TEST Engineering & Management

July - August 2020
ISSN: 0193-4120 Page No. 5471 - 5478

Research on Fragmented Data Analysis and Integration Based on Computer Software

Zhiqiang Ma[1,*]

[1]School of Information Engineering, Henan Vocational College of Agriculture, Zhengzhou, Hean, China, 451450

Article Info
Volume 83
Page Number: 5471 - 5478
Publication Issue:
July - August 2020

Article History
Article Received: *25 April 2020*
Revised: *29 May 2020*
Accepted: *20 June 2020*
Publication: *28 August 2020*

Abstract

The era of big data, as an era in which people re-understand the world and change the way information is obtained and processed, while providing convenience for people, it also brings some problems. In order to solve the fragmentation of system information and data, fragmented data integration is proposed. Relying on computer software, using data sharing for data integration analysis, hoping to bring reference for efficient fragmented data integration.

Keywords: *Computer Software, Data Fragmentation, Data Interaction, Data Sharing;*

1. Introduction

In the era of network information, as an era in which people can obtain information and processing methods, it also brings various problems when it provides various conveniences[1-3]. For the education industry, the advent of the network information age is not only an opportunity for the future, but also a challenge to face difficulties[4-6]. Therefore, in order to compensate for the phenomenon of data fragmentation in the Internet age, a processing model of fragmented information integration for this phenomenon is proposed.

2. Main problems

With the development of information technology, knowledge, information and data are showing a trend of fragmentation. The medium of university office and teaching has gradually shifted from paper medium to electronic medium, and the information system has penetrated into every aspect of campus. Outsourcing is the most common form adopted by various colleges and universities. Due to many differences in system brands, operating platforms, business logic, data storage, etc., the more the number of systems and the longer the use period, the more likely it is that there will be differences in shared data between the systems. Large, resulting in fragmented data. The above problems are mainly manifested in the following aspects:

2.1. Poor data interaction

Since each business system comes from different construction units, the system structure and data standards are not uniform, the data cannot be directly used by other systems, and the degree of automation is low; cross data between multiple applications is usually manually exported, imported or operated by multiple systems at the same time The method achieves the goal of basically consistent data (as shown in Figure 1). The data is related but the system does not interact, which ultimately leads to the production of "data islands".

Published by: The Mattingley Publishing Co., Inc.

July - August 2020
ISSN: 0193-4120 Page No. 5471 - 5478

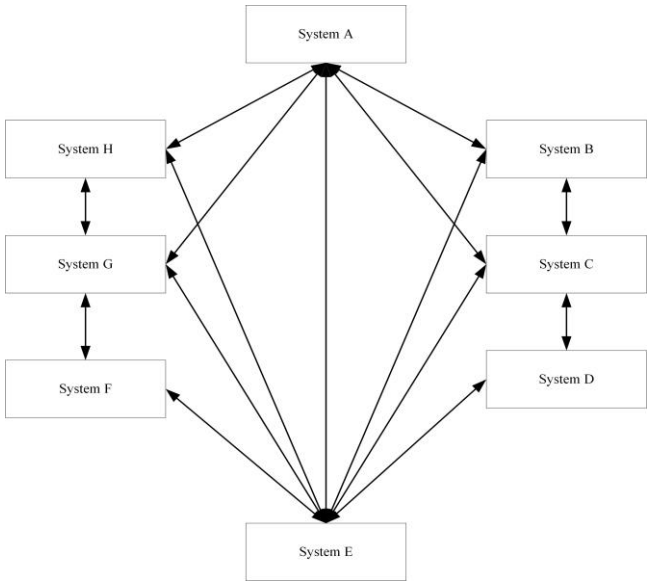

Figure 1. The chaos of the docking of various business systems in the traditional mode.

2.2. Low information utilization

In the process of digital campus construction, universities lack top-level design. The independence and pertinence of many business systems are more prominent, and the data of different businesses of the same entity is scattered in the business system. Affected by the lack of interaction between systems, information utilization and decision support capabilities are low, and the value of data has not been well reflected.

