

Received: August 22, 2017

Revision received: July 14, 2018

Accepted: July 20, 2018

Copyright © 2018 ESTP

www.estp.com.tr

DOI 10.12738/estp.2018.6.228 • December 2018 • 18(6) • 3247-3253

Research Article

Construction and Management of Computer Professional Education Laboratory in the Context of Big Data

Chunying Shi¹*Shandong Agriculture and Engineering University*HuiQin Li²*Shandong Agriculture and Engineering University*Chuanzhong Mao³*Shandong Agriculture and Engineering University*Xiufang Shi⁴*Shandong Electronic Information Products Inspection Institute*Yanjun Zhang⁵*Shandong Agriculture and Engineering University*Jin Zhang⁶*Shandong Agriculture and Engineering University*

Abstract

The advent of the mobile Internet industry has brought about exponential growth of data. The enterprises' demand for big data talents has become more urgent. However, enterprises generally reflect that the big data-related talents cultivated by universities cannot meet the requirements of enterprises. In view of the current insufficient support for big data in college teaching, this paper, aimed at applied talent training, analyses the core work of big data analysis laboratory construction and its specific content. Then, it proposes a big data laboratory construction plan based on industry-university-research integration, which was applied in one university by comparing the learning effects of students in different grades. The results show that the lab plan can effectively improve students' ability to analyse and apply data, and provide strong support and useful reference for college big data education and scientific research.

Keywords

Big Data • Teaching Research • Applied Talents • Computer Education

¹Correspondence to: Chunying Shi (MSE), Shandong Agriculture and Engineering University, Jinan 250100, China. Email: shichunying@sina.com

²Shandong Agriculture and Engineering University, Jinan 250100, China. Email: lihuiqin608@163.com.

³Shandong Agriculture and Engineering University, Jinan 250100, China. Email: 362361522@qq.com

⁴Shandong Electronic Information Products Inspection Institute, Jinan 250100, China. Email: sdbzit@126.com

⁵Shandong Agriculture and Engineering University, Jinan 250100, China. Email: zyj8033@163.com

⁶Shandong Agriculture and Engineering University, Jinan 250100, China. Email: 116405921@qq.com

The advent of the big data era has changed all aspects of economic and social life. The demand for big data talents has been increasing day by day. Also, the huge commercial value is contained in the explosive data generated by industrial application systems. In this context, all sectors of society have been aware the profound impact of big data on the future economic development and technological advancement (Ramaswamy, Rastogi & Shim, 2000). Therefore, big data has been valued by the state and enterprises. In 2014, big data was listed in the Chinese government report as a national key support industry. In 2015, the Standing Committee of the State Council adopted the *Action Plan on Promoting the Development of Big Data*, clearly proposing to promote the development and sharing of domestic big data. The statistical results show that in 2015, China's big data market exceeded 10 billion yuan, and by 2017, it has reached 20 billion yuan just within two years.

The colleges and universities are responsible for cultivating big data talents. They should not only focus on the teaching of theoretical knowledge, but also emphasize on training students' hands-on practical ability, since big data is a partial application discipline. Besides, the big data has emerged only for a short time, so the construction of the laboratory is in its infancy, and the laboratory construction of many universities cannot meet the training needs of big data disciplines, which has brought about unfavourable factors for the cultivation of practical and innovative talents required by the enterprises (Lynch, 2008). For this, this paper firstly analyses the current construction status of computer education laboratories, and then states how to build high-quality big data laboratory form the following aspects: the application field of big data, the core needs of enterprises as the guiding direction, the composition of big data lab, and the teaching practice of computer labs, etc. This shall build a platform for big data practice for teachers and students, and thus enhance the competitiveness of big data talents.

