Study of Response Time in Cloud Computing

Jitendra Singh
PGDAV College, University of Delhi-110065
Email: jitendra.singh0705@gmail.com

Abstract—Performance of a cloud is a critical issue and dependent on various factors that include load balancing used to distribute the load at data center. Proliferation of the cloud paradigm is profoundly dependent on the performance of the cloud. Therefore, to examine the cloud performance, it is significant to identify the factors that govern cloud performance. This work is focused to identify the factors that may have significant impact on the cloud performance. To identify the performance factors, we have reviewed the major publications of the last five years as well as the major real time cloud performance tools offered. Study revealed that response time is the major factor that has the significant impact on cloud computing performance. Finally, we have conducted an experiment to establish the dependency of response time on various load balancing techniques and number of data center. Experiment revealed that the round robin is directly proportional to the number of data center

Index terms—Cloud performance, load balancing, response time, data center, scheduling, Broker policies.

I. INTRODUCTION

Cloud computing is a utility based model and offers the ubiquitous access of resources [1, 2, 3]. In cloud paradigm, resources are managed and maintained by the cloud providers; hence users are free to concentrate on business instead on Information technology (IT) [4, 5]. In cloud computing resources are accessed from centralized location similar to that of mainframe [6]. These centralized resources that facilitates in resource provisioning are known as data center.

In cloud computing, resources are offered as a service. Correspondingly, Infrastructure as a Service (IaaS), platform as a Service (PaaS), and Software as a Service (SaaS) are three prominent types of services offered by the cloud providers [7]. These services are offered with the help of Private, Public, Community and Hybrid cloud. User can subscribe to either of them or in combination. Among the offered deployment models, public and private are the most popular cloud deployment models and subscribed by the wide variety of users.

Presently, a number of cloud providers that include Amazon, Google, Microsoft, IBM, etc. are offering cloud services. Cloud subscribers have the great expectation from this paradigm and expect the experience similar to

that of own managed resources. However, providing the rich cloud usage experience is a huge challenge due to the distance involved from user to the cloud data center, existing bandwidth, Service policies and Load balancing techniques, etc.

Arrival and departure of subscriber in public cloud cannot be predicted accurately. Therefore, load varies from one data center to another. As a result, some of the data centers are utilized to their optimum level whereas others are underutilized. To distribute and balance the load, broker service policy and load balancing techniques are some of the popular methods utilized. This work has the following objectives:

- To identify the factors having significant impact on cloud performance.
- To determine the dependency of response time on service broker policy
- To determine the effect of number of data center on response time.

Rest of the manuscript has been organized as:
Section II describes the data center, their architecture and locations opted by the prominent cloud providers. Section III identifies the performance factors by reviewing the literature. Section IV describes the prominent researcher's work that is related to us in any way. Section V describes the experimental environment needed for our work. Results obtained of our experiment are discussed in section VI. Finally, paper concludes with section VII by enumerating the conclusion.

II. CLOUD DATA CENTER

Cloud data centers are widely utilized for the provisioning of resources. To provide better performance, reliability and availability, majority of the cloud providers have opted for multi-locations of the data centers. Presently, developed countries such as USA, UK, Singapore and countries of European Union have been selected as preferred locations for data center. Location of a data center is selected by considering the various factors such as reputation of the country, peace, stability prevailing, etc. Selection of stable and peaceful country avoids any loss due to terrorist activities, theft, natural calamities, etc. Data center locations opted by the prominent cloud providers have been depicted in table 1

Table 1: Data Center Locations of prominent cloud providers

S.No.	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

Each data center consists of numerous nodes which are in the order of hundreds. Nodes in each data center offers basic on demand storage and computational services over the internet. Computational resources are provisioned in the form of virtual machine (VM), an abstract unit of computation in the cloud. Offered VM differs with one another in configuration such as memory, CPU, bandwidth, etc. Data center architecture is illustrated in fig. 1. To cater the need of wide variety of users, considering their computational need and pricing options. VM resources, such as RAM, processor and storage capabilities can be subscribed by the cloud users as per their need. Selection of appropriate VM resources is extremely important for better response time and rationalization of the billing.