3. The mode of fragmented data integration in universities

Data integration means planning, sorting, and storing scattered, single, and unplanned data according to criteria such as importance and priority. At the data level, it is easy to use and share, which can better reflect its value. Combining the above existing problems, it is necessary to realize the construction of digital campus in colleges and universities, rationally integrate fragmented data and application resources, optimize business processes, complete work process reconstruction, and realize data standardization and business process standardization. At the same time, in-depth data mining and analysis are carried out to meet the different needs of various departments of the university, and further improve the school's management efficiency and decision-making support capabilities.

From the analysis of the technical route of data storage, exchange and sharing, the data integration mode can be roughly divided into the following three types: data centralized mode, distributed resource mode, and data sharing platform mode.

The above three modes have big differences in implementation complexity, input cost and operational performance.

The data concentration model is limited by the high cost of outsourcing application system customization requirements, and the adoption rate is not high, so in-depth analysis is not performed.

Test methods and steps: (1) Server A simulates an application server and carries all data; Server B provides distributed resource indexing services and data sharing platform services, where the data sharing platform service configures 40 fields in the data of server A; Server C Simulate the access side, use wget to get the two scripts on server B. The scripts respectively request the field resources inside and outside the central database on server A through the distributed resource indexing service and data sharing platform service, and then calculate and output to the decimal point. 3 bit running time. (2) Server C accesses the distributed resource indexing service three times, restarts all servers after recording the results, accesses the data sharing platform service three times and records the results.

The test results are shown in Table 1. It can be seen that when requesting services from the data sharing platform, the fields configured in the central database are slightly behind the distributed resource indexing service due to the forwarding and storage during the first access, and the later efficiency is much higher. Distributed resource index mode; there is not much difference in efficiency between the two fields for non-central storage.

Published by: The Mattingley Publishing Co., Inc.

July - August 2020
ISSN: 0193-4120 Page No. 5471 - 5478

Table 1. Performance test results of distributed resource mode and data sharing platform mode.

Test serial number	Distributed resource model service	Data sharing platform mode service	
	Test time for all fields of the target table/s	The target table is transferred to the central library field test time/s	The target table is transferred to the central library field test time/s
1	10.0209	11.4313	9.9967
2	10.0758	3.7744	10.0858
3	10.0664	3.4775	10.0233

The above test is based on an experimental environment with a unified character encoding and a unified database environment. If the diverse database environment, character set, field size and other factors of the school application system are taken into account, the advantages of the data sharing platform model will be more prominent.

In summary, this research will take the construction of a data sharing platform model as an example to study the implementation path of fragmented data integration in universities and demonstrate the application results.

4. The technical route to realize the data sharing platform model

Integrating fragmented data and realizing data interaction and sharing is an effective solution to realize application integration and data sharing. This section will focus on the unification of data standards, the aggregation and centralized storage of shared data, and data The exchange and sharing of the three parts to realize the data sharing platform mode, its technical route model is shown in Figure 2.

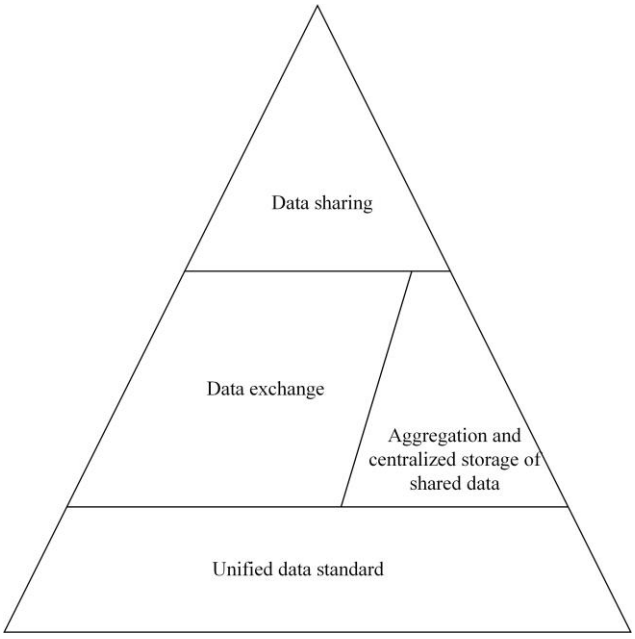

Figure 2. Data sharing platform model technical route model.

