

Research on the Learning Platform of Big Data Management and Application Specialty based on Ability of Practice and Innovation

Xiaogao Yu^{a,*}, Xingjun Liu^b, Xiaobo Zhu^c, Ruiqing Peng^d

School of Information Management and Statistics, Hubei University of Economics, Wuhan, China

^a 469554298@qq.com, ^b tecom_sam@163.com, ^c 470721862@qq.com, ^d 670777852@qq.com

*Corresponding author

Keywords: Practice, Innovation, Specialty, Learning platform

Abstract: In order to improve talents training quality of big data management and application specialty and cultivate talents with practical and innovative ability, big data technology and network technology are applied to professional talent training. The learning platform of big data management and application is designed, the space-time limit has been broken, high quality learning resources are dynamically tracked and collected, a knowledge system that meets the needs of society is established, personalized learning space with practice and innovation ability is constructed. Through application analysis, the professional learning platform has effectively improved the teaching quality and talent training quality.

1. Introduction

Today, data is growing exponentially and the age of big data has arrived^[1], big data strategy has become the focus of countries. A series of plans, policies and reports such as big data research and development plan, technologies and policies supporting data-driven innovation and big data: seizing opportunities and guarding values were officially released by the United States one after another, other countries have also issued a series of strategies, such as the industrial strategy: building a UK adapted to the future issued by the UK, and so on^[2]. China attaches great importance to the big data strategy and has successively issued the big data industry development plan (2016-2020), the big data industry development plan in the 14th five year plan, the digital economy development plan in the 14th five year plan and other documents^[3].

In foreign countries, big data specialty was born in the form of data science specialty, for example, North Carolina State University in the United States established a master's degree in data analysis in 2007, the University of California Berkeley opened an introduction course in Data Science in 2011, and New York University officially established a master's degree in Data Science in 2013. The "big data management and Application Specialty " of undergraduate education in Chinese Colleges and universities is still in its infancy and development stage. So far, the Ministry of education has approved 140 colleges and universities to set up the specialty of big data management and application, the construction of this specialty started late and has a weak foundation, accounting for a small proportion of the total number of colleges and universities in China^[4]. There are some problems in the teaching process of this specialty, such as unclear enterprise needs, limited experimental and training gains, unclear industry status, underdeveloped laboratory construction and unclear employment direction of students^[5]. In order to solve the existing problems and improve the training quality of big data professionals, this paper designs a learning platform of big data management and application specialty based on ability of practice and innovation. The platform organically integrates professional basic knowledge and ability training, and focuses on students' practical ability and innovation ability.

2. Professional Training Requirements

This specialty is to cultivate professionals who master the basic knowledge of big data

management and application and the ability to solve practical problems.

Professional basic knowledge includes big data, data processing and data management. The basic knowledge of big data is taught from the aspects of professional introduction, program design, data structure, operating system, database system and application, computer network, management information system, information security, etc. Data processing knowledge is taught from python programming, data acquisition, data cleaning, ETL, distributed computing, data storage, API and other aspects. Data management knowledge is taught from the aspects of data governance system, data quality, data risk, metadata management, master data management, data architecture, data control, etc.

The cultivation of professional talents' ability to solve practical problems is carried out from the aspects of data modelling and analysis, data display, industry application and so on. The training of data modelling and analysis ability is carried out from the aspects of data warehouse, data mining, data analysis, data modelling and so on. The cultivation of data presentation ability is carried out from the aspects of big data visualization, data warehouse, graph database, correlation analysis, data mining and so on. Industry applications cover a wide range of industries. According to the characteristics of different students, it provides sufficient personalized space for choosing industries, such as finance, production, manufacturing, circulation, media, e-commerce, government, etc.

3. Learning Platform Design

3.1. Architecture Design

The learning platform of big data management and application specialty is constructed by big data technology and Internet technology. It collects high-quality teaching resources at home and abroad, cleans, converts, stores, processes and mines them, refines various resources required for the training of professionals in big data management and application, realizes the teaching space and personalized learning space of big data management and application^[6,7], recommends high-quality teaching resources to teachers and students^[8], and dynamically monitors the training of professionals as a whole. Its architecture is shown in Figure 1.

