Performance Analysis of Cloud Computing

Rajat[1] Dr. Ajay Kumar Bharti[2]
[1]M. Tech.Student [2]Professor
[1,2]Department of Computer Science
[1,2]Maharishi University of Information Technology, Lucknow, India

Abstract— Cloud computing is a method or technique for enabling convenient, on demand network access to a shared pool of computing resources (such as computer networks, servers, applications, storage, and services) that can be continuously provisioned and released with minimum management efforts. In this paper we have used some of the existing performance monitoring tools and techniques popular among variety of users. During this study a number of factors such as response time, bandwidth and latency etc. have been determined. Considering the frequency of factors appearing from the review work it has been inferred that the cloud response time is very crucial for the cloud performance and selected for further study.

Key words: Cloud Computing, Load Balancing, Cloud Management

I. Introduction

Complexity in cloud computing has increased with the evolution of cloud computing and with the popularity of Hybrid cloud deployment model. To facilitate the IT administrator in managing the cloud, automated cloud management tools are very useful for automating the management task in the cloud. Considering this importance we have been discussed some of the prominent cloud management tool. This work has made significant contribution by differentiating the cloud management tools considering the criteria such as open source and proprietary, cloud deployment and operating system supported etc. This will be very useful for the users in making the decision for selecting the right type of cloud management tools that are applicable for them. The main objectives of the present research work are:

- To determine factors affecting the performance.
- To compare the cloud management tools.
- To study response time dependency on service broker policies, load balancing techniques, number of data center and number of users.

A. *Performance Analysis*

Performance can be measured considering the various aspects. This work is an effort to determine the factors which have a huge impact on the performance of the cloud computing. Following are the methods which are used to determine the performance factors:

- Applications that monitor the performance.
- Performance factors that are monitored by the automated applications.
- Performance factors that has been considered by previous researchers.

B. *Cloud Management*

A number of automated cloud management tools exist which can be used for performance monitoring and management. This section discusses the comparison on cloud management tools existing by comparing their features such as open source/ proprietary, type of cloud used (i.e. private or public cloud) and operating system supported by cloud management tools.

C. *Response Time Dependency*

Data center are using different service broker policies and load balancing techniques. Current work studies the performance of some of the prominent service broker policies and load balancing techniques considering the following objectives:

- To find out the dependency of response time on various service broker policies.
- To find out the dependency of response time and data processing time on load balancing techniques.
- To determine the effect of user base on response time.
- To determine the effect of number of data center on response time.
- To determine the effect of region distance on data center response time.

We have focused on the issues which are directly under the control of cloud provider. None of them has considered the dependency of load balancer on broker service policies and what is the best combination of scheduling algorithm, broker policy and data center location in cloud computing. Hence, there is a need to undergo holistic study on the above factors related to cloud performance. The current work focuses to determine the best service broker policy and scheduling algorithm considering the distance of data center.

II. Review of Literature

Science Cloud (Keahey et al., 2008) is a Cloud testbed infrastructure constructed by the scientific community for service performance evaluations. The infrastructure is configured with the Nimbus toolkit to enable remote releasing of resources in similar manner as EC2 services.

OpenCirrus (Avetisyan et al., 2010) is a large scale Cloud testbed comprising of federated heterogeneous distributed data centers. It enables researchers to exchange data sets and develop standard Cloud computing benchmarks. Virtual machine management in OpenCirrus can be done by different services as long as they are compatible with the EC2 interface.

Open Cloud (Grossman et al., 2009) is a research Cloud testbed that is designed to support computations that span more than one data centers. Data centers in Open Cloud are interconnected with a dedicated high-speed wide area network. Sakellari and Loukas (2013) made a more detailed review of currently available research Cloud testbeds.

Stantchev (2009) presented a general approach for evaluating non-functional QoS properties of individual Cloud services. This approach is based on an architectural transparent" black box" methodology and comprises the following steps: identifying benchmark, identifying configuration, running tests, analyzing results, and making recommendation.

Hill and Humphrey (2009) conducted a quantitative performance analysis of high performance computing on EC2 infrastructure and their obtained results show that EC2 may offer reasonable service performance for small-scale HPC applications. However, the evaluation reported (Jackson et al., 2010) indicates that EC2 is slower than a typical mid-range Linux cluster and much slower than a modern HPC system when supporting realistic super-computing applications. The above studies primarily focused on tightly-coupled HPC applications, for example Message-Passing Interface (MPI) programs. Workflows are loosely-coupled parallel applications that consist of computational tasks interconnected through data and control dependencies. Performance of EC2 Cloud service was evaluated from the perspective of scientific workflows (Juve et al., 2009) and compared with a typical HPC system.

