ISSN: 0976-3104
SPECIAL ISSUE: Computer Science

ARTICLE

QUALITY OF SERVICE ON PERFORMANCE EVALUATION- A SURVEY

Deepa Mani[1*], Anand Mahendran[2]

[1]*School of Information Technology and Engineering, VIT University, Vellore-632014, Tamilnadu, INDIA*
[2]*School of Computing Science and Engineering, VIT University, Vellore-632014, Tamilnadu, INDIA*

ABSTRACT

Cloud environment is altogether different from conventional processing environment, and along with these, the execution of cloud performance is extra fundamentals. The development of information in the cloud is quick. Consequently, it requires that resources and framework accessible at removal must be similarly experienced. Infrastructure level implementation in cloud includes the execution of servers, system and capacity which go about as the absolute completeness for driving the whole cloud business. This paper aims at supporting investigation around the cloud computing and thereby giving an overview of the best in the class of QoS demonstrating methods reasonable for cloud frameworks. Our objective is to study overview of current and forthcoming investigation on QoS methods in cloud computing.

KEY WORDS
Dynamic power management; cost optimisation; Energy efficiency control; Reservation cluster;, systematic mapping

Received: 19 November 2016
Accepted: 16 December 2016
Published: 5 January 2017

INTRODUCTION

Cloud computing has created in popularity starting late because of specialised and moderate favourable circumstances of the on interest limit organisation model. Various cloud executives are presently dynamically accessible, giving advertising, which includes Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS) arrangements [4]. The development stack of the cloud has an additional standard in initiative data centres, where private and hybrid cloud structures are progressively included [2]. Despite the way that the cloud has essentially reworked the breaking point provisioning process, it represents to a couple of novel troubles in the region of Quality-of-Service (QoS) management [1]. QoS suggests the levels of execution, steadiness, and accessibility offered by an application and by the stage or framework that hosts it. QoS is essential for cloud customers, who foresee that supplier will pass on the advanced quality properties, and for cloud suppliers, who need to locate the right trade-offs between QoS levels and operational costs [3].

To accomplish indeterminate workload and to be exceedingly accessible for clients, anyplace at whatever point resource over-provisioning is a normal condition in a cloud system. Regardless, most power dependant facilities will unavoidably encounter the evil impacts of unmoving times or low use for some days or months. Since there, as a rule, have reduced activity realized by a method for unpredictable entries [5]. QoS properties have consistent consideration well before the initiation of cloud computing, implementation diversity and resource quarantine systems of cloud stages have altogether had QoS investigation, expectation, and certification [6]. Clustering is an essential and financially profitable podium for executing parallel applications that process vast measure of information with the hubs of a cluster through the interconnected system. Clustering is customarily utilised as a part of numerous information mining applications to gather together the measurably comparable information components [7].

In case, all the above components have been satisfied we can say that cloud computing is working with the unequalled force. There is a chance that any of the factors comes up short or neglects to fulfill the client needs then we need to run further with other procedures of distributed computing. In the accompanying portions, we have examined about every one of the elements of Quality of Service of our best.

In this paper, we have explained a summary of the existing models that target to quantify and recover the power and energy consumption, workload, cost estimation and availability of data centres and cloud hosts. In section 2, the study on performance enhancement modeling is given. Section 3, describes performance evaluation modeling. This survey shelters performance enhancement modelling, power and energy consumption modelling, workload modelling, Cost optimisation modelling, Reservation cluster modelling, and the duality of performance modelling. Section 4 deals the different methods to achieve the low cost, energy consumption and high availability. Finally, we conclude in section 5.

MATERIALS AND METHODS

Performance enhancement modeling

Performance modelling aims to perform the evaluation or forecast of the reaction time not just central for little applications additionally for the vast applications that are running on an open cloud. The principle responsibility of the cloud server centers to offer the nature of administrations, paying little mind to the exuberant method for the cloud where the consignment changes were made [8]. To satisfy the quality need, demands encouraged on the public cloud ought to check for their presentation i.e. reaction time and preparing time so execution variables are inside the lenience boundary [9].

