Exploration on Construction of Big Data Laboratory

Ji Kan[1], Pu Heping[2], Chen Xiaojun[3]

[1]Southwest Petroleum University, Network and Information Center, Chengdu , Sichuan, China

Abstract: *Based on the actual situation of Southwest Petroleum University(SWPU), this paper analyzes the necessity of the construction of the big data laboratory, explores its feasibility and puts forward the construction scheme. The Network and Information Center (NI) provides construction site and professional technicians and introduces New Cape as the big data experimental platform. The big data laboratory can provide a strong support for cultivating big data talents skilled in both big data technology and industry application.*

Keywords: big data laboratory, experimental platform, construction scheme

1. Introduction

Big data, which has been widely applied in recently years as an increasingly mature technology, has attracted intense attention from education and information industry. Since Data Science and Big Data Technology was first set up as an major in 2015, there has been 35 Universities which has set up the major of Data Science and Big Data Technology in 2016, according to the information provided by the ministry of education[1]. For students majoring in Data Science and Big Data Technology in SWPU the following courses including Probability Theory and Mathematical Statistics, Multivariate Statistical Analysis, Data Warehouse and Mining Technology, Algorithm Analysis and Technology, Data Visualization Technology, etc. are offered. To learn these courses well, students need more than theoretic studies. Experiments also are required in order to deepen their understanding of theoretical knowledge and improve their practical hands-on ability[2]. Therefore, it's of great necessity to set up laboratories with not only professional equipment and basic software environment, but also computing environment and practical cases of big data. Although School of Computer Science has established I9000 big data application development platform, the system has not been widely used due to its complex operation and lack of professional guidance. The survey shows that software and hardware environment required for data mining and big data analysis experiments cannot be completed in existing laboratories. Therefore, the construction of specialized big data laboratories is the inevitable choice in the era of big data, and the only choice for the development of laboratory science in SWPU.

2. Demand Analysis of the construction of Network Information Center Big Data Laboratory

2.1 Demand of big data application

In the process of information construction, colleges have deployed many business systems, which produce a large amount of data every day at an increasing rate. These data include not only structured data generated by MIS such as teaching, scientific research, personnel, finance, etc, but also unstructured data such as logs, videos, pictures, as well as behavior data by card consumption and campus network online. Although these data have been effectively managed and stored, they are inadequately analyzed, mined and applied due to the lack of big data laboratories, big data technology and data application experience[3].

2.2 Demand of scientific research and teaching in universities

As Big Data is an emerging major, some universities have zero experience in the teaching of big data majors. The teaching of big data majors is in the exploratory stage and focuses more on theory teaching. Due to the lack of big data experiment environment, the students don't have strong experimental ability, and they have low proficiency in big data technology. Big data laboratory can not only meet the teaching needs, but also provide support for scientific research in universities. In order to obtain technical achievements in big data-related scientific research, computer laboratories in universities should build a teaching and scientific research platform centered on big data technology [4]. Massive data is the premise of big data teaching and scientific research. Universities do not have access to massive data, nor methods to obtain data according to law. Moreover, there is lack of real data and cases of enterprises as comprehensive data for big data experiment and training.

2.3 Demand of big data experiment environment and hardware equipment

Generally speaking, students majoring in computer can complete various professional learning tasks with one machine, but experiments of Big Data courses cannot be completed without at least two machines. It is very difficult to carry out big data experiments, due to high degree of relevance and closeness of big data major courses, high requirements for preserving experimental results, high hardware requirements of big data experiment environment, and the lack of professional big data experiment environment.

2.4 Demand of big data market

It is estimated that the global data generation volume will reach 50.5ZB in 2020, with a year-on-year increase of 23%. During the period of 2008-2020, it is predicted that the overall revenue scale of big data market will maintain an annual growth of about US$7 billion and the revenue scale of global big data market will reach 56 billion US dollars. The amount of data generated in my country will increase from 7.6ZB in 2018 to 48.6ZB in 2025, and CAGR will reach

30.35%, exceeding the amount of data generated by United States during the same period by about 18ZB[5].

