Design and Development of University Smart Campus Platform Based on Big Data

Daigen Huang

543948894@qq.com

Sichuan University of Culture and Arts, Mianyang, Sichuan, China

Abstract: Under the background of the in-depth development of digital society, the informatization process of colleges and universities will gradually change to intelligence and intelligence, and the new generation of information technology represented by big data technology has become the core force of building a "smart campus". In this regard, this paper takes the overall architecture of "Smart Campus" as the research object, and comprehensively integrates the practical characteristics of big data technology, network information technology and computer application technology. On the one hand, Hadoop cluster is used to build a data analysis and processing server. On the other hand, we will design a Web-based intelligent campus application service platform based on J2EE technical specifications. The whole platform adopts B/S architecture design, and the Web Server completes the control of various business logics and the configuration of corresponding API data interfaces, which facilitates users to query and call many data resources such as teaching, scientific research, student behavior, campus management, etc. through simple interactive operation of front-end pages. It also relies on data mining algorithm models such as K-means and Person to complete the analysis and processing of massive data, which optimizes the teaching management mode and improves the level of educational services in colleges and universities, thus making benefits for the construction of smart campus in colleges and universities.

Keywords: Big Data, Smart Campus, Hadoop, Data Analysis, Computer Application.

1 INTRODUCTION

With the release of China's "14th Five-Year Plan" and the outline of the 2035 long-term goal, building a high-quality education system in an all-round way has become the starting point and focus of further deepening the reform and innovative development of China's higher education, and it is also the only way to meet the needs of education modernization and achieve a strong education country. Under the high-quality education system, the concept, culture and ecology of education are constantly innovating and changing. Especially under the influence of the new generation of information technologies such as network communication, big data and artificial intelligence, the education mode and talent training objectives have undergone unprecedented changes, and the development of educational informatization has entered a critical period of transformation and upgrading. [6]

At present, the deployment of information infrastructure in colleges and universities is becoming more and more perfect, and various education and teaching systems and campus

BDEDM 2023, January 06-08, Changsha, People's Republic of China

DOI 10.4108/eai.6-1-2023.2330349

management service systems are everywhere. However, the original intention of most designs and constructions is based on the concept of "on demand, independence and one by one", so that the data application caliber among various systems is not uniform, and there is a lack of data correlation and interaction, thus forming an isolated island structure of data information, hardware equipment and application services, which not only increases the application difficulty of the system, but also makes the management and operation and maintenance of the system complicated and chaotic. [10] Traditional and single information means can't meet the development requirements of campus comprehensive management and knowledge resource service in the era of big data and artificial intelligence. However, the proposal of smart campus can fully perceive the campus environment with the support of big data technology, and highly unify and correspond the physical space and data space, fully contain all kinds of characters, events and tasks and data information on campus, and manage and monitor all kinds of information and data on campus in various directions by using data modeling and data mining technologies, so as to realize the reconstruction of campus development ecology. In view of this, this paper holds that colleges and universities should adhere to the strategy of educational modernization and high-quality development driven by educational informatization, accelerate the innovation and integration of the new generation of information technology and the practice of educational and teaching activities, give full play to the application advantages of big data technology, network information technology and computer application technology, and build a smart campus application service platform in colleges and universities. Based on the interactive application of data processing center and visualization, the platform of "Smart Campus" comprehensively covers three major sections: teaching, scientific research and daily management, completes the functional integration of each subsystem, enhances the comprehensive utilization of data resources, improves the service level of campus, and makes a useful attempt to construct a paradigm of smart education in colleges and universities.

2 INTRODUCTION OF KEY TECHNOLOGIES

2.1 Big Data Technology

Big Data generally refers to the data collection whose data volume has exceeded the normal range, and users can't use ordinary software tools and methods to process it. Big data comes from the accumulation of network communication technology, which is the inevitable result of the application of various digital technologies. Since big data was put forward, it has formed a huge and dynamic concept after years of development. On the one hand, it represents a huge amount of data information resources, on the other hand, it represents the general name of a series of data value technologies. [9] Big data technology is the foundation of data value, and it is also the key to the transition of big data from single data level to application level. Big data technology is a collection of technologies covering the whole life cycle of big data, covering a series of links such as data collection, transmission, cleaning, storage, analysis, processing, presentation and application. It provides a complete processing paradigm for the acquisition and reflection of the value of big data, and also expands the application scenarios of big data. With the wide application of big data, the development speed of big data technology is constantly accelerating, and many technical means are gradually gathering and forming a systematic and ecological big data technology stack.

