



# Research on the Construction and Application of University Data Middle-End Under the Background of Education Digital Transformation

Beibei Wang<sup>(✉)</sup>

Shandong Institute of Commerce and Technology, Jinan, People's Republic of China  
57182339@qq.com

**Abstract.** Under the background of the current digital transformation of education, the data governance system with the main goal of improving the value of data should run through the whole process of education informatization construction. Based on the in-depth analysis of the pain points of data management in the current smart campus environment of universities and the problems existing in traditional shared data center, this paper puts forward the architecture system of university data middle-end, which realizes the closed-loop management of the whole life cycle of data. And illustrate the service improvement obtained after the implementation of the data middle office through specific cases. It is hoped that through the author's combing, it can provide certain theoretical references and method references for other universities in the construction of data middle-end.

**Keywords:** digital transformation of education · data middle-end · smart campus · data governance

## 1 Introduction

With the continuous development of smart campus, university data is increasingly diversified. In the 2022 National Education Work Conference and annual work points report, the Ministry of Education proposed the “Implementation of Education Digitalization Strategic Action” to improve data governance capabilities, strengthen data mining and analysis, and build a new model of data-based education governance [1]. Under the background of the digital transformation of education, the focus of informatization work in universities has gradually changed from daily management to data services, and the demand for data of various business departments has changed from simple business management to data analysis and decision-making judgment, from internal management of a single department to cross-departmental process services. The open sharing and application needs of teachers and students for data are becoming more and more prominent, and data privacy and security are becoming more and more important.

The traditional data warehouses and data sharing center can no longer meet the current needs of universities for in-depth data mining, analysis and utilization in terms

of data volume, data quality and supply mode, which seriously hinders and restricts the promotion of smart campus. Aiming at the pain points of data management in the smart campus environment of universities, this paper puts forward the architecture system of university data middle-end, which solves the closed-loop of data collection, cleaning, governance, storage, use, management and other links, and promotes the continuous improvement of data value through the cycle process of data governance and control with data assetization as the core.

## 2 The Traditional Shared Data Center Construction Mode and Existing Problems

Limited by the construction background and technical level of the digital campus era, universities have built a master data management platform with the goal of realizing data integration and sharing [2], as shown in Fig. 1. This model adopts the idea of centralized construction. It extracts, replicates, and integrates all business systems in the school into a shared data center through data integration tools such as data exchange platform ODI and API service bus and so on. Then, the master data management platform distributes the data to the corresponding business system, and builds a relatively primitive data exchange and sharing system integrating data collection, storage, management and use.

However, with the continuous advancement of smart campus, application scenarios continue to diversify and become instantaneous. The technical architecture of the traditional shared data center can no longer meet the current demand for data services in the construction of smart campus and the improvement of school governance ability, and there are many problems [3], which are summarized as follows:

- The traditional data center use black-box operation, which has high technical requirements, difficult maintenance and heavy workload, and usually requires technical vendors to assist in management, and has high operating costs.
- Data integration is not comprehensive, only structured data can be collected, semi-structured data and unstructured data integration is not supported.
- Data quality is not guaranteed. It is difficult to provide credible data support for leadership decision-making, support data statistics and analysis, and support complex cross-departmental business collaboration.
- Data sharing is inadequate. When departments need data, operations are complex and cannot be provided accurately.

