Research on Application Platform of College English Teaching Based on Hadoop Data Bus

Lei Shen[1(✉)], Meichun Yang[1], Wentong Zhang[2], and Haie Jia[3]

[1] Accounting School, Tong Ling University, Tongling 244000, Anhui, China
SL05622628@163.com

[2] School of Foreign Languages & International Business, Guangdong Mechanical and Electrical Polytechnic, Guangzhou, Guangdong, China

[3] Tibet Agricultural and Animal Husbandry College, Nyingchi 860000, Tibet, China

Abstract. We have studied the application platform of college English teaching based on Hadoop data bus. The working principle of college English teaching application research is to use artificial intelligence (AI) to analyze the content of each curriculum plan for learners, and then use this information to create a personalized learning plan with customized exercises. These customized courses are provided through interactive video lectures driven by artificial intelligence, which are synchronized with students' personal devices or computers so that they can learn at their own pace. The research is divided into four parts: (1) Investigation of existing Hadoop based applications; (2) Develop a prototype application to collect and analyze big data using MapReduce framework in Hadoop; (3) Analyze performance characteristics, such as scalability and throughput; (4) Comparison with other frameworks (such as RDBMS).

Keywords: Hadoop data bus · English teaching · Platform research

1 Introduction

In the information age, information technology has been introduced into many fields. With the support of mobile devices, mobile applications, which have been popularized in recent years, have been introduced into education. Because convenient ways and mobile devices have become a part of public life, they have become a hot favorite in the education industry. Nowadays, English teaching in colleges and universities is being innovated, so it is necessary to give play to the educational value of the platform, innovate the teaching mode, build an English learning platform suitable for the development of college students, provide strong support for college students' English learning, and create more convenient and efficient learning ways [1]. When reading widely the core papers of theoretical research such as college English teaching method, it is not difficult for the researchers to find that the scholars or expert professors who wrote the papers are working in key universities, and their research results are mainly about foreign language theories. They analyze and explore the problems found in their teaching and research,

Y. Zhang and N. Shah (Eds.): BigIoT-EDU 2023, LNICST 580, pp. 565–570, 2024.
https://doi.org/10.1007/978-3-031-63130-6_63

which are theoretical and provide sufficient analytical data. However, there are few examples of specific problems in college English teaching, which is not conducive to the reference of local colleges. In contrast, there is an obvious gap between college English teaching in local colleges and universities, especially classroom teaching research and that in key colleges and universities. There is a lack of relevant theoretical guidance, and various problems are encountered in teaching practice. Even if the foreign language core journals publish the teaching research results of experts from famous universities, these practices are not suitable for the teaching and practice of local universities. The quality of students in different colleges and universities varies greatly, which determines the overall quality and ability of students and affects foreign language classroom teaching. In addition, due to the insufficient research on classroom teaching complications caused by the application of modern information education technology in the teaching papers of foreign language core journals and the emphasis on pure theory, attention and reflection should be paid [2]. This paper combines the advantages of the platform in English teaching and the feasibility of its application, and develops strategies for building a college English learning platform, hoping to play a role in promoting the development of higher education.

2 Related Work

2.1 Big Data Analysis Technology

With the rapid development of the Internet, cloud computing and mobile applications, today's society has entered the era of big data. According to Gartner's definition, big data can be understood as a large amount of data that traditional methods and tools cannot handle. It defines data sets that exceed the normal processing range and size and force users to adopt non-traditional processing methods. Wal Mart is a typical example [3]. This retail giant handles more than 1 million customer transactions per hour, and its database is estimated to contain more than 2.5 PB of data, equivalent to 167 times the information contained in all books of the Library of Congress.

Big data has four typical characteristics:

(1) The data volume is large and can reach the level of TB or even PB.
(2) There are many kinds of data, including structured, semi-structured, unstructured data.

    It also shows strong heterogeneity in its storage coding.
(3) Speed. Because of the large amount of big data, the calculation time will become longer, so the processing speed is required.
(4) The value density is low (Value), the amount of data is large, the useful information in the data is relatively small, and it is difficult to extract information.

3 Hadoop Distributed Processing Framework

In the big data scenario, data analysis and mining generally rely on various distributed computing frameworks, such as MapReduce, Prege, Dryad, etc., and are generally process based. For example, the platform mentioned in uses the BPEL process execution

engine. The traditional data analysis architecture can no longer meet the needs. For example, the platform in is based on Google App Engine, and data is transmitted to the cloud server through the network. This architecture is not suitable for large amounts of data, and the data is stored on an open service platform, which brings potential data security problems, which is unacceptable to many companies and individuals.