Response time is not only crucial for small applications but also for larger applications that are running on a public cloud. Cloud data centers have the responsibility to provide the quality of services, despite of the dynamic nature of the cloud where the load varies all of a sudden [8-9]. To fulfill the quality requirement, applications hosted on public cloud need to be checked for their performance i.e. response time and processing time so that performance factors are within the tolerance limit. Before provisioning of new resources, it is desired that existing resources are used to their optimum level, to avoid unnecessary investment on resources. Selection of appropriate type of scheduling techniques, broker service policy and load distribution are some of the methods that can be applied to optimize the usage of cloud resources. These techniques are extremely useful in distributing the load evenly.

III. IDENTIFYING PERFORMANCE FACTORS

To improve the cloud performance, it is extremely important to identify the performance factors having significant impact on cloud performance. To accomplish this objective, we have reviewed the major publications of the last five years. In addition, many of the real time cloud performance monitoring tools offered by the cloud provider or third party vendors have been reviewed.

Study revealed that some extra ordinary works have already conducted in the performance area.

A. Literature review

Authors such as [10] had suggested the significance of cloud performance monitoring tools by considering both the cloud provider as well as cloud user's perspective. Author has monitored the response time, CPU, storage usage, etc. in peak and off peak hours. [11] measured the performance of five (05) dominant cloud providers (Google, Rackspace, Salesforce.com, Amazon and terremark) for their performance by hosting a website. Author measured the response time and latency of each of the above cloud providers. [12] suggested a new approach for infrastructure management to determine the performance in virtualized environment by monitoring infrastructure response time. [13] has emphasized to monitor the performance factors, such as latency, response time and execution time, and declared that these factors are extremely risky for the overall performance of the cloud. [14] Emphasized to monitor the cloud performance by monitoring the factors like response time, and latency by using the automated tools to enable the cloud users to know about the exact status of the cloud.
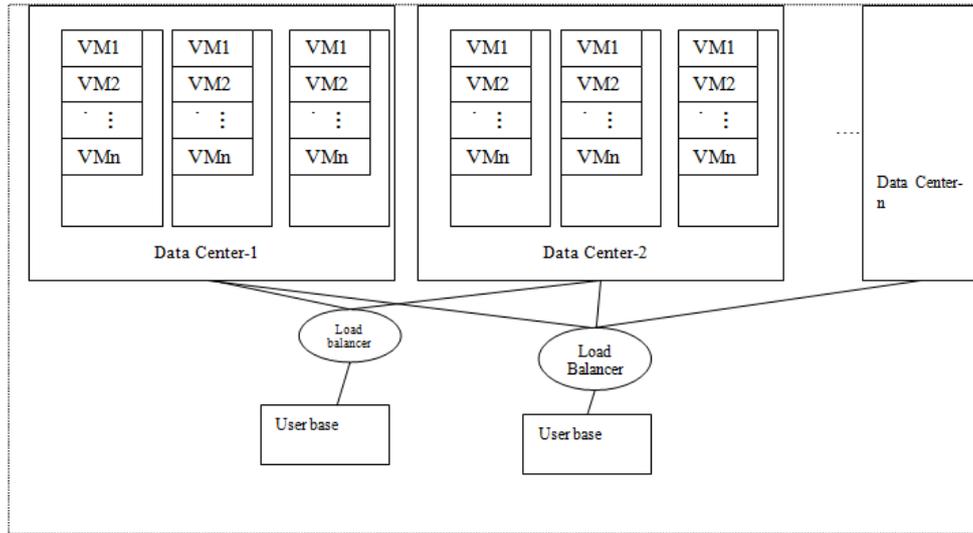
Fig. 1: Data center and user base connection

Table 2: Providers and monitoring factors

S.no	Provider	APM Tool	Performance functions
1.	Netinst[15]	Observe reporting system	• Page response time • Transactions processed • Network error, latency & utilization
2.	Copper Egg[16]	RevealUptime	• URL, port, & site latency, response time, uptime, health. • Worldwide real time data collection coverage, analytics, alarms. • Troubleshooting, multi-user access
3.	Hyperic [17]	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 • Monitors 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 all the applications, individually. • 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).

Importance of cloud performance is also realized by the corporate sector; therefore they are offering automated performance monitoring tools. Monitoring tools are extremely useful in determining the cloud status at specific time and in analyzing the historical performance of the cloud. Cloud monitoring tools can be offered by the cloud providers themselves (such as cloudwatch by Amazon) or by third party, such as, revealuptime and cloudstatus, etc. Monitoring tools offered by the cloud providers are not exhaustive; therefore, they are suitable only for limited usage, whereas third party tools are exhaustive and provide detailed information related to the cloud status and their performance. Various real time cloud performance tools

considered for the study is enumerated in table 2.

