ERP Virtual Simulation Practice Platform based on Big Data

Jiang-Hui LIU¹ and Man-Jia CHEN²,*

¹Network and Information Center, Experimental Teaching Center Guangdong University of Foreign Studies, Guangzhou, Guangdong, China
²School of Business, Guangdong University of Foreign Studies, Guangzhou, Guangdong, China
*Corresponding author

Keywords: Big Data, Data Analysis and Mining, ERP, Virtual Simulation, Experimental Platform

Abstract. The large amount of data accumulated in the ERP practice course system of universities has an impact on the quality of teaching. And the purpose of this paper is to expound and design the intelligent ERP virtual simulation Big Data analysis and mining platform. Its principle is to explore the relationship between massive data to promote the reform of teaching programs and realize the rational use of teaching resources.

1. Introduction

Enterprise resource planning (ERP) is an enterprise information management system for the integrated management of material resources, capital resources, and information resources for the manufacturing industry. This system is widely used by most contemporary enterprises. In order to realize the teaching mode of combining theory with practice, universities have also set up related ERP courses, which are designed to improve students’ practical ability to meet the needs of enterprises for compound talents.

With the rapid increase of information, people's needs for analysis and processing of massive data are increasing. This paper aims to introduce big data technology into the virtual teaching practice platform for university ERP teaching. This new platform helps teachers to analyze different problems by collecting and processing the data generated by students using ERP system. And the analyzed information will facilitate reasonable changes in teaching plans to meet the needs of the times.

2. Big Data Technology Overview

With the arrival of the third wave of information technology, cloud computing, the Internet of Things and big data have gradually entered people's horizons and are widely used in all aspects of life. The rapid development of information technology has brought us massive amounts of information, and various types of data have become increasingly complicated, which has brought us into the era of Big Data.

Big Data is a massive information asset that spans systems, disciplines, time and space. Under the new processing mode, it can accurately, meticulously, objectively and timely reflect all aspects of people and things. It can also provide services for human production and life. Big Data has 6 V characteristics, that is, volume, velocity, variety, veracity, and visualization. [1]

Big Data technology has a complete and complex system. Its main technologies involve data collection, preprocessing, and distributed storage. The establishment of NoSQL databases, data warehouses, and the realization of data through various technologies such as parallel computing and visualization are also the key technologies of Big Data [2]. These technologies are widely used in enterprise management activities. Enterprises can process massive data generated in the production and operation process in batches and efficiently, and then establishing data models to provide advice for enterprises.
3. Virtual Simulation Technology

Virtual simulation technology is a system that uses one virtual system to imitate another real system. It mainly includes virtual reality technology and simulation technology, and has the following four characteristics: immersion (to obtain an immersive experience), interactivity (people and the environment can react with each other), illusion (to simulate the real environment and the environment that cannot exist in reality), and fidelity (as if in the real world). [3]

The virtual simulation experimental platform is based on virtual reality technology, multimedia technology and Internet technology. It combines simulation technology with virtual reality technology to integrate a production and business activity virtual system into reality and applies into actual teaching. Through this virtual simulation, a computer network course teaching platform can be constructed, which provides a basis for the teaching and practice of university courses, so that students can improve their practical ability while learning theoretical knowledge.

4. Design of ERP Virtual Simulation Experiment Platform based on Big Data

This paper refers to the Big Data technology architecture and combines the current data processing requirements of the ERP virtual simulation experiment platform, hoping to apply Big Data technology to the intelligent ERP virtual simulation system. Overall, the technical architecture includes five levels, and the hierarchical structure is shown in Fig. 1.

![Figure 1. Hierarchical Structure of Virtual Simulation Practice Platform.](attachment:image)

In the data source layer, it will collect various types of data (generate large amounts of data after a long-term use) generated by the ERP platform through the Big Data collection tools, and provide massive data sources for big data analysis.

In the acquisition layer, data from the data source layer will be collected and extracted through ETL tools such as Kettle and Datestage, which will remove some useless data, and integrate the scattered, disorderly, and inconsistent data into the analysis layer.

768
Analysis layer, the most important part of the platform, provides storage, backup, query, and calculation functions for data from the acquisition layer, including two parts: the Big Data foundation platform and the fusion database. The Big Data foundation platform is the key to data analysis. The platform will use the open source Hadoop technology foundation to analyze and mine the data processed by the acquisition layer. The fusion database is used to store the analysis results, application data and mining data. It is a data repository after Big Data analysis. The fusion database uses the open source Sqoop and Big Data foundation platform to establish data transfer functions.

Relationship layer and application layer are service layers that directly face to users, provide users with query results of data analysis, and give them with decision suggestions.

5. **Implementation of ERP Virtual Simulation Practice Platform Based on Big Data**

5.1 **Big Data Basic Analysis Platform**

The open source software Hadoop under the Apache Software Foundation is the key technology to implement the ERP virtual simulation Big Data analysis platform. The Big Data basic analysis platform will integrate the Hadoop ecosystem and the Spark ecosystem to process, analyze, and mine the data of the ERP virtual simulation practice platform.

