Performance Analysis of services in Cloud Computing

[1]SeemaChowhan,[2]Ajay Kumar,[3]Shailaja Shirwaikar
[1]Assistant Professor,[2]Director,[3]Associate Professor
[1]Department of Computer Science,
[1]Baburaoji Gholap College, Sangvi,Pune,India

*Abstract:*Cloud computing is internet based technology in which cloud provider provides the operating system, application software on the top of hardware as resources to its customer and delivers it as a service to the end users over the internet. Cloud makes it possible to access application and data any time anywhere full filling Quality of service parameters (QoS) like performance, availability and reliability. Service provider supplies infrastructure including hardware and software as a Virtual Machine. The Challenge for a customer is to identify requirements at different workloads and choose an appropriate set of VMs that give a satisfying performance to its end users in terms of response time. To overcome this, we present an architecture able to monitor the load (mainly web-browser applications) with single computing virtual machines. This is accomplished by monitoring the system for single user and multi user load variations. The analysis of this data can be used by customer to negotiate and provision resources appropriately as per demand and type of services.

***Index Terms* - Quality of Service, Virtual Machines.**

I. INTRODUCTION

Cloud computing is a paradigm shift to computing [1]. Stakeholders of cloud computing are providers, cloud computing customers and end-users [2]. Cloud provider provides resources from their Data-Center to cloud customers and cloud customer uses this resources to provide services to its customers. End users use services of cloud customers. For example any online shopping website owner uses AMAZON EC2 for hosting their website. The people who do online shopping are end users. Customer uses resources from a cloud Data-Center instead of organization‘s own on-premises servers. This increases reliability and availability of services and also improves performance thus providing a satisfying experience to its end-users. The performance measures used by customer are response time, throughput and so on. Performance depends on many factors such as type of service, the number of users using the service as also the load at network level as well as cloud provider's end.

For example during festival seasons, number of end users increase for on-line shopping service. To sustain end user experience, good provider provisions and releases resources as per load. Resource provisioning depends on the resource requirements envisaged by the customer. Service level Agreements between cloud provider and the customer specify the resources and quality levels required for the execution of job in order to minimize the cost from customer perspective and to maximize the resource utilization from provider’s perspective [3]. The customer need to understand its requirements for cloud based services. The physical resources of a physical machine (PM) such as CPU, memory, network and disk I/O required by a service decides the cost of service. Software resource requirements include operating system and plethora of other services. Virtualization Technology has been the backbone of cloud computing as it bundles hardware and software resource requirements into a Virtual Machine(VM) of appropriate capacity so that it can be used as a unit of resource allocation.

An important challenge for providers of cloud computing services is the efficient management and monitoring of Virtual Machines. Monitoring includes observing server's system resources like CPU Usage, Memory Consumption, I/O, Network, Disk Usage, Process etc. The Challenge for a customer is to demand VMs that match its fluctuating requirements and yet provide satisfying experience to its end users. Before getting into Service level agreements, customer need to identify parameters that dictate the resource loads.

In this paper, an analytical model is presented for customer to understand customer requirement for provisioning resources for web-based application which is hosted on cloud. As a case study a document management system designed on Amazon public cloud is used for running different services and to monitor the performance of VM server for different CPU usage parameters. Rest of the paper is organized as follows,Section I contains the introduction , Section II contain the background and related work on resource monitoring and prediction, Section III contain architecture and implementation, Section IV contains experimental results, Section V concludes research work with future directions.

II. BACKGROUND AND RELATED WORK

Cloud is Independent but commonly available online utility on demand, irrespective of location, that is accessible through web .Cloud computing is browser based application to access software, platform and infrastructure as a service that is hosted on multiple servers which possesse massive amount of storage capacity [4] [5]. Cloud computing technology is used in our day to day life as Google applications and email like Gmail, Yahoo and Hotmail. Quality of service parameters documented in a contractual form agreed between provider and customer, are called service level agreements (SLAs) [6], [7], [8] which ensures delivery of QoS parameters, such as availability, reliability, response time and throughput to the users as per signed agreement