B. Study on Performance tools

To monitor the performance in the cloud computing, numerous automated performance monitoring (APM) tools can be utilized. These tools are capable enough to monitor the data in real time as well as to analyze the historical data. Monitoring tools can be accessed using wide variety of devices including, smart phones, desktop, laptop, etc. Some of the prominent cloud performance tools, their respective providers, performance functions have been discussed in table 2

C. Findings

From the review of publications, and the real time automated performance tools, it can be inferred that response time, data processing time, throughput and uptime are the significant factors having significant impact on cloud performance. Considering the significance of the response time, it has been considered for further study.

IV. RELATED WORK

Some of the major work related to cloud performance has already been carried out. Majority of them were focused on Scheduling algorithm, broker service policies, etc. Some of these major research works in the area of cloud performance has been discussed as follows:

An analytical model for interaction aspects between on-demand request and allocation of virtual machine from the infrastructure owner to the user has been proposed by [18]. Comparison and numerical analysis of several scheduling algorithm has been performed considering the blocking probability of on demand request. Paper concluded with the possibilities of energy saving in the operational range (where on-demand requests do not face unpleasant blocking probability) with the allocation of virtual machines to physical servers based on the priority of different allocation strategies.

In web 2.0, applications composed of several components that run indefinitely and need to be available to end users throughout their execution life cycle. To avoid failure, [19] proposed a solution to find the optimal number of component types needed on nodes so that each component is present on every allocated node. Further, more nodes cannot exceed a maximum threshold and the total running cost of the applications needs to be minimized. Authors have suggested both optimal and sub-optimal solution. Both solutions were relying on genetic algorithms to achieve their goals. The efficiency of the suboptimal algorithm was studied with respect to its success rate, i.e., probability of the schedule and to provide highly available applications in case all but one node fails.

To improve the performance in IaaS based multiple clouds, an algorithm for resource optimization with pre-emptable task execution that increases the utilization of the cloud was proposed by [20]. Algorithm was capable to adjust the resource allocation dynamically, based on the updated information of the actual task executions. Efficient and reliable access limitations of traditional CPU- oriented batch schedulers were discussed by [21]. To improve the performance, authors proposed Stork data placement scheduler in data intensive computing. Lucas-[22] considered to improve the performance in multi cloud having fragmented interfaces, pricing schemes and virtual machine offerings. To minimize the cost, work had proposed a broker based on optimization algorithms to interface with multi cloud. Broker's algorithm was capable to decide how to select the cloud for the deployment of a service and how to distribute the different components of a service among different clouds,

or even when to move a given service component from one cloud to another to satisfy opted optimization criteria. Paper has considered different scheduling strategies for optimal deployment of virtual services across multiple clouds, based on different optimization criteria (e.g. cost optimization or performance optimization), different user constraints (e.g. budget, performance, instance types, placement, reallocation or load balancing constraints), and different environmental conditions (i.e., static vs. dynamic conditions, regarding instance prices, instance types, service workload, etc.).

To provide affordability and to improve the performance in 'cloud storage', [23] proposed a Meta CDN based approach. Proposed system exploits 'Storage Cloud' resources, creating an integrated overlay network that provides a low cost, high performance CDN for content creators. MetaCDN removes the complexity of dealing with multiple storage providers by intelligently matching and placing users' content on to one or many storage providers based on their quality of service, coverage and budget preferences. Work considered the access efficiency of cloud storage during accessibility from anywhere and the capability demanded by the users. Authors further proposed an efficient cloud storage model for heterogeneous cloud infrastructures to improve the efficiency in cloud.

To evaluate the cloud performance considering the diversity of user requests and nature of cloud data centers. [24] proposed a novel approximate analytical model for cloud server farms. The model allows cloud operators to determine the relationship between the number of servers and input buffer size, on one side, and the performance indicators such as mean number of tasks in the system, blocking probability, on the other hand probability that a task will obtain immediate service.

Considering the heterogeneity and proliferation of cloud eco system [25] proposed a brokering approach that optimizes placement of virtual infrastructures across multiple clouds and also abstracts the deployment and management of infrastructure components in these clouds. To minimize the end-to-end delay for single-input applications, maximizing frame rate for streaming applications and to arbitrary node reuse [26] proposed dynamic programming-based optimal solution and proved the NP-completeness of the problems, for each of which, a heuristic algorithm based on a similar optimization procedure was proposed.

