



Design and Research of Big Data Experimental Teaching Platform

Li Dai and Jialong Liu(✉)

Teaching Affair Office, Leshan Normal University, Leshan 614000, Sichuan, China
qingreansd84416@163.com

Abstract. In order to meet the actual needs of big data experimental teaching in the new era, this paper explores the design of big data experimental teaching platform based on the current mainstream computer technology. Through architecture design, database design and functional module deployment, the design effect is tested. The test results show that the performance of the platform has reached the expected goal and achieved initial success. It is expected that the platform will be gradually popularized and applied in the future big data experimental teaching.

Keywords: Big data · Experimental teaching · Teaching platform

1 Introduction

In recent years, with the further deepening of the application of big data technology, it has been popularized and applied in the experimental teaching of various majors, and researchers have also carried out some research on it. Based on the experimental big data learning platform, researchers such as Huang Hailong [1] applied the technology of big data experimental teaching platform in computer hardware experimental teaching, and integrated the flipped classroom mode, which significantly improved students' practical ability; Cao Tianze [2] and others used the existing old PC equipment to build a big data experimental teaching platform to carry out teaching, and also achieved good teaching results. The above research shows that the advantages of big data experimental teaching platform are remarkable, but most of these studies are broad and lack of specific research on the design of big data experimental teaching platform, so this research will focus on making up for the above shortcomings.

2 Platform Basic Requirements Analysis

Considering the actual experimental teaching needs, in this design, the basic requirements of the big data experimental teaching platform are divided into the following three aspects.

Firstly, the experiment cluster management, including: (1) Cluster environment management: mainly designing the communication network between users, and importing

some tools and data sets needed for the experiment; (2) Cluster deployment management: it is mainly used to build a big data experiment environment and provide a big data cluster for typical experiments; (3) Cluster change management: data can be saved to restore the experiment at any time, and the cluster can be added or deleted.

Secondly, platform teaching management, including: (1) according to different account levels, determine different platform permissions; (2) manage the platform users and divide the user’s experimental resources; (3) Upload, modify and download relevant materials; (4) Provide the function of publishing experimental tasks for teachers and uploading homework for students.

Thirdly, platform monitoring, including: (1) Physical resource monitoring: mainly used to monitor the experimental resources of each key node; (2) Virtual resource monitoring: its main purpose is to realize the unified distribution and management of virtual resources; (3) User behavior monitoring part, which mainly captures the operation of student users during the experiment.

3 Platform Module Design

3.1 Basic Architecture of the Platform

Considering the needs of development speed, practical application and later visit, this system development is based on the mainstream B/S architecture, and adopts the four-tier architecture model [3, 4], as shown in Table 1.

3.2 Functional Module Design

3.2.1 Experimental Cluster Management Module

To realize the function of this module, in the design of this link, firstly, create a cluster network environment, manually create Linux Bridge and name it br-ex, and bind it with enp1s0f1 network card to create an external network, then create the internal network of vxlan, and connect them with each other by creating virtual router router1. Secondly, the deployment management of the experimental cluster is carried out. This link adopts the mode of “one-click cluster deployment” [5], including: (1) Check the resources of the experimental cluster pool of the platform. If the resources allow users to conduct

Table 1. Overall architecture of platform

levels and ranks	major function
User layer	Used to perform related operations.
application layer	Used for business logic processing
Monitoring layer	Used for monitoring user resources and experimental behaviors.
Virtualization layer	Used for file storage and related service deployment.

(In this architecture, the virtualization layer is the lowest layer and the user layer is the highest layer.)

Table 2. Role table

Role ID	Role type
1aef408baf8b48958d43ac8533ecdf62	Admin
1507c10a8c2e4e128a54d02c8701f2e5	Teacher
1cc316316490410e9b5fa32748a1e007	Student

experiments, obtain the id and experimental project of the current student user, and assign them a virtual network device and IP; (2) importing the pre-made experimental tools and data sets, and transmitting them from the server to the virtual machine; (3) Show cluster information and virtual machine information to users, and provide operations of adding, deleting and checking. Include increasing or decreasing the number of nodes in the cluster, querying the cluster information through the cluster name or state, verifying the state of the cluster, and deleting the cluster owned by the user.

On the other hand, in order to achieve a more scientific and reasonable quota, user quotas are allocated based on the following formula:

$$\sum_{i=1}^N (CPUNumi * CoreNumi) * OverR * CPUR$$

$$\geq UNum * UVCPUNum$$

In this formula, OverR refers to the ratio of physical CPU to virtual CPU, which is the number of the *i*th physical CPU and refers to the number of cores; The number of users refers to the virtual CPU quota of users.

