

## Computer Information Processing Based On Big Data

Qunwei Lv, Yi Jin

East China Jiaotong University, Nanchang Jiangxi, 330013, China

**Keywords:** Big Data Background, Computer Information Processing, Storage Mode, Hadoop File System

**Abstract:** Along with our country's information technology, communication technology and the rapid development of computer technology, it has been widely used in many fields, in the era of big data, of computer information processing ability put forward higher requirements, and how to deal with large data brings challenges and opportunities, is need to face the key issues in computer field. In order to further improve the ability of computer information processing, this paper studies the development of computer information processing technology under the background of big data. Based on Hadoop file system as the object, this paper studies the distribution processing architecture and data storage, and the process of data read and write operations, refer to learning a variety of computer application in practical production to the design of the distributed file system, based on the architecture of some tuning, improve the storage and stability of HDFS file system, make the system to meet the demand of computer information is stored. The results show that the computer information processing capacity has increased significantly, and the storage rate has increased by 12.28%, which proves the effectiveness of the technology.

### 1. Introduction

In the new era, the technology related to big data develops rapidly and has been widely applied in many professional fields, but it also puts forward stricter technical requirements for modern computer information technology. At present, the efficient data processing capacity of computer information technology cannot meet the application demand of big data in China [1-2]. Therefore, under the new background based on big data, it is necessary to make continuous effective technological innovation on computer data processing capacity technology [3-4].

The information processing program includes image acquisition instruction, search target detection instruction, shape recognition instruction, event occurrence instruction, virtual image generation instruction and display control instruction. Kazufumi Kobashi says the search target detection instruction allows the computer to detect the search target from an image of the subject obtained according to the image acquisition instruction. The event occurrence instruction enables the computer to make the event happen in the virtual space according to the shape of the search target[5]. Takagi Yasuhiro introduces an information processing device, including: a communication unit that communicates with an image forming device; A storage unit; A receiving unit receives a set value of the image forming device; The acquisition unit obtains information about the state of the image forming device from the image forming device to change the setting of the set value [6]. Mikinori Ehara according to the present invention an information processing device project configuration for a given area in a three-dimensional images to a flat surface consists of a two-dimensional image of an object, and leads to a display to display a projection error range, including the projection area, the way is to cover the range of error on the 2 d images. Therefore, users can effectively search for regions corresponding to regions of interest in three-dimensional images in two-dimensional images [7].

In order to better study the computer information processing technology in the background of big data, this paper started from the computer information processing technology and established a distributed Hadoop cluster to increase the caching mechanism to meet the needs of data processing [8-9]. By establishing the auxiliary information database on the processing server, the corresponding relationship between the data file name and the auxiliary information is established

to realize the traceability function and efficient access of the system, and technical Suggestions are provided to improve the information processing capacity of the computer [10].

## 2. Big data and information processing technology

### 2.1 Big Data

Big data technology has the following major characteristics: First, it has a huge amount of data. Based on information processing data, big data takes a large number of important data types as its development premise. The second is the data type changing, along with the modern computer information technology and mobile Internet vigorous development of information technology, People's Daily life, work and technology further the dependence of information processing technology, in the era of information industry, the big data information to present the geometric distribution of speed increasing, the data processing speed and the amount of data storage is also important need to ask for height, in the information technology people face a lot of industry data, need timely and effective data analysis on the data, to get more valuable data information in time.

### 2.2 Information Processing Technology

In order to meet the needs of the development of The Times, information processing and acquisition technology should be innovated. On the premise of ensuring the integrity of information collection, the information should be scientifically sorted and analyzed to ensure the clear target of information collection, and the target source should be tracked and supervised as necessary, and the data flow should be controlled to ensure the rationality of data collection. HDFS (Hadoop Distributed File System) is a Distributed computing File System based on Hadoop. The essence of the file system is a way to determine the partitioned files or disks for the OS. It is also a special data structure used to consolidate files on disk. It is also used to partition storage files and to store both disks.

