Research on Application and Development of Financial Big Data

Allam MAALLA, Guang-Yu WU* and Shao-Qiang LI

Research Institute of Open Data Commercial Application, School of Information Technology and Engineering, Guangzhou College of Commerce, Guangzhou, China, 511363

*Corresponding author

Keywords: Big data, Curriculum framework model, Corporate recruitment.

Abstract. This paper discuss the basic thinking, methods and tools of software engineering, masters the financial business knowledge, the analysis and design theory and methods of financial information systems, and has the ability to analyze, design, implement and maintain financial information systems, and can be used in financial applications. IT companies such as development companies and financial information system providers engage in the analysis, design and implementation of software, or engage in the analysis, design, implementation, maintenance and management of financial information systems in the IT departments of various financial institutions such as banks, securities and insurance. A financial information-based composite software talent with a solid professional foundation, broad knowledge, and the ability to adapt to the future development of information technology. Specifically, it is reflected in four aspects: knowledge system, professional skills, project experience and comprehensive quality:

Introduction

On July 1, 2017, the General Office of the State Council issued the “Opinions on Strengthening the Service and Supervision of Market Subjects by Using Big Data”; on July 4, the State Council issued the “Guiding Opinions on Actively Promoting the “Internet+” Action” On September 5, the State Council issued the "Outline for the Promotion of Big Data Development." The intensive introduction of these heavy documents marks the official establishment of China's big data strategic deployment and top-level design. Benefiting from the rapid expansion of the big data market, the demand for related IT support has exploded. Among them, enterprises that provide big data infrastructure, big data software technology services, and industry big data content consulting services have brought unprecedented Customer group.

IDC predicts that by 2020, the company's expenditure based on big data computing and analysis platform will exceed 500 billion US dollars, and the compound growth rate will reach 34.1% in the next 5 years; in the next 3 to 5 years, China needs 1.8 million data talents, but currently only About 300,000 people. At the same time, China's colleges and universities in cloud computing, data science and other majors are still in their infancy, and the talents cultivated each year are far from meeting the needs of the industry. Therefore, it is imperative to open a big data major and accelerate the cultivation of talents.

The financial industry is the industry that relies most on data and is the easiest to realize data. In recent years, emerging financial institutions such as consumer loans and P2P are the products of the combination of big data technology and finance. At present, the demand for big data talents in finance is extremely strong in China. Only Internet finance is one year, and the growth rate is 3-5 times per year. It is generally believed that there will be a gap of 1 million talents in Internet finance, and the most lacking is big data risk control talents, including data mining and statistical modeling talents from primary to advanced.

Core and Featured Courses

Cloud Computing and Introduction to Big Data

As an introductory course in the direction of this major, this course introduces students to the
concepts, technologies and applications related to cloud computing and big data, and enables students to establish a preliminary understanding of the relevant knowledge, technology and development prospects of the profession, as a guide for the follow-up course.

**Distributed Computing Framework Foundation**

As the foundation and core technology course of this major, this course introduces students to the basic concepts, installation and configuration of the Hadoop distributed computing system, distributed programming model (MapReduce), distributed file system (HDFS), and related scheduling. Monitoring and maintenance tools enable students to build a basic understanding of distributed computing systems, master the primary distributed application design and implementation methods, and lay the theoretical and practical foundation for subsequent in-depth courses.

**Distributed Database Management and Development**

As a core technical course in this major, this course introduces students to the basic concepts of distributed databases, installation and configuration, management and maintenance, data access and development. The course focuses on NoSQL databases such as HBase, MongoDB, Redis, etc., and describes their use and development in a distributed environment. To enable students to establish a basic understanding of distributed databases, master the primary distributed database application system design and development methods, and lay the theoretical and practical foundation for the subsequent in-depth courses.

**Distributed Computing Framework Component Technology**

As a core advanced course in this major, this course introduces students to mainstream components on the Hadoop distributed computing platform, including Hive, Pig, Sqoop, Flume, Kafka, Zookeeper and more. Enable students to have a complete Hadoop ecosystem-based design and implementation of big data applications.

**Real-time Calculation and Memory Calculation**

As a core advanced course in this major, this course introduces students to high-performance distributed computing frameworks, including Storm and Spark, as a more powerful alternative to the Hadoop framework.

**Data Visualization Technology**

As an elective course in this major, this course introduces students to the basics of data visualization and the design and use of platforms and development tools, including Excel, Reporting Services, Chart.js, D3.js, Tableau, etc. Through this course, students will be able to present the results of big data processing in an efficient, flexible and friendly manner.