To scale the application in intercloud, to improve the performance in case of resource limitations in cloud computing and to provide the quality of service to the user as per the user service level agreement [27] proposed an architecture for cloud co-coordinator. The proposed co-coordinator enables scaling of application across multiple, independent Cloud data centers, following market-based trading and an extensible design that allows its adoption in different public and private Clouds.

[28] provided an overview of the current state of high performance cloud computing technology, customer's requirements. Work advocated use of virtual cluster from different cloud providers to provide the effective

utilization of the resource and to avoid either underutilization or overloading.

Considering the performance dependency of IaaS clouds on a large set of parameters, e.g., workload, system characteristics and management policies [29] proposed a scalable stochastic analytic model for performance quantification of Infrastructure-as-a-Service (IaaS) Cloud. Specifically, a class of IaaS Clouds that offer tiered services by configuring physical machines into three pools with different provisioning delay and power consumption characteristics. To minimize the interactivity work further suggested a multi-level interacting stochastic sub-models approach where the overall model solution is obtained iteratively over individual sub-model solutions and the proposed method compared with a single-level monolithic model to determine its efficiency.

[30] Considered the dependency of HPC on scalability for efficient support of network communications in virtualized environments. The scalability of High Performance Computing (HPC) applications in IaaS cloud depends heavily on the efficient support of network communications in virtualized environments. Paper analyzed the main performance bottlenecks in HPC application scalability on the Amazon EC2 Cluster Compute platform by evaluating the communication performance on shared memory and a virtualized 10 Gigabit Ethernet network; assessing the scalability of representative HPC codes, the NAS Parallel Benchmarks, using number of cores, up to 512; analyzing the new cluster instances (CC2), both in terms of single instance performance, scalability and cost-efficiency of its use; suggested techniques for reducing the impact of the virtualization overhead in the scalability of communication-intensive HPC codes, such as the direct access of the Virtual Machine to the network and reducing the number of processes per instance; and proposing the combination of message-passing with multithreading as the most scalable and cost-effective option for running HPC applications on the Amazon EC2 Cluster Compute platform.

Multimedia communication research and development often requires computationally intensive simulations in order to develop and investigate the performance of new optimization algorithms.

Depending on the simulations, they may require even a few days to test an adequate set of conditions due to the complexity of the algorithms. The traditional approach to speed up this type of relatively small simulations, which require several develop–simulate–reconfigure cycles, is indeed to run them in parallel on a few computers and leaving them idle when developing the technique for the next simulation cycle. This work proposed a new cost-effective framework based on cloud computing for accelerating the development process, in which resources are obtained on demand and paid only for their actual usage. Issues are addressed both analytically and practically running actual test cases, i.e., simulations of video communications on a packet lossy network, using a commercial cloud computing service. A software

framework has also been developed to simplify the management of the virtual machines in the cloud. Results show that it is economically convenient to use the considered cloud computing service, especially in terms of reduced development time and costs, with respect to a solution using dedicated computers, when the development time is longer than one hour. If more development time is needed between simulations, the economic advantage progressively reduces as the computational complexity of the simulation increases.

From the above work, it can be inferred that majority of the above works were focused on improving the performance by means of various techniques, including, scheduling algorithm, balancing techniques, or scheduling to reduce the energy consumption. Hence, it is apparent that there is a need to determine dependency of factors such as number of data center, broker service policy and load balancing on response time

V. EXPERIMENTAL ENVIRONMENT

To determine the dependency of response time on number of users, data center location and broker service policies as mentioned in section 1, a series of experiments have been conducted using cloud analyzer simulator. This simulator is built on CloudSim. Results obtained from CloudSim environment are very close to real environment. It is widely used by researchers as well as by the corporate to simulate the cloud environment and to obtain the result for analysis purposes. Following sub-sections describe the environment used to conduct this experiment.

A. Region Division

In this experiment the whole earth is divided into five regions to match with the classification of cloud analyzer. Region divisions and the countries falling under these regions have been summarized in table 3.

B. Data Center configuration

Data center is the key component in the cloud computing and offers the various infrastructure to the users, for instance, computational and storage. Resources offered may have different configuration, for instance, processor architecture, Operating system and Virtual machine manager. In our experiment, total 05 data centers have been considered for the simulation environment.