4.1. Unification of data standards

Establishing a unified data standard is a prerequisite for achieving fragmented data integration. A data standard is a rule that regulates attributes such as data composition, type, and length. It has a great impact on data collection, processing, and exchange. It is the foundation for data sharing and system integration. At present, most colleges and universities have basically completed the construction of digital campus business systems, and are gradually turning to the construction of system collaboration and resource sharing. The goal is to integrate all business systems and realize data exchange and sharing.

As a basic condition for data interaction and sharing, a unified data standard can ensure the

Published by: The Mattingley Publishing Co., Inc.

July - August 2020
ISSN: 0193-4120 Page No. 5471 - 5478

standardized classification and storage of data in the collection, processing, exchange and transmission process, and maximize the comprehensive benefits of data resources. The unification of data standards starts from the following key points:

(1) Standardized data standards

The establishment of the data standard system is guided by user needs, and follows the relevant national and provincial (autonomous region) standards such as the "University Management Information Standards" and "CELTS-33 Higher School Management Information Standards", combined with the actual situation of the universities themselves. Improve and determine a series of unified, standardized, and flexible data standard systems on campus, such as information coding, information systems, information management, and information security, in line with the characteristics of universities.

(2) Compile data standards

A complete set of data standards includes basic data element standards, data exchange format standards, information classification and coding standards, and data authority standards. The compilation of data standards should be based on reality and have scalability.

4.2. Aggregation and centralized storage of shared data

Big data is characterized by its wide range of sources and complex types. Shared data refers to data used by multiple or all business systems, such as user accounts and basic information. After standardizing and sorting out the data of each business system in accordance with the data standards, in accordance with the importance of the data in the business logic, the degree of coupling, the degree of activity and other conditions, while considering the characteristics of data continuity and traceability, the following three categories are established Storage space: create basic database, create business shared database, create historical database. Shared data is an active element in the mass data. It is classified into storage, centralized management, unified source and content version, which is conducive to the circulation and sharing of different systems.

4.3. Data exchange and sharing

Data exchange is a means of data sharing, and data sharing is the purpose of data exchange. According to the structure and characteristics of different business systems, exchange means are selected to complete data comparison, transfer, iteration and storage, and achieve the goal of data closed loop.

(1) Data exchange

Data exchange refers to the transmission and sharing of data from multiple business subsystems through certain middleware, program interfaces or services, to improve data utilization, and to ensure data connectivity between heterogeneous systems(As shown in Figure 3).

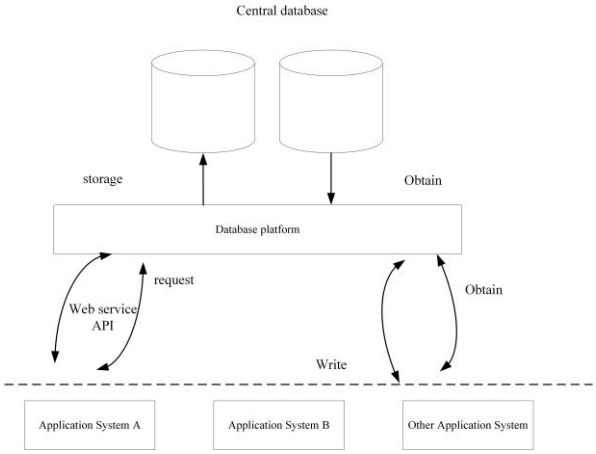

Figure 3. Data exchange process.

However, the danger of using this method cannot be ignored. Multiple applications reading and writing a data table at the same time are prone to serious consequences such as data conflict, error, loss, and even data damage.