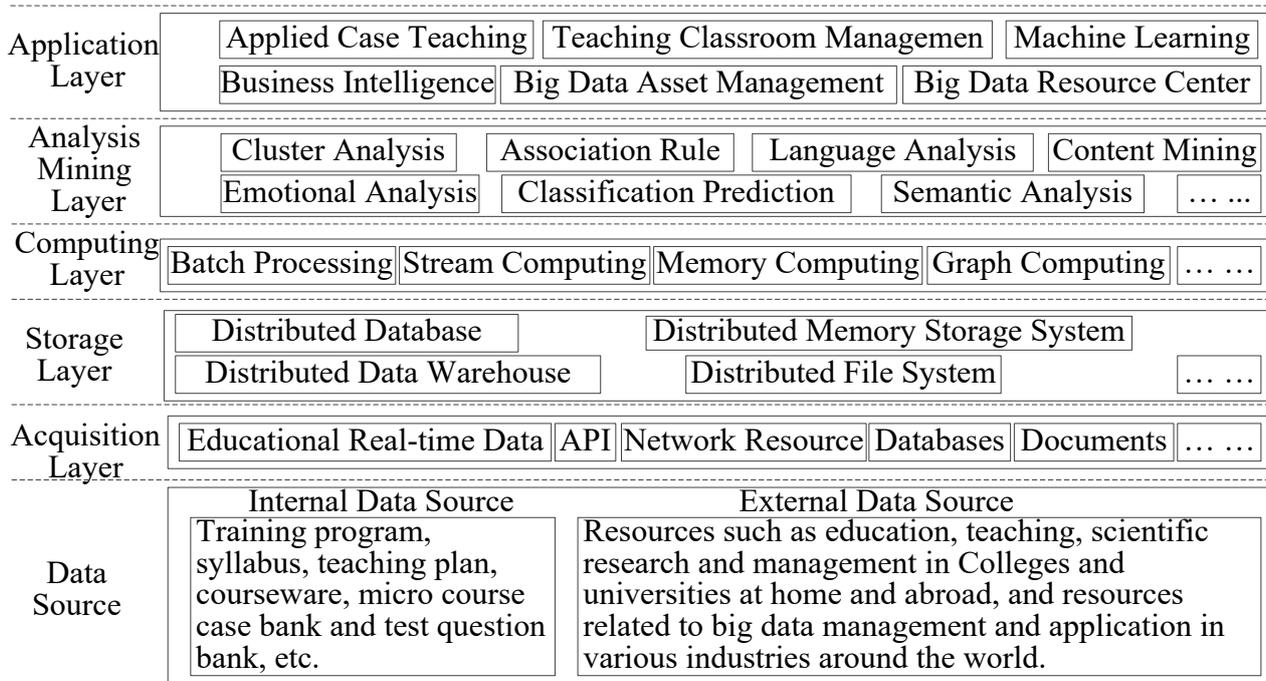


Figure 1 Big data management and application learning platform architecture

- Data Source

The data source is composed of external data source and internal data source. The external data source uses Internet and other tools to collect domestic and foreign college education, teaching and

other resources and high-quality resources related to the cultivation of professional talents in various industries; Internal data sources collect professional education and teaching resources in Colleges and universities.

- Acquisition Layer

In order to meet the needs of timeliness, continuity, integrity and naturalness of resources required for professional teaching, the acquisition layer can be divided into two forms according to the data acquisition method: (1) Batch collection. Copy and import the collected professional data files, data records, professional teaching resource files, etc. (2) Real time acquisition. Adopt intelligent technology to capture and transmit various professional data resources in real time.

- Storage Layer

This layer adopts distributed storage architecture to solve the problem of big data storage. The storage methods mainly include file storage, data warehouse, database, memory storage, etc.

- Computing Layer

This layer provides efficient computing functions for big data processing, mainly including batch processing, stream computing, memory computing, graph computing and other computing engines^[9,10].

- Analysis Mining layer

The focus of big data analysis is to use big data technology to solve the problems existing in the training of big data management and application professionals; The focus of big data mining is to design new algorithms and build new models. A new method is designed to solve the problem of talent training of big data management and application specialty based on clustering, association rules, time series analysis, prediction, regression analysis and other technologies and combined with the characteristics of talent training of big data management and application majors.

- Application Layer

This layer applies big data analysis and mining technology to give full play to the advantages of the Internet. This layer is mainly reflected in the following six aspects: application case teaching, teaching classroom management, machine learning, business intelligence, big data asset management and big data resource center.

3.2. Functional Design

3.2.1. Applied Case Teaching

The learning platform carries out auxiliary teaching, course case, big data analysis case teaching and business analysis case teaching based on the course case base and industry case base. It can integrate theory with practice and cultivate students' practical and innovative ability. The course case base includes professional course knowledge such as big data acquisition technology, cleaning and processing technology, analysis and mining technology, visualization technology and big data development technology^[11]. The cases in the industry case base are abstracted and summarized from the actual combat projects in the industry, these cases are combined with the big data resource center to realize the actual project application environment and cases in the whole process, and can complete the practical application of algorithm models such as data acquisition, cleaning, processing and transformation, visual mapping, data label portrait, analysis and prediction.

