Journal of Electronics and Information Science (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/jeis.2025.100203
ISSN 2371-9524 Vol. 10 Num. 2

Data Processing System Based on Computer Software Engineering Technology

Jiashan Zhao[1,a,*], Li Chen[1,b]

[1]*Digital Campus Construction Center, Chang'an University, Xi'an, Shaanxi Province, China*
[a]*jszh@chd.edu.cn*, [b]*chenli@chd.edu.cn*
[*]*Corresponding author*

Keywords: Data Processing System; Computer Technology; Software Engineering; System Test

Abstract: Under the background of big data, the data processing system based on computer software engineering technology can provide effective support for big data processing. This paper analyzed the data processing system based on computer software engineering technology, and put forward relevant suggestions, hoping to promote the better development of big data work in China. At the same time, people also put forward higher requirements for society, environment, life and other aspects. In this context, big data work has also been widely used. In the context of the continuous maturity, popularization and development of Internet technology, big data has become a very important and irreplaceable factor of production in all fields of society. A series of benefits and impacts generated by big data application are also very significant. The analysis and research on big data application in big data platform have greatly promoted the interconnection and exchange between various fields of society. Therefore, this paper studied the data processing system based on computer software engineering technology, and used wavelet threshold filtering algorithm to optimize processing. The research results showed that the data processing system based on computer software engineering technology could complete data processing and write into the database in about 30s under the same other conditions. The conventional system needed about 40s, which indicated that the relationship between computer software engineering technology and data processing system was positive.

1. Introduction

With the development of computer technology, the working mode of data processing has changed greatly. More attention is paid to data information, especially the reliability of information. For example, data processing software system can not only improve efficiency and accuracy, but also reduce repetitive work. At present, many enterprises have applied computer technology to establish their own computer systems and communication systems.

With the continuous improvement of technical level, computers are becoming more and more important to enterprises. Therefore, if enterprises want to achieve good economic and social benefits, they must constantly develop and innovate in today's society. In the age of big data, people

have higher and higher requirements for data processing. If enterprises want to gain an advantage in the fierce market competition, they must improve the efficiency of data processing software. Therefore, how to improve the computer data processing system in today's big data era has become a problem that every enterprise needs to think about.

This paper mainly introduces a large database system and its main functional modules, application principles and other aspects. In addition, it also puts forward the corresponding measures to improve the system, and elaborates the implementation of network communication technology, information management and storage technology based on software engineering technology.

2. Related Work

With the continuous development of computer technology, its application in various fields is more and more extensive, especially in data processing. For this reason, a large number of scholars have studied it and produced a large number of research results. Among them, with the rapid expansion of the technology field, it is both important and challenging to firmly grasp the content that the core technology can provide, especially in terms of data processing capacity. Dinh Tien Tuan Anh introduced BLOCKBENCH, a benchmark framework for understanding the performance of private blockchains under data processing workloads [1]. Pastorello Gilberto introduced this enhanced dataset and processing method in detail. These datasets were provided as open source code, making the dataset more accessible, transparent and replicable [2]. Muangprathub Jirapond developed an optimal irrigation system for agricultural crops based on wireless sensor networks, aiming to design and develop a control system through smart phones and web applications, which used node sensors in wheat fields for data management [3]. However, these studies are all analysis of data processing methods, and there is no systematic research on them. Therefore, a scientific technology is urgently needed to analyze the data processing system.

In view of the above problems, the application of computer software engineering technology to data processing system analysis has become a hot topic in today's society, and a lot of research has been carried out in the relevant neighborhood. Because of historical and technical reasons, the frequently used statistical data has always occupied the dominant position of empirical data analysis in the past and is still empirical software engineering, which is a pity. Due to many shortcomings of frequently used statistics, such as lack of flexibility, intuition and hard to explain results, the efficiency of processing heterogeneous data is reduced, and these data are increasingly applied to practical software engineering. Furia Carlo A pointed out the above shortcomings and suggested the use of Bayesian data analysis technology, which could provide practical advantages without losing detail and robustness [4]. Agrawal Amritanshu put forward a new method to implement the research and practice of software engineering: using data mining technology and optimization technology for empirical analysis. For example, a data mining program could generate a model for optimizing the exploration of the program. In software analysis, it was possible, useful and necessary to combine data mining with optimization [5]. The above research shows the applicability of software engineering in data processing, and provides rationality for the application of computer software engineering technology to data processing systems.

3. Construction of Computer Software Engineering Technology and Data Processing System

3.1 Concept of Computer Software Engineering Technology

The development history of computer software engineering technology in China is not long. In the process of computer development, software development is also a very important part. With the

acceleration of China's economic and social development, people have higher and newer requirements for science, technology and productivity. As an important part of scientific and technological innovation, computer engineering technology can not only promote the continuous improvement of science and technology, but also promote the sustainable development of the national economy. Therefore, computer software engineering technology has far-reaching significance to society and economy.

