

Research on large scale data processing technology based on cloud computing

Wen Qiuhua

Academy of Overseas Chinese Studies In Jinan University, Guangzhou 510632, PR China

qiuhuawen80@sina.com

Key words: fusion, cloud computing, big data processing, deployment strategy, fusion processing

Abstract. To improve the single problem in large data processing system based on cloud computing technology, the research focuses on the impact of network data processing model based on hybrid architecture, fusion processing, and decentralized processing. This research can be used in the next generation of cloud platform processing system, which can provide a solution for the future cloud platform processing technology.

Introduction

In broadband Internet access technology and intelligent terminal of the rapid popularization of today, the network data capacity and the processing of data volume / data intensity increased significantly faster than any period of time; the big data era has quietly come. Big data era is accompanied by the development of cloud technology, how to carry out security, effective and accurate data processing in the cloud environment, has become a hot area of research.

The people of data to deal with the problem of research, mainly based on cloud environment of large-scale distributed deployment, the architecture of network integration depth and dynamic real-time data processing the three cloud technology research direction of expansion, also made some achievements, the problems under the environment of the cloud data processing solution is also beneficial. However, the complexity of the cloud environment and the problems are often intertwined, the research process also encountered great difficulties, mainly in the following 3 aspects: (1) the data structure is very complex, processing methods are not unified, the need for parallel processing; (2) the processing of large data, in order to improve the system's real-time and processing resources security, effectiveness, accuracy; (3) system

The flexibility, which needs to be transparent to the developers and users, and the existing system, only provide limited interface for developers and users to enjoy, which also indirectly affect the performance of data processing.

In this paper, a new model of large data processing is proposed, which is based on the idea of technology integration. This paper attempts to provide a new method for solving large data processing problems in cloud environment.

Large data integration scheme based on Fusion Theory

In cloud computing environment, human computer interaction, data processing and network processing technology are always in a state of deep integration. The starting point of the fusion idea is to organize the network resources in the dispersed state in the cloud environment to organize in a cooperative way, and then make the integration, and give full play to the comparative advantage of the resource of various decentralized state. As an organic integration of various comparative

advantages, the integration of the integration scheme has broad application prospects.

Under the cloud computing environment, the research of big data processing mainly focuses on the construction of large data processing system, the cooperative mode of distributed data and the key technology. In macro terms, the above mentioned points can be divided into 2 aspects: the mixed treatment and the mixed management:

The research points of hybrid management are mainly based on the processing mechanism, data sharing mechanism, resource sharing mechanism, distributed data management mechanism, processing synchronization mechanism, etc., and hybrid processing mainly focus on the system operation model and related supporting technology.

Fusion theory is a kind of thinking of solving the problems related to cloud computing, and it is a model of large data processing model, which mainly studies the cooperation of big data and the data processing. It involves how to carry out the work among the members, how to carry out the interaction, operation coordination, collaboration and cooperation between the members, the main session model, the meeting model, process model, activity model and hierarchical model.

Session model is the basis of the communication and coordination between the various factors, and through the implementation of a specific speech act (such as the request, promised, etc.). The meeting model is another basic form of the group to work together. The participants communicate with each other through the shared information space. The process model divides the complex tasks into small steps, the activity model is further extended to the process model, and the task is divided into explicit sub task, defining the relationship between the sub task and the task of the task. In general to solve large data problem can only deal with the composition of integration and coordination between factor to solve; to adopt a variety of models of hybrid, through the integration of ideas to be integrated, often must be by a dicing, graded approach is described.

Application analysis of large data processing

In cloud environments, there may be a huge difference in the structure and size of the large data in the cloud environment. After the abstract, the basic elements of the processing framework include: processing the members' rights, public processing resources, processing events and processing activities. Member disposed jurisdiction describes a data processing flow within the system operation and authority to deal with, according to the processing needs of data flow confluence of scribing and grading to give the appropriate permissions; common processing resources describes the cloud environment the Ziyun the shared processing channel, signaling and command processing mechanism; processing activities described by treatment in the process of transmitting, processing and receiving of all the dynamic behavior of the overall features; event processing describes the processing activities in the progress, the index set for coordination between the various factors.