Lu et al. (2010) evaluated the performance of Azure service for running Windows applications. Elastic service provisioning is one of the distinguishing features of Cloud computing. With a certain level of QoS guarantee, a Cloud data center automatically allocates more resources when the workload increases beyond a certain threshold (scaleup or scale out), and releases unused resources when the load reduces (scale-down or scale-in). The impact of resource scaling on service performance has also been analyzed as an important aspect of Cloud performance evaluation.

Ghoshal et al. (2011) compared the I/O performance of Amazon EC2 confronted with Magellan, a private Cloud platform, and an HPC cluster. The obtained results highlight the overhead and variability of I/O performance in both public and private Cloud solutions and also indicate that NFS performance of regular EC2 instances is many orders of magnitude worse than that of the parallel file system installed in the HPC cluster. In order to improve I/O performance to support data-intensive applications, Amazon lunched the storage-optimized instance family that includes High I/O quadruple extra-large (HIl) instance and High Storage eight extralarge (HSl) instance.

Exp´osito et al. (2013) evaluated performance of two Amazon EC2 Cluster Compute instances the quadruple extra-large and eight-extra-large instances. They found that the scalability of HPC applications on public Cloud infrastructures relies heavily on the performance of communications, which depends on both the network fabric and its efficient support in the virtualization layer. Cloud services may achieve quite different levels of performance under various workloads generated by diverse applications. Unlike computation and communication intensive applications, data-intensive applications typically show strong demands for high-performance I/O and storage access in a Cloud infrastructure.

Li et al. (2013) employed a systematic literature review method to outline the stateof- the-practice of evaluating commercial Cloud services by classifying the published works according to the following categories: evaluation purposes, the evaluated Cloud services, aspects and properties of evaluated Cloud services, the metrics measured for evaluation, benchmarks used for testing, and the experimental environment setup. In order to provide a guideline for implementing different types of measurement experiments for evaluating the numerous and diverse Cloud services, the researcher (Feng et al., 2014) developed a taxonomy of IaaS Cloud service performance evaluation and proposed a conceptual model that generalizes the existing measurement-based performance evaluation practices.

Atas¸ and Gungor (2014) developed a framework for evaluating PaaS performance and proposed a set of benchmark algorithms that can help determine the most appropriate PaaS provider based on different resource needs and application requirements. Commercial PaaS services such as Cloud Foundry, Heroku, and OpenShift, were tested (Atas¸ and Gungor, 2014) and the obtained results were analyzed by the authors using two evaluation methods: the Analytical Hierarchy Process (AHP) and Logic Scoring of Preference (LSP).

Leitner and Cito (2016) conducted literature review to collect and codify existing evaluation results of public commercial IaaS Cloud services with a performance of an IaaS Cloud service can be estimated in advance and how stable the performance will be. In addition to using commercial Cloud infrastructures as the testbeds for evaluating service performance, some research Cloud testbeds have also been constructed and utilized for Cloud service performance evaluation.

III. CLOUD PERFORMANCE & LOAD BALANCING

In cloud computing paradigm, application and data are stored in data center of cloud provider which is located in diverse geographical locations. Presently developed countries such as USA, UK and countries from European Union are selected as preferred locations for data center by the cloud providers. These locations have been selected considering the peace and stability existing in these countries, so that any loss due to terrorist activities can be avoided. Data center locations of some of the prominent cloud providers have been depicted in table.

S.N	Region	Data Center Location		
		USA	EU	Asia
1.	Amazon Web Service	East(Northern Region), US West(Northern California), Brazil	Ireland, Netherlands, Germany	Japan, China, Singapore
2.	Microsoft	Quincy, Washington, San Antonio, Texas Chicago, Illinois	Ireland, Netherlands	Singapore, Hong Kong
3.	Google	USA (Quilicura, Oklahoma, Lenoir, The Dalles, Oregon)	Finland, Belgium	Singapore Taiwan, Hong Kong

Table 1: Data Center Locations of Prominent Cloud Provider

Each data center consists of a number of nodes; these nodes are in the order of hundreds. Data center offers basic on demand storage and computation over the internet. Provisions of these computational resources are in the form of VM, which is an abstract unit of computational in cloud. These VM differs in configuration such as memory, CPU and bandwidth etc. Due to the dynamic nature of cloud environments, diversity of user's requests and time dependency of load, cloud centers must provide expected quality of service at widely varying loads (Xiong & Perros, 2009; Baker et al., 2011).