(1) Important role of computer software engineering technology in the current economic society

With the development of science and technology in China, computer technology and network communication technology are developing faster and faster. The emergence of these technologies has played a positive role in promoting China's society and economy. Computer software engineering technology mainly refers to the recombination of traditional science and technology by using a series of new science and technology and means such as computers and networks. It is managed, stored and developed through new information processing means, so as to achieve great changes in people's lifestyle and production mode. It plays an increasingly important and irreplaceable role in social life.

(2) Development of computer software engineering technology

As an interdiscipline, computer software engineering technology also needs a certain amount of comprehensive discipline knowledge. In order to better carry out software engineering work, it is necessary to understand the relevant professional knowledge. At the same time, in order to better carry out software engineering technology and project management, it is also necessary to conduct analysis and research from multiple perspectives. Finally, as far as computer software engineering technology is concerned, it needs to understand the specific business processes and formulate relevant work plans and planning strategies based on actual needs. By understanding the relevant knowledge of computer software engineering technology, its development direction and scale can be better determined, so that the development of computer software engineering technology can be more targeted and effective.

(3) Application of computer software engineering technology

With the continuous development of computer technology, the application scope of computer software engineering technology is also expanding, mainly in engineering design, project management, e-commerce and other fields. In a word, as an important part of the computer field, the development of computer software engineering technology in China has far-reaching significance. Its specific application is shown in Figure 1.

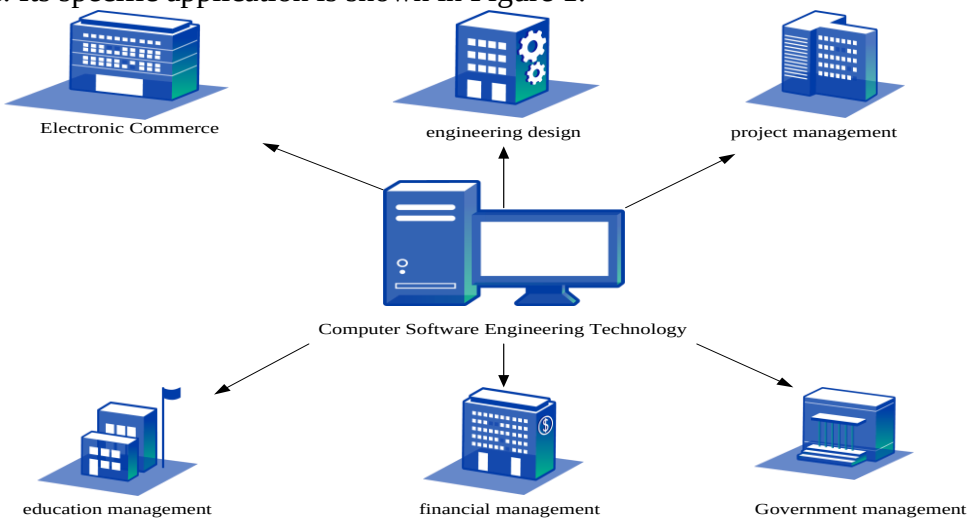


Figure 1 Specific application of computer software engineering technology

3.2 Wavelet Threshold Filtering Algorithm

Due to the limitation of the range of spatial noise control methods, noise causes the loss of the real information of the image when denoising, and the range of noise reduction is very small. The frequency domain transformation can separate the actual information of the smoother low-frequency image from the image noise signal with sharp high-frequency changes. Therefore, with the development of spectrum analysis technology, the denoising technology in image frequency domain has attracted more and more attention. Wavelet threshold filtering technology has been more and more widely used in the development of wavelet analysis technology. Based on this, this paper uses it to process the noise of the data processing system and further improve the optimization of the system [6-7]. The specific steps are as follows:

The key to the application of wavelet analysis in image processing is discrete two-dimensional wavelet transform, which represents the image signal to be processed as:

$$g(A_1, A_2) \in Z^2(T^2) \tag{1}$$

In Formula (1), A_1 and A_2 are horizontal and vertical coordinates.

First, the formula of two-dimensional continuous wavelet transform can be expressed as:

$$EU_g(x, y_1, y_2) \le g(A_1, A_2), \zeta_{x,y_1,y_2}(A_1, A_2) \ge \frac{1}{x}\iint g(A_1, A_2)\zeta\left(\frac{A_1 - y_1}{x}, \frac{A_2 - y_2}{x}\right) fA_1\, fA_2 \tag{2}$$

Among them, $\frac{1}{x}\zeta\left(\frac{A_1 - y_1}{x}, \frac{A_2 - y_2}{x}\right)$ is obtained by changing the two-dimensional mother wavelet $\zeta(A_1, A_2)$, and $\frac{1}{x}$ can ensure that the function energy do not change before and after the transformation.

