Scholars Journal of Engineering and Technology

OPEN ACCESS

Abbreviated Key Title: Sch J Eng Tech
ISSN 2347-9523 (Print) | ISSN 2321-435X (Online)
Journal homepage: https://saspublishers.com

Research on Big Data Resource Management and Optimization Based on Cloud Computing Technology

Zhang Dong-Liang[1*]

[1]School of Computer Science & Technology, Zhoukou Normal University, Zhoukou, China

DOI: 10.36347/sjet.2023.v11i03.001 | **Received:** 23.01.2023 | **Accepted:** 28.02.2023 | **Published:** 04.03.2023

***Corresponding author:** Zhang Dong-Liang
School of Computer Science & Technology, Zhoukou Normal University, Zhoukou, China

Abstract | **Original Research Article**

This paper points out the problems of irrational, complex and low utilization rate of massive data storage in most industries at present, analyzes the meaning and characteristics of big data, thus determining the necessity of distribution and management of big data, providing reference ideas for the management of Chinese big data resources, and further promoting the big data resources to play its due role. To provide support for the transformation of science and technology management decision-making, to achieve social and technological progress. On this basis, from different aspects of the mass data management methods and strategies and the significance and efficiency of these methods to achieve, improve the utilization rate of cluster resources, increase the ability of data sharing, a variety of computing frameworks can also share a distributed storage data, finally through the experiment demonstrated the effectiveness and security of the above strategies.

Keywords: Massive data; Distribution and management; Strategy; Computing framework.

1. INTRODUCTION

Big data, as one of the core technologies in the Internet + era, plays an important role in modern society. Both enterprise production and operation and People's Daily life are closely linked with big data. Correspondingly, big data resources have become a hot resource in current social production. Compared with natural resources, big data resources have the characteristics of unlimited replication and convenient transmission, which gives birth to the problem of distribution and management of big data resources. In the era of big data, how to ensure the rational utilization of big data resources and how to guarantee the security of big data resources has become a key content to be studied in the field of computer science in the new era. Therefore, it is an inevitable trend for the utilization of big data resources to make use of Internet technology to allocate and manage big data resources from the actual situation. However, there are still many problems in the process of big data resource allocation and management. It is necessary to enhance the efficiency and quality of big data resource allocation and management based on the actual problems. As the main product of the rapid development of big data technology in the era of "Internet Plus", big data resources play a huge role in modern social production, commercial operation and people's life. However, with the rapid development of big data resources, the security of people's personal information is also seriously threatened because of big data. Telecom fraud cases emerge in endlessly, and the amount involved is appalling. Therefore, it is necessary to carry out effective distribution and management of big data resources, and the implementation or supervision of relevant state departments, to ensure that big data resources are applied to the right way.

2. RELATED CONCEPTS AND ADVANTAGES OF BIG DATA

Big data resources, as the name implies, are Internet resources obtained through big data technology. Big data collects resources on the Internet and accurately paints users according to their Internet footprints, so as to realize commercial push or user description that better meets users' needs. Big data resources, born with the birth of big data technology, play an important role in the Internet + era. Big data resources have changed the consumption pattern of our social consumers to some extent, driving the upgrading and progress of social consumption level. At the same time, big data resources also allow users to enter a larger information cocoon, bringing certain negative impacts. At the beginning of its birth, the purpose of big data resources is mainly to paint portraits of people. It is

to predict their possible behaviors in the future through the traces left by people in the history, so as to provide reference materials for user behavior prediction. To this day, in our country, big data resources have become a very common resource, and the commercial push of various businesses is closely linked with big data. The work of the Party and the state has gradually realized big data, and the use of big data resources can more accurately meet the needs of users, so as to improve the utilization rate of resources and ensure that resources are obtained by those who need them more.

Big data resources were born in the Internet and are special Internet resources. Therefore, big data resources not only have some characteristics of traditional information resources, but also have their unique characteristics. Specifically, big data resources have three characteristics.

