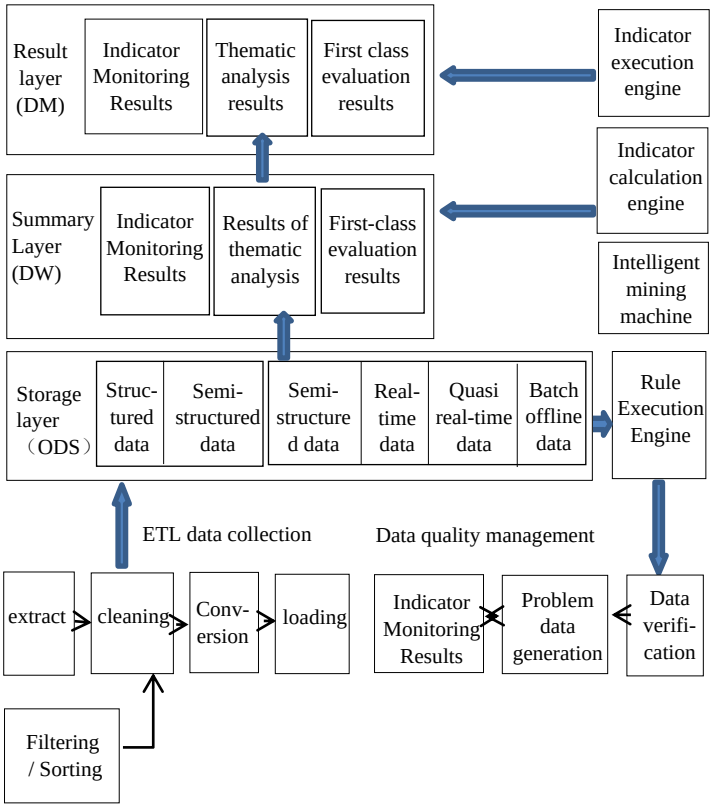

Figure 2. Data stream processing technology

The system data source extracts, cleans and loads the data through the data integration tool, synchronizes it to the data mart, and then obtains the data and validation rules through the data quality management, inputs them to the rule execution engine, and outputs the problems to the data source for rectification. From the ODS cache layer to the DW summary layer and DM result layer, the index execution engine, intelligent mining and machine learning are used for processing. Finally, the system displays applications and external systems through large screen display tools, intelligent analysis reports, external interfaces, and provides processing result output through data services, intermediate libraries, interface integration, etc.

5.5 System deployment mode

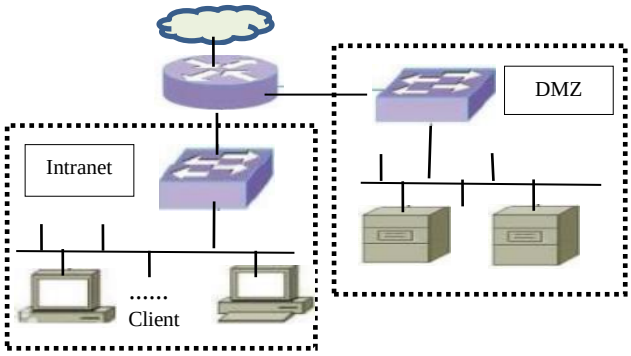

Figure 3. Deployment structure diagram of the system

5.6 Safety technology scheme

According to the Basic Requirements for Classified Protection of Information Security Technology Network Security (GB/T 22239-2019) and the expansion requirements of cloud computing, mobile Internet, the Internet of Things and industrial control systems, the construction, grading, protection, evaluation and filing of the system are carried out to ensure the security of key information infrastructure. According to Basic Requirements for Password Application in Information Systems (GM/T0054-2018), domestic password application is realized. Through the information security operation and early warning system, the centralized monitoring and security situation awareness of the system are carried out to achieve dynamic network security management and control that integrates monitoring, early warning, protection, detection, response and recovery.

6 BENEFIT ANALYSIS OF SYSTEM CONSTRUCTION

6.1 Economic benefits

Big data brings great technical challenges, but also great technological innovation and opportunities. The accumulated big data contains a lot of in-depth knowledge and value that is not available when the amount of data is small. Big data analysis and mining will realize a variety of high value-added services for universities, reduce costs and increase efficiency, and produce economic benefits.

6.2 Social benefits

The financial big data analysis platform fully embodies the development concepts of digital operation, intensive development, fine management and standardized construction, realizes the transformation of management from extensive to lean, enhances the core competitiveness of the school, sets up the leading image of the school in the industry, and generates social benefits.

7 CONCLUSION

Financial big data analysis platform is a necessary tool for the transformation from financial accounting to management accounting, which has good economic and social benefits. With the development of information technology, more and more universities will build financial big data analysis platform. Using big data technology and Python development tools, this paper proposes the construction scheme of financial big data analysis platform by establishing analysis model. This scheme provides a reference for universities to build a financial big data analysis platform, and also solves the problem that a large amount of data in colleges and universities cannot be effectively analyzed and used.

ACKNOWLEDGMENT

This research is supported by the Science and Technology Research Program of Chongqing Municipal Education Commission, China (Grant No.Kj202102858562385).

REFERENCES

[1] Then Hu. et al. Accounting big data analysis and processing technology: Boosting data to empower a new future of finance [J]. Finance and Accounting, 2022 (10): 33-38.

[2] Liu Songyue. Et al. Big Data and University Research [J]. Chongqing Higher Education Research, 2022 (10): 7-19.

[3] Zhang Min.et al. [Big Data Financial Analytics - Python based], China Renmin University Publishers, 303-305 (2022).

[4] Zhang Hang. Discussion on Application Technology of Python Computer Software [J]. Computer knowledge and technology, 2020 (16):96-97.

[5] Xu Yongjia. et al. Designing Relational Database with Unified Modeling Language (UML) [J]. Journal of Kunming University of Science and Technology. 2001(26):61-64.