IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-5, Issue-5 E-ISSN: 2347-2693

A Framework for Management and Monitoring of QoS-based Cloud Services

K. J. Modi[1*], D.P. Chowdhury[2]

[1]Department of Information Technology, U V Patel College of Engineering (Ganpat University), Gujarat, India
[2]Department of Information Technology, U V Patel College of Engineering (Ganpat University), Gujarat, India

Corresponding Author: kiritmodi@gmail.com

Online Available at www.ijcseonline.org

Received: 20/Apr/2017, Revised: 28/Apr/2017, Accepted: 21/May/2017, Published: 30/May/2017

Abstract— Cloud computing provides dynamic and scalable environment for the benefit of end users to offer services which are known as cloud services. Due to on- demand, scalability low operational cost benefits the customers using cloud service is increasing day by day. Thus, the biggest and major challenges in the cloud environment is maintaining the Quality-of-Service (*QoS*), where cloud provider has to manage the resources in such a way that they will able to provide promised QoS metrics such as, performance, availability, reliability and cost. Cloud services are offered depends on Service Level Agreements (SLAs), an official agreement between cloud provider and user. To measure what quality of the services is offered by the cloud provider, we require monitoring the cloud computing services. Monitoring is important for both customer and service provider. The monitoring status helps cloud provider to improve their services and also helps the customer to know whether they are receiving the promised QoS or not. In this paper, we proposed the general framework and addressed the different aspects related to management and monitoring of QoS-based cloud computing services. We also discuss the key issues and challenges of QoS-based cloud services.

Keywords—Cloud Service Monitoring, Service Level Agreement, Qos, Cloud Service, QoS Management

I. INTRODUCTION

Cloud computing has received attention in past few years; this is due to the economic and technical advantages of the resource management model [1][2]. There are many cloud vendors, which are provides different type of services such as, Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS) solutions[3]. The cloud computing has also played important role in enterprise data centers, where private, public and hybrid cloud architectures are rapidly adopted by the different organizations.

Despite of the cloud computing has simplified the capacity provisioning process, it has many issues and challenges in the field of Quality-of-Service (QoS) management. QoS means the levels of performance, availability and reliability offered by cloud applications as well as by the platform or infrastructure that hosts it. Customers expect that cloud providers will deliver the promised QoS, and for cloud providers, need to maintain the right balance between QoS promised in SLA and operational costs. But to maintain this optimum balance is really very difficult task for cloud service providers. If the cloud provider unable to provide the promised QoS they have to pay the economic penalties as specified in SLA[5].

The rest of the paper is organized as follows. Section I contains the introduction of cloud computing and QoS Management in cloud computing environment. Section II describes different metrics for QoS evaluation for cloud service. Section III contains the issues and challenges in QoS for cloud services. The section IV contains the architecture for QoS management of cloud services. The Section V describes the importance of QoS monitoring in cloud environment. Section VI concludes the requirement of QoS monitoring and management in cloud service.

II. METRICS FOR QOS EVALUATION OF CLOUD SERVICE

In this section, we describe different metrics for QoS evaluation for cloud service and associated with three aspects of the cloud computing services are performance, economic security and general [2].

A.*Performance* Metric

There are several different options provided by cloud providers to their customers. The decision of selecting the right option by the customers is very important regarding to performance, service latency and precision of the cloud computing services. Performance can be measured in many contexts such as:

1) Response time: The duration needed from the beginning to the completion of task. It can be measured using following formula [2].

Response Time= Finish Time of task – Submission Time of task

2) *Throughput*: It means how much a number of task completed per unit of the time. It can be measured by using following formula [2].

$$T = \frac{N}{Execution\ time\ of\ N\ tasks\ + Total\ delay\ of\ \ N\ tasks}$$

Where N= total number of executed task, T=Throughput

3) Timeliness: It is the capability to meet deadlines, i.e. To complete the task with in the promised time period.

There are several features of performance aspects identified by many researchers are summarized as shown in Table 1.