(1) Great amount of resources

First of all, big data resources are huge, which is the inevitable result created by big data technology. Just for a user, big data can collect all his footprints and information on the Internet, including browsing record, consumption record, search record, etc. These data are collected and organized to form a portrait of the user. And the data of a single user is extremely rich. In the era of Internet +, almost every Internet user is exposed to massive resources every day. Even if the user can only select part of them to pay attention to, the long-term accumulation will also be an astronomical number. And big data resources need to collect and integrate all these data, so big data resources are bound to be massive resources.

(2) Strong replicability

As an Internet resource, big data resources are replicable, which is one of its basic characteristics. Currently, big data resources in China are mainly stored in cloud servers of various platforms, and users who have access to the cloud can obtain the big data resources, instead of data resources that cannot be replicated due to server problems.

(3) Diversification of uses

Another significant feature of big data resources is the diversification of uses. Big data resources can be applied in all aspects of production, life and enterprise management in modern society. In the production process, relying on big data resources, we can judge the market prospect and demand of products, adjust the production plan, improve the use efficiency of production funds, and obtain greater profits for enterprises.

3. PROBLEMS EXISTING IN BIG DATA AND THE NECESSITY OF MANAGEMENT

In the current allocation and management of big data resources in China, there are still three main problems: insufficient management technology, security guarantee defect and immature management system. These three problems restrict the use efficiency of big data resources and bring great security risks to the development and utilization of big data resources. In the process of big data resource allocation and management, we must first solve these problems.

3.1 Backward management technology

The deficiency of management technology mainly refers to the current domestic management technology of big data resources still has a lot of defects, on the one hand, there is no complete big data resource allocation and management technology system. Because big data resources are not provided by one platform or one enterprise, many enterprises are collecting and providing big data resources, and these enterprises have a variety of big data resource allocation and management technologies, not all enterprises have sufficient level of big data resource allocation and management technologies. Moreover, the allocation and management technology of big data resources can be divided into algorithm optimal solution, enterprise optimal solution and user optimal solution. Three different technology categories will bring different management results, but also have different defects. Therefore, if an enterprise chooses one of the three technologies, there will be deficiencies in management technology.

3.2 Management system is not mature enough

At present, the problem of big data resource allocation and management is still caused by the immature management system of big data resources. The domestic big data resource allocation and management lack of institutional management methods. There is neither a unified management standard and system for the industry, nor enough supervision and management departments to supervise and manage the allocation and utilization of big data resources, and there is also a lack of a unified technical system for the allocation and utilization of big data resources. Therefore, various big data resource suppliers choose distribution and management methods and technologies according to their own needs, and the results are bound to be uneven. It also causes a lot of problems in allocation and management of big data resources.

3.3 Improve the utilization efficiency of big data resources

The total amount of big data resources is very large, and the user of big data resources needs to choose the most suitable resources for their own use from these massive resources, which is itself a complex process, which needs to go through screening, testing, feedback and other links. If the big data resource allocation method is adopted, the big data resource center will also use big data technology to allocate resources to the most suitable demanders. To meet the needs of users of big data resources accurately and efficiently. Big data resources need to go through several processes from the

platform to the demand side, which is not a simple transmission. The use mode of big data resources under the distribution can greatly shorten the process and time of resources flowing to the demand side, so that big data resources can flow to the best demand objects in the shortest time. At the same time, the management of big data resources is also the key to improve its utilization efficiency. After resource allocation, if there is no follow-up management, there will be confusion of resources. Different demand parties may exchange big data resources with each other, but this lack of management makes the role of distribution greatly weakened. Enterprises still face the problem that big data resources are chaotic and cannot be quickly utilized. Therefore, follow-up management of big data resources must be carried out to ensure efficient utilization of big data by demanders and quickly meet their own needs by using big data resources.

3.4 Meet diversified needs of users

The allocation and management of big data resources also plays a positive role in meeting the diversified needs of Internet users. For Internet users, big data resources are an important reference basis for predicting their next behavior, and only by providing enterprises with more high-quality big data resources can enterprises carry out a higher level of push according to the needs of users. For example, if a user has browsed a large amount of laptop information on the Internet, the content pushed by big data must be computer. However, after the user has placed an order to buy a computer, it is a waste of big data resources to push relevant information. Instead, the subsequent push of computer accessories should be carried out to better meet the actual needs of users. The realization of this capability depends on the allocation and management of big data resources.

