International Journal of Computer Applications (0975 – 8887)
Volume 64– No.15, February 2013

Performance Analysis of Web Applications on IaaS Cloud Computing Platform

Swapna Addamani
Dept of Computer Science & Engg.-R&D Centre
East Point College of Engineering & Technology, Bangalore, India.

Anirban Basu, PhD.
Dept of Computer Science & Engg,- R&D Centre
East Point College of Engineering & Technology, Bangalore, India

ABSTRACT

This paper proposes a queuing model for performance analysis of web applications on a Cloud Environment. In this model, the IaaS cloud computing platform is modeled as multiple queues and the virtual machines VMs are modeled as service centers. The instances act as virtual machines and VMs run on servers, its number decided *a priori* before running an application. The model is based on the Reserved Instances behavior of applications on Amazon EC2 [1]. The web applications running on the Cloud platform were analyzed to determine the statistical distributions of the relevant parameters and experiments were conducted for validation of the model by running the web applications on EC2. Results show that the model though simple can predict the performance accurately.

Keywords

Cloud Computing, Performance Evaluation, Benchmarking, Closed Queuing Network, JMT Tool.

1. INTRODUCTION

Cloud computing provides utility oriented IT services on "pay-per-use" basis [1]. "The Cloud" has a large pool of resources and provides a development platform which can be reconfigured dynamically to adjust to a variable load for better performance and for optimum utilization. Users submit their requests for computing resources such as CPU, RAM, disk, application, infrastructure software, etc. which are provisioned in the cloud platform without the users being aware about the details of execution environment. The important entities involved in Cloud computing are [1]:

- Cloud Server (CS) or Cloud Service Provider (CSP): Is a pool of resources in a data centre and provides huge processing power and variety of services to the Clients.
- Clients who demand services from CS/CSP. The services may include computing power, storage resources, applications, and processes.

Cloud based services have three service models which are:

- Infrastructure-as-a-Service (IaaS),
- Platform-as-a-Service (PaaS)
- Software-as-a-Service (SaaS).

Examples of IaaS cloud include Amazon EC2, IBM Cloud while Microsoft Azure and Google AppEngine are examples of PaaS cloud. Examples of SaaS include Gmail, Google Docs. Understanding the characteristics of computer system performance has become critical for service applications in cloud computing. Commercial success of any Cloud Computing platform depends upon its ability to deliver guaranteed Quality of Services (QoS). Evaluating the performance by measurements, simulation or by modeling is becoming important.

This paper presents an approach for studying performance of IaaS cloud computing platform and proposes a queuing model for performance analysis of web applications modeled as closed class of jobs. The model has been analyzed using JMT tool [19] and validated on EC2. The web applications were analyzed to determine the statistical distributions of the relevant parameters and experiments conducted for validation of the model. Results show that the model though simple can predict the performance accurately.

2 EARLIER WORK

Measurement is used for selecting the hardware and software configuration of any computer system prior to procurement and done by running Benchmark programs. With the increasing usage of cloud computing, more and more users are seeking answers to questions on which type of cloud service they should choose and from which vendor. While work is going on designing appropriate benchmarks, so far Phoronix Test Suite [3] has gained prominence. Performance evaluation of Cloud Computing platforms by measurement using Benchmark programs has been conducted. Jorg Schad, Jens Dittrich, Jorge-Arnulfo Quian´e-Ruiz carried out a study of the performance variance of widely used Cloud infrastructure Amazon EC2 from different perspectives [4]. Micro benchmark programs were used to measure the performance variance in CPU, I/O and Network. In [5] authors focus on performance measurement and analysis of network I/O applications (network-intensive applications) in a virtualized cloud. To maximize the benefit and effectiveness of server consolidation and application consolidation, the authors show that performance improvement can be high by strategically collocating network I/O applications. Results of performance analysis of Cloud Computing Services for Many-Tasks Scientific Computing on four commercial Cloud Computing platforms have been reported in [6] and performance measurement of scientific applications on Amazon EC2 has been carried out in [7]. In [8], Huberet.al. Present experimental results on two popular state-of-the-art virtualization platforms, Citrix Xen Server 5.5 and VMware ESX 4.0, as representatives of the two major hypervisor architectures. Based on these results, they propose a basic, generic performance prediction model for the two different types of hypervisor architectures. Performance Evaluation of Amazon EC2 focusing on non functional properties has been reported in [9].