3. Feasibility of the construction of Network and Information Center Big Data Laboratory of in SWPU

3.1 Policy Guarantee

The Outline of the 13th five-year Plan puts forward the implementation of the national big data strategy, "taking big data as basic strategic resources, comprehensively implementing actions to promote the development of big data, speeding up the opening up and sharing of data resources, and its development and application, and facilitating the industrial transformation and upgrading and social governance innovation[6]. In the key tasks of the document "Key Points of Educational Informatization in Sichuan Province in 2019", it is clearly proposed to continue to explore and develop the application of educational big data, further strengthen the in-depth mining and analysis of educational big data, and actively promote the thematic application of big data in the content of improving the basic environment construction of educational informatization. SWPU will take education informatization 2.0 as the goal, the principle of "adapting measures to local conditions, steady advancement, characteristic development" as the guidance, and strengthen the connotation construction, and in-depth implementing of Thirteenth Five-Year Plan for Information Construction of Southwest Petroleum University, according to Southwest Petroleum University Informationization Key Points in 2019. It points out that based on smart teaching systems, SWPU will conduct research on the application of big data in teaching informatization, explore the process evaluation of teaching and learning, and vigorously promote the construction of educational informatization, and strive to build an information environment that is compatible with the construction of first-class disciplines.

3.2 Personnel Guarantee

There are 45 on-the-job staff, including 23 with master's degree or above, 5 with deputy senior titles and 23 with intermediate titles in Network and Information Center of SWPU[7]. It has formed a technical backbone with relatively stable and reasonable educational background, age and title structure. Apart from that, the research backbones of School of Computer Science in big data are joined to form a team of both full-time and part-time teachers.

3.3 Data Guarantee

In the construction process of digital campus, structured and unstructured data generated by the business system, as well as real cases and data provided by big data experiment platform, provide powerful guarantee and support for both teachers and students in teaching and scientific research.

4. Construction Scheme of Network and Information Center Big Data Laboratory

4.1 Construction Goals

It is aimed to fully implement the mode of the integration of production, learning, research and application, and establish a big data laboratory that focuses on cultivating professional and characteristic talents from multiple perspectives of theory, practice, scientific research and application[8]. The goal is to improve the effectiveness of teaching and the quality of talent cultivation based on project-driven principle, with teaching as the priority, and the updating of supporting management services. Relying on the technical support of the Network Information Center and the faculty and technical resources of the research team of School of Computer Science, course resources of big data will be used to integrate theoretical teaching, experimental teaching and practical projects of big data enterprises, so as to improve students' big data technical level step by step. Professional big data analysis experimental environment can provide support for teachers' scientific research innovation; students can apply theoretical knowledge such as algorithm analysis and design to actual data analysis process so that they combine theoretical knowledge with practice, and therefore improve their project practice ability, making good preparation for future employment.

4.2 Construction scheme

(1) Hardware environment of Big Data Laboratory
Hardware of Big Data Laboratory[9] consists of servers, terminal computers etc. Server is the core hardware platform of Big Data Laboratory, providing sufficient storage space, memory space, cloud computing resources, mirror scheduling and other resources to support achievement accumulation of Big Data Laboratory. Terminal computers mainly support teachers, students and scientific researchers to carry out big data experiment courses and scientific research work. The cloud computing cluster access device connects the storage server to the user, and the user invokes the mirror resource in the mirror warehouse through the container scheduling center. Big data experiment platform adopts B/S structure. Teachers and students use terminal computers to access resources of big data experiment platform through browsers. The network topology of big data laboratory construction is shown in Figure 1.
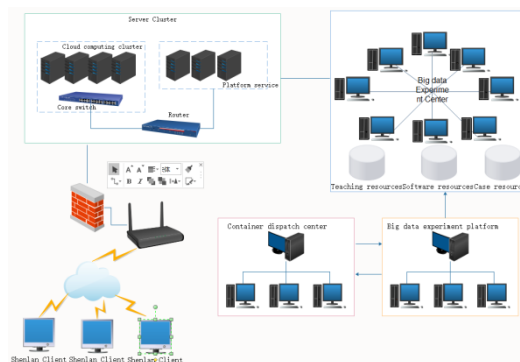

