

# Analysis and Design of Advance Scalable QoS Based Resource Provisioning Framework

Pallavi Shelke<sup>a,1</sup>, Dr. Rekha Shahapurkar<sup>b</sup>

<sup>a</sup>*Oriental University, Indore, Madhya Pradesh, India*

<sup>b</sup>*St Vincent Pallotti College of Engineering & Technology, Nagpur, Maharashtra, India*

**Abstract.** In today's growing cloud world, where users are continuously demanding a large number of services or resources at the same time, cloud providers aim to meet their needs while maintaining service quality, an ideal QoS-based resource provisioning is required. In the consideration of the quality-of-service parameters, it is essential to place a greater emphasis on the scalability attribute, which aids in the design of complex resource provisioning frameworks. This study aims to determine how much work is done in light of scalability as the most important QoS attribute. We first conducted a detailed survey on similar QoS-based resource provisioning proposed frameworks/techniques in this article, which discusses QoS parameters with increasingly growing cloud usage expectations. Second, this paper focuses on scalability as the main QoS characteristic, with types, issues, review questions and research gaps discussed in detail, revealing that less work has been performed thus far. We will try to address scalability and resource provisioning problems with our proposed advance scalable QoS-based resource provisioning framework by integrating new modules resource scheduler, load balancer, resource tracker, and cloud user budget tracker in the resource provisioning process. Cloud providers can easily achieve scalability of resources while performing resource provisioning by integrating the working specialty of these sub modules.

**Keywords.** Quality of Service, Scalability, Resource Provisioning, QoS parameters, Cloud Computing

## 1. Introduction

Now a day's Cloud computing is a technology evolving in developments and it is a very important part of today's life. Cloud provides us mobility, virtualization, it is easy to maintain from home, reliability, and service provisioning as per our demand most organization and businesses switch to the cloud. The cloud environment is made up of collective resources to provide services to its user over the internet. [1] Three basic layer structures are used in a cloud environment, named Software as Service (SaaS), Platform as Service (PaaS), and Infrastructure as a service (IaaS). It will easily possible on SaaS environment, to access type of software like educational and business software which is present in cloud and user can access it through the internet on SaaS Platform. As cloud users, we can design various software-based programs and services through hardware, storage, also willing to do certain computation as well as hosting in that case also IaaS platform help us.

<sup>1</sup>Pallavi Shelke PG scholar, Department of CSE, India  
Email: pwankhede@stvincentngp.edu.in.

Generally, most of the cloud user demands about hardware parts as its to buy it personally is very expensive. IaaS platform provides virtualized part of underlying cloud hardware. In the cloud Environment 70% of people make use of the IaaS platform and 30% people used PasS and SaaS [2]. In the cloud environment, everyday cloud infrastructure providers and cloud users face big problems in terms of resource management. Generally, management of various types of resources is done through Resource provisioning and Resource scheduling mechanism. Resource Provisioning Mechanism helps cloud providers to find out the best resources within the required time to their user or client, whereas resource scheduling helps to schedule the resources, map them to the workload then perform real execution. Resource scheduling is always performed after the resource provisioning. State of art of network provisioning strategies fails to reconcile cloud user and cloud service vendor benefits. It is the very important automatic identification of suitable resources as per customer 's request because it directly affects service response time and cost. To overcome this issue, successful design of resource provisioning framework is very important in cloud computing environment [2] This paperwork mainly contributes towards Quality of service (QoS) in the resource provisioning process, work completed by researchers in same field discussed in detail. In this paper mainly we target scalability as the main QoS parameter.

The paper is structured as follows: In Section 2, background and related work highlighting the importance of the Resource Provisioning process in cloud environment, analysis existing resource provisioning proposed models along with scalability as the main QoS parameter, its motivation, importance in resource provisioning is discussed. Literature survey on QoS-based resource provisioning presented in Section 3. In section 4 under discussion, we have identified shortcomings and challenges while reviewing earlier research paperwork. In Section 5, the design of the proposed model is presented. Challenges and future work are discussed in section 6. Lastly, section 7 is based on the conclusion.

