

A Review on Cloud Performance Techniques

Rajat¹, Dr. Ajay Kumar Bharti²

M. Tech., Dept. of Computer Science, Maharishi University of Information Technology, Lucknow, India¹

Professor, Dept. of Computer Science, Maharishi University of Information Technology, Lucknow, India²

Abstract: Cloud is defined as “A bunch of commodity computers networked together in the same or different geographical locations, operating together to serve a number of customers with different need and workload on demand basis with the help of virtualization”. Cloud services are provided to the Cloud users for some of the useful services including telephone, electricity and water via “pay-as-you-use business model”. Administering the Quality of Service is the primary task of the Cloud Applications. However, maintaining the performance, scalability, availability, reliability and dependability is the challenging task for the Cloud Service Providers.

Keywords: Cloud, Cloud Computing, Performance Analysis

I. INTRODUCTION

One of the latest developments in the computer world is Cloud Computing. The features of this application have attained inherent value and have a role in managing ones everyday life. Kleinrock (1969) predicted that As of now, computer networks are still in their infancy. But as they grow up and become more sophisticated, we will probably see the spread of 'computer utilities' which, like present electric and telephone utilities, will service individual homes and offices across the country.

The advancements in Grid Computing Technologies formed a basis for the evolution in Cloud Computing (Girichidis et al., 2011). Eric Schmidt, the CEO of Google is the first to promote the term Cloud Computing in 2006. Though Cloud Computing has its roots in the past with a new social and technical perspective, the out sprout is very recent. If the design of a cloud is considered, it is built naturally on an existing grid based structure.

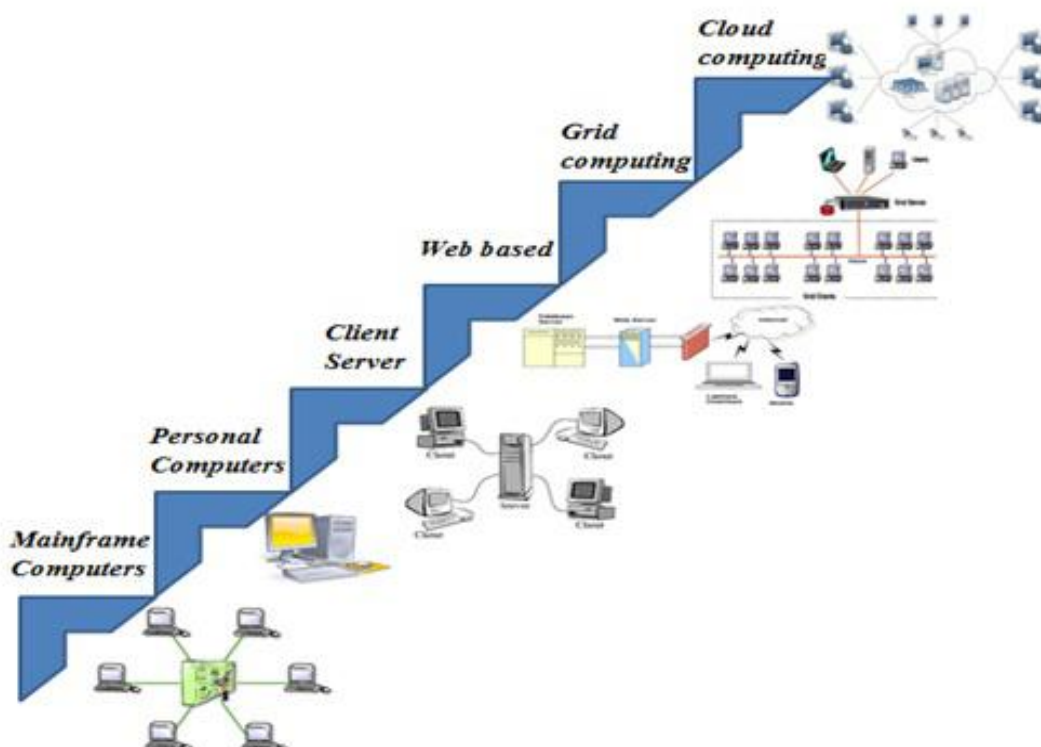


Fig: Evolution of Cloud Computing

A paradigm shift that took place in the world of computing during the last fifty years is demonstrated in Fig.1.1. The figure classifies, in particular, six different and distinct stages. In the first stage, terminals were used to connect to