3.2.2. Teaching Classroom Management

This function realizes the professional course management, adopts online examination, makes excellent courses, provides practical cases, and guides students to carry out vocational certification and employment practice.

3.2.3. Machine Learning

It supports the mainstream deep learning framework, has rich algorithm components, and realizes visual modelling and real-time task monitoring. It can support tensorflow, caffe, pytorch and other mainstream deep learning frameworks. It has built-in rich machine learning algorithm components such as feature engineering, statistical analysis, deep learning, text analysis and

machine learning. It can turn complex machine learning algorithms into visual algorithm components, greatly reduce the difficulty of students' learning, and realize real-time task monitoring functions such as node monitoring, online and offline management and scheduling management ^[12].

3.2.4. Business Intelligence

It supports tables, line charts, biaxial charts and other charts. It can generate reports with one click, realize intelligent recommendation, and help students clarify their ideas and master analysis methods in the face of sea data. It can intelligently answer the questions of different students and solve the confusion of students in learning professional knowledge. It has the function of intelligent mining and can support rich data links such as files, relational databases, non relational databases and big data platforms.

3.2.5. Big Data Asset Management

This function integrates the five systems needed to complete data governance, adopts the visual model design for the hierarchical management mode of data warehouse, realizes the development of multi-source fusion data, makes the ETL process transparent and visible by using the visualization technology, and improves the learning efficiency by one-stop rapid development.

- **Data Governance**

It has five systems: data access, data governance, data development, data assets and data services to complete data governance. Among them, the data access system completes two tasks: data exploration and data access; The data governance system completes five tasks: data standard management, data quality audit, data quality evaluation, metadata management and data warehouse model design; The data development system completes three tasks: offline data development, real-time data development and workflow scheduling; The data asset system completes five tasks: master data management, asset catalog management, data asset retrieval, label system construction and label application; The data service system completes five tasks: resource catalog preparation, data resource details, data resource subscription, data service encapsulation, service monitoring and flow restriction.

- **Visual Model Design**

It is a visual model design based on the data warehouse oriented hierarchical management mode, which can provide visual and hierarchical modelling tools. In the work area, users can follow the data warehouse construction process to carry out visual modelling by dragging components, and the blood relationship of data is displayed in the form of business attribute, which can accurately express the business logic.

- **Development of Multi-source Fusion Data**

During data processing, it is necessary to build real-time processing environment and offline data processing environment respectively, resulting in high environmental cost; Due to a large number of manually written codes, there are high technical requirements for personnel and can not be effectively managed, resulting in high maintenance costs in the later stage^[12]. These factors will inevitably make the cost of data processing high.

In order to reduce the cost of data processing and improve the teaching effect, the module uses the visual batch stream integrated data fusion technology to process massive data, and uses the pipeline to explore the interactive data, which can improve the speed of data processing, effectively improve the computing performance and reduce the difficulty of data processing. The module has the following four functions: (1) Data Management. This function can manage the version of data processing and has a release and online mechanism; It supports Oracle, mysql, Preto, hive and other multi-source data fusion computing and has a variety of development modes. It can meet the needs of different scenarios and different people. (2) Data Processing. This function supports the creation, maintenance and online debugging of multilingual scripts; It has the version management function of manageable and traceable data processing process. (3) data handling. It has the function of uniformly scheduling and managing workflow tasks; It has rich data processing components, which can quickly complete data cleaning, conversion, processing and other operations; It has a drag and drop process designer, which can easily drag and drop data processing components to define

complex workflow.

- Visualization Technology

ETL process can be transparent and visible through big data visualization technology. The following will be introduced from the perspectives of big data development and big data management. (1) Big data development perspective. The method of graphical process design is to establish mapping by creating source definition, target definition and transformation in the data flow designer to realize the graphical operation interface. Heterogeneous data can be easily integrated by dragging and dropping with the mouse; It has built-in rich data conversion function components, encapsulated a large number of function and parameter call rules, and the flexibility of calculation expression has been greatly improved; It has the function of automatic task scheduling, and can apply the set different task scheduling rules to the design, planning and execution scheduling of tasks. (2) big data management perspective. It can centralize the resources such as Spark, Hive, Python, MRjob, shell and so on, cooperate with the execution of environmental scheduling tasks, achieve unified management of scheduling tasks, and achieve global and transparent scheduling. It centrally monitors the execution log and task exceptions, visually presents the task execution results and exception alarm information, uniformly coordinates the inventory of abnormal problems in the production environment, and realizes the unified monitoring of task nodes.

