The Analysis and Design of Data Sharing Platform for Universities Based on Blockchain Technology

Shengnan Wang[1,a], Xiyan Lv[2,b]

[1]School of the Economics and Management, Beijing Jiaotong University, Beijing 100044, China

[2]School of the Economics and Management, Beijing Jiaotong University, Beijing 100044, China

[a]17125474@bjtu.edu.cn, [b]Lvxiyan@bjtu.edu.cn

Abstract: With the developing of information technology, a large proportion of universities have accumulated lots of data, and individual data centers have been built. However, the data standards are inconsistent and formed lots of information islands objectively. So, the problem of data integration and data sharing need to be solved urgently. In this study, a data sharing platform for universities combining the concept of data port and blockchain technology has been proposed. The feasibility and efficiency of the platform is demonstrated and analyzed by special examples.

Keywords: Blockchain, Data sharing platform, Data port.

1. Introduction

With the developing of information technology, the construction of informatization in universities is on the rise. The emergence of such technologies, for instance, big data, artificial intelligence, Web of things and cloud computing offer new opportunities for universities so that digital campus constructions are able to make new achievements. A large number of business information systems such as OA, personnel systems, teaching and administrative systems, asset systems and card systems which improve the efficiency of staff [1] have been built in more and more universities. Therefore, the interconnection between systems obtained by data sharing platform is urgently needed. Numerous literatures involved the construction of data sharing platform for universities, but most of them are essentially centralized based on the data warehouse. If judging only by appears, from a technical point of view, it seems that the phenomenon of "information islands" has been solved, but from a practicall point of view, lots of challenges in operation arising. The inconsistent data standards bring challenges to the data integration and data sharing. Therefore, a data sharing platform for universities combining the concept of data port and blockchain technology that provides reference value for the data sharing in universities has been proposed in this study.

2. Overview of Data Sharing in Universities

Universities have achieved fruitful results in the early stage of information construction. For instance, financial systems and teaching systems which benefit utilization of students and also contribute to the improvement of teachers' work efficiency. Since the construction of technology is immature, lacking of uniform standards and the operation of different system independently, so that information islands are formed gradually. In order to maximize the value of the data, majority of universities have carried out the construction of data sharing center which could realize the interconnection and interoperability of various systems.

2.1 Characteristics of the data

As we all know, data is distributed in various colleges in the university. The data shows feature of dispersion, heterogeneity, high autonomy and low share ability.

2.1.1 Dispersion

Majority of school-level data produced from the college-level in accordance with the hierarchical relationship. Therefore, information islands are gradually formed, the data shows a characteristic of dispersion.

2.1.2 Heterogeneity

Heterogeneity often manifests itself in two aspects: system heterogeneity and pattern heterogeneity. The so-called system heterogeneity which is composed to identify the difference between application systems and database systems; pattern heterogeneity is the heterogeneity of storage schemas[2]. Each college develops and purchases business system and data center respectively according to its own business process respectively, this leading to the existence of heterogeneous systems. Moreover the form of data storage in various systems are differ, such as DAS direct additional storage, NAS network additional storage etc.

2.1.3 High autonomy

In universities, individual colleges create their own business system based on business processes. And specialists are in charge of managing their own data centers for the usage of all departments within the college. Therefore, there are high degree of autonomy existing.

2.1.4 Low share ability

Generally speaking, the share rate of data among colleges is extremely low due to lacking of data and information exchange between colleges. However, data exchange and sharing exist between school and colleges.

2.2 The Present Situation of Data Sharing in Universities

In order to solve the problem of information islands, data sharing centers have been set up in majority of universities. Tongji University has built a data warehouse platform which stores a large amount of historical data after extraction, conversion and loading (ETL), and used OLAP, data mining and other techniques, so that they could analyze and process to support decision-making[3]. Shanghai Ocean University has established a "comprehensive service platform" which integrates the data scattered in different information systems into a common database and realize the sharing of data and information[4]. Most universities are using data warehouse-sharing mechanism which can integrate data that originally scattered in different systems. It can be likened to bringing small streams of data into a huge data lake.

2.3 Problems in data Sharing

The method above solves the problem of data sharing technically and structurally, but brings challenges to managers during operation as well. The process of data sharing is very chaotic.

2.3.1 Highlight the problem of data responsibilities

Most schools are using the form of centralized data management for data sharing. Data is unified and managed in school and the colleges can get data from school if necessary. Due to the large amount of data, in the management process data loss and redundancy might occur. If errors or duplication of the data were found, it's hard to determine which part of the transmission process was went wrong. Therefore the problem of data responsibilities needs to be solved urgently.

2.3.2 Increase the difficulty of management and maintenance

Data sharing center is established without taking into account the linkages and constraints between different systems, and data is concentrated blindly. The data center gathers a lot of data in a short period of time. As each system uses different software and different data standards, platform managers require to know both database management knowledge and business knowledge. So it not only poses a challenge to the operation personnel but also increases the difficulty of operation and maintenance.

