

Design and Development of Enterprise Marketing Data Analysis System Based on Big Data Technology

Yi Zhou¹, Yu Chen²

¹304810996@qq.com

²550417230@qq.com

¹Chongqing College of Architecture and Technology, Chongqing, China

²Chongqing College of Architecture and Technology, Chongqing, China

Abstract: With the rapid development of information technology and Internet technology, network marketing platforms are constantly enriched, and the integration of marketing data is getting higher and higher, which has had an impact on the traditional marketing strategy of enterprises and brought opportunities for transformation and development. In order to promote enterprise marketing into a new stage of development, this paper uses Hadoop ecology to effectively collect, analyze and process information and data with the help of big data-related technical means, and constantly innovate marketing strategies. Combining the challenges faced by enterprise development under the background of big data development, this paper discusses the countermeasures for the innovative development of marketing strategy in the era of big data, and develops three data analysis function modules of existing customer data, potential customer data and target market data, which can effectively help enterprises stabilize Existing customers, mining potential customers and target markets are conducive to the innovation and development of small and medium-sized enterprises in the era of big data for marketing strategies.

Keywords: big data; Hadoop; enterprise marketing; data analysis

1 Introduction

With the development of the Internet and the popularization of Internet devices, there are more and more ways of information dissemination, the rapid development of Internet information technology, and people's social modes have also undergone major changes. The rise and promotion of various software has not only changed people's social interaction, but also changed people's shopping, living and other habits. Various types of software such as social networks and intelligent platforms can already solve most of the needs of life, and generate a large amount of data while meeting the needs or solving problems. For example, Taobao can generate about 20TB of data every day. The explosive growth of Internet data has led to the emergence of big data technology. Big data technology is a technology to quickly obtain valuable data information from different types of data, and the data era has arrived.

Changes in user habits and needs will inevitably bring about changes in corporate marketing models. Marketing refers to companies discovering or excavating quasi-consumer needs, promoting and selling products from the overall atmosphere and their own products, digging deep into product connotations, and meeting consumers' needs. Demand, so that consumers can deeply understand the product, and then buy the product, is the most important stage to effectively improve the profit of the enterprise. Before planning marketing promotion activities, there is no accurate and reasonable data to support the activities, which may easily lead to the phenomenon of high input and low output, so that the marketing promotion activities do not achieve the expected results. In addition, the single promotion activities of traditional enterprises lacking interaction are not effective, so that enterprises can no longer focus on traditional publicity methods as in the past. The future promotion strategies of enterprises should also be guided by market trends and pay attention to innovative research on marketing strategies.

If enterprises want to make their enterprises have greater commercial value now, they must actively respond to the era of big data, be good at using big data, conduct data analysis, and fully enter the development model of the dual combination of online marketing and physical marketing [1]. It is a major development opportunity to actively change the marketing concept and find an effective development model; if the enterprise cannot change the traditional marketing attitude and thinking, it is likely to lose its foothold under the impact of the big data wave place [8]. Big data requires enterprises to pay attention to online marketing, and the primary premise of online marketing is to grasp potential customers, have a precise positioning of customer groups, and analyze consumers' needs and orientations according to the browsing records of consumers. When companies analyze big data, it is difficult to find out the real consumer groups through data analysis, or the analysis of the real life or purchase needs of the consumer groups is not accurate, which eventually leads to a failure of online marketing. Due to the lack of professional talents, enterprises also have shortcomings in the division and setting of customer objects, so it is difficult to formulate reasonable marketing strategies and programs. Therefore, enterprises should keep up with the pace of the digital economy era. To ensure the quality of marketing, enterprises must start with big data, do a good job in the collection, cleaning, and processing of marketing data, and predict future development trends based on data analysis results. Give full play to the value of data and promote the development of enterprises [4].

Based on the above status quo, the author of this paper believes that an enterprise marketing data analysis system based on Hadoop big data technology and aided by RFM algorithm data mining technology can effectively solve the above problems. This system can effectively help corporate users to judge the current situation of the marketing market, perceive existing user classification groups, understand the information of potential customers, and formulate targeted marketing strategies based on such information, giving a boost to the development of small and medium-sized enterprises, thereby promoting. The domestic marketing market is booming.

