

## Data Science and Big Data Technology Professional Talent Demand and Training System Construction

Feng Liu<sup>1, a</sup>, Weiwei Guo<sup>1, b</sup>, Hui Wang<sup>1, c</sup> and Xiaoyan Li<sup>1, d</sup>

<sup>1</sup> Heilongjiang University of Technology, JiXi, 158100, China

<sup>a</sup>liufeng8038@163.com, <sup>b</sup>gwwguoweimei@163.com

**Keywords:** Data Science and Big Data Technology; Talent Cultivation; Practical Ability

**Abstract.** Data science and big data technology are emerging interdisciplinary subjects. They are facing problems such as insufficient training of talents and industry needs, inaccurate curriculum systems, and inaccurate personnel training. Based on these many problems, combined with the needs of new engineering talents training, combined with the regional characteristics of the Northeast region and the development of the big data industry, the article has in-depth research on the cooperation model of cooperation with enterprises, from the talent training objectives and Starting from the training of standards, we have established a training system for talents in data science and big data technology to improve the quality of personnel training.

### Introduction

"Data Science and Big Data Technology" is an emerging interdisciplinary subject that integrates computer science, mathematical statistics and information systems, and collects, organizes, analyzes and processes data, and uses relevant knowledge in the industry. Analyze and manage to get valuable information. With the continuous development of information technology, the era of big data and artificial intelligence is getting closer and closer to us, and data science has become the core technology of social development. In 2015, the national policy strongly supported the development of the information technology service industry. The State Council issued the "Opinions on Promoting the Innovation and Development of Cloud Computing and Cultivating New Formats of the Information Industry", and the Ministry of Industry and Information Technology issued the "Guidelines for the Construction of a Comprehensive Standardization System for Cloud Computing". These policies provide a good environment for promoting the rapid growth of the information technology service industry. In 2016, with the gradual implementation of the above policies, relevant supporting policies and local policies have been introduced one after another, which will play a positive role in promoting the development of the information technology service industry. Especially in the emerging service sector, under the guidance of national policies, local governments will speed up the introduction of corresponding policies and measures in conjunction with their respective advantages and needs, and increase support for cloud computing and big data services. With the continuous enhancement of the role of information technology services in innovation and support services for economic and social development, deepening industry applications and improving information security will remain the focus of future policy attention. In 2017, the Ministry of Industry and Information Technology formulated the "Development Plan for Big Data Industry", which strongly supported universities to explore the establishment of professional talents and cross-border compound talent training mechanisms in the field of big data. In 2016, three national 985 colleges and universities established undergraduate majors in data science and big data technology. In 2017, the Ministry of Education approved the addition of data science and big data technology to 32 institutions. In 2018, the Ministry of Education announced new data. 248 colleges of science and big data technology; as of February 2019, there are 283 undergraduate colleges in China that offer data science and big data technology<sup>[1-3]</sup>.

In Heilongjiang Province's demand list of talent shortages in 2017-2018, data analysis and big data application talents have reached a scarcity level. According to the Heilongjiang Provincial Party Committee Organization Department and Heilongjiang Provincial Human Resources and Social

Security Department, July 2016 In the Heilongjiang Province Key Industries (Industry) Urgently Needed Talents Catalogue (2017-2018), the gap in the new generation of information technology professionals in Heilongjiang Province is 71%. Among them, Java software engineers have great needs in the province, not only in relatively developed cities such as Harbin and Daqing, but also in the eastern regions of Heilongjiang, including Mudanjiang, Jixi, Qitaihe and Shuangyashan. Lack of relevant talent. Among them, data analysis engineers, demand analysis engineers, and WEB front-end development engineers are particularly scarce in Harbin, Mudanjiang. How to use existing resources to cultivate big data talents for localities is a problem that local universities must study.

### **Talent demand characteristics**

"Big Data" is one of the hot words in the IT industry. The use of big data business value has gradually become the focus of profits that industry professionals are vying for. The application of big data is gradually expanding from the Internet and finance to education and government affairs. The major social and economic fields, such as transportation, logistics, and medical health, have profoundly affected China's future social and economic development. The global big data technology and services market has revenues of more than \$23.8 billion in 2016; according to McKinsey's forecast, by 2020, US big data can create a value of \$380 billion to \$ 790 billion [4-5].