## **2. Background and Related Work**

Quality of Service (QoS) plays a crucial role in the cloud computing world, and researchers are working hard to design their proposed model from the perspective of both cloud users and cloud providers to reach deadlines, perfect execution times, and budget constraints.

**User Perspective:** cloud user wants a variety of services and resources as per their time with different expectations [3]

**Provider Perspective:** cloud provider wants to host many clients as per at a time to used resources available with them [3].

**QoS Perspective:** service often plays an important role between user and provider, where the assessment of the overall performance of service matters. Adequate QoS cannot be provided to cloud users until provisioning is made key functionality of resource offered. Therefore, to provide effective resources QoS based resource provisioning strategy is required [4].

## **2.1. Emphasizing the Importance of The Resource Provisioning Process in Cloud Environment**

Resource management is an umbrella practice that covers all cloud resource characteristics and utilization.[4] So under resource management, we have a resource provisioning process as the first step provision multiple resources by checking availability type resources available in the resource pool. In the cloud Resource Provisioning cloud user, cloud resource provider, Workload Resource manager play a major important role. In a cloud environment, the workload resource manager maintains Resource Description, QoS Metric, SLA Measure Under cloud resource manager resource provisioned work. The resource manager's main responsibility is to mapped resources to workload based on the Quality-of-service requirement of the user. At the same time, many users demand different types of resources and services. To manage the bulk of workload firstly workload queue is maintained in the cloud environment, to serve user requests as per priority. One's a user request for resource his /her request is shifted to the workload queue. As multiple type resource requests are there from cloud users, it shifted to workload analyzer where workload cluster is maintained where the type of the same request is the club to gather in one block. Then cloud user request checks for QoS measure and SLA measure. once this process is completed resource provider shifts the user request to the resource scheduler. Then Resource scheduler identified and detects the type of resource available in Resource pool. Once this step is complete required resource is shifted back Cloud resource provisioner. then there will rechecking of QoS and SLA measure before sending the required resource to cloud user by Cloud workload manger [4]. With a basic understanding of resource provisioning process, we have noted the following key point

- Identification of enough resource available in the Resource pool
- Provision Resource as per cloud user fluctuating demand
- Classification and clustering of different type of Resources and workload
- Nourishment of Quality of Service and required resources at a very high level
- Maintain a high-quality agreement at the service level
- Reduce Waiting time of cloud user requests.
- Serve multiple requests at the same time during peak demand by cloud users.

These key points will help in designing an advanced Resource Provisioning Framework

## **3. Literature Review**

The following table shows a comparative analysis of the work done from 2016 to 2020 and QOS parameters targeted till year 2020.

**Table 1:** Comparative Analysis of work addressed by various researchers in field of QOS based Resource Provisioning from year 2016 to 2020.

Author Name, publication Year	Techniques /framework /algorithm used	Distinguished work completed	Findings	QOS parameters target
Hala Haasan et al /year- 2020 [5]	QOS based Trust Model	i). Trust value dynamically updated at reach transaction ii) fake users will be identified on basis of the covariance mathematical Technique.	type of environment suitable for model not clearly mentioned.	Performance, cost, security, capacity, network
Aishwariya Chakraborty et al /year -2020 [6]	SensOrch: QoS-aware resource orchestra- tion scheme	1. Maintain high QOS and portability of sensor as a service 2.with threshold value enhance the performance of sensor- cloud and provide higher network-life time 3. Fair revenue distribution among sensor owner is possible	sensor cloud interaction as well as cost aspect missing	Resource Utilization
Shivangi Dhariwal et al/ year-2020 [7]	Resource optimization, Profit maximization technique	Maximizing the profit of the cloud server with a limited amount of resources. Compute size and speed of the server	worked only on assumed value. real time value needs to be considered to improve server-side performance	cost
A. Meenakshi et al /Year -2019 [8]	k-means clustering, gray wolf optimization partitioning technique.GSO-AGA algorithm	Reduce the high load on the server Allocates resources with the least amount of waste and provides the maximum benefit. minimum memory storage and minimum time will be possible with this algorithm	not clearly monitions which QOS parameter are improving with proposed model	cost, execution time
Sukhpal Singh Gill et al /year- 2018 [9]	SCOOTER framework	1.Optimized QoS parameters 2 Managing the resources automatically	Framework work only for fixed requirement or old resource requirement.	cost, resource Utilization, execution time
PVinothiyalakshmi et al /year - 2017 [10]	E-MCA Technique	Opens the way for the most efficient workload-resource pair to analyze workload clearly and distinctly by applying auction technique and	Need to check QoS attributes in cost and time by integrating proposed model in a real cloud	cost, Resource availability, response time (performance)

