Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 2, Issue 1: 102-109, DOI: 10.25236/AJCIS.010023

Design of College Students' Physique Monitoring and Disease Warning System Based on Hadoop

Ying Zou, Leian Liu*, Xisheng Chen, Xixin Huang and Minfeng Wei

College of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China
**Corresponding author e-mail: lla@zhku.edu.cn*

ABSTRACT. *In this paper, based on college students' physique detection data and the big data processing tool, Hadoop, a system for college students' physique monitoring and disease warning based on Hadoop is designed. The system consists of Hadoop distributed system architecture, Hadoop-based Hive data warehouse tool, Sqoop data transmission tool, database technology and data visualization system, etc. The system takes Hadoop architecture as the core, mainly through the storage, statistics, analysis and visualization management of the physique monitoring data to monitor the health of the students, to achieve prediction and early warning of susceptible diseases, and to generate analysis reports to help developing more reasonable sports programs aiming to improve the physique of the students and improve their health.*

KEYWORDS: *Hadoop, data analysis, data visualization, physique monitoring, disease warning*

1. Introduction

College students, as reserve talents for national development, shoulder the historical responsibility of national prosperity and rejuvenation. Their physical health is not only closely related to their own lives, but also directly related to the future development of China and the rise and fall of the nation. The improvement and enhancement of their physical health are conducive to the economic and social development of the country. For a long time, the Chinese government has always attached importance to the physical health of adolescents [1-3]. As early as in the 1950s, China began to conduct physical fitness tests for college students. In the late 1970s, it began to explore the management model of college students' physical fitness test. In the 1990s, China has established a national physical fitness monitoring and management system, and began to conduct college students'

ISSN 2616-5775 Vol. 2, Issue 1: 102-109, DOI: 10.25236/AJCIS.010023

physical fitness test management. In 2002, the Ministry of Education piloted the "National standards for students' physical health" and established a national monitoring network for students' physical health. At present, domestic scholars are also very concerned about the problems related to college students' physical health, and have conducted research on these issues. The literature [1] took Guangzhou as an example to discuss the common problems of the current college students' physique health management mechanism, and put forward a new idea of constructing a new college students' physique health management mechanism. The literature [2] investigated the physical health status of students in Nanchang Hangkong University in 2014 by taking the four grades who participated in the physical fitness monitoring as the research objects. Based on the specific situation of the students in Nanchang Hangkong University, the factors affecting the physical health of the students in the university were analyzed from the aspects of environmental factors and self-factors. The literature [4] studied the issues related to the monitoring and development of college students' physique, and constructed the college students' physique monitoring system, physique data network management service platform and scientific construction service platform. The literature [5] analyzed the current health status of college students and the main problems existing in the monitoring mechanism of college students' physical health, and focused on the innovative measures of the monitoring mechanism. The literature [6] discussed the reasons for the decline of the physical health status of college students in China, and explored the implementation path for improving the physical health of college students. The literature [7] analyzed the current situation of monitoring and management of college students' physique in China, and proposed strategies to improve college students' physique health. The literature [8] studied the policies and regulations on the physical monitoring management of college students in China from the perspectives of development process and implementation methods. In general, these studies show that the physique of college students in China has been declining for many years. However, as far as the author knows, there is no system that uses college students' physique monitoring data for big data analysis and uses the results of big data analysis for disease warning. At present, in response to the requirements of national policies, colleges and universities arrange one or two physical fitness testing programs every year to record the indicators of students' physical conditions. However, in the current situation, the data collected by colleges and universities are only used to complete the evaluation of grades, without in-depth mining and processing, so as to maximize the value of the data. With the rapid development of cloud computing, mobile Internet and Internet of things technology, various industries have entered the era of big data, and the development of big data applications has attracted the attention of various industries. Hadoop distributed infrastructure mainly provides HDFS distributed file system to store data and map-reduce parallel computing framework. Its high reliability, high extension, high efficiency, high fault tolerance, low cost and other advantages enable developers to easily develop and run applications processing massive data on its basis. In addition, other technologies based on Hadoop, such as Hive and Sqoop, respectively provide data statistical analysis, data transmission and other functions. Data visualization technology can present data in a graphical way to make the data clearer for deeper

ISSN 2616-5775 Vol. 2, Issue 1: 102-109, DOI: 10.25236/AJCIS.010023

observation and analysis. In this paper, based on the constitution detection data of college students in our university and the big data processing tool, Hadoop, a constitution monitoring and disease early warning system based on Hadoop technology is designed to provide strong technical support for the constitution health of college students in our university.

