International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 09 , September 2017
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

QUALITY OF SERVICE IN CLOUD COMPUTING: A SURVEY

Shiva Prakash

Department of Computer Science and Engineering,

Madan Mohan Malaviya University of Technology, Gorakhpur, Uttar Pradesh (India)

ABSTRACT

Cloud computing is utility based IT services based on the model as pay per usage for cloud computing environment transported to user's crossways sphere in excess of the internet. Load balancing is an essential feature for quality of service in cloud computing environments without proper load balancing we cannot expect better response time, the order for the cloud has improved because individuals and enterprises have now wandered to the cloud and cloud benefactors need. A major issue associated with cloud computing is QoS management that is allocation of capitals to the requests to assurance service based on concert, convenience, and dependability. Various allocation policies are available and they have their own pros and cons

This paper presents a literature review on quality of service in cloud computing environment. This survey is performed by inspecting the characteristic of available approaches for providing QoS. Merits and demerits of various approaches is also considered. We have gone through many good researches and found that there are some issues needed to be resolved.

Keywords: QoS Management, Guarantee Service, Resource Allocation, Availability, Reliability

I. INTRODUCTION

Cloud Computing is an rising technology which changes the way service is provided. Many cloud service provider companies are there in the market to provide cloud services such as Platform as a Service (PaaS), Infrastructure as a Service (IaaS) and Software as a Service (SaaS). Star.t-up companies are using the cloud as they do not want to spend on infrastructure. Cloud Computing provides different kind of service such as hardware and software via internet or browser. Cloud Computing uses the concept of virtualization as the task in Cloud Computing require minimum completion time, optimum utilization of resource and better performance. Different types of task scheduling algorithms are used in cloud computing to provide QoS. Lot of challenges in cloud computing are as follows optimal cost services, response time, security [12], quality of services requirement etc, an internetworking of large number of remote servers is internet based computing in cloud to allow data storage in centralised manner and online access of services or resources. Load Balancing [11] is essential for efficient operations in distributed environments to provide better QoS. In cloud computing data centres receives various client requests. These requests come to load balancer and load balancer is effectively balance the load across the system and ability to enhances the quality of service [13].

In this research paper we study what is QoS and what are different techniques of cloud computing that will help us to enhance QoS. Cloud computing deals with shared computer resources and data and provide the data to

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

millions of users over the internet. Cloud computing is more better than the normal computing system like Mobility, mobility term means that spreading service in a wide space that you can use this service from anywhere you have access to the internet. You can access your documents which you have uploaded to cloud storage services like Drop Box. Cloud computing provides increasingly storage, so you will not be worry about running out of space on your memory. Cloud computing is inexpensive comparing to the other memory storage. The software already installed online, so no need to install it by manually. Many cloud computing providers provide spaces free like Drop Box. Though it has lots of strong points there are some drawback also for example if we talk about security and privacy, your information may be access by unauthorized users. For protection against unauthorized access, the concept offers password based protection and operate on secure servers with data encryption technology.

Rest of paper organized as section 2 presents techniques and applications of cloud computing, section 3 admission control, section 4 we provide review of related some existing works, section 5 show comparative study of well-known techniques and finally we provide conclusion in section 6.

II. TECHNIQUES TO PROVIDE QoS OF CLOUD APPLICATION

Quality of Service (QoS) depends on various factors as resource allocation with proper load balancing results to reduction of congestion, minimization of delay, enhances availability, reliability and overall the level of performance. There are many techniques to provide quality of service to the cloud applications. Scheduling, admission control and dynamic resource provisioning are some techniques used to achieve that goal.

A. **Scheduling:**

Main goal of development is to improved consumption of the assets with no disturbing the services provided by the cloud. The main aim of using job scheduling algorithm is to enhance the QoS of cloud computing and provide predictable output on instance. The main task to maintain efficiency and fairness among the jobs with proper utilization of resources results to obtain high performance.

=In the mechanism of given that these services it is essential to progress the consumption of data centre resources which are operating in most dynamic workload environments. There are so many algorithms for scheduling in cloud computing. Some of them are as First Come First Serve (FCFS) basis means execution follow the sequence of task arrival, Round-Robin algorithm (RRA) means allocate a time slice in circular way to each task, Min-Min Algorithm (MMA) means execution based increasing order of task size and Max-Min algorithm MMA) means execution of task decreasing order of task size.