2 Key technologies

2.1 Hadoop

Hadoop is a distributed system integration framework developed by the Apache Foundation, and now a complete Hadoop ecosystem has been formed. Hadoop is composed of modules such

as distributed file system HDFS, parallel computing MapReduce, and unified resource management Yarn. Hadoop is characterized by scalability. Hadoop provides stable and reliable expansion capabilities, and provides high storage and processing capabilities by expanding the number of cluster-related technical nodes. The second is low cost (Economical). Hadoop is an open source project that can be deployed to conventional servers, and can be distributed and processed through conventional servers to form server clusters without requiring servers with specific performance to deploy. In addition, it has the characteristics of high efficiency. With the help of the parallel computing model, Hadoop can process data in parallel on the current data storage node, which greatly speeds up the processing speed. Finally, Hadoop is reliable, uses the distributed file system HDFS, and maintains 3 copies of data by default. Today, Hadoop has formed a complete Hadoop ecosystem as shown in Figure 1, which mainly includes distributed file system HDFS, distributed computing framework MapReduce, distributed column storage data warehouse Hive, column-oriented distributed database Hbase, distributed Flume, a massive log collection system, Zookeeper, which provides distributed application coordination services, Sqoop, which is used for data transfer between Hadoop and traditional databases, and Oozie, a Hadoop workflow scheduling engine, etc [2].

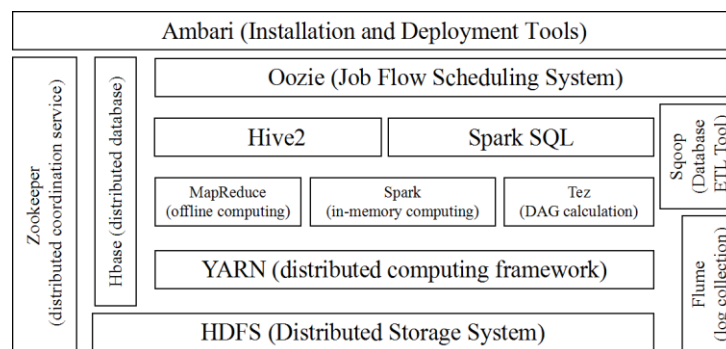


Figure 1: Hadoop Architecture Diagram

MapReduce: MapReduce in Hadoop is a simple software framework that allows applications to run on large clusters of thousands of machines and to process datasets larger than terabytes with reliable fault tolerance. The reason why MapReduce is so popular is that MapReduce is easy to program. By simply implementing some interfaces, it is possible to send running programs to a large number of cheap PC machines in a distributed manner. This means that writing a distributed program is the same as writing a simple serial program. It is because of this feature that MapReduce programming has become very popular. It is highly fault-tolerant, and if one of the machines in the cluster hangs, it can transfer the computing task it is running to another node and run the task. The most important concepts of the MapReduce computing framework are Map and Reduce. Hadoop decomposes a large task into many subtasks, and multiple subtasks can be executed simultaneously to achieve parallel computing. Fundamentally, MapReduce borrows the characteristics of functional programming. Its implementation is a Map operator function that maps key-value pairs to new key-values. It is then passed to the Reduce operator function and the value is merged with the same intermediate form key, as shown in Figure 2 [3].

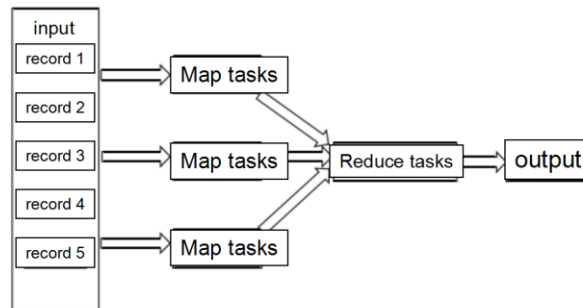


Figure 2: MapReduce Workflow

Flume: Flume is one of Apache's top projects - Flume provides a highly available, highly reliable distributed system for large-scale log collection, transmission and aggregation. Flume is a tool or service that collects and aggregates data resources like logs or events. The design principle of Flume is to store it centrally by streaming data from various web servers to HDFS, HBASE, KAFKA, etc. Flume can efficiently collect, transmit and store log data information from a large number of web servers. Flume supports various types of source input data as well as output data, supports the data transmission of various policies and paths, and can also be extended horizontally.