Table 1. Performance Metrics [4]

Characteristics	Particulars	Metrics
Communication	Denotes the data exchange between internal cloud services instances or between external clients and the cloud providers.	• Error Rate in Connection • Packet Loss Frequency
Computation	It indicates the task (job) processing or computing data in the cloud systems.	• CPU Load (%)
Memory	The reuse of temporary stored information in cache improves the data access time in cloud systems	• Mean Hit Time (s) • Response Time (ms)
Time	Completion of task within promised time (meet deadline) and at the same time maintaining the QoS is very important.	• Computation time • Communication time

B.Economic Metric

For customers, economic matric is most important aspect while selection of cloud provider. Different cloud service provider offers variety of options at different cost. Cloud service provider offers different virtual machine depending on customers' plans and requirements. Customers compare all the available options and select the most economical service provider and plan that meet the customer's requirements.

1. Security Metric

Information security and privacy are crucial problems in cloud environment. In cloud computing, services are accessed by the customers via internet. But internet is the untrusted and insecure network. Thus, there is always security concerns while storing and sharing company's confidential data. The features such as data security and authentication are of security aspects identified by many researchers are explained in Table 2

Table 2. Security Metrics [4]

Features	Description	Metrics
Authentication	Authentication usually means verifying the whether the user is authenticated or not. Only authenticated user has right to access the data,	• Effectiveness • Sensitivity • Confidentiality
Data Security	Cloud provider offer different plan and option to secure the data in the cloud at different cost. Customers have to select the plan depending on what level of security they needed to secure their data in cloud.	• Communication Latency over SSL • Is SSL Applicable • Audit ability

C.General Metric

The general QoS metrics of cloud computing services are specified below:

4) Availability: Availability represents the idea of anywhere and anytime access to cloud service. The availability can be measured by the following formula [2].

$$A = \frac{(commited\ hour - outage\ hour\ during\ commited\ time\ ) * 100}{commited\ hour}$$

Where A=Availability

5) Reliability: Reliability means the capability of the system to provide service to customer without malfunction. Thus, we can say that it the capability of the system to provide service to customer even when there is some component failure (i.e. fault tolerant).
6) Adaptability: The ability of system to adapt the changes in responsive manner. It can be done by dynamic reallocation of computing depending on workload, it is also called elastic computing.
7) Scalability: It is the capability to expand the cloud computing resources when you scale up/down.

III. ISSUES AND CHALLENGES OF QOS BASED CLOUD SERVICES

Maintaining the good QoS is the important aspects in cloud computing but there are several issues and challenges to provide better QoS to cloud customers which are explained below:

1) Scalability is the one of the advantage of cloud but when number of user increases rapidly is difficult to provide promised QoS to customers by the cloud providers.
2) Cloud services are available to customer at any time and from anywhere through internet. When there technical problems occur, it is difficult for cloud providers to provide promised response time and availability to cloud customers
3) IOT is the emerging field in the digital world. The tiny wireless sensors are the important part of the IOT architecture. These tiny wireless sensors are having power limitations, low data processing capabilities and limited communication range.
4) The cloud computing services are accessible to the users via internet. Since internet is untrusted network the security is one the major issue.

IV. QOS MANAGEMENT FOR CLOUD SERVICE

In cloud, service level agreement (SLA), is the legal agreement between cloud providers and customers in which all QoS parameter is specified. If cloud provider unable to fulfill the promised QoS parameters, then they have to pay the penalties specified in SLA. Thus, it is required to monitor the different QoS parameters and manage the QoS to provider customer satisfaction. We discuss the models that are used to provide better QoS to the users. We categorized the QoS management into two models.

1) Workload Modelling: In workflow modelling we predict or assess the arrival rates of requests and of the demand for resources (e.g., CPU requirements) The capability to forecast the resource demand are used for many enterprise applications. Several investigations have been done by many researches in last to decade for the problem of estimating resource demand. In [9] presents a regression-based analytic model for dynamic resource provisioning.
2) System Modelling: In system modelling we measure the performance of a cloud system, either at runtime or at design time. Several models (such as queueing models, stochastic activity networks, stochastic process algebras, stochastic reward nets) are used to forecast the particular QoS values such as availability, response time or reliability. In this paper, we focuses the only queuing model in deep. Queueing network is a collection of queues interacting through incoming request and outgoing request. In queuing network, each queue indicates a software buffer (e.g., admission control, or connection pools) or a physical resource (e.g., CPU, network bandwidth, etc.). Applications in cloud are often tiered and queueing networks can used to store information about the interactions between tiers. Also, In [10] represents a multi-tier application in cloud using queueing network and to present a resource allocation policy which is SLA aware.