Simulation has also been one of the evaluation techniques for performance evaluation of Cloud Computing platforms. CloudSim[10] is one of the most popular simulation tools. Performance Evaluation on Open Cirrus Cloud Computing test bed located at the University of Illinois using tools to mimic distributed users and cloud facilities was carried out in 2009 as reported in [11] Application of CloudSim to study Cloud Computing environments has been reported in [12].

Modeling of a Cloud Computing platform for the performance analysis has been done by few researchers. In [13]. Brebner and Liu have used a Service Oriented Performance Modeling Technology for modeling the performance and scalability of Service Oriented Applications architected for a variety of platforms. In [14] Hamzeh et. al. discuss use of M/G/m queues for studying performance and for deriving analytical solutions. In [15], the authors describe a Markov chain based approach for the performance and availability analysis of cloud provided services for IaaS deployment model. In [16], Chen and Li propose a queuing-based model for performance management on cloud and carried out measurements on an experimental set up. Analysis of a cloud system using Queuing Theory and π- calculus for improving the cluster utilization keeping security aspects in mind has been discussed in [17]. Kaiqi Xiong and Harry Perros in [18] have proposed a Queuing Network model for studying the performance of a Cloud Computer especially the response time.

This paper presents a simple queuing model which can accurately characterize the performance of IaaS cloud computing platform for statistical distributions relevant to the web applications constituting the workload.

3. QUEUING MODEL FOR THE CLOUD

An application runs on the virtual machines on cloud using Virtualization software such as Xen, KVM etc. [4]. It is through virtual machines (VM) that the applications share computing resources on a cloud. As shown in Figure 1, when an user sends a request to a web application on cloud, the request is sent to the Cloud Controller. The dispatcher in Cloud Controller forwards the request to the queue of the target application. The instances of the target application run on VMs which act as servers to process the requests in the queue.

The system is modeled as a queue and the VMs as servers. The number of instances needed to be used in the form of VMs is decided *a priori* by the Web Application Developer. The user requests coming from an application join the queue and await service. A Cloud Computing platform can be modeled as consisting of multiple queues for different applications and the Web Application Provider uses the Reserved Instances feature of Amazon EC2 [12] to specify the number of VMs for each application to deliver the QoS as per the SLA with the users. The user requests can be served in many possible orders, such as first come first served, last come first served, shortest processing time first, random order, round robin, and so on. However, the first come first served is still the predominant scheduling method.

When an application is deployed on a cloud, the Cloud Controller forms a queue for it to hold the client requests. VMs are created by Cloud Controller on Cloud Node(s) as decided by the Web Application Provider. The number of VMs created can be specified by service level agreement or an empirical value can be used in case nothing is specified in service level agreement. All the VMs of an application run either on a single Cloud Node or multiple Cloud Nodes, and each of them has the same computing resource. Regardless of the physical location of VMs, there is an instance of the web application running on each of them.

For example, in Figure 1, application A has *n* instances each of which is running on a VM. Similar to A, application B has *m* instances. All the VMs of A and B can run on a single Cloud Node or run on multiple Cloud Nodes, and they constitute two clusters. In Figure 1, Queue A and Queue B on Cloud Controller are associated with applications A and B respectively.

The Closed Queuing Model shown in Figure 2 is used to analyse the behaviour of such a system

4. APPLICATION BENCHMARKS

Two web applications: Tourism Management System and Product Search Engine were run to measure parameters pertaining to the Queuing Model viz, the Think Time and Service Time. These were measured by using software monitors at the appropriate places in the applications. However in this paper, the validation of the model is discussed with reference to results of one application Product Search Engine described below.

4.1 PRODUCT SEARCH ENGINE

In this application, user is allowed to search for the different products in the web after completing registration formalities. If user wishes to purchase any product he/she can proceed further. The product details are stored in the database of application. Once user registers in this web page, he/she can search the products based on interest. This application gives separate login page for each user.

4.2 MEASUREMENT OF PARAMETERS

The web application was run and data collected on parameters which were intrinsic characteristic of the applications and independent of the environment of their execution. The parameters of interest were Think Time and Service Time and these were measured by using software instrumentation techniques. The values collected on the two applications were analyzed for different operations and their characteristics determined using Minitab software [20]. The result of the analysis shows that the Think Time for both the applications follow Weibull Distribution [21] with different values of shape and scale parameter whereas the Service Time follows Exponential Distribution with different values of scale parameter.