Therefore, in the two-dimensional continuous wavelet, the expression of its inverse transform is:

$$g(A_1, A_2) = \frac{1}{v_c}\int_0^{+*} \frac{fx}{x^3}\iint EU_g(x, b_1, b_2)\zeta\left(\frac{A_1 - y_1}{x}, \frac{A_2 - y_2}{x}\right) fy_1\, fy_2 \tag{3}$$

The rotation scale factor X in two-dimensional continuous wavelet transform is changed to $|X| = x_{11}x_{22} - x_{12}x_{21}$. Then, a_{ok} is converted into an integer, and the discretization parameters X and $\bar{y}$ are changed into X_0^k and $X_0^{k^-}m$, so that:

$$\begin{aligned} &EU_g(k, \overline{m}) \le g(\overline{A}) \\ &\zeta_{k,\overline{m}}(\overline{A}) \ge |X_0|^{-k}\int_{T^2} g(\overline{A})\zeta\left[X_0^{-k^-}A - \overline{m}\right] f\overline{A} \\ &= |X_0|^{-k}\iint g(A_1, A_2) \bullet \zeta\left[x_{11}^k A_1 + x_{12}^k A_2 - m_1, x_{12}^k A_1 + x_{22}^k A_2 - m_2\right] fA_1\, fA_2 \end{aligned} \tag{4}$$

Among them, $\zeta_{k,\overline{m}}(\overline{A}) = |X_0|^{-k}\zeta\left[X_0^{-k}A - \overline{m}\right]$.

Wavelet threshold method is to decompose wavelet into image, and concentrate its energy and noise energy in the coefficients of each frequency domain component. Then, the selected threshold filtering wavelet coefficients are used to remove the noise energy. Finally, the processed signal is reconstructed to restore it to an image, thus obtaining the corresponding noise reduction effect. The algorithm flow is shown in Figure 2.

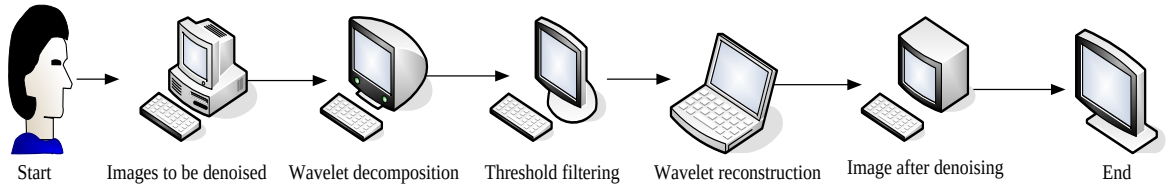


Figure 2 Flow chart of wavelet threshold denoising algorithm

After wavelet transform, the high frequency sub frequency band in the image should be filtered by threshold. Among them, the selection of threshold has a great impact on wavelet threshold filtering. There are usually two methods, one is the hard threshold method of direct filtering coefficient, and the other is the soft threshold method. Their expressions are:

$$\theta = \begin{cases} \theta & |\theta| \geq \mu \\ 0 & |\theta| < \mu \end{cases} \tag{5}$$

$$\theta = \begin{cases} \mathrm{sgn}(\theta)(|\theta| - \mu) & |\theta| \geq \mu \\ 0 & |\theta| < \mu \end{cases} \tag{6}$$

Because the soft threshold rule causes the loss of image edge details and blur, a threshold based compromise threshold filtering method is proposed. The threshold filtering of this method can be expressed as:

$$\theta = \begin{cases} (1-i)\xi + i * \mathrm{sgn}(\theta)(|\theta| - \mu) & |\theta| \geq \mu \\ 0 & |\theta| < \mu \end{cases} \tag{7}$$

Among them, i is the weighting factor, and its calculation formula can be expressed as:

$$i = \frac{\mu}{|\xi| * \exp\left\{\sqrt{\frac{|\xi| - \mu}{|\xi| + \mu}}\right\}} \tag{8}$$

Among them, the threshold μ is determined by $\mu = \varsigma\sqrt{2\log M}$ using the unified threshold method. Among them, M is the image size and ς is the noise estimation, which is generally defined as $median(|JJ_1|)/0.6745$. $(|JJ_1|)$ is the high-frequency subband with a lot of noise energy obtained after the first layer wavelet decomposition. The above formula can be used to reduce noise,so as to further optimize the digital processing system [8].