2. System basic structure

The basic structure of the system is shown in figure 1. As the lowest level architecture, Hadoop server cluster, HDFS distributed file system is responsible for receiving and storing external data, and MapReduce combines with Hive data warehouse for data cleaning and statistical analysis. Sqoop connects Hive with Mysql database for data transfer. At last, the data results are visualized through the website pages, and then the information prediction and scheme formulation are carried out.

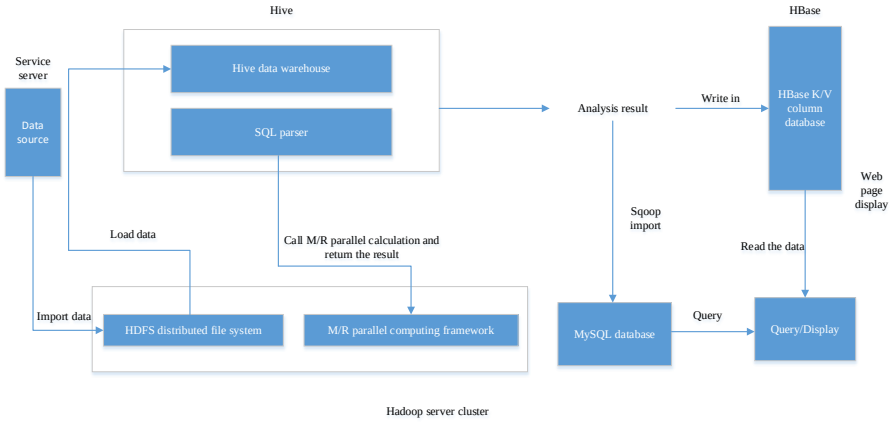

Figure. 1 System basic structure diagram

3 System data processing module

3.1 Data storage

The data collected in each physical examination will be saved to the local file system. The file will be uploaded to the HDFS distributed file system using the shell command. HDFS will split the data into fixed size data blocks and save them to different DataNode data nodes. The data storage process is shown in figure 2.

ISSN 2616-5775 Vol. 2, Issue 1: 102-109, DOI: 10.25236/AJCIS.010023

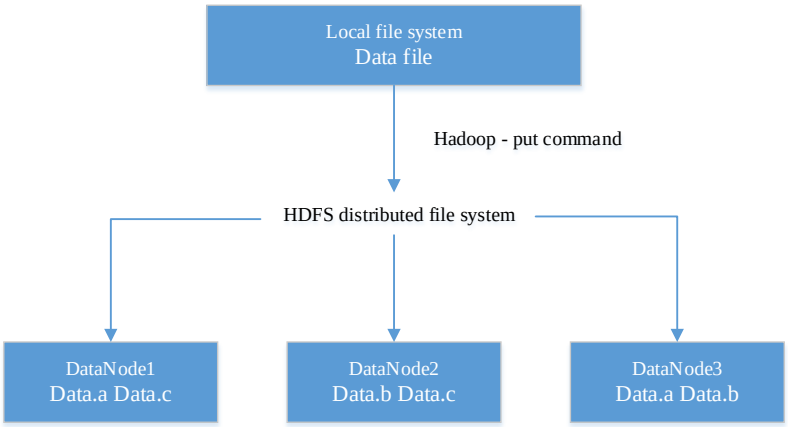

Figure. 2 Data stored procedure diagram

3.2 Data cleaning and conversion

The data stored in the HDFS distributed file system is the original physical examination data. In addition to the body-related data of the tested students, it also includes useless data such as project numbers, which need to be cleaned out. MapReduce is a programming model that uses *map( )* and *reduce( )* functions to write operational programs with distributed operation on Hadoop, a distributed system, which can process large-scale data. The *map( )* function preanalyzes and preprocesses the original data, and the *reduce( )* function merges the data after the analysis of the *map( )* function and draws the conclusion of induction. The functions of data cleaning and data conversion are as follows.

● Data cleaning: students' information and body data, such as height, weight, lung capacity, endurance, flexibility, speed and dexterity, are retained, and the date of birth and project numbers are cleaned.

● Data conversion: the physical examination data are mainly divided into three aspects: body shape, body function and physical quality. According to the implementation method of "National standards for students' physical health", the corresponding scores of each data are calculated, and the total scores are calculated according to the proportion of corresponding scores.

The data conversion process is shown in figure 3.