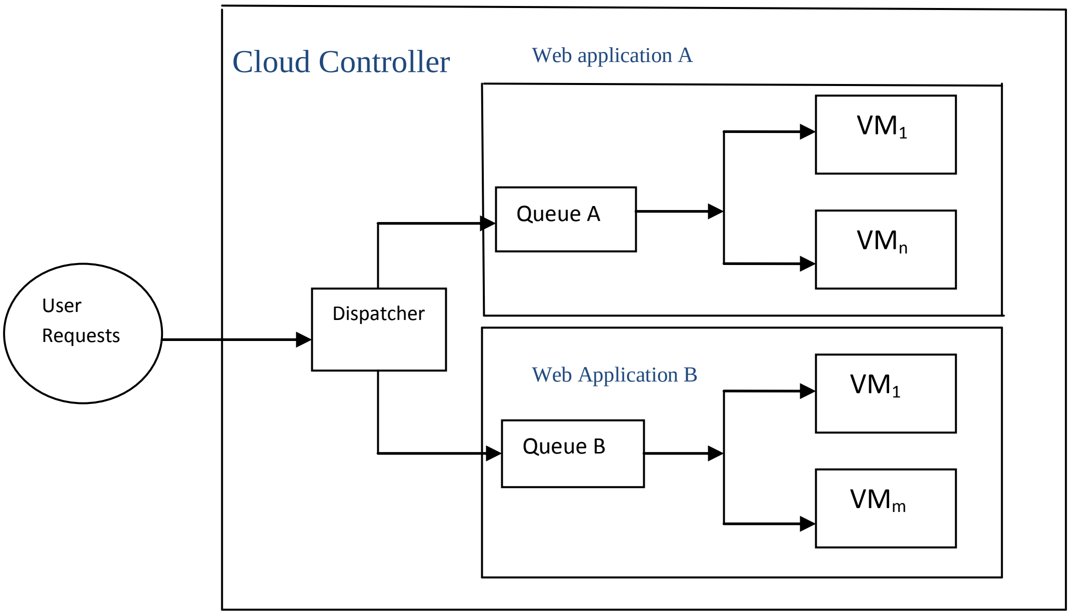

Figure 1:Applications Running on Cloud

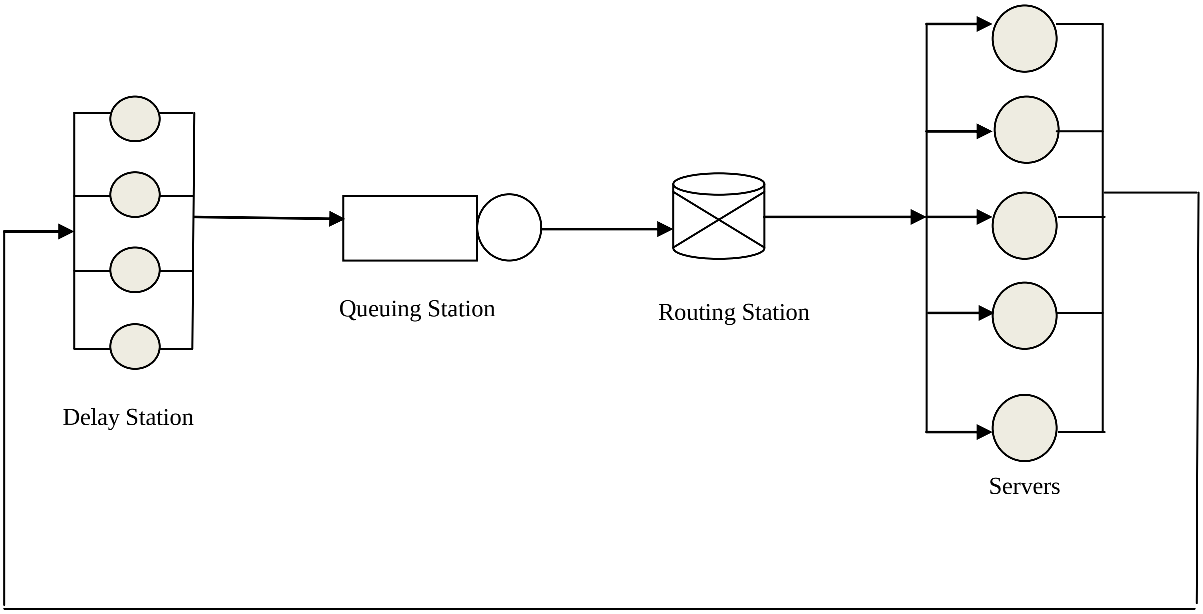

Figure 2: Closed Queuing Model

5. Performance Analysis Using Queuing Model.

The closed queuing model was analyzed using JMT with different distributions of Think Time and Service Time using multi server queuing model shown in Figure 2 using FCFS routing discipline. The number of servers and the number of users were varied and the model analyzed for different values.