Table 3: Countries and their regions

S.No.	Region	Countries
1.	R_0	USA
2.	R_1	Countries of North America
3.	R_2	Countries of European union
4.	R_3	Countries of Asia like china, India
5.	R_4	Country from Africa
6.	R_5	Australia

Configuration of each data centers has been given in table 4.

Table 4: Configuration of data center

S.No.	Memory size(Mb)	Storage Details(Gb)	Available BW (Kbps)	No. of Processor	Processor Speed (Mhz)
1.	2048000	1000000	10000	20	10000
2.	2048000	1000000	10000	4	10000
3.	2048000	1000000	10000	4	10000

In our experiment, we have considered the homogeneous parameters such as pricing, storage and memory for all the data centers. It has been considered similar for all the data centers, so that comparison can be accomplished with one and another. Each Data centers have three hardware units as per the following details:

a) User Base

Simulation is carried out considering 15 user base (UB) located in different regions to access a particular data center. Each user base has the following configuration:

b) Application Deployment

Each server deployed has the following architecture

Table 5: application deployment architecture

VM	10
Memory(each VM)	512 MB
Bandwidth	1000

c) Service Broker Policy

Three service policies which have been considered for evaluation:

- Closest Data Center
- Optimize response time
- Reconfigure dynamically with load

Closest Data Center Based Routing (CDC): CDC is based on the quickest path available from the user base to the data center with minimum latency. Service broker will transmit the traffic to DC with minimum transmission delay.

Optimum Response Time (ORT): ORT Service broker policy actively monitor the performance of all the data center and directs the traffic to a datacenter which estimates to give the best response time to the end user at the time it queried.

Reconfigure Dynamically with Load (RDL): This service policy is an extension to CDC and deals with the scaling of resources as per the load it receives. In case of overload it increases the number of VMs in the data center and reduces the VMs in case of less load.

Table 6: Hardware units of Data center

S.No.		
1.	Architecture	X86
2.	Operating System	Linux
3.	Virtual Machine Manager (VMM)	Xen

Table 7: User base Configuration

Configurations factors	Number
User grouping factor in user bases (Equivalent to number of simultaneous users from a single UB)	1000
Request grouping factor in Data Center (Equivalent to number of simultaneous request a single application server instance can support)	100
Executable Instruction length per request These user bases are not homogeneously distributed to evaluate the effect of response time of data center as per the user base.	1000

VI. RESULTS AND DISCUSSION

This work is primarily focused to find out the dependency of response time on:

- service broker policy
- load balancing techniques and
- Number of data center.

Results obtained are discussed in the upcoming sub-sections.

Dependency of response time on service broker policies.

To find out the response time dependency on broker service policy, three different policies are used while the number of user base remained the same for all the broker policies. Broker policies that are used for this experiment has been given in table 8.

Table 8: Service Broker Policy

S.No	Service Policy
	Closest data center (CDC)
	Optimum response time (ORT)
	Reconfigure dynamically with load(RDL)

Result obtained has been shown in Fig. 2. In the given condition ORT has performed better than any other algorithm considered, whereas RDL has the worst performance

A. Effect of number of data center on response time

To determine the effect of number of data center on response time, we have started from 03 data center and increased one data center in the consecutive experiment. In this experiment, other than the data center no other factors has changed.
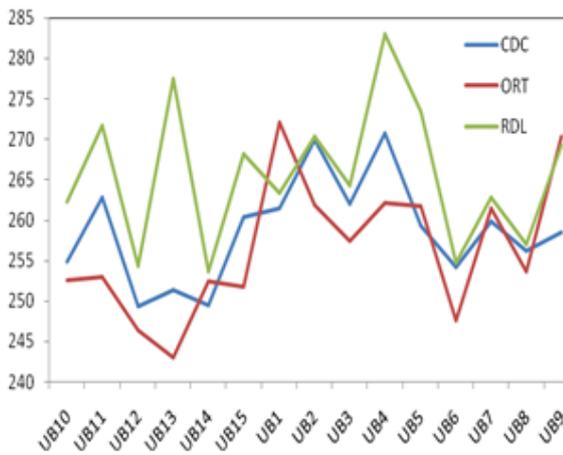
Fig. 2: Response time dependency on broker service policy

It is revealed that response time reduced drastically upto 06 data center beyond which it is reaching towards constant value. Same is illustrated in Fig. 3.

