Research and Practice on Campus Big Data Foundation Platform

Shengzhong Yuan ¹, Hong He ²,*

¹Network and Information Management Centre, Shandong University, Weihai, China
²School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, China

*Corresponding author e-mail: hehong@sdu.edu.cn

Abstract: Universities at home and abroad have reached a consensus on the importance of big data research and application. The importance of education data has been promoted to the national strategic level. The construction of education data centre, the sharing and application of regional data have become the main practice mode. Our work is on the research and exploration from the shared database to the big data foundation platform in the university information construction. In addition to the existing educational administration system, we design and develop a series of application subsystems based on big data foundation platform, which can be presented to readers as cases, hoping to serve as an inspiration to each other.

1. Introduction

In the era of vigorous development of new technologies such as Internet, big data, etc. the use of cool techs has made today's universities a mini-city of wisdom. It not only brings more convenient in the study, work and life for teachers and students, but also make campus management easier.

International famous universities have made great progress in big data application development in recent years. Arizona State University [1] has made innovative improvements in curriculum, campus life, technology and facilities to meet the needs of students and teachers in the past three years. The biggest action of the school is the forward-looking "smart campus" plan launched in 2016. Arizona State University hopes to provide teachers and students with high quality, personalized learning and life experiences by combining the Internet of things, big data and their analysis. Transformation and upgrading the Sun Devil Stadium can be said a representative case. The 75000-seat stadium has sensors and cameras in every corner. The former can be used to collect data on site temperature, voice decibels, and the operation of each faucet, while the latter can be used to track queues at franchises and bathrooms. Data analysis will give the stadium more insight into its operation.
The Internet of things at the University of Nebraska, Lincoln, in addition to the lockers, self-checkout systems, and control systems [2], has ushered in a new "member"—Print IT, an intelligent kiosk based on cloud systems since three years ago. By 2017, there had been 28 Print IT all over the campus which allowed students to copy, print, scan, fax from any device, as well as from cloud accounts and USB drives. Print IT are installed in crowded areas of the campus, such as dormitories, libraries, cafes and teaching buildings. In this way, students and teachers will no longer need to extract the required documents from the computer in the laboratory, but will be able to press printing keys in their own room and print documents from any Print IT on campus later.

A smart grid project at the University of Washington [3], which installs more than 200 smart meters in campus buildings, tracks energy use in real time. By analyzing the collected data, the computer software can determine the energy consumption of students, and finally put forward the measures of reducing energy consumption and saving costs. Another goal of the project is to determine which energy-saving strategies are more effective in achieving long-term carbon reduction targets.

Domestic famous colleges and universities in recent years also attach great importance to the application of big data research [4, 6]. Peking University established the financial big data undergraduate cross major in 2016. Tsinghua University set up the Institute of data science and announced that it will launch the Master of Science project of multidisciplinary training. Sichuan University, University of Foreign Economics and Trade, and Shandong University have laid out the big data analysis platform for students' learning management in the construction of wisdom classroom.

In the construction of colleges and universities in China, the big data of training of talents and campus card [4, 5] data is highly concerned, but it is rare to see the management system of college students based on big data. After more than 10 years of information construction in Colleges and universities, the campus network is complete and the application of campus card and smart phone leads to a large number of student data produced every day. These big data reflecting the situation of the students are the important basis for the management of students. The actual application of data in the colleges and universities level is not enough, the overall construction model has not been formed. There are many problems, such as data concentration, low degree of construction and application of departmental information system, low quality of data and insufficient demand for data analysis at all levels. In a few universities with shared databases, there is also a lack of overall management of universities data. The application of big data (Xi'an Jiaotong University, Fudan University, Central China normal University) for students and (Zhejiang University) for subject are being studied [6].

In the wisdom classroom construction, we layout the campus big data foundation platform and designed several applications to support common business and serve teaching and management. This platform and application cases described later will be a reference for the construction of fraternal colleges and universities.

2. Construction of Campus Big Data Foundation Platform
The campus big data foundation platform obtains initial data from admissions management. Freshmen's data are pretreatment into the shared database [4, 7, 8] through allocation of student numbers, binding of mobile number, bank account number and other attributions. With the students' growth, the big data foundation platform works to record all of the student data. These data include student scores from the teaching management system, including records of punishment for cheating in
exams. Moreover, there are student performance data from the student management system that include subject competition awards, overseas visits, scholarships, student loans, work-study programs, and disciplinary violations, etc. See Fig1.