- One Stop Rapid Development

In order to improve students' learning efficiency, one-stop rapid development is realized around the methodology of data asset management. It has a one-stop development system that runs through the whole process of data asset management and sharing; Visualization of the whole development process can be completed with basic knowledge of SQL; The navigation mode can be used to deal with complex processes without paying attention to the underlying complex logic; Batch operation can be adopted to realize key functions, and the efficiency of data development and governance delivery has been greatly improved.

3.2.6. Big Data Resource Centre

In order to realize industry data sharing, provide data services and realize digital innovation, the big data resource center realizes the functions of industry data collection, cleaning, management and service, and provides data support for the professional learning platform.

3.3. Platform Features

- The development environment can meet the needs of big data teaching

It adopts R/Python online development environment to realize a strong interactive training system of learning and practicing in experimental teaching, which solves the problems that the learning conditions can not meet the requirements and the integration of teaching and practice in traditional teaching.

- Rich resources

It integrates cases with rich functions, reduces the difficulty of learning, closely follows the actual needs of enterprises, improves students' ability to solve practical problems, and solves the problems that traditional experimental training attaches importance to code, learning is too boring and pointless.

- Academic resources for teaching and research

It can track and collect the latest achievements in the industry, keep up with the cutting-edge trends, solve the differences between different colleges and universities, and shorten the distance between enterprise needs and teaching.

- Practical cases can meet the needs of industry

It uses a variety of methods to collect the latest resources in the industry, build a case base integrating academic and practical work in the industry, solve the problem that teaching can not be combined with practice, and improve students' ability to solve practical work.

4. Conclusion

The learning platform has been applied to the teaching of big data management and application. Students can establish personalized learning space on the platform. Online and offline teaching can be combined with each other to give full play to students' learning enthusiasm and carry out independent learning. The platform can break through time and space constraints, help teachers impart professional knowledge, give full play to students' subjective initiative in learning professional knowledge, broaden their horizons, effectively improve teachers' teaching quality and students' learning efficiency, and cultivate students' practical ability and innovation ability.

Acknowledgements

This research was financially supported by the special project of Digital Curriculum Resources Research of China Association of Higher Education: Prediction and Evaluation of Teaching Effectiveness Driven by Big Data (No.: 21SZYB01).

References

- [1] Xu xukan, Xue Mengyao. Research on Data Literacy Ability Evaluation Index System for Big Data Management and Application Major. *Information Studies: Theory & Application*, 2021, 44(09):50-56.
- [2] Wang Xiwei, Li Yueqi, et al. Research on the Training Mode of Big Data Management and Application under the Background of New Liberal Arts. *Library and Information Service*, 2021,65(17):45-56.
- [3] Li Chunyu. Exploration on the construction of curriculum system for Big Data management and application specialty. *wuxian hulian keji*, 2021,18(05):114-115.
- [4] Yu Xiaogao, Yu Jichao. Research on Prediction of "Four Poor Students" Based on Big Data. *Information & Communications*, 2019(09):81-83.
- [5] Yang Li, Xu xukan, Li Yiming. Research on the Construction of Data Literacy Education System for Big Data Management and Application Major. *Information Studies: Theory & Application*, 2020,43(10):45-49.
- [6] Yu Xiaogao, Yu Xiaopeng. Research on training methods of internationalized talents based on Internet plus and big data. *Information & Communications*, 2018(11):287-289.
- [7] Yu Jichao, Yu Xiaogao. Research on Personalized Learning Space in Internet+and Big Data. *Software Guide*, 2019,18(12):245-248.
- [8] Peng Zhen, Dong Yu, Wang Xiaowei. Research of the Major of Big Data Management and Application of Its Curriculum System. *China Modern Educational Equipment*, 2020(07):49-51.
- [9] Yu Xiaogao. Research on international curriculum teaching. *Journal of Higher Education*, 2016(16):17-19.
- [10] Yu Xiaogao. Research on international teaching of management information system under big data environment. *Software Guide*, 2016,15(05):216-218.
- [11] Yu Xiaogao. Research on prediction of high-risk students based on big data. Xiamen University Press, May 2019.
- [12] Zhou bukun, Geng Ying. Reform and Practice of Application-oriented Personnel Training Mode Based on the Integration of Innovation and Entrepreneurship and Professional Education. *Research and Exploration in Laboratory*, 2020,39(10):220-223.