Current status of computer laboratory teaching

Big data lab infrastructure

With the explosive growth of cloud computing and mobile Internet in social and economic life, Big Data has attracted more attention and become an important information wealth. Meanwhile, there exist huge differences in the processing and management between massive data and the conventional data, so the construction of big data-based laboratories has started in the computer speciality of various universities (Jagadish *et al.*, 2014). However, the construction of big data labs is significantly different from traditional computer labs. In terms of lab construction content, in addition to the requirements of conventional hardware devices, sufficient powerful software is needed to support the analysis and application of large-scale data. In the laboratory teaching concept, the laboratory needs to provide strong industry-university-research support, study real business cases, and analyse different data of multiple industries. It's applied in subject research, project development and implementation, and provides massive industrial data for research and teaching (Picciano, 2012). A list of equipment is required for a basic big data lab as follows:

Table 1
Basic Big Data Library Hardware List

Product	Quantity
Bigdata server	10
Key value machine	1
Switch	2
Data node server	2
Data analytic server	2
Data management server	2

Problems in college big data labs

Due to the short construction period of colleges' big data labs and the lack of uniform standards, there exist the following problems: 1) There is a lack of basic experimental equipment; it requires high big data experimental hardware configuration, high CPU, and large memory clusters for data calculation, but many schools of computing in colleges cannot afford high costs to build a big data cluster experimental environment. 2) The basic experimental environment is weak. The deployment, configuration and maintenance of big data distributed software systems need higher costs. Also, they are easy to be modified during the use process, and the experimental environment is not easy to maintain. The big data distributed software system is difficult to install and configure, and the experimental environment is easily destroyed, so professional laboratory teachers are required to maintain the software hardware environment. But these professional teachers also bring certain labour costs. 3) The experimental teaching standards are missing. Big data hardware and software updates are fast and iterative. From Hadoop to tensor flow (Daniel, 2015) over the years, the higher requirements for textbooks and teachers' knowledge update have been proposed, and also some problems have been caused, such as the great difficulty in making experiment, various experimental changes, and the lack of uniform standards in experiment teaching. 4) The project practical training cases are scarce. Because the real big data is mostly owned by the enterprise, it is usually more expensive to obtain such data, which results in insufficient experimental data and further restricts the experiment. So, the university lacks big data project training, making it difficult to cultivate the professional project ability required for practical talents (Reyes, 2015). 5) The curriculum system is to be perfected. Big data teaching and research is difficult at the beginning, with high mathematics requirements; the big data speciality of different universities are in their infancy, and the talent training curriculum system lacks systematicity; the big data teaching and research resources are scarce, and professional teachers for guiding experimental environment are insufficient (Tulasi, 2014).

Laboratory Construction Management

Laboratory requirements in the context of big data

In the context of big data, the traditional teaching and training model cannot meet the requirements of enterprises and society for talents. During the teaching process of computer profession, it is necessary to further increase the theory and practice of big data and improve students' data use and innovation ability. For this, Big

Data lab is required to implement an open, mobile, joint mechanism, and improve data processing capabilities through integrated application of in-memory database, library warehouse, distributed CPU + GPU cluster solution. Besides, it should complete a series of complete closed-loop big data processes, including data collection, storage, processing, and application, while meeting the massive data continuous processing capability and performance index such as peak factor etc. (Sugimoto, Ding & Thelwall, 2013). Especially for data storage, it is necessary to calculate the storage capacity according to the data analysis of the PB level, which is taken as the design standard of platform, equipped with the corresponding high-speed network, and mass storage device (Johnson, 2017).

In addition, the construction of big data experiments needs to follow the following principles: 1) Theory is combined with practice. Practice is the core competitiveness of big data talents. Compared with the construction of big data discipline that started late in China, it should focus more on cultivating students' practical ability and innovating in practice. 2) Adapt to the needs of market and enterprises. The training of big data talents needs to take the actual needs of social posts as the fundamental starting point, deeply understand the needs for big data talents in all walks of life and enhance students' ability to learn in market segments. 3) Industry-university-research integration. In the construction process of a big data laboratory, we will learn from the ideas of industry-university-research cooperation to better cooperate with enterprises and combine education, scientific research and production. While improving students' practical ability, the research results can also create a certain income for the laboratory and further reduce the cost of big data management (Kim & Cooke, 2017).

