

Study on Big Data Mining of Universities by the Fully Integrated Campus Datacenter Platform

Liwen Zhu ^a, Qingpeng Nie ^b

Informatization Office, Rizhao Polytechnic, Shandong 276826, China

^a zhuliwen@rzpt.edu.cn, ^b nqp@rzpt.edu.cn

Abstract. Mining big data of university and providing better services for users in university has become a key problem in data services phase of university informatization. Taking Rizhao Polytechnic as an example, this paper analyzes and studies this problem. A solution for the fully integrated campus datacenter platform is proposed, and its objectives and contents are also illustrated. The system framework of the integrated campus datacenter platform and its key technologies are detailed and discussed. Finally, the effects of this platform in the future are evaluated and summarized.

Keywords: big data, shared data center, data cleaning and integration, information criteria, comprehensive college situation analysis.

1. Introduction

In recent years, with the continuous advancement of information construction in Colleges and universities, informatization has brought high-speed and convenient services to teaching, scientific research, learning and life. At the same time, it has also produced a large number of data recording the daily management and teaching work of colleges and universities. These data include structured data and unstructured data. These data resources have high value. How to make the data "speak", play the value of the data, and better serve the various work of the university is a problem that every college must face.

2. Research Environment and Problem

With the rapid development of technologies such as big data and cloud computing, the application fields of big data are becoming more and more wide. From astronomy, climatology, information science, finance to e-government, public services, business intelligence, business management, marketing and so on, data are excavated and utilized. At present, the state attaches great importance to the application of big data in Colleges and universities. In the Guidelines on Further Promoting the Development of Vocational Education Informatization issued by the Ministry of Education, it is proposed that "the application level of modern information technology such as big data and cloud computing should be generally promoted in the decision-making, management and service of vocational colleges". The modernization of China's education needs informatization to support. The modernization of education and informatization inevitably require the realization of quality control, quality management and quality evaluation in technology. All of these require the analysis and processing of large data to provide support.

With the continuous advancement of information construction, the construction of colleges and universities is moving from digital campus to intelligent campus. In order to facilitate the teaching, scientific research and management of the school, many colleges and universities have built educational administration system, online teaching platform, scientific research management system, student management system, logistics services to provide convenient services. However, with the accumulation of time, these business systems are constantly generating a large amount of data at all times. In the past, the use of these data only stayed in simple applications such as monitoring, query, statistics, and so on. There is no deep mining of the value behind the data. How to let big data "speak" so as to provide services for teaching, scientific research, management, life and consumption of teachers and students in the most convenient way with valuable information in big data, and to

transform information-based data into intelligent service is a new and higher requirement for the construction of information-based campus.

Rizhao Polytechnic is one of the earliest vocational colleges to carry out the construction of digital campus. In the previous construction, 25 independent business systems, such as educational administration system and student-worker system, have been established. With the passage of time, more and more data deposits have been generated. Because the data between these systems are not interconnected, data inconsistencies have also occurred in the operation process. At present, the focus of school construction is changing from management as the core to service as the core. Building a fully integrated campus data center platform is very important for the development of the school. Based on the national information construction standard, the school establishes a fully integrated campus data center platform to realize data exchange and sharing among heterogeneous business systems and consolidate the data base. Through the construction of "flexible data center", all kinds of data resources will be centralized and effectively managed, data quality will be improved, and data resources will be monitored to ensure that information construction has data, data can be managed, data quality is high and data can be monitored.

3. Exploration of Big Data Solution in Colleges and Universities

To make big data "speak" requires the establishment of a fully integrated campus data center platform, which covers all aspects of teaching, scientific research, management, assets, finance, books, archives, employment and other data. Through data cleaning and integration platform, data are standardized according to information standards, making it the most authoritative shared data center database in the whole school.

3.1 The Construction Target of Fully Integrated Campus Data Center Platform.

(1) Realize the data sharing and fusion of each business system. Full integration of campus data center platform will standardize and standardize the data generated by various business systems, breaking the information island between business systems.

(2) Provide a comprehensive analysis and decision-making model of university big data. By using big data technology, real-time data collection and summary analysis in all aspects of school operation are realized, analysis model is established, data visualization is realized, and reference basis for school management decision is provided.