![Figure 1. The data trend in campus](image)

2.1. Data sources and big data acquisition
All of the big data applications based on the campus information and perfect data integration. The big data source mainly includes the data of various service systems of colleges and universities, the log data of machines and the data of the external Internet. It is needed to open the service data, develop and use the existing machine data, and at the same time incorporate the Internet data. Here we need to collect all kinds of data, after cleaning and merging into the big data foundation platform.

2.1.1. Service data
Service data mainly includes the service data coming from the campus card system, student management system, entrance guard system, educational administration system, scientific research system and library system, etc. It is the basic data stored in the relational database corresponding to each service system. See Table 1.

| Nama  | ID number       | Student Number | Department         | Dorm  | Mobile phone  | Events | recorder |
|-------|-----------------|----------------|--------------------|-------|---------------|--------|----------|
| Jack  | 3703****201     | 201800****23   | Information Engineering | 8-3-102 | 178****678   | 10     | Sara     |

In order to be safe and convenient, besides retrieving data from existing business databases, schools can use sensors and cameras to collect data from the entire campus, thus enhancing the students' daily life experience on campus. Classroom is a concentrated embodiment of school education thought. Students' performance data in classroom can support the process management of teaching. For example, using location based sensors and "virtual Bluetooth technology" to determine whether students took part in a course. Those students who are always absent from class may fail. By collecting and analyzing past data on students' learning activities, we found that those who did not regularly attend tutoring classes for freshmen were more likely to drop out of school in their sophomore years. Fig 2 shows how the student portrait system acquires service data from the databases of the existing management information system.
Student parents applications

School administrator applications:
Analytical decision-making and early warning systems

Employer applications

| Teaching administration platform | Intelligent classroom management system | Student management | Innovative and entrepreneurial platform | Campus Card system | Entrance guard system | Library system |
|---------------------------------|----------------------------------------|--------------------|----------------------------------------|-------------------|----------------------|---------------|

**Figure 2.** Big data foundation platform acquires service data from the existing databases

Service data are regularly push into the big data foundation platform of MySQL cluster by ETL engine. The ODI tools are usually used to push service data, shown in Fig 3.

Particularly, in order to maintain data consistency in the big data foundation platform, we pay attention to keeping the uniqueness of the data source. That is to say, there is only one role is “write” permission, and other roles have only “read” permission in data access.

**Figure 3.** ODI tools schematic diagram

The mapping relationship between the source database and the target database is established by using the GUI interface of ODI, and the data is pushed into the big data platform regularly by using ODI Agent [9].

2.1.2. **Machine data**

Machine data mainly includes college wireless log, internet log and URL log. It certainly includes the logging of devices such as mainstream vendors that records user authentication information and user connection information recorded by wireless AP devices. Through the open source software Flume to receive, and stored in the NFS cluster. Using Ruby language to write a program to read the data from the NFS cluster in real time and then write it into the big data base platform of the MySQL cluster and Elastic search cluster after the cleaning transformation. Fig 4 shows a sample of machine data.
By processing machine data, we can observe the time and habits of students online. Students who browse online during class hours are easy to drop out. There are many students online during class hours, indicating that the teacher’s lectures do not attract students' attention.

2.1.3. Internet data
Internet data mainly includes various portals, news, Weibo, social, forums, posts, and other data. The data can be collected by crawler engine and stored in university big data platform after being cleaned. Figure 5 shows a sample of Internet data [10]. For the external data on the Internet, the content is crawled regularly through the Ruby crawler engine, and then cleaned and converted into the big data foundation platform of the MongoDB cluster. This part mainly aims at the university public opinion monitoring system.