According to the different mapping functions, the global mapping image compression algorithm can be divided into linear compression and nonlinear compression. Among them, the most representative linear compression algorithm is the direct linear compression method, which USES a unified linear function to compress the image. The algorithm is simple and easy to implement, and its application is also widespread. For example, automatic gain control, whose mapping relationship is as follows:

$$g(x, y) = g_{\min} + \frac{g_{\max} - g_{\min}}{f_{\max} - f_{\min}} \times (f - f_{\min}) \quad (1)$$

Taking the method of compression mapping based on S-curve function as an example, its specific expression is:

$$g(x, y) = \frac{c \cdot f(x, y)}{f(x, y) + e^{b-a \cdot f(x, y)}} \quad (2)$$

## 3. Design of the experiment

### 3.1 Experimental Background

The rapid progress and wide application of big data information technology provides a broad space for the development of modern computer information technology and puts forward higher technical requirements for information processing. By constantly strengthen the data of computer application system research of information processing technology, can effectively improve the efficiency of the computer system operation management, improve the computer system has the function of information processing of huge amounts of data collection, and as an important technology to realize effective data processing, premise condition is a normal must ensure that the data information collection processing job smoothly. Therefore, the computer information system is required to collect and sort all kinds of data according to different ways of information before

starting to collect and process all kinds of data in real time, so as to greatly improve the accuracy and efficiency of data processing.

### 3.2 Experimental Design

To some food production enterprise daily production information as the experimental sample, according to upload the information stored by the computer time, the information such as the length, in combination with post label equipment, plane reservation to check equipment and monitoring content, the video data are named according to the information, and then into the Datanode server, data storage, Datanode to backup data. The Namenode server records where blocks of data and backups are stored. When the client traces the video according to the production date or production batch number, the transaction processing server queries the name and storage location of the video file according to the information and returns it to the client, which directly reads the data to the Datanode where the data resides. The experimental results are shown in Table 1.

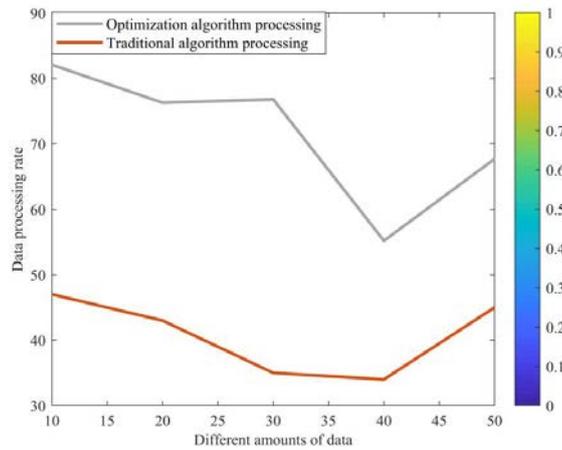
**Table 1.** Experimental results

	MRLLF	SGLLF	FLLF
Sequence file 1	0.2183	0.2134	0.2213
Sequence file 2	0.4211	0.4266	0.4228
Sequence file 3	0.7724	0.7213	0.7436
Sequence file 4	0.5664	0.5142	0.2614