IV. PERFORMANCE FACTORS DETERMINATION

Monitoring the cloud performance is important for both the cloud user and the cloud provider in real time and analyzing the historical performance of the cloud. A large number of applications running in public cloud require huge and powerful resources. Applications known as "big data" consist of workload such as digital media collection, virtual worlds, simulation traces, data obtained from scientific instruments, and enterprise business databases are resource hungry application. Response time for these applications is very important, lacking which leads to unpleasant experience to the user. Applications hosted on public cloud should be checked for its performance i.e. response time and processing time so that these factors are within the tolerance limit. Before provisioning the more resources, it is important that resources available are used at its optimum level to avoid further investment in these resources.

V. PERFORMANCE MONITORING APPLICATIONS

To monitor the performance in the cloud computing, a number of automated performance monitoring (APM) tools from various vendors are available which provide the monitoring in real time as well as on historical data. Majority of them enable the user to monitor the details of their cloud using wide variety of devices such as smart phones to laptop devices. Some of these tools have been defined as in table.

S. N.	Provider	APM Tool	Factors monitored
1.	Netinst (http://www.netinst.com/-products/observer-reportingserver/index.php)	Observe reporting system	End user page response time Transactions Processed Network error, latency & utilization
2.	Copper Egg (http://copperegg.com/revealuptimewebsite-monitoring/)	RevealUptime	URL, port, & site latency, response time, uptime, health. Worldwide real time data collection coverage, analytics, alarms. Troubleshooting, multiuser access
3.	Hyperic (http://www.hyperic.com/products/cloud-status-monitoring)	Cloudstatus	Monitor service availability, response time, latency, and throughput Provides real-time Reports Application availability and performance
4.	CA technologies	CA virtual assurance	Infrastructure response time, latency time Monitor the Table 1.2: Provider and factor for monitoring performance in real time Deployed as virtual appliances.
5.	Virtual instruments	Virtual wisdom	Measures the response time of individual fiber channel. SAN Latency time and load information for all fiber channel traffic.
6.	Xangati	Xangati	Infrastructure response time for each application Monitors data in real time Storage performance for only IP attached devices
7.	Akorn	Akorn	Infrastructure response time is collected from end to end. Polls the entire infrastructure once in 15 minutes (not in real time).

Table 2:

Hence, from the above discussion and table it can be inferred that response time, data processing time, throughput and uptime are important factors which plays pivotal role in cloud and hence, required to be monitored to determine the cloud performance.

VI. CONCLUSION

In cloud computing performance, some of the work has already been carried out. Such related work has been discussed review of literature highlight the development that took place in cloud computing performance area. From the related works it is evident that majority of them were focused in some particular area such as data center location or scheduling, virtual machine or migration of virtual machine or managing the cloud at application level. While our work has considered almost all the factors such as data center locations, scheduling algorithm, broker service policy, number of virtual machine and number of processors. We have focused on the issues which are directly under the control of cloud provider. None of them has considered the dependency of load balancer on broker service policies and what is the best combination of scheduling algorithm, broker policy and data center location in cloud computing. Hence, there is a need to undergo holistic study on the above factors related to cloud performance. The current work focuses to determine the best service broker policy and scheduling algorithm considering the distance of data center.