2.2 Hadoop Technical Framework

Hadoop is an open source technology framework realized by Java language, and it is also the most widely used big data core technology at present. Hadoop is a distributed system architecture, and its interior is mainly composed of HDFS (Hadoop Distributed File System) and MapReduce, which can be used as the core of the underlying storage and analysis to provide users with reliable, scalable and distributed computing big data services. [5] In addition, the high scalability of Hadoop framework can support the deployment and installation of various functional components, and they are compatible with each other, forming an independent application system and gradually evolving into Hadoop ecosystem, as shown in Figure 1. Hadoop supports cluster deployment, which is convenient for users to quickly complete the design and development of large-scale data analysis applications, and realize the parallel high-speed operation and complex call of big data.

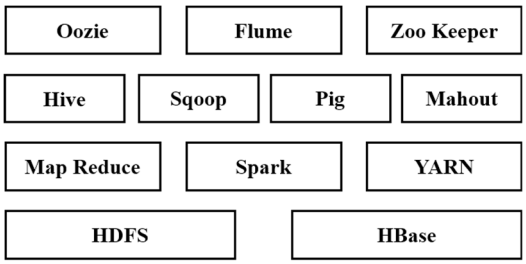

Figure 1: Hadoop ecosystem (original)

2.2.1 HBase

HBase is a distributed database with column storage, but HBase itself is not directly involved in file storage, and its actual functions are still realized by HDFS under Hadoop framework. The design core of HBase is to realize random and real-time read/write access of HDFS system. As the representative of non-relational database, its main functional modules include HMaster, Region Server and ZooKeeper. [2] Among them, HBase supports users to complete various operations in HBase database through Java API under HBase Client, and HMaster is responsible for the distribution and management of Region Server, realizing the overall load balance of the system and the control of various permissions.

2.2.2 Spark

Spark is a more efficient distributed parallel computing framework. Compared with MapReduce, the whole calculation process of Spark framework is based on memory, which greatly reduces the reading and writing time of disk, and enables RDD (Resilient Distributed Datasets) as a new data structure to realize a more brief and concurrent calculation process. Spark framework contains four parts: tool layer, calculation layer, storage layer and resource scheduling layer, in which the calculation layer (Spark Core) is the running foundation of the whole framework, which is mainly responsible for the scheduling of distributed computing tasks and the basic control of input and output interfaces. [8] In addition, the Spark framework will be extended based on Spark Core, integrating four core components of Spark SQL, Spark

Streaming, MLlib and GraphX to enrich the application scenarios of Spark framework in different fields.

2.3 Javaweb

JavaWeb is the sum total of technologies that use Java language to complete the development of dynamic Web applications, which can help developers quickly solve various problems of Web client or Web server. JavaWeb is subordinate to J2EE technical specification, which can divide Web application into four parts: presentation layer, control layer, business logic layer and persistence layer, and support the development and deployment of each part by JSP page, Servlet, JavaBean or EJB and JDBC. In addition, with the continuous expansion of the scale of Web applications, the related functions and requirements are becoming more and more complex, and JavaWeb has entered the framework stage. Many frameworks represented by Structs, Spring, Hibernate and Mybatis have been integrated and changed to form SSH or SSM, which can greatly speed up the development process, reduce time and cost, and realize the agile development of web applications.

2.4 Data Mining

Data Mining (DM) is a process of extracting hidden information with certain potential value from a large number of disordered data information with the help of computer application technology. [3] The essence of data mining is to build a data analysis and processing model, taking data information as the research object, and providing help for subsequent decision-making through analysis and prediction. The process of data mining can be simply divided into four steps: problem definition, data preparation, data mining and result analysis. Among them, the choice of data mining algorithm is the core of the whole data mining work, and it is also the key to build the corresponding data mining model. Common data mining methods include classification analysis, prediction analysis, cluster analysis, valuation analysis and correlation analysis, as shown in Table 1. [7] As for the final result analysis, it is more inclined to the correlation between data, the trend and trend of data and the characteristic expression of specific objects.

Table 1: Type of common data mining algorithms (original)

Classify	Data mining method	Data mining model
Guided data mining	Classification analysis	Decision Tree, Random Forest, Neural Network
	Predictive analysis	Regression tree, Rough Set method
	Valuation analysis	SVM, Bayesian method
Unsupervised data mining	Correlation analysis	Pearson correlation coefficient, Aprior algorithm

2.5 Development Process

According to the application requirements of the above related application technologies, complete the configuration and deployment of the development environment of the application service platform of smart campus in colleges and universities. The development content of the system is divided into two parts. One is to build Hadoop cluster, complete the collection and storage management of all kinds of campus data information, and build a data analysis and

mining engine with the help of Spark framework. Secondly, under the Java development environment, the SSH framework is used to complete the development of Web Server, and the development and deployment of various lightweight service bus interfaces are completed, thus forming a standard Web application.