3.3 Construction of Data Processing System

(1) System composition and functions

Software engineering is a method of software development based on computer technology. The system mainly includes two parts: database and application program. The data processing system is one of the core and key links of the platform, and its functions include data collection, data management, data mining, data analysis and application [9-10]. As the foundation, database needs to be designed and developed, and the application program needs to optimize the management of the database after realizing the application of functions. First of all, during data collection, it is necessary to ensure that user permissions and donkey safety control are handled well, and relevant

technicians are trained. The data information shall be processed according to the actual needs. Then, various computing technologies are used to process and calculate data information. Finally, a variety of database technologies can also be used to complete relevant functional requirements to achieve optimal management of big data processing system platform [11-12].

(2) Implementation function

During the design, the system should be analyzed according to the actual needs to ensure that there are no problems in the process of data processing. Specifically, the system includes data acquisition, storage and display functions. For some large databases, the system can realize query, import and other functions. The system needs some operations in actual use, mainly including data input and output and various data storage. For a more complex information system, it is not only necessary to maintain and manage it, but also to meet the requirements of information acquisition. Based on this, relevant standards should be applied to the system and effectively implemented in the design. For some large data, corresponding technical standards need to be selected and applied according to actual needs [13-14].

(3) System advantages

The system has the following advantages: it can select the appropriate database according to the actual needs; it has good security; it can provide high reliability and stability; it can be able to effectively connect the database information with the internal data information of the enterprise; it can realize classified management and summary of data information [15]. With the development of society, computer technology plays an important role in all walks of life. It is particularly important to improve the efficiency of data processing in the context of social informatization.

(4) System improvement measures

1) Adopting reasonable data management methods

2) Selecting appropriate technology to improve the efficiency of data processing

3) Improving database reliability

4) Reasonably controlling variables and appropriately limiting the ability of computer system operators

5) Strengthening data processing to ensure its accuracy and integrity

4. Data Processing System Test Based on Computer Software Engineering Technology

4.1 System Test Environment

The test environment of the system includes software and hardware. The operating system, tool software and other aspects adopt relatively stable official versions to ensure the performance of the system. Software operating system: Linux (Centos6.9), hardware memory 36G, hard disk 5TB, which ensures the reliable operation of the system under computer software engineering technology. The configuration list of the system test environment is shown in Table 1.

Table 1 List of test environments

CDH cluster (8 nodes)	Software	CDH5.12, Spark1.5, Hive0. 13
	Hardware	10 core memory
Storm cluster (3 node)	Software	Kafka0.11.0.1, Storm0.9.3
	Hardware	10 core memory
Redis node (2 nodes)	Software	redis
	Hardware	10 core memory
Kettle node (2 nodes)	Software	spoon5.14
	Hardware	10 core memory

4.2 System Functionality Test

This part adopts the black box test method to test the functions of each system through test cases. The specific test process includes: data collection, offline data processing, data mining, message queuing, stream processing functions, etc. The functional test evaluation criteria of the system are shown in Table 2.

Table 2 Evaluation criteria of system function test

Function Description	Function test of data processing system based on computer software engineering technology	
Case Purpose	Ensure the normal operation of the main functions of the data processing system and ensure the accuracy of the data	
Content	Methods and Steps	Critique
Streaming	Real time data collection is transmitted to Storm through a message queue for stream processing, and then written to the database	The data statistics in the database are correct
Data Warehouse	Get the source data, and then preprocess it and import it into the data warehouse. More businesses need to make statistics on relevant reports	It can complete the collection, analysis, preprocessing, import of hive form, collection of complete logs, extraction, conversion and loading of hive form of MP logs
Real Time Bidding	Real time quotation, real-time collection, sending message queue to Storm for stream processing, and then writing to the database	Establish efficient and reasonable data calculus model

(1) Data acquisition function test

The test data collection module includes: database data collection, offline log collection and streaming data collection module testing. Table 3, Table 4 and Table 5 are the test cases of the three modules. According to the offline data of the system, the three modules have been tested respectively, and good results have been achieved. The performance indicators of the system have been verified through experiments. (A: case number, B: case name, C: purpose, D: case level, E: process, F: expected result, G: actual result)

Table 3 Test cases of database data collection function

A	1
B	Database data collection
C	Ensure that the data in MySQL and Hive tables are the same
D	Functional test
E	a. Clear the data in the hive table b. Set Sqoop script and related connection test parameters c. Use Sqoop script and synchronize MySQL data to Hive table
F	Both data are the same without error
G	System success

