Research on Distributed Storage Technology of Database Big Data Based on Cloud Computing

Duan Hongwei¹,*, Bi Ligetu²
¹Chifeng University, Inner Mongolia Chifeng 024000 China
²Chifeng Industrial Vocational Technical College, Inner Mongolia Chifeng 024000 China

*Corresponding author: 123569138@qq.com

Abstract. With China entering the era of big data, the storage and management of big data has become a hot issue. The proposal and use of distributed storage technology has achieved good results. The continuous development of Internet communication technology and the popularization of computer information technology promote the evolution of traditional data model. The distributed storage system for big data has strong expansibility. It can optimize the existing storage space, realize the optimal allocation of space resources, and reduce the cost of data storage. Cloud computing as a resource sharing business computing model, it receives a huge amount of data every day, and the amount of data is still expanding. This has caused great pressure on distributed storage in cloud computing environment. This paper studies the distributed storage technology of database big data based on cloud computing, in order to strengthen the data management in the era of big data and improve the security and utilization of data.

Keywords: Big data, database, cloud computing, distributed storage

1. Introduction
With the continuous development of network science and technology, the number of Internet users is increasing, so it has evolved into the era of big data explosion [1]. In the era of big data, data has the characteristics of high speed, diversity and large scale. In the context of a large number of data, the data capacity has exceeded the storage space, which easily leads to data management out of control [2]. From the current various industries and fields, we can know that computer technology has been widely applied to all kinds of data management, and realized the unified management of resources and data sharing [3]. In order to meet the needs of information management in the information age, distributed database came into being. In the era of traditional centralized database technology management, there are few information storage and information users, and the centralized database can meet the basic data management needs [4]. The application of database technology to big data can effectively deal with the above problems, realize the rapid integration and processing of data, and effectively improve the level of data processing [5]. The core feature of big data is massive data. The data capacity exceeds the
storage space of existing hardware. IT industry workers must ensure that the speed of data storage is consistent with the speed of growth, otherwise data management may be out of control [6]. In the context of the Internet, the increase of network data types and quantity puts forward strict requirements for data preservation and management, but the traditional data mode, that is, the operator builds a database system for data preservation and management [7].

With the increase of information users, the requirements for information storage and transmission management are constantly improving. In the aspect of information management, large capacity and efficient database technology are needed to ensure the high quality and efficient application of information [8]. Distributed database is opposite to centralized database. By combining multiple data storage units into the same database and storing them on different data storage nodes, we can realize super large capacity storage and access of massive data [9]. In the field of data sharing, the Chinese government will further develop and gradually use it, so in the period of continuous development of computer network technology, taking scientific measures is the premise to ensure the development of database technology [10]. As a resource sharing business computing model, cloud computing receives a huge amount of data every day, and the amount of data is still expanding [11]. This has caused great pressure on distributed storage in cloud computing environment. The distributed storage system based on big data has good scalability, which can maximize the use of existing storage space, realize the optimal allocation of resources, and effectively reduce the cost of data storage [12]. This paper studies the distributed storage technology of database big data based on cloud computing, in order to strengthen the data management in the era of big data and improve the security and utilization of data.

2. Application of distributed storage technology in the era of big data

2.1 Distributed storage technology of extreme data processing
The core advantage of extreme data distributed storage technology is its powerful data storage function. The main application is to process a large number of data, but its data mode operation is also very cumbersome. High quality data storage performance is the core of extreme data distributed storage technology, which is mainly used in data mode with large data processing capacity and complicated operation. Due to the influence of the traditional database system on the modern database from the perspective of big data, various types of databases will appear incompatible in the process of using. This kind of reason will lead to the information resources in the database can not be effectively used, so that the resources are wasted or relatively idle, but also can not make the user's needs to achieve the ideal state. Distributed storage technology in cloud computing environment is mainly divided into two forms: centralization and decentralization, both of which have their own advantages and disadvantages [13]. For example, the advantage of centralized structure is simple structure, simple operation, but prone to single point of failure, while decentralized structure will not appear single point of failure. Relatively, its operation is much more complex than centralized structure.

Under the function of distributed database, it can realize the storage and management of mixed data. Mixed data storage has obvious advantages in data query, statistics and management, and has been widely used in big data. The flow of data mining technology in cloud computing database is shown in Figure 1.

The purpose of the application of distributed database technology in large databases is to solve the problem that the traditional centralized database technology needs insufficient data storage and management analysis. Distributed object storage is an access unit, which can read and write objects under the network protocol. The core idea is to separate data and control channels, and build a storage system based on the use of object storage equipment. Through the application of distributed database technology, data collection and collation are realized with the cooperation of multiple units, and the collected data are settled and saved [14]. In today's Internet age, the data is growing rapidly and the number of data users is increasing. Under this background, the traditional centralized database technology is difficult to meet the scattered, diversified and massive data management. Cloud computing is a new commercial computer model with resource sharing, which has strong
comprehensive ability, including cluster technology, virtualization technology, parallel computing, utility computing and information management technology.