Laboratory construction plan

Based on the section 3.1, the construction of a big data lab needs to follow the construction plan: 1) Lab hardware consists of windows computer and mac for students, virtualization system hardware configuration, experimental data storage, hadoop server cluster, data computing hardware resources. 2) Software resources. A visual big data platform should be established, e.g., using the Unity3D engine to achieve high-dimensional simulation of 3D interface, and visually display the results of data analysis in WYSIWYG. For the multi-source and easy-to-construct mass data, the platform, through data processing, storage management, visual interaction analysis and other technologies, achieves graphical data query, visual association analysis, evidence chain and intelligence cues exploration. Thus, it can visually depict the state machine for data analysis, understand overall operating conditions, analyse data fluctuation reasons, make real-time detection of data changes, conduct an accurate judgment and measurement of data running status, and improve students' ability to understand data. 3) Curriculum system. According to different majors and directions, different course learning paths are set. By providing the teaching materials such as experimental instruction manuals, video teaching cases, resource documents, experimental codes, and homework demonstrations etc. for each course, an integrated teaching system can be established for big data teaching, in which the students can make the experimental operations in multiple levels, including on-line learning, principle demonstration, comprehensive analysis, and independent design etc. 4) Construction of the teaching staff. The stable and professional instructors capable of big data

analysis, organized, and development should be trained, and hardware engineers be trained to regularly manage and maintain hardware devices in big data labs.

Laboratory management plan

The management of big data labs can be carried out in the following dimensions. 1) Data security is a fundamental part of big data lab management. Since the data of the big data lab is mostly from the enterprise, it has high commercial value and confidentiality, so it is necessary to strengthen and guarantee the data security. In terms of hardware, the access control system is used to replace the traditional door lock system. All personnel need to swipe their cards in and out, and the video surveillance system is used to control and supervise the laboratory personnel. Besides, the remote computer control and alarm device are installed, and then it can quickly alarm to protect the laboratory when the monitoring personnel finds something abnormal with the laboratory. In terms of software, the laboratory computer should periodically update the password, and a multi-password + token verification mechanism are established to control the use of personnel. During use of data, different types of data are encrypted so that teachers and students need to obtain an account password through authentication. It is also forbidden to use USB, hard disk and other devices to copy data. 2) Network-based management of laboratory equipment. The big data lab involves a large number of experimental equipment, which requires a strict management system to ensure the full-network openness of laboratory equipment in storage, use, and maintenance. An equipment reservation system should be built to ensure efficient use of large computing devices. At the same time, the analysis tool is used to make regular statistical analysis for the utilization rate of the equipment, and provide effective data for future instrument procurement. 3) Lab talent management upgrade. With reference to the principle of industry-university-research integration, a group of high-level lab management personnel should be trained to ensure safe and efficient operation of the experiment. Lab technicians can use the data system of the laboratory to assist researchers in technical development and instrument upgrade.

Effect analysis of lab construction plan

One university in Sichuan Province was selected to apply the plan proposed above in this paper. According to this plan, 15 sets of 32-core 64G servers are configured, and connected through a FC switch and a disk array, which can support the operation of 300 virtual machines (480 core CPU, 960G memory, and 15T hard disk space) at the same time. On this basis, the Hadoop cluster is built, and the master node and the slave node are built. The cluster can complete the Hadoop experiment, MapReduce data analysis, HBase massive data storage, and the HIVE data warehouse etc. The entire virtual machine is configured as shown in the Table 2.

