

Statistics Analysis and Visualization for Big Data of E-commerce Platform Sales Evaluation

Wei Zhan^{*}, Jinhui She, Yangyang Zhang, Chenfan Sun

School of Computer Science, Yangtze University, Jingzhou, Hubei, China

**Corresponding Author.*

Abstract

With the rapid increase in the sales scale of e-commerce platforms is accompanied by the rapid growth of consumer evaluation data on commodities at the same time. How to use big data analysis and visualization technology to mine the valuable information in the massive consumers evaluation data is an urgent issue in promoting the development of e-commerce platforms. However, the amount of e-commerce evaluation data is huge, growing fast, and mostly unstructured data, which is typical big data. In order to efficiently realize the visualization of e-commerce evaluation big data, this paper proposes an end-to-end four-layer framework for data visualization system. The data acquisition layer uses the Webcollector crawler to crawl a total of 420,000 mobile sales evaluation data on the JD website and stores them in the MySQL database; The data import layer uses the Sqoop tool to import MySQL data into the Hadoop platform; The data processing layer uses HDFS and MapReduce to process and analyze big data; The visualization implementation layer uses Jsp+Servlet+JavaScript+echart integrated technology to visualize the big data of distribution of mobile phone sales, user purchase impressions, and user mobile phone portraits. Which helps consumers choose their favorite mobile phones conveniently, and provide decision-making support for e-commerce companies to more accurately launch products, benefiting both parties.

Keywords: E-commerce platform, sales review, big data analysis, visualization

I. Introduction

Big data 387 technology is developing rapidly and its data scale has reached the level of PB (1024TB). Facing such a huge amount of data, its value cannot be estimated, but we can not quickly process the set data in a specific time. How to extract the value of its storage is the first thing in front of people. In China, big data technology **Error! Reference source not found.** research is distributed in cloud computing, data mining, distributed computing, parallel computing and other fields to achieve the purpose of rapid acquisition, analysis, processing and extracting value from them, focusing on the application of data in various regions. In order to improve the stable operation of power system, Qu Haishan [0] proposed to solve the problem of inaccurate data due to the traditional method of loading data by using data visualization technology in big data analysis. Xing Yue [0] put forward that enterprises should analyze and think about the relevant situation within enterprises by analyzing big data and strong analytical ability, and promptly optimize and adjust the marketing methods and business content of enterprises, so as to formulate marketing strategies that are more in line with market development and promote the sustainable development of enterprises. while in foreign countries, it focuses on the basic model design and algorithm analysis and other aspects. After data processing, data sharing, data retrieval, data analysis, data visualization and other operations can be carried out to mine its internal relations and play different values. Gagan, et al [0] Through the analysis of big data such as consumption patterns, emotions and feedback of bank customers, banks can better understand customers, so as to develop personalized solutions for customers and guide customer loyalty. Researchers at the University of Massachusetts at Lowell have identified metabolic inhibitors and promoters by using bigdata analytics technology to study antibodies, improve the productivity of antibody and optimize the production process of (mAb) [0]. Vijayakumar, Tamil selvan, et al **Error! Reference source not found.** Based on Wald adaptive prefetching boosting classification, czezanowski similarity mapping reduction (wapbc-csmr)

technology solves the problem of accuracy and time complexity of query processing. Different metrics are used to analyze wapbc-csmr data on large data sets, which achieves the goal of improving the accuracy of query processing. It reduces the query time and error rate. Figure 1 is a big data ecosystem diagram:

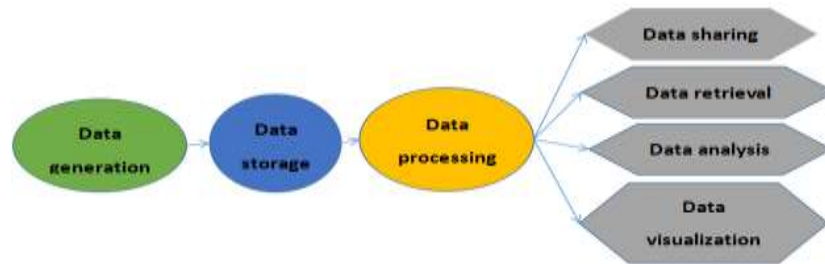


Fig 1: Big data processing

With the rapid development of Internet technology in China, e-commerce companies such as Jingdong and Taobao are online rapidly. the phenomenon of online shopping has been everywhere, the development of e-commerce industry has begun to have a certain impact on the operation of physical stores. Most users will post some comments after purchasing the goods on the Internet. These comments can help the merchants to accurately locate the consumers' situation, and also help other consumers can more easily choose their favorite mobile phones. Mining users' evaluation of goods is of great significance, but for a huge number of unstructured users' evaluation of goods, traditional data analysis and data mining methods have been difficult to meet the explosive growth of sales evaluation data analysis and mining potential value and data visualization function existing in the state of big data. Moreover, ordinary technology cannot analyze and mine it quickly and accurately, and cannot perform computational processing on a single computer. Therefore, how to efficiently analyze commodity big data, mine the intrinsic value of data, and visualize data becomes a key issue.

Aiming at the current situation of big data of commodity sales evaluation, based on Hadoop **Error! Reference source not found.** distributed file system, NoSQL database platform based on HBase massive data, MapReduce distributed data mining analysis and calculation framework, this paper excavates, analyzes and processes the user commodity evaluation data crawling from Jingdong shopping platform, and uses visualization technology to mine, analyze and process the user commodity evaluation data The graph representation method of mobile phone users' portraits intuitively displays 400000 mobile phone users' product evaluation data in Jingdong mobile phone mall, and helps customers quickly and comprehensively understand their favorite products by referring to the displayed user product evaluation information when purchasing products. Obtain the intrinsic value of the data and provide decision support for e-commerce to launch products more accurately. Greatly promote the development of e-commerce.