To realize the Big Data analysis platform, we must first set up a Hadoop distributed storage environment for the persistent storage of data. Use the Centos Linux system host to build the Hadoop cluster environment, and install Java and Scala languages to support the operation of Hadoop and Spark.

5.2 **Virtual Simulation Operation**

The general process of virtual simulation operation is to build a model and write code. Establishing a model is the key to virtual simulation operations. The modeling method is mainly to build a model through model parameter acquisition, mechanism analysis method and model actual measurement. By obtaining the data running on the ERP virtual simulation platform, the open-source KNIME data mining software is used to analyze the data to build a production management model, procurement management model, order management model, and financial management model, and deeply explore the potential between each model contact. Writing code is to write mathematical formulas into computer languages and turn abstract models into executable programs. Simulation calculation combined with real-time data for fast and efficient analysis, making full use of the massive data of the ERP virtual simulation experimental platform, organically integrating the mastery of multiple internal business processes such as finance, sales, purchasing, inventory, production and human resources. The powerful computing power provided by the platform enables dynamic analysis of the entire process.

5.3 **Implementation of Platform Information Resource Scheduling**

The information resource scheduling of the actual ERP virtual simulation experiment platform needs to meet the different needs of users. By collecting and extracting data, the platform needs to analyze the data characteristics to provide users with a decision basis.

To perform Big Data analysis on the ERP virtual simulation practice platform, we must first obtain various data in the ERP. The ERP virtual simulation Big Data experimental platform encapsulates various types of platform data through virtualization technology and forms an ERP data pool. When users want to analyze certain data, they make a request to the data source pool, and pass the target data and the processing result of the data to a set of requirements. Then the task debugging degree of the virtual data is performed in the data pool. First, the virtual data is processed. The current load status of the mapped data resources is evaluated and calculated. Based on the user's requirements for the target data, the ERP virtual
simulation big data experimental platform selects the resource scheduling options that can meet the user's requirements, thereby allocating data for the user, as is shown in Fig. 2.

Figure 2. The Process of Resource Scheduling.

The data generated in the ERP system is mainly stored by the Nginx server. You can use the streaming data collection tool Flume in the Hadoop ecosystem to deploy multiple proxy nodes on the server to collect the collected data in HDFS distributed file storage and make them persistent storage in the system.

Data stored in HDFS can be analyzed in three major types: offline analysis, real-time data query, and data mining.

First of all, the offline analysis can use MapReduce to perform offline analysis of batch iterative algorithms. In addition, it can also use the data warehouse products Hive and Pig on the Hadoop platform for partial offline analysis. As for the Hive data warehouse, it can establish various dimensions for a large number of historical data, analyze and establish a data relationship model and provide users with decision analysis.

Real-time data query is implemented through the open source distributed database HBase. The amount of data in a single table in HBase can usually reach terabytes or petabytes, but in most cases data can be read in milliseconds, because the mechanism of HBase is that data is written to RAM first. When the amount of data reaches a certain amount, it will then be written to disk. Based on this mechanism, real-time data query can be realized by reading existing data in RAM.

The Spark calculation engine directly reads the data in HDFS to form the initial RDD. During the reading process, the RDD is automatically divided into several partitions part1 ... part n (n is the total number of partitions for parallelized calculation) based on the HDFS storage partitions. The next process is to perform characterization processing, establish various latitude relationships and use zipWithIndex() function to binary change part of the feature vectors. Finally, it will perform integrated operations on the data to deeply dig the hidden information of the data [4]. The specific process is shown in Fig.3.

Figure 3. The Specific Process of Data Analysis and Mining of Intelligent ERP Virtual Simulation Big Data Practice Platform.
YARN in Hadoop provides resource scheduling management functions for the entire Hadoop cluster, allocates resources for data analysis, real-time data query, and data mining, and provides underlying computing resources to enable various functions to be implemented.

6. Conclusion

ERP Big Data virtual simulation practice platform is proposed to make good use of the platform data. Through the analysis and mining of the data, this platform can make the abstract data be converted into effective and intuitive information, which provides the theoretical basis improvement of the university ERP practice teaching.

References

[1] Y. M. Luo, F. T. Liu, M. X. Chen, W. L. Tang, Y. L. Yang, & X. L. Tan. A machine learning model based on initial gut microbiome data for predicting changes of Bifidobacterium after prebiotics consumption. Journal of Southern Medical University 38.3(2018):251-260.

[2] X. F. Zhang, X. T. Huang, and F. Wang. The construction of undergraduate data mining course in the big data age. 2017 12th International Conference on Computer Science and Education (ICCSE) IEEE, 2017.

[3] Y. N. Wang, and G. L. Feng. The Application of Virtual Instrument Technology in the Teaching of Microcomputer Principle and Interface Technology. Software Engineering (2017).

[4] Na, Bao, Q. Wang, and Y. J. Jia. Research of Characteristics of Stimulation Methods and Application of Acupoint in Auricular Needle Therapy Based on Data Mining. 42.4(2017):372-376.