

# Construction of Big Data Precision Marketing System Based on Hadoop

Daifen Tang

Chongqing College of Architecture and Technology, Chongqing, 401331, China

**Keywords:** Big Data, Accurate Marketing, Hadoop, Network Construction.

**Abstract:** In order to promote the information economization process of small and medium-sized enterprises in China, promote the prosperity and development of market economy, and improve the problems of poor information management conditions and weak internet background marketing ability of traditional enterprises, this paper combines big data technology to establish an accurate marketing application software. In this system, web crawler technology is used to capture URL data of web pages, and hadoop platform is used to collect, clean, calculate and process the data, and javaweb technology is used to realize data visualization. The system uses apriori association rule algorithm and clustering algorithm to effectively help e-commerce enterprise managers understand customers' needs and consumption preferences, accurately analyze old and new consumption data to improve consumer loyalty, stabilize existing consumer groups, increase potential customer groups, improve sales and marketing plans of enterprises, improve sales performance of enterprises, and help enterprises obtain more ideal economic benefits.

## 1 INTRODUCTION

In the current era, network technology has been popularized, and Internet application technology has developed rapidly. With the birth of online payment means, people's production and lifestyle are closely related to the Internet. As many traditional small and medium-sized enterprises are not aware of the importance of Internet technology combined with marketing, it is difficult to obtain the complete data information of consumers in the current era, which leads to the problem that marketing strategies do not match the market trend and environment, and they cannot fully meet consumers' demand for goods, resulting in low conversion rate of consumers, and eventually losing money until the enterprises disappear. Moreover, the explosive growth of consumer information data also makes more and more local server databases of enterprises overwhelmed and difficult to support. But there are important commercial values behind the huge data information, and small and medium-sized enterprises need to make marketing decisions based on these effective data analysis reports. (Zhang, 2020)

The rational use of big data technology can first help enterprises to integrate a large amount of consumer information data for unified management, and use various algorithms in data mining technology to

understand consumers' precise needs and preferences according to the data, so as to improve the current enterprise marketing strategy and promote the economic benefits of enterprises.

According to the above analysis, the author believes that in order to meet the needs of today's e-commerce platform enterprise managers, an e-commerce precision marketing system based on Hadoop platform web technology and web crawler technology came into being under the background of big data. The users of hadoop-based precision marketing application system are managers of small and medium-sized enterprises. According to the needs of users, it can help enterprise managers to collect consumer data from various channels to establish data centers managed by enterprises independently. With the advantages of big data technology, the data of different types, different channels and based on different communication protocols are unified and integrated, and the data exchange is realized.

## 2 KEY TECHNOLOGIES

### 2.1 Python Crawler Technology

The web crawler technology refers to the technology of automatically extracting web information by a

class of programs according to Internet rules. Python is the most commonly used development language for this technology. The principle of web crawler technology is realized by setting up new crawling rules and setting the URL of the portal. (Ji, 2017) Firstly, the developer selects a certain amount of seeds according to the requirements and saves the corresponding URLs. Then, a URL queue to be grabbed is set by the algorithm to save the selected URLs. After that, the program starts to download the contents corresponding to these URLs and grab the key information, and the processed URLs will be saved in the new grabbed queue. In the meantime, DNS resolution data and webpage download data generated by URL resolution will also be saved in the downloaded webpage database.

## 2.2 Hadoop Processing Platform

As a development and application ecosystem, Hadoop platform can support data-intensive applications, and its component team is growing with time. The most important components are distributed file system HDFS and parallel programming model MapReduce. The HDFS is responsible for the distributed storage of massive data, while mapreduce is to realize centralized parallel computing of distributed data, and the two complement each other. Hadoop ecosystem has many subprojects including Ambari, Hive, HBase, Zookeeper, Flume, Mahout, etc. besides Hadoop and mapreduce. With the cooperation of multiple components and clear division of labor, even inexperienced developers can use the advantages of clusters to deal with big data conveniently and quickly. (Li, 2017)

## 2.3 J2EE Framework

The J2EE is a simplified javaweb development platform designed and developed by SUN Company, which can develop a series of application software platforms. In order to simplify the application software development program of large enterprises, J2EE has specially developed a reusable component module to improve the development efficiency. Besides, it has also built a structure that can automatically handle the level, thus reducing the skill requirements of developers in developing application software. (Ma, 2022)

## 2.4 Development Environment

In this paper, the author briefly introduces the related technologies of platform development and use. In the

big data precision marketing system, Hadoop is used as a big data server cluster to process data and store it in MySQL database, and the corresponding application platform is developed by using JavaWeb technology.