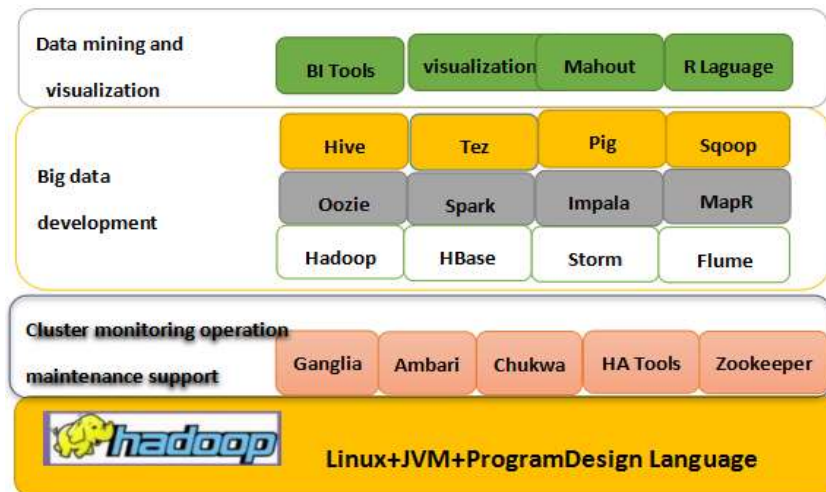


Fig 2: Big data ecosystem

Figure 2 is the big data ecosystem. The bottom layer is the big data processing platform built by Hadoop platform and Linux +JVM + programming language. HA Tools and Zookeeper etc. realize the monitoring and operation and maintenance of the cluster. The second top layer is the big data development layer with many tools such as Hadoop, HBase, etc. through these tools, data mining, analysis, storage and other operations make the structure complex , the massive data that is difficult to extract becomes simple and can be used by people. The top layer is to realize data mining and visualization through Bi tools, data visualization tools, mahout and R language.

The key technologies used in this paper, such as webcollector crawler architecture, data ETL Technology, data processing mode based on Hadoop distributed file processing platform, MapReduce 0,0,Error! Reference source not found.distributed data processing platform, data visualization, etc, domestic and foreign scholars have done a lot of research work and achieved a lot of results. Hadoop Distributed File System (HDFS) 0 is a distributed file system with high reliability and high scalability, which can provide massive file storage capacity. The HDFS working mode in big data is streaming data access. Based on common hardware, it mainly supports one write and multiple reads, and is a file system designed to store large files 0 .The MapReduce distributed computing framework is a large-scale massive data platform, which adopts a non shared large-scale cluster system with good cost performance and scalability. MapReduce has the characteristics of simple, easy to understand and easy to use models. In the process of processing large-scale data, MapReduce hides a lot of tedious details and simplifies the development work of programmers. MapReduce can implement a large number of data processing problems such as machine learning and data mining algorithms 0.

With the rapid development of big data, people generate a large number of data every day, among which e-commerce data accounts for a significant proportion. Therefore, in order to obtain these data, the crawling technology is catching people's eyes. Web crawler 0-0 is an application that can automatically extract web page information from search engines. The webcollector crawler used in this paper also belongs to the category of network crawler. It can implement different functions of webcollector crawler by writing plug-in modules based on the webcollector kernel. Mining and analyzing the buyer's evaluation data to explore the consumer's love for the goods. The traditional mining algorithm has a good effect on the small data set, but for the user evaluation of the big data set, using Hadoop distributed processing tools can more efficiently process the data 0. The development of scientific visualization 0 mainly focuses on the two research directions: information visualization and data visualization. Through visualization technology, a large amount of non-intuitive and abstract data and information can be displayed in the form of graphic and image information. Data visualization technology 0 needs to solve the problems of determining the data space, developing data, analyzing the data and visualizing analysis results.In this paper, we use the visualization technology to study the evaluation of mobile phone sales in achieve a model that benefits both sellers

and consumers.

II. The Design of System Architecture

The system framework of this project is divided into four layers and the contents of each layer are as follows:

The data source layer is the comment information of the mobile phone purchase of Jingdong Mall, and the webcollector crawler project is imported into eclipse to crawl the mobile phone evaluation information of Jingdong client. The ETL layer works together with MySQL and sqoop to import the crawled data into the HDFS distributed file system. The data is stored in the processing layer. Based on the Hadoop platform, MapReduce is used to calculate the data, which is initially stored in Hbase, and phoenix is used to map and store the data. The web presentation layer uses the technology of JSP + servlet + Javascript + echart to visualize the stored data. The system architecture diagram is shown in Figure 3:

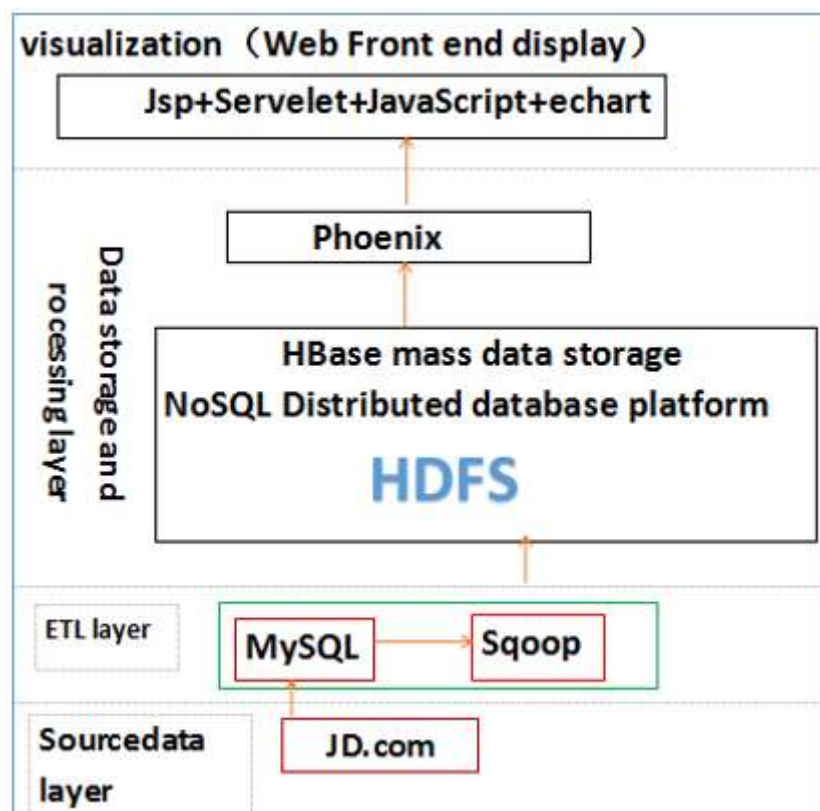


Fig 3: The system architecture

Figure 4 shows the running process of this task:

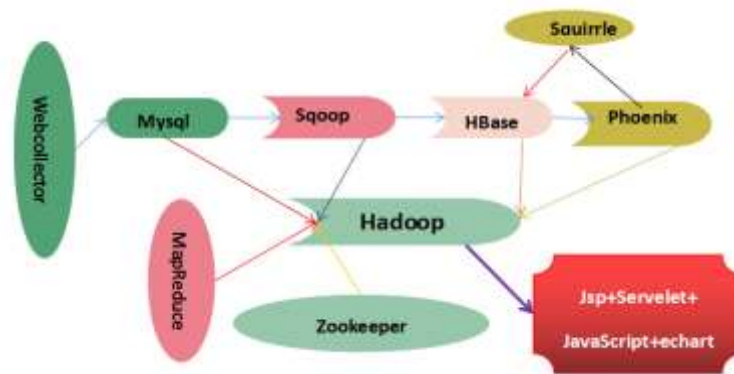


Fig 4: The process of task running

Webcollector: Distributed crawler architecture for crawling 400,000 mobile phone sales data of Jingdong Mall.