First of all, Hadoop cluster architecture needs the support of hardware and software. The underlying operating system is Linux, CentOS 6.7(x86_64) is the version, and jdk-8u291-linux-x64 is the JDK version. According to the application requirements of the system, Hadoop will be deployed in a completely distributed cluster. There are 7 nodes in the cluster, which are named Master1, Master2, Slave1, Slave2, Slave3, Slave4 and Slave5 respectively. Hadoop is version 2.7.7, which is installed in each node, and components such as Yarn, HDFS, Zookeeper, HBase, Sqoop and Kafka are also deployed in each node. Among them, Sqoop component can import all kinds of structured data into HDFS system under Hadoop architecture to realize distributed storage and form raw data. Kafka can import the operation logs and unstructured data of the system into HBase for storage.

Secondly, for the data mining analysis function, the system will use Spark framework to read all kinds of data in HDFS. After data preprocessing, it will complete the data call processing by constructing the corresponding data mining model. As shown in Figure 2, the K-means clustering algorithm code is implemented for Spark.

```
object Kmeans {
 def main(args:Array[String]) = {
 val conf= new SparkConf().setAppName("K-Means").setMaster("spark://master:7077")
   .setJars(Seq("E:\\Intellij\\Projects\\SimpleGraphX\\SimpleGraphX.jar"))
 val sc = new SparkContext(conf)
 val data = sc.textFile("hdfs://master:9000/kmeans_data.txt", 1)
 val parsedData = data.map(s => Vectors.dense(s.split(" ").map(_.toDouble)))
 val numClusters = 2
 val numIterations = 20
 val model = KMeans.train(parsedData, numClusters, numIterations)
 println("Cluster centres:")
 for(c <- model.clusterCenters) {
  println(" " + c.toString)   }
```

Figure 2: K-means algorithm model building code (original)

Finally, for the development of Web application server, the basic development language is Java, MyEclipse 2018 is the integrated environment, Tomcat 8.0 is the Web server, and MySQL 5.7 is the database server. And the project object model (Maven) is used to manage the project structure. Maven chooses Apache-Maven-3.2.1 version. In the process of building the overall development environment, the installation of JDK and the configuration of environment variables are completed first to build the foundation of Java application development. Secondly, the installation of MyEclipse and the installation of Tomcat, the Web server, and the configuration of Tomcat is completed in the Preference option under MyEclipse. Then, based on SSH architecture, complete the integration and encapsulation of the whole system. Through the introduction of the above key technical theories, the overall environment of the system development, the configuration of related software and tools are

determined, and the technical feasibility of the overall project of the application service platform of smart campus in colleges and universities is also clarified.

3 FUNCTION REALIZATION

3.1 Data Management

The platform has a unified initialization login interface, which makes the previous decentralized education and teaching system and campus management service system business platform integrated. The platform uses big data technology to establish a perfect data sharing center, collect the historical data of various existing systems on campus, complete the lossless transplantation and optimized integration of data resources, and realize the distributed storage of data information with Hadoop cluster to realize the comprehensive management of campus data. [4] Under the data management function module, the system will give different data usage rights according to different user roles. For example, student users can independently check the semester and academic year plans, keep abreast of the learning progress, and know their own credits and grades. Teacher users can quickly handle all kinds of administrative, party-mass and other daily work, realize the joint search and use of data and information, and effectively improve work efficiency.

3.2 Teaching Service

According to the actual application needs of students and teachers, the platform will set up two sub-functions of education and teaching management and scientific research achievement management under the teaching service module. Under the function of education and teaching management, teacher users can upload teaching materials, publish relevant teaching plans and curriculum tasks, arrange online homework and tests, and make statistical analysis of students' learning situation. The functions of student users include online learning, data downloading, online homework and testing.

As for the management function of scientific research achievements, teacher users have the right to complete scientific research project application, scientific research project audit, personnel information query, scientific research projects and achievements query, etc. in the platform, which can reduce a lot of complicated and trivial information collection, improve efficiency and ensure quality.

3.3 Campus Management

As the core application of "Smart Campus" platform, it will take students' daily behavior on campus as the research object, conduct multidimensional data analysis and mining, and provide scientific decision-making basis for the management of various affairs on campus. Taking the student behavior analysis model as an example, the data features are selected from five dimensions, including classroom learning, online learning, other ways of learning, campus life and daily entertainment, and the corresponding data sources come from the internal database of the platform, campus video surveillance system and campus "all-in-one card" system. As shown in Table 2, the information table of students' behavior characteristics, the platform will synthesize all kinds of characteristics to construct students' behavior patterns, that is, student user portraits.