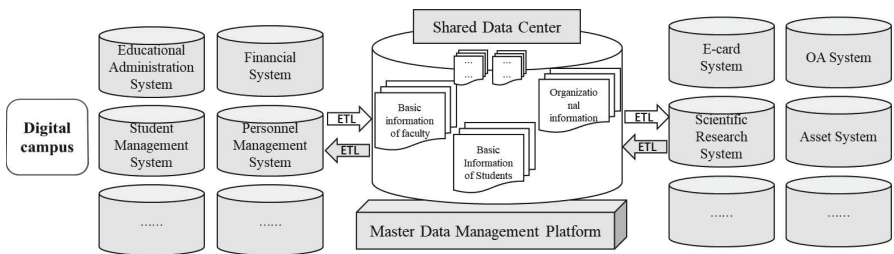


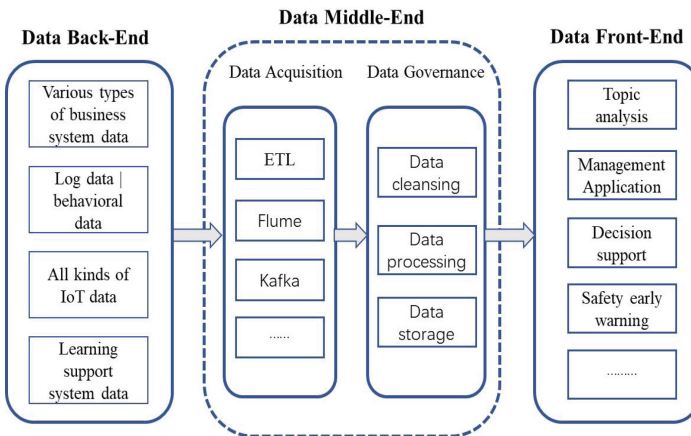
Fig. 1. Traditional shared data center for universities

### 3 The Relevant Concepts and Construction Significance of the Data Middle-End

The concept of data middle-end was first raised by Alibaba in China [4], and then gradually used by various industries. It has been hailed as the next stop for big data. It is defined as a set of sustainable “putting the data of the enterprise to use” mechanism, a strategic choice and organizational form. It is based on the unique business model and organizational structure of the enterprise, through mature products and implementation methodology support, to build a set of mechanisms to continuously turn data into assets and serve the business. Its underlying logic is to abstract data into services by digital means to respond to rapid changes in front-end services.

The application of the new data management architecture of the data middle-end to universities, many domestic experts have carried out relevant research. Hu Rui and Rui Zhong proposed a university data governance system based on data middle-end. Wang Xiaojing pointed out that big data technology should be integrated into the construction of smart campus data middle-end, and it is proposed to build and maintain a public data platform.

After collecting, storing and processing the whole-domain data in the school, the data middle-end implements standardized management of unified calculation caliber, system logic and management process to form standard data, and then stores it to form data assets. Finally, the data is transferred to various front-end application scenarios such as service, decision-making, and support, providing efficient services for the front-end. The data middle-end links the data back-end and the data front-end. It brings together the back-end data scattered in various business systems to form big data. Through data governance, effective data can be sorted out to flexibly and efficiently solve the personalized needs of the front-end, as shown in Fig. 2. Its core is the output of data service capabilities, which directly empowers various business applications of the school through data services.



**Fig. 2.** Diagram of the data middle-end

## 4 Architecture Design of University Data Middle-End

According to the whole life cycle process of data, the construction of data middle-end generally includes a series of projects such as data collection, data computing, data storage, data governance, and data services. It covers all business systems horizontally and vertically manages the entire data lifecycle. In this paper, the overall architecture of the data middle-end is divided into five parts: data acquisition layer, data storage layer, data management layer, data open layer and data application layer [5].

### 4.1 Data Acquisition Layer

The data acquisition layer mainly completes the aggregation and integration of full data. On the basis of structured data, we can collect semi-structured and unstructured data. For business systems based on structured data, such as personnel, scientific research, educational administration and so on, through the ETL integration tool regularly extracted data from the business system into the intermediate database, and stores it in the master database after some data integration operations.

Data in offline spreadsheets, log data output by business systems (mainly semi-structured data output by data streams), unstructured data such as files, videos, pictures, and the Internet of Things are collected into the platform through tools such as data filling and log processing. Most of the log data is in the form of syslog and stored in Kafka. Unstructured data for all kinds of the Internet of Things are stored in Hadoop-Hbase [6].

### 4.2 Data Storage Layer

The data storage layer is responsible for the storage of all data and undertakes the data transformation tasks of data assets. It stores the collected data into a uniformly designed school-level data model to improve data availability. Through the relevant collection tools, the whole-domain data of universities first enters the operational data store (hereinafter referred to as ODS) according to the scheduling of the collection cycle such as timing, polling, and triggering. ODS is used to collect and store raw business data.

The original heterogeneous data is first classified and stored, source governance, traceability data lineage and quality relationship, and then stored in the data warehouse DW after the data format is unified. The data warehouse is the main body of the data asset, which forms the data asset by sorting out and redefining the original data of the ODS layer, and transforming it according to the standard. Finally, in order to meet the needs of specific departments or users, it is stored in a multi-dimensional way, including defining dimensions, indicators to be calculated, and the level of dimensions, etc., to generate a data mart for decision analysis needs.