In past few years, due to importance of QoS in cloud many researches are focused on the better QoS management. Zhang et al. presented a QoS aware system for mobile cloud computing, in which authors proposed a framework for adaptive QoS management in mobile cloud computing environment which is based on fuzzy cognitive map (FCM) [7]. In [8] proposed a reputation-aware QoS provisioning scheme, which can minimize the operational cost of computing resources at the same time fulfilling the desired QoS parameters. One of the approaches is automatic QoS management which provide intelligent environment for service management. In such environment, we use self-management of resources depends on the domain knowledge of the cloud computing components. The emerging semantic and ontology technology can be utilized to automate the QoS management. Kourtesis, Dimitrios, Jose María Alvarez-Rodríguez, and Iraklis Paraskakis surveyed the semantic-based QoS management. They discuss the issues, challenges and current status of sematic based QoS management [6].

V. QOS MONITORING FOR CLOUD SERVICE

Now a day's organization is migrating towards the cloud computing services. However, such migration has its drawback like Performance degradation, data center outage, loss of connectivity. Due to this migration of applications (e.g., scientific data processing application, multi-layered business application, multi-media application, etc.) to clouds one of the major challenge is performance unpredictability. In 2012, Amazon EC2 cloud was crashed, which took down the applications of many SMEs. Some applications were out of service for hours, others for days [6]. Thus, we require to monitoring of cloud computing services [11]. The task of monitoring is important for both cloud providers and consumers. For cloud service providers, it helps them to control and manage the resources so that will able provide promised quality of services. For customers, continuous monitoring helps to know the SLA violation [8]. If SLA is violated then cloud service provider has to pay the penalties as specified in SLA. Thus, monitoring is clearly fundamental for all the tasks covered by the role of "cloud auditor". The Monitoring of the cloud computing service have following three phases:

1. Data Collection: In this phase, the QoS parameters are collected such as response time, availability etc.

2. Processing: In this phase, the parameters that are collected in data collection phase are checked whether SLA is violated or not.
3. Reporting: In this if SLA is violated then the alert signal is sent to report the SLA violation.

A.Framework for QoS Management of Cloud Service

In this section, we presented a general framework for management and monitoring of QoS-based cloud services. The three key components of this architecture are explained as below:

1. Customer: It may be user or computer that uses the cloud computing services through web portal. These cloud services are provided by the cloud vendors situated anywhere in the world. Customer specifies their requirements in SLA and based on this monitoring is done.

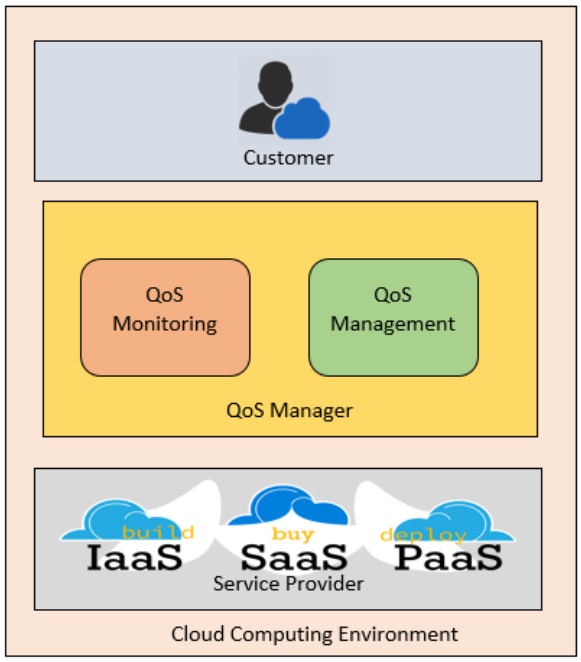

Figure 1. General Architecture for Management of QoS-based Cloud Services [12]

2. QoS Manager: The QoS Manager is designed to be responsible for two key functionalities:
   - QoS Monitoring: The QoS Manager monitors the all the services provided to their users. It generates statistical data for all QoS metrics.
   - QoS Management: The QoS Manager stores the data generated by the QoS monitor and uses theses data to take decisions (such as resource allocation, resource migration etc.) to give better quality of service to the cloud customers. The QoS Manager takes decision based on the either workload modelling or system modelling.
3. Service Provider: Service providers are those which deliver services to the cloud customers through web portal. The examples of different cloud providers are Google, Amazon, Microsoft Azure, Rackspace, iWeb, CloudSigma, yahoo, salesforce, IBM etc.