According to the data volume and overall operation requirements of the system, this paper chooses to build a Hadoop3.3.1 cluster with three nodes. Then, the distributed collaboration system zookeeper-3.4.1, distributed file system HDFS 2.6.5, flume1.9.0, Hive 0.13.1 and Hbase2.6.5 are installed and deployed in these three nodes synchronously, and the initial construction of hadoop cluster is completed. The cluster will be developed under Linux system. This paper selects Centos6.5 Server release version of Linux operating system. The version of the web crawler framework Scrapy is 2.5, and Python3.8 is chosen as the development language. (Lin, 2016)

In this system, the front-end development tool of JavaWeb application is bootstrap+jquery, and the development language is JavaScript+HTML+CSS. The back-end Java development tool is IDEA 2021.1.3 (Ultimate Edition), the development environment is JDK 1.8, and the J2EE framework of Tomcat+Spring MVC+Spring+MyBatis is used in the implementation of this system. The development language is Java, and MySQL 8.0.28 is selected to help manage data.

## 3 OVERALL DESIGN

According to the needs of enterprises, hadoop-based Big Data Precision Marketing System establishes a top-down one-stop data collection, analysis, processing and visualization system. The main functions of data collection, data storage, data cleaning, data query and data analysis are supported by hadoop ecological cluster, and visualization is realized by javaweb technology.

First of all, collect data from three sources. One is the collection of local enterprise server data by flume, two is the URL data collected from the product details page by python web crawler technology, and the last is the access to Taobao, Weibo and other shared data through external JDBC interface. These data will be preliminarily cached in HDFS distributed storage. And the data of the crawler set is stored by redis. The data calculation module is implemented by mapreduce, which analyzes the preliminary data and manages the crawler results, and uses data mining techniques such as association rule algorithm to achieve the portrait of consumers. After processing, the data will be saved in HDFS and hive.

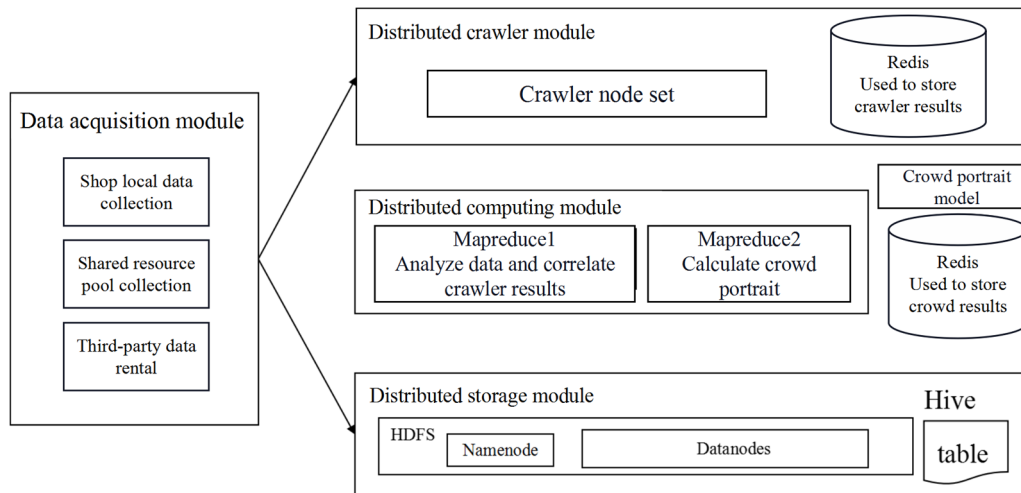


Figure 1: Data pool establishment code.

```

<bean id= "dataSource" class= "com. zaxxer: hikari. HikariDataSource" >
<property name= "driverClassName" value= "${jdbc. driver}" />
<property name= "jdbcUrl" value= "${jdbc. url}" />
<property name= "username" value= "${jdbc. username}" />
<property name= "password" value= "${jdbc password}" />
</bean>

```

Figure 2: Data pool establishment code.