[9]. Virtual machine is a logical partition or container that runs on a host machine. Multiple VMs can run on the physical server. Resources are dynamically assigned to VMs from available resource pool. VMs are used as physical server with all functionality such as accessing the Operating systems, CPU, and hard disk from common resource pool. As per user requirement cloud provider scales up by allocating more VMs and scales down by deallocating provisioned VMs. Vilaplana et al [10] presents a cloud-based system architecture that emphasises on the scalability problem in cloud-based systems. Changing computing requirements for web-based application are managed by considering response time as QoS parameter with web-based load variability. Resource management in cloud computing environment includes resource discovery, resource allocation, task scheduling and load balancing. Thus for better system performance it becomes important for cloud provider to do accurate monitoring of the resources which are consumed by the customer to provide satisfactory services. The essential element of resource management is the discovery process which involves searching for the suitable resources that match the application requirements of customer [11]. The resource discovery and allocation process is managed by the cloud service provider where user will not be aware of performance of resource in advance. So it becomes equally important for customer to monitor resources before provisioning so that it can ask for the resources as per its requirement rather than provider providing predefined resources. This will minimize resource cost for customer.

The researchers emphasises on possible resource prediction associated to virtual infrastructure based on the analysis of resource log data in cloud Data-Center. Mallick et al proposed Markov Chain model which uses a set of historic data to predict the system usage and resource requirement in future [12]. Ge, et al proposed a frame work for resource monitoring by a script running on physical server and monitored through VM[13]. Chenet al implements an adaptive resource monitoring framework and prediction mechanism using vector auto regression prediction algorithm [14].

III. ARCHITECTURE AND IMPLEMENTATION

Parameters that assist customers to select number of VMS as per their requirements are described in Table 1 for a document management system. Some of them are type of services, time slot of service and single or multi-user service. The parameters can vary as per application requirement of users.

Table 1. Usage load parameters for a Document Management system

Parameters	Service
Type of service	• Upload • Download • Delete • Search
Time slot of service	Time of the day service is being used.
Single user/multi user	The services can be performed either single user or multiple users

A. Experimental setup

Amazon EC2 provides wide range of VM instances which include varying mixtures of CPU, memory, storage, and networking capacity. It provides flexibility to user application to select mix of resources. For the experiments, Amazon micro instance with following configuration is used.

Processor: Intel(R) Pentium(R) CPU G620 2.60GHz
RAM:1 GB
Space: 8 GB
System Type: 64-Bit operating system

On the above mentioned VM instance Ubuntu 16.04 operating system and software's PHP, HTML, MYSQL version-14.14, JS,Css, putty, pscp, winscp, puttygen and Apache2 is installed. Application for document management system is developed which includes different user with different permissions on web services created for upload, download, Search and delete. In the system multiple users can perform different web operations on various documents as per their need. The system architecture and working is presented in Figure-1

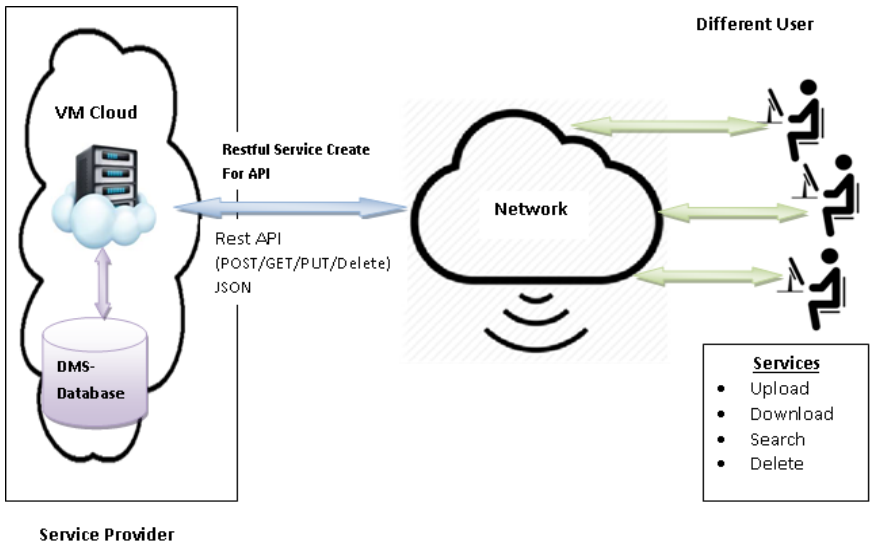

Figure-1 System Architecture

The performance of different types of services, used on the cloud, are monitored using Linux monitoring commands such as Sar, mpstat, and VMstat.