mainframe computers which were shared by numerous users, and in those days terminals were bigger than monitors and keyboards. In the second stage, PCs i.e., stand-alone Personal Computers turned out to be more powerful and satisfied the everyday work of the users. The advantage of these computers is, a mainframe need not be shared with anyone. In the third stage, numerous computers were allowed to connect to one another in computer networks. Works can be carried out in a PC and the resources from the PC can be shared to other computers by connecting them through local networks. In the fourth stage, other local networks were connected by making use of the local networks in order to establish a global network to enable the users to connect with the internet with a view to operating distant applications and to utilize the available matters. In the fifth stage, electronic grid concept was brought to disseminate distributed computing storage and power resources. PCs were utilized to acquire a network of computers very transparently. Presently, at the sixth stage, we are able to get access to all the resources available on the Internet due to the arrival of Cloud Computing in a simple as well as scalable manner.

II. REVIEW OF LITERATURE

Cloud frameworks can solely turn into tenable substitutions for conventional endeavor information systems whether a legitimate level of performance is contributed. For that reason researchers initiated assessing performance of Cloud services instantly alike services become effective in the interest to render a principle to potential consumers for building resolutions for adopting the Cloud service. Most of the experiments conducted by means of test-bed on a Cloud framework for analyzing the measurement based performance. The experiments administered for measurement-based methods for evaluating the performance on a cloud test-bed involves the subsequent widespread technique. Initially the researchers must state the intention and extent of the assessment, and next detect the characteristics of the services provided by cloud to be valued. The subsequent step is to define the performance metrics to be examined and applicable benchmarks applications for analysis to be proven. At the moment, experimental setup is shaped then the experiments can be executed.

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. Various performance metrics may be used for evaluating different features of Cloud services.

High Performance Computing (HPC) for scientific applications typically requires large amounts of computational and storage resources as well as network bandwidth; thus are particularly challenging to Cloud services. Therefore evaluating performance of commodity commercial Cloud services for supporting high-performance scientific applications attracted much research attention since Cloud services appeared. Evangelinos and Hill (2008) tested Amazon EC2 service performance and found that EC2 may support small sized responsive on-demand HPC applications.

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. The authors found that although performance of EC2 is not equivalent to a traditional HPC system, it is reasonable given the resource available. Iosup, Ostermann, Yigitbasi, Prodan, Fahringer and Epema (2011) analysed the performance of Cloud services for Many- Task Computing (MTC) workloads, which are also loosely coupled applications. Performance of four commercial Cloud services, Amazon EC2, GoGrid, ElasticHosts, and Mosso, were measured. The obtained results indicate that the tested Cloud services need an order of magnitude in performance improvement for supporting high performance MTC scientific applications.

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.

A representative work on Cloud performance evaluation with elastic scaling strategies was presented (Hwang et al., 2016). In this research work, the authors tested IaaS Cloud services under scaleout and scale-up workloads through benchmark experiments conducted on Amazon EC2. The obtained results indicate that scaling-out instances have much lower overhead than those experienced in scale-up experiments, while scaling up is more cost-effective under sustaining heavier workload. Evaluation of the performance of Cloud computing platform services may need to handle some special issues, including performance metrics and benchmarks appropriate for PaaS.

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 launched the storage-optimized instance family that includes High I/O quadruple extra-large (H1) instance and high storage eight extra large (H1) instance.

Expósito 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 state-of-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 Open Shift, 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).

The rapid developments of Cloud computing technologies together with success of the Cloud business model have enabled a wide spectrum of Cloud services offered by a large number of vendors. These services have been employed by numerous users for supporting various applications, including business and e-commerce applications in addition to scientific computing, which have highly diverse performance requirements.

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 test beds for evaluating service performance, some research Cloud test beds have also been constructed and utilized for Cloud service performance evaluation.

Science Cloud (Keahey et al., 2008) is a Cloud test bed 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 test bed comprising of federated heterogeneous distributed data centers. It enables researchers to exchange data sets and develop standard Cloud computing benchmarks. Virtual machine management in Open Cirrus 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 test bed 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 test beds.