Fig. 1 Data mining technology process

2.2 Distributed object storage technology

Usually, the distributed database itself has the function of dynamic expansion, and under the multi-unit structure, it is combined with the network platform to coordinate with the complete commands, so that the data can be transmitted smoothly. Because each node can only apply local resources and cannot access other resources. Object storage can provide an abstract interface, and it can support the storage of any object in the range of bytes to several terabytes, and the business can divide data more flexibly. The cost of object storage is very low, because object storage is based on standard hardware facilities and only needs ordinary servers and hard disks. Distributed database itself has intelligent index function, that is, under the guidance of database coarse-grained index, data packets are automatically formed during data download, so as to achieve the effect of data collection and screening. Distributed database technology shows that the data storage function is distributed on different computer network nodes. Based on the traditional centralized database management technology, it also has the functions of data distribution acquisition and calculation, and constructs a distributed relational framework. Traditional storage technology has great drawbacks. For example, when the data is viewed heavily, it will bring great load to the system, reduce the network speed, reduce the utilization rate of broadband, and lead to the waste of network resources.

Distributed storage in cloud computing no longer relies on client/server structure to provide services, but uses new computer technologies such as virtualization and load balancing to serve users. The network infrastructure of distributed storage system is shown in Figure 2.

Fig. 2 Network infrastructure
Big data technology has become the key to analyze and deal with problems. However, the traditional centralized database processing technology can not meet the needs of data growth, and can not meet the requirements of efficient management of big data. Distributed database technology has the advantages of high scalability, high concurrency and high availability in the era of big data. Its application in big data technology can make up for the low efficiency of data management and analysis, low information storage function, low reading and writing function and low operational reliability of traditional centralized database technology in the era of big data. Driven by intelligent indexing, it doesn't need to occupy a lot of space, and its expansion ability is ideal. It can ensure that it won't be affected by various factors in the process of data indexing, and its response is timely and fast. The most fundamental advantage of customized distributed storage technology is that it can provide personalized storage services according to the actual needs of the system, meet the actual needs of software, and make adaptive improvements in the system, which perfectly meets the diversified changes of data [15]. The high scalability of distributed database technology provides more possibilities for the linear expansion of information capacity. The expansion of data storage can be realized by adding data storage nodes, which is suitable for the development background of information explosion in the era of big data. The application of distributed database technology in big data is of great significance for improving the efficiency of big data management and analysis and ensuring the safety, stability and reliability of data in the process of data storage, reading and transportation. It is the driving force for the rapid development of big data technology, and has a milestone significance for promoting the application of big data technology.

3. Key technology structure of distributed storage in cloud computing environment

Each computer system that constitutes a distributed database system can be stored in one place independently, and each computer can store and manage various data information in one database. It is a very traditional method to use the switch as the data center. Before the distributed storage technology in cloud computing environment appeared, the switch was the core structure of the traditional storage technology. Distributed database can make all kinds of data be transmitted quickly, and in the process of data transmission, realize the supervision of data transmission, find out existing problems and deal with them in time, and ensure the safety of data transmission. In distributed database, external connection is usually used to control and manage the overall situation of the database. On the premise of internet equipment, the server can also help the switch to process, forward and store data packets. In this way, the advantages of server as data center and switch as data center can be brought into full play. In the remote database system, the database integration is realized by using the connection technology, and the database units distributed in various positions are collected and processed to form a whole.

![Database encryption module system structure](image)

**Fig.3** Database encryption module system structure

In the distributed database system, it has the function of supporting hybrid network expansion, that
is, under the action of hybrid network protocol, it realizes the conversion and processing of network protocol. In the distributed database system, all kinds of data can be managed and saved uniformly to form an advanced data cluster, which can ensure the security and integrity of the internal information of the database and provide support for the subsequent user data application [16]. The system structure of database encryption module is shown in Figure 3.

In the process of database application, distributed database is like a complete database service platform, and users use this platform to realize data access and query. In order to solve the technical problems of database in the field of big data, the measures taken are to establish unified database management for different types of databases. At the same time, the retrieval and management of data and information resources need to be improved effectively, so that customers will be satisfied gradually and the work efficiency will become more efficient, and the operation process will be simplified in the process. The security performance of database technical information itself is more important, so the database technical security should be managed at all times. First, strengthen the construction of the center of the database, and optimize the anti-invasion ability of the database in terms of the terminal, so as to improve the security of the whole database technology. Then, the database with key information storage can adopt the function of automatic identification and storage to implement the security plan. The value of database technology in actual use depends on sharing the data and information resources of the database itself. Usually, users can quickly and conveniently obtain the resources they want through the computer network platform, and gradually enhance the efficiency of information use.