2.2 RFM Model

RFM model is widely used in user value analysis, which can well measure customer value and customer spending power. R represents the time since the last customer purchase, F represents the total number of customer purchases in a specific time, and M represents the total amount of customer purchases in a specific time. These three indicators constitute the RFM model and evaluate the value of measuring customers. . The RFM model can dynamically display the complete outline of customer value and customer spending power, providing the basis for personalized service. By observing the dynamic fluctuation of customer value, the long-term value of customers can be predicted. Support more marketing decisions by improving the status of three metrics. The application of the RFM model aims to create a user behavior report that will become an important indicator of customer retention. Therefore, the RFM model is widely used at the marketing level to stimulate new users to continue to spend and retain. It can also be used as an important indicator for monitoring the health of business users [10].

2.3 Development Environment

The author of this paper briefly introduces the relevant technologies used in platform development. An enterprise marketing data analysis system based on big data builds a Hadoop cluster. Four Windows X86 servers are used. The CPU configuration of the four servers is the same as the hard disk configuration. The memory configuration of the four servers is 64GB and the other three are 32GB. 3 are Datanodes. Due to the large memory required by the Namenode, the Namenode node uses a server with a memory size of 64 GB, and the Datanode uses a machine with a memory size of 32 GB. ETL data processing is performed on the data through Hadoop components such as Sqoop, Flume, HDFS, and Mapreduce on the cluster. This paper studies the development of a big data enterprise marketing data analysis system on the Windows

10 system. It combines CSS+HTML+Javascript language to use Bootstrap development framework for front-end development, and introduces Echarts to visualize the data of this system for chart display.

The functional logic backend part of the system is developed based on the MTV mode on the Django 2.0.13 framework. The system is developed in Python 3.7.4. The system integrated development tool is Pycharm 2021.3, using MySQL 8.0.28 database as a relational database cache, and using Nginx 1.21.5 and uWSGI2.0.8 to complete the online system server deployment. The choice of system architecture and key implementation technologies is very important. According to the analysis of relevant technologies in the current era, it is believed that it is technically feasible to build an enterprise marketing data analysis system platform based on the above technologies.

3 Demand analysis

3.1 Functional Requirements

With the advent of the era of big data, big data has had a profound impact on the development of society. The big data-based enterprise marketing data analysis system developed in this paper is aimed at the user group for small and medium-sized entities to operate enterprises. This paper conducts functional demand analysis based on the needs of enterprise users. This system uses big data analysis to formulate corresponding marketing strategies to help companies perceive existing user classification groups, and then provide better services for existing customers. Understand the information of potential customers, and then gain a more precise insight into the needs of potential consumers. The system also needs to help enterprise users judge the current situation of the domestic marketing market, help enterprises to adjust targeted marketing strategies according to such information in a timely manner, and achieve a leap in marketing capabilities. In addition, the system needs to have the characteristics of ease of use, the system should have a unified operating standard, and each module is simple and standardized, so that the enterprise management personnel with relatively insufficient computer knowledge can also successfully realize the operation of the system. At the same time, it is also necessary to present the data through visual charts to make the data more intuitive and visual, so as to solve the difficulties of users in understanding and using the data.

3.2 Overall Design

The author of this paper uses a variety of technologies to serve the big data-based enterprise marketing data analysis system. Collect business data in real time from the original information system through Flume and obtain external data through Scrapy technology, and use Distributed File System (HDFS) technology for file storage, then use Mapreduce programming technology for distributed computing, and use Hbase technology for unstructured data. The data is stored, and the log information obtained by Kafka is transmitted to the Hive data warehouse to complete the data collection, conversion, cleaning and storage functions. Then use Python language technology to perform statistical analysis and in-depth mining on the data in the warehouse according to the preset standards and principles, and store the analyzed structural data in the relational database MySQL library of the enterprise marketing data analysis system for subsequent follow-up. Design provides rules. According to the design at the application level,

the Django framework is used to write back-end logic functions + the Bootstrap template is used to write web front-end pages using HTML+CSS+JavaScript, and Echarts is used for chart visualization. The system is built to the Nginx+uWSGI server to complete the overall system architecture. The above process is shown in Figure 3 [7].