ISSN 2616-5775 Vol. 2, Issue 1: 102-109, DOI: 10.25236/AJCIS.010023

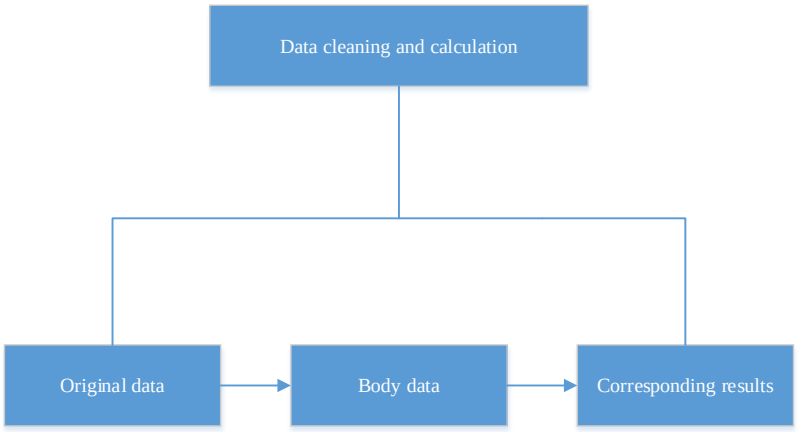

Figure. 3 Data conversion diagram

3.3 Statistical analysis of data

Hive data warehouse can map data files into a database table and provide SQL query function. It can quickly realize simple MapReduce statistics through SQL statements, which is very suitable for the statistical analysis of data warehouse. After preliminary data cleaning and calculation of MapReduce program, data sets, Dataclean and Datacount, are saved in HDFS. Load the above two data sets into the partition table built in the Hive data warehouse, so as to store and analyze the data according to the upload time. Dataclean and Datacount are the raw data of physical examination and the score data, respectively. The detailed field descriptions are shown in figure 4.

```
hive> describe dataclean
OK
stunum  string
sex     string
age     string
height  string
weight  string
bmi     string
vc      string
stamina string
fp      string
sn      string
time    string
Time taken: 0.13 seconds
```

```
hive> describe datacount;
OK
stunum  string
sex     string
bmi_grade       string
vc_grade        string
stamina_grade   string
fp_grade        string
sn_grade        string
total_points    string
time    string
Time taken: 0.144 seconds
```

Figure. 4 Hive data table field descriptions

On the basis of these two tables, the following data table is obtained through further statistical analysis.

ISSN 2616-5775 Vol. 2, Issue 1: 102-109, DOI: 10.25236/AJCIS.010023

Table 1 Data field description table

Name	Content
num_of_people	Number of people
bmi_count	Body mass index (BMI) number distribution
clean_avg	Average medical data of each item
count_overall_avg	Average score of each item
count_grade	Distribution table of total grades

3.4 Export data to MySQL database

Sqoop is a tool for transferring data between Hadoop and relational database. It can import data from a relational database (such as MySQL, Oracle, Postgres, etc.) into HDFS of Hadoop, and also import data from HDFS into relational database. First, the mysql_t table is created in the MySQL database to hold the hive_t table data exported by Hive. The Sqoop command is then executed for the data export operation as follows.

Sqoop export –connect jdbc:mysql://hadoop1:3306/hive –username root –password hadoop –table mysql_t -export-dir /hive/cda.db/hive_t --fields-terminated-by'\t'

4. Data visualization and report generation module

The data visualization used in this system is to display the website as a means of data table information in the database through the back-end program call, combined with the front-end page to achieve data visualization.

4.1 Development tools

The data visualization tool used in this system is d3.js. The Web development technologies used are PHP, HTML5, CSS, JavaScript. The full name of D3 is "Data-driven Documents", a javascript library about data-driven documents. D3.js is mainly used to manipulate data. It injects life into the data through the use of HTML, SVG and CSS, and transforms the data into various graphics that are both easy to understand and gorgeous. PHP is an open source and free web development language, with the advantages of simple editing, strong practicality, strong cross-platform, high efficiency and object-oriented, etc. In particular, PHP is not only good support for dynamic image processing, but also extremely close connection between PHP and MySQL database.

ISSN 2616-5775 Vol. 2, Issue 1: 102-109, DOI: 10.25236/AJCIS.010023

4.2 Data visualization design

The data visualization principle of this system is to use PHP program to query the data table information in MySQL database, and load the query results into the front-end page for dynamic display. D3.js, combined with HTML and CSS, can make interactive graphics of web pages more dynamic and clearly and effectively convey the deep meaning of data. The specific steps of data visualization function implementation are shown in figure 5.