VI. CONCLUSION

Cloud computing is now a most adapted paradigm for diverse IT needs for many organizations and institutes. Presently, there are several of cloud providers that provide various cloud computing solutions with different price. With the increase of cloud provider, it is difficult cloud customers to select the ideal cloud provider who can fulfil their QoS requirements such as performance, response time, security, availability etc. at best price. With the increase popularity of cloud computing the customers using cloud solutions is also increasing, thus it is become really difficult for cloud providers to provide promised QoS. In this paper, we presented the QoS aspect in cloud service and we describe the different metrics for QoS evaluation. We explain the issues and challenges in QoS management and the role of monitoring in QoS management of cloud services. We also presented a general framework for management and monitoring of Qos-based Cloud services. QoS management and monitoring are very important aspects to evaluate the performance of cloud environment. If cloud provider is unable to fulfill the promised QoS to the customers, then there should be a provision to pay penalties specified in the SLA.

REFERENCES

[1] D. Ardagna, G. Casale, M. Ciavotta, J.F Pérez, W. Wang. "*Quality-of-service in cloud computing: modeling techniques and their applications*", Journal of Internet Services and Applications, Vol.25, Issue.1, pp.1-11, 2014.

[2] S. Ayyub, D. Roy, "*Cloud Computing Characteristics and Security Issues*", International Journal of Computer Sciences and Engineering, Vol.1, Issue.4, pp.18-22, 2013.

[3] M. Lagwal, N. Bhardwaj, "*A Survey On Load Balancing Methods and Algorithms in Cloud Computing*", International Journal of Computer Sciences and Engineering, Vol.5, Issue.4, pp.46-51, 2017.

[4] RP. Sarang, RK. Bunkar, "*Study of Services and Privacy Usage in Cloud Computing*", International Journal of Scientific Research in Computer Science and Engineering, Vol.1, Issue.6, pp.7-12, 2013.

[5] F. Jrad, J. Tao, A. Streit, "*SLA based Service Brokering in Intercloud Environments*", CLOSER, Portugal, pp.76-81,2012.

[6] D. Kourtesis, J.M. Alvarez-Rodríguez, I. Paraskakis, "*Semantic-based QoS management in cloud systems: current status and future challenges*", Future Generation Computer Systems, Vol.3, Issue.2, pp. 307-323, 2014.

[7] P. Zhang, Z. Yan, "*A QoS-aware system for mobile cloud computing*", 2011 IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS),US, pp.518-522, 2011.

[8] KJ Modi, DP. Chaudhury, G. Garg, "*An Ontology- based Approach for Automatic Cloud Service Monitoring and Management*", International Conference on Big-Data and Cloud Computing (ICBDCC'17), NY, pp153,157, 2017

[9] Q. Zhang, L. Cherkasova, E. Smirni, "*A regression-based analytic model for dynamic resource provisioning of multi-tier applications*", Fourth International Conference on Autonomic Computing, Jacksonville, pp 27-27, 2007.

[10] H. Goudarzi, M. Pedram", *Multi-dimensional SLA-based resource allocation for multi-tier cloud computing systems*", IEEE 4th International Conference on Cloud Computing, Washington, pp. 324-331, 2011.

[11] V. Patidar, M. Kumbhkar, "*Analysis of Cloud Computing Security Issues in Software as a Service*", International Journal of Scientific Research in Computer Science and Engineering, Vol.2, Issue.3, pp.1-5, 2014.

[12] B. Addis, D. Ardagna, B. Panicucci, M. S. Squillante, L. Zhang, "*A Hierarchical Approach for the Resource Management of Very Large Cloud Platforms*", IEEE Transactions on Dependable and Secure Computing, Vol. 10, No. 5, pp. 253-272, 2013.