2) Create an intermediate table of data.

In practice, it is more common to use an intermediate table to complete data exchange, because it can take into account the different needs of data while keeping the application system structure unchanged.

3) Development program interface.

Data exchange is the support for the information-based business system to realize data

Published by: The Mattingley Publishing Co., Inc.

July - August 2020
ISSN: 0193-4120 Page No. 5471 - 5478

resource interaction. Through data exchange, heterogeneous system integration is realized. In practical applications, universities need to choose the data exchange method according to the structure and function of the business system. Under normal circumstances, interface docking is preferred, followed by the creation of data intermediate databases and tables, and direct access to business databases should be the final alternative.

(2) Data sharing

Data sharing is the process of communicating necessary data between various business systems and external parties through the exchange of data. It carries the important task of data circulation in many business systems in the school, and realizes data circulation and sharing between different applications through data exchange.

There are three common activation methods for data exchange behavior: periodic events, as shown in Figure 4, passive events, and manual events.

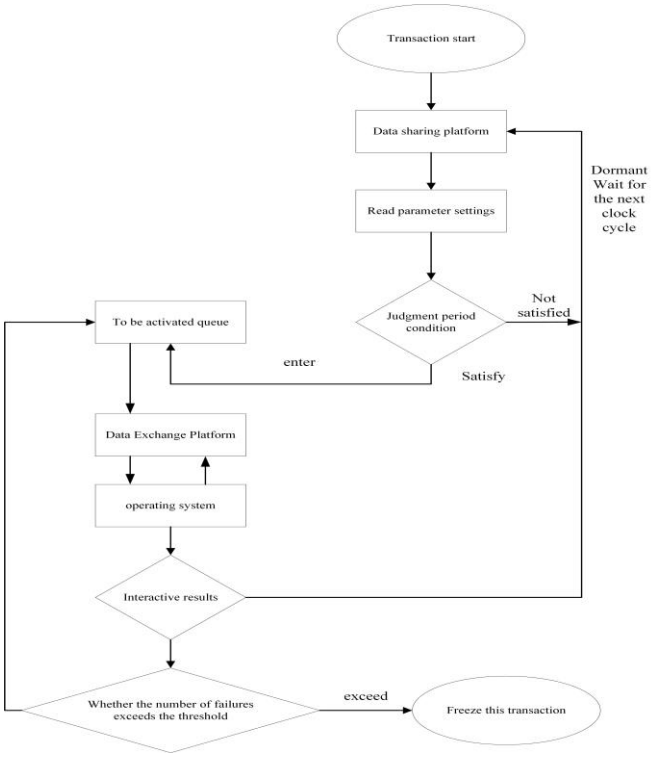

Figure 4. Implementation of the heartbeat behavior of the data sharing platform.

By building a data-oriented data sharing platform in colleges and universities that encapsulates data conversion, exchange, storage, management and other functions (as shown in Figure 5), the sharing of heterogeneous system data information can be achieved, and the data that is common in the construction of digital campuses can be solved to the greatest extent Poor interaction, low utilization, and weak security.

Published by: The Mattingley Publishing Co., Inc.