Figure 1: Network topology of big data Laboratory construction

(2) Software environmentof Big Data Laboratory

Big Data Laboratory introduces the new big data experiment platform of Cape, including cloud computing management system, big data course management system, big data online experiment system and big data cloud storage system. The big data experiment platform is shown in Figure 2.
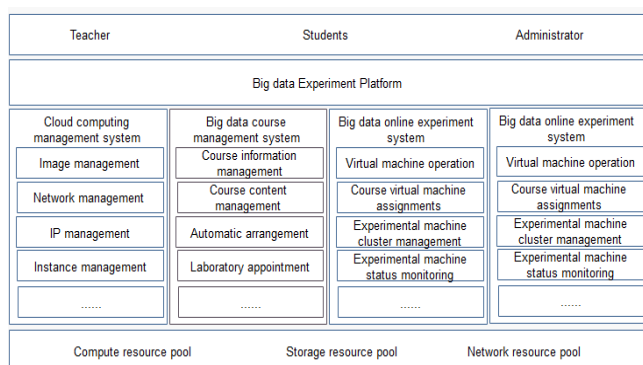

Figure 2: Big data experiment platform

The cloud computing management system is mainly used for real-time monitoring of total resource consumption of cloud servers, as well as management and operation of use of mirrored warehouses.

The big data course management system sets up course content corresponding to big data, such as Hadoop, R language, Python, etc, and can also innovate course content.

The big data online experiment system provides big data experiment cases and supports experiment manuals. A variety of big data development environment are nested within the system, including a variety of big data experiment courses.

The big data cloud storage system mainly manages user storage space, uploading, and downloading various types of files.

(3) Data Resources

Big data is inseparable from massive data resources. Under the background of digital campus construction, after years of information construction, Southwest Petroleum University has formulated information management norms and technical standards, perfected network security and data security guarantee systems, built the unified data sharing and exchange platform, unified identity authentication and authorization management platform, unified information portal, OA, financial system, scientific research system, personnel system[10], educational administration system, equipment asset system, academic engineering system and other cross-departmental school-level business systems. The full coverage of school-level business system has basically been achieved. Years of operation of network and business system have accumulated a large amount of business data, operation data and logs. Sufficient mining and reasonable analysis of these data can yield a lot of valuable information. It is also a good time for the implementation and deployment of Big Data Laboratory. While fully mining school data resources, the Big Data Lab also collects data from other industries, such as real estate and medical care.

(4) Construction of big data experiment course system

In order to ensure close combination of theory, practice and the cultivation of students' knowledge application ability[11], the Big Data Laboratory is also equipped with the big data experiment course system, including teaching outline, experimental documents, teaching PPT, practical cases and other supporting resources to support teaching and scientific research. Course System of Data Science and Big Data Technology is shown in Figure 3.
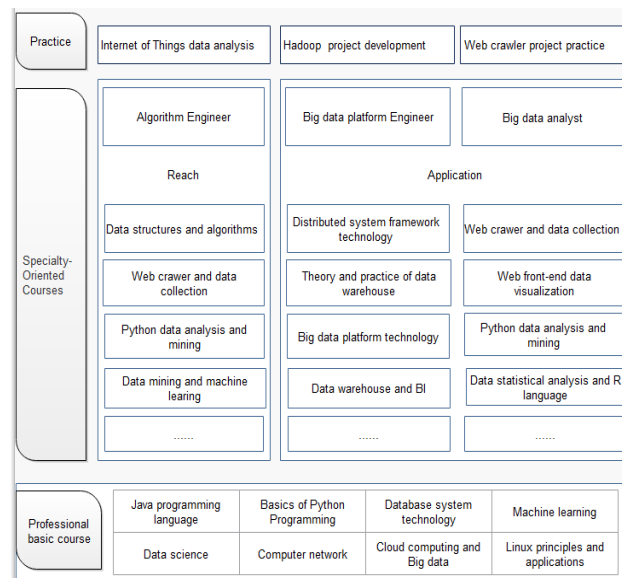

Figure 3: Curriculum system of Data Science and Big Data Technology

(5) Construction of teacher training system

Big Data professionals are needed to manage big data laboratories. The company's practical project experience in big data and Big Data talent training system built by Network Information Center can jointly meet the needs of related professional teacher training, technical exchanges and technical competitions. Technicians of the Network Information Center will undertake training tasks and laboratory management work related to big data, and make full use of the laboratory to meet the needs of technical exchanges and technical competitions.