Hadoop: This article implements Distributed File System (HDFS) and MapReduce Distributed Computing Framework based on Hadoop big data platform.

MySQL: Import the sales evaluation big data crawled in the Webcollector in the linux environment;

Sqoop: Extracts, cleans, transforms, loads, and imports data from MySQL through MapReduce tools and imports them into HBase.

Hbase: Real-time distributed database for accessing sales evaluation data after ETL.

MapReduce: Distributed computing framework for sales evaluation data under Hadoop platform, Sqoop for data ETL operations.

Phoenix: Phoenix provides a standard Sql way to manipulate Hbase data. Perform association mapping on the data in the HBase table, query, insert, modify, and so on.

Zookeeper: A distributed collaboration service that monitors the movement of Master and Slave nodes in the system.

Visualization Tools: For the sales evaluation big data under the Hadoop platform, Jsp+Servlet+JavaScript+echart integrated technology is used to achieve visualization. Figure 4 shows the running process of this task.

III. Data Sources

Open the official website of Jingdong mobile phone, input mobile key words, you can browse a large number of mobile phone information, click commodity evaluation to see detailed evaluation information, as shown in Figure 5:



Fig 5: The commodity evaluation chart

WebCollector is a JAVA crawler framework (kernel), which can analyze the web pages and write the required plug-ins to realize the powerful function of directional data collection. The kernel diagram is shown in Figure 6:

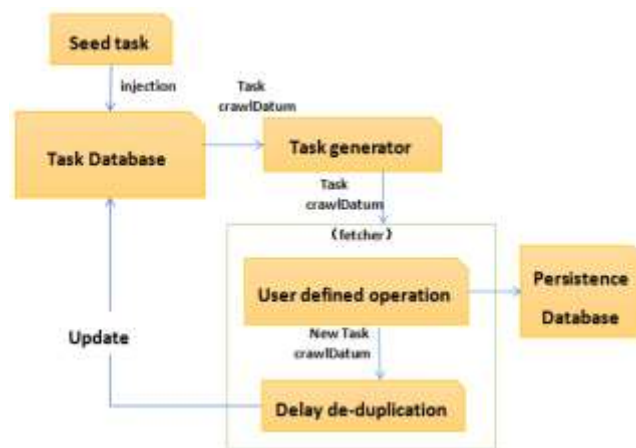


Fig 6: The kernel schematic of web collector

Install eclipse in the window system and import the WebCollector crawler project; install MySQL and create jd_database database in MySQL, then create spider1 table and design table field: id indicates that the product evaluation number is set as the main key, platform represents the product platform crawled by the crawler program, xinhao Indicates the phone model, title of the title evaluation, content indicates the user's evaluation of the mobile phone, memberlevel: user membership level, fromplatform: which platform the user purchased from, area: user area, userimpression: user's impression of the purchased mobile phone, color: user The selected phone color, price: phone price, productSize: phone size, creationTime: the time the user commented on the phone, zhuaqutime: the time to use the Webcollector crawler to crawl information, label: tag. As shown in Figure 7:

```
Database changed
mysql> CREATE TABLE `spider1` (
-> `id` int(11) NOT NULL AUTO_INCREMENT,
-> `platform` varchar(255) DEFAULT NULL,
-> `xinhao` varchar(255) DEFAULT NULL,
-> `title` varchar(255) DEFAULT NULL,
-> `content` text DEFAULT NULL,
-> `memberlevel` varchar(255) DEFAULT NULL,
-> `fromplatform` varchar(255) DEFAULT NULL,
-> `area` varchar(255) DEFAULT NULL,
-> `userimpression` varchar(255) DEFAULT NULL,
-> `color` varchar(255) DEFAULT NULL,
-> `price` varchar(255) DEFAULT NULL,
-> `productSize` varchar(255) DEFAULT NULL,
-> `creationTime` varchar(255) DEFAULT NULL,
-> `zhuaqutime` varchar(255) DEFAULT NULL,
-> `lable` varchar(255) DEFAULT NULL,
-> PRIMARY KEY (`id`)
-> ) ENGINE=MyISAM AUTO_INCREMENT=19712 DEFAULT CHARSET=utf8;
Query OK, 0 rows affected (0.07 sec)
```

Fig 7: The table of fields design

Configure the database user name and password. To start the webcollector crawler and access the mobile phone related information on the JD e-commerce platform, you need to enter the ID and model of a mobile phone. Webcollector starts the crawlDb.DBManager process for content crawling, the maximum depth of crawling is 2, and the average time spent crawling each message is 1 seconds. The crawled data is directly imported into the spider table in jd.db in MySQL. See the running effect in the server as shown in Figure 8:

```
-07-18 15:49:33 INFO cn. dmic.webcollector.fetcher.Fetcher - close segmentwriter:cn.
07-18 15:49:33 INFO cn. dmic.webcollector.crawlDb.DBManager - start merge
07-18 15:49:33 INFO cn. dmic.webcollector.crawlDb.DBManager - merge fetch database
-07-18 15:49:33 INFO cn. dmic.webcollector.crawlDb.DBManager - merge link database
-07-18 15:49:33 INFO cn. dmic.webcollector.crawlDb.DBManager - end merge
-07-18 15:49:33 INFO cn. dmic.webcollector.crawler.Crawler - depth 2 finish:
total urls: 0
total time: 1 seconds
07-18 15:49:33 INFO cn. dmic.webcollector.crawler.Crawler - start depth 1
```

Fig 8: The web collector of crawl information

IV. Data ETL

The ETL 0 data is extracted, cleaned, transformed, and loaded and finally data warehouse is built. The main process is to extract the required data from the data source, load the data into the data warehouse according to the data warehouse model defined in advance after the data cleaning process, to realize the data conversion function.