Hive: Hadoop based data warehouse can support structured storage of big data. Hive maps structured data to database tables, stores table structure information in traditional relational databases, and provides simple SQL query interfaces. After the user executes the SQL query, Hive converts the SQL into MapReduce calculation and submits it to the Hadoop system for execution.

Mahout: Mahout is an algorithm library for data mining and machine learning written using MapReduce programming framework [4]. It provides many scalable classical data mining algorithms, including clustering, classification, collaborative filtering recommendation algorithms, and so on.

Sqoop: Sqoop can be used to batch transfer data between traditional databases and HDFS. It is a Hadoop based batch data import and export tool.

Oozie: This is a scalable, multi tenant based, secure and easy-to-use workflow management system developed specifically for Hadoop. It can manage the scheduling between multiple tasks. Oozie connects these tasks with directed edges, adds necessary control nodes, and finally generates a directed acyclic graph (DAG). Oozie uses XML to define these tasks and their relationships.

4 Research on College English Teaching Application Platform Based on Hadoop Data Bus

The goal of this paper is to design and implement a Hadoop based big data analysis college English teaching application development platform. Developing a one-stop big data analysis college English teaching application development platform can greatly speed up the development of data analysis college English teaching applications and save resources. A perfect and one-stop big data analysis college English teaching application development platform should provide various functions related to big data analysis college English teaching.

Big data analysis college English teaching applications generally run on distributed clusters, and these resources are generally shared [5]. Therefore, the development platform should first support multi tenancy (multiple users can use the tools simultaneously and independently). Each user has an independent personal workspace in which users can store data resources, develop data analysis college English teaching applications, design reports, etc. Figure 1 shows the hadoop framework.

For general data analysis college English teaching needs, the data analysis college English teaching application development platform should integrate most of the commonly used data analysis college English teaching algorithms [6]. These algorithms can be selected by users, and users can also submit their own algorithms to the development platform. The architecture of Hadoop based big data analysis college English teaching application development platform is shown in Fig. 2. In the user layer, users can see the component list provided by the development platform, and use these component lists

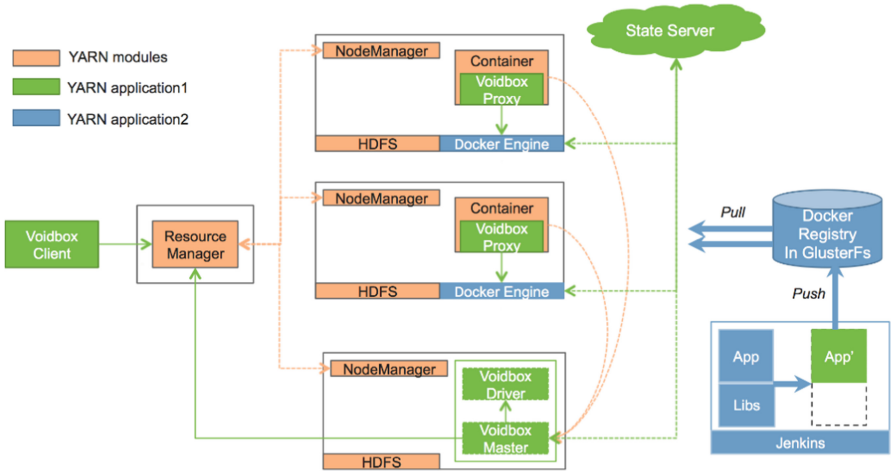

Fig. 1. Hadoop framework

to build data analysis college English teaching process. The component list of the user layer is consistent with the algorithm library of the processing layer. After the user has designed the data analysis college English teaching business process, submit the process to the tool background, and the model layer will translate the user's data analysis college English teaching business process into an Oozie based execution process. Then, the data analysis college English teaching task submission engine submits Oozie's execution process and related executable components to the platform execution layer—Hadoop distributed cluster [7]. The algorithm in the algorithm library of the tool can use the MapReduce programming framework and call Mahout, Weka, R and other algorithm libraries.

5 Implementation Method of English Teaching Application Platform

The data analysis college English teaching component model is an abstract description and implementation of common functions and algorithms such as data processing, data mining, data statistics, etc. At the demand and business level, that is, from the user perspective, each component contains.

Segment function description and configuration interface. For example, data cleaning, clustering algorithm, or classification algorithm. From the perspective of execution, each component is an executable node, including an executable code and a callable interface. From the perspective of component developers, they should not only define component descriptions and configuration interfaces, but also write executable code and call interfaces.