Table 4 Test case of offline log data collection function

A	2
B	Offline log data collection
C	Get the same data file as each server node from Kettle's job
D	Functional test
E	a. Determine test related parameters for Kettle scripts b. Output date, execute Kettle script, and extract data from each node to the hard disk in the service storage c. Run the Sqoop script to upload the data file to the path specified by HDFS
F	Use Kettle's script to collect log files that are consistent with those on each server. The file size and content are the same. The file is not lost and no error occurs
G	System success

Table 5 Test case of streaming data collection function

A	3
B	Streaming data collection
C	The data generated in Kafka is consistent with the original record, and has time guarantee
D	Functional test
E	a. Set test related parameters for each node Flume b. Start running Flume program and run the cluster c. Write the log file to the test data of Flume monitor
F	The real-time data on the server is completely consistent with the data in Kafka, without any omission. In 5S, it can almost ensure the smooth transmission
G	System success

(2) Offline data processing function test

Its function test is divided into two parts, namely: data preprocessing and data statistical analysis. Table 6 and Table 7 are the test cases of the two modules. Through the verification of the test cases of the two modules, good results have been achieved, and the system meets the expected functional requirements.

Table 6 Test cases of data preprocessing function

A	4
B	Offline data preprocessing
C	Extract a specific field data from the original log file and store it in a hive table
D	Functional test
E	a. Upload MP test data and DSP test data to the specified directory in HDFS b. Run HiveSQL program in DSP to preprocess test data c. Run MapReduce and HiveSQL programs to process data in MP test
F	Generate data in specified format in Hive table and compare with test data. There is no difference between the two
G	System success

Table 7 Data statistics analysis function test cases

A	5
B	Statistical analysis of offline data
C	Make statistics on the preprocessed data according to different needs and store it in the hive table
D	Functional test
E	a. Upload test data in the relevant form of Hive b. Import parameters and run HiveSQL script for statistical analysis
F	DSP and MP data are connected, and the results are correct according to different dimensions
G	System success

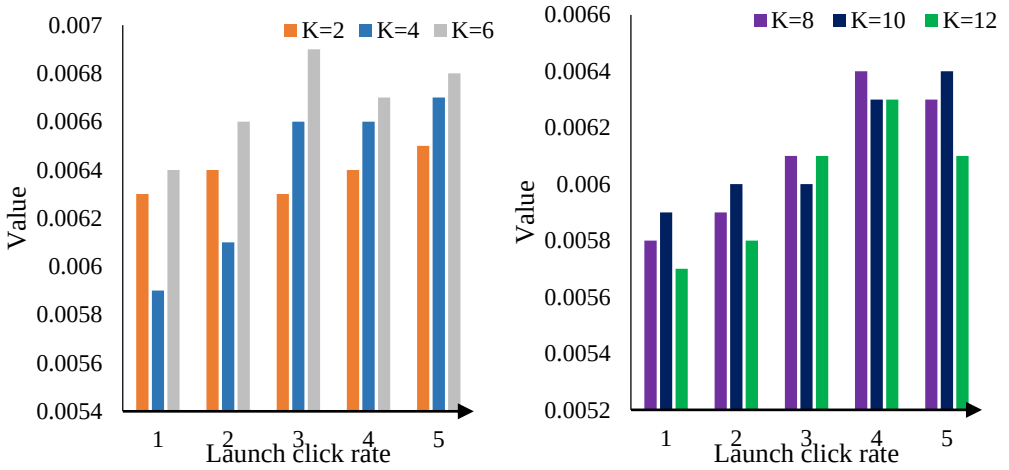
(3) Data mining function test

It mainly tests the establishment of high-value group model and CTR (click through rate) model, as shown in Table 8.

Table 8 Data mining function test cases

A	6
B	Data model construction
C	Based on the characteristics of data, select appropriate algorithms to construct CTR and high-value group models
D	Functional test
E	a. Pre process the test data b. Substitute the processed sample data into the CTR and high-value group models respectively c. Evaluate and optimize CTR and high-value group model d. Online data model, test results
F	Both models were successfully built
G	System success

When the launch click rate is 1, 2, 3, 4, 5, the launch effect of different K values is shown in Figure 3.



(a) Release effect when K=2, 4, 6 (b) Release effect when K=8, 10, 12

Figure 3 Release effect under different K values

It can be seen from Figure 3 (a) that when K=2, the release effects are 0.0063, 0.0064, 0.0063, 0.0064 and 0.0065 respectively. When K=4, the release effect is 0.0059, 0.0061, 0.0066, 0.0066,

0.0067 respectively. When K=6, the release effect is 0.0064, 0.0066, 0.0069, 0.0067 and 0.0068 respectively. As shown in Figure 3 (b), when K=8, the release effects are 0.0058, 0.0059, 0.0061, 0.0064, and 0.0063 respectively. When K=10, the release effect is 0.0059, 0.006, 0 006, 0.0063, 0.0064. When K=12, the release effect is 0.0057, 0.0058, 0.0061, 0.0063 and 0 0061. It can be seen from Figure 3 that, regardless of the launch click rate of 1, 2, 3, 4 and 5, when K=6, the launch effect is the best.

