Research on the Optimal Deployment and Scheduling Strategy of Big Data Application in Cloud Computing platform

Lu Lihua

Northwestern Polytechnical University, Shaanxi, China,710129

Abstract: With the progress and development of science and technology and society, all walks of life in China have started to actively apply big data. Big data has become the most important part and plays an indispensable role in the development of today's society. In recent years, in the application of big data based on cloud computing platform, the flexibility of big data has been fully reflected in the resource allocation. At the same time, in the actual application of big data, cloud computing platform can make corresponding reflection based on the requirements of big data for relevant platforms. Moreover, the charging behavior based on cloud service will be carried out in accordance with the actual demand. For users in today's era, this is completely in line with their consumption pattern, especially for small and medium-sized users in today's market. The advantages and characteristics of this applicable business are particularly prominent. However, in the application process of big data technology based on cloud computing platform, there are still some problems that need our attention. This paper makes an in-depth discussion on the performance of big data in the cloud computing platform to make a brief analysis on the optimization deployment and scheduling strategy of big data application in the cloud computing platform.

1. Big data and cloud computing

The 21st century is an era of information. With the wide application of Internet in all walks of life, e-commerce, social network and other Internet technologies have also been widely used, among which the most important application is data processing. Many diversified applications in today's market make the scale and types of today's data keep rising, so it is necessary to improve the data analysis and processing technology based on the current situation. The application of big data processing framework has effectively improved the quality and efficiency of data processing and reduced the difficulty and pressure of data processing. However, in the actual application of these data processing frameworks, there are very high requirements for relevant software and hardware and we must adopt a computer cluster with a certain scale as effective support. Therefore, if the user scale is small and the resources are small, it will inevitably increase the difficulty of work.

Based on this situation, cloud computing platform is provided for these users in big data processing, which has obvious characteristics of economy and applicability. The main function of cloud computing platform is to take service as the core and provide computing resources and functional services for the majority of users so that users can browse and pay according to their actual needs to solve the problems caused by resource and hardware limitations during the process of accessing the platform.

In addition, in the application of cloud computing platform, the function of big data application migration can be realized to effectively make up for the defects in the application of traditional big data:

In terms of technology, cloud computing platform is mainly based on virtualization technology. First, the limitation of hardware and software in big data processing can be effectively reduced. Secondly, the resources of the underlying platform can be coupled with the application services of

the upper layer, which can effectively improve the flexibility of resource scheduling and better meet the needs of customers.

In terms of business model, cloud computing platform can conduct centralized processing of big data and as a third party, cloud computing platform can provide more professional management, maintenance, service and other work. When charging fees based on platform service projects, users do not need to maintain and manage big data applications by themselves, but only need to choose the services provided by the platform, which not only effectively reduces the cost of users in the use process, but also greatly expands the range of available resources for users.

2. Study on the optimized deployment and scheduling strategy of big data application in cloud computing platform

2.1 Resource layer optimization deployment strategy research

In the actual process of resource layer optimization, the content of optimization is mainly the performance of virtual computing resources in the cloud computing platform. However, due to the obvious differences in the objectives to be optimized, there are also differences in the optimized deployment. We can divide our deployment optimization efforts into two categories. The first one is to optimize the communication performance of virtual cluster and improve the efficiency and quality of data transmission between virtual machines.

The second type is to optimize the computing performance of the deployment of virtual machines. The main work of this optimization deployment is to improve the data processing ability of virtual machines.

2.1.1 Optimize deployment resource layer communication performance

Optimizing the communication performance of resource layer, that is, optimizing the data communication performance mainly aims at the configuration of relevant network resources in the cloud computing platform and optimizes the computing nodes in it to achieve the virtual cluster communication performance of the platform. The main content of optimization is the data transferred between virtual machines while the communication time between virtual machines in traditional resource layer allocation is not in the main content of research.