IV. EXPERIMENTAL RESULTS

In this section several experiments are performed on VM server to understand significance and to get values for different CPU parameters for monitoring and observation.

A. Multiple users performing different services

In this experiment ten users performs upload, download services on file size of 1 MB and 4 MB to get time required for user running application on VM. The time required for different users are plotted in Table-2 for upload service, Table-3 for download services.

Table-2 %User time for upload service

Users	%User-Time (Seconds)	
	File-Size (1MB)	File-Size (4MB)
1	0.294	0.28
2	0.13	0.15
3	0.16	0.207
4	0.18	0.19
5	0.25	0.39
6	0.37	0.36
7	0.28	0.406
8	0.48	0.45
9	0.43	0.53
10	0.49	0.543

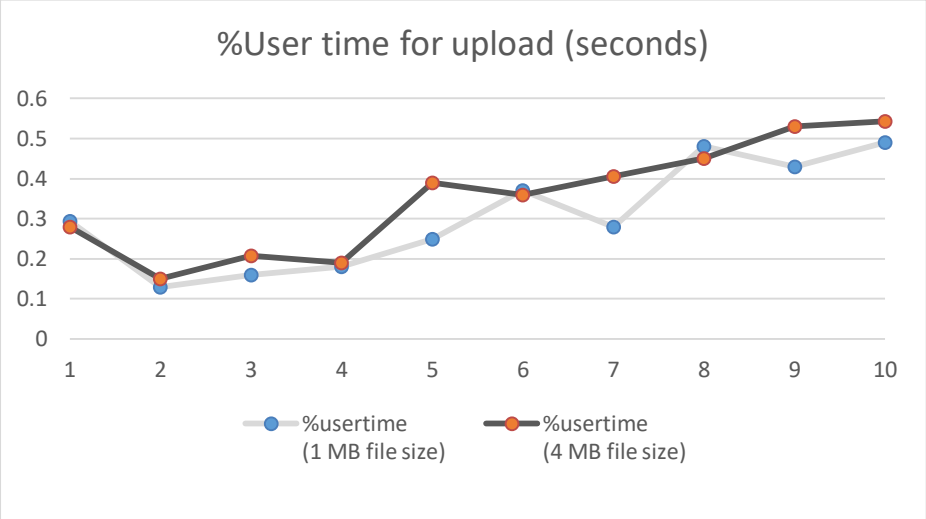

Figure-2:%User-Time for upload service

Table-3: %User-Time for download service

Users	%User-Time (Seconds)	
	File-Size (1MB)	File-Size (4MB)
1	0.23	0.1
2	0.03	0.17
3	0.13	0.87
4	0.17	0.19
5	0.27	0.39
6	0.29	0.42
7	0.31	0.44
8	0.38	0.55
9	0.43	**0.52**
10	0.49	0.57