This paper uses the ROC (Receiver Operating Characteristic) curve of the CTR model to analyze the data processing system based on computer software engineering technology and the conventional data processing system. The horizontal and vertical axes in the curve represent the ratio of false positive and true positive respectively. The AUC (Area Under Curve) value represents the area below the ROC curve. The larger the area, the better the model effect. The specific results are shown in Figure 4.

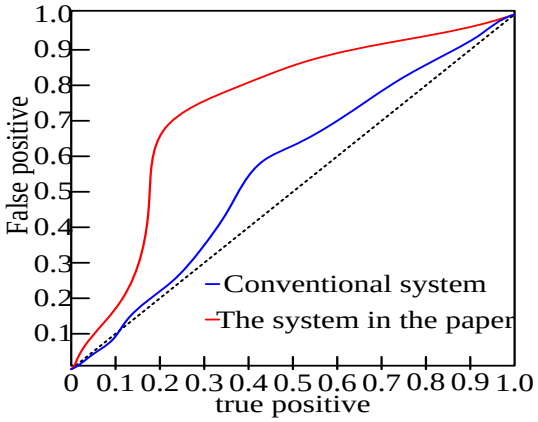


Figure 4 ROC curve of CTR model of two systems

It can be seen from Figure 4 that the AUC value area of the system in this paper is significantly larger than that of the conventional system, and the test results show that the system meets the expected requirements of this function.

(4) Message queue function test

The message queue function test mainly tests the construction process of the data queue, as shown in Table 9.

Table 9 Message queue function test cases

A	7
B	Message data queue construction
C	Kafka obtains data from Flume program and stores it in different topics for Storm to read
D	Functional test
E	a. Start Kafka running program b. Start Flume program and boot different topics of the test program in Kafka c. Read data from different topic of Kafka
F	Flume data is successfully written to the topic specified by Kafka. Storm can correctly read Kafka data and accurately match human data with consumption data
G	System success

(5) Streaming function test

The streaming processing function test is mainly the test of the Storm data processing function module, as shown in Table 10.

Table 10 Test case of streaming processing function

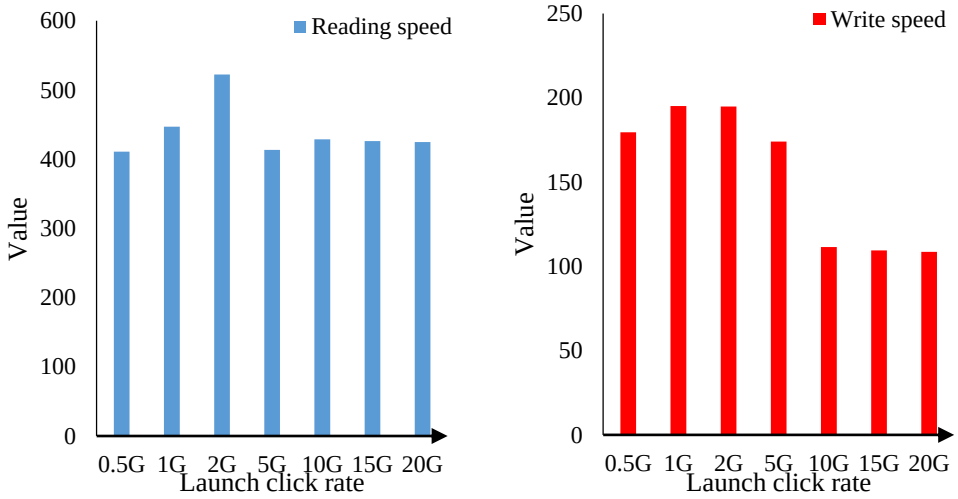
A	8
B	Storm data processing
C	Storm program reads data and writes statistics in the database
D	Functional test
E	A. Submit the data handler to the Storm cluster B. Generate and transmit real-time test data to Kafka
F	The data in the database has been successfully updated and conforms to the statistics
G	System success

4.3 System Non Functional Test

The non functional requirements of the whole system is tested, including Hadoop and Storm cluster performance tests.

(1) Hadoop

The performance test of Hadoop cluster mainly considers the following issues. One is the reading and writing speed of the cluster, and the other is the processing of small data. To test the file read/write performance of the Hadoop cluster, first the cluster section of the Hadoop cluster is set to 7. Then, the test document of the specified size is wrote and read to the cluster, and the corresponding read throughput rate is obtained, as shown in Figure 5.



