ENHANCING CLOUD PERFORMANCE

Sandeep Gautam (14), Ganesh Karki(16), Sudip Maharjan(20), Udip Parajuli(26)
Department of Computer Science & Engineering, School of Engineering,
Kathmandu University

Abstract

Performance is the act of performing or executing any operation, usually with regard to effectiveness. Likewise, Cloud Performance is the measure of how well a Cloud system is functioning and delivering hosted services over the internet. Cloud, which has brought flexibility, capacity, and power of processing, has the internet on its backbone. To improvise and monitor cloud activities, we need to contemplate cloud performance. The performance of a cloud is a critical issue and dependent on various factors that include the Load Balancing and Service Broker Policy. We have reviewed the significant publications as well as the primary real-time cloud performance tools offered to identify the performance factors. Findings revealed the criticality of low response time. In this paper, we provide an overall perspective on cloud evaluation criteria and highlight it with the help of experiments conducted over Cloudsim simulation.

KEYWORDS: *Cloud computing, Performance factors, Load balancing, Broker policies, Response time.*

1. Introduction

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources [1]. Cloud experience sometimes isn't at the level which the service provider aims to be. Cloud management, at some point, needs to appreciate some tangible or intangible changes/modifications. And for this reason, effective cloud performance must be noted. Cloud performance can help determine how well a cloud system is functioning. Furthermore, cloud performance evaluation can help find the improvements which can be made to the system. Factors are affecting the cloud performance determined using automated applications, the discussion of which is carried out in this paper.

In cloud computing, resources are offered as a service. Several cloud providers like Google, Microsoft, Amazon, IBM, and many more are offering Cloud Services. However, it is a huge challenge to provide the rich cloud usage experience due to the implemented Service Policies, Load

Balancing techniques, and the limitation of the Cloud data center.

2. Literature Review

Due to the popularity and progress of cloud in different organizations, cloud performance evaluation is of particular importance, and this evaluation can help users make the right decision. Service level agreements, network bandwidth, storage capacity, buffer capacity, disk capacity, fault tolerance, number of users, location of the data center, usability, latency, processor power and response time are the factors affecting Cloud performance as seen in Niloofar Khanghahi and Reza Ravenmehr's research study[2].

To examine Cloud performance, it is significant to identify the factors that govern Cloud performance. Upon identification of those factors, essential changes can be conducted to make a better experience of Cloud service. Response time is the primary factor that has the most significant impact on Cloud computing. The number and the distance of the data center also have a substantial effect on the performance of the cloud. The research conducted by Jitendra singh[3] over the study of Response time in Cloud computing also suggests that this is true.

V. Vinnetha (2012) had suggested the significance of Cloud performance monitoring tools. The consideration of both the Cloud provider as well as the Cloud user's perspective was emphasized. The author has monitored the response time, storage usage and processor consumed, etc. in both peak and off-hours. Thus, it emphasized following the Cloud performance by monitoring the factors by using the automated tools to enable the Cloud users to know about the status of the cloud.

It can be inferred from the review of publications that performance factors have the most significant impact on Cloud performance.

3. Related Work

Some of the significant research works in the area of Cloud performance has been discussed as follows:

Renowned Cloud service providers like Google, Oracle, IBM, etc. also provide their Cloud environment with their characteristics and architectures. And hence, the actual Cloud environments such as Google, Amazon EC2, Gogrid [4], Elastichosts [4], Mosso [4], Azure[5], etc. are suitable for the Cloud service performance evaluation as well [6].

With an objective to provide affordability and to improve the performance in Cloud Storage, [7] proposed a Meta CDN based approach. The proposed system exploits 'Storage Cloud' resources were creating an integrated overlay network that provides a low cost, high-performance CDN for content creators. Meta CDN removes the complexity of dealing with multiple storage providers by intelligently matching and placing users' context on to one or many storage providers based on their quality of service, coverage, and budget preferences. The author further proposed an efficient Cloud storage model for heterogeneous Cloud infrastructures to improve efficiency in the Cloud.

Jitendra Singh (2014) conducted a series of experiments using a cloud analyzer simulator to determine the dependency of response time on several users, data center location, and broker service policies [3]. Findings are justified in this paper.

4. Cloud Computing Performance Evaluation

The higher performance of the Cloud means higher productivity and rich cloud experience for the user.

We need to find out those essential factors that can result in degraded or enhanced Cloud experience to adjust the Cloud performance per our requirements

4.1 Factors Affecting Cloud Performance

Computational performance of a Cloud server, Cloud storage performance, and Cloud server network performance, all sum up to provide the Cloud performance as a whole. All the factors that affect Cloud performance generally fall under one of the components, as mentioned earlier. Some of the critical factors considered in this paper are as follows:

- Network bandwidth: If the bandwidth is too low to provide service to customers, performance will be low too [8].
- Buffer Capacity: If servers cannot serve a request, it will be buffered in temporary memory. Therefore, buffer capacity has its effects on Cloud performance as well.
- Number of users: If the data center has more users than the rated capacity, then there will be reduction in performance.

- Distance from the data center: Further away the data center, there is more chance of poor Cloud performance.
- CPU and RAM: Computer performance has an effect over Cloud performance as well.

Besides, usability, workload, redundancy, latency too have critical impact over the Cloud performance.

4.2 Response Time Dependency

- Response time is the time elapsed between an inquiry in the system and its response to that inquiry.
- Study revealed that response time is the major factor that has the most significant impact on Cloud computing performance. Thus, can be used as a unit for measurement of cloud performance.
- Low response time may be critical to successful computing.