In the "Big Data" report, McKinsey pointed out that the shortage of big data talents will seriously restrict the development of big data industry, especially the professionals in statistics and machine learning, and the talents who know how to use big data to run enterprise management and analysis. In the US market alone, the talent gap for big data talent and senior analytics experts in 2018 will be as high as 190,000. In addition, US companies need 1.5 million big data-related management talents who can ask the right questions and use big data analysis results. China's big data research and big data application talents demand and job seeks have a large gap for several consecutive years. The rapid development of the new economy has led major companies to develop big data businesses and have high demand for professional big data talents. At present, the market demand for big data talents is characterized by large demand, high salary levels, and an upward trend. From the domestic and international market demand and development situation, the most popular positions in the big data industry are data scientists, data engineers and data analysts. The training and employment direction of big data talents are gradually developing in the following four aspects [6].

1. Research-oriented transformation to applied talents. At present, many big data-related big data technologies have begun to be applied. Different from the initial big data, a large number of researchers are needed in the early days of big data. At the current stage, the demand is mainly based on applied talents. Big data technology needs to be penetrated into various fields.

2. The focus of demand shifts to big data analysts. In the big data application phase, data analysts are the focus of demand, because data analysis is an important way to reflect the value of data, and traditional industries pay more attention to the field of big data analysis. The demand for data analysis will drive the needs of big data development and maintenance personnel, as well as the demand for artificial intelligence and IoT talent.

3. The development trend of "Big Data + Industry" is obvious. The initial stage of big data is mainly to train big data technology talents, including platform construction, big data framework design, algorithm design, data analysis, and results visualization. With the deep integration of big data and traditional industries, big data Development will be closely integrated with the industry. Big data talents with industry background will be the focus of enterprise needs. Having industry knowledge will be an important part of big data application.

4. Big data and artificial intelligence, Internet of Things and other technologies will be deeply integrated. With the development of 5G communication networks, the future of Internet of Things, mobile Internet, big data, artificial intelligence, and traditional industries will be deeply integrated. Therefore, for the employed, it is necessary to understand the talent needs of the future market and master relevant technical knowledge.

## Training Objectives

In the context of new engineering, the design of data science and big data technology professionals training needs to respond to change and shape the future as the construction concept. The main path of inheritance and innovation cross, integration, coordination and sharing is to cultivate diversification and innovation. Excellent engineering talents provide intellectual and talent support for the future. Data science and big data technology major, cultivate a good mathematical foundation and logical thinking ability, have high information literacy, master the basic theories, methods and skills of computer science, big data science and information technology, and receive systematic scientific research training. It includes basic knowledge of mathematics, statistics, computer and other disciplines for big data applications, data modeling, efficient analysis and processing, basic theories, basic methods and basic skills of statistical inference. Knowledge of big data in applied fields such as natural sciences and social sciences, strong professional ability and good foreign language application ability, competent for data analysis and mining algorithm research and big data system development, research-oriented, technical talents, advanced compound , application-oriented talent [7].

Taking the talent demand of the big data industry as the starting point, combining the actual school running situation of the colleges and universities and the characteristics of the local economic development of Jixi, focusing on cultivating engineering application talents for the big data engineering and information technology industry, highlighting the school-enterprise cooperation and the integration of production and education. "The characteristics of running a school, together with Zhongke Remote Sensing, Dana Education Group, Beijing Huayu Xingye Oracle Data Center, Lianchuang Group, etc. to carry out the top-level design of big data professional construction, from talent training objectives, curriculum system, discipline team building, practice platform Long-term, all-factor planning and construction of the major in the five major areas of scientific research construction and service local construction. On the basis of repeated investigations and arguments, we will identify the orientation of running schools, relying on the development strategy of "two black and one green" industrial clusters in Jixi City, and positioning the talent training target as the basis for the social and economic transformation and development of Jixi City, focusing on training "Big Data + Green Agriculture "Big Data + Coal Industry" "Big Data + Smart City" is a shortage of big data application talents.