5. Features and highlights of Network Information Center Big Data Laboratory

First, the big data experiment platform teaching basic experiment course of New Cape Electronics Co is introduced., Ltd, to build a big data basic experiment course management platform. The big data experiment teaching system is introduced to build the practical training platform, to integrate the theoretical knowledge, practical teaching and big data practical project, so as to improve students' hands-on operation ability. High-end professional equipment provided by New Cape company is used to provide software and hardware as well as practice environment for experiment teaching. The professional big data environment built within the framework of the mainstream big data software can help students apply theoretical knowledge to practical analysis and development of big data, improving their practical project ability.

Big data analysis experiment platform has the characteristics of openness, scalability, security, stability, simplicity, ease of use, and easy management and maintenance. It not only develops different businesses in view of different research groups, but also meets the needs of training students of multiple colleges and multiple disciplines, to help cultivate innovative talents. A big data laboratory provides a professional experimental environment for big data teaching and scientific research and case teaching of real application scenarios.The big data scientific research results provide new management methods for students' grid management and students' precise poverty alleviation.

6. Conclusion

Combined with the actual situation of SWPU, the target of Network Information Center Big Data Laboratory was clarified. The feasibility was analyzed in view of the needs of the construction of Big Data Laboratory,. Finally, the construction scheme of big data laboratory of Network Information Center was proposed. The big data laboratory is to provide the most basic knowledge of teaching and research and development environment for big data talent training. Teachers rationally design big data experimental teaching programs to enable students to master big data technology proficiently, thus guaranteeing the training for big data talents. There are still some aspects not fully considered in the construction of big data laboratory in this article, such as the construction of big data research platform, which is also a direction for further exploration and research in the next step.

References

[1] Li Fujuan. Research on construction and application of big data laboratory[J]. Experimental Technology and Management,2018,35(5):243-245

[2] Baidu Wenku-Construction scheme of TipDM big data cloud service laboratory [EB/OL].(2016-06-28). https://wenku.baidu.com/view/daca3b265fbfc77da369b103.html

[3] Douban-Application of big data technology in university[EB/OL].(2019-05-31).https://www.douban.com/note/720713149/

[4] Liao Jun, Zhang Yi, Wang Chengliang, etc. Research and analysis of university big data laboratory construction system[J]. Experimental Technology and Management,2018,35(11)：241-244

[5] China Industry Information. Market size forecast and development pain point analysis of global and Chinese big data industry in 2019-2025[EB/OL].(2020-02-20).https://www.chyxx.com/industry/202002/835772.html

[6] Language TIPS. Xijinping: implementing the National Big Data Strategy to speed up the building of digital China[EB/OL].(2017-12-12) http://language.chinadaily.com.cn/2017-12/12/content_35280418.htm

[7] Network and Information Center. Introduction[EB/OL]. https://www.swpu.edu.cn/nic/zxgk1/zxjj.htm

[8] Construction scheme of big data laboratory in university[EB/OL].(2020-01-16).https://wenku.baidu.com/view/4a908f3342323968011ca300a6c30c225901f039.html

[9] Li Hui, Zhang Biao, GaoWanlin. Thinking and exploration on construction of big data [J]. Experimental technology and management,2016,33(7):195-199

[10] Ji Kan, Deng Yun etc. Analysis of college students' behavior based on big data[J]. Journal of Fujian Computer, 2018,34(09):137-137+133

[11] Li Xiaodan, Liu Yunxiang etc. Construction on big data analysis laboratory of application-oriented computer majors[J]. Experimental Technology and Management,2017,34(9):236-238

Author Profile

Ji Kan received the master degree in Computer software and theory from Southwest Petroleum University in 2012, and have been working in the Network and Information Center of Southwest Petroleum University since my graduation in 2012.