3.2 Construction Contents of Fully Integrated Campus Data Center Platform.

Fully integrated campus data center platform is constructed according to the principle of unified planning and step-by-step implementation, which includes four modules: information standards, shared data center, data cleaning and integration platform, and comprehensive school situation analysis.

3.2.1 Information Standards.

Based on the national standard, ministry standard, line standard and the existing information standard of the college, taking into account the compatibility, consistency and expansibility of the standards, the standards of the school are constructed and perfected. And the specifications of information classification and coding are given. A set of management information standards conforming to the actual situation of the school is constructed, so as to realize the unification of data standards and accurate information of various business departments. Form a unified specification of information in the process of collection, storage, processing, exchange, user access and transmission.

3.2.2 Shared Data Center.

To build a shared data center platform for the whole school, which can store the standardized data cleaned and transformed according to the information standards through the data cleaning and integration platform, solve the problem of information islands, ensure the accuracy and consistency of data, realize the authoritative data source of "who produces, who maintains, who is responsible",

and provide it to various business systems for invocation, while supporting to be large. Data analysis and application provide basic data sources.

3.2.3 Data Cleaning and Integration Platform.

Data cleaning and integration platform, based on ETL middleware technology, is the infrastructure for data exchange and sharing services between business systems and shared data center platforms, realizes data exchange and integration between heterogeneous systems, and ensures real-time updating and high consistency of data.

3.2.4 Comprehensive School Situation Analysis.

Using big data technology, the data of data center is analyzed and applied to show the school situation in an all-round way and provide services for all work of the school.

4. Construction System of Fully Integrated Campus Data Center Platform

4.1 Architecture of Fully Integrated Campus Data Center Platform.

Fully integrated campus data center platform adopts service-based multi-tier architecture. The system is divided into three layers: data storage layer, business logic and service layer, user interaction layer and so on. The data storage layer uses Oracle to store the basic information of the system, industry standard information of the Ministry of Education, business data sharing and metadata of the central database and other structured information. Business logic and service layer mainly aim at business logic of information standard, metadata, Central Library and basic data management. They provide business process infrastructure components and provide them to upper logic reference, other component reference and external system access in the form of service interface. User interaction layer mainly provides friendly access interface and data presentation mode for end users. Fully integrated campus data center platform integrates the shared data of the school business system into the central database of the platform through the DCI data integration tool through extraction and cleaning. The system architecture of Fully Integrated Campus Data Center Platform is shown in the following figure.