Table 2
Virtual Machine Setting

Hardware	Spec
cpu	480 Core
Memory	960G
Disk1(sda)	10T
Disk2(sdb)	5T
Virtual network1	1000M
Virtual network2	1000M

In the actual teaching, the lab mainly undertakes the experiments of undergraduate and graduate courses such as data mining and data warehouse, advanced database management, Hadoop practice, MapReduce practice, probability theory and statistics, big data algorithm, etc. The main achievements of the lab during operation are as follows:

1) Through the experimental part of the above course, the undergraduate students in the lower grades can fully understand the process of data management, big data collection, analysis, cleaning and integration, and the actual ability of big data is enhanced. Besides, senior students and graduate students can also complete the construction of a large-scale data processing platform in the laboratory, and make processing of real big data in enterprises. 2) The built big data lab can accommodate more than 200 students to conduct experiments simultaneously, fully meeting the students' internship training needs. Because the laboratory has reached the enterprise level standard of big data production environment in terms of comprehensive functions, hardware equipment and personnel professionalization, it can take over some big data outsourcing work of some small and micro enterprises, so as to achieve a win-win situation between student practice and business needs. 3) The built laboratory is significantly conducive to improving the teaching and research capabilities of the university's big data discipline, and promoting the discipline construction of the related majors. The employment rate of undergraduates and postgraduates in the direction of big data has been significantly improved. According to the feedbacks of enterprises, this university's big data majors have a better advantage in terms of practice and innovation ability than similar universities.

Conclusions

The rapid development of Mobile Internet and the Internet of Things has brought about exponential growth of data, and the enterprises' demand for big data talents has becoming more urgent. As the main source of training for social talents, colleges and universities undertakes the teaching of basic big data talents. However, it's generally reflected that the big data-related talents trained in colleges and universities do not meet the requirements of enterprises. In view of the current insufficient support for big data in college teaching, a big data lab construction and management plan based on industry-university-research integration was proposed in this paper, in order to improve students' ability to innovate big data applications. To this end, this paper, based on the goal of applied talent training, analyses the core work of big data analysis laboratory construction and the related specific content. The plan proposed in this paper has been adopted and implemented by one selected university. After the laboratory was built, the learning effects of different grade students were compared in this paper. It's found that this plan can effectively improve the students' ability to analyse and apply data, improve the employment rate of graduates majoring in big data, and the satisfaction of enterprises; especially the graduate students of this university in big data direction have showed a more comprehensive practical ability of big data application and received favourable comments from employers. The results show that the proposed plan has strong effectiveness and practicability, and it can provide strong support and useful reference for big data education and scientific research in colleges and universities.

References

- Daniel, B. (2015). Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904–920. <http://dx.doi.org/10.1111/bjet.12230>
- Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., & Ramakrishnan, R., Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, 57(7), 86-94. <http://dx.doi.org/10.1145/2611567>
- Johnson, V. (2017). Leveraging technical library expertise for big data management. *Journal of the Australian Library and Information Association*, 66(3), 271-286. <http://dx.doi.org/10.1080/24750158.2017.1356982>
- Kim, Y., & Cooke, L. (2017). Big data analysis of public library operations and services by using the Chernoff face method. *Journal of Documentation*, 73(3). <http://dx.doi.org/10.1108/JD-08-2016-0098>
- Lynch, C. (2008). Big data: How do your data grow?. *Nature*, 455(7209), 28-29. <http://dx.doi.org/10.1038/455028a>
- Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, 16(4), 9-20. <http://dx.doi.org/10.24059/olj.v16i3.267>
- Ramaswamy, S., Rastogi, R., & Shim, K. (2000). Efficient algorithms for mining outliers from large data sets. *ACM Sigmod International Conference on Management of Data*. 427-438. <http://dx.doi.org/10.1145/342009.335437>
- Reyes, J. A. (2015). The skinny on big data in education: Learning analytics simplified. *Techtrends*, 59(2), 75-80. <http://dx.doi.org/10.1007/s11528-015-0842-1>
- Sugimoto, C. R., Ding, Y., & Thelwall, M. (2013). Library and information science in the big data era: Funding, projects, and future [a panel proposal]. *Proceedings of the American Society for Information Science & Technology*, 49(1), 1-3. <http://dx.doi.org/10.1002/meet.14504901187>
- Tulasi, B. (2014). Significance of big data and analytics in higher education. *International Journal of Computer Applications*, 68(14), 21-23. <http://dx.doi.org/10.5120/11648-7142>