***Corresponding Author**
Email: mdeepa@vit.ac.in
Tel.: +919789305010

Power and energy consumption modelling

Power and energy consumption and estimation modelling involve gauging the vital utilisation of virtualized hardware. Resource dispute of numerous applications attempting to utilise hardware can bring about significant overheads [18]. Resource sharing can prompt a decrease in energy utilisation, just like the case in sharing a multicore CPU for instance. The shared memory of the CPU reserves the various centres which regularly prompts a performance increment, while less energy is being expended [11].

Workload modelling

Workload modelling demonstrates how to build a distinctive cloud workload pattern and prompts to more knowledgeable choices to achieve better resource management [12, 15]. Likewise, workload modelling in cloud computing empowers presentation scrutiny and reproduction, which conveys advantages to cloud suppliers and analysts as it permits. (i) the evaluation, through re-order, of supply organisation procedures allowing the change of cloud organisations' QoS.(ii) the assessment of these methodologies without sending and implementation of the requests in immoderate expansive scale circumstances; (iii) the imitation of sensible cloud situations using organised change, alteration, and redundancy [13,14].

Cost estimation modelling

Cost estimation modelling is requested turns for various sorts of resources over the period among which the in-house servers has displaced with virtual machine occurrences (hereon called the arranging time frame) [16]. In continuation to that, there are resources has been sorted like CPU, RAM, storage and network [17]. (i) Define the demand curve for CPU that is, the average number of runnable processes over the period. (ii) Define the demand curve for RAM that is memory usage of the server (iii) Define the network demand, i.e., some data dragged in and dragged out from the network. (iv) Define the storage demand, i.e., the amount of data deposited in the hard drives.

Availability modelling

Availability modelling in the cloud services is fundamental for upholding client certainty and staying away from income misfortunes because of Service Level Agreement infringement consequences [21]. Since the software and hardware segments of cloud foundations may have constrained dependability, the utilisation of unessential segments and numerous clusters might be mandatory to accomplish the anticipated level of reliability while also the vibrant increase in the gaining and the computational costs [20].

Performance Assessment modelling

Cloud computing assets must be perfect, superior and capable. High performance is one of the cloud preferences which must be acceptable for every administration. Higher execution of services and anything identified with cloud have an impact on clients and administration suppliers. Henceforth, execution assessment for cloud suppliers and clients is essential. There are numerous strategies for execution forecast and assessment; we utilise the accompanying techniques in our assessment [22].

- Assessment taking into account based on criteria and attributes
- Evaluation has given as re-enactment [23].

Variables affect on execution

These days, the tenure "performance" is an exemplary idea and incorporates more broad ideas, for example, reliability, energy effectiveness, scalability and so on. Because of the degree of cloud computing situations and the extensive quantity of undertakings and typical clients who are utilising cloud setting, numerous components can influence the execution of cloud computing and its capitals [10]. A portion of the essential elements deliberated in this paper are as per the following:

Safety

The effect of safety on cloud presentation may appear to be delicately weird; however, the effect of security on network infrastructure has demonstrated. For instance, dispersed denial-of-service (DDoS) bouts wide effect systems execution, and it will essentially diminish systems execution besides effective on reaction time also. Therefore, if this threat and any same threats undermine cloud domain, it will be a major anxiety toward clients and suppliers. Many researchers have been proposed in information security [23], insurance and access control to improve the security [24]. Cloud computing security gives a model to the customer centered information encryption for expanding the unwavering quality.