The model was analyzed for Exponential distributions of Think Time and Service Time and reported in [22] Since JMT does not cater to Weibull Distribution, for performance analysis, the Think Time was approximated by Gamma Distribution, while Service Time followed Exponential Distribution. The variation of Response Time with Number of Users and with Number of Servers are shown in Figures 3 and 4. The variation of response time with the two parameters is intuitively correct. The model can be used to determine the number of users and number of servers (cloud instances) for an optimal value of the response time.

6. VALIDATION OF THE MODEL

The web applications were run on Amazon EC2 to validate the queuing model formulated. As the Cloud Computing platform EC2 provides IaaS, it displays details of resources consumed by the application and the utilization of resources. The Web Application, Product Search Engine was run and utilization measured for one and two VMs which were compared with the values obtained from JMT for both Exponential Distribution and Gamma Distribution of Think time and for Exponential Distribution of Service Time. It is found that utilization values obtained from EC2 compare favorably with values obtained from JMT with Gamma Distribution of Think Time. Screen shot of the result obtained by running on two VMs is shown in Figure 5.

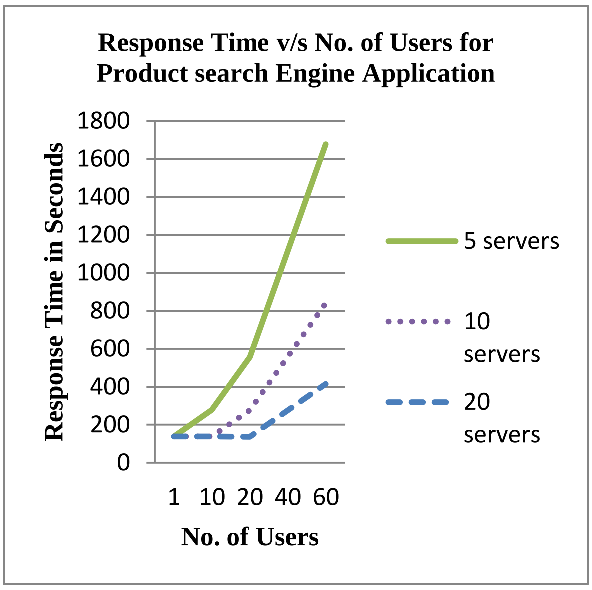

Figure 3 for Gamma Distribution of Think Time

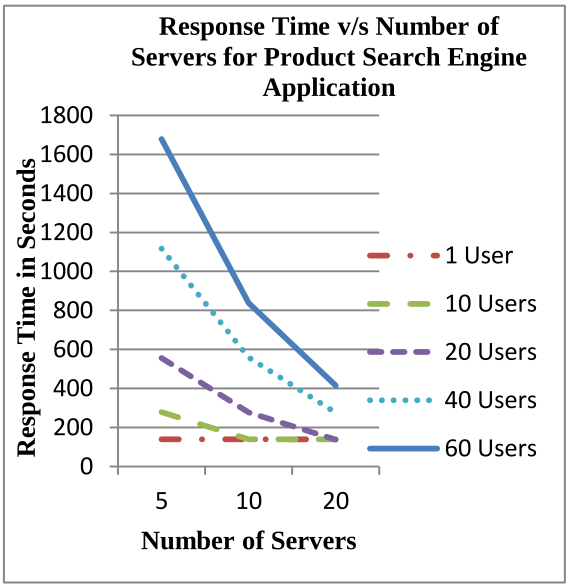

Figure 4 for Gamma Distribution of Think Time

Table 1. Summary of the results obtained is given below

Application	Utilization on JMT for VM=1 with Gamma Distribution of Think Time	Utilization on EC2 for VM=1	Utilization on JMT for VM=2 with Gamma Distribution of Think Time	Utilization on EC2 for VM=2
Product Search Engine	0.98	1.00	0.49	0.50

The results are encouraging and show that the values obtained by analysis of simple Queuing Model closely matched the actual values obtained by running on Amazon's EC2 Cloud for Gamma Distribution of Think Time and Exponential Distribution of Service Time.

