Pisco Med Publishing

Original Research Article

Research on Big Data Teaching Oriented to the Cultivation of Engineering Ability

Xiqiang Duan[1,2], Lei Zhang[2]

1.Taishan University, Tai'an, Shandong province, 271000, master, lecturer, research area: Big Data.

2.Taishan University, Tai'an, Shandong province, 271000, master, lecturer, research direction: Data Security.

Abstract: In view of the current problems in the teaching of elective courses for graduate students, such as teachers'lectures, students'writing and reading reports, and the lack of systematic training in engineering literacy, in the course of "Big Data Processing Technology", probe into the teaching reform method of the elective course of specialty. While teaching the knowledge of big data, through the design and implementation of group projects, training students in system analysis and design, as well as team cooperation and the ability to solve engineering problems. Practice has proved that ability-oriented teaching, the way of good results, students highly recognized.

Keywords: New Engineering; Big data; Engineering capability; Teaching research

1. Course teaching mode

At present, many universities offer courses related to big data analysis, different universities open, set big data courses, the focus of their knowledge are different. But, basically, it falls into the following categories: first, courses on data analysis and data mining; and second, courses on parallel and distributed programming architectures and models, hadoop MapReduce and the Spark ecosystem; NoSQL and NewSQL database technologies for NON, structured data processing and storage management; and cloud storage oriented courses. By consulting the big data related courses of world-class universities, the following knowledge points are adopted as the main teaching content of the Big Data Application Technology Course: Big Data Storage, and management technologies, big data programming models, big data analytics and visualization technologies, as well as new big data applications such as business intelligence and Data Lakes. Curriculum learning is an important learning process faced by graduate students. Professional courses can enable graduate students to master more extensive professional knowledge and play an important role in future research. The teaching of elective courses for graduate students should adopt different teaching methods and models according to the nature of courses and the characteristics of students. The teaching mode should not be single, it needs to adopt various teaching methods according to the characteristics of the course. In the course of "Big Data Processing Technology", teaching, technical lectures, course experiments, seminars, course group projects and other ways.

2. Course objectives

2.1 Theoretical objectives

The course covers the basic concepts of big data technology, big data ecosystem, big data storage model, big data programming model, Big Data Analysis Technology, big data, visualization technology and mainstream application technology of big data.

2.2 Engineering capability development objectives

The cultivation of students'ability to solve engineering problems is achieved through curriculum experiments, design, and comprehensive engineering projects in the curriculum. The problem of engineering ability development is not only a personal problem, but also a team work. Big data, process technology aims to develop students'ability to work in teams, communicate effectively, manage projects, and learn throughout their lives through project teams. (1) work as a team. Divide students into discussion and project teams, because a project can not be solved by one person and requires collaboration with others. (2) effective communication, discussion class, written and oral students, report, we discuss together to form a final solution. (3) active learning and life-long learning, so as to enable students to take the initiative in learning and reduce classroom teaching by teachers. Students are required to design experiments to verify and solve problems, to collect information, to seek resources, and to improve their professionalism. (4) project management, where students try to execute projects, plan tasks and time.

3. Implementation of curriculum teaching

3.1 Class schedule

"Big Data Processing Technology"is an elective course for graduate students. It has a total of 48 hours. The course consists of lectures, course experiments, literature reading, group discussions and course groups, projects, final exams, and so on. The teacher

doi: 10.18282/l-e.v10i6.2795

spent 32 class hours on the core big data ecosystem architecture and key technologies, and the students in the optional course, 2 -- 4, formed a project team that had to complete a project related to big data processing technology, projects. The teacher arranges 4 class hours to review the students'Literature Reading Report and the project opening report that the group plans to carry out, the work load, the difficulty degree, the experimental data source, the key technology and the algorithm and so on carry on the gatekeeper. Each project team will report on the project, the construction of the experimental environment, the design of the key technologies and algorithms, the programming implementation, the system testing, the problem description and the solution, etc. , to discuss. By each graduate student group members to explain or demonstrate, a total of 6 times, 12 class hours, teachers to participate in the discussion and guidance, students encounter problems to give solutions, suggestions.