The overall design of javaweb of this system chooses B/S mode and adopts MVC for development. The architecture is designed and developed by the traditional three-tier architecture of J2EE, which is the control layer, the business layer and the persistence layer. The business logic design of the core function of the whole system is developed by spring, the control layer is used to design the interactive function of client display, which is designed by spring-mvc, and the data persistence layer uses mybatis, as shown in Figure 1.

#### 4 FUNCTION REALIZATION OVERALL DESIGN

The enterprise marketing data analysis system based on big data technology is aimed at small and medium-sized entities operating enterprises. When the user logs in to the system through the account password, you can see three major data analysis function modules: existing customer data, potential user data and target market data, which are developed according to the user's needs.

#### 4.1 Existing Customer Data

In the existing customer data, the data is mainly the data information of existing customers stored locally, and the main form is static data. In addition to the most basic personal information of customers' age, gender and region, there are also consumption information of each customer, such as consumption amount, consumption amount and consumer goods category. For the core function of the user system is the search engine function, which requires that the front end of the application system will send out a large number of request instructions for interaction with the back end of the database server. In this process, JDBC needs to be built and destroyed continuously, which wastes the resources and memory of Internet databases. Therefore, this system also introduces a subsystem of database connection pool, which allows applications to reuse database connections. The partial implementation code is shown in Figure 2.

#### 4.2 Potential Customer Data

The data of potential users is the key function of this system, through which users can obtain the information of potential users' consumption characteris-

tics, and enterprise customers can accurately deliver to the promotion groups according to this information.

According to the local enterprise server data and the URL data collected by python web crawler technology, the system obtains the Internet shared data such as Taobao and Weibo through the external JDBC interface for unified integration to form the data of potential consumers. Establish data model through Hadoop to build portraits of consumer groups. According to the consumer information data, the crowd portrait model can help enterprise marketing managers accurately determine the key consumer groups of products according to the characteristics of consumer groups. In the realization algorithm of this function, k-mens clustering algorithm is used to subdivide users. The system obtains data according to the ID of consumers for integration, and the consumption records are the key data set used by k-means algorithm. The system sets these data records as feature vectors for analogy cluster analysis, and the more consumers are used, the more accurate it is. The formula is defined as follows, where P is the standard point, and E is the sum of squared errors of the selected object data, which is the mean value.

$$E = \sum_{k=1}^k \sum |p - m_i|^2$$

Firstly, through the existing data, we can obtain the Boolean mark list of consumers' goods according to the user ID to obtain the initial data. The system obtains the strong correlation information between commodities by calculating the confidence. For example, if the execution degree of commodity A and commodity B is 0.5, half of consumers who buy commodity A will also buy commodity B. Therefore, the system will advise users to increase the promotion of commodity B among consumers of commodity A, and the buyers of commodity A are potential users of commodity B. The quantitative definition of strong association rules is the minimum confidence threshold, named min\_conf. In this system, apriori rule analysis algorithm is used to find frequent itemsets, and this algorithm is used to collect commodity data information of association rule algorithm.

## 5 CONCLUSION

Due to the limited ability and time of the author, there are still many shortcomings in this research.

The depth and breadth of the research content need to be improved, and the follow-up work needs the support of more experts and scholars. The data research of this paper stays in qualitative research but lacks quantitative analysis. Meanwhile, the article is not specific to a certain industry, so although it has certain applicability to various industries, the specific conditions of different industries are different, and their application strategies are also different. It lacks systematic marketing recommendation method and deeper user segmentation method, which needs further improvement.

## REFERENCES

- Ji Xiaoyan. Research and Implementation of Personalized Recommendation in E-commerce Based on Hadoop. [D]. Lanzhou Jiaotong University.2017.06.
- Li Zhang. Design and Implementation of Internet Data Marketing System Based on Hadoop. [D].University of the Chinese Academy of Sciences.2017.04.
- Lin Qingpeng. Research on Precision Marketing Strategy Based on Big Data Mining. [D]. Lanzhou University of Technology.2016.04.
- Ma Xiaohong. Application and Value Overview of Big Data Precision Marketing in E-commerce. [J]. Electronic commerce.2022.01.
- Zhang Wenhui. Research on BJ Company's Precision Marketing Strategy under the Background of Big Data. [D]. Hebei GEO University.2020.12.