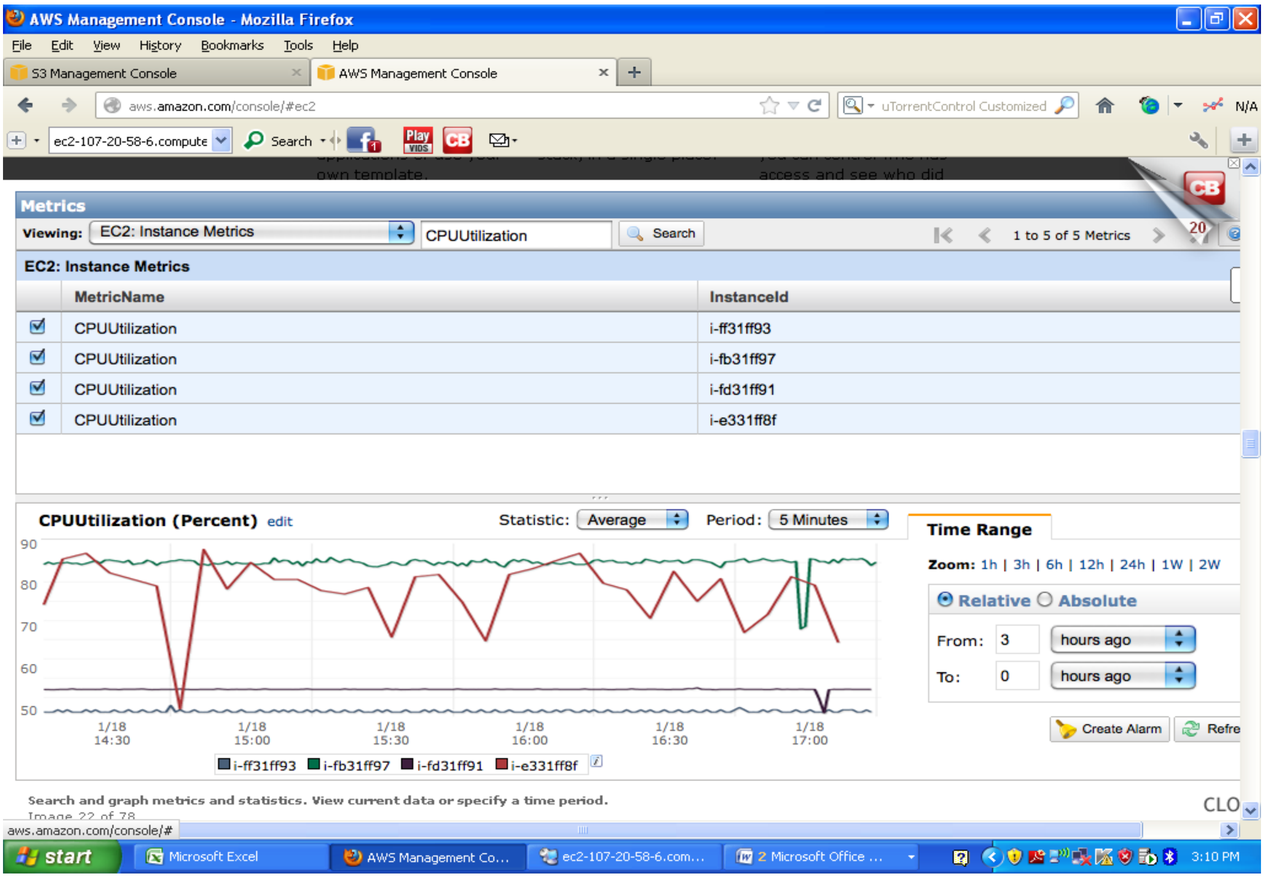

Figure 5 Utilization Graph in Amazon EC2 Cloud

7. CONCLUSIONS

The paper presents a closed Queuing Model for performance analysis of IaaS Cloud Computing Platforms. Software monitoring was used to measure model parameters on two web applications and the model was analyzed using JMT tool. The variation of the response time with the number of users and with the number of instances (or Virtual Machines) was studied. The model was validated by running the application on EC2 and results of validation are encouraging. From the analysis it can be concluded that the Closed Queuing Model proposed in this paper can be used to determine the optimal number of users and the number of VMs for a desirable value of the response time for IaaS Cloud Computing platform.

8. ACKNOWLEDGEMENTS

The project was carried out with assistance from Computer Society of India's Minor project funding scheme 2012.