3.2.2 Platform User Management

To realize the management of platform users, in the research of this link, user accounts are first created, and their roles are shown in Table 2.

After the user account is created, the newly added role is authenticated based on Keystone service, and the authentication rule name corresponds to the user instruction function. The specific method is to add the authentication rule names in turn in the json-based permission authentication policy file, as the basis of whether there is permission to call this function. If you have permission, call related functions for processing.

3.2.3 Teaching Assistant Management

In order to ensure that this platform can meet the needs and management of students in the teaching process, this teaching assistant management module is designed to realize a series of teaching management functions such as the release and display of experimental information, the management and download of experimental data, the submission and grading review of homework [6]. Specifically, it is mainly divided into the following parts.

First, the release of the experiment. After the teacher selects this function, the IssueExpView function is responsible for processing the request, and it is redirected to the web page form pointed by IssueExpForm, which provides an input field for experimental information. After the teacher user inputs the relevant information, the system uses JavaScript to process the relevant information and increase the mouse response event, so that the experimental content can be viewed by hovering the mouse pointer over the experimental title.

The second is the submission of homework. In this section, the class of homework file is Lab_report, which has similar attributes to the experimental data, but adds three types of data: student name, review status and grade. When the job is submitted for the first time, the review status, score and comments are all empty, and this is the status to be reviewed. The attribute will change after being scored.

The third is homework evaluation. This module provides teachers and users with an assignment evaluation button, which corresponds to the CheckReportView function. After obtaining the assignment information, the information is transferred into the CheckReportForm, which provides a result input field and a comment input field. After submission, the assignment will be returned to the assignment display interface, and the right side of the assignment will display the scores and comments, and the assignment review status will be changed to “Reviewed”.

3.3 Database Design

In the database design of this experimental teaching platform, the design is mainly based on MySQL2019, and the detailed database design is carried out according to different data information involved in the system platform. The specific data tables are shown in Tables 3, 4 and 5.

At the same time, in order to improve the security of data, in this design, the encryption algorithm of RSA combined with AES is used for data encryption design, and its basic flow chart is shown in Fig. 1.

On this basis, in order to ensure the effective application of the encryption algorithm, DS420j storage is selected in this design. Firstly, the relevant data are converted into uniform data, then these data are output through the interface, and finally encrypted based on this encryption algorithm.

Table 3. Platform User Data Sheet

field name	Chinese name	data type
zhID	Account ID	Int
Zhmc	Account name	varchar
Zhmm	Account password	varchar
Project ID	Project ID	int

Table 4. Data Sheet of Document Information

field name	Chinese name	data type
wdID	Document ID	Int
Wdmc	Document name	varchar
Wdlx	Document type	varchar
Wdzt	Document status	varchar
Fs	mark	Int
Py	comment	varchar

Table 5. Data Sheet of Experimental Equipment

field name	Chinese name	data type
xnjID	Virtual machine ID	Int
xnjdlm	Virtual machine login name	varchar
Xnjzt	Virtual machine status	varchar
IPdz	Ip address	Int
CPUsl	Number of CPU	Int
Nczl	Total memory	Int
Cpzl	Total disk	Int

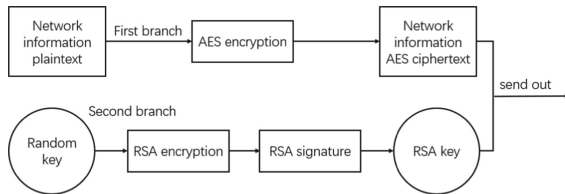


Fig. 1. Flow chart of RSA fusion AES encryption algorithm

3.4 Data Processing Module of Experimental Platform

In order to improve the efficiency of the experimental platform in big data processing, the designer adopts the distributed computing mode, in which multiple computers are used to complete the computing tasks together. Considering that the big data teaching platform needs both computing efficiency and algorithm efficiency in the later application, in this design, the designer integrates Hadoop and Spark, which are commonly used in distributed computing, to make them in a cluster, so as to build a Map/Reduce distributed computing model based on Hadoop framework and incorporating Spark computing mode. Among the Map/Reduce distributed computing models, the main functions of Spark computing model are: (1) In the preprocessing of big data, some raw data are

called in for data cleaning and weighting; (2) In the process of big data analysis and calculation, multi-node parallel calculation is used to complete the calculation task scheduling of the algorithm, and the transaction set division and pruning strategy are combined to reduce the complexity of data processing. At the same time, in this link, because the process of scanning Value is completed directly in the Spark computing framework, the common problems such as insufficient memory in the traditional big data experimental teaching platform have been significantly improved.