For the application, the big data processing mainly focus on the application system construction, the current main solution ideas around the following 3 points:

(1) the integrated architecture is actually a client / server architecture. The server is responsible for the management, control and scheduling of the whole application system, and all of the application logic, data, etc., the client refers to the user's input and output. When the user wants to perform the relevant tasks, the need to send a request to the server through the customer point, the server is completed. Collaborative application framework for the integration of simple, easy to maintain, but the application of the cooperative application of a strong dependence on the server, it is easy to become the bottleneck of the whole system.

(2) distributed architecture, the nodes are in the same position, and the control and management modules are distributed in the client. The client can achieve a better system of autonomy, its

robustness, flexibility, openness and versatility are better, but the distribution of data storage and operation of the respective processing, making the global consistency is difficult to maintain, real-time synchronization between nodes and user's dynamic registration is also difficult to achieve.

(3) the hybrid structure combines the advantages of the first two structures, and the data information is maintained by the server, the user management and the information are transmitted and distributed. This structure makes full use of the computer resources, greatly reducing the server's work pressure, and reducing the "bottleneck" effect, which can make the system have good flexibility, robustness and short response time, and guarantee the consistency of the system information.

Research on large data processing system based on Cloud Computing

Large data processing platform based on cloud computing platform has three characteristics, which are nodes distributed, dynamic data processing and data source. Combined with the large data processing model, it is necessary to construct a large data processing platform, which is based on the construction of data processing platform, which is based on the elements, public processing resources, processing events and processing activities.

The core idea of data processing system is to deal with the data of dispersed state. The key factor is to extract and distribute data resources. First, we need to provide interface for external application, data management part provides the application data interface to the data resources, and then the fusion of data resources, in a certain degree of tolerance, the integration of the comparative advantages of the data processing mechanism. In addition, the data processing center of the data processing center should adopt the centralized mode, the unified distribution of data resources, thus achieving the integration of data processing.

The author from the large data processing system in a certain information industry park, combined with the demand of information enterprise cluster, the existing scattered data resources for mining (mainly ERP, SCM system), for the next step in the development of industrial park to provide strategic advice.

This paper proposes the system deployment using the integration of ideas, combined with the cloud environment of data fusion processing and data fusion management, to achieve the big data resource management, system deployment architecture is as follows: data source is the ERP, SCM and other application data system, after the server into a distributed n database, then merged into the big data management module, and finally provide data query, data decision-making and other services.

Conclusion

In the research of large data processing system based on the current cloud environment, the research on the large data processing system is very little, especially the information industry, the information industry is not realized. In this paper, we analyze the key and difficult problems of large data processing in cloud environment, and then analyze the large data processing, especially the impact of the system architecture on big data processing.

Acknowledgements:

This is subject of the overseas Chinese Affairs Office of the State Council "wisdom of the overseas Chinese Affairs" with the No.gqbw2015002 and the "Overseas Chinese high level talents for the investigation conditions of service for the country " with the No.gqbw2015001 .

References

- [1] Sang Woo Han, Jong Won. A multi-agent-based management system for pervasive collaborative computing environment[C]. IEEE international conference on computing and communications. USA: Institute of electrical and electronics engineers, 2009:1-6.
- [2] Han Xu, Cao Yongcun. The application of computer-supported collaborative technologies in web-based teachers professional development system[C]. The 5th international conference on computer science and education. USA: Institute of electrical and electronics engineers, 2010,30(9):842-845.
- [3] Han J, Kamber M. Data mining concepts and techniques[M]. San francisco: Morgan kaufmann, 2006. Massachusetts: MIT press, 2004:191-211.