III. CHARACTERISTICS CLOUD COMPUTING

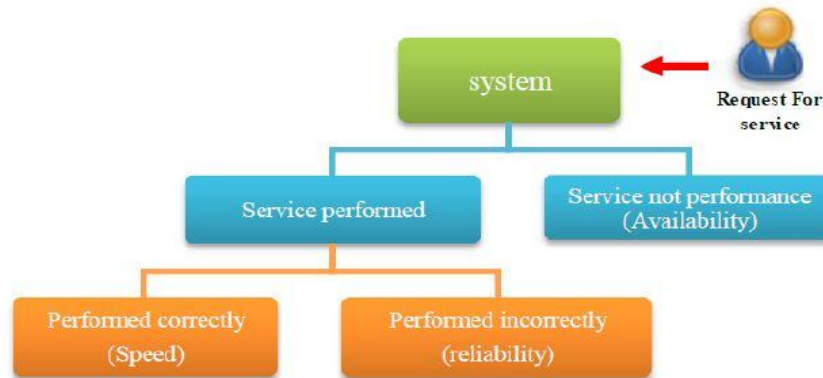
Cloud computing contest 5 essential characteristics that delineate key functions offered by cloud, 3 service paradigms that provide level of service, 4 deployment paradigms that specify where cloud is organized and who can access to it. Key distinctiveness of cloud is as follows:

- **On-demand self-service:** customers belong to cloud handle computing sources in on-demand parameter that they only pay for resources what they utilize. A fundamental clause that a cloud provider has got to accomplish is the capacity to carry computing sources on every occasion the customer requests them. As of the customer's position of view accessible computing possessions are virtually unlimited (i.e., customer is not restricted group of server machines positioned at one location. It is conscientiousness of cloud supplier to have adequate resources to convince necessities of the entire 4 patrons). Make the most of computing sources on demand is one of mainly preferred qualifications for huge number of business enterprises because it exterminate prerequisite for planning forward, procuring, mounting resources they will demand at specific point in view. This assists customer to let alone building superfluous honest savings in server systems. Moreover, as soon as comparing cloud model of computing with time-honored model of holding servers, cloud computing will prevaricate expenditure of having underneath consumed resources.
- **Broad network access:** Computing sources provided via cloud could be accessed by every common service using thin otherwise thick client-machines like laptop, Personal Digital Assistant, mobile-phones etc.
- **Resource Pooling:** Cloud supplier supply group of computing-sources to numerous consumers as per claim. Consumer with admission facility does not contain information about accurate site of cloud however could be proficient to supply site by superior generalization plane like country, state, data center (Himanshu et al., 2009).
- **Rapid-elasticity:** Computing-sources supplied via cloud are extremely scalable. Consumer could quickly scale up computing-sources they want, scale them down if there is no require to utilize it any longer. Cloud Scalability furnishes factual modularity to cloud. Furthermore computing-sources materialize like countless, There could be no need for consumers to craft schedule for provisioning (Nick et al., 2010) (QinglingWang et al., 2011).
- **Measured-service:** Computing sources supplied by cloud are guarded according to computing sources abilities. Their usage can be supervised, accounted to be competent to afford lucidity for both supplier, consumer of the resources (Shicong et al., 2013) (Noorshams et al., 2011).
- **Off-premise:** Service is hosted, delivered from a site that belongs to service provider
- **Flexible billing (Pay as you go):** An additional novel feature of cloud computing is application of utilization based billing form. Fees can be levied on a subscription basis or can be tied to actual consumption of resources. Consumer pays only for short-term utilization of processors or disk-storage.
- **Universal/ubiquitous access (Access anywhere, anytime):** pooled resources are available to anyone authorized to utilize them.
- **Maintenance and advancement:** Since cloud supplier moderately than customer keeps computing source, there is effectual outsourcing of upholding jobs. Thus cloud supplier retains, renews resources, whether resource is hardware or software. Consequently total maintenance and replacement of principal hardware resources are clear to consumer, because they do not persuade customer's acquaintance. Even though this might be true in outstanding illustration, there may be undersized intermission when a consumer's illustration is reallocated from one hardware platform to another to take out safeguarding otherwise restore of given physical proposal, for the extent of this period consumer may perhaps not take account of resources attached with this illustration presented. Organization uses various cloud services as IaaS, PaaS, SaaS and the models like public, private, hybrid. These models and services has various cloud security issues. Each service model is associated with some issues. Security issues are considered in two views first in the view of service provider who insures that services provided by them should be secure and also manages the customer's identity management. Other view is customer view that ensures that service that they are using is secure enough.