		QoS-based clustering.	environment, need to implement in a real-world cloud environment	
Xiaoyong Xu, et al/ year -2017 [11]	event-driven resource provisioning framework	1.It detects environmental changes in the form of Events, and agreements with events to reduce the expense of recruiting VMs 2. Reduce running cost of large computations	Real-time execution of proposed work required	Cost, scale up and scale down algorithm used
Tao Chen et al /year- 2017 [12]	online QoS model approach, used hybrid dual-learners technique	1. Produces better overall accuracy while having acceptable overhead 2. eliminates the need for heavy human intervention, which can be complex and error-prone.	To get the best result need to test & try the same proposed model on the new application	Response time
Himadri Shekhar Mondal et al/ year-2017 [13]	Fuzzy logic concepts	Improving QOS by Balancing the load with help if-then rule of fuzzy logic	This model can be improved by adding costing, reliability, more fruitful results Processing rate, etc. using the logic of Fuzzy.	Speed of processor, response time
Jolly Upadhyaya et al/year -2017 [14]	QOS Innovative model	Defined QOS parameters in terms of user oint of view help the cloud provider to maintain quality of service and also prepare their service. user and provider know in advance QOS variable and parameter	Proposed model only design for the education sector, QoS parameter or variable targeted by innovative model not mention.	Response time
Xianrong Zheng1, et al/ year- 2017 [15]	Spearman coefficient approach	Used to predict both QoS ratings and rankings for cloud services.  To help the cloud Providers to improve their brand and consumers to the cloud Identify services that satisfy the requirements of their QoS	Provide less accurate rating.	QoS-based cloud service recommendation, not targeted specific parameter

Tamal Adhikary et al /year -2016 [16]	QoS aware VM provisioning, VCA based resource provisioning method,	Decreases energy consumption, helps to increase resource usage across a range of multimedia social applications.	Type of SLA requirement violated is not mentioned in this paper	response time on average time as QOS parameter
Sukhpal Singh/-year 2016 [17]	QoS-based resource provisioning and scheduling framework (QRPS)	work in both environments, Homogenous and heterogeneous	Not focus on energy-saving parameters. provisioning must be done on maintaining the SLA as per user requirement	Energy, cost and execution time

According to our comparative study of resource-based QoS processes, it was found that a majority of work performed by researchers on QoS parameters is only targets cost, response time, execution time, and resource utilization. Also, after studying various proposed models, we found that scalability is the most important parameter to consider, as it will indirectly fulfill fundamental QOS parameters such as cost, response time, and resource utilization. Still, there is a lack of automating scalable real-time framework which suits all cloud computing platform. Few automated scalable frameworks available still facing a lot of issues which are discussed in the next paragraph.

#### 4. Discussion

Since after working more on scalability, we discovered that majority of research work is often centered on the auto scalability concept and we have noted the following reason of less work on scalability and its issues.

##### 4.1 Issues on Scalability Based Resource Provisioning Process

Researchers have come up with a novel concept of auto-scaling approach to improve resource provisioning work with scalability, where research paper [18], highlighted certain unique challenges like, a monitoring tool and an auto-scaling mechanism is required that meets quality of service standards and is also compatible with all deployment models at the SaaS and PaaS level also auto scaling should have a high level of fault tolerance

##### 4.2 Review Questions for Handling Scalability Issue

We have identified following questions based on above listed issues

- What will be the computation rate?

- How fast cloud user will get service benefits
- How the storage process will happen?
- How often do things /requirements change?
- What will be the success boundary with scalability?