### 4.3 Data Management Layer

The data management layer is inseparable from the data storage layer, which takes data governance, data security and other service capability components as the core, and

builds unified metadata management and data quality management based on unified data standards, master data, and data algorithm models [7], forming the foundation of data assets and providing data support for data sharing and application on smart campus. Its core data management capabilities are as follows:

#### **4.3.1 Data Standard Management**

Standard management of data is mainly based on data elements, data items, and data sets, unifying data structure and code structure, and realizing data standardization.

#### **4.3.2 Metadata Management**

Through the management of metadata, the data lineage analysis between different tables is realized, which provides support for the establishment of data standards.

#### **4.3.3 Data Quality Management**

Through data quality inspection of data integrity, consistency, standardization, accuracy, uniqueness, relevance, etc., data quality report is formed to complete the check of data quality.

### **4.4 Data Open Layer**

This layer provides various forms of data open services such as API, database push, and file export based on the data resource directory. It provides standard data for business systems and data services downstream of data, completes data sharing, and forms a closed loop of data processes.

Taking the configuration API system as an example, data is retrieved through a unified API, and each API can provide data for different application development systems. Users independently select data resources to initiate data application, and data management personnel will authorize the API after the approval. The authorization can be fine-grained to the field level. When the API provides external data services, sensitive fields can be protected by encryption. It implements proactive data service push to meet various application scenarios.

As a unified management portal and data release service platform, the data open layer realizes unified data management and release, reduces the difficulty of data coordination and calling, improves the efficiency of data docking, and realizes autonomous, automated, and real-time data open services.

### **4.5 Data Application Layer**

The data application layer is located in the upper part of the data middle-end, and provides different types of data applications and services by defining the dimensions of service objects, data application scenarios, and specific data uses. Such as the following:

- Based on the data visualization platform, it can start from the roles of school leaders, administrative management, teaching staff and other roles to realize the real-time presentation of important indicators in terms of personnel, finance, materials,

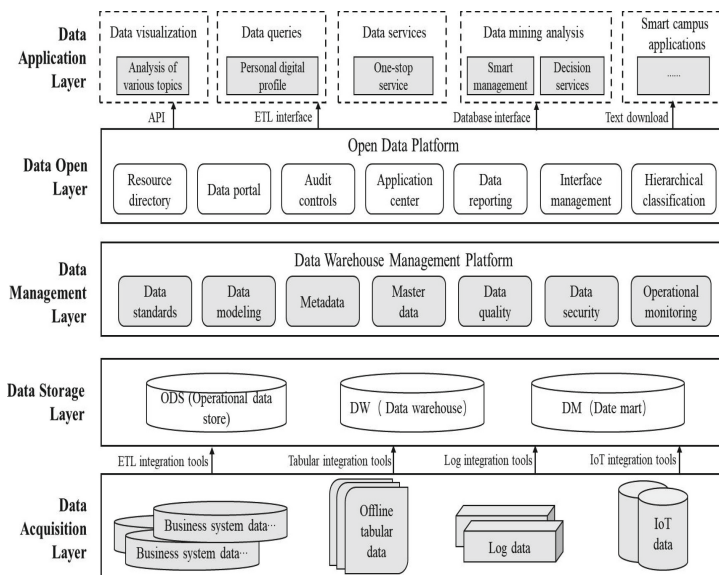


Fig. 3. Data middle-end architecture

teaching, learning, and research. The diversity and correlation of data assets across departments and businesses provide decision-making services at different levels for decision-makers.

- It can provide open catalogs of data assets at different levels. Under the premise of ensuring data security, it is convenient for users to obtain data and redefine data, and efficiently and quickly complete the daily data reporting work of the school such as high-base reports and talent training work status data, so that users can enjoy fast and convenient data services.

The designed data middle-end architecture is shown in Fig. 3. The whole system fully supports the data sharing and exchange of university data center and the comprehensive applications of smart campus.

## 5 Practice and Effect of Data Middle-End Construction in Universities

At present, the introduction of data middle-end architecture in the construction of universities smart campus can ensure that the entire digital campus system has great significance of capability openness and service sharing. This paper takes a university in northern China as an example to study the service improvement obtained after the implementation of the data middle-end architecture system [8].

