

Prediction of Cloud Ranking in a Hyperconverged Cloud Ecosystem Using Machine Learning

Nadia Tabassum¹, Allah Ditta², Tahir Alyas³, Sagheer Abbas⁴, Hani Alquhayz⁵,
Natash Ali Mian⁶ and Muhammad Adnan Khan^{7,*}

¹Department of Computer Science, Virtual University of Pakistan, Lahore, 54000, Pakistan

²Department of Information Sciences, Division of Science & Technology, University of Education, Lahore, 54000, Pakistan

³Department of Computer Science, Lahore Garrison University, Lahore, 54000, Pakistan

⁴School of Computer Science, NCBA&E, Lahore, 54000, Pakistan

⁵Department of Computer Science and Information, College of Science in Zulfi, Majmaah University,
Al-Majmaah, 11952, Saudi Arabia

⁶School of Computer and Information Technology, Beaconhouse National University, Tarogil, Lahore, 53700, Pakistan

⁷Faculty of Computing, Riphah School of Computing & Innovation, Riphah International University Lahore Campus,
Lahore, 54000, Pakistan

*Corresponding Author: Muhammad Adnan Khan. Email: adnan.khan@riphah.edu.pk

Received: 12 October 2020; Accepted: 01 January 2021

Abstract: Cloud computing is becoming popular technology due to its functional properties and variety of customer-oriented services over the Internet. The design of reliable and high-quality cloud applications requires a strong Quality of Service QoS parameter metric. In a hyperconverged cloud ecosystem environment, building high-reliability cloud applications is a challenging job. The selection of cloud services is based on the QoS parameters that play essential roles in optimizing and improving cloud rankings. The emergence of cloud computing is significantly reshaping the digital ecosystem, and the numerous services offered by cloud service providers are playing a vital role in this transformation. Hyperconverged software-based unified utilities combine storage virtualization, compute virtualization, and network virtualization. The availability of the latter has also raised the demand for QoS. Due to the diversity of services, the respective quality parameters are also in abundance and need a carefully designed mechanism to compare and identify the critical, common, and impactful parameters. It is also necessary to reconsider the market needs in terms of service requirements and the QoS provided by various CSPs. This research provides a machine learning-based mechanism to monitor the QoS in a hyperconverged environment with three core service parameters: service quality, downtime of servers, and outage of cloud services.

Keywords: Cloud computing; hyperconverged; neural network; QoS parameter; cloud service providers; ranking; prediction



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

In this era, computers are viewed as core machines vital in every field of life to store data, process documents, and share information with others. As the global user-base of computers continues expanding, the need to have a more cohesive and distributed platform to ensure connectivity, sharing, and security is increased. Cloud computing has emerged as an answer to this problem with the provision of a distributed, connected, and secure environment along with revolutionary facilities such as shared infrastructure and software. It not only reduces the infrastructure cost but also enables use of applications and services without having to own them. Global software giants thus developed a completely new ecosystem, with multiple cloud service providers appearing (such as Microsoft, Google, Amazon, IBM, and Rackspace). The services provided by these CSPs are discussed in the following section [1].

1.1 Cloud Service Models

Service can be described as the endpoint of a connection; it should be well-defined, self-explained, and independent of context. In a digital ecosystem, many services are available for users as independent, well-describing, and stateless modules that perform a discrete unit of work. Cloud computing provides on-demand computing resources and services (see Fig. 1).

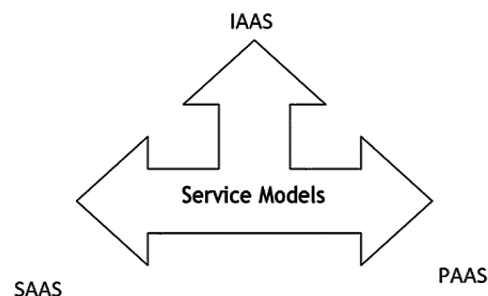


Figure 1: Cloud service models

1.1.1 Software-as-a-Service

Software-as-a-service can be defined as a process by which application service providers (ASP) provide different software applications over the Internet for users to rent. Due to the cloud architecture, users do not need to buy the license of a specific application and install it on a device; instead, they can use the application on the cloud, perform their desired tasks, and pay for their usage. Considering global demands, ASPs provide a rich collection of applications addressing general, government, business, and scientific requirements. Software-as-a-service (SaaS) is being used on variant digital devices like cell phones, laptops, and desktops, in domestic as well as professional environments. SaaS offers complete applications as a service on demand, and hence is also referred to as on-demand services [2].

1.1.2 Platform-as-a-Service

Cloud computing provides platform-as-a-service (PaaS) in which consumers only need to implement their applications without managing the configuration of servers and storage. This results in high scalability, elasticity, availability, reliability, and optimized performance. Consumers may configure, develop, deploy, and test their applications using PaaS in a runtime environment. This service has become essential for individuals as well as the corporate sector with a less

complex and ready-to-use working environment. Google App Engine, Amazon Web Services, and Windows Azure are a few PaaS examples.

1.1.3 Infrastructure-as-a-Service

In IaaS, the customer gets access to hardware infrastructure and can implement their applications on the cloud platform without having the responsibility of infrastructure management. Moreover, with the help of virtualization, resource provisioning is possible with control over the configuration of each virtual machine. Infrastructure-as-a-Service offers basic on-demand infrastructure through an application programming interface (API), which interacts with hosts, switches, and other resources [3]. Virtualization is an abstraction of logical resources that includes virtual machines, virtual storage, and virtual networks.

1.2 Deployment Models

Four deployment models are currently used in cloud computing.

1.2.1 Public Cloud

The public cloud model describes the traditional meaning of cloud computing. The public can access a general cloud service easily. The public cloud is also known as an external cloud or multi-tenant cloud because it makes the environment of the cloud more accessible. Homogeneous data, such as common policies, shared resources, and rented infrastructure, usually reside in this model. Furthermore, the public cloud has a large economic scale. Google, Amazon, and Microsoft offer many public cloud services like `ge.tt` [4].

1.2.2 Private Cloud

A private cloud is usually made for one organization. It may be controlled and managed by an organization (end-to-end) or a third party. It is also referred to as an internal cloud or on-premise cloud because of its limited resources accessible by consumers. A private cloud usually contains heterogeneous data and its policies are customized [5].

1.2.3 Hybrid Cloud

A hybrid cloud is a combination of two (public and private) or more clouds (public, private, and community) that are bound together but remain unique. A hybrid cloud has more scalability, security, flexibility, and cost-efficiency than other clouds. It also provides application and data portability. Eucalyptus Software is a big example of this type of model [6].

1.2.4 Community Cloud

The community cloud is shared by many organizations and supported by a specific group or community. This model is available for a specific class, group, or community of people with high specifications and security concerns. Resources are only shared with those who are part of that group or community. Its cost is lower than that of the public cloud but higher than that of the private cloud [7].

1.3 Hyperconvergence

In traditional systems, all modules needed a different skill to be managed, and all entities were configured and tested separately. In the era of convergence, hardware-defined infrastructure was introduced along with monitoring software and backup, as shown in [Fig. 2](#). In a hyperconverged infrastructure, all server components on a single unit are integrated through a software-defined environment. Furthermore, all components are readily available and ready to use.

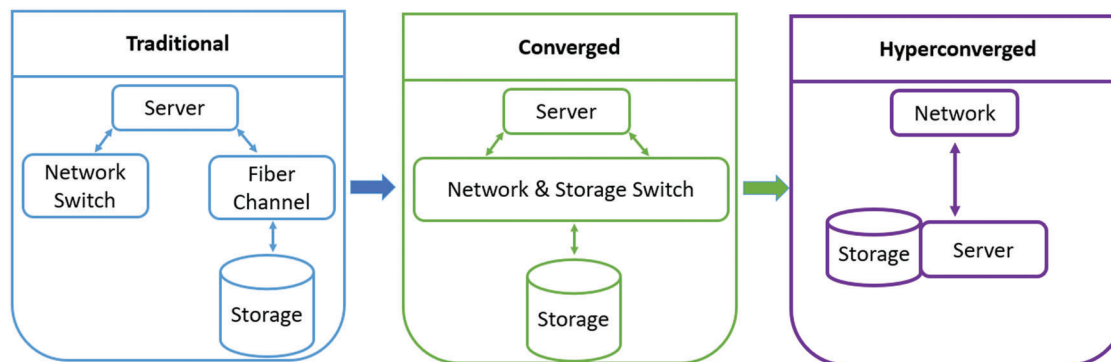


Figure 2: Traditional, converged, and hyperconverged infrastructures

2 Related Work

Ranking means assigning certain values and then sorting that choice according to its value. Generally, the lowermost value symbolizes the best choice and is considered the top rank. In cloud computing, the ranking is slightly transformed due to the pre-existing cloud infrastructure and naming conventions [8]. The ranking is thus based on quality parameters of services offered by cloud service providers, while infrastructure quality is based on the capacity and configuration of the hardware. This capacity and configuration debate transformed into the hyperconverged virtualization that brought variant facilities and features to the cloud community.

One example is using the RNN to analyze and predict an application's execution operation codes for ARM-based IoT devices. To train models, researchers have used an IoT application dataset including benign ware and malware, and then tested the trained model using 100 new IoT malware samples with three different long short-term memory (LSTM) arrangements [9].

The purpose of hyperconvergence is to simplify the operation and management of data centers by converging the computing, storage, and networking components into a single, software-driven appliance. The hyperconverged infrastructure is defined as an IT infrastructure framework in which storage, virtualized computing, and networking are tightly integrated within a data center. The software-based architecture that is the centerpiece of hyperconvergence is what makes the integration possible. In a hyperconverged environment, all of the servers, storage systems, and networking equipment are intended to work together through the appliance [10].

Hyperconvergence evaluation and performance is based on a hyperconverged infrastructure that focuses on mixed modeling of a wide variety of VMs used for a utility as a service. Under a hyperconverged model, cloud computing services can achieve better performance in a true hybrid cloud environment and also manage container-based applications in an efficient way; this results in increased efficiency and greater scalability. A well-established technique represents provision of data and, specifically, in Query Optimization of Big Data [11].

An emerging and challenging problem in cloud platforms is how to deal with the computer's capacity and workload, the latter of which is complex due to the variety of data types and ad-hoc devices; therefore, the cloud is the ideal platform to deal with mixed complexities. Researchers have discovered the potential use of a recurrent neural network (RNN) and deep learning techniques to discover cloud-based IoT malware. Furthermore, various methodologies using machine learning, deep learning, and combinations of neural networks have been presented for data prediction.

3 Cloud Ranking

Giving ranks mean assigning some value and then sorting those values. Normally the lowest value represents the best choice. The lower the value, the better the rank. Ranking in cloud services is gaining popularity. In a cloud infrastructure, ranking is slightly different compared with other cloud service models because of naming conventions and the pre-existing cloud infrastructure. As cloud services are growing swiftly, cloud service providers are keen to offer new and advanced services to attract more consumers, which is a competitive trend that can be especially beneficial for consumers; however, these swiftly introduced new services can also create a chaotic scenario for the consumers. Deciding on a service that fulfills the consumer's requirements [12] is a convoluted process.

Due to the complex nature of service types and quality parameters, there is currently no dedicated framework for cloud service indexing and cloud ranking.

The different levels of quality of service (QoS) in cloud computing are shown in Fig. 3.

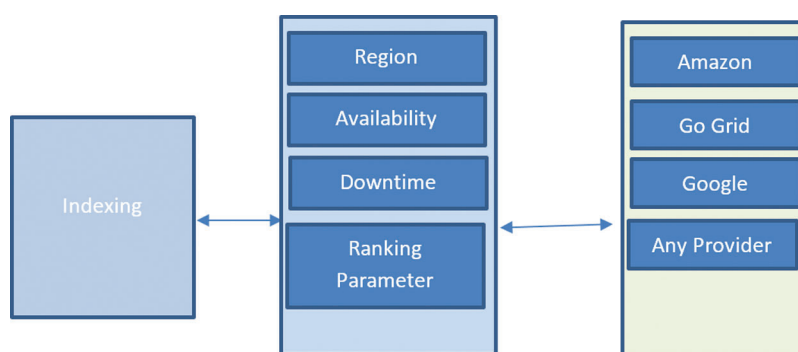


Figure 3: Indexing manager

The key factor in indexing is that the requirement of the user should be satisfied; Fig. 3 displays how to manage indexing services. The indexing manager will receive information and process it according to ranking parameters such as, region, availability, and downtime. The indexing manager will then identify the best-ranked cloud service based on the user's requirements. In the indexing module, the indexing manager will also be responsible for other activities as well, like taking parameters for ranking and keeping a record of indexed results [13].

The indexing of different cloud services must manage the status