2.1.2 Resource layer computing performance optimization deployment

This paper focuses on the optimization deployment strategy of resource layer computing performance. In the process of research, there are obvious differences based on big data operational framework and big data application and the characteristics are very obvious. In the actual optimization process, the parameters of virtual machine optimization configuration, location affiliation and other aspects are quite different, so the cloud computing platform must be taken as the basis for the corresponding services (as shown in Figure 1).
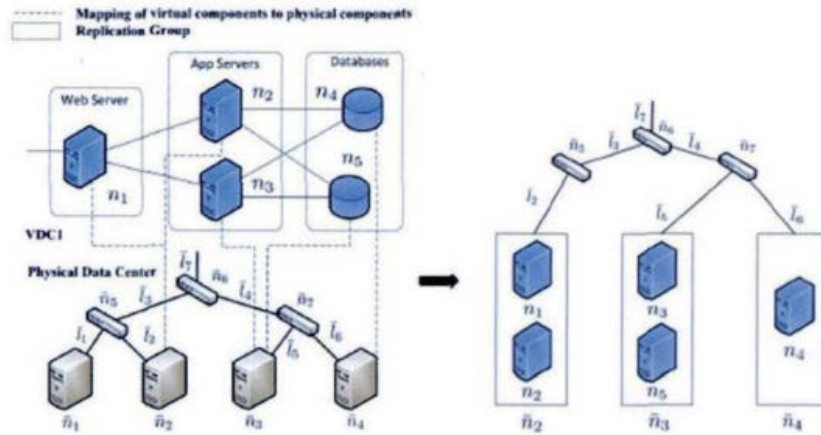
Figure 1 Local compute node application deployment

After detailed analysis, the main optimization content is virtual machine allocation, data node location and other parameters. After optimizing these parameters, the goal of optimizing resource layer can be finally achieved. However, there are still some defects in the optimized deployment.

First, although the optimized deployment of the resource layer has been clarified, there is still the problem of fragmentation. In the actual optimized deployment process, too much emphasis is placed on the computing and communication performance contained in the computer cluster, but the impact on the overall optimized deployment strategy is neglected, for example, whether it can lead to a reduction in the efficiency of data processing, transmission requirements, etc. However, if the application optimization of big data in cloud computing platform is analyzed, the optimization content does not have regional characteristics. Therefore, in the process of overall performance optimization, if partial optimization deployment is carried out, it will not have a good applicability and will also generate a large cost.

2.2 Platform layer optimization deployment strategy

In the process of optimizing the deployment of the platform layer, the focus of the work is mainly to deploy the corresponding software application image under the cloud computing platform and allocate virtual resources based on the deployment of the resource layer to form a virtual platform that can perform specific tasks. In addition, in the process of optimizing the deployment platform layer, the core content of the research is the optimization deployment strategy of application copy. However, the difference of platform deployment mode leads to the optimization deployment strategy of the actual platform layer divided into two main directions:

2.2.1 Local compute node optimization deployment strategy

First, the application deployment based on the local computing node must integrate the application copy directly in the virtual image, promote the corresponding application copy to work with the virtual machine and then carry out the corresponding jobs. In addition, it is necessary to optimize the deployment according to the logical architecture of the software application, but application components will inevitably have some differences, which will lead to differences in the optimized deployment of the platform. Ultimately, improving the optimization effect can improve its reliability.

2.2.2 Storage platform optimization deployment strategy

In the process of formulating the optimized deployment strategy for storage platform, it is necessary to take the cloud computing platform under different constructions owned by independent replicas as the main optimization content and carry out corresponding optimization work based on the deployment mode. In the actual running process of virtual machine, if the application is the same, it can perform repetitive execution. It just needs to be implemented by hanging the application copy in the virtual machine so that its flexibility in execution can be greatly improved and its efficiency can be enhanced in the process of virtual machine.

Conclusion

At present, big data based on cloud computing platform is widely applied in various social fields, but there are still some problems to be solved. Based on the application of big data in cloud computing platform optimization, this paper first discusses the application of cloud computing in the platform resource layer and platform layer and then puts forward corresponding strategies and suggestion, hoping that it can provide relevant reference for the development of big data-related application optimization in China.

References

[1] Luo Gongyin, Zhou Chengcheng, Dong Zhongzhong, Zhang Mingzhao,Li Fanhua. Research on the framework of power grid cloud computing system based on Spark [J]. New Communication in China,2018,20(01):234-237.

[2] Zhang Bin. Zibo hydrological cloud computing platform and big data application [A]. Hohai university, China water conservancy society. Exploring "smart water conservancy" to promote scientific and technological innovation -- 2017 (the fifth session) Chinese water conservancy information technology BBS proceedings [C]. Hohai University, China water conservancy society: Beijing water consulting co., LTD,2017:8.

[3] Lin Bo. Xinhua global cloud computing platform standards and practices [A]. China journalism and technology workers' federation. 2015 China journalism and technology workers' federation excellent collection of papers on journalism and technology [C]. China federation of journalists and technicians: China federation of journalists and technicians,2015:7.