## Training specifications

The "Data Science and Big Data Technology" major is an interdisciplinary subject involving applied mathematics, statistics, and computer science. Therefore, the major emphasizes the development of big data talents with multidisciplinary skills. The graduates of this major should have the following abilities [8]:

### 1. Professional quality

(1) Have good professional ethics and professional ethics, and have a strong sense of organization and teamwork;

(2) Understand the laws and regulations related to design and research and development related to big data occupation and big data industry;

(3) Having the ability to adapt to development and a correct understanding of lifelong learning and continuous learning ability;

(4) Have a certain sense of innovation, innovation and innovation, and have certain humanities and artistic accomplishments;

### 2. Professional ability

(1) Have the ability to design and implement engineering experiments, and be able to analyze the experimental results; have the basic ability to comprehensively apply the theoretical and technical means to analyze and solve big data engineering problems;

(2) It has the ability to design systems and processes by comprehensive application of theory and technical means. The design process can comprehensively consider economic, environmental, legal, security, ethical and other constraints;

(3) Master solid data science and big data technology basic knowledge and basic theoretical knowledge of the major, have a systematic engineering practice learning experience, understand the frontier development status and trends of the major, and have the ability to solve complex engineering problems;

(4) Have the mathematics knowledge required to work in this profession, and be able to understand and flexibly apply the model in data science;

(5) Master the basic methods of data inquiry, literature retrieval and the use of modern information technology to obtain relevant information; have certain scientific research and practical work ability;

### 3. Social ability

(1) Have certain organizational management ability, strong ability to express and interpersonal skills, and the ability to play a role in the team;

(2) Engineering practice ability: personnel management, time management, technical management, process management and other capabilities;

(3) Ability to communicate, compete and cooperate with international vision and cross-culturalism.

## Curriculum Structure

Based on the engineering practice education for new engineering, we will carefully study and analyze the new national economy and new industries, combine the needs of talent cultivation in the northeast region, and jointly develop advantages with cooperative enterprises, jointly build a talent training system, curriculum system, and strengthen curriculum resources. Construction, including the teaching resources of the curriculum and the data resources of the students in the practice link, strengthen the close integration of the theory and practice content, combine the industry data resources provided by the enterprise, and carry out the training mode of “big data + industry” to make the work tasks of the industry The curriculum is designed to enable students to graduate and work, and to seamlessly connect with the company. Strengthen the practice link, establish a curriculum system with distinctive characteristics, dynamic adjustment and reasonable structure, so that talent cultivation can closely follow the social needs, and the curriculum setting reflects practicality [9].

Table 1 Main Course Semester Assignment Table

Serial number	Course Title	Hours/credits	Self-study	Semester
1	Introduction to Big Data Science and Technology	15/1	10	1
2	Python programming	56/4	20	2
3	Probability Theory and Mathematical Statistics	56/4	20	3
4	Object-Oriented Programming	56/4	20	3
5	Matlab programming	42/3	20	2
6	Computer composition principle	56/4	24	4
7	computer network	56/4	24	3
8	Data mining and analysis	48/3	14	5
9	data visualization	48/3	14	6
10	Big data development technology (Hadoop/Spark)	36/2.5	10	6
11	Cloud computing and data center	24/2	8	6

### 1. Talent training system

This professional course consists of four parts: 1) general education courses, including general education courses, ideological and political courses, military theory courses and level-level courses; 2) subject courses, subject platform courses and majors The core curriculum consists of two parts; 3) multiple training elective courses, including professional application elective courses, professional academic elective courses, compound application elective courses, innovation and entrepreneurship

classes, public elective courses and second class courses; 4) practice and Graduate design (thesis) link. The implementation of multi-category talent training is specialized application, professional academic, entrepreneurial innovation, and composite application.

## 2. Main course

The main courses offered by Data Science and Big Data Technology are: Introduction to Big Data Science and Technology, Python Programming, Probability Theory and Mathematical Statistics, Object-Oriented Programming, Matlab Programming, Computer Composition Principles, Computer Networks, Data Mining and Analysis, Courses in data visualization, big data development technology (Hadoop/Spark), cloud computing and data centers. The course hours and corresponding semester are shown in Table 1.

## Practice Teaching

The data science and big data technology talents are mainly applied, and the practical teaching links make an important part of the whole talent training process. In the actual teaching operation process, the practical environment and practical data resources are scarce. The teacher ignores the importance of the practical content in the teaching process, and the practice time is too small, which leads to the teaching effect is not obvious. In order to achieve better teaching results and personnel training purposes, we should strengthen the understanding of practical content. Therefore, the development of personnel training in data science and big data technology must improve the practical teaching links and teaching, and pay attention to the cultivation of students' practical ability<sup>[10]</sup>.

### 1. Practice platform construction

The construction of the practical platform is mainly to build an experimental training platform with the Beijing Huayu Xingye Oracle Data Center. The entire platform consists of three parts, namely the teaching resource platform, the training project and the data resource management platform and monitoring platform.