4. STRATEGIES FOR BIG DATA RESOURCE ALLOCATION AND MANAGEMENT

For the distribution and management of big data resources, it is necessary to start from the perspective of modern technology, adopt unified technical standards and management system of the industry to manage, and improve the management system, improve the security system, improve the utilization efficiency and security of big data resources, so that big data resources can play their due role.

4.1 Resource Scheduling and management platform

Mesos is a distributed management and scheduling system that can perform fine-grained scheduling control over distributed clusters. The Mesos project was created at UC Berkeley and is now a top Apache project. The Mesos system uses a two-level scheduling structure. The two-level scheduling is that the Mesos system allocates all resources to the computing framework, but the second-level scheduling is that the computing framework allocates resources to a specific task system through its own resource dispatcher.

Mesos consists of Mesos-Master, Mesos-Slave, Framework, and Executor. The Mesos-Master, on the other hand, is the center of the whole system, responsible for connecting all the Mesos design frameworks and slaves, and assigning all the data on the slaves to the Framework according to specific policies. The Mesos-slave transfers its resources to the Mesos-master, and after receiving instructions from the Mesos-master, allocates resources to the Framework. Framework refers to external computing architectures, such as Hado1op and Spark. These architectures have their own resource schedulers, which are used to allocate tasks within the framework. The Executor is primarily used to start tasks within the architecture.

Mesos is widely used by Twitter, e Bay, Apple, and many others because of its scalability, fault tolerance, and robustness, as well as its ability to give immediate feedback and adapt to different frameworks.

YARN is a universal resource management system. Developed from Hadoop's Map Reduce V1 architecture, YARN can manage and schedule upper-layer applications system-wide and supports resource management and data sharing through clusters. Before YARN appears, integration management is often integrated into MPv1, resulting in single point of failure and simpler computing architecture. In addition, this structure makes Job Tracker unable to more efficiently adjust Map Reduce work status on more than 1,000 hosts [8]. YARN was created to solve these problems. Its basic idea is to separate the two main functions of Job Tracker: resource management and job scheduling/monitoring. YARN consists of three modules: the global Resource Manager, the Node Manager of a node, and several application-specific Application masters. The Resource Manager is responsible for managing the resources of the entire cluster and managing resource allocation for applications. The Application Master is responsible for the scheduling and coordination of specific applications, and the Node Manager is responsible for the maintenance of each node.

4.2 Resource Scheduling and Management Standards

In the process of big data resource allocation and management, it is crucial to establish a unified standard for the industry. The big data resource industry is in chaos because of the lack of a unified standard for the industry. In order to gain profits and enhance their own industrial competitiveness, enterprises will choose to relax the management of big data resources, resulting in the leakage of big data resources. Companies are just turning a blind eye. Therefore, in the subsequent big data resource allocation and management process, unified standards must be established in the industry,

and enterprises in the industry must be strictly restricted in accordance with the standards. At the same time, it is also necessary to establish the supervision and management system of the starting industry and introduce the supervision and management of government departments to ensure the reasonable allocation and management of big data resources.

4.3 Big data resource configuration strategies based on cloud computing

Cloud computing technology is also an effective way to allocate big data resources. Cloud computing technology is the product of the current development of the Internet and computers, and has very superior performance. The use of cloud computing to transfer data to the cloud server for storage and distribution can effectively improve the allocation effect of big data analysis resources. In the process of using cloud computing technology to realize big data resource allocation, the core problem is the setting of the objective function of resource allocation. However, considering that in the big data center, because each node performs different tasks of big data center resource allocation, there is no balanced allocation, and some network nodes are allocated too many resource tasks. As a result, the nodes consume too much energy and thus fail to assign tasks or fail to function. To effectively prevent the occurrence of this phenomenon, the minimum objective function of the delay of the distribution node is set as min L, and formula (1) can be obtained:

$$\min \mathrm{L} = \sum_{i=1}^{M} A_i + \sum TR_{iJ} \text{ ..... (1)}$$

In Formula (1), M is the resource set of big data center. I is the number of resource allocations. T is big data center resource allocation time. R is the signal-to-noise ratio at the receiving end of big data resources. According to formula (1), the objective function of big data resource allocation is proposed.