**Data Statistics and Analysis**

As a core advanced course in this major, this course introduces students to statistical analysis techniques based on Python and R. Including data file editing and finishing, basic statistical analysis, parameter estimation and hypothesis testing, non-parametric testing, analysis of variance, correlation analysis, regression analysis, cluster analysis, discriminant analysis, factor analysis, correspondence analysis, reliability analysis, survival Analysis, time series analysis, and the drawing of statistical graphs enable students to master the processing and analysis methods of typical industry business data.

**Knowledge System**

Mathematical basis: Including calculus, linear algebra, probability statistics, numerical analysis, etc.

IT foundation: Including operating systems, networks, databases, software engineering, programming techniques, data structures and algorithms, etc.
Knowledge base in the financial sector. Including international finance, marketing, insurance, securities investment, etc.

**Professional Skills**

Database system management and development: MySQL, MongoDB, Redis, HBase, etc. Big Data Application Development Language: Java as the core, supplemented by Python, Scala, R, etc.

Construction, configuration, development and deployment of big data processing frameworks: Hadoop, Storm, Spark, etc. Use of data analysis and presentation tools: reporting tools, D3.js, etc.

**Project Experience**

Familiar with enterprise software project life cycle, development process, specification, etc.

Understand and implement software quality requirements: performance, security, scalability, maintainability, reliability, etc. Understand the financial industry: industry background, business model, market characteristics, and how data and IT systems are used in the financial industry

**Comprehensive Quality**

Good professional basic qualities: document writing, presentation reporting, business communication, etc. Strong learning ability and study habits, has a certain degree of micro-innovation, data awareness

**Course Settings Table**

The professional competence-course structure derivation process mainly includes two stages: “computation ability theme” and “capability-curriculum structure transformation”. The main process of the first phase of the "computational power theme" is as follows:

![Diagram](image)

Figure 1. The data analysis process of "ability topic calculation."

**Hadoop Big Data Integrated Experiment System**

This experimental system is designed to provide students with a complete set of Hadoop and its environment, design, development, monitoring, maintenance tools, software and services. With this experimental system, the experimental and training environment requirements of the core technology courses of this major can be met.

This experimental system is divided into two major components:

A virtual lab environment for students to learn big data. The environment is carried out by means of the aforementioned virtualized desktop teaching system, and the network administrator configures the big data learning virtual machine in advance for the students to use.

The real environment for research or large-scale case presentations. This environment is carried out through several servers.

**Main Function**

*Basic platform:* The basic platform for big data storage and processing, which can realize the storage and management of massive data, support common components of platforms such as Hive, Impala, Pig, Spark, and Yarn, and provide support for data analysis services on the platform. These
common components increase the ease of use of platform data, making data manipulation and data analysis easier to use, saving labor and reducing labor time.

B. Data integration: Support the unified storage of massive structured data, semi-structured data, and unstructured data, deepen the expansion of enterprise intelligence and service capabilities, and improve the decision-making level of enterprises. We can use enterprise-level data ETL tools or open source ETL tools. For example, Flume, Sqoop, Kafka, etc., integrate externally structured, semi-structured and unstructured data into big data platforms. Through this platform, faster and more convenient data integration can be achieved.

C. Data analysis and processing: Provide different data analysis and processing modules for different industry applications. Through data analysis and processing, enterprises can clarify the scope of their intelligence and service capabilities, as well as the requirements and solutions implied by the data. Solutions and the ability to meet the deep data mining and analysis needs of different industries.

D. Platform security: The security of the platform is mainly reflected in the operations of personnel, data management, data disaster recovery and transfer. Personnel operations: unified access control, comprehensive auditing to prevent administrators from viewing business system data files; data management: overall data protection solution, automated Kerberos deployment (big data SSO); data disaster recovery and transfer: platform has high Availability, fast fault recovery, data security, automatic detection and data repair and fault tolerance in data detection.

E. Development service layer: The development service layer visualizes the results of data analysis and processing through visual management and application, and has a good human-machine interface and user experience.

**Summary**

The case part is a set of teaching cases formed by extracting suitable business scenarios from the real projects of the enterprise and carrying out the necessary simplified processing. Each version of the case is based on the business scenario, combined with the software engineering and software system architecture of the university. Courses such as technology, as well as technical skills required for project construction such as teamwork development and project resource version management, are designed to help students fully implement the entire process of enterprise-level application system construction. This platform will provide online office system and gold audit project management system as a student training case for students to carry out training.
Acknowledgement

This research was financially supported by the 2017 Higher Education Teaching Research and Reform Project of Guangdong Province, "The Big Data Talent Cultivation Practice Teaching System of Business Schools" (Project No. 2017SJJXGG01), and Characteristic Key Subject of E-commerce Construction Project of Guangzhou College of Commerce Foundation, (Project No.TSZDXK201601).

References