4.1 Building a Massive Data Store NoSQL Database

Build the Hadoop platform, configure the HDFS function required for the cluster to transfer files, configure the zookeeper monitoring cluster, and build a massive HBase data storage platform. When the Hadoop master terminal executes the JPS command, in the printing result, it can be seen that the four main node processes of ResourceManager, JPS, namenode and secondarynamenode, as well as the Hadoop slave terminal input the JPS command, and there will be three processes of nodemanager, datanode and JPS, In addition, after HBase runs, the main node adds the hmaster and hquorumpeer processes, indicating that the cluster has been completely set up, as shown in Figure 9,10:

1616 nameNode
2544 HMaster
1987 ResourceManager
2805 Jps
2475 HQuorumPeer
1822 SecondaryNameNode

Fig 9: The hadoopmaster processing

1729 HRegionServer
1411 DataNode
1539 NodeManager
1994 Jps

Fig 10: The hadoopslave processing

4.2 Data import and export

Sqoop 0 has a data exchange system to help data exchange between relational databases and Hadoop. The imported data source is a relational database. Before importing, jdbc will be used to connect to the database. The target is the big data system where HDFS and HBase are located. By extracting the records of the tables imported in MapReduce, they will be written to HDFS 0. Based on the Hadoop big data environment, this article uses the sqoop tool to import the crawled comment data stored in MySQL into the HDFS / HBase cluster through MapReduce parallel computing. The working principle is shown in Figure 11.



Fig 11: The working principle of sqoop

In the case where the Hadoop platform is built, all processes are required to be in a startup state. Enter the bin directory of HBase to start the HBase shell operation editing process, and create the PINGJIA.SPIDER table, and the column family is f1, which is used as the table for receiving data. Enter the Sqoop directory, in the bin directory, enter the connection code to import and export data. First, Sqoop connects to MySQL through the primary node address. Sqoop's import tool will run a MapReduce job to read the data in the table, connect to the PINGJIA.SPIDER table with the column family f1 and the primary key is id in Hbase and import the data. As shown in Figure 12:

```
[ZYY@master sqoop - 1.4.5.bin_hadoop-2.0.4-alpha]$ bin/sqoop
import --connect jdbc:mysql://172.18.0.2:3306/jd_db --username
hadoop --password hadoop --table spider --hbase-table
PINGJIA.SPIDER --column-family f1 --hbase-row-key id
--hbase-create-table -m 1
```

Fig 12: The process of MySQL connects HBase via Sqoop

The following figure demonstrates how Sqoop interacts with mobile sales evaluation data from MySQL into HBase.

Data import process: Sqoop is written in JAVA and provides JDBC access to data stored in MySQL. Before importing data, Sqoop uses JDBC to retrieve all columns and their SQL data types. Sqoop selects the primary key column of the table as the metadata and determines the query to be executed for each map according to the maximum and minimum values in the primary key. The MapReduce job started by Sqoop acts on the InputFormat class, reads some contents of the table from the MySQL database through JDBC, and start three Map tasks in Hadoop on the Master and Slave nodes to divide the query results. After the query, Sqoop sends the job to MapReduce cluster, deserializes the data in the ResultSet, generates an instance of the class, and writes it to HDFS.

Data export: Similar to the import process, Sqoop selects JDBC method before data is exported. A Java class is defined according to the instance of the class, and the value of the same type of record insertion is parsed from the text. Start the MapReduce job, read the metadata file from HDFS, and start the three Map tasks in Hadoop on the Master and Slave nodes to parse query results and execute export method. Import into a table created by HBase. As shown in Figure 13:

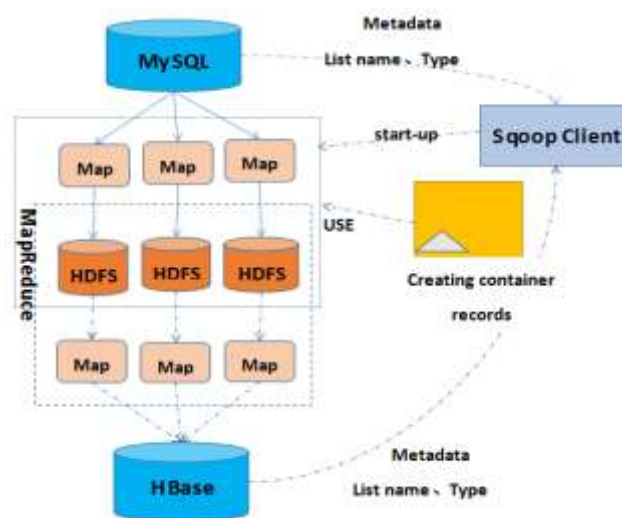


Fig 13: The import and export Process via Sqoop Data

V. Data storage and processing

5.1 HDFS provides Hbase with high reliable underlying storage support

This paper is based on visualization of sales analysis statistics based on the Hadoop + HBASE architecture. At present, Hadoop+HDFS 0 solves unstructured data storage in the Hadoop ecosystem 0, and HBase 0 solves semi-structured data or structured data storage.

HDFS can expand the storage capacity of the system horizontally by adding machines; the system is highly available; cost-effective; suitable for data distributed computing; suitable for storing unstructured data. and so on.

The HBase database is based on HDFS, and it can be used for data storage management by column mapping, which can realize the function of fast query in large tables. HDFS is a distributed file processing system in Hadoop and allows users to store data in the form of files, which provides great convenience for crawling data to be stored in hbase. HBase reads files from data memory. The data generator can write files to HDFS and Hbase, and HBase and HDFS can read and write data. The HDFS and hbase read and write file modes are shown in Figure 14.

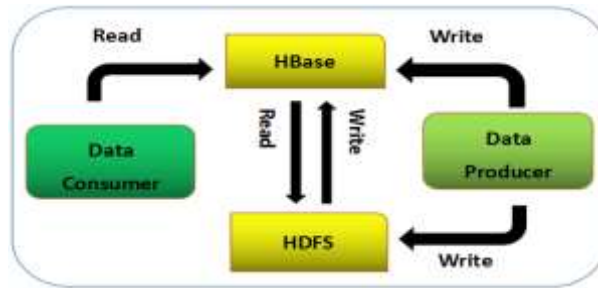


Fig 14: The read-write File Mode about HDFS and HBase

5.2 MapReduce provides high performance computing power

MapReduce 0-0 is a parallel programming mode for processing big data, which is mainly used for parallel calculation of big data. MapReduce is composed of several maps and reduce, and through the high concurrent processing function of MapReduce computing model, the speed of data analysis, de duplication and calculation is greatly improved. The most important two phases of the MapReduce calculation model are Map and Reduce. In the Map phase, calculation model will call user-defined map function to process several key value pairs of user data mapping. In the Reduce phase, it will traverse all the sorted wells. By calling user-defined reduce function, data with the same key value will be merged for production Key value pair 0. Its working principle is shown in Figure 15:

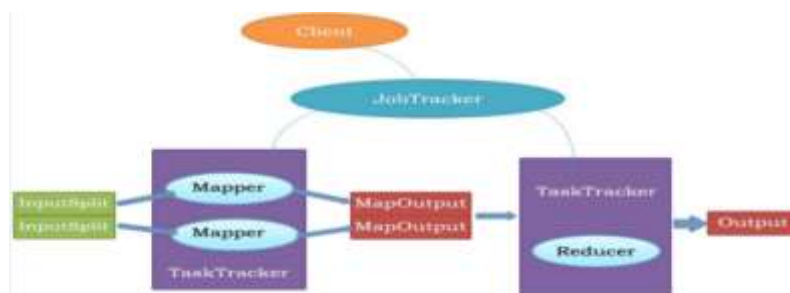


Fig 15: The working flow of mapReduce

MapReduce performs parallel calculation on the crawled sales data and counts the calculated data. Using HBASE to calculate the calculated data, a total of 399,968 product evaluation data is crawled, and the CPU time is 59,730 ms. MapReduce calculates the total. The duration is 235.3618 seconds, as shown in Figure 16.

```

Total megabyte-milliseconds taken by all map tasks=188900352
Map-Reduce Framework
  Map input records=399968
  Map output records=399968
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=4303
  CPU time spent (ms)=59730
  Physical memory (bytes) snapshot=216494080
  Virtual memory (bytes) snapshot=194310536
  Total committed heap usage (bytes)=143654912
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=0
09/26 16:47:02 INFO mapreduce.InputJobBase: Transferred 0 bytes in 235.3618
seconds (0 bytes/sec)
09/26 16:47:02 INFO mapreduce.InputJobBase: Retrieved 399968 records.

```

Fig 16: The Statistics pfocess about data evaluation

5.3 Mapping tables in HBASE using the Phoenix middle key

When solving the sales evaluation data storage and query of the e-commerce platform, the existing HBase is a database, but its query statement is not very easy to use. Although it can access unstructured data, but it needs to frequently use of the MapReduce computing framework when reading and writing again, which consumes a lot of hard disk resources and seriously affects the computing efficiency. This article uses ZooKeeper for cluster monitoring, HBase + Phoenix 0.0 storage warehouse to operate unstructured sales evaluation data through MapReduce computing framework. The HBASE+phoenix storage warehouse architecture is shown in Figure 17:

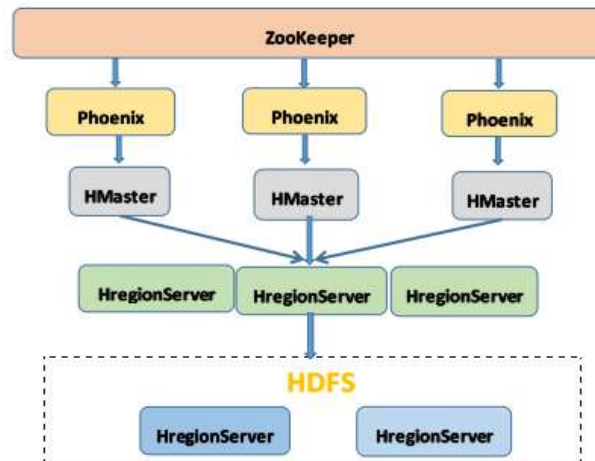


Fig 17: The framework of HBase and phoenix

Create a corresponding 'PINGJIA.SPIDER' table in hbase in Phoenix, where f1 is the column family of the HBase table, as shown in Figure 18:

```

create view PINGJIA.SPIDER(
  id VARCHAR primary key,
  f1 platform VARCHAR,
  f1 xinhao VARCHAR,
  f1 title VARCHAR,
  f1 content VARCHAR,
  f1 memberlevel VARCHAR,
  f1 fromplatform VARCHAR,
  f1 area VARCHAR,
  f1 userimpression VARCHAR,
  f1 color VARCHAR, x
  f1 price VARCHAR,
  f1 productSize VARCHAR,
  f1 creationTime VARCHAR,
  f1 zhuaqutime VARCHAR,
  f1 label VARCHAR
) as select * from PINGJIA.SPIDER;

select * from PINGJIA.SPIDER limit 10;

```

Fig 18: The table fields about phoenix mapping Hbase

Phoenix 0 is characterized by high integration and completeness, fault tolerance, high scalability, support for heterogeneous systems, intelligence and ease to use. Based on HBase to build a standard SQL operation, phoenix has a low-latency feature, which will compile the SQL into a series of Hbase Scan operations, and then generate the standard JDBC result set online transaction processing. Phoenix supports table creation and versioning incremental changes through DDL commands, and the corresponding metadata is stored in the Hbase table. Metadata is used to describe the relationship mapping between the phoenix tables and Hbase. Figure 19 shows the Phoenix vs. hbase map.

```
0: jdbc:phoenix:master:2181> select * from PINGJIA.SPIDERT limit 10;
```

ID	platform	xinhao	title
1	jingdong	Huawei honor 4A	[Deluxe package version] 4A (SCL - AL00) 2
GB + 8GB memory version	White Mobile		
10	jingdong	Huawei honor 4A	[Deluxe package version] 4A (SCL - AL00) 2
GB + 8GB memory version	White Mobile		
100	jingdong	Huawei honor 4A	[Deluxe package version] 4A (SCL - AL00) 2
GB + 8GB memory version	White Mobile		
1000	jingdong	Huawei honor 4A	[Deluxe package version] 4A (SCL - AL00) 2
GB + 8GB memory version	White Mobile		
10000	jingdong	Huawei P8 standard	Huawei P8 standard Cloud dust Mobile Unicom 4G phone
Dual SIM			
100000	jingdong	Huawei maimang 4	Huawei maimang 4 Morning gold Mobile Unicom Telecom 4G phone
Dual SIM			
10001	jingdong	Huawei maimang 4	Huawei maimang 4 Morning gold Mobile Unicom Telecom 4G phone

Fig 19: The phoenix mapping to HBase

5.4 Squirrel for Data Manipulation

Squirrel SQL 0,0 is a data tool that can connect multiple data. This article is used to connect to Phoenix database. In a way, it is a better database client tool. Under the Lib of squirrel installation directory, add the jar package required by phoenix, then enter the operation interface to add the phoenix driver, add host information configuration to complete the connection. As shown in Figure 20:

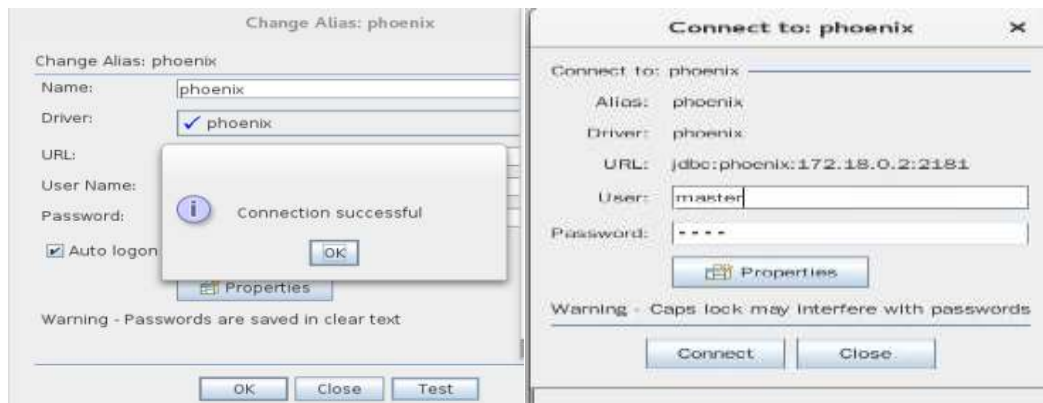


Fig 20: The connection between Squirrel and Phoenix

After accessing the Phoenix interface with JDBC unified database, a unified user interface is used to operate the database to realize data splicing, and SQL language is written to realize data stored in HBase table to be displayed in squirrel interface. The effect is shown in Figure 21.

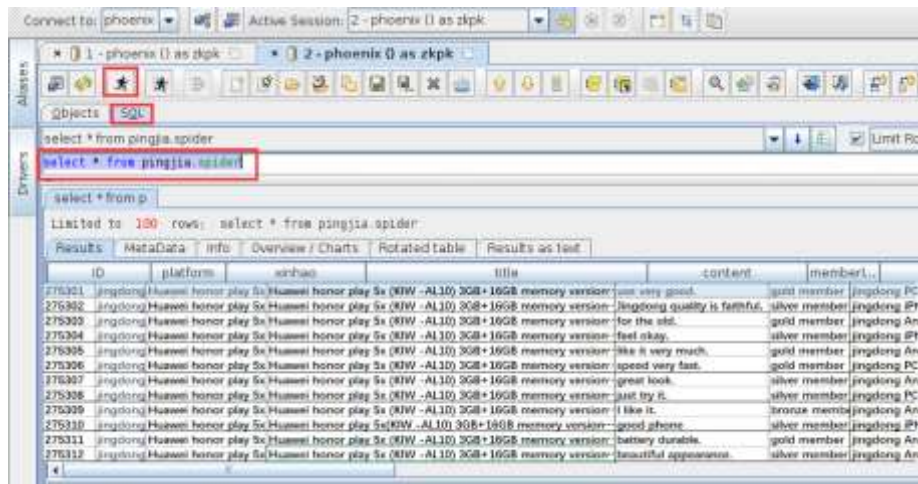


Fig 21: The Squirrel connection process of HBase

VI. The Visual Implementation

Based on the Hadoop file distributed system, this paper uses Jsp+Servlet+JavaScript+echart 0]technology to visualize the sales data after phoenix calculation and processing, tThrough the mapping of the data value, spatial coordinates, the relationship between different location data and other visual channel elements such as mark, location, shape, etc. Through the analysis of visual graphics such as regional map, histogram, line chart, user's mobile phone portrait, etc., it can clearly and effectively convey and communicate information and carry out interactive analysis of data at the same time,so as to help merchants adjust products according to sales conditions , achieve better sales of goods,and help customers to more intuitively see the characteristics of each product so as to choose their favorite products.

As shown in the following three meter diagrams:

we crawled through the evaluation data of mobile phone sales of one e-commerce platform, the evaluation data of 14 mobile phones in Jingdong Mall, and the total evaluation data of mobile phones was 400000. As shown in Figure 22:

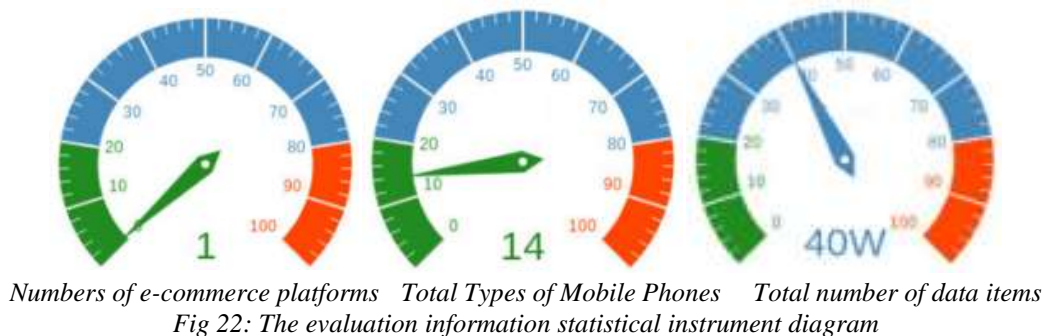


Fig 22: The evaluation information statistical instrument diagram

User purchase impression:

From the column and line chart, the horizontal axis represents the type of evaluation, and the vertical axis represents the number of people who published such evaluation. We can see users' different impressions of the purchased products. This kind of graphics can help the merchants improve the quality of the products, and also help the users to better choose the products that are suitable for them. Figure 23, 24 are the users' products of the purchase of the

goods:

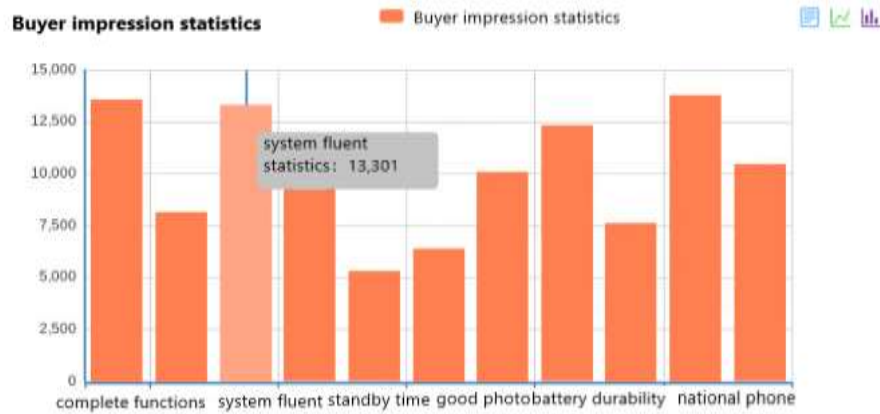


Figure 23 The Histogram of User Purchase Impression

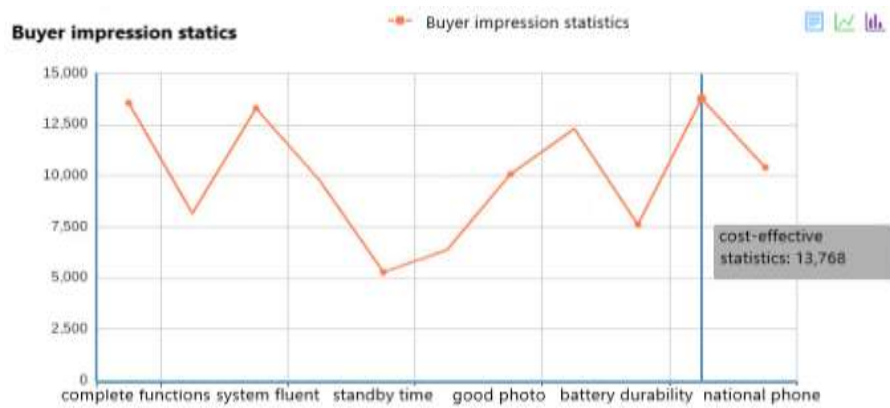


Fig 24: The impression line diagram of customer purchasing

Buyer membership level:

From the broken line chart of buyer member level, we can intuitively see the sales tendency of products to members of different levels, and also understand the mobile phone purchases of different levels of buyer members in Jingdong Mall. Figure 25 shows the rating of the buyer member:



Fig 25: The level diagram of user membership

User's mobile phone portrait:

According to the user's evaluation of the product, the user's mobile phone portrait is taken. From the figure, we can clearly see the user's intuitive evaluation of the mobile phone, which can help the business to provide sales quality and help other users to better select the product that suits them.

User's color preference:

Intuitively reflects the user's love for the color of a certain brand of mobile phones. Help the merchant to locate user's purchase needs more accurately. The user's mobile phone color preference is shown in Figure 26, 27:

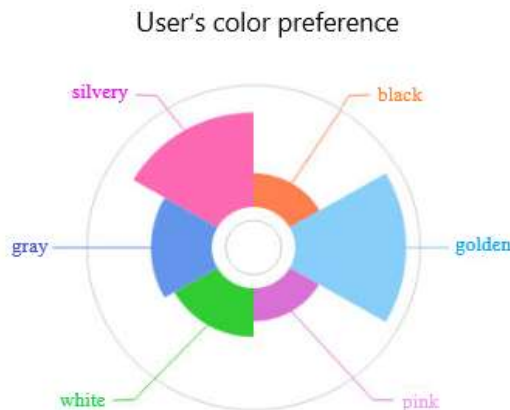


Fig 26: The char of user color favorites

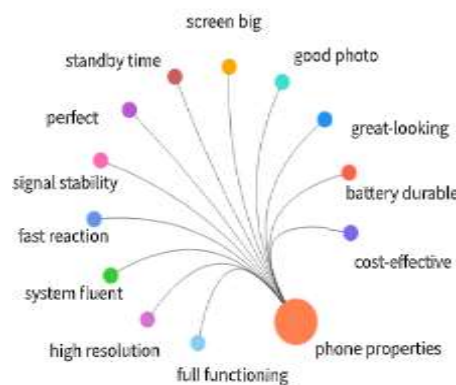


Fig 27: The portrait of user cell phone

VII. Conclusion

Throughout the design process, this paper describes the research background of the visualization of e-commerce platform sales data analysis and statistics, and briefly introduces the platform and big data technology used to realize the analysis of e-commerce platform sales evaluation data, as well as the visualization; Secondly, it analyzes and introduces the whole process of building the system: first, it introduces the system architecture composed of data source layer, data ETL layer, data processing and calculation layer, data visualization implementation layer; and shows the system in the form of figure and text description from using webcollector crawler to crawl the big data of mobile phone sales evaluation in Jingdong Mall; the crawled data is imported into MySQL database, then the data is imported to the NoSQL database platform based on HBase massive data through sqoop; finally, under the big data environment, the data is calculated and preprocessed through HDFS and MapReduce, and the processed data is finally realized through the data visualization tool. The following sections

introduce the core technologies such as Sqoop, Hadoop, HBase, and Phoenix, which are used in the system, including the basic concepts and operating principles of these technologies and their functions in this project. At last, the information content of user's commodity evaluation can be expressed intuitively by means of instrument chart, map, line chart and mobile user's portrait.

The visualization of JD mobile phone sales evaluation data realized in this paper, on the one hand, extracts and cleans the large amount of disorderly and unorganized data to make it organized, and finally stores it in the database, and then visualizes the data to help consumers see JD mobile phone more intuitively, and help consumers to quickly and accurately select their favorite products. On the other hand, the visualization of consumer data helps merchants to intuitively understand the needs of consumers and more accurately deliver their products and this technology can also be applied to the visualization of College Students' information.

Although this paper designs a complete system to realize the visualization of big data analysis and statistics of e-commerce platform sales evaluation, there are still some problems in the overall design of the system.

Data Crawling Technology: This article uses the Webcollector crawler technology to crawl the JD mobile phone sales evaluation big data under the Windows system, but crawler crawling speed is not very satisfactory. Using Hadoop-Webcollector distributed crawler technology to crawl sales evaluation big data under Linux system can speed up data crawling.

Data storage technology: Due to the large number of mobile phone evaluation data in Jingdong Mall, the load requirements on this system are relatively high, and MySQL technology may not be able to quickly store the crawled data under windows. Therefore, using the Hive database in Linux system is more conducive to improving the storage speed.

Computation Engine: MapReduce distributed computing framework is used in this paper. Using MapReduce computing framework requires HDFS reading and writing process. However, spark, a fast general-purpose computing engine specially designed for large-scale data processing, can avoid HDFS reading and writing process. Therefore, the use of Spark can be better applied to the data mining processing of big data of JD mobile sales evaluation.

Big data is a very popular word nowadays. When building a big data platform, you may encounter many problems during this period, including computer configuration problems, the proficiency of the Linux language, the number of machines in the design cluster, how to configure the role each node needs to play, the setup of the cluster, the installation and configuration of big data software, the startup sequence and maintenance of the cluster, file format and the realization of sales data visualization by using Java technology. Just because these complex and tedious steps and the unknown problems determine why big data technology has powerful functions and the difficulty of deep learning big data technology, big data technology is full of Psychedelic color.

Acknowledgement

This research was funded by the China University Industry-University-Research Innovation Fund "New Generation Information Technology Innovation Project" (2019ITA03004) and Jingzhou 2018 Science and Technology Development Program "Research and Development of New Methods of Unmanned Aerial Vehicle Visual Data Deep Learning Recognition Technology in Urban Management" (Project No. 2018024).