4.3 Load Balancing Techniques

It is hard to predict the number of users accurately using the public Cloud at a time. Therefore, the load varies from one data center to another. As a result, some data centers are utilized to their optimum level, whereas others are underutilized.

Providing rich Cloud usage experience is a massive challenge due to the variable distance and load involved from users to the Cloud data center. Load Balancing techniques are for maintaining a proper distribution and balance in the load.

4.4 Service Broker Policy

A service broker decides which data center should provide the service to the requests coming from each user. And thus, the service broker controls the traffic routing between the user and datacenter [9].

In simple words, the service broker policy is the datacenter selection policy. When requests come from the users, then the service broker policy helps to decide which data center will provide service for upcoming requests.

There are three service broker policies already existing in Cloud Analyst [10]:

- Closest Datacenter Policy.
- Optimize Response Time Policy.
- Dynamically Reconfigurable Routing with Load Balancing.

Closest data center gives the best results in terms of cost and response time compared to other two policies.

5. Simulation Experiment

To determine the dependency of response time on the Service Broker policy, Load Balancing techniques and the number of

data centers, [3] conducted a series of experiments using a Cloud analyzer simulator built on Cloudsim. In this experiment, the whole Earth is divided into different regions to match the classification of the Cloud analyzer.

S. No.	Region	Countries
1.	R0	USA
2.	R1	Countries of North America
3.	R2	Countries of European Union
4.	R3	Countries of Asia
5.	R4	Countries from Africa
6.	R5	Australia

Table 5: Countries and their regions

In the experiment, homogeneous parameters have been considered such as pricing, storage and memory for all the data centers. Three service policies which have been considered for evaluation are:

- Closest Data Center (CDC)
- Optimize Response Time (ORT)
- Reconfigure Dynamically with Load (RDL)

CDC is based on the quickest path available from the user base to the data center with minimum latency. ORT service broker policy actively monitors the performance of all the data centers and diverts the traffic to a data center which estimates to give the best response time to the end user at the time it is queried. RDL is an extension to CDC.

5.1 Results and Discussion

To find out the response time dependency on Broker Service policy, three different policies are used while the number of user bases remained the same for all the broker policies. Broker Policies that are used for this experiment has been given below:

- Closest Data Center (CDC)
- Optimum Response Time (ORT)
- Reconfigure Dynamically with Load (RDL)

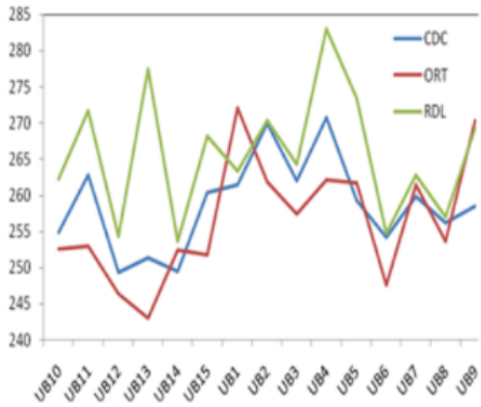

Figure:5.1 : Response time dependency on broker service policy

Figure:5.1 suggests that, in any given condition ORT has performed better than any other algorithm considered.

Further, the experiment's result suggested that response time depends upon the number of data centers and Load Balancing techniques as well.

6. Conclusion

Performance in Cloud computing is a crucial issue and can be controlled at various levels. And, increasing the power and speed of the data center is not always efficient, and sometimes it only adds additional cost. So, efficiency/performance should not be increased more than what is required.

Response time has dependency on broker service policy, load balancing technique, and scheduling algorithm. Response time can be substantially reduced by selecting the appropriate type of broker service policy.

7. Reference

[1] P. M. Mell, T. Grace, "NIST definition of cloud computing", Technical Report, SP 800-145,available at <https://csrc.nist.gov/publications/detail/sp/800-145/final>[Accessed 6 April 2020].

[2] N. Khanghahi, R. Ravanmehr, "CLOUD COMPUTING PERFORMANCE EVALUATION: ISSUES AND CHALLENGES", International Journal on Cloud Computing, Vol3, No.5,October 2013

[3] J. Singh, "Study of Response Time in Cloud Computing", available at : <http://www.mecs-press.org/>

[4] Alexandru Iosup & Simon Ostermann & Nezih Yigitbasi (2010) "Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing", IEEE TPDS, MANY-TASK COMPUTING

[5] Donald Kossmann & Tim Kraska & Simon Loesing, (2011), "An Evaluation of Alternative Architectures for Transaction Processing in the Cloud", ACM

[6]Sergej Poltorak, (2011) "Cloud Computing: Meet the players. Performance Analysis of Cloud Providers", BASEL UNIVERSITY COMPUTER SCIENCE DEPARTMENT

[7] J. Broberg, R. Buyya, and Z. Tari, "MetaCDN: Harnessing Storage Clouds for high performance content delivery", Journal of Network and Computer Applications, Vol.32., 2009, pp. 1012-1022.

[8] Keith R. Jackson & Krishna Muriki, (2010), "Performance Analysis of High Performance Computing Applications on the Amazon Web Services Cloud", IEEE International Conference on Cloud Computing Technology and Science

[9] Hetal V. Patel, Ritesh Patel, “International Journal of Advanced Research in Computer and Communication Engineering”, Vol. 4, Issue 1, 2015.

[10] DhavalLimbani and BhaveshOza, “A proposed Service Broker Strategy in CloudAnalyst for Cost-effective Data Center Selection”, International Journal of Engineering Research and Applications, India, Vol. 2, Issue 1, pages. 793-797, Feb. 2012.