2.2. Data storage and calculation
The data is stored on the private cloud of the campus network. According to the scale and category of the data, the data calculation is divided into offline data processing and online real-time data processing. The offline data processing means using the traditional Hadoop MapReduce calculation method. The mass data is decomposed to a single nodes by a large number of cheap hardware servers, then the results are aggregated to the client. For some of the more real-time data, such as the log class,
the Elastic Search cluster needs to be used for real-time processing, and the results are fed back to the client. Fig 6 shows the framework.

| Data Application: | Smart Campus, Teaching and Practice, Query, search, Statistics, Analysis, early warning … |
|-------------------|--------------------------------------------------------------------------------------------------|
| Data Analysis:    | Visualization, model, Self-help graphical reporting, algorithm … |

| Distributed computing engine | Data Integration | Data access | Security | Operation | Distributed database, Compatibl e SQL | Memory database | Big data software and hardware integrated professiona l services and safety control |
|-----------------------------|------------------|-------------|----------|-----------|--------------------------------------|-----------------|-----------------------------------------|
| Data acquisition            | schedulin g of resources | Administratio n | Authenticatio n | Monitoring deployment | | | |
| Sqoop                       | HDFS             | Audit data protection | | | | |
| Kafka                       | Data Storage     | | | | | |
| Flume                       | X86 adaptation   | HPC adaptation | virtualization adaptation | | |
| WebHdfs                     |                  | | | | Container |
| Environmenta l adaptation   |                  | | | | | |

**Figure 6.** Storage and computing framework deployed on the campus network

### 3. Application Cases Based on Campus Big Data Foundation Platform

Big data assets are the foundation of colleges and universities information construction. According to the business needs, we can develop various applications based on big data foundation platform to serve students and teachers. In particular, the following subsystems based on big data foundation platform are new addition to the existing teaching system and student management system. They form the overall solution of campus information together.

**3.1. Information Fusion Subsystem of College Students Facing Honesty Education**

The subsystem is designed for the following reasons:

1. Student management information is an important part of education. The Ministry of Education promotes student management information because it is still the weak link of university management.
2. The system is the indispensable components of the overall framework of information construction.
3. It can solve the problem of isolated information islands in student management after the construction.
4. The establishment of a platform for the whole society, including parents and employing unit, to participate in students' education.
5. It is beneficial for students' management to pre-process and analyse problems in students. This information fusion subsystem is equivalent to the portrait subsystem of college students. The daily performance data of students, including academic performance, class attendance, examination discipline, innovation activities and subject competition awards, etc. are all gathered into the
comprehensive quality evaluation system that forms a kind of data honesty for objective examination of students. See Fig 1. Like the bank's credit system, good data record students' good performance. The data of students' academic misconduct, forging, altering, or lending documents or materials, spreading rumors on the Internet or attacking others by hacking techniques are recorded in the big data foundation platform by the relevant departments. In this way, the cost of students making mistakes and dishonesty will become very high. The system monitors their behavior everywhere like a working Skynet, stifles their lucky psychology of making mistakes, and cultivates the habit of honesty. Students can constantly regulate their behavior and improve their own quality by querying their honesty. Not only can the information fusion subsystem help to find and solve the problems in student management in time, but also help the parents and employers find out how to deal with the problems encountered in the process of growth.

3.2. An analysis of card consumption and precision poverty alleviation subsystem

This subsystem uses the relevant data of the information fusion subsystem to generate accurate poverty alleviation results. For example, when anonymous donors ask for funding for needy students, we can immediately give them a list of students who consume in the canteen, but who have a low level of consumption. Of course, they can also add other conditions, such as preferences for college or major, or failing in grades.

3.3. Teachers and students comprehensive portrait subsystem

Classroom is the epitome of the thought of college education. The wisdom classroom is an important infrastructure of modern university education, which promotes the transformation of teaching concept from teaching-centered to learning-centered and teaching process management. The main functions of the wisdom classroom are to facilitate interactive teaching, complete the online recording of the classroom, broadcast live online across platforms (Windows, IOS, Android, etc.) and realize the online interaction of courses inside and outside the school, and record the whole process of classroom teaching and learning and its behavior analysis. Because each lesson of the teacher is recorded for reference after class, the teacher will prepare lessons carefully as far as possible. At this point, the teacher welcomed by the students is the teacher who is good at teaching. Nowadays, there is a general situation among teachers that emphasis on teaching is less important than scientific research. Some teachers do not concentrate on improving the teaching level. The big data of the wisdom classroom will accurately reflect the professional quality of teachers. Of course, students' grades in peacetime are also more objective.

After the data of the wisdom classroom system enters the big data foundation platform, together with other data, it can support the teacher-student portrait system, which describes the individual performance of the teacher and student more comprehensively.