The big data analysis college English teaching application development platform should be open and customizable. In this way, we can accumulate some excellent algorithms with the help of open source and grow into a mature and sound tool. According

to the principle of openness, the system supports user-defined components. Users follow the specification, develop their own components and upload them through the client. After uploading, users can see and use this component on the client. In the same way, the system should also support the user-defined components uploaded by users before uninstalling through the client, so that users cannot use the components on the client [8]. The scalability of components is very important, which directly determines the availability of the system. Users can build a domain oriented big data analysis college English teaching application development platform by customizing domain components.

In Oozie, different types of execution nodes have different parameter configurations and triggering methods. This article gives the parameter information that Java program and Hive program need to configure in Oozie configuration file. From the XML definitions of the two nodes, you can see the Java type nodes. You need to define the class of the main function and give the input parameters of the Java program. The Hive node does not need to specify this information. In addition, the Hive node needs to specify the Hive script to be executed [9]. The configuration parameter information is shown in Fig. 2.

```
<xs:complexType name="JAVA">
    <xs:sequence>
        <xs:element name="job-tracker" ... />
        <xs:element name="name-node" ... />
        <xs:element name="prepare" ... />
        <xs:element name="job-xml" ... />
        <xs:element name="configuration" ... />
        <xs:element name="main-class" ... />
        <xs:choice ... >
            <xs:element name="java-opts" ... />
             <xs:element name="java-opt" ... />
        </xs:choice ... >
        <xs:element name="arg" ... />
        <xs:element name="file" ... />
        <xs:element name="archive" ... />
        <xs:element name="capture-output" ... />
    </xs:sequence>
</xs:complexType>
```

Fig. 2. Configuration parameter information

In Java, a concept is generally represented by a Java class, so here we design an abstract class AbstractModule to represent components, and use the methods in the class to constrain the interfaces that each component must provide [10]. In addition, the development platform uses Java polymorphism to realize different representations of different types of components.

The user first chooses to inherit an AbstractModule subclass according to the running mode of the algorithm, then adds the Java annotation @ Module, defines the component's meta information according to the annotation specification, and then uses Eclipse

or Java Jar tools to package the necessary files into Jar packages and submit them to the platform [11]. The platform obtains different component parameters and XML configuration templates according to different component types at runtime, and adopts different translation strategies.

6 Conclusion

The research on college English teaching application platform is a research conducted by the organization, which aims to find out the problems faced by students in learning English and how to solve these problems. Many organizations have been studying this problem, some of the most famous organizations include: National Language Acquisition Research Center (NCRLA), Modern Language Association, etc. They all have different ideas about what needs to be done, but they all agree that there should be better English teaching methods so that it can help students learn faster. The purpose of such research is to ensure that students have all the skills they need for their future careers so that they can communicate as much as possible with others from different countries.

Acknowledgements. The innovation project of university students in Anhui (No: AH20140383095).

References

1. Qian, Z., University, Y.: On the teaching of college English writing based on the platform of digital writing. J. Shandong Agric. Eng. Univ. (2016)
2. Wu, Z.: Research on the innovation of hybrid teaching under the platform of rain classroom—Take "Advanced Business English Course" as an example. In: 2019 3rd International Conference on Data Science and Business Analytics (ICDSBA) (2019)
3. Wang, D., Zhang, Y.: A study on the flipped classroom model in college English listening and speaking teaching based on Longman interactive English teaching platform. J. Anshan Normal Univ. (2016)
4. Jiajia, L.: A college English teaching mode based on a computer network platform. Agro Food Ind. Hi Tech **28**(1), 616–619 (2017)
5. Hui, Q.: Design of the innovative teaching platform for college English linguistics teaching based on the network technology (2020)
6. Zhang, Y.: Research on college English online learning platform model based on big data technology. J. Phys: Conf. Ser. **1648**, 042090 (2020)
7. Hua, F., Liu, C.: The application of big data analysis in college English classroom vocabulary memory and learning efficiency teaching (2021)
8. Xiao-Hong, Q.: An analysis of application of register theory in teaching college English writing **19**(10), 6 (2022)
9. Gonzalez-Valero, A., Reeves, A.G., Page, A.C.S., et al.: An Activity-based Oxaziridine platform for identifying and developing covalent ligands for functional allosteric methionine sites: redox-dependent inhibition of cyclin-dependent kinase **4** (2022)
10. Yuan, Y.: An empirical study of the hybrid teaching model based on network teaching platform in college English teaching. J. Hubei Open Vocat. College (2019)
11. Meng, N.: Flipped classroom teaching of college English based on Wechat platform. J. Changzhou College Inf. Technol. (2017)