Status of Cloud Computing in Data Intensive Data Center

Dalin Zhao
Liaoning Jianzhu Vocational University, Liaoning, China

*Corresponding author e-mail: 183187172@lnjz.edu.cn

Abstract. With the advent of big data era, data intensive computing has been widely concerned by the industry, but the research on data intensive center is still in its infancy. With the development of science and technology, the emergence of cloud computing has brought great convenience to people, and has become a research hotspot for a time. This paper analyzes the application status of cloud computing in data intensive data center. In the research, we analyzed the application status of cloud computing in data intensive data center. Taking data decision-making as an example, we studied the application effect of cloud computing in data intensive data center by analyzing the influence of different number of data sets and data centers on the time cost of data transmission. The results of this paper show that by 2019, the corresponding rate of cloud computing in data decision-making, data security and data management has reached 88%, 92% and 95% respectively. Moreover, the research results show that the application of cloud computing in data intensive data center can effectively reduce the time cost of data transmission and achieve good application effect.

Keywords: Cloud Computing; Data Intensive Computing; Data Center; Data Transmission Time Overhead

1. Introduction
With the rapid development of mobile Internet, Internet of things, e-commerce and social media, we have entered the era of big data. The scientific community and other sectors of society all feel the huge impact of massive data [1-2]. Cloud computing is a new computing mode, which allows users to configure and access the computing resource pool independently according to their own needs and environment. Cloud computing realizes people's desire for large-scale data processing and storage in a low-cost and efficient way [3-4].

Data in the data intensive environment has the characteristics of massive, high-speed change, distribution, heterogeneous, semi-structured or unstructured. The data volume is generally TB or even Pb level. The traditional data storage and indexing technology in relational database management system can not fully describe these characteristics [5-6]. On the other hand, massive distributed data also brings great difficulties to the traditional database storage. The high-speed changing data makes the data management technology which can only handle static data or data snapshot powerless. The emergence of cloud computing technology provides a new method for the development of data intensive applications [7-8]. Analyzing the application status and application effect of cloud
computing in data intensive data center is of great significance for the application of cloud computing in data intensive data center [9-10].

In this paper, firstly, cloud computing and data intensive data center are briefly described, then the application status of cloud computing in data intensive data center is analyzed, and taking data decision as an example, the application effect of cloud computing in data intensive data center is studied. The research results of this paper show that the application rate of cloud computing in data intensive data center is higher and higher, and has achieved good application effect.

2. Cloud Computing and Data Intensive Data Center

2.1. Cloud Computing

Cloud computing is to centralize the computing resources of cloud providers and open the services to users by means of user payment. Users can enjoy the same computing services as large enterprises with less cost.

Cloud computing has several remarkable characteristics:

(1) Large scale

Cloud computing not only needs to provide support for different types of business at the same time, but also needs to provide support for massive cloud computing users in each specific task.

(2) Isomerism

Cloud computing can not only be compatible with different basic resources, including database, stand-alone operating system and other software infrastructure resources, as well as the hardware infrastructure resources with network, computing and storage as the main equipment in network environment, but also provide service support for specific business needs defined by users based on their own needs.

(3) Virtualization

Virtual technology is an important part of cloud computing technology. Through the use of virtualization technology, users can obtain the resources they need from cloud computing without understanding the process of service provision and mastering the operation ability of equipment only relying on the use of Internet.

(4) Economic applicability

In cloud computing, massive and cheap server resources are gathered together to form a cloud data center for users to use through Internet technology, which can reduce the user's cost and ensure high performance at the same time with low cost.

2.2. Data Intensive Data Center

The term data intensive comes from "data intensive computing", which refers to a computing method for massive data. In the data intensive data center, set the data center set as:

$$DC = \bigcup_{i=1,2,...} \{dc_i\}$$

Then the unit data transmission cost matrix between data centers in DC is as follows:

$$DataCenterCost = \begin{bmatrix}
dcc_{11} & dcc_{12} & \ldots & dcc_{1|PC|} \\
dcc_{21} & dcc_{22} & \ldots & dcc_{2|PC|} \\
\vdots & \vdots & \ddots & \vdots \\
dcc_{|PC|1} & dcc_{|PC|2} & \ldots & dcc_{|PC||PC|}
\end{bmatrix}$$

Define the collection of data intensive application workflow as follows:

$$PS = \{ps_1, ps_2, ps_3, \ldots\}$$
Then the unit transmission cost cost from data set to each data center, and the unit transmission cost cost matrix between data set and each data center is as follows:

$$DataToCenterCost = \begin{bmatrix}
    d_{ccc1} & d_{ccc2} & \ldots & d_{ccc_{|PC|}} \\
    d_{ccc1} & d_{ccc2} & \ldots & d_{ccc_{|PC|}} \\
    \vdots & \vdots & \ddots & \vdots \\
    d_{ccc_{|oSt|}} & d_{ccc_{|oSt|2}} & \ldots & d_{ccc_{|PC|_{|oSt|}}}
\end{bmatrix}$$

(4)

Generally speaking, data intensive has the following essential elements: (1) the research object is big data; (2) the source of data channel is computer generated through simulation experiments or captured by instruments; (3) data processing tools are all kinds of software; (4) data storage tools are mainly computers; (5) The acquisition of knowledge theory is to draw new conclusions or theories based on data without theoretical premise and hypothesis.