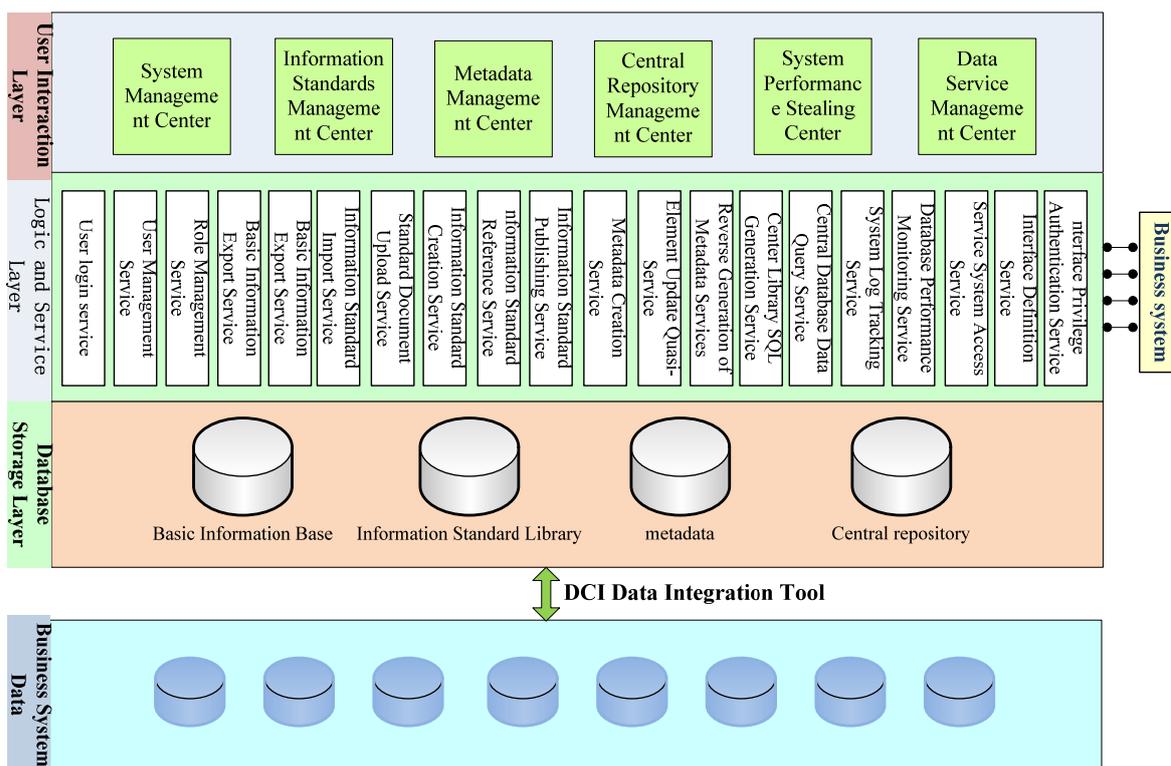


Fig. 1 System Architecture of Fully Fused Campus Data Center Platform

4.2 Key Technologies of Fully Integrated Campus Data Center Platform.

The platform mainly adopts DCI data integration technology. And DCI is divided into three subsystems: DCI-MS, DCI-Studio and DCI-Engine. DCI-MS focuses on integrated management and monitoring of data integration, scheduling data integration tasks; DCI-Studio focuses on task creation and design; DCI-Engine focuses on task parsing and execution, and provides support for clustering and distributed solutions. Each of the three systems has its own functions. They work together but do not depend on each other. The loosely coupled design can adapt to changes flexibly. On the other hand, when the internal structure and implementation of one system change, it will not affect other systems.

5. Summary

After the completion of the Full Integration Campus Data Center, most of the important data of the school are integrated into the shared data center database through the platform. The inconsistency of data among business systems is eliminated, and the problem of information islands is solved through the unified storage of data. The platform can provide data services for all levels of the school, guide the scientific, refined and standardized management of the school, in order to comprehensively enhance the core competitiveness of the school.

References

- [1]. Yingtao Q U, Office I, University H E. Construction and research of big data platform in Colleges and Universities[J]. Applied Science & Technology, 2018.
- [2]. Yang W, Jun Y, Zhenhua L. Research on the Planning and Construction of Platform of Comprehensive Informatization Service in Colleges and Universities[J]. Procedia Engineering, 2012, 29(4):69-73.
- [3]. Levy M, Raviv D. A framework for data center site risk metric[C]// Ubiquitous Computing, Electronics & Mobile Communication Conference. 2018.
- [4]. Mecca G, Papotti P, Santoro D. Schema Mappings: From Data Translation to Data Cleaning[M]// A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years. 2018.
- [5]. Sowmya R, Suneetha K R. Data Mining with Big Data[C]// International Conference on Intelligent Systems and Control. IEEE, 2017:246-250.
- [6]. Sztipanovits J, Bapty T, Koutsoukos X, et al. Model and Tool Integration Platforms for Cyber-Physical System Design[J]. Proceedings of the IEEE, 2018, PP (99):1-26.
- [7]. Naimi A I, Westreich D J. Big Data: A Revolution That Will Transform How We Live, Work, and Think[J]. Mathematics & Computer Education, 2014, 47(17):181-183.
- [8]. Chen H, Chiang R H L, Storey V C. Business Intelligence and Analytics: From Big Data to Big Impact. [M]// Business intelligence and analytics: From big data to big impact. 2012.
- [9]. Wu X, Zhu X, Wu G Q, et al. Data mining with big data[J]. IEEE Transactions on Knowledge & Data Engineering, 2013, 26(1):97-107.
- [10]. Chen C L P, Zhang C Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data[J]. Information Sciences, 2014, 275(11):314-347.