## 4. Research on computer information Processing under big data

### 4.1 Analysis of Computer Information Processing Based on Big Data

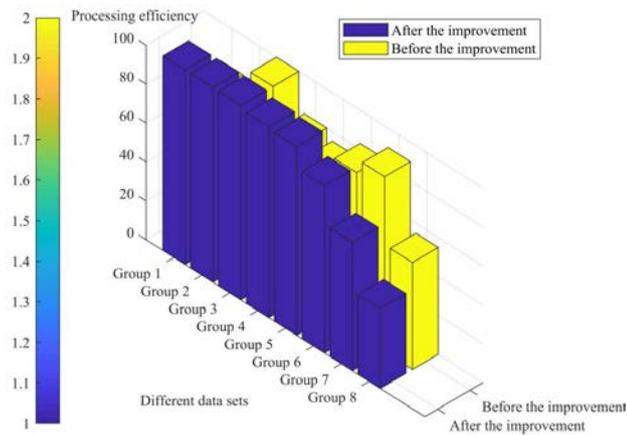
As shown in Figure 1, the SSIM value of the algorithm in this paper achieved the optimal value in the simulation, indicating that for different data groups, the algorithm in this paper kept the original features to the greatest extent. Since HDFS has only one Namenode and is usually deployed on cheap servers, there is a certain probability of hardware failure or software error. When a problem occurs, the Namenode can no longer provide metadata to the system, rendering the entire file system corrupted. Although running Secondary, Namenode can assist Namenode to restore operation, it is not hot backup, but cold backup, which is time-consuming to switch over. The Namenode has two important files, the metadata Mirror file (Fsimage) is responsible for keeping the directory tree of the file system, and the metadata Operations Log (EdITS) is responsible for modifying the directory tree. Keep an up-to-date metadata image in memory, namely the Fsimage file and the Edits file. HDFS supports a hierarchical tree file structure, which is no different from a traditional file system. Users can create a directory, put the files in the directory, and add, delete, query, and so on. Due to the multi-copy policy of HDFS, the file system is not good at modifying files. The namespace node is the core of the system, responsible for managing the system's namespace and recording changes to the namespace or properties in the system. The namespace node is also responsible for managing the system's copy placement policy, and users can set the number of copies of files.



**Figure 1.** Data processing rate of optimization algorithm in simulation experiment

The Datanode is the slave node of the file system and is the location where the actual data blocks are actually stored in the HDFS system. When the connection between the data node and the Namenode is established, the heartbeat will be maintained continuously with the Namenode, and all the existing Block information will be sent to the Namenode periodically. The message returned by the heartbeat contains the Namenode's instructions to the data node, such as copying the data Block to another data node or deleting the data. The Datanode can also interact directly with the client. When the client obtains the location of data storage from the Namenode, it can directly establish a connection with the Datanode to write and read data from the Datanode. When a client makes a data backup, the Datanode creates an exercise to make a copy of the data according to the rules and other data nodes. The communication mechanism between datanodes also ensures that when the file is written in the cluster, the consistency of the data of multiple copies is analyzed for the HDFS architecture and read-write process. It can be seen that there is only one Namenode server in HDFS providing metadata service for the entire system, and the performance, reliability and scalability of the entire HDFS depend on the Namenode server. In HDFS for most application scenarios, lightweight Namenode node is enough to meet the needs of the file system, but in large-scale applications, with the increasing amount of data, the expansion of storage, system data block node Datanode, increasing the amount of metadata is correspondingly increase, the Namenode nodes may fail, the whole system will not operate properly.

As shown in Figure 2, the EPI value of this algorithm is only 4.2% and 10.3% lower than the optimal value of the two algorithms. To sum up, the computer information processing technology used in this paper is more superior in objective evaluation, ensuring that the computer system can obtain high-quality computer images suitable for subsequent production and processing. For food production enterprise in experiment in this paper, first of all, to ensure the integrity of the information collection, the science of food producing all kinds of information and analysis, the specific goals and ensure that information collection, and the target source in carrying out the necessary follow up and supervision, to control the data flow, ensure the rationality of the data acquisition; Secondly, after the completion of information collection, the data and information should be scientifically processed according to customer needs, and the information needed by customers should be separated and provided to customers. During the information processing and transmission, attention should be paid to the security protection of information, and advanced technologies should be adopted to prevent information theft and leakage.



**Figure 2.** Acquisition efficiency of traditional information processor and optimized information processor

#### 4.2 Suggestions on Computer Information Processing Based on Big Data

Under the background in the era of big data, it is necessary to put forward higher requirements, the key technology in the enterprise information, due to the information in the database system database and information usually has the basic characteristics of interconnected, so if a key link of the whole process of information processing, security problems will directly lead to the use of the whole system database information security is greatly reduced. In order to realize the effective security protection of enterprise database, more attention should be paid to the basic research of enterprise information security and technology. Enterprises should not only strengthen the research and protection of database security, but also pay attention to the security of the information stored in the database.