July - August 2020
ISSN: 0193-4120 Page No. 5471 - 5478

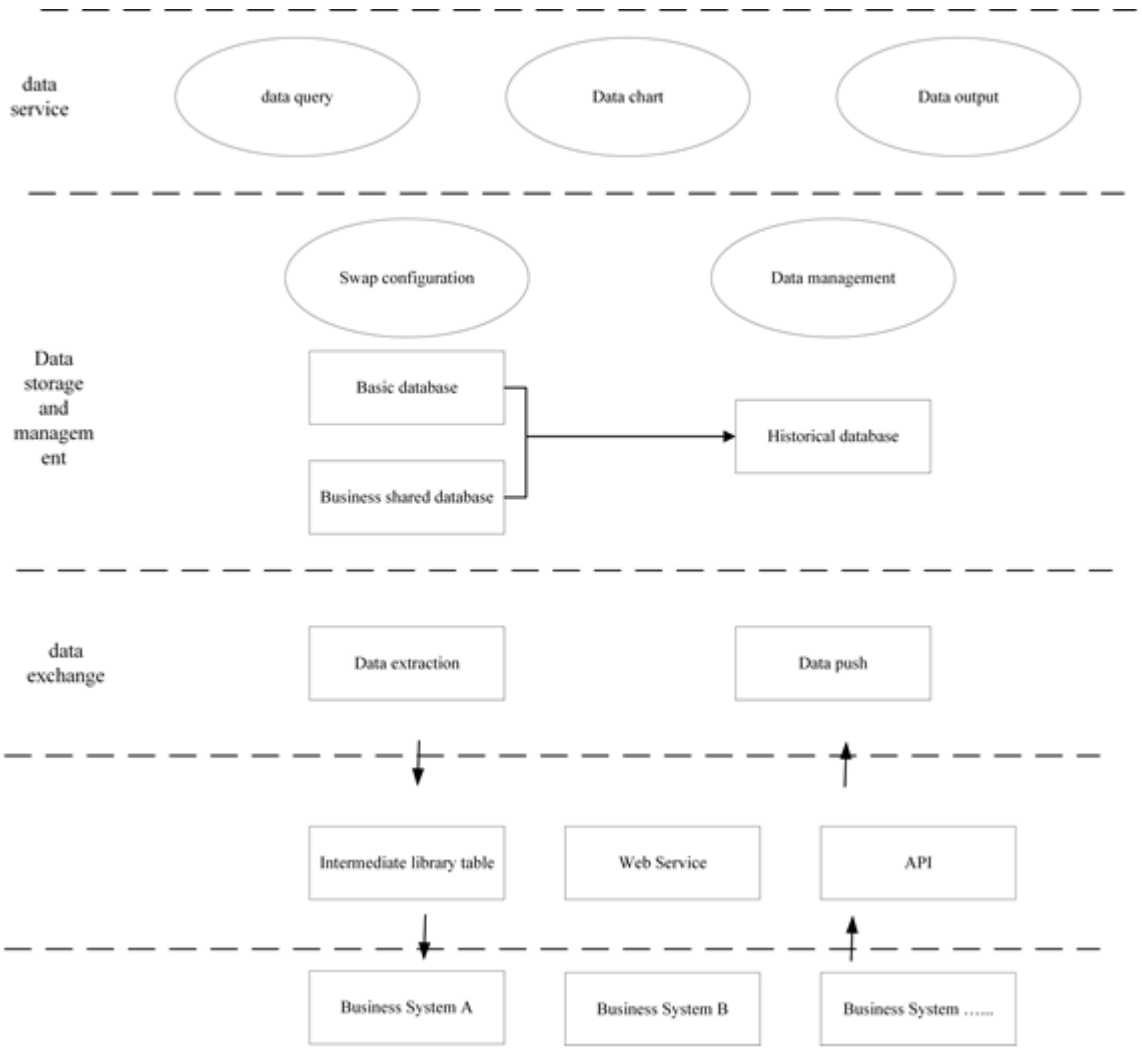

Figure 5. Conceptual diagram of data sharing platform.

The realization of data sharing in universities is an effective means to improve the utilization of data resources and avoid repeated waste in data collection, storage and management. It can realize resource allocation more rationally, improve the top-level design of digital campuses, save construction costs, and continuously promote university informatization Construction develops in depth.

5. Experimental analysis and effect

According to the technical route of the data sharing platform model, firstly, the statistics and analysis of the various business system data in the school are carried out, and a series of documents have been promulgated and implemented according to the needs of each user department and combined with the information construction of the school to standardize the school administration The coding standards for teaching and scientific research have formed a data standard system that conforms to the characteristics of the school. This system is now fully applied in the practice of integration of existing business systems, new business systems and data sharing platforms.