IV. PERFORMANCE ANALYSIS ON CLOUD COMPUTING

Performance is an outlet for any system. A definition of performance has been proposed based on the ISO 25010 scale where the performance of cloud computing system is determined by analysing the properties involved in the performance of an effective and reliable service to meet the needs mentioned under and within the maximum transaction system boundary conditions. Measures and standards are applied to the three elements of users, suppliers and supervisors. Figure (1) shows that each output system is based on several possibilities, that is, the system may provide a service correctly or incorrectly, or the service may be altogether rejected and this affects the performance of the system. Cloud computing is one of the distributed systems which deals with dispersed resources and is based on processing big size data and this need called for the performance control and analysis. techniques to develop them, that is to say cloud control is important for both users and service providers. It provides information and indicators of performance, where that performance level affects the adoption of cloud services by users. Performance is among the

advantages that should be available in the cloud services because performance has an impact on users and service providers. To evaluate performance, we must take into account several criteria to evaluate the factors that affect the performance of cloud computing, including the average response per unit time and the average waiting time per unit time and others in the properties of performance measures of the cloud.



- (SaaS) Evaluation is made by users directly depending on the performance measures, the speed of response, reliability of technical services and availability.
- (PaaS) Evaluation is made by users directly or indirectly depending on the performance measures in response to technicality, productivity, reliability, technical service and middleware capability.
- (IaaS) The performance measures are determined depending on the infrastructure performance, capacity, reliability, availability, and scalability.

CONCLUSION

Cloud computing is an innovative technology as well as progress quickly furthermore it is hard to contest a high-quality characterization of cloud computing technology. From the above analysis we observed that, security affects performance and on the other hand, the level of performance by the user, the provider and the components of the cloud affect the security. They both affect the customer's confidence in the adoption of the cloud, and this is an indicator that affects the development and progression of the cloud and the light of this, the level of security and performance is considered a challenge and an obstacle to the progress and development of cloud computing.

REFERENCES

- [1]. Expósito, R. R., Taboada, G. L., Ramos, S., González-Domínguez, J., Touriño, 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]. 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.
- [3]. 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.
- [4]. 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.
- [5]. Expósito, R. R., Taboada, G. L., Ramos, S., Touriño, J. and Doallo, R. (2014), 'Performance evaluation of data-intensive computing applications on a public iaas cloud', *The Computer Journal* p. bxu111.
- [6]. 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.
- [7]. 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.
- [8]. 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.
- [9]. Li, Z., Zhang, H., O'Brien, L., Cai, R. and Flint, S. (2013), 'On evaluating commercial cloud services: A systematic review', *Journal of Systems and Software* 86(9), 2371–2393.
- [10]. 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.
- [11]. 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.
- [12]. 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).
- [13]. Sharma, G. P., Singh, S., Singh, A. and Kaur, R. (2016), 'Virtualization in cloud computing'.
- [14]. Sharma, S. K. and Kumar, R. (2012), 'A markovian feedback queue with retention of renege customers and balking', *Adv. Model. Optim* 14(3), 681–688.

- [15]. 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.
- [16]. 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.
- [17]. Xing, Y. and Zhan, Y. (2012), Virtualization and cloud computing, in 'Future Wireless Networks and Information Systems', Springer, pp. 305–312.
- [18]. Xiong, K. and Perros, H. (2009), Service performance and analysis in cloud computing, in 'Services-I, 2009 World Conference on', IEEE, pp. 693–700.
- [19]. Yang, B., Tan, F. and Dai, Y.-S. (2013), 'Performance evaluation of cloud service considering fault recovery', The Journal of Supercomputing 65(1), 426–444.
- [20]. Yu, Z., Liu, M. and Ma, Y. (2010), 'Steady-state queue length analysis of a batch arrival queue under n-policy with single vacation and setup times', Intelligent Information Management 2(06), 365.

BIOGRAPHIES



Rajat, M. Tech., Scholar in Department of Computer Science & Engineering at Maharishi University of Information Technology, Lucknow (U.P.) India. His research interest is in Cloud Computing, Performance Analysis and Security.



Dr. Ajay Kumar Bharti, Professor, Computer Science & Engineering at Maharishi University, Lucknow (U.P.) India. His research interest is e-Governance, Service Oriented Architecture and Knowledge Based System. He has published number of research papers in reputed journals and conferences.