3.2 Theoretical content

3.2.1 Big data storage and management

In the past ten years, the growing demand for large-scale data storage and processing has been driving the rapid development of data storage technology, parallel computing technology and database technology. From local storage to cluster, from distributed storage to cloud database, storage, data storage methods have been greatly enhanced. Database systems have evolved from traditional, traditional relational database systems to the latest NewSQL database systems based on NoSQL and support for database, integrity constraints. In addition to structured data, big data involves a lot of Unstructured data, NoSQL and NewSQL are new ways of storing unstructured, structured data. Big data storage is built on top of distributed data store, and HDFS is the most widely used distributed data store. NoSQL database, a variety of documents, graphs, column clusters and key values. The course will cover big data storage models and management techniques. Most NoSQL database systems are open source, so they can be easily incorporated into the experiments of the big data course.

3.2.2 Big data programming model

The distributed computing model is the core of the big data application architecture. Big Data Programming Model provides a general interface for writing Big Data Application System, and provides a simple way for building big data application system. Big data, often involving distributed parallel computing on a cluster. MapReduce is the primary, stream-oriented framework for parallel computing, which supports a variety of data-intensive programs and is the de facto programming model for big data. The core of big data processing is to adopt two basic process objects, Mapper and Reducer, which are processed on different nodes of the cluster by two special computing methods, Map and Reduce. Another widely used algorithmic framework for big data is the Resilient Distributed Dataset RDD (Resilient Distributed Dataset) based on the implementation of the Spark platform. By using RDD, the user does not have to worry about the distributed nature of the underlying data, but only needs to express the specific application logic as a series of transformation processing, thus avoiding the storage of intermediate results, greatly reduces the overhead of data replication, disk I/O, and data serialization. The class will provide an overview of the different programming models in the big data frameworks Hadoop and Spark through more specific code examples.

3.2.3 Big data analytics and visualization

Visualization has become a core technology to extract meaningful information and bring value from the results of big data analysis. Effective tools and technologies that can easily, visualize large amounts of messy data, and improve business decision-making capabilities by visualizing big data. For powerful data analysis and excellent visualization, using the R language on Hadoop will provide a flexible platform for data analysis. R provides the most popular open source statistical analysis software package, Hadoop's powerful data processing capabilities and R's powerful visual analysis capabilities of the combination of RHADOOP, provides a good platform for big data analysis. Visual tools such as the R language and Tableau will be introduced.

3.2.4 Big data applications

With the continuous updating and iteration of big data technology, data management tools have been developed rapidly, and related concepts emerge in endlessly, such as from the initial decision support system (DSS) to business intelligence (BI) , data warehouse, Data Lake, data center, and so on. A Data Lake is a single store of all data in an enterprise, including raw copies of the source system data and the transformation of data for reporting, visualization, analysis, and machine learning tasks. Data Lakes can include a variety of types of data: relational database structure, data, semi-structured data (CSV, XML, JSON) ; unstructured, data (e-mail, documents, PDF) and binary data (images, audio, video). Hadoop is the most common technique for deploying data lakes. Data Center is an intelligent data processing platform that takes on technology, leads business and builds standard definition. Many businesses are already experimenting with using big data and cloud technologies to build data, lakes, and data centers, and to support data driven intelligent decision making. Today, the world's top companies and large corporations have deployed or are deploying data lakes and data centers. The course will introduce students to these new technologies and give them an understanding of the real state of big business, according to the application.

References:

[1] Wei Linghua, Xu Chengzhen, Hu Guoliang. Research on the practical teaching system of data science and Big Data Technology. Journal of Huaibei Coal Industry Teachers College, 2021,42(04) : 83-87.

[2] Dai Pei-shan, Fan Min. Exploration on the teaching reform of big data specialty -- R language data analysis and programming as an example [J]. Computer Literacy and technology, 2021,17(30) : 191-193. DOI: 10.14004 J. No, no, no. CKT. 2021.2931.

[3] Victor-gabriel Lemaire. A probe into the teaching reform of computer application in colleges and universities in the era of big data. Journal of Zhejiang Institute of Water Conservancy and hydropower, 2021,33(04) : 87-90.

[4] Zhao Yanfang, Pan Wenlin. A preliminary study on the teaching of programming course in data science and Big Data Technology. Journal of Yunnan Nationalities University, 2021,30(03) : 298-300.