The process of using cloud computing to realize big data resource allocation is as follows: start → set objective function → set boundary function through proportional fair algorithm → use cloud computing to find the optimal solution → end. Through this process, the optimal allocation of big data resources can be realized. It is also one of the main ways of big data resource allocation in the new period.

5. EXPERIMENT

This chapter mainly demonstrates the experiment of big data resource allocation. Through a series of tests and judgments on proportional fairness algorithm, the experimental results are intuitively displayed through test tables.

5.1 Configuration of the experimental environment

The hardware environment is Intel central processing unit (CPU) i7-12700K, motherboard Microstar MAG B660M, SSD WD Blue SN570 1TB, memory DDR4 3200 16G; Software environment is Windows 10 64-bit operating system, integrated development environment JetBrains PyCharm 2021.

5.2 Experimental process and results

Big data sources can be classified into labeled data and unlabeled data. The relevant scheduling data used in this experiment should be labeled data. One difference is that the labels are not all marked in advance, but need to be dynamically marked in the training process according to the chemical learning algorithm.

The input is composed of a number of two-dimensional matrices, each matrix needs to contain information about system resources, job size, job execution, etc., in the process of job scheduling.

There are 2 resources in the whole system, and the number of each resource is 3. At the same time, an input can observe 4 time steps. There are three job slots that can be directly called by the system. Each job slot records the number and execution time of various resources required by the job. When too many jobs are reached, it puts them in a backup queue. A batch of data generated according to this kind of strategy realizes the resource allocation of big data.

6. SUMMARY

To sum up, big data resources is one of the most important Internet resources in the Internet + era. It has exerted an important influence on Chinese social development in many aspects. To a great extent, the allocation and management of big data resources are determined to play the role of big data resources. However, with the increasingly perfect industry system and further breakthroughs in security technology, it is bound to realize the distribution and management of big data resources, ensure the security and effectiveness of the use of big data resources, and promote the further development of Chinese big data technology.

REFERENCE

- Minglei, F., & Xianya, G. (2021). Research on Resource Allocation Method of Big Data Center Based on Cloud Computing [J]. *Journal of Langfang Teachers University (Natural Science Edition)*, 21(2), 10-13.
- Liu, W. (2021). Big Data Information Resource Allocation Method Based on Nash Equilibrium Solution [J]. *Information and Computer (Theoretical Edition)*, 33(11), 180-182.
- Tongle, L., & Shanshan, W. (2021). Resource allocation Method of Big Data Center Based on Improved Genetic Algorithm [J]. *Yangtze River Information and Communication*, 34(3), 68-70.

- Zhang, H., & Huang, G. (2020). Distributive justice Challenge and Regulation Principle in the Era of Big Data [J]. *Journal of Shandong University of Science and Technology (Social Science Edition)*, 22(2), 22-27.
- Yi, B., Wang, X., Huang, M., & Zhao, Y. (2020). Novel resource allocation mechanism for SDN-based data center networks. *Journal of Network and Computer Applications*, *155*, 102554.
- Liuping, C. A. I., & Long, Y. (2019). Big Data Resource allocation Algorithm Based on Genetic Algorithm [J]. *Journal of Chongqing University of Technology (Natural Science)*, 33(4), 134-140.
- Wang, L., & Chen, Z. (2019). Parallel Distribution Optimization of Big Data Information Resources in Bottom Network [J]. *Journal of Computer Research and Development*, 36(2), 267-270.
- Zhang, C. (2015). Research on Key Technologies of Dynamic Allocation of Memory Resources in Map/Reduce Massive Data Processing Platform [D]. *Beijing: Beijing University of Technology*.
- Guoqing, Y., & Lanrong, Z. (2017). Research on Virtual Machine Resource Efficient Allocation in Big Data Cloud Center [J]. *Computer Measurement and Control*, 25(8), 272-274.
- Xiaonong, L. (2020). Research on Cloud Computing resource Allocation Based on Artificial technology. *Modern Electronic Technology*, 43(21), 178-181.
- Luo, P., & Wang, X. (2018). Research on the resource allocation algorithm based on forecasting in mobile cloud computing. *Journal of Intelligent & Fuzzy Systems*, *35*(2), 1315-1324.