REFERENCES

[1] Amazon EC2, http://aws.amazon.com/ec2

[2] Rajkumar Buyya and Karthik Sukumar, Platforms for Building and Deploying applications for Cloud Computing, CSI Communications, May 2011, Vol. 35. No. 1. pp. 6-11

[3] www.phoronix-test-suite.com

[4] Jorg Schad, Jens Dittrich, Jorge-Arnulfo Quian´e-Ruiz, "Runtime Measurements in the Cloud: Observing, Analyzing, and Reducing Variance", Proceedings of VLDB Endowment. Vol.3. No. 1. pp. 460- 471.

[5] Y Mei. L Liu, X Pu, S Sivathanu. "Performance Measurements and Analysis of Network I/O Applications in Virtualized Cloud". Proc. IEEE International Conference on Cloud Computing, Miami, Fl, July 2010

[6] Alexandru Iosup, Simon Ostermann, M. Nezih Yigitbasi, Radu Prodan, Thomas Fahringer, and Dick H.J. Epema, "Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing", IEEE Trans. On Parallel and Distributed Systems, Vol. 22, No.6. June 2011, pp. 931-945.

[7] Keith R. Jackson, Lavanya Ramakrishnan, Krishna Muriki, Shane Canon, Shreyas Cholia, John Shaif, Harvey Wasserman and Nicholas Wright, " Performance Analysis of High Performance Computing Applications on the Amazon Web Services Cloud". Proc. IEEE Second International Conference on Cloud Computing Technologies Science, 2010.

[8] Nikolaus Huber, Marcel von Quast, M Hauck and S Kounev " Evaluating and modeling virtualization performance overhead for cloud environments", 2011.

[9] Vladmir Stantchev, "Performance Evaluation of Cloud Computing offerings". Proc. 2009 Third International Conference on Advanced Engineering Computing and Application in Sciences, pp. 187-192

[10] Rajkumar Buyya, Rajiv Ranjan and Rodrigo N. Calheiros1,"Modeling and Simulation of Scalable Cloud Computing Environments and the CloudSim

Toolkit: Challenges and Opportunities", Proc. of International Conference on High Performance Computing & Simulation, 2009, June 2009. pp. 1-11

[11].Ahmed Khurshid, Abdullah Al-Nayeem, and Indranil Gupta, " Performance Evaluation of the Illinois Cloud Computing Testbed", Technical Report, Department of Computer Science University of Illinois at Urbana-Champaign, 2009

[12] Rodrigo N. Calheiros, Rajiv Ranjan, and Rajkumar Buyya," Virtual Machine Provisioning Based on Analytical Performance and QoS in Cloud Computing Environments", Proc. International Conference on Parallel Processing (ICPP), Sept2011pp.295-304.

[13].Paul Brebner and Anna Liu,"Performance and Cost Assessment of Cloud Services", Proc of Int. Conference on Service Oriented Computing, 2010, pp. 39-50

[14] Hamzeh Khazaei,. Jelena Misic and Vojislav B Misic, "Modelling of Cloud Computing Centers Using M/G/m Queues", Proc. 2011 31st International Conference on Distributed Computing Workshops, pp.87-92

[15] K S Trivedi, Rahul Ghosh and Vijay K Naik, "Performance and Availability Analysis for Infrastructure as a Service Cloud", Proc. CONSEG 2011, International Conference on Software Engineering, Bangalore, Feb 2011, pp. 106-111

[16] Hao-peng CHEN, Shao-chong LI,"A Queuing-based Model for Performance Management on Cloud", Proc. 2010 6th International Conference on Advanced Information Management and Service (IMS), pp. 83-88

[17] A M Alijohani, D R W Holton, I Awan and J S Alanazi, "Performance Evaluation of Local and Cloud Deployment of Web Clusters". Proc. 2011 International Conference on Network-Based Information Systems, pp.274- 278

[18] Kaiqi Xiong and Harry Perros, "Service Performance and Analysis in Cloud Computing", Proc. of the IEEE Congress on Services-IEEE Cloud 2009.

[19] Marco Bertoli, Giuliano Casale, Giuseppe S, JMT: performance engineering tools for system modeling, ACM SIGMETRICS Performance Evaluation Review, Volume 36 Issue 4, March 2009, pp. 10-15

[20] www.minitab.com

[21] K S Trivedi, Probability and Statistics with Reliability, Queuing and Computer Science Applications, Wiley, 2001

[22] Sapna Addamani and Anirban Basu "Performance Analysis of Cloud Computing Platform", International Journal of Applied Information Systems ISSN:2249-0868, Vol.4, No. 4, October 2012, pp.29- 33