#### **4.3 Examples Based on Work Completed On Scalability Based Approaches /Frameworks with main findings**

A lightweight approach is proposed in paper [19], that gives the fine-grained scaling at resource level itself which help to meet QoS requirements which will work efficiently over traditional VM-level cloud service scaling that overuse resources while still rising cloud provider operating costs. Finding: However, in this paper we found that more research is needed to establish how resources might be planned among applications with varying QoS requirements. A New auto-scaling mechanism was described in the paper [20], where auto-scaling method efficiently completes all jobs inside the user-defined deadlines .it helps to minimize costs for different workload patterns. The main findings of this paper are that workload forecasts will produce better results. This technique assumes a cloud customer with an unlimited budget, but it's better to consider it from the perspective of a user with a limited budget. Author presented an automatic resource provisioning approach in paper [21], for auto-scaling resources based on reinforcement learning with Marko decision process ((MDP). Here author tried to reduce the SLA violation increase stability but they do not focus on the type of services only work on fixed services but the status of the workspace will be altered based on its use not provided by the proposed approach, with the prediction of incoming load it can be possible with reinforcement learning but that not completed in the proposed approach. To address the issue of vendor lock, the author [22] described a model-driven approach for connecting a cloud platform-independent service model with cloud-specific operations. To provide auto-scaling deployment across clouds, the author makes use of cloud management tools and demonstrated data of different applications on multiple clouds. As a future work author would like to use machine-learning techniques to expand the model's predictive and constructive auto-scaling methods.

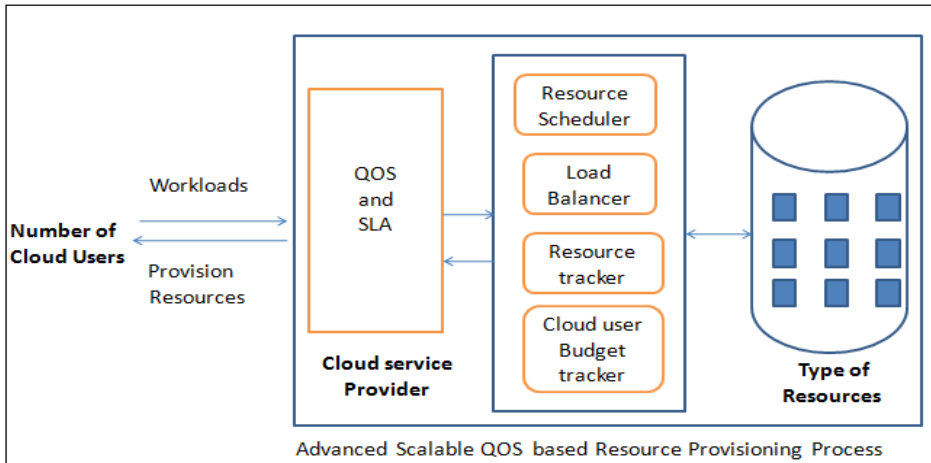
In reference to the above-mentioned research papers study, we have identified the following major challenges, which we will attempt to address through the proposed framework.

1. Still there is a lack of automating scalable real-time framework which suits all cloud computing platform.
2. Cost will impact scalability as per the on-time demand of cloud users.
- 3.To scale up and scale down resources as per resource availability in the resource pool automated resource pool tracker is required before start provisioning and scheduling the resources which save cloud user waiting time and cloud user provider's resource allocation time
4. There should be no delay in servicing resource requests as there is a heavy demand from cloud users.
5. There is a need for resource scheduling with various QoS requirements that must be met in one place

6. The framework should still take into account users with limited budgets also class of cloud users.
7. In the design context, the incoming user request should always be forecast

## 5. Proposed Model

Based on an existing research study on the above-mentioned resource provisioning proposed model, comparative analysis of QoS parameters, scalability issues, and auto scalability based completed research work, we proposed a scalable QoS based advanced resource provisioning framework, as shown in Figure 1, through which we will aim to overcome the challenges listed above.



**Figure 1.** Advanced Scalable QoS based Resource Provisioning Framework

The following sub-module will play a key role in the resource provisioning phase, assisting in the reduction of the difficulties and limitations addressed in the previous sections.

- **Resource Scheduler:** can constantly coordinate workload or application requests coming from cloud users. It can also aid in coordinating demands depending on the availability of resources in the resource pool.
- **Load Balancer:** A load balancer can assist the cloud provider in balancing customer workload according to priority which will reduce massive request traffic
- **Resource Tracker:** A resource tracker can assist with keeping track of resource waste if any, as well as searching for over-provisioning and under-provisioning of resources regularly.
- **Cloud User Budget Tracker:** The budgetary requirements will be taken into consideration by Budget Tracker.

We would aim to satisfy the needs of cloud users and cloud resource providers with this proposed framework by integrating different QoS criteria as well as SLAs (service Level agreement) in one place



## 6. Challenges and Future Work

The demand for global computing services is rising day by day in today's world. The cloud service provider must get ready with sufficient resources with QoS requirements to handle several cloud user requests also will get benefits when they start increased resources. But what if the resources are limited? How to deal with limited resources. Even though there was a considerable amount of research completed to develop a dynamic and self-managed cloud system, still there is a lack of adequate QoS-based resource provisioning approach required. In this paperwork, we have studied research work completed by researchers which mainly focused on scalable QoS parameters in resource provisioning. Guaranteeing QoS requirements is a need of today's resource provisioning approach. In future work, we will implement the above scalable QoS based resource provisioning framework which will overcome the following challenges and in our next research paper we will present the implementation and results of the above-proposed model

- As cloud user resource requirement or service requirement or workload changes
- on time therefore an automatic provision framework is required at the cloud provider side so that he can self-configure the available resource.
- Resource Provisioning System needed at the user side also just to fulfil their QoS expansion needs less pay.
- Flexible approach is required handling workloads that are rising repeatedly
- Scale-up and scale-out strategies are required to handle massive user request traffic.

## 7. Conclusion

In this Paperwork, we have discussed the importance of QoS-based resource provisioning and classified existing approaches proposed by researchers who worked on the same subject. Existing research work papers help us identify resource provisioning with many QOS-based parameters. Many researchers target one or more important QoS parameters, but after surveyed we have found that very few of them focus on scalability as QOS's main parameter/attribute. We have also included information on scalability types, the need for scalability, and analysis questions for dealing with scalability issues. Our paper concentrated on the scalability parameter, which is critical for both cloud providers and cloud users in terms of resource management and utilization. As future work in our research paper we will the implement proposed work which is based advanced scalable QoS-based resource provisioning framework that assists in overcoming the above difficulties while also attempting to close the study gaps.

## References

- [1] Ahuja S. P, Deval, N. , "On the performance evaluation of IAAS cloud services with system-level benchmarks", *International Journal of Cloud Applications and Computing*, pp80–96, 2018.
- [2] Durga S, Mohan S, Dinesh Peter J, Martina Rebecca Nittala. "Towards Benefiting Both Cloud Users and Service Providers Through Resource Provisioning " , *International Journal of Information Technologies and Systems Approach* , volume 12 ,2019.

- [3] M.H. Ghahraman, Mengchu Zhou, Chi Tin Hon, "Towards Cloud Computing QoS Architecture :Analysis of Cloud System And Cloud Services", IEEE/Caa Journal Of Automatica Sinica, Vol. 4, No. 1, 2018.
- [4] Sukhpal Singh, Inderveer Chana, (2016). Cloud resource provisioning: survey, status and future research directions. Springer- Verlag London, 2016.
- [5] Hala Hassan, Ali Ibrahim El-Desouky, Abdelhameed Ibrahim, El-Sayed M. El-Kenawy. "Enhanced QoS-Based Model for Trust Assessment in Cloud Computing Environment", IEEE Access, 2020
- [6] Aishwariya Chakraborty, Sudip Misra, Ayan Mondal, and Mohammad S. Obaidat., "SensOrch: QoS-Aware Resource Orchestration for Provisioning Sensors-as-a Service", 978-1-7281-5089-5/20/\$31.00 ©IEEE, 2020.
- [7] Shivangi Dhariwal, Dr. Hemant Makwana, "Maximize the Cloud Profit to Improved QoS in Cloud Computing: Design and Analysis", International Conference on Inventive Systems and Control, IEEE Xplore Part Number: CFP19J06-ART; ISBN: 978-1-5386-3950-4, 2019
- [8] A. Meenakshi, H. Sirmathi, J. Anitha Ruthi, "Cloud computing-based resource provisioning using k- means clustering and GWO prioritization. J. Soft Computing, Springer-Verlag GmbH Germany, part of Springer Nature, 2019.
- [9] Sukhpal Singh Gill, Rajkumar Buyya. "Resource Provisioning Based Scheduling Framework for Execution of Heterogeneous and Clustered Workloads in Clouds: from Fundamental to Autonomic Offering", J Grid Computing, <https://doi.org/10.1007/s10723-017-9424-0>, Springer, 2018
- [10] P. Vinodhialakshmi, R Anitha, "Efficient Dynamic Resource Provisioning using Auction based Technique with Improved QoS in Cloud Environment", ICCIDS, 978-1-5090-5595- ©IEEE, 2017
- [11] Xiaoyong Xu, Maolin Tang and Yu-Chu Tian, "QoS-guaranteed Resource Provisioning for Cloud-base MapReduce in Dynamical Environments. Future Generation Computer Systems, 78(Part 1), pp. 18-30, 2017
- [12] Tao Chen, Rami Bahsoon, "Self-Adaptive and Online QoS Modeling for Cloud-Based Software Service. IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 43, NO. 5, 2017
- [13] Himadri Shekhar Mondal, Md. Tariq Hasan, Taposh Kumar Karmokar, Shamlendu Sarker, "Improving Quality of Service in cloud computing architecture using Fuzzy Logic", ICAEE, 28-30 September, Dhaka, Bangladesh, IEEE, 2017
- [14] Jolly Upadhyaya, Dr. Neelu Jyoti Ahuja, "Quality of Service in Cloud Computing in Higher Education: A Critical Survey and Innovative Model", (I-SMAC). IEEE, 2017
- [15] Xianrong Zheng, Li Da Xu, Sheng Chai. "QoS Recommendation in Cloud Services. Digital Object Identifier 10.1109/ACCESS.2017.2695657. IEEE Access, 2017
- [16] Tamal Adhikary, Amit Kumar Das, Md. Abdur Razzaque, "Majed Alrubaian Mohammad Mehedi Hassan, Atif Alamri, "Quality of service aware cloud resource provisioning for social multimedia services and applications", Springer, 2016.
- [17] Sukhpal Singh, Inderveer Chana, "Resource provisioning and scheduling in clouds: QoS perspective", J Supercomputer, 72:926–960 DOI 10.1007/s11227-016-1626-x, Springer, 2016.
- [18] Hanieh Alipour, Abdelwahab Hamou-Lhadj, Xiaoyan Liu, "Analyzing Auto-scaling Issues in Cloud Environments. CASCON 24<sup>th</sup> Annual International conference on computer Science and software Engineering, 2014
- [19] R. Han, L. Guo, M. M. Ghanem, and Y. Guo, "Lightweight resource scaling for cloud applications In Cluster, Cloud and Grid Computing (CCGrid)", 12th IEEE/ACM International Symposium. pp. 644–651, 2012
- [20] Ming Mao, Marty Humphrey, "Auto-Scaling to Minimize Cost and Meet Application Deadlines in Cloud Workflows", SC11 November 12-18, Seattle, Washington, USA Copyright 2011 ACM 978-1-4503-0771-0/11/11, 2011
- [21] Bahar Asgari, Mostafa Ghobaei-Arani, Sam Jabbehdari, "An Efficient Approach for Resource Auto-Scaling in Cloud Environment", International Journal of Electrical and Computer Engineering (IJECE), Vol. 6, No. 5, pp. 2415–2424 ISSN: 2088-8708, 2016
- [22] Hanieh Alipour, Yan Liu, "Model Driven Deployment of Auto-scaling Services on Multiple Clouds. IEEE International Conference on Software Architecture Companion, 978-1-5386-6585-5/18, © IEEE, 2018.