There are many high-availability clusters in the database, which is mainly because the data information in the same database is transferred to other platforms to provide more database instance information under the action of cache interaction technology. The global structure is shown in Figure 4.

Combining the characteristics of multimedia data and information, we can solve the problems in the introduction of multimedia technology into database technology in time to ensure the accuracy and timeliness of information, and at the same time greatly improve the security of data. In the global transaction concurrency control technology, the cache replication technology is used to copy the information of each node, so that each node should be unified with dictionary data, and all transaction information can be viewed, so that it can give full play to its own functions. The absolute security of database information can't be guaranteed only through the database security system. On this basis, important data of users should be kept confidential to ensure the data security. Finally, a system security protection system should be established. In order to meet the special needs of some users, the full integration of object-oriented method and database information can effectively guarantee the flexible use of data and the security of application technology in the process of data transmission. With the continuous development of computer technology, database technology is gradually innovating, and the means of object-oriented technology are constantly integrated into the corresponding database technology.

4. Conclusions
With the advent of the era of big data, strict standards are put forward for data processing. It is very
necessary to update and apply database technology in order to better meet users' own requirements. In this era of explosive data growth, the birth of cloud computing is in line with the trend of the times. Cloud computing has received wide attention because of its large storage space and strong data processing capability, and thus has been rapidly popularized. In the era of big data, distributed database technology meets the needs of rapid growth and advanced management of big data, well adapts to the needs of multi-type, quantitative and decentralized management of big data, and solves many problems faced by traditional centralized database management technology. The application of distributed database technology in big data can realize the classification and processing of big data information, help each user realize data analysis and processing, and effectively solve the problems existing in the data application process. The application of distributed database technology in big data expansion, load support, data loading and business management explains its capabilities and advantages in information management in the era of big data, and provides more management opinions for future big data management.

References
[1] Wang Shanshan. Research on High Security Distributed Storage System. Electronic World, vol. 604, no. 22, pp. 196-197, 2020.
[2] Hou Guiyun, Chen Guiying, Lu Zhiqiang. Research on Distributed Data Storage Simulation in Cloud Computing. Computer Simulation, vol. 35, no. 7, pp. 330-334, 2018.
[3] Li Hao, Fan Penghua. Analysis on key technologies of distributed storage in cloud computing environment. Electronic World, vol. 578, no. 20, pp. 189-190, 2019.
[4] Wu Yanbo, Xue Qin, Xiang Dawei, et al. NoSQL distributed big data storage technology and application under the cloud platform. Modern Electronic Technology, vol. 464, no. 9, pp. 52-55+60, 2016.
[5] Wang Linghui, Xie Yunyue, Zhou Meihua. Performance analysis of Hadoop distributed storage architecture. Modern Electronic Technology, vol. 41, no. 18, pp. 92-95, 2018.
[6] Gu Qin. Research on Distributed Storage Technology for Big Data. Wireless Internet Technology, vol. 187, no. 15, pp. 159-160, 2020.
[7] Wang Liqing, Zhang Huamin. Research on Redundancy of Big Data Distributed Storage. Computer Knowledge and Technology, vol. 12, no. 19, pp. 1-2, 2016.
[8] Zhang Chi. Privacy protection mechanism based on big data distributed storage logistics personal information. Computer knowledge and technology, vol. 16, no. 8, pp. 292-294, 2020.
[9] Xie Mengyi. Design of distributed big data asynchronous migration system under hybrid cloud storage architecture. Electronic Design Engineering, vol. 27, no. 23, pp. 45-49, 2019.
[10] Wu Yanbo, Xue Qin, Xiang Dawei, et al. NoSQL distributed big data storage technology and application under the cloud platform. Modern Electronic Technology, vol. 39, no. 9, pp. 44-47, 2016.
[11] Geng Jiajie. New energy big data system platform construction practice. Electronic World, vol. 544, no. 10, pp. 40-41, 2018.
[12] Tang Guwen, Han Jiafu, Li Hongsheng. Distributed storage strategy for spatial big data. Computer Technology and Development, vol. 29, no. 3, pp. 194-197, 2019.
[13] Yang Chengyue, Song Qianyun, Xiang Longgang, et al. Research on global spatiotemporal big data organization and management based on non-relational database. Geographic Information World, vol. 26, no. 6, pp. 116-122, 2019.
[14] Wang Jun, Yu Airong, Niu Yanjie, et al. Research on New Distributed Big Data Storage Technology in Web Applications. Research Information Technology and Application, vol. 9, no. 3, pp. 20-25, 2018.
[15] Chen Yuan, Hu Hang, Yue Shihua. Research on storage technology of distributed big data platform for building information model based on IFC standard. Computer Application and Software, vol. 36, no. 2, pp. 131-136+179, 2019.
[16] Wang Xiaoting, Sun Liping, Yang Jingchao. Analysis of distributed storage strategy under spatial big data. Computer and Information Technology, vol. 27, no. 4, pp. 72-74, 2019.