2.3.3 Difficulties of data quality guarantee

The school data sharing center is available for making statistics by school and exchanging information between school and colleges. However proper monitoring and recording of the data flow still missing. In the process of continuous use of data, the problem of data tampering, data loss and others are easy to occur. So data quality is difficult to guarantee.

3. Blockchain - based University Data Sharing Platform

In view of the shortage of the above sharing mechanism, in this study, a data sharing platform in university based on the blockchain technology has been proposed. It adopts the decentralized

architecture mode and bring in the concepts of data certificate and data port. The architecture plays the features of decentralization, autonomy, Security of blockchain to achieve efficent sharing of the data.

3.1 Introduction of Blockchain technology

In 2008, the blockchain technology is first mentioned in Satoshi Nakamoto's paper 《Bitcoin: A Peer-to-Peer Electronic Cash System》. In this paper, blockchain which is the basic technology for building the bitcoin data structures and encrypting transmission of transaction information has been adressed[5]. The technology mainly uses the decentralized approach and the principle of cryptography to ensure the safety and reliability of the data.

Blockchain technology has the characteristics of decentralization, openness and transparency, strong autonomy and no tampering. It has attracted many scholars to join in the research of this technology. Domestic and foreign researchers have shown that blockchain technology can not only be applied to the financial sector, but also can be applied to other areas to solve security and other issues. For example, blockchain technology in the Internet insurance industry can protect privacy information of customers effectively and reduce the risk of information asymmetry, meanwhile reduce the cost of Internet insurance[6]. Blockchain technology can provide highly reliable platforms and secure shared account books for the interaction between people and machines, machines and machines in the web of Things[7].

3.2 Construction of Data Sharing Platform in universities Based on Blockchain

Aiming at the heterogeneity of business systems of each college, as shown in Fig.1, a data sharing platform is established. Data sharing platform is divided into three parts, including the college level, blockchain service center, school level.
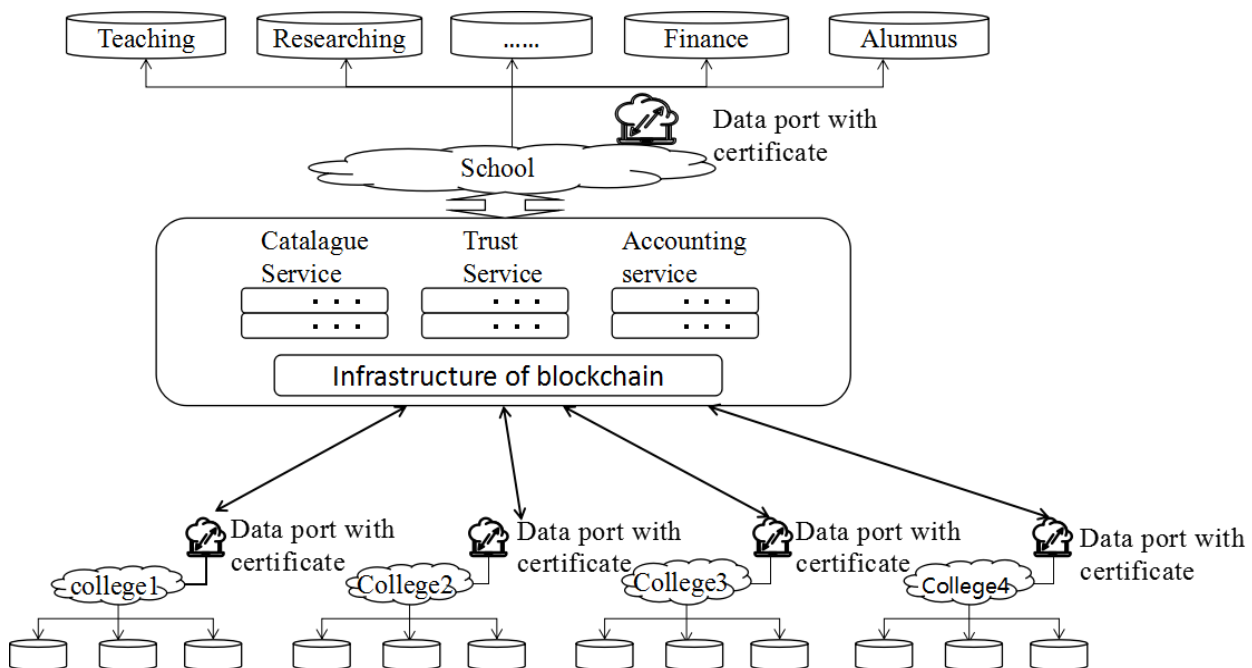

Fig. 1 Blockchain - based data sharing platform for a university

3.2.1 The college level

In Figure 1, small cylinder below the college represents different business systems, and each college has its own different business systems. In a decentralized approach, each college establishes a credentialized data port. Data port is a small database used to store data. The credential data port means a credential management mechanism is set up in the data port. Each college data center will put the data which can be shared into the data port to facilitate the sharing of data.