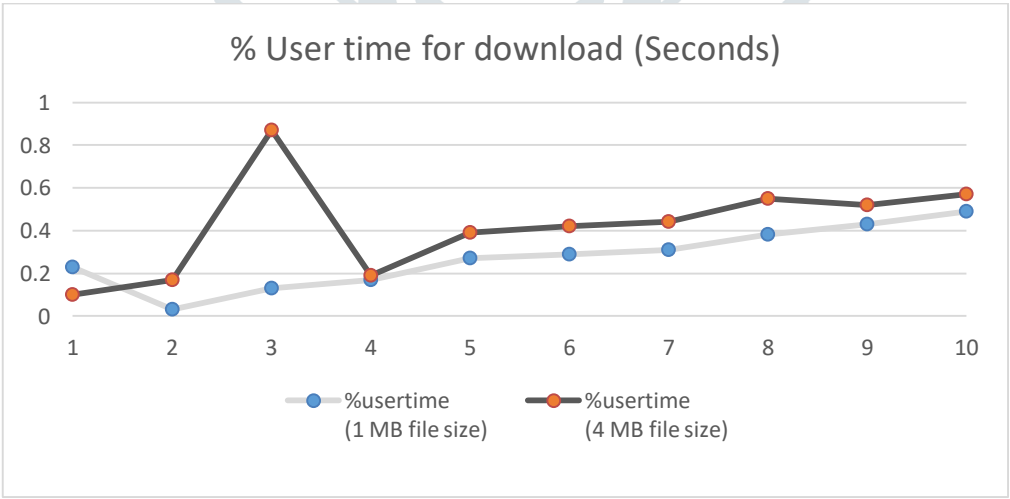

Figure-3: %User- Time (Seconds) for download service

Observation: As presented in Figure-2 and Figure-3 CPU time required for upload and download services depend on file size and number of users simultaneously accessing application. %User-Time is more for 4MB file than 1MB file.However the relationship is almost linear with respect to number of users. The variations are less for download service as compared to upload service.

B. Day wise Traces of different services:

Five day wise traces are taken for different services like upload and download on different file size ranging from 50KB to 1950KB with difference of 300KB.

Table-4: Five day traces of %User-Time for upload service

Days	%User-Time (Seconds)					
	U-50KB	U-350KB	U-650KB	U-1250kb	U-1650KB	U-1950KB
1	0.762	0.208333	3.795	0.134	0.152	0.1875
2	8.213333333	0.128333	8.3	0.12	0.116	**37.292**
3	8.738571429	0.1475	**42.53**	24.2625	0.152	7.498
4	4.33	0.144	26.6875	14.0675	**46.585**	0.1375
5	2.06	0.192	0.115	28.5975	0.128	0.115

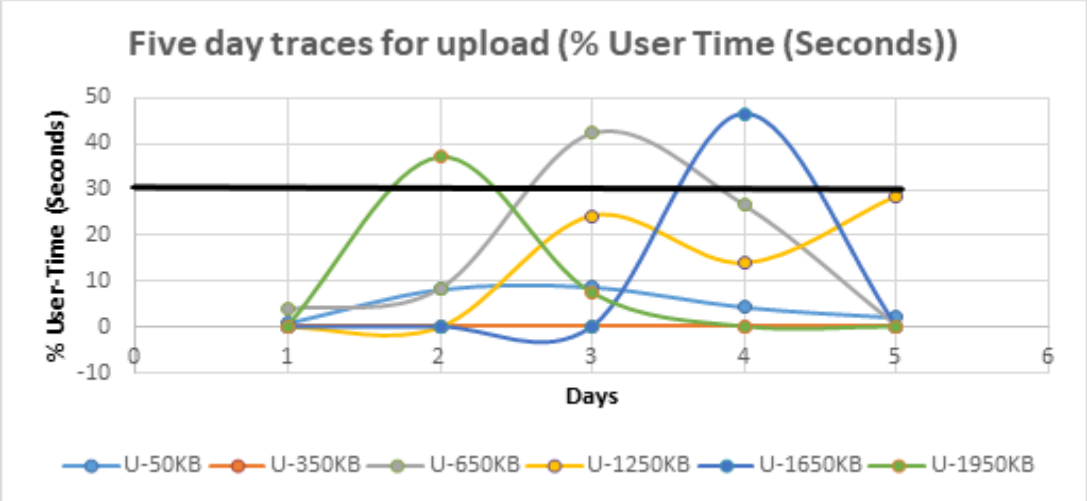

Figure-4: Five day traces for upload (%User- Time (Seconds))

Table-5: Five day traces of %User-Time for download service

Days	%User-Time (Seconds)					
	D-50KB	D-350KB	D-650KB	D-1250KB	D-1650KB	D-1950KB
1	13.43	6.6975	10.634	4.34	8.321	25.67
2	29.568	29.188	25.45	29.77	22.96143	47.98333
3	28.28333333	17.632	21.99	26.22	23.9525	23.55
4	18.58	30.7675	29.0825	17.975	**43.2325**	31.9825
5	16.25	**52.066**	**37.54**	**37.264**	24.106	37.54