Overall, the Map/Reduce distributed computing model designed this time is mainly used to parallelize the algorithm, and its processing steps are as follows: (1) Using the Map/Reduce distributed programming model under Hadoop framework, the data in the whole database is decomposed into several continuous data segments, and each data segment corresponds to a storage device; (2) In the way of parallel scanning, the data sets calculated locally by each storage device are counted, and each data set is processed by the corresponding mapper. The processed results are first stored in the distributed cache, and then the processed results of each division will be collectively mapped to a global node and sorted in descending order.

3.5 Design of Monitoring Module

In the design of the monitoring module, the monitoring of the following links is mainly realized by creating a Dashboard class called monitor:

The first is to monitor the physical resources. The monitoring script is `physical_collect.sh`, which calls the file `physical_collect.py` written by python, so as to realize the retrieval of CPU, memory, disk and other related information, and calculate its utilization rate. In order to improve the real-time performance, it is set to obtain the current information every ten seconds, and all the obtained information is stored in MySQL database.

The second is to monitor virtual resources. In this link, `get_instance_info()` function is used to obtain all virtual machine information, which is stored in the virtual machine data table, and update the virtual machine table on the platform with `update(host,sql)` function. At the same time, the resource occupation of each virtual machine is monitored based on the `connect_instance()` function.

The third is to monitor the user's account. On the one hand, it monitors the user's resources. Based on the virtual resource monitoring module, it monitors through the `get_user_monitor_data(self)` function to view the current user resources in real time. On the other hand, it monitors the user's experimental behavior, which uses threads to start the process of monitoring the user's experimental behavior.

On the other hand, considering that the big data experimental teaching platform has high requirements for the accuracy of applied data information, in order to avoid some irrelevant information affecting the normal operation of the teaching platform, in this design, cloud log analysis technology is introduced, and a deep neural network model (Fig. 2) using LSTM is built to realize real-time monitoring of relevant data information. This technology is mainly realized by clustering algorithm, which is mainly based on the following steps: First, the logs are grouped based on the log length. Each new original log should be divided into word sets to improve the accuracy of analysis. The second is to get the log string, which helps to solve the problem of the same log length but different

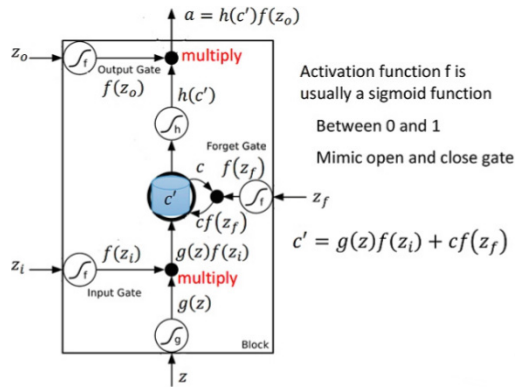


Fig. 2. Schematic diagram of LSTM depth neural network model

templates. At the same time, in order to avoid too many log strings, it is stipulated that “*” should be used to replace numbers or logos, and then these strings and “*” should be combined into a string according to the position of their respective characters, which forms the final log string. Thirdly, clustering the formed log strings. In order to reduce the time complexity, before extracting the log template, a simple clustering process is carried out. On the premise of not distinguishing the case, the logs with the same log strings are classified into one class, and then extracted to improve the efficiency. Fourthly, the log matching tree is built based on Tric concept. Through Tric concept, the log template can be extracted efficiently and accurately, and the dynamic log matching tree can be realized. The newly entered log is used as the node with the same query item as the log in the tree.

3.6 Installation and Configuration of Platform

After all the above functional modules and basic parts are designed and deployed, the platform will be installed and configured. This link is mainly divided into the following steps. First, install and deploy the master node. This link mainly deploys the real-time monitoring tools and their clients, as well as the mirroring, network and access services of OpenStack. Through the installation and deployment of the master node, teachers can manage the background of the whole platform through the master node after logging in to the administrator account, including checking the platform status, monitoring and managing various resources such as CPU, memory and hard disk, and also managing the student accounts on the platform. Ordinary student accounts can create instances, mirrors, snapshots, etc. on the platform, and view the detailed information of these resources, as well as the quota of experimental resources they have applied for.