The teaching resource platform includes Linux operating system, distributed file system, distributed computing framework, distributed database, data warehouse HIVE, Python language programming, data analysis and data mining, distributed computing framework Spark, data collection and other big data. Professional theory class and corresponding 15 experimental packages (experimental data, experimental manual, experimental video)

The training project and data resource management platform includes all the processing flow from data collection, data cleaning, data set management, data processing to data mining, and supports the whole life cycle of the training project and the management of intermediate data products. The technical platform for the training of teaching. The training program also includes the movie website customer value analysis case, the online and consumer relevance analysis case and the customer's preference movie recommendation case. The real project can effectively improve the students' practical ability through the training of these items.

The monitoring platform includes big data hardware monitoring and big data software monitoring. During the teaching operation, the hardware device running status can be monitored in real time and alarm linkage information can be generated, and the running status of big data software can be monitored in real time and alarm linkage information can be generated. Through the construction of the big data platform, the rapid integration of internal data and external data is realized, which brings excellent data calculation and analysis and mining capabilities, and can realize functions of data management, data standard management, data quality management, and data security.

### 2. Practical content

The practical teaching content is guided by work tasks and constructs a practical teaching system. Through the in-class experiments and off-campus internship training, students will develop practical hands-on ability, industry adaptability and innovation ability. The school cooperates with the company to provide professional lectures for the students in the school, and guides the students' experiments throughout the course to enhance the students' practical ability. The enterprise sets up real project cases in the enterprise for students, through the experiment, curriculum design, time content of the enterprise internship series, builds a corporate practice platform for students, and

combines the training of industry data resources to effectively ensure that the students who are trained can meet the enterprises of the big data era. Requirements for talent.

## Conclusion

Based on the needs of data science and big data technology talents, this paper combines industry big data and real projects of enterprises to build talent training and curriculum system, and explores a training model suitable for big data professionals in Northeast China. The training system is constantly in operation. Perfecting and dynamic improvement, so as to continuously improve the practical and innovative ability of students in data science and big data technology, it is not only the requirements of economic and social development for the quality of talents, but also provides some valuable references for the training of big data talents.

## Acknowledgements

This work was supported by Heilongjiang Provincial Department of Education Project "Research and Practice of Problem-Driven Teaching Model Based on Computational Thinking in Software Technology Professional Language Courses" (SJGY20170323).

## References

- [1] Ou Weihua. Research on the Construction of Professional Practice Teaching System of "Data Science and Big Data Technology"[J]. Software Guide, 2018, v.17; No.187(05):111-113.
- [2] Hou Xiaofang, Li Wei, Li Jie. Research on the Construction of Big Data Foundation Course[J]. Journal of Beihua University of Aeronautics and Astronautics, 2018, v.28; No.118(03):46-48.
- [3] Dong Zhaowei, Wang Suzhen, Sun Jieli. Research on the training of big data talents in the field of management based on the new engineering concept[J]. Journal of Hebei University of Economics and Technology (Comprehensive Edition), 2018, v.18; No.70(02):77-81.
- [4] Hao Shuixia, Guo Yunxia. A New Exploration of the Training Mode of Information and Computing Science Talents in the Age of Big Data[J]. Journal of Jiangsu Normal University(Natural Science), 2016, 34(2): 72-75.
- [5] Yang Hong, Li Zhiyao, Zhang Zhiqiang. Exploration and Practice of the Training System of Practical Ability of Data Science and Big Data Technology[J]. Journal of Chengdu University (Social Science Edition), 2018, No.177(03):111-117.
- [6] Chen Xin, Zhou Longfu, Cao Yaoqin. Exploring the Construction of Applied Undergraduate Data Science and Big Data Technology Talents Training Program[J]. Modern Industrial Economy and Informatization, 2017.
- [7] Liu Qiming. Application of Modern Apprenticeship System in Talent Cultivation of Big Data Technology and Application Specialty[J]. Computer Fan, 2018, No.94(05):160.
- [8] Zhang Zhiwei, Fang Aidong, Cui Lin. Research on the Construction of Big Data Major in the Background of New Engineering[J]. Journal of Panzhihua University, 2018, 35(05): 113-117.
- [9] Tao Wan, Zhang Wei, Yang Dan. Reform of Big Data Technology and Application Direction of Information Management Major under the Background of New Engineering[J]. Journal of Weinan Teachers College, 2018, v.33; No.258(10):42-48.
- [10] Zhang Zhigang, Zhao Liqiang, Gao Tianying. Research on Computer Science and Technology Professional Training Program for Big Data Application[J]. Tianjin Education, 2017(1): 32-33.