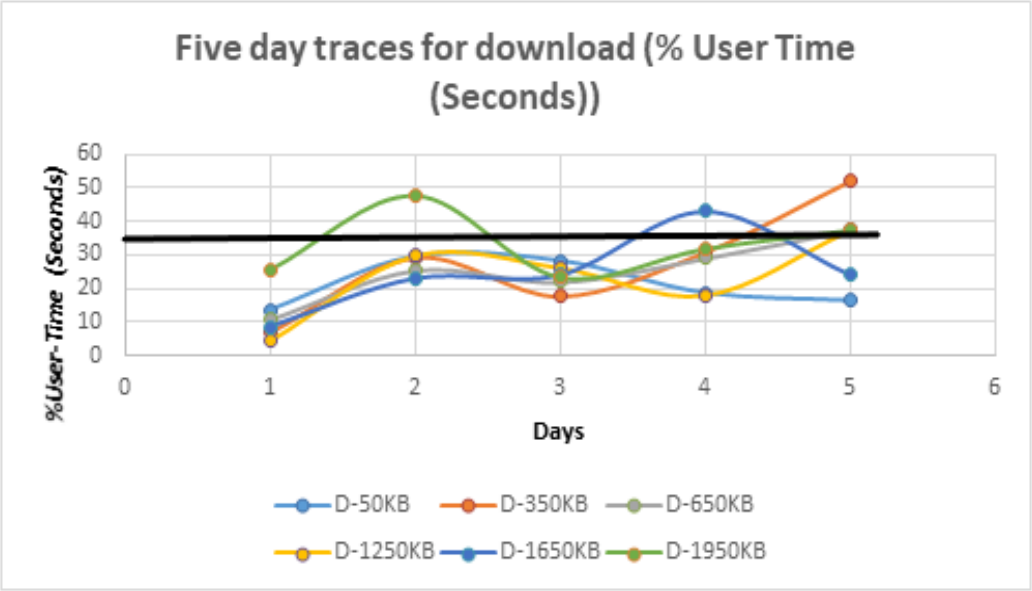

Figure-5: Five day traces for upload (%User- Time (Seconds))

Observation:

The performance of upload service are presented in Table-4 and Figure-4. Upload services perform poorly for file size 1950KB on day2,650KB on day3 and 50KB on day4.For rest of file size performance is equally well with slight difference in %User-Time.

The performance of download service are presented in Table-5 and Figure-5. Upload services perform poorly for file size 1950KB on day2,350KB on day5 and 50KB on day4.For rest of file size performance is equally well with slight difference in %User Time.

The upload and download operation performance doesnot depend on file size or its day of the week but also on the time of the day. This is because the time taken by the operations depends on transmissiontime, processing time, traffic condition on network and bandwidth. Upload and download speed is also affected by the distance between computer and the server that is either sending or receiving data.Though one cannot predict the time taken directly as an action of file size or day of the week but the values lie in a certain range.Almost all the values are within the range of 30 seconds with few outliers as shown in Figure-4 and Table-5

C. Time Instance Traces of different services

The upload, download, search and delete services are executed by single user on VM server at different time intervals and % User-Time are monitored for file size of 50KB in half an hour time slots.