B. Resource provisioning:

The resource provisioning techniques are of two types first is static and second is dynamic. Dynamic resource provisioning is the process of assigning available resources to the cloud application. Resource allocation will make services suffer if the allocation not managed in the right way. This resource provisioning technique is used to meet desire QoS.

The optimal resource allocation strategy (RAS) should avoid the following the criteria as following:

- **Resource contention:** Suppose two cloud application are requesting the same resource at the same time, there will be conflict that requested resource will be allocated to which application; this situation is called resource contention.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

- **Resource Scarcity:** number of resources are fixed (let say K) and the demand on these resources (m) is very high (m>K) the scarcity of resources arises.
- **Resource fragmentation**: Suppose K number of resources will be divided in n number of smaller chunks (i.e. k1, k2, k3, …, kn and k1< K, k2<K, …, kn<K). Now if an application is requesting form number of resources (where m<K), system we not allocate the requested number of resources to that application even though enough resources are available. This situation known as resource fragmentation.
- **Over Provisioning:** Suppose an application is requesting for n number of resources, but system is providing m number of resources to that application (where m>n), then over provisioning will arise.
- **Under provisioning:** Suppose an application is requesting for n number of resources, but system is providing m number of resources to that application (where m<n), then under provisioning, will arise.

III. ADMISSION CONTROL

The main purpose of admission control is to provide strong performance. At admission control time, the Infrastructure Provider (IP) must consider not only the fundamental computational and networking necessities but also the extra requirements that may be required to be added at runtime so it becomes elastic. For example, if multiple users are working on cloud with high variations, the number of VMs are required more and that may be added at runtime many times multiple of the number of the basic ones.

IV. LITERATURE REVIEW

This section presents the work related to quality of service in cloud commuting environment. We consider the works which directly focused on enhancement of network performance with quality of service is one of the main concerns and also focuses on the challenges and limitations of existing well-known authors works.

C. Hershey et al. [2] proposed a SoS method for proper responding of enterprise QoS monitoring, management architecture in cloud computing, it is extension of previous work to provide topology for which find out points in administrative blocks in which QoS metrics can be managed and monitored. For example, to provide new SoS approach to real word scenario via DDoS-distributed denial of service. The method is very effective but it was not applied to federated clouds in real time. M. Salam et al. [3] proposed a federated QoS-oriented approach in cloud computing where different independent service provider may cooperate continuously to provide QoS–assured services. The key elements for enabling cloud federation used were Cloud Coordinators (CC) and Federation Coordinators (FC). The different feature of the planned federation context is its QoS-orientation that can managed the reactive resource provisioning diagonally several providers, henceforth to take full advantage of quality of service targets and resources practice, remove SLA destructions and improve SLA validation. However, composite facilities were not created by means of a combination of facilities from changed cloud providers and no any facility was finished for distributed DoS outbreaks.

M. Hassan et al. [4] studied and tested the assignment the group of emblematic Big data tasks on Amazon cloud EC2 in succession of Big data. They created a great replication setup and compared the projected technique with other approaches. Though, the proposed approach was cost effective, concert metrics for instance throughput, delay variable and delay were not occupied into deliberation.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Lee et al. [7] planned an architecture that working based on the agent technology to grip the one-to-one care of requested Quality of Service requirements and service level arrangements, to provision confirmation and endorsement. Furthermore, the agent technology dynamically analysed resources allocation and deployment. This work's weak point was lack of self-learning algorithm to determine the timing of automatic allocation of system resources.

Bin et al. [8] proposed a novel QoS-aware dynamic data replica delete strategy for disk space and maintenance cost saving purpose. Experimental results demonstrated that the DRDS algorithm can save disk space and maintenance costs for distributed storage system while the availability and performance quality of service requirements are ensured. However, increased overhead on update and inconsistency of data is usually associated with data replication.