3. Research Methods and Design

3.1. Research Methods

(1) Literature research method

Firstly, through the method of literature analysis, this paper collects relevant domestic and foreign literature according to the topic selection, and summarizes and sorts out the core points as the theoretical basis of this study.

(2) Case law

In this paper, cloud computing is applied to data intensive data center, and its data decision-making is taken as an example.

3.2. Experimental Design

(1) Data sources

In this paper, the data used in the analysis of the application status of cloud computing in the data intensive data center are all from the data published by the relevant national authoritative websites.

(2) Application analysis

In this paper, cloud computing is applied to the data decision-making of data intensive data centers. The advantages of cloud computing are highlighted by analyzing the impact of the number of data sets and the number of data centers on the time cost of data transmission.

1) Experimental environment

The experiment environment is: Inter(R) Core(TM)2 Duo2.93GHz, RAM 3GB, hard disk 320GB, 100MB network bandwidth.

2) Experiment set settings

In this experiment, we set the number of data sets as 20, 40, 60, 80, and 100, and the number of data centers is set to 5, 10, 15, 20, 25 and 30, respectively.

4. Analysis and Discussion of Research Results

The data of data intensive data center has three characteristics, mainly in the following aspects:

(1) Magnanimity

The data scale of such applications has reached terabyte (1024gb) or even petabyte (1000tb) level, so it needs to provide efficient computing power and sufficient storage capacity.

(2) Distribution

Because people all over the world cooperate and share the data resources of such applications, the data of such applications may be distributed in different geographical regions.

(3) Relevance
There may be correlation between the data of such applications. At present, data intensive applications are deployed in the grid system, because the grid system can provide it with efficient computing power and sufficient storage capacity, but the establishment of a grid system needs high cost, and the grid system can not be shared by people all over the world.

The emergence of cloud computing technology provides a new method for the development of data intensive applications.

Some characteristics of cloud computing system also meet the requirements of data intensive applications

1) Like grid system, cloud computing system can provide efficient computing power and enough storage capacity for data intensive applications. However, different from grid system, cloud computing is cheaper than grid system because it is composed of data packets composed of many hardware clusters.

2) Cloud computing system based on Internet has created a new mode, so that people all over the world can cooperate and share data resources. Therefore, when data intensive applications are deployed in the "cloud", people can download and publish the data of such applications all over the world, that is, to share the data.

Through cloud computing technology, data intensive applications can get more extensive use rights. Therefore, this paper analyzes the application status and application effect of cloud computing in data intensive data center.

4.1. Application Status of Cloud Computing in Data Intensive Data Center

With the development of cloud computing, cloud computing is more and more widely used in data intensive data centers. Now, we analyze the application status of cloud computing in data intensive data centers in recent five years, and the results are shown in Figure 1.

![Figure 1. Analysis of cloud computing application in data intensive data center](image)

As can be seen from Figure 1, the application of cloud computing in data intensive data center is very extensive. Generally speaking, the application of cloud computing in data intensive data center mainly includes three aspects: data decision-making, data security and data management. We can see that with the passage of time, the application rate of cloud computing in data intensive data centers is getting higher and higher. By 2019, the application rate of cloud computing in data decision-making, data security and data management will reach 88%, 92% and 95% respectively. Thus, cloud computing is becoming more and more popular in data intensive data centers.
4.2. Application Analysis of Cloud Computing in Data Intensive Data Center

After this study, we have known that cloud computing is more and more used in data intensive data centers. How about the application effect of cloud computing in data intensive data centers? We will analyze this. In the analysis, we take the data decision of cloud computing in data intensive data center as an example.

4.2.1. Analysis of the impact of data set changes on time overhead

In the study, we analyze the impact of data set changes on the time cost, and the results are shown in Figure 2.

![Figure 2. Analysis of the impact of data set changes on time overhead](image)

It can be seen from Figure 2 that when cloud computing is applied in data intensive data centers, the time cost of data transmission will be reduced. When the number of data sets is 20, the time cost of data transmission with cloud computing is 60s, and that of data transmission without cloud computing is 100s. When the number of data sets is 40, the time cost of data transmission with cloud computing is 80s, and that of data transmission without cloud computing is 220s. When the number of data sets is 60, the time cost of data transmission using cloud computing is 120s, and that of data transmission without cloud computing is 310s. When the number of data sets is 80, the time cost of data transmission with cloud computing is 260s, and that of data transmission without cloud computing is 400s. When the number of data sets is 100, the time cost of data transmission with cloud computing is 290s, and that of data transmission without cloud computing is 480s. It can be seen that with the increase of the number of data sets, the time cost of data transmission is also increasing, and the application of cloud computing to it can effectively reduce the time cost.