Table 9: Number of data center

S.No.	Number of data Center
1	03
2	04
3	05
4	06
5	07
6	08
7	09
8	10
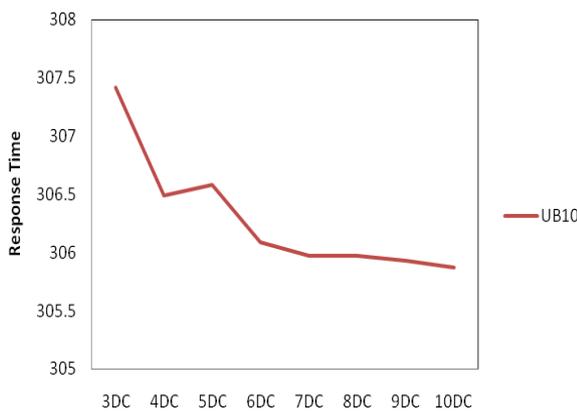
Fig. 3: Response time dependency on number of user base

VII. CONCLUSION

Performance in cloud computing is a critical issue and can be controlled at various level. Literature review revealed that response time is a significant factor that governs the cloud performance. Response time has dependency on broker service policy, load balancing technique and scheduling algorithm. Experiment conducted to determine the dependency of response time on broker service policy and number of data center. Results obtained revealed that among three broker service policies i.e. closest data center, optimum response time and re-configure dynamically with load; closest data center has the best performance over other broker policies considered. Whereas, to determine the data center performance, when we increased the number of data center, response time reduced proportionally upto a certain level before reaching to almost constant value. Response time can be substantially reduced by selecting the appropriate type of broker service policy. Consequently, performance of cloud can be improved significantly.

REFERENCES

[1] R. Buyya, C.S. Yeo, & S. Venugopal, "Market-oriented cloud computing: vision, hype, and reality for delivering it services as computing utilities" in: High performance Computing and Communications, 2008, HPCC'08, 10th IEEE International Conference on, 2008, pp. 5–13. doi:10.1109/HPCC.2008.172.

[2] R. Buyya, C. Yeo, S. Venugopal, J. Broberg, & I. Brandic, "Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility", Future Generation Computer Systems, Vo1.25, 2009, pp. 599–616.

[3] G. Boss, P. Malladi, D. Quan, L. Legregn "Cloud computing, in: High Performance On Demand Solutions", (HiPODS), IBM, (2007).

[4] L.M. Vaquero, L. Rodero-Merino, J. Caceres, M. Lindner, " A break in the clouds:towards a cloud definition", ACM SIGCOMM Computer Communication Review, Vol. 39, No.1, 2009, pp. 50–55.

[5] H. Khazaei, J. Misic, and V.B. Misic "Performance Analysis of Cloud Computing Centers," Proc. Seventh Int'l ICST Conf. Heterogeneous Networking for Quality, Reliability, Security and Robustness (Qshine), 2010.

[6] Luis M. Vaquero, Luis Rodero-Merino, Juan Caceres, Maik Lindner, "A Break in the Clouds: Towards a Cloud Definition", ACM SIGCOMM Computer Communication Review, Vol. 39, Number 1, 2009.

[7] P. M. Mell, T. Grace, "NIST definition of cloud computing", Technical Report, SP 800-145, available at http://www.nist.gov/index.html.

[8] Xiong, K., and Perros, H. (2009), "Service Performance and Analysis in Cloud Computing", Proceeding of IEEE World Conference Services, pp. 693-700.

[9] J. Baker, C. Bond, J. Corbett, J.J. Furman, A. Khorlin, J. Larsonand, et. al. "Megastore: Providing Scalable, Highly Available Storage for Interactive Services," Proceeding of Conference on Innovative Data Systems Research (CIDR), pp. 223-234, Jan. 2011.

[10] V. Vineetha,"Performance monitoring in cloud ", Infosys view point, pp.1-8, 2012.

[11] A. Avram, "Measuring and comparing the performance of 5 cloud platforms", available at http://www. infoq.com/news/2010/07/Benchmarking-5-Cloud Platforms, accessed on (12 Jul 12).

[12] B. Harzog, "Infrastructure Performance Management for Virtualized Systems", White Paper, APM Experts, 2010 pp. 1-18.

[13] P.L. Bannerman, "Cloud Computing Adoption Risks: State of Play", 2010, available at www.nicta.com.au/pub?doc=4387(accessed on 12 Aug 2012).