Table 2: Information table of students' behavior characteristics (original)

No.	Data dimension	Data features	Example
1	Classroom learning	Course characteristics	Course name, course teacher
2		Performance characteristics	Normal grades, comprehensive grades
3	Online learning	Learning duration characteristics	System usage time
4		Job completion characteristics	Job completed, job not completed
5	Other ways to learn	Book borrowing characteristics	Borrowing time, book name
6	Campus life	Characteristics of dining in canteen	Location, amount
7		Shopping characteristics	Time, amount
8	Daily entertainment	Participation characteristics	Activity time, activity name

Through the user portrait, data mining analysis can be carried out for the corresponding problems. For example, in the analysis of association rules between students' behaviors and students' achievements by Apriori algorithm, the calculation formulas of association rules are shown in formulas 1 and 2. S stands for support, that is, the proportion of student behavior (X) and student achievement (Y) in all data sets; And c stands for confidence, that is, the ratio of student behavior (x), student achievement (y) and student behavior (x) support. (Gao, 2022) The results of the final calculation are shown in Table 3, and the platform can also display the corresponding data results in the form of charts, as shown in Figure 3, which is a scatter plot of the correlation between achievement and shopping characteristics.

$$S(X \Rightarrow Y) = \frac{\text{count}(X \cup Y)}{|D|} \tag{1}$$

$$C(X \Rightarrow Y) = \frac{S(X \cup Y)}{S(X)} \tag{2}$$

Table 3: calculation formula for some indexes (original)

No.	Consequent	Antecedent	Support (S)	Confidence (C)
1	Good grades	Learning duration characteristics	19.77	86.55
		Book borrowing characteristics		
		Participation characteristics		
2	Poor grades	Characteristics of dining in canteen	17.69	66.25
		Participation characteristics		
		Course characteristics		

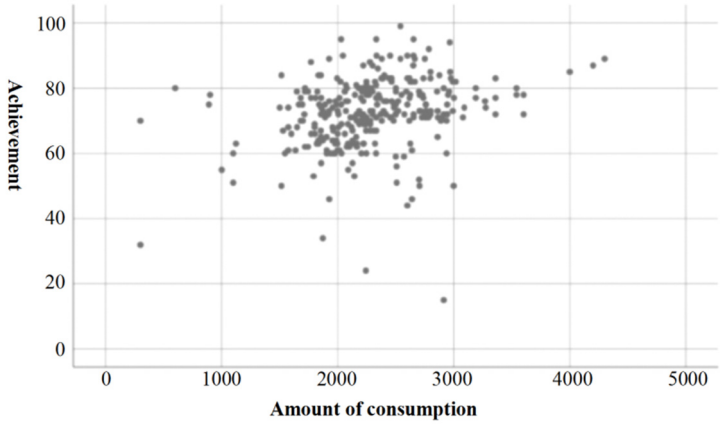

Figure 3: Scatter chart of correlation between achievement and shopping characteristics (original)

4 CONCLUSIONS

With the aim of building a "smart campus" in the new era, this paper builds an application service platform for smart campus in colleges and universities based on the functional characteristics of big data technology, network information technology and computer application technology. The function of the platform combines the current campus education and teaching systems and campus management service systems, completes the sharing and interaction of all kinds of data, and provides a brand-new data analysis system. It is convenient for users to realize the application and management of teaching, scientific research and daily management with concise, convenient and efficient operation, which improves the comprehensive utilization rate of data resources, improves the service level of campus, and makes a beneficial attempt to construct the wisdom education paradigm in colleges and universities.

REFERENCES

[1] Gao Shengxuan. (2022) Research on the Correlation between College Students' Daily Behavior and Achievement Based on Apriori Algorithm[D].Shenyang Normal University.05.

[2] Li Qingyun. (2015) Research and Implementation of Application Platform Based on HBase[D].Beijing University of Posts and Telecommunications.03.

[3] Meng Weicheng. (2022) Development and Application Analysis of Computer Data Mining Technology[J].Electronic Time.01.

[4] Niu Yongliang. (2022) Practice and Exploration of Smart Campus Construction in Colleges and Universities[J].Network Security Technology & Application.05.

[5] Shi Fangxia, Gao Yi. (2021) Hadoop Big Data Technology Application Analysis[J].Modern Electronic Technology.09.

[6] Tang Manqing. (2022) Research on Informatization Construction of University Education Management in Big Data Era[J].Neijiang Technology.11.

[7] Wang Mingxing. (2014) Research and Application of Data Mining Algorithm Optimization[D].Anhui University.04.
[8] Wu Xindong, Ji Shengwei. (2018) Comparison of MapReduce and Spark in Big Data Analysis[J].Journal of Software.02.
[9] Zhao Peng, Zhu Yilan. (2022) Overview and Development Prospect of Big Data Technology[J].Astronautical Systems Engineering Technology.01.
[10] Zhou Li. (2021) Research Status and Trend Analysis of Smart Campus in China's Universities[J].Journal of Huaiyin Institute of Technology.10.