4.2.2. The impact of the number of data centers on time cost

The analysis results of the impact of the number of data centers on the time cost are shown in Table 1.
Table 1. The impact of the number of data centers on time cost

| Number of data centers | Time cost of application cloud computing | No application of cloud computing time overhead |
|------------------------|------------------------------------------|-----------------------------------------------|
| 5                      | 105s                                     | 240s                                          |
| 10                     | 150s                                     | 262s                                          |
| 15                     | 165s                                     | 275s                                          |
| 20                     | 170s                                     | 281s                                          |
| 25                     | 184s                                     | 288s                                          |
| 30                     | 192s                                     | 296s                                          |

It can be seen from Table 1 that with the increase in the number of data centers, the time cost of data transmission is also increasing. The application of cloud computing to it can effectively reduce the time cost. Among them, when the number of data centers is 5, the time cost of data transmission with cloud computing is 105s, and that of data transmission without cloud computing is 240s. When the number of data centers is 10, the time cost of data transmission with cloud computing is 150s, and that of data transmission without cloud computing is 262s. When the number of data centers is 15, the time cost of data transmission with cloud computing is 165s, and that of data transmission without cloud computing is 275s. When the number of data centers is 20, the time cost of data transmission with cloud computing is 170s, and that of data transmission without cloud computing is 281s. When the number of data centers is 25, the time cost of data transmission with cloud computing is 184s, and that of data transmission without cloud computing is 288s. When the number of data centers is 30, the time cost of data transmission with cloud computing is 192s, and that of data transmission without cloud computing is 296s. Therefore, the application of cloud computing in the data decision-making of data intensive data center can effectively reduce the time cost of data transmission.

5. Conclusions
With the rapid development of mobile internet, internet of things, e-commerce and social media, we have entered the era of big data. The scientific community and other sectors of society all feel the huge impact of massive data. More and more data begin to develop towards the intensive type. The data in the data intensive environment has the characteristics of massive, high-speed change, distribution, heterogeneous, semi-structured or unstructured. The traditional data processing methods can not meet their needs. The emergence of cloud computing brings a new direction to the processing of dense data. This paper analyzes the application status of cloud computing in data intensive data center, and takes data decision as an example to study the application effect of cloud computing in data intensive data center. The results show that cloud computing is more and more widely used in data intensive data centers, and has achieved good results.

References
[1] Abbas H, Maennel O, Assar S. Security and privacy issues in cloud computing[J]. Annals of Telecommunications, 2017, 72(5-6):233-235.
[2] Al-Dhuraibi Y, Paraiso F, Djarallah N, et al. Elasticity in Cloud Computing: State of the Art and Research Challenges[J]. IEEE Transactions on Services Computing, 2018, 11(99):430-447.
[3] Li J, Zhang Y, Chen X, et al. Secure attribute-based data sharing for resource-limited users in cloud computing[J]. Computers & Security, 2018, 72(JAN.):1-12.
[4] Xia Z, Wang X, Zhang L, et al. A Privacy-Preserving and Copy-Deterrence Content-Based Image Retrieval Scheme in Cloud Computing[J]. IEEE Transactions on Information Forensics & Security, 2017, 11(11):2594-2608.
[5] Wei W, Fan X, Song H, et al. Imperfect Information Dynamic Stackelberg Game Based
Resource Allocation Using Hidden Markov for Cloud Computing[J]. IEEE Transactions on Services Computing, 2018, 11(99):78-89.

[6] Maqueira-Marin J M, Bruque-Camara S, Minguela-Rata B. Environment determinants in business adoption of Cloud Computing[J]. Industrial Management & Data Systems, 2017, 117(1):1-19.

[7] Kannan S, Gavrilovska A, Gupta V, et al. HeteroOS: OS Design for Heterogeneous Memory Management in Datacenter[J]. ACM SIGARCH Computer Architecture News, 2017, 45(2):521-534.

[8] Nguyen M, Alesawi S, Li N, et al. A Black-Box Fork-Join Latency Prediction Model for Data-Intensive Applications[J]. IEEE Transactions on Parallel and Distributed Systems, 2020, 31(99):1983-2000.

[9] Abdullah M, Khana S A, Alenez M, et al. Application Centric Virtual Machine Placements to Minimize Bandwidth Utilization in Datacenters[J]. Intelligent automation and soft computing, 2020, 26(1):13-25.

[10] Usman M J, Ismail A S, Chizari H, et al. Energy-efficient Virtual Machine Allocation Technique Using Flower Pollination Algorithm in Cloud Datacenter: A Panacea to Green Computing[J]. Journal of Bionic Engineering, 2019, 16(002):354-366.