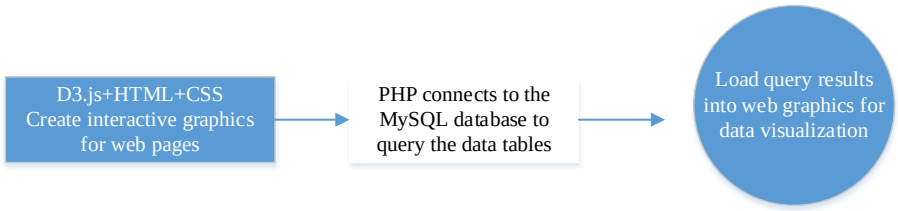

Figure. 5 Data visualization process

4.3 Analysis report generation function

After cleaning, calculation and statistical analysis of the original measurement data, the final results are visualized in the form of graphics on the web page. The data analysis report is the final result of a series of procedures of the system, which is mainly composed of the list of vulnerable diseases and the disease prevention and improvement plan, reflecting the potential value of the data. Physical test data are mainly related to the body function, body shape and physical quality of the human body in three aspects. According to the relevant research conclusions in the medical field, combined with the data analysis results, the diseases prone to occur are speculated, and the disease list is generated. In view of the high and low incidence of diseases in the list, the main disease prevention measures and physical health promotion programs are generated for the reference of relevant departments of the school to develop physical exercise courses for students.

5. Conclusion

For college students' physique detection data, the system designed in this paper use the big data processing technology, Hadoop, to solve the problem of data waste, and the visual display of data makes the physique monitoring more clear and convenient. Based on the analysis of the data results and the reasonable combination of the three factors affecting human health, this paper gives an early warning of the vulnerable diseases, and generates a targeted improvement plan to help the our university develop more suitable physical education courses and activities for the students.

ISSN 2616-5775 Vol. 2, Issue 1: 102-109, DOI: 10.25236/AJCIS.010023

Acknowledgements

This work was financially supported by the provincial-level characteristic specialty-"Network Engineering" and the provincial teaching team-"Teaching team of basic core course of computer major" (Official document by Department of education of Guangdong province ([2017] 214)), 2018 higher education teaching reform project of Guangdong province-"Reform and practice of the training mode of network engineering talents based on the cooperation of school-school and school-enterprise" (Official document by Department of education of Guangdong province ([2018] no. 180)), the university-level high-quality resource sharing course project-"Network Security Technology" and "Wireless Sensor Network and RFID Technology", the university-level education and teaching reform project-" Research on training mode of innovative network engineering talents" and "Network technology core teaching team", the Science and Technology Planning Project of Guangdong Province under Grant (2017A070709012), and the National University Students' innovation and entrepreneurship projects (201611347010, 201711347004 and 201811347008).

References

[1] Q. Kuang (2015). Innovation research of the health management mechanism of college students' physical. Master thesis of South China University of Technology.

[2] W. Wang (2016). Research on the current situation and influencing factors of college students' physical health-Take Nanchang Hangkong University as an example. Master thesis of Nanchang Hangkong University.

[3] H. Y. Yu (2014). Evolution course and stage feature of student physical fitness test in China. Journal of Beijing Sport University, vol. 37, no. 10, p. 113-118.

[4] W. L. Cai and H. Zhang (2013). A study on the monitor and development of university students'physique. Journal of Henan Institute of Education (Natural Science Edition), vol. 22, no. 3, p. 75-77.

[5] H. S. Liu, X. L. Zhang, and J. Shou (2014). Innovative analysis on the monitoring mechanism of college students' physical health. Journal of Wuzhou University, vol. 24, no. 3, p. 90-93.

[6] L. Chen, W. Zhou, and T. Y. Pu (2018). On the implementation path to physique health promotion of college students under the national fitness program. Sports Research and Education, vol. 33, no. 1, p. 48-52.

[7] J. S. Lai (2016). On analysis and countermeasure to current situation of monitoring and management of chinese college students' physical quality. Journal of Southwest China Normal University (Natural Science Edition), Vol. 41, No.4, p. 162-168.

[8] J. S. Lai and W. P. Yu (2016). Researches on the current situation of chinese college students' physique supervision and management policies and regulations. Journal of Jilin Sport University, vol. 32, no. 2, p. 36-40.