Secondly, install and deploy the control node, mainly configure the management network card and network related information of the node, IMM management network configuration, NTP server and DNS service related configuration, etc. Through this link, we can realize the control service of various resources, realize the flexible allocation and sharing of resources, and provide services for the upper master node. Thirdly, the computing and storage nodes are installed and deployed, and this link is deployed by the way

of Fuel automatic installation. After the above steps are completed, start all the computers, first start the management software, and then the students start the corresponding software on the teaching computer respectively. After entering the corresponding IP, user name and password, they can log on to the platform for practical training.

4 Testing and Analysis

In order to explore whether the performance and requirements of the big data experimental teaching platform designed this time meet the requirements, the platform is tested in this link. The test is mainly divided into the following two parts.

First, based on the “one-click deployment” method, we test the deployment of the user’s experimental environment. This link simulates the cluster deployment of multiple user accounts, and the deployment of different user scales is shown in Table 6.

From the data in Table 6, it can be seen that the deployment speed and success rate of the big data experimental environment of this platform are relatively high, and it is also superior to the previous traditional deployment mode.

On this basis, a typical big data experiment is taken as a case to further test the performance of the platform. In this link, 10 million data sets from a large shopping website are introduced as test cases to simulate the unfavorable situation of 30 users conducting experiments at the same time. The test shows that the resource occupancy is shown in Table 7.

According to the data in Table 7, the big data experimental teaching platform designed this time can receive multiple users to conduct typical big data experiments at the same

Table 6. Deployment Results Table

subscriber number	Average longest deployment time/s	CPU utilization/%		Number of first successes
		Control node	Computing node	
one	2.9	45	15	one
five	3.8	50	24	five
10	6.2	fifty-two	34	10
20	8.7	66	47	20
30	25.4	93	67	29

Table 7. Analysis of resource occupation

Index	parameter
CPU utilization/%	77.3
Memory occupancy/%	58.2
Hard disk occupancy/%	32.0

Table 8. Comparison of results after using big data experimental teaching platform

Group	total number of people	average score	excellent rate
Experimental group	35	85.3	91.0
control group	35	82.1	83.4

time, and its performance is relatively excellent, which is expected to have high practical value.

On this basis, in order to further explore the practical application effect of the big data experimental teaching platform designed in this paper, after discussion and decision of the teaching and research working group, the platform is applied to the practical teaching work. In a computer major of a university, two classes with similar average scores are selected, namely the experimental group and the control group. The experimental group introduces the big data experimental teaching platform designed in this paper for auxiliary teaching, while the control group uses the traditional experimental teaching mode for teaching. After 16 teaching weeks, the results of the two groups are shown in Table 8.

From the data in the Table 8, it is not difficult to see that after the introduction of the designed big data experimental teaching platform, the average score and excellent rate of the students in the experimental group are significantly higher than those in the control group. This shows that the big data experimental teaching platform designed this time has certain practical application value and is expected to be further popularized and applied in the future.

5 Conclusion

All in all, the big data experimental teaching has higher requirements for the comprehensive performance of the teaching platform, and it is also an indispensable content to optimize its design based on the latest computer technology. Through this research and the actual test, it is shown that this research has effectively improved the actual effect of the big data experimental teaching platform, and has potential application value. Of course, as the platform is still in the initial stage of operation, there is bound to be a large room for expansion. Therefore, in the future work, it is still necessary to continuously optimize and adjust it in combination with the actual situation to ensure that its superior functions can be fully exerted.

References

1. Huang H, Li Q (2019) Research on the application of experimental big data learning platform in experimental teaching of computer hardware. *Mod Comput* 32:43–46
2. Cao T (2020) Research on the design of an economic and open-source big data experimental teaching platform. *Heilongjiang Sci* 11(13):94–95
3. He Y, Duan L, Zou Z (2022) Design of Internet traffic big data application teaching inquiry platform. *Exp Sci Technol* 20(02):117–121

4. Deng N (2021) Research on the construction quality of adaptive experimental teaching platform based on cloud computing. *Popular Standardization* 24:16–18
5. Li T (2020) Design of experimental teaching platform of Cloud Computing and Big Data based on multi-computing framework. *Inf Comput (Theoret Ed)* 32(24):190–192
6. Du W (2020) Research on the application of big data platform in professional experimental teaching. *Software* 41(06):32–34

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