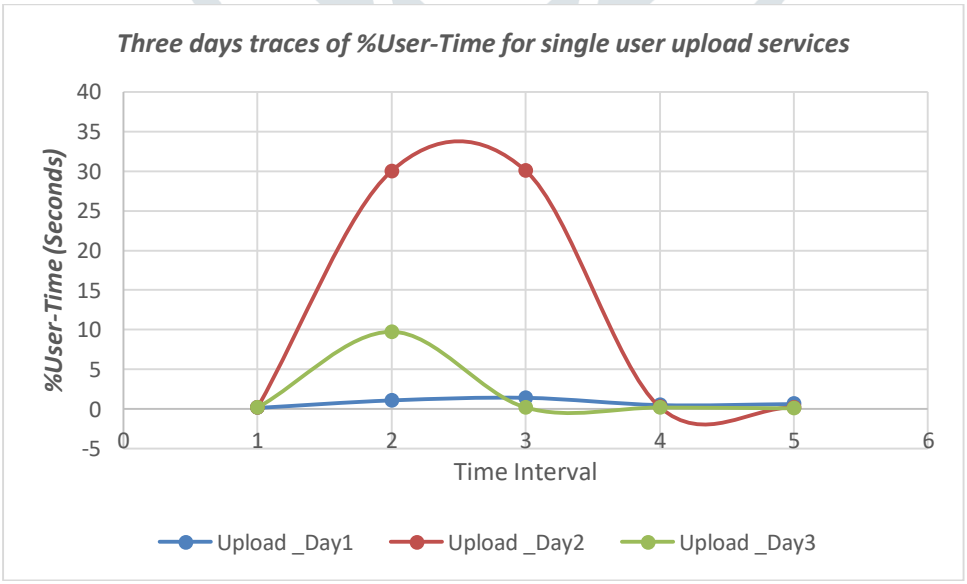

Figure-6: Three days traces of %User-Time for single user upload services

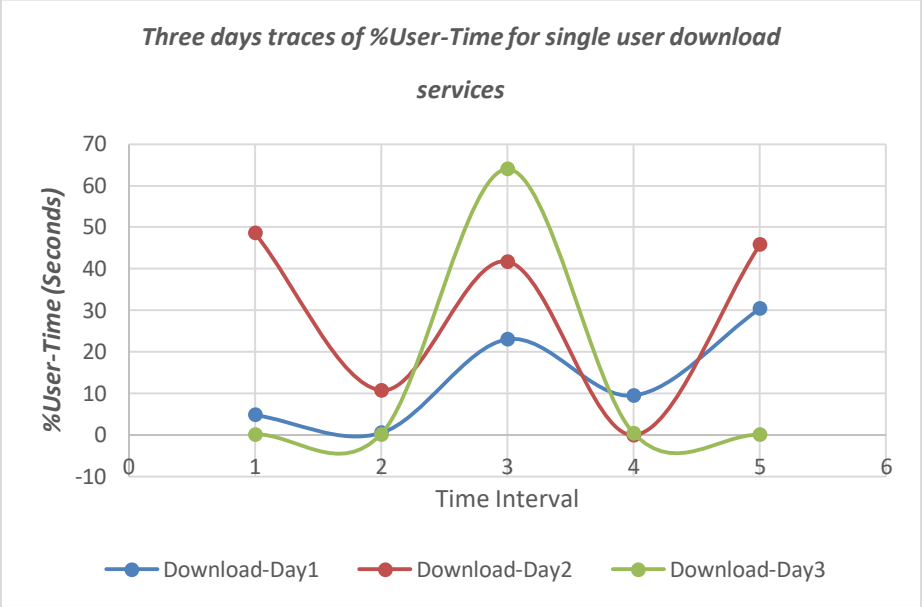

Figure-7: Three days traces of %User-Time for single user download services

Observation:Upload operationas presented in Figure-6 gives better %User-Time on day1 as compared to day2 and day3. Upload operations give poor results on day2.Download operationas presented in Figure-7 gives better %User-Time on day1 as compared to day2 and day3. Download operations give poor results on day3. Search and delete operation time depends on the content and are I/O bound.

V. CONCLUSION

The monitored data is analyzed by customer for planning and forecasting resources for provisioning as per application requirement. It allows to see at a high load if the hardware resources are being used fully, how much memory is active so that more VMs can be provisioned.The experiment results shows that while there is a linear relationship between number of active users and the time taken, however it does not directly depend on file size, time of the day and type of service. One of the limitation of experiment was limit on the file size.

REFERENCES

[1]Voas, J., & Zhang, J. (2009). Cloud computing: New wine or just a new bottle? *IT professional*, *11*(2), 15-17.

[2]Michael Armbrust Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy H.Katz, Andrew Konwinski, Gunho Lee, David A. Patterson, Ariel Rabkin, IonStoica, and MateiZaharia. Above the clouds: A berkeley view of cloud computing. Technical Report UCB/EECS-2009-28, EECS Department, University of California, Berkeley, Feb 2009.