3.4. Energy saving subsystem of public lighting sites

Teaching buildings often appear bright lights at night and small number of self-study, energy waste is serious. Energy saving subsystem learns the number of students by collecting the video data of the classroom during the evening study.
3.5. Forecasting and warning subsystem
Through the analysis of the big data, a prediction model for students’ achievement is established, the subsystem automatically prompts or reports to the relevant departments for the absence or missing persons without any reason, the low consumption personnel, the abnormal Internet access and Internet addicts, the illegal use of electricity, and so on.

3.6. Campus public opinion monitoring subsystem
It observes and guides the trend of public opinion inside and outside the campus, intervenes in the dissemination of false information in time, avoids the occurrence of adverse public impact events, ensures the reputation of the school, and stabilizing the order of the campus.

3.7. The big data experimental management subsystem
The big data experimental management subsystem [12] aims to provide a unified platform for managing all course teaching materials, handouts, experimental manuals, experimental data sets, experimental assignments, experimental reports, experimental results management, user management (including student roster management and teacher Information management). The subsystem serves teachers and students. Platform administrator and virtualization cluster administrator are the same, responsible for the creation and allocation of virtual machines. Teachers can upload, download, update resources, issue experimental steps, assign assignments, and their grading standards, manage experimental results, and have the highest authority. Students are able to manage their own personalized materials, including lab reports, experimental results, and their own learning and experimental materials, while allowing online laboratory operations and periodic test assignments. Check the experimental data of big data technology uploaded by teachers, and upload, move, update, delete, interactive edit, copy and other operations.

Teachers and students use a unified login entry to enter the big data experimental platform, the experimental management software platform docked behind the virtualization management cluster and the physical cluster transparent to users, supporting the distribution of virtual machines on demand. Easy and fast access to experimental data and data sets.

4. Conclusions
College campus is a big data source. These data are the objective basis of the whole staff, the whole process and the all-round education. Using the mature campus network infrastructure to build a private cloud, the campus big data solution can be deployed at a lower cost. With the development of business, the application development based on big data foundation platform is the sustainable content of campus information construction.

After the completion of the campus big data foundation platform, the data after partial desensitization can be used for big data course system and student training. Courses that can be offered include big data basics, the key technologies (such as deep learning, introduction of Hadoop ecology, high efficiency computing, data mining algorithms analysis, etc.), and big data technology practical training.

With the development of information construction in colleges and universities, relying on information resources, providing more value-added services for teachers and students is becoming more and more important to the supporting institutions of information construction, which requires the
system R & D personnel to have a deeper understanding of the business of the main customers. With the increasing collection of information on big data basic platform, the application of modern information technology in big data is becoming deeper and deeper, and the constant development of value-added services has put forward higher requirements for database construction, security and management in colleges and universities.

References
[1] Information on https://fullcircle.asu.edu/students/nsf-graduate-research-fellow-engineering-solutions-big-data-challenges/, NSF Graduate Research Fellow engineering solutions to big data challenges, Posted by Lanelle Strawder, July 27, 2018.
[2] Information on https://www.unl.edu/, NIC to house Nebraska's first self-driving shuttle, by Scott Schrage, Aug, 2018.
[3] Information on http://www.washington.edu/, July, 2018
[4] XIN Ying, Research on Digital Campus Based on Data Centre, Master's thesis, May 2013.
[5] CHI Xingying, YUAN Yong, Prospect on Application Management and Development of Campus Card system [J], Computer knowledge and Technology, Nov 2014. Pp. 2475-2477.
[6] Information on http://www.sugon.com, Aug 2018.
[7] WANG Yongming, Basic platform Construction and Application function Analysis of Campus Card system [J]. Intelligent Building and Urban Information, Jan 2009. Pp.101-105.
[8] ZHOU Zhaoyong, Construction of Campus Card in the Environment of Digital Campus [J]. China education network, Jan 2014. Pp. 45-48.
[9] Information on http://www.bubuko.com/infodetail-61701.html, Aug 2018.
[10] Information on https://blog.csdn.net/myboyliu2007/article/details/51557438, writing from Flume to HDFS. Jan 2016.
[11] CHENG Cheng, OU Shuyang, The Status Quo and Problems of the Building of China’s Social Credit System and Suggestions [J]. International Business and Management. Feb 2014. Pp. 169-173.
[12] Information on http://www.inspur.com/lcjtww/2315750/2315754/2322269/index.html, Sept 2017.