References

- [1] L. Xie, L. Tong. "Research on optical fiber sensing spectral multiplexing technology of big data analysis technology," *Laser Journal*, vol. 42, no. 1, pp. 149-153, 2021.
- [2] X.D. Wang, X.Y. MA. "Big Data Analysis with SVM Model Algorithm for Fault Diagnosis of Marine Equipment," *Journal of Shanghai Scientific Research Institute of Shipping*, vol. 44, no. 1, pp. 18, 2021.
- [3] H.S. Gagana, H.N. Roja, H.S. Gouthami. "A Survey on Bigdata Analytics using in Banking Sectors,"

- Journal of Trend in Scientific Research and Development, vol. 4, no. 4, 2020.
- [4] T. Selvan, S. (Erode Sengunthar Engineering College, Perundurai, Erode, India); Balamurugan, P.; Vijayakumar. Prefetched wald adaptive boost classification based Czekanowski similarity MapReduce for user query processing with bigdata, Distributed and Parallel Databases, Article in Press, 2021.
 - [5] W. Wang, X.J. Chen. "A review of university big data research," China Science and Technology Information, vol. 16, pp. 109-110, 2018.
 - [6] W.J. Tu, Y. Feng, B.H. Chao, X.M. Ji, L.D. Wang. "Stroke-unit care for stroke patients in China: the results from Bigdata Observatory platform for Stroke of China," Journal of Neurology, 2021.
 - [7] D. Hou, G. Li, W.Q. Zhao, L. Yang. "Application of Big Data Analysis and Visualization Technology in Power Grid Company," Smart Grid, vol. 3, no. 12, 2015.
 - [8] AlNuaimi, B. Khamis (College of Business, Abu Dhabi University, Abu Dhabi; P.O. Box 59911, United Arab Emirates); Khan, Mehmood; Ajmal, Mian, "The role of big data analytics capabilities in greening e-procurement: A higher order PLS-SEM analysis," Technological Forecasting and Social Change, vol 169, August 2021.
 - [9] Y.X. Li, S.L. Li. "Advances in Intelligent Systems and Computing," vol 1185, pp 18-22, 2021, Recent Developments in Intelligent Computing, Communication and Devices-Proceedings of ICCD 2019.
 - [10] S.R. Zhao, H.Q. Tao, Y.S. Zhang, T. Xu, K. Zhang, Z.K. Hao, E.H. Chen. Neurocomputing, vol 448, pp276-289, August 11, 2021.
 - [11] F. Wang, Y.H. Lei. "Model Analysis of Hadoop Distributed File System ," Telecommunications Science, vol. 12, pp. 5-99, 2010.
 - [12] J.C. Chen, Z.M. Ding, Y. Gao. "A review of big data hotspot technology," Journal of Beijing University of Technology, vol. 3, pp. 358-367, 2017.
 - [13] X. Q. Cheng, X.L. Yan, Y.Z. Wang, et al. "Review of big data systems and analysis techniques ,". Journal of Software, vol. 9, pp. 1889-1908, 2014.
 - [14] L.Z. Zhou, L. Lin. "A review of research on focusing reptile technology," Computer Applications, vol. 9, pp. 1965-1969, 2005.
 - [15] J.L. Li. "Linux-based python multi-threaded crawler program design," Computer and Digital Engineering, vol. 5, pp. 861-863+876, 2015.
 - [16] X.Y. Wang, L.M. Zhang. "Research on data mining engine based on big data," Electronic Design Engineering, vol. 15, pp. 31-34, 38, 2017.
 - [17] S. Liu, M.Y. Cai, X. Liu. "Visualization of Taobao Cultural Industry Data," Science and Education Wenhui, vol. 19, pp. 183-185, 2018.
 - [18] Q.Q. Hu. "Research and application of data visualization technology based on Hadoop," North China University of Technology, 2016.
 - [19] X.W. Ding, S.L. Jie, J.Y. Li. "Parallel ETL based on Spark [J]. Computer Engineering and Design," vol. 9, pp. 2580-2585, 2017.
 - [20] J.L. Yu, Z.X. Zhu, X.J. Liang. "A Data Exchange System Based on Sqoop," Internet of Things Technology, vol. 3, pp. 35-37, 2016.
 - [21] B. Yang. "Application of Sqoop data collection and storage system," Electronic Production, vol. 21, pp. 38-39, 2017.
 - [22] J.R. Chen, J.J. Le. "Overview of big data solutions based on Hadoop ecosystem," Computer Engineering and Science, vol. 10, pp. 25-35, 2013.
 - [23] B.Y. Song, J.L. Wang, W. Wang. "Research on HDFS optimized storage strategy based on Vandermonde code," Chinese Journal of Computers, 2015, (9): 1825-1837.
 - [24] C. Cui, L.J. Zheng, F.P. Han, et al. "Memory-based HBase secondary index design," Computer Applications, vol. 6, pp. 1584-1590, 2018.
 - [25] J.D.S. Ghemawat. "MapReduce: Simplified Data Processing on Large Clusters," vol. 1, no. 1, pp. 107-113, 2008.
 - [26] C.R. Yan, X. Yan, Z. Yu. IncMR: Incremental dataprocessing based on MapReduce. Proc of Cloud 2012.

- [27] Ludwig, A. Simone. "MapReduce-based fuzzy c-means clustering algorithm: implementation and scalability," vol. 6, no. 6, pp. 923-934, 2015.
- [28] T. Kong, B.Z. Cao, H.H. Qiu. "Research on Video Reptile System Based on MapReduce," Journal of Huazhong University of Science and Technology (Natural Science Edition), vol. 5, pp. 129-132, 2015.
- [29] C. Zhang. "Research and application of data processing methods for large-scale log files," Guilin University of Electronic Technology, 2014.
- [30] D. Meng, J.F. Zhan, L. Wang, et al. "Integrated Machine Group Operating System Phoenix," Computer Research and Development, vol. 6, pp. 979-986, 2005.
- [31] X.L. Tian H.B. Liu, C. Liao, et al. "Application of Phoenix+HBase storage warehouse in mobile population statistics," Journal of Xi'an University of Posts and Telecommunications, vol. 1, pp. 111-115, 2017.
- [32] Q.Y. Li, W. Yu. "Conversion design of relational database to HBase," Information Network Security," vol. 1, pp. 51-55, 2015.
- [33] A. Demichelis "Squirrel 3.0 Reference.Manual".
"Squirrel Standard Library 3.0". Alberto Demjchelis.
- [34] X.L. Hao, H.P. Zhang. "Design of Web-based Multidimensional Data Visualization System," Software Guide, vol. 8, pp. 133-136, 2018.
- [35] A.M. Nosari, M. Caira, M.L. Pioltelli, et al. "Hema e-Chart Registry of invasive fungal infections in haematological patients: Improved outcome in recent years in mould infections," Clinical microbiology and infection: European Society of Clinical Microbiology and Infectious Diseases, vol. 8, pp. 757-762, 2013.