A large amount of information and data increase in geometric form, which makes the new computer data storage and processing technology more strict. However, at present, there are still some technical defects in the data storage and processing capacity required by computers, which cannot fully meet the continuous development and application requirements of industrial big data. Therefore, the data storage capacity technology system needs to be gradually optimized and continuously improved. First of all, we should pay attention to constantly improve the data storage of modern computers and provide enough storage space for a large number of information data acquisition based on the application characteristics of current big data. Secondly, the way to store a large amount of information should be orderly and reasonable as far as possible, so as to enhance the efficiency and scientificity of data collection and analysis of a large amount of data query information. The orderly information arrangement method should be used to make up for the gaps in data query and prevent various problems of loss of a large amount of data information.

Computer is mainly supported by network hardware operating system and network software infrastructure, which can ensure the timely and effective collection and processing of data in network information. However, in the current era of big data, the speed of automatic information processing capacity of computers cannot fully meet the demands of its rapid development. However, with the continuous development of mobile cloud computing technology, it has partially replaced the automatic information processing capacity that traditional computer information systems cannot have. Therefore, in order to fully meet the new requirements for the development of modern big data applications, need to be further improved to optimize network information processing speed, will be closely integrated network cloud data of computer and network technology fully, make network information processing more efficient, reasonable and orderly, provide a new technical foundation for the development of large data application support.

#### Conclusions

With the continuous development of computer technology and network technology, the

emergence of massive data information, the world has been transferred from the era of data to the era of big data. The advent of the era of big data has brought great impact on computer information processing technology. Based on this, this article embarks from the large data under the background of computer information processing, information processing technology are introduced, and research, analyzes the working principle and configuration method of Hadoop, as the core of the underlying data storage service technology, build HDFS storage platform, and in view of the actual project need to be configured tuning, completed the rational the goal of highly efficient storage and processing of computer data.

## References

- [1] Nintendo Co., Ltd. Computer-readable storage medium having stored therein information processing program, information processing apparatus, information processing system, and information processing method[J]. *Natural Language Engineering*, 2017, 259(3):e40.
- [2] Jiang Y. Research on linguistic group decision making method based on computer information processing[J]. *Agro Food Industry Hi Tech*, 2017, 28(1):803-807.
- [3] Brother Kogyo Kabushiki Kaisha. Image information processing method, image information processing apparatus and computer-readable recording medium storing image information processing program[J]. *Sony*, 2016, 4(1):85-96.
- [4] Canon Kabushiki Kaisha. Medical imaging apparatus, information processing method, and computer-readable storage medium[J]. *Procedia - Social and Behavioral Sciences*, 2015, 141(1):781-790.
- [5] Kazufumi Kobashi. Information processing apparatus, control method for information processing apparatus, and program[J]. *Marine & Petroleum Geology*, 2017, 81(6):97-111.
- [6] Takagi, Yasuhiro, Mori, Kazuhiro. Processing device, squealing sound generation method in information processing device, and computer program product[J]. *Journal of the Acoustical Society of America*, 2017, 132(4):2769.
- [7] Mikinori Ehara, Takezo Fujishige. Information processing apparatus, method, and computer product, for file naming[J]. *Neural Computation*, 2018, 14(9):2181-2200.
- [8] Ryo Ishikawa, Takaaki Endo, Kiyohide Satoh. Information processing apparatus and information processing method having communication function[J]. *Journal of Materials Processing Technology*, 2016, 14(9):2181-2200.
- [9] Shida, Muneyuki. Storage medium storing computer program, information processing system, and information processing method[J]. *Progress in Geography*, 2017, 259(3):e40.
- [10] Hashem I A T, Yaqoob I, Anuar N B, et al. The rise of "big data" on cloud computing: Review and open research issues[J]. *Information Systems*, 2015, 47(jan.):98-115.