Data Integrity

Data integrity gives similarly adaptable, position-autonomous, low-cost stage for the customer. Data integrity comprises of two viewpoints. One is the productivity and safety in which to create people in general and private key plainly and proficient to do the secret key era. The bounds are determined, and it is productive since a lot of information encryption is done and keeps from attacks [24]. Since we scramble

information to keep from unauthorised clients; data integrity is kept up [25]. In information storage framework, customers store their data in the cloud for the accessibility of reports and the security must be guaranteed. One of the critical subjects is the Byzantine failure and has overcome in the distributed frameworks which may achieve the capacity issues. It comprises of survey and document appropriation, and homomorphic token is acquainted with doing the encryption [26].

Scalability

It is the capacity of the system to play out the predetermined functionalities which characterise its ability. Methodologies like horizontal scalability and vertical scalability have acquainted with enhancing the scalability of the system [27]. Cloud computing handles expanding requests. There are various sorts of scaling available they are vertical scaling is restricted by the way that we can just get the large size of the server, horizontal scaling manages the capacity to scale more extensive to manage traffic and corner to corner scaling [28]. Virtual machine versatility is constrained in the event of the TCP message workload contrasted with different threads. The condition of-workmanship engineering system permits various virtual machines to scale the length of memory breaking points. It is done until it achieves its cut-off points [29].

Performance

Execution change is the estimation of results of a specific procedure. An expansive measure of information can be isolated into lumps so that hacking of information can avoid without manipulating the whole information [30]. The elite presentation tests in two distinct fields, for example, supercomputing on a committed group and a group of virtual machines running in the cloud and different architectures has been proposed to enhance the execution rate [31].

System models

Power and energy consumption modeling

Measuring the energy utilisation of virtualized equipment is a long way. Resource dispute of various applications attempting to utilise hardware can prompt bring about noteworthy overheads [11].The resource sharing prompts a failure in energy utilisation, similar to the case in sharing a multicore CPU for instance. Getting to shared memory CPU stores prompts a performance increment, while less vitality has consumed. In this section, we make a refinement between usage based energy prediction models and performance monitoring counter-based (PMC) expectation energy models.

In usage-based models, direct regression is frequently connected to a particular arrangement of use hardware statistics and the deliberate energy utilisation. In PMC-based models, logged (virtualized) in which occasion hardware counters are utilised to frame a prediction model. It is less demanding to acquire OS-provided inputs, for example, CPU; disk and memory usage to attain the hardware event counters [33]. Be that as it may, these models can frequently not be widespread to all hardware setups, as proficiency of hardware segments contrast extraordinarily. In this manner, utilising relative PMC (rPMC) as is done in [32] can offer a more nonspecific and less blunder inclined model for evaluating the hardware utilisation for workloads.

Estimation modeling

The power utilisation of PCs is not about the work they fulfill [33]. The measure of unmoving nodes must be kept to a base in server farms since they consume an accommodating measure of energy while not performing any operations. The facts can confirm that a server needs to run, yet the just low execution is necessary. Dynamic voltage and frequency scaling (DVFS) assistance in diminishing the energy cost, permitting it to keep running in a lower power mode by downscaling the CPU voltage and frequencies. Virtual Machine Consolidation can likewise be viewed as a multi-target streamlining issue, taking the minimization of power utilisation and resource wastage as the destinations, as depicted in [34]. Resource wastage has seen as the whole of staying unused resource of a physical host.

The objective is to acquire a non-commanded set of arrangements, the Pareto set, utilising Ant Colony Optimization heuristics. Ant Colony Optimization (ACO) is a metaheuristic propelled by the perception of genuine ant colonies and based upon their aggregate scavenging performance [34]. The goal to do this, linear regression models for resource wastage and power utilisation framed, of which the outcomes utilised as the two destinations of the improvement algorithm. Based on the algorithm a requested set of VMs and physical hosts as info and figures the attractive quality. The likelihood of moving a VM to a particular host while ensuring that every host does not surpass its resource use limits. At the point when the Pareto ideal set has computed, VM consolidation can happen to utilise the subsequent VMs to hosts mapping.