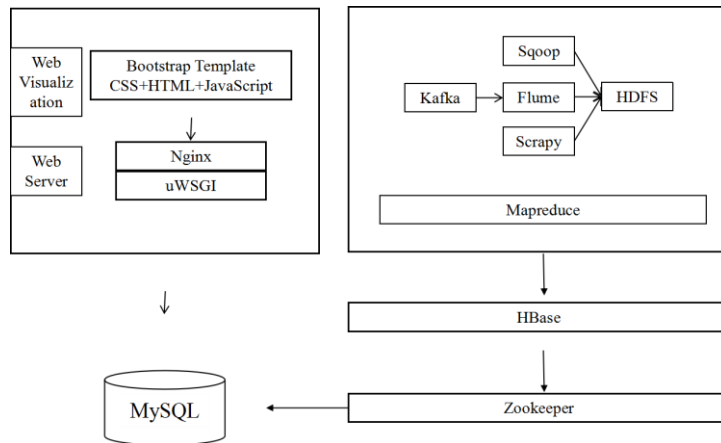


Figure 3: Overall Architecture of the System

4 Function realization

The enterprise marketing data analysis system based on big data technology is aimed at user groups for small and medium-sized entities to operate enterprises. After the user logs in to the system through the account and password, he can see the three main data analysis function modules developed according to the user's needs: existing customer data, potential user data and target market data.

4.1 Analysis of Existing Customer Data

Click to enter the existing customer data analysis module, users can see the group classification of existing customers of the enterprise, and the specific classification of customer groups according to the level of RFM value is shown in Figure 4. The customer group classification uses a pie chart to display the proportion of customers. Click to enter each classification to view the customer list and customer specific consumption information in each classification. In some marketing scenarios, it is sometimes inappropriate to treat different customers the same way, so we use the RFM model to analyze user behavior and consumption trends based on user data, and label the data results accordingly to apply to different consumers. The R value is divided into five grades: 1 month, 2 months, three months, half a year, one year, and more than one year according to the time (Recency) data of the last consumption, 4, 3, 2, 1, F value is the frequency of consumption (Frequency) data, according to the consumption frequency $[20, +\infty]$, $[10,20]$, $[4,9]$, $[2,3]$, $[1]$ five gears in descending order. Similarly, the M-value consumption amount (Monetary) data is divided into 5 grades in descending order, specifically classified as the top 40% of the consumption amount, the top 40%-60% of the consumption amount, and the top 60%-80% of the consumption amount., users with the top 80% - 90% of the consumption

amount and users with the top 90% - 100% of the consumption amount. Here, the average value of the RFM value is 2.3 by weighting the number of users. Users with an RFM value greater than 2.3 are high-value users in this dimension, and users with an RFM value less than 2.3 are low-value users in this dimension. Figure 5 shows the specific code for obtaining the number of days from the latest purchase, the number of historical cumulative purchases, and the historical cumulative consumption amount of daily registered users in Hive using HQL language. With limited company resources, we will give priority to serving the most important customers of the company, and the stickiness between customers and us will be higher, and a loyal cooperative relationship will be established with both parties [6].

R value	F value	M value	customer type
High	Low	High	Important to keep customers
High	Low	Low	generally keep customers
High	Low	High	Important to keep customers
Low	Low	High	Important customer retention
Low	Low	Low	General customer retention
High	Low	High	Important to keep customers
High	High	High	Important value customers
High	High	High	General Value Customers

Figure 4: Consumer RFM Classification Diagram

```

create table if not exists user_consume as
select t1.update_time,t1.user_id,last_pur_day,his_pur_count,his_sale_room
from
(select user_id,his_pur_count,his_sale_room,update_time
from dws_user_users_his_sum
where update_time=concat('${YEAR}',-,'${MONTH}',-,'${DAY}')) t1
join
(select user_id,last_pur_day,update_time
from dws_user_users_last_time_sum
where update_time=concat('${YEAR}',-,'${MONTH}',-,'${DAY}')) t2
on (t1.user_id=t2.user_id);

```

Figure 5: RFM Value Hql Acquisition Code

4.2 Potential Customer Data Analysis

Click to enter the potential customer analysis module, and you can see the portrait of the potential customer character model created by the system based on the data information obtained from major mainstream media. This system obtains the business data required by enterprises by purchasing API interface services of mainstream social media platforms such as Weibo and Douyin. Obtain a list of users who search for keywords of a company's products on mainstream media platforms, and obtain the age group, gender breakdown data, and regional breakdown

data of relevant users' fans. These data are aggregated, and potential consumer customers are profiled to help companies target accurate consumer groups to focus on advertising on the most suitable social media for high-quality and efficient marketing [5].