### 5.1 Uniform Data Standards have been Formulated

On the basis of the investigation of the existing business system in the school, the data standards of the school are formulated in combination with the national standards, the

**Table 1.** All kinds of data provided by the data middle-end

<i>Categories</i>	<i>Title</i>	<i>Number</i>	<i>Title</i>	<i>Number</i>
Information encoding	Define data standards	658	Enable data standards	471
	Define the standard field numbers	10231	Enable the standard field numbers	6578
	Total amount of code	30151	Code mapping /transformation	120
Master data (key business objects)	Number of built data objects	484	Number of built data fields	7812
	Total amount of data	19608923	Average amount of object data	40514
	Data business systems	48	Acquisition business system	26
Data asset quality	Detect data objects	484	Detect fields	7812
	Involves business areas	14	Discover problem data	2715
Data asset openness	Data resource service catalog	87	Number of published APIs	35

standards of the Ministry of Education and the industry standards. This can ensure that the university has unified specifications in the process of data modeling, information collection, processing and data exchange.

## 5.2 A School-Level Data Asset Platform has been Built

The university has completed the extraction of multiple business data and various unstructured data. The quality inspection of the extracted data was carried out by adopting the governance strategy of “internal diagnosis and treatment + external traction”. At the same time, data quality reports are generated regularly and pushed to the corresponding data administrators to modify erroneous data. The well-governed data forms the school’s data assets, and provides API services for various business topics to the upper layer, which strongly supports various information services and data analysis needs in the smart campus. The various types of data provided after the completion of the data middle-end are shown in Table 1.

## 5.3 A Data Application Service System Has Been Constructed

- The business module based on the data middle-end can quickly develop light applications. For example, during the epidemic, developers quickly launched applications such as “Daily health reporting for teachers and students” and “Application for non-school personnel entering the school” based on basic data.

- The university has completed the data analysis and display for five major scenarios, including educational administration, personnel, student management, scientific research, and library, serving daily teaching, scientific research and management.
- The construction of one form was carried out to reduce the repeated filling of basic information by teachers and students, and realize the data “filling in once and sharing in multiple places” [9].

## 6 Conclusions

As the fifth major factor of production after land, labor, capital and technology, data plays an increasingly important role in the process of informatization and modernization of universities. The construction of the data middle-end is not only a technological innovation, but also an update of the organization and management model. Its introduction fundamentally improves data integration and data governance capabilities, and provides the possibility for “data governance”. But at the same time, we should also realize that the improvement of university governance capabilities must not only rely on the data middle-end system, but also require close cooperation in the strategic planning, management system, business process and personnel data literacy of the university. In this way, the maximum value of data can be realized, and the continuous innovation of campus information construction and application can be promoted [10].

## References

1. Key points of the Ministry of Education’s work in 2022[EB/OL]. [http://www.moe.gov.cn/jyb\\_sjzl/moe\\_164/202202/t20220208\\_597666.html](http://www.moe.gov.cn/jyb_sjzl/moe_164/202202/t20220208_597666.html), 2022–02–08.
2. CHEN Shiyuan, “Research on the construction method of university data center,” *Computer Era*, 2022,(04), 126-130.
3. LÜ Zhongting, CUI Wei, LIU Yang, ZHANG Jie, “Design and Analysis of the Construction Scheme of University Data Mid-End,” *Microcomputer Applications*, 2022,38(02), 19-21.
4. LI Guangqian, “What is a Data Middle-End?,” *Information China*, 2019,(06), 72-75.
5. LI Xiaolan, “Design and Implementation of Shared Data Center Platform in Higher Vocational Colleges,” *China Computer & Communication*, 2022,34(01), 122-124.
6. XU Bin. “Exploration of the construction of data middle platform in colleges and universities,” *Information Technology and Informatization*, 2020,(09), 246-248.
7. YANG Sen, ZHANG Ruiting, “Research on the Construction of Data Center in University,” *Office Informatization*, 2022,27(13), 27–29+36.
8. XIAO Jiongen, LIU Xinrong, “Research on Data Medium Platform Construction and Data Governance of Smart Campus,” *Journal of Tianjin Vocational Institute*, 2021,30(02), 82-86.
9. WANG Beibei, “Research on the Construction of Data Governance System in Vocational colleges under the Perspective of Big Data,” unpublished.
10. WEI Jianhang, LIU Yuanzhi, LUO Chao, WANG Linan, CAO Wena, “Research on data governance system of universities based on data middle platform”, *Information Technology and Informatization*, 2022,(06), 98-101.



**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