[14] Virtual appliances, "Private cloud computing performance", white paper, 2012.

[15] http://www.netinst.com/products/observer-reporting-server/index.php.

[16] http://copperegg.com/revealuptime-website-monitoring/.

[17] http://www.hyperic.com/products/cloud-status-monitoring.

[18] T.V. Doa, and C. Rotter, "Comparison of scheduling schemes for on-demand IaaS requests", The Journal of Systems and Software, Vol.85, 2012, pp. 1400– 1408.

[19] M.E. Frincu, "Scheduling highly available applications on cloud environments", Future Generation Computer Systems, Volume 32, March 2014, Pages 138–153.

[20] Jiayin Li , Meikang Qiu , Zhong Ming , Gang Quan , Xiao Qin , Zonghua Gu, "Online optimization for scheduling preemptable tasks on IaaS cloud systems", Journal Parallel Distributed Computer, Vol. 72, 2012,pp. 666–677.

[21] T. Kosar, and M. Balman,"A new paradigm: Data-aware scheduling in grid computing", Future Generation Computer Systems, Vol. 25, 2009, pp. 406-413.

[22] J. L. Lucas-Simarro, R. Moreno-Vozmediano, R.S. Montero,I.M. and Llorente, ,"Scheduling strategies for optimal service deployment across multiple clouds", Future Generation Computer Systems, (2012) (in press).

[23] 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.

[24] H. Khazaei, J. Misic, and V.B. Misic "Performance Analysis of Cloud Computing Centers," Proc. Seventh Int'l ICST Conf. Heterogeneous Networking for Quality, Reliability, Security and Robustness (Qshine), 2010.

[25] J. Tordsson, R.S. Monterob, R. Moreno-Vozmedianob, and I.M. Llorente ,"Cloud brokering mechanisms for optimized placement of virtual machines across multiple providers", Future Generation Computer Systems, Vol. 28, 2012, pp. 358–367.

[26] Q. Wu and Y. Gu, "Optimizing end-to-end performance of data-intensive computing pipelines in heterogeneous network environments", Journal Parallel Distributed Computer, Vol. 71, 2011, pp. 254–265.

[27] R.N. Calheiros, A.N. Toosi, C. Vecchiola, and R. Buyya,"A coordinator for scaling elastic applications across multiple clouds", Future Generation Computer Systems, Vol. 28, 2012, pp. 1350–1362.

[28] V. Mauch, M. Kunze, and M. Hillenbrand, "High performance cloud computing," Future Generation Computer Systems, vol. 29, no. 6, 2012, pp. 1319–1644.

[29] R. Ghosh, F. Longo, V.K. Naik, K.S. Trivedi," Modeling and performance analysis of large scale IaaS Clouds," Future Generation Computer Systems, Vol. 29 Issue 5, July, 2013, pp. 1216-1234.

[30] R.R. Expósito, G.L. Taboada, S. Ramos, J. Touriño, and R. Doallo, "Performance analysis of HPC applications in the cloud", Future Generation Computer Systems, Vol. 29, 2013, pp. 218–229.

[31] Amazon Elastic Compute Cloud, User Guide, API Version ed., Amazon Web Service LLC or Its Affiliate, http://aws.amazon.com/documentation/ec2, Aug. 2010.

Author Profile

Jitendra Singh born in 1975 at Kanpur, India. He has completed his masters in computer science from Madurai Kamaraj University and 'C' Level (M. Tech) from National Institute of Electronics and Information technology (NIELIT) India. It is followed by PhD (Computer Science) in the area of cloud computing. He has also qualified the prestigious UGC-NET examination conducted by the UGC of India in the year 2006.

He has more than 11 years of teaching experience during which he has taught to the students of Bachelor and Master level. He is also engage with the Stratford University, USA, India Campus, as a part time faculty from last three years. Currently, he is teaching in PGDAV College, University of Delhi, India. During his teaching career, he has taught various subjects that include Data Structure, C++, Java Core, Networking, Routers, Virtual Private Network and firewalls, DBMS, etc. He has contributed as an author in more than dozen of research papers. Many of them have been published in reputed journals. In addition, he has also written a book titling 'Cloud computing for beginner to researcher' that can be accessed from Amazon.

He is member of various associations that include Institute of Electronics and Telecommunication Engineers, International Association of Engineers, The Society of Digital Information and Wireless Communications (SDIWC). His research interests include cloud computing, Networking, Network Security, etc.