4.3 Target Market Data Analysis

Users can click to enter the target market data analysis to see the three sub-modules of e-commerce sales market data, domestic industry development trends and industry policy dynamics. In the e-commerce sales market data sub-module, you can see the sales data of enterprise-related products in the e-commerce market obtained from large e-commerce platforms such as JD.com and Taobao through web crawler technology. This system will collect and organize these data to help enterprises and businesses conduct marketing analysis. These data include the distribution of product sales in different price ranges, which are displayed using bar charts to help physical companies identify the price positioning range of products sold online. The domestic industry development trend sub-module can also obtain the monthly, quarterly and yearly total sales data of various commodities in the national public data of the National Bureau of Statistics through web crawler technology. An intuitive line chart can be generated to help companies judge the market development trend. The industry policy dynamic module timely updates the latest policy trends released by the Ministry of Commerce and local government departments on the relevant industries of enterprises, helping enterprises to adjust and optimize business policies in a timely manner according to the changing trends of the national market [9].

5 System test

Now part of the test cases are described. Item number 001 is the MySQL synchronization to Hive data, the test process is the business background MySQL data discovery and modification, and the test results add new records corresponding to the data of Hive data warehouse. Item number 002 is the data cleaning, the test process is the business background MySQL increase two identical data, only one record after Hive data warehouse.

The enterprise marketing data analysis system is tested in the given test case, and the test results are summarized:

- (1) Functional test results are: The data collection, cleaning and data analysis functions of the retail enterprise marketing big data analysis system are normal, achieving the expected effect.
- (2) The scalability test results are: only add Hadoop cluster nodes, and the processing speed is accelerated. Scalability achieves the desired effect.
- (3) The performance test results are as follows: the human intervention of the system operation process does not affect the normal operation of the system, and the performance achieves the expected effect.

6 Conclusions

This paper first introduces the background and significance of designing a data marketing system based on big data. In the context of massive data, it is particularly important to know what kind of people the users behind each terminal are and what he is interested in. It is difficult for the face of traditional technical architecture is difficult to meet the needs. Therefore, this paper proposes the project of Internet data marketing system based on Hadoop to more efficiently process and analyze massive data. Although the system has been deployed and put online, there are still many details to be continuously improved in the future, which need to continue the in-depth research and continuous optimization in the future work.

References

- [1] Gao Jie (2018). Analysis of the predicament of the marketing mode of small and medium-sized enterprises in the era of big data. *Modern marketing*.11.
- [2] Jiang Shangbin (2019). Design and implementation of a retail enterprise marketing big data analysis system. *Zhengzhou University*.04.
- [3] Liu Bing (2021). Analysis of Enterprise Big Data Marketing Strategy under the Development of the Internet Era. *China Storage and Transport*.07.
- [4] Liu Fang (2021). Analysis of Enterprise Marketing Informatization Management in the Background of Big Data Era. *China Management Informatization*.01.
- [5] Li Zhang (2017). Design and Implementation of Internet Data Marketing System Based on Hadoop. *University of Chinese Academy of Sciences*.04.
- [6] Ren Xitao, Wang Jian (2020). Research on Enterprise Marketing Innovation Based on Big Data Collection and Analysis Technology. *China Market*.09.
- [7] Wu Wenyu (2019). Integration and innovation path analysis of enterprise marketing data management in the era of big data. *Modern Information Technology*.02.
- [8] Yang Bowen (2021). Analysis of the essence of enterprise marketing and marketing strategy under the background of big data. *Marketing Planning and Management Edition*.12.
- [9] Yang Xiaoying (2021). Analysis of Enterprise Marketing Strategy Innovation Based on Big Data Era. *National Circulation Economy*.03.
- [10] Zhang Haixiang (2021). Design and implementation of user tag system based on Hadoop technology. *Southeast University*.06.