REFERENCES

[1] Exp´osito, R. R., Taboada, G. L., Ramos, S., Gonz´alez-Dom´ınguez, J., Touri˜no, J. and Doallo, R. (2013), ‘Analysis of i/o performance on an amazon ec2 cluster compute and high i/o platform’, Journal of grid computing 11(4), 613–631.
[2] Accorsi, R., Lowis, L. & Sato, Y.(2011),“Automated Certification for Compliant Cloud-based Business Processes”, Business & Information Systems Engineering, Vol. 3, No.3. pp. 145-154, DOI: 10.1007/s12599-011-0155-7.
[3] Hill, Z. and Humphrey, M. (2009), A quantitative analysis of high performance computing with amazon’s ec2 infrastructure: The death of the local cluster?, in ‘Grid Computing, 2009 10th IEEE/ACM International Conference on’, IEEE, pp. 26–33.
[4] Grossman, R., Gu, Y., Sabala, M., Bennet, C., Seidman, J. and Mambratti, J. (2009), ‘The open cloud testbed: A wide area testbed for cloud computing utilizing high performance network services’, arXiv preprint arXiv:0907.4810 .
[5] Chieu,T.C., Mohindra,A., and Karve,A.A.(2011),"Scalability and Performance of Web Applications in a Compute Cloud",Eighth IEEE International Conference on e-Business Engineering,pp.317-323.
[6] Feng, Y., Li, B. and Li, B. (2014), ‘Price competition in an oligopoly market with multiple iaas cloud providers’, IEEE Transactions on Computers 63(1), 59–73.
[7] Exp´osito, R. R., Taboada, G. L., Ramos, S., Touri˜no, J. and Doallo, R. (2014), ‘Performance evaluation of data-intensive computing applications on a public iaas cloud’, The Computer Journal p. bxu111.
[8] Atas¸, G. and Gungor, V. C. (2014), ‘Performance evaluation of cloud computing platforms using statistical methods’, Computers and Electrical Engineering 40(5), 1636–1649. Avetisyan, A. I., Campbell, R., Gupta, I., Heath, M. T., Ko, S. Y., Ganger, G. R., Kozuch, M. A., O’Hallaron, D., Kunze, M., Kwan, T. T. et al. (2010), ‘Open cirrus: A global cloud computing testbed’, Computer 43(4), 35–43.
[9] Ghosh, R., Trivedi, K. S., Naik, V. K. and Kim, D. S. (2010), End-to-end performability analysis for infrastructure-as-a-service cloud: An interacting stochastic models approach, in ‘Dependable Computing (PRDC), 2010 IEEE 16th Pacific Rim International Symposium on’, IEEE, pp. 125–132.
[10] Ghoshal, D., Canon, R. S. and Ramakrishnan, L. (2011), I/o performance of virtualized cloud environments, in ‘Proceedings of the second international workshop on Data intensive computing in the clouds’, ACM, pp. 71–80.
[11] Li, Z., Zhang, H., OBrien, L., Cai, R. and Flint, S. (2013), ‘On evaluating commercial cloud services: A systematic review’, Journal of Systems and Software 86(9), 2371–2393.
[12] Lu,W., Jackson, J., Ekanayake, J., Barga, R. S. and Araujo, N. (2010), Performing large science experiments on azure: Pitfalls and solutions, in ‘Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on’, IEEE, pp. 209–217.
[13] Madan, K. C. (2014), ‘On a single server queue with arrivals in batches of variable size, general service in three fluctuating modes, balking, random breakdowns and a stand-by server during breakdown periods’, Investigaci´on Operacional 35(3), 189–201.
[14] Seleman, Y. E. A. (2016), ‘Efficient e-government services, constraints and problems of technical applications and software and the transition to smart government’, Global Journal of Computer Science and Technology 16(1).
[15] Sharma, G. P., Singh, S., Singh, A. and Kaur, R. (2016), ‘Virtualization in cloud computing’.
[16] Sharma, S. K. and Kumar, R. (2012), ‘A markovian feedback queue with retention of reneged customers and balking’, Adv. Model. Optim 14(3), 681–688.
[17] Shea, R., Wang, F., Wang, H. and Liu, J. (2014), A deep investigation into network performance in virtual machine based cloud environments, in ‘INFOCOM, 2014 Proceedings IEEE’, IEEE, pp. 1285–1293.
[18] Xia, Y., Zhou, M., Luo, X., Zhu, Q., Li, J. and Huang, Y. (2015), ‘Stochastic modeling and quality evaluation of infrastructure-as-a-service clouds’, IEEE Transactions on Automation Science and Engineering 12(1), 162–170.
[19] Xing, Y. and Zhan, Y. (2012), Virtualization and cloud computing, in ‘Future Wireless Networks and Information Systems’, Springer, pp. 305–312.
[20] Xiong, K. and Perros, H. (2009), Service performance and analysis in cloud computing, in ‘Services-I, 2009 World Conference on’, IEEE, pp. 693–700.