3.2.2 Blockchain service center

Building blockchain service infrastructures based on Internet technologies, including catalogue service, trust service, accounting service.

a) Catalogue service

The catalogue service is mainly used to show the available data content clearly. The data provider uploads the data catalogue to the catalogue service in the clouds. And the catalogue service automatically partitions the data according to the college to which the data belongs. Each college forms a block so that the data user can find the data desired efficiently.

b) Trust service

Trust services are aimed to ensure the security of data and the censorship of data users during data sharing. In each data transaction process, trust service system perform two tasks. One is to examine the certificate; the other is to encrypt the data according to asymmetric encryption technology of blockchain to ensure the security and untamperability of the data.

c) Accounting service

Accounting services are used to record transactions. According to the nounce value and hash value validation of blockchain, it can form transaction timestamp which is used to display the time generated by the block. The time will accurate to seconds accordance with the standards of the Unix era[8]. Accounting services realizes the traceability and verifiability of the data.

3.2.3 School level

In Figure 1, the small rectangular column above the school refers to the major business systems established by the school in accordance with its own business rules. Like colleges, the school also establishes a credential data port for the review and transmission of the data. Colleges can also get the data from school according to their needs.

3.3 The Process of data Sharing

Data sharing process is divided into two blocks, one is the data publishing process of data provider; the other is the process from the data user send out share demand to the data usage. Data sharing process diagram shown in Fig. 2.
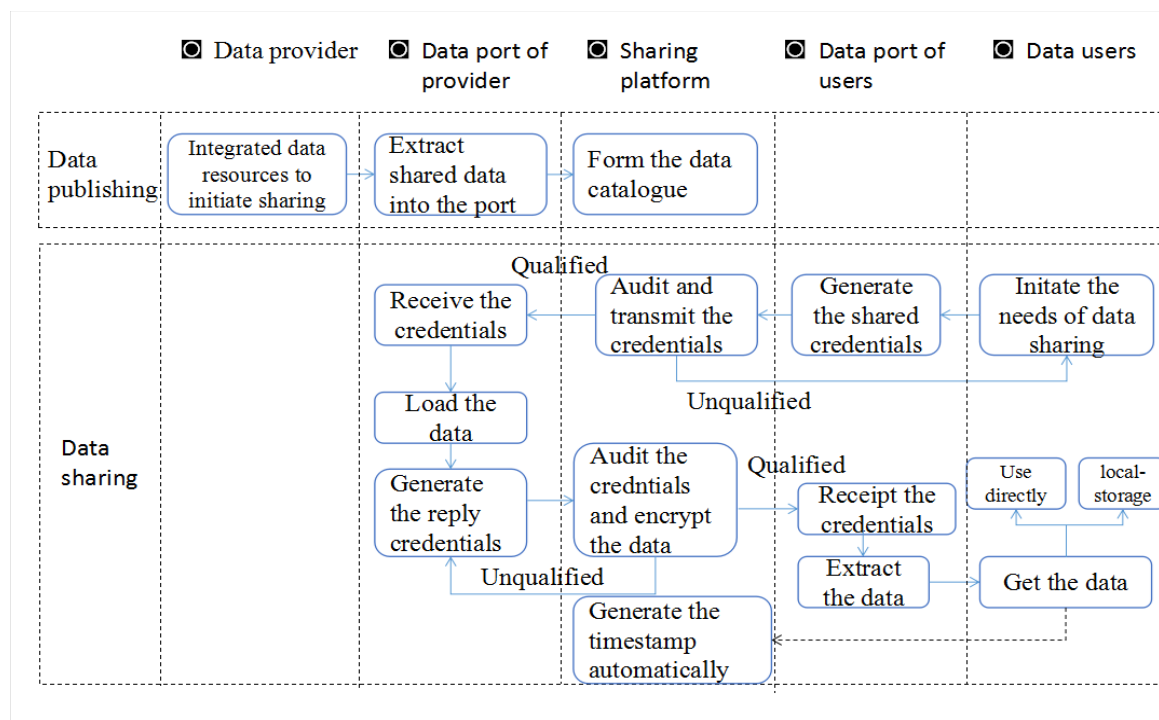

Fig. 2 The process of data sharing

3.3.1 Data publishing process

The data provider first integrates the data resources that can be shared. Then the provider puts the data into the data port by extracting and converting (ETL). Finally, provider uploads the data content to the cloud data catalogue service so that the data user can see the shared data information.