Secondly, in the process of business system database analysis, the data is unified into basic categories and business sharing categories and stored in the central database to provide sufficient basis for subsequent exchange and sharing; finally, the school uses the data sharing platform model as a guide and adopts independent research and development The data sharing platform and the two

Published by: The Mattingley Publishing Co., Inc.

July - August 2020
ISSN: 0193-4120 Page No. 5471 - 5478

lines of the business system reconstruction are parallel as a solution, which realizes the closure and circulation of the data loop in the school.

Let's take the entry of school faculty and staff as an example to illustrate the application effect of data integration in solving actual business problems.

The faculty and staff under the traditional mode need to go back and forth between multiple business departments, and the management personnel of each system are entered into the business system separately, and the procedures are quite cumbersome. At the same time, due to the high degree of manual intervention, the probability of data differences and errors is also greater.

After the data sharing platform intervenes, newly recruited faculty and staff only need to report to the personnel department, and the manager will enter the basic information such as personnel number, name, ID number, etc. into the personnel management system, and then they can take up their posts. The data sharing platform automatically dispatches basic personnel information according to the established period according to the exchange methods and parameters configured by the management personnel of the information management department, obtains and judges the changed data entries and fields, and stores them in the central database after conversion according to the school coding rules. In the same way, other business systems such as the educational administration system and the academic system are also automatically dispatched from the central library to the application by the data sharing platform. In this way, a complete data exchange and sharing process of basic information of faculty and staff is completed. Data pull and push do not require manual intervention at all, and the total time is no more than 10 seconds.

In summary, after integrating the data of each business system, relying on the data sharing platform, the faculty information is automatically distributed and synchronized to each business system to realize automatic data flow.

Through the construction of a data sharing platform, a college has completed the batch standardized conversion of different types of data to ensure the safe interaction and sharing of data between heterogeneous systems. At the same time, the third-party reporting system is used to provide more accurate and detailed information summary and analysis for the decision-making layer, and to achieve structured and informed information management.

6. Conclusion

With the continuous development of information technology, the degree of network intelligence has become an important indicator for judging the level of universities, and data fragmentation is a common problem faced by the construction of digital campuses in universities. The integration of fragmented data is an effective way to eliminate the island effect, improve the construction of digital campus, and realize the interactive sharing of data. Through the fragmented integration and integration of university data resources, improve data user management and authority management, optimize business processes, promote the development of university electronic school affairs, realize education information and management standardization, and provide scientific and accurate data basis for university leaders to make decisions.

Acknowledgements

Department of education of Henan Province, Application and research of teaching diagnosis and improvement based on information system fragmentation in the construction of "Double Universities"(No. 2019SJGLX704).

References

[1] Mcdonald A D, Lee J D, Aksan N S, et al. The language of driving: advantages and applications of symbolic data reduction for analysis of naturalistic driving data [J]. Transportation Research Record, 2018, 2392(1): 22-30.

[2] Hetz G, Datta-Gupta A. Integration of time-lapse seismic and production data:

July - August 2020
ISSN: 0193-4120 Page No. 5471 - 5478

analysis of spatial resolution [J]. Transport in Porous Media, 2020(2): 1-10.

[3] Kemper V G, De M F, Emmerling T C, et al. High resolution data analysis strategies for mesoscale human functional MRI at 7 and 9.4T [J]. Neuroimage, 2018, 164(5): 48-58.

[4] Tian X, Zhang G, Shao Y, et al. Towards enhanced metabolomic data analysis of mass spectrometry image: Multivariate Curve Resolution and Machine Learning [J]. Analytica Chimica Acta, 2018, 5(2): 211-219.

[5] Exner T E, Dokler J, Bachler D, et al. OpenRiskNet, an open e-infrastructure to support data sharing, knowledge integration and in silico analysis and modelling in risk assessment [J]. Toxicology Letters, 2018, 3(2): 102-110.

[6] Dias L S, Ierapetritou M G. Integration of planning, scheduling and control problems using data-driven feasibility analysis and surrogate models [J]. Computers & Chemical Engineering, 2020, 134(4): 1-17

Published by: The Mattingley Publishing Co., Inc.