(a) Read speed (b) Write speed

Figure 5 System throughput performance test report

As shown in Figure 5 (a), the read speeds are 410.814MB/s, 446.854MB/s, 522.328MB/s, 413.52MB/s, 428.725MB/s, 426.125MB/s and 424.681MB/s respectively. As shown in Figure 5 (b), the write speeds are 179.505MB/s, 195.123MB/s, 194.864MB/s, 174.007MB/s, 111.504MB/s, 109.454MB/s and 108.648MB/s respectively. It can be seen from Figure 5 that under different file sizes, the system' write speed does not change much, and can be controlled within the range of 100-200 MB/s. However, it has a great impact on read performance. The difference between the fastest and slowest read speeds has exceeded 100MB/s. With the increase of data, the reading speed is relatively stable. In general, Hadoop clusters can handle a large number of files.

In the scalability test of Hadoop cluster, TeraSort, a built-in instance test tool, is used. The file size is set to 50G, and the time spent sorting files under different node numbers are tested, as shown in Figure 6.

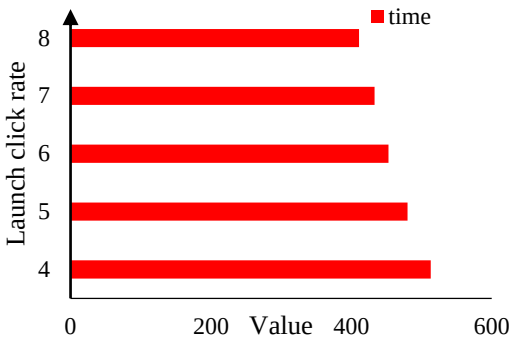
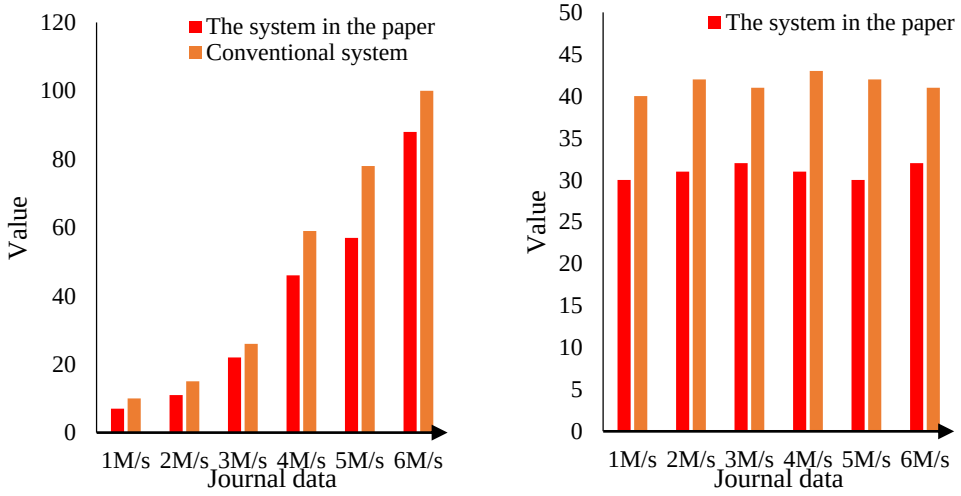


Figure 6 System expansion performance test report

It can be seen from Figure 6 that the computing time decreases with the increase of the number of nodes, indicating that the computing power of the cluster increases. It shows that its scalability is good, and cluster computing capacity can be enhanced by expanding nodes.

(2) Storm

The performance of the Storm cluster mainly tests the throughput of the system in processing data. It mainly aims at the overall time consumption of cluster processing data under different data flows. It is ensured that the Storm cluster can still complete the logic processing function within the specified time when the data traffic increases dramatically. In this paper, the journal data is 1M/s, 2M/s, 3M/s, 4M/s, 5M/s, and 6M/s, and the corresponding number of journal entries is about 10000 entries/s, 20000 entries/s, 30000 entries/s, 40000 entries/s, 50000 entries/s, and 60000 entries/s, respectively. They are brought into this system and the conventional system for testing. The results are shown in Figure 7.



(a) Processing duration (b) Writing duration to database

Figure 7 Storm performance test results

It can be seen from Figure 7 (a) that the Storm cluster processing time of this system is 7ms, 11ms, 22ms, 46ms, 57ms and 88ms respectively. The processing time of Storm cluster in conventional system is 10ms, 15ms, 26ms, 59ms, 78ms and 100ms respectively. As shown in Figure 7 (b), the system writes to the database for 30s, 31s, 32s, 31s, 30s and 32s respectively. The normal system writes to the database for 40s, 42s, 41s, 43s, 42s, 41s respectively. It can be seen from Figure 7 that the processing time of both systems increases with the increase of data volume. However, the system in this paper can complete data processing and write to the database in about 30s. The conventional system needs about 40s, which shows that computer software engineering

technology is conducive to the optimization of data processing system.