[3]Byun, E. K., Kee, Y. S., Kim, J. S., &Maeng, S. (2011). Cost optimized provisioning of elastic resources for application workflows. Future Generation Computer Systems, 27(8), 1011-1026.

[4]Xiong, K., &Perros, H. (2009, July). Service performance and analysis in cloud computing. In *2009 Congress on Services-I* (pp. 693-700). IEEE.

[5]Zhang, Q., Cheng, L., &Boutaba, R. (2010). Cloud computing: state-of-the-art and research challenges. Journal of internet services and applications, 1(1), 7-18.

[6] John, M., Gurpreet, S., Steven, W., Venticinque, S., Massimiliano, R., David, H., & Ryan, K. (2012). Practical Guide to Cloud Service Level Agreements.

[7] Wu, L., &Buyya, R. (2012). Service level agreement (sla) in utility computing systems. IGI Global, 15.

[8] Liu, F., Tong, J., Mao, J., Bohn, R., Messina, J., Badger, L., & Leaf, D. (2011). NIST cloud computing reference architecture. NIST special publication, 500(2011), 292.

[9] Bianco, P., Lewis, G. A., & Merson, P. (2008). Service level agreements in service-oriented architecture environments (No. CMU/SEI-2008-TN-021). Carnegie-Mellon Univ Pittsburgh Pa Software Engineering Inst.

[10]Vilaplana, J., Solsona, F., Mateo, J., &Teixido, I. (2013, December). SLA-aware load balancing in a web-based cloud system over openstack. In *International Conference on Service-Oriented Computing* (pp. 281-293). Springer, Cham.

[11] Buyya, R., &Ranjan, R. (2010). Federated resource management in grid and cloud computing systems. *Future Generation Computer Systems, 26*(8), 1189-1191.

[12] Mallick, S., Hains, G., & Deme, C. S. (2012, July). A resource prediction model for virtualization servers. In *2012 International Conference on High Performance Computing & Simulation (HPCS)* (pp. 667-671). IEEE.

[13] Ge, J., Zhang, B., & Fang, Y. (2010, October). Research on the resource monitoring model under cloud computing environment. In *International Conference on Web Information Systems and Mining* (pp. 111-118). Springer, Berlin, Heidelberg.

[14] Chen, H., Fu, X., Tang, Z., & Zhu, X. (2015, July). Resource monitoring and prediction in cloud computing environments. In *2015 3rd International Conference on Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence* (pp. 288-292). IEEE.

Authors Profile

Ms. Seema Chowhan is working as a faculty and head in subject of computer sciencein Baburaoji Gholap College Pune, India affiliated to Savitribai Phule Pune University, Pune.She has 18+ years of experience in teaching UG and PG courses. She has completedM.Phil. (CS).Her research interests include Cloud Computing and Networking.

Dr. Ajay Kumar experience covers more than 26 years of teaching and 6 years of Industrial experience as IT Technical Director and Senior Software project manager. He has anoutstanding academic career completed B.Sc. App. Sc. (Electrical) in 1988, M.Sc. App.Sc. (Computer Science-Engineering and Technology) in1992 and PhD in1995. Presently,working as Director at JSPMs Jayawant Technical Campus, Pune (Affiliated to Pune University). His research areas are Computer Networks, Wireless and Mobile Computing,Cloud computing, Information and Network Security. There are 74 publications at National and International Journals and Conferences and also worked as expert, appointedby C-DAC to find Patent-ability of Patent Applications in ICT area. Six commercialprojects are completed by him for variouscompanies/ Institutions. He holds variety ofimperative position like Examiner, Member of Board of Studies for Computer and IT,Expert at UGC.

Dr. Shailaja Shirwaikar has a Ph. D. in Mathematics of Mumbai University, Indiaand worked as Associate Professor at Department of Computer Science, Nowrosjee WadiaCollege affiliated to Savitribai Phule Pune University, Pune for last 27 years. Her researchinterests include Soft Computing, Big Data Analytics, Software Engineering and CloudComputing.