W. C. Chu et al. [3] proposed a prescribed model to support not only the ECC facilities, strategy and building through SaaS, PaaS, IaaS but also the concurrent one-to-one care and active examination on the QoS issues for the possibilities from QoS facility providers and the SLA for several ECC clients. Created the prescribed model, investigation and testing the model was produced to provision involuntary testing in addition to runtime intensive care to declare the gratification to the necessities/SLA restraints. This work had some limitations such as not adapting the features and solutions of IOT into the framework as well as the field experiment

R. Karim et al. [6] proposed a mechanism to map the users QoS necessities of cloud facilities to the true QoS provisions of SaaS formerly plan them to the greatest IaaS facility that suggestions the best QoS assurances. The end-to-end QoS values was considered because of the planning. They proposed a traditional direction to accomplish the mappings procedure. The QoS conditions was hierarchically modelled using the analytic hierarchy process (AHP) approach. The the analytic hierarchy process formed model helped to enable them aping process across the cloud layers and to rank the candidate cloud services for the end users. A case study was used to illustrate and validate the explanation method. No performance of evaluation was done based on actual QoS data groups of cloud service station.

P.Zhang et al. [7] presented a QoS context for cloud computing with mobility and an reactive QoS supervision process to accomplish QoS assurance in mobile cloud computing environment. Y. Xiao et al. [10] presented an well-organized reputation-based QoS provisioning system, to minimize computing resources cost, even though sustaining the anticipated QoS metrics. They considered the numerical possibility of the reply time as a hands-on metric somewhat than the distinctive mean reply time. More so, QoS provisioning algorithm was not used to integrate security and privacy metrics. M. Xu et al. [11] introduced diverse QoS with multiple workflows necessities a multiple QoS inhibited development strategy of multi-workflows (MQMW) to discourse the issue of many workflows with different QoS requirements. The projected approach could agenda numerous workflows which were started at any time though QoS constraints such as availability and reliability were not added to workflows.

V. COMPARATIVE STUDY OF QOS TECHNIQUES

The comparative study as shown in table 1 has based on quality of service techniques in cloud computing on the following parameter- author and year of publication, what tools and/or technique used and their advantages and disadvantages. Most of the work is based on prediction based quality of service, response architecture, QoS oriented cloud computing, and minimization of operational cost, response time and data processing time.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Table 1: Comparative Study of Qos Techniques in Cloud Computing

Author and year of publications	Techniques used	Advantages	Disadvantages
P. C. Hershey, 2015 [2]	Enterprise Monitoring, Management Response Architecture in Cloud Computing (EMMRA CC) System of Method (SoS)	Enhanced Performance and Prevents distributed denial of (DDoS) of attacks	Result cannot be applied to federal clouds because cloud providers‟ servers were not integrated in real time
M. Salam, 2015 [3]	QoS Focused on Cloud Computing Framework, Federated Coordinators (FC) and Cloud Coordinators (CC)	Defend the Benefactors from any possible SLA destruction. different QoS requirements	No delivery was finished for distributed DoS attacks. Composite services were not created by means of a mixture of services from diverse cloud providers.
W. C. Chu, 2014[4]	Multi agent model	Established an integrated cloud data service to support QoS and SLA manipulation.	The features and solutions of IOT was not adapted into the framework as well as the field experiments.
M. M. Hassan, 2014[5]	Heuristic algorithms	Cost effective and dynamic Vm allocation model to handle big data tasks.	Performance metrics such as delay, delay variation and throughput were not taken into consideration
R. Karim, 2013[6]	AHP based ranking algorithm	Presented end to end computing in very special way in cloud environment	No performance evaluation was done considering datasets of real QoS
P. Zhang, 2011[7]	Fuzzy cognitive map and QoS Prediction Algorithm	Facilitates QoS prediction, establishment, assessment and assurance	No good model with suitable configurations was generated.
S. Lee, 2012[8]	Agent based technology	Enhance communication between layer control data with reverence to few service station performance such as SLA.	No self-learning procedure was considered the allocation of resources automatically
Y. Xiao, 2010[10]	Dirichlet Multinomial model	The proposed management	QoS provisioning algorithm was not
M. Xu, 2009[11]	Scheduling Algorithm	Produced better scheduling results	QoS constraints such as reliability and availability was not added to workflows.