Economic cloud federation

Cloud federations take into consideration relocating VMs between numerous data centres of various suppliers at different geological areas [11]. A cloud federation comprises of regularly slighter participating cloud suppliers that go for manageability, lessening carbon dioxide outflows, and bringing down power costs [35]. As little and medium cloud suppliers can frequently not resist with the huge cloud suppliers, for example, Amazon, Google, and Rack space, they shape a league that gives every member a chance to acquire an external resource from each other, while minimising the energy expenses and carbon-dioxide emissions of their data centres.

Power utilisation can minimise by taking the neighborhood power costs into thought. Each VM is allocated an energy spending plan for every economic time interim. In each economic time interim, the budget plans for all VMs in the following economic time interim are ascertained utilising an estimation of resource use for the VMs utilising linear regression.

Workload modeling
Workload categorization

The workload is the measure of handling that the PC has been given to do at a given time [42]. The workload comprises of some measure of utilisation programming running on the PC and typically some number of clients associated with and communicating with the PC's applications. The workload is characterised like computation, memory, networking, and storage [35].

Deployment environment: Factual characterizations of observational data are valuable in QoS [15]. Some of the QoS model parameters, e.g., network transfer capacity fluctuation, virtual machine (VM) start up times, begin failure probabilities used to estimate the practical qualities. Perceptions of performance variability have accounted for various sorts of virtual machine occurrences [35]. Different works describe the variability in VM start-up times , which is connected specifically with working operating system picture size.

Workload Implications

Regression Techniques. A typical workload derivation approach includes evaluating just the mean interest set by a given sort of requests on the resource [32]. The method depends on contrasting the performance measurements (e.g., throughput and usage of resource) anticipated execution model against estimations accumulated in a controlled test environment [4].

Cost estimation modelling

A cost capacity gives the details about the expenses of force utilisation, framework clog and server start-up. The impact of vitality productivity controls on reaction times, working modes and acquired expenses are all illustrated [33]. Our goals are to locate the ideal service rate and mode-exchanging confinement, to minimise the cost of a response time ensures under differing entry rates. A productive green control (EGC) algorithm is initially proposed for taking care of obliged optimisation issues and making costs/performance trade-offs in systems with various power sparing policies [5].

Amortization

It is critical to comprehend the commitment of IT base expenses. Consequently, amortisation limit is ascertained for servers and different amenities so that reasonable ascription of expenses for different IT resources (software/ hardware) can be realised [11]. The limit is mandatory to compute the month to month deterioration cost (amortisation expense) of every infrastructure. These things have introductory purchase cost, the expense of which is ascertained on the duration over which the speculation is amortised at the expected interest rate. Studies have uncovered that the expense of CPU storage and transfer speed twofold when the expenses have amortised over the duration of the substructure [35].

Cost of servers

It is acknowledged that each one of the servers has relative setups and servers mounted on racks. This statement is made to inspire the estimation of the cost of the server (without amortisation).Henceforth, the cost of the server has calculated with the combination of quantity of servers in a firm and the expense per server. The amortizable parameter for server determined in the past part will be used to choose the amortised server cost and amortizable Parameter for Server [34].

Availability modelling

There are some mechanisms which can help to protect against the failures, [32] have considered those mechanisms into groups like fault tolerance mechanisms, protective redundancy and overload protection.

Fault lenience mechanisms

Fault tolerance can be achieved, based on robustness and dependability of the system. It can classify into two types, i.e., proactive and reactive. The Proactive adaptation to non-critical failure arrangement is to

keep away from deficiency, blunders and failure by anticipating them. It will proactively displace the suspected part implies recognise the issue before it comes. Responsive fault tolerance to non-critical failure arrangements decreases the exertion of failure when the disappointment adequately happens. These can further portray by two sub-strategies like error taking care of and fault treatment. The main intentions of error handling are to expel the errors from the computational state. Fault treatment drives for keeping issues from being reactivated [31].

CONCLUSION AND FUTURE DIRECTION

In this paper first, we discussed facts that are involved in performance enhancement modeling and evaluation modeling. QoS approaches to receive a key part in the change of distributed computing to ensure that customers can trust cloud administrations. There has been a growing energy for QoS approaches in cloud computing among modern pros and analysts. The work comprehensively surveys the different performance enhancement modeling like power consumption modeling, cost estimation modeling, and availability modeling and so on. However, the technologies used in these modeling are somewhat difficult to analyze their corresponding QoS, from the service provider point of view. We have done detailed survey in workload and system modeling to QoS management.

CONFLICT OF INTEREST

There is no conflict of interest

ACKNOWLEDGEMENTS

None

FINANCIAL DISCLOSURE

None

COMPETING INTEREST

No competing interest

REFERENCES

[1] Armbrust M, Fox A, Griffith R, et al. [2010] A view of cloud computing. Commun ACM 53(4):50–58.

[2] Zhang Q, Cheng L, Boutaba R. [2010] Cloud computing: state-of-the-art and research challenges. J Internet Serv Appl 1(1):7–18.

[3] Akpan, Helen Anderson, B RebeccaJeya Vadhanam. A Survey on Quality of Service in Cloud Computing."*International Journal of Computer Trends and Technology (IJCTT) volume* 27: 58-63.

[4] Ardagna, Danilo, et al.[2014] Quality-of-service in cloud computing: modelling techniques and their applications. *Journal of Internet Services and Applications* 5.1 : 1.

[5] Chiang, Yi-Ju, Yen-Chieh Ouyang, Ching-Hsien Robert Hsu.[2015] An Efficient Green Control Algorithm in Cloud Computing for Cost Optimization.*IEEE Transactions on Cloud Computing* 3.2: 145-155.

[6] Petcu D, 0 Macariu G, Panica S, Craciun C. [2013] Portable cloud applications - from theory to practice. Future Generation Comput Syst 29(6):1417–1430.

[7] Malathy G, Rm Somasundaram.[2012] Performance enhancement in cloud computing using reservation cluster." *European Journal of Scientific Research, ISSN* : 394-401.

[8] Singh, Jitendra. "Study of response time in cloud computing." *International Journal of Information Engineering and Electronic Business* 6.5 (2014): 36.

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

[10] Grozev, Nikolay, and Rajkumar Buyya. [2015] "Performance modelling and simulation of three-tier applications in cloud and multi-cloud environments."*The Computer Journal* 58.1 :1-22.

[11] Warnaar, Martin.[2016] Cloud Energy Consumption Measurement and Reduction: an Overview of Methods.

[12] Magalhães Deborah, et al.[2015] Workload modeling for resource usage analysis and simulation in cloud computing. *Computers & Electrical Engineering* 47:69-81.

[13] Feitelson DG. Workload modeling for computer systems performance evaluation. Cambridge University Press; 2015. In press, available online at http://www.cs.huji.ac.il/~feit/wlmod/wlmod.pdf.

[14] Moreno I, Garraghan P, Townend P, Xu J. An approach for characterizing workloads in Google cloud to derive realistic resource utilization models. In: Proceedings of 7th international symposium on service oriented system engineering (SOSE). IEEE; 2013. p. 49–60.

[15] Chen Y, Ganapathi AS, Griffith R, Katz RH. Towards understanding cloud performance tradeoffs using statistical workload analysis and replay. Technical Report. University of California at Berkeley; 2010. URL: http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-81.html.

[16] Truong, Hong-Linh, Schahram Dustdar.[2010] Composable cost estimation and monitoring for computational applications in cloud computing environments.Procedia Computer Science 1.1 : 2175-2184.

[17] Singer, Georg, et al.[2010] "Towards a model for cloud computing cost estimation with reserved instances." *Proc. of 2nd Int. ICST Conf. on Cloud Computing, CloudComp 2010*..

[18] Vilaplana, Jordi, Francesc Solsona, and Ivan Teixidó.[2015] A performance model for scalable cloud computing. *13th Australasian Symposium on Parallel and Distributed Computing (AusPDC 2015), ACS*. Vol. 163.

[19] P.Latchoumy, P.Sheik AbdulKhader. [2011] Survey on fault tolerance in grid computing" IJCSI International Journal of Computer Science Issues, 2(4)

[20] Dantas, Jamilson, et al.[2015] Eucalyptus-based private clouds: availability modeling and comparison to the cost of a public cloud." *Computing* 97(11) :1121-1140.

[21] Nguyen, Tuan Anh, Dong Seong Kim, Jong Sou Park. [2016] Availability modeling and analysis of a data center for disaster tolerance." *Future Generation Computer Systems* 56: 27-50.

[22] AYMAN G. FAYOUMI, [2011] PERFORMANCE EVALUATION OF A CLOUD BASED LOAD BALANCER SEVERING PARETO TRAFFIC, Journal of Theoretical and Applied Information Technology, 32(1).

[23] Lar, S-U., Xiaofeng Liao, and Syed Ali Abbas. "Cloud computing privacy & security global issues, challenges, & mechanisms." Communications and Networking in

China (CHINACOM), 2011 6th International ICST Conference on. IEEE, 2011.
[24] Kulkarni, Gurudatt, et al. [2012] A security aspects in cloud computing." Software Engineering and Service Science (ICSESS), 2012 IEEE 3rd International Conference on. IEEE.
[25] Chalse, Rajkumar, Ashwin Selokar, Arun Katara.[2013] A New Technique of Data Integrity for Analysis of the Cloud Computing Security. Computational Intelligence and Communication Networks (CICN), 2013 5th International Conference on. IEEE,
[26] Ramaiah Y, Govinda G, Vijaya Kumari.[2013] Complete Privacy Preserving Auditing for Data Integrity in Cloud Computing. Trust, Security and Privacy in Computing and Communications (TrustCom), 2013 12th IEEE International Conference on. IEEE.
[27] Lee Jae Yoo, Soo Dong Kim. [2010.]oftware approaches to assuring high scalability in cloud computing." e-Business Engineering (ICEBE), 2010 IEEE 7th International Conference on. IEEE,
[28] Hassan, Shoaib, and Farooque Azam[2014]. Analysis of Cloud Computing Performance, Scalability, Availability, & Security." Information Science and Applications (ICISA), 2014 International Conference on. IEEE,
[29] Jamal, Muhammad Hasan, et al.[2009] Virtual machine scalability on multi-core processors based servers for cloud computing workloads." Networking, Architecture, and Storage, 2009. NAS 2009. IEEE International Conference on. IEEE,
[30] Behal, Veerawali, and Anil Kumar. "Cloud computing: Performance analysis of load balancing algorithms in cloud heterogeneous environment." Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference-. IEEE, 2014.
[31] Keville, Kurt L, et al. 2012] Towards Fault-Tolerant Energy-Efficient High Performance Computing in the Cloud.", Cluster Computing (CLUSTER), 2012 IEEE International Conference on. IEEE,.
[32] P Xiaoa, Z Hub, D Liua, G. Yana, and X. Qua.[2013.] Virtual machine power measuring technique with bounded error in cloud environments. Journal of Network and Computer Applications, 36(2):818{828,
[33] A Kansal, F Zhao, J Liu, N Kothari, A.A Bhattacharya. [2010] Virtual machine power metering and provisioning. In Proceedings of the 1st ACM symposium on Cloud computing, pages 39{50).
[34] Y. Gao, H. Guan, Z. Qi, Y. Houb, L Liuc. A multi-objective ant colony system algorithm for virtual machine placement in cloud computing. Journal of Computer and System Sciences, 79(8):1230{1242, 2013.
[35] M Giacobbe, A Celesti, M. Fazio,
[36] M Villari, A. Pulia.[2015]to. An approach to reduce carbon dioxide emissions through virtual machine migrations in a sustainable cloud federation. In Sustainable Internet and ICT for Sustainability (SustainIT), pages 1(4).