5. Conclusions

For enterprises, they have encountered many problems in data processing, mainly reflected in low data quality and low efficiency. The data processing system of computer software engineering technology mainly includes three aspects. The first is to clean and sort out the data to ensure its correctness and integrity. The second is to ensure the accuracy of data and improve efficiency in the use process. The third is to further improve the practicability of the system by optimizing the system and designing each module and function. From the current situation, computer software engineering technology has been widely used in all walks of life, laying a solid foundation for enterprises to achieve high-quality development. In this paper, the functions of the new system were tested, and the system run well. Compared with the conventional system, the positive effect of computer software engineering technology on the data processing system was verified.

References

[1] Dinh Tien Tuan Anh, Liu Rui , Zhang Meihui , Chen Gang , Ooi Beng Chin , Wang Ji . "Untangling blockchain: A data processing view of blockchain systems." IEEE transactions on knowledge and data engineering 30.7 (2018): 1366-1385.

[2] Pastorello Gilberto, Trotta Carlo , Canfora Eleonora , Chu Housen , Christianson Danielle , Cheah You-Wei ,et al. "The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data." Scientific data 7.1 (2020): 1-27.

[3] Muangprathub Jirapond, Boonnam Nathaphon,Kajornkasirat Siriwan,Lekbangpong Narongsak,Wanichsombat Apirat,Nillaor Pichetwut. "IoT and agriculture data analysis for smart farm." Computers and electronics in agriculture 156 (2019): 467-474.

[4] Furia Carlo A., Robert Feldt, and Richard Torkar. "Bayesian data analysis in empirical software engineering research." IEEE Transactions on Software Engineering47.9 (2019): 1786-1810.

[5] Agrawal Amritanshu, Tim Menzies, Leandro L. Minku, Markus Wagner, Zhe Yu. "Better software analytics via "DUO": Data mining algorithms using/used-by optimizers." Empirical Software Engineering 25.3 (2020): 2099-2136.

[6] Roh Yuji, Geon Heo, and Steven Euijong Whang. "A survey on data collection for machine learning: a big data-ai integration perspective." IEEE Transactions on Knowledge and Data Engineering 33.4 (2019): 1328-1347.

[7] Ienca Marcello, and Effy Vayena. "On the responsible use of digital data to tackle the COVID-19 pandemic." Nature medicine 26.4 (2020): 463-464.

[8] Assarroudi Abdolghader, Heshmati Nabavi Fatemeh , Armat Mohammad Reza , Ebadi Abbas , Vaismoradi Mojtaba . "Directed qualitative content analysis: the description and elaboration of its underpinning methods and data analysis process." Journal of Research in Nursing 23.1 (2018): 42-55.

[9] Ho Joses, Tumkaya Tayfun , Aryal Sameer , Choi Hyungwon , Claridge-Chang Adam . "Moving beyond P values: data analysis with estimation graphics." Nature methods 16.7 (2019): 565-566.

[10] Belotto Michael J. "Data analysis methods for qualitative research: Managing the challenges of coding, interrater reliability, and thematic analysis." The Qualitative Report 23.11 (2018): 2622-2633.

[11] Mahdavinejad Mohammad Saeid, Rezvan Mohammadreza,Barekatain Mohammadamin,Adibi Peyman,Barnaghi Payam,Sheth Amit P.. "Machine learning for Internet of Things data analysis: A survey." Digital Communications and Networks 4.3 (2018): 161-175.

[12] Elliott Victoria. "Thinking about the coding process in qualitative data analysis." The Qualitative Report 23.11 (2018): 2850-2861.

[13] Turkbey Baris, Rosenkrantz Andrew B.,Haider Masoom A.,Padhani Anwar R.,Villeirs Geert,Macura Katarzyna J.,et al. "Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2." European urology 76.3 (2019): 340-351.

[14] Oussous Ahmed, Benjelloun Fatima-Zahra,Lahcen AyoubAit ,Belfkih Samir. "Big Data technologies: A survey." Journal of King Saud University-Computer and Information Sciences 30.4 (2018): 431-448.

[15] Reichstein Markus, Camps-Valls Gustau , Stevens Bjorn , Jung Martin , Denzler Joachim , Carvalhais Nuno ,et al. "Deep learning and process understanding for data-driven Earth system science." Nature 566.7743 (2019): 195-204.