3.3.2 Data usage process

When data users want to use the data, firstly, users send out the data sharing requirement, and the

data consumer data port retrieves the data catalogue and generates the shared certificate according to the demand and then sends it to the data sharing platform for review. If the certificate is unqualified, it sent back to the data consumer; if it is qualified, the certificate will be sent to the data provider data port. Data port receipt certificate and pack the data, and then the generated response certificate carrying data packets will be sent to the data sharing platform. The shared platform uses the asymmetric encryption technology of blockchain to encrypt the data to prevent tampering, and then verify the certificate. If the certificate is unqualified, it returns to data provider. When the certificate is passed, the certificate and the data packet are transmitted to the data port of data user. Data users can store the data either locally or use directly according to the needs. After the data consumer obtains the data, it means that a data transaction is completed. The sharing platform generates a time stamp based on the blockchain technology, and at the same time records transaction time and transaction information. In the described above, a cycle of data sharing is completed.

3.4 Characteristics

The data sharing scheme in universities described in this article has subverted the previous centralized data sharing model. It adopts a decentralized distributed architecture and uses blockchain technology. The implementation of certificate management makes that there is the trust in a shared exchange and easy to manage with a low cost. It can effectively alleviate the current problems such as college data sharing management difficulties, data security which is difficult to guarantee and etc.

3.4.1 Decentralized

This model forms a distributed database management via the establishment of data port with decentralized features. It realizes the data storages at home and services exist in the cloud. To some extent, it improves the autonomy of each business system and solves the integration problems of multi-source heterogeneous systems in universities. Meanwhile, it relieves the pressure of database managers and make the platform easy to operate.

3.4.2 Credential sharing

Based on the blockchain, the university data sharing platform implements credential management to control the sharing process strictly. Using asymmetric encryption technology of blockchain to ensure that data can not be tampered with during data transmission. One data comes from one source makes the responsibilities clearer so that ensure the data's quality. This method effectively solves the current situation that the data flow disorderly in universities and the problems which can not be known. With the assistance of certificates, the data providers offer data consumers a clearer understanding of the blood relationship of the data.

3.4.3 Low cost and easy to construct

The biggest advantage of the data sharing platform is convenient and less time consuming. Moreover, the catalogue service makes data to be found easily. The chained shared process standardizes the data sharing and improves sharing efficiency. Decentralized mode reduces investment of managements so that saving capital investment. The deployment patterns of service cloud greatly reduce the difficulties of deployment and capital investment of the program.

4. Conclusions

The construction of digital campus is in full swing, while the demand to the quality of data sharing is getting higher and higher. The primary purpose of data sharing is to change the status quo of information islands and maximize the value of data assets. So the construction of data sharing platform in universities not only meets the demand of business personnel, but also ensures the safety and efficiency of data sharing. At present, there are problems with data sharing in universities which is difficult to manage and the responsibilities to the data are not clear. Based on the characteristics of data in universities and weakness of the existing data sharing platform. In this study, a college data sharing platform based on blockchain which has the characteristics of decentralization, Credential sharing and easy to construct has been proposed. It realizes the data storages at home and services exist in the cloud. The certificate carries data that brings a new idea to data sharing in universities. Blockchain-based data sharing technology is able to be applied to other areas in the future to solve

data sharing problems.

5. Acknowledgments

6. References

[1]Jingqi Li, Yijie Bian, Guobao Zhang. Design of data sharing platform in universities based on business process [J]. China Education Information, 2016 (11): 93-96.

[2]Hanlai Wan. Realization of data exchange and sharing platform in digital campus of universities [J]. Science and Technology of Communications, 2016,8 (11):72-73.

[3]Hao Zhang, Bo Liu, Henghua Zou, Weisheng Xu. Exploration and practice of university Data Governance [J]. Electronic Technology and Software Engineering, 2017 (21): 181-183.

[4]Yongjun Li, Lin Peng, Cheng Lin, Qiuyu Lu, Yuefang Xia. Research on big data governance in college information management [J]. Journal of Management Informaiton in China, 2016, 19 (03): 185-187.

[5]NAKAMOTO S. Bitcoin. A peer-to-peer electronic cash system [R / OL]. [2017-04-13]. https: //bitcoin.org/bitcoin.pdf.

[6]Dawei Zhao. Discussion on the application of blockchain technology in Internet insurance industry [J]. Journal of Financial Development Research, 2016 (12): 35-38.

[7] BAHGA A, MADISETTI V K. Blockchain platform for industrial internet of things [J]. Journal of Software Engineering and Applications, 2016 (9): 533-546.

[8]Huiqin Yang, Lei Sun, Xichao Zhao. Building a mutual trust and win-win supply chain information platform based on blockchain technology [J / OL]. Science and technology progress and decision-making: 1-11 [2018-01-28].