VI. CONCLUSION

In this paper, we surveyed various QoS issuers and compared various techniques based on QoS in cloud computing and determine the extent to which QoS challenge has been resolved. Many researchers have provided scheduling techniques, admission control, traffic control, dynamic resource provisioning, etc to handle the issue of QoS in cloud computing. Resource provisioning and time scheduling techniques really help to enhance utilization of resources correctly in time. By adopting these technologies in cloud computing it will improve the response time and availability of the resources that are required for users. The excess dynamic local assignment to be distribute using the concept of proper load balancing in the entire cloud to accomplish the fulfilment of a high user resource consumption ratio as well as improved QoS.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

REFERENCES

[1.] Roshni Singh, Ataussamad and Shiva Prakash, "Enhancement of Resource Allocation using Load Balancing in Cloud Computing" International Journal of Advanced Research in Computer Science(IJARCS), *ISSN: 0976-5697,* Vol. 8, No. 4, May-June, 2017, pp.11-18.

[2.] P. C. Hershey, S. Rao, C. B. Silio and A. Narayan, "System of systems for Quality-of-Service observation and response in cloud computing environment", IEEE Systems Journal, Volume: 9, Issue:1, 2015, pp. 1-5.

[3.] M. Salam and A. Shawish, "A QoS-oriented inter-cloud federation framework", IEEE Systems Journal, 2015, pp. 642-643.

[4.] W. C. Chu, C. Yang, C. Lu, C. Chang, N. Hsueh, T. Hsu, S. Hung, "An approach of quality of service assurance for enterprise cloud computing (QoSAECC)", in Proc. Int. Conf. on Trustworthy Systems and their Applications, 2014, pp. 7-13.

[5.] M. M. Hassan, B. Song, M. S. Shamin, and A. Alamri, "QoS aware resource provisioning for big data processing in cloud computing environment", in Proc. Int. Conf. on Computational Sc. and Computational Intelligence, 2014, pp. 107-112.

[6.] R. Karim, C. Ding, A. Miri, "An end-to-end QoS mapping approach for cloud service selection", in Proc. IEEE Ninth World Congress on Services, 2013, pp. 341-348.

[7.] P. Zhang, and Z. Yan, "A QoS-aware system for mobile cloud computing", in Proc. of IEEE, 2011, pp. 518-522.

[8.] S. Lee, D. Tang, T. Chen, W. C. Chu, "A QoS assurance middleware model for enterprise cloud computing", in Proc. IEEE 36th Int. Conf. on Computer Software and Application Workshops, 2012, pp. 322-327.

[9.] L. Bin, Y. Jiong, S. Hua, N. Mei, "A QoS-aware dynamic data replica deletion strategy for distributed storage systems under cloud computing environments", in Proc. Second Int. Conf. on Cloud and Green Computing, 2012, pp. 219-225.

[10.] Y. Xiao, C. Lin, Y. Yiang, X. Chu, X. Shen, "Reputation-based QoS provisioning in cloud computing via Dirichletmultinomial model", IEEE ICC Proceedings, 2010, pp. 1-5.

[11.] M. Xu, L. Cui, H. Wang, Y. B. Bi, "A Multiple QoS constrained scheduling strategy of multiple workflows from cloud computing", IEEE International Symposium on Parallel and Distributed Proceeding with Applications, 2009, pp. 629-633.

[12.] Vikas Kumar and Shiva Prakash, "Modified Active Monitoring Load Balancing with Cloud Computing", IJSRD - International Journal for Scientific Research & Development, ISSN: 2321-0613, Vol. 2, Issue 9, September 2014, pp-184-189.

[13.] Ataussamad and Shiva Prakash, "Enhancement of Security in Cloud Computing using Steganography", International Journal of Advanced Research in Computer Science(IJARCS), ISSN: 0976-5697, Vol. 8, No. 4, May-June, 2017, pp.5-10.

[14.] Aditya Narayan Singh and Shiva Prakash, "Comparative Study of Various Load Balancing Technique with Energy Efficient in Cloud Computing", IJSRD - International Journal for Scientific Research & Development, ISSN: 2321-0613, Vol. 3, Issue 05, August, 2015, pp-1021-1024.

[15.] Vikas Kumarh and Shiva Prakash, "A Load Balancing Based Cloud Computing Techniques and Challenges", published in International Journal of scientific research and management (IJSRM), ISSN: 2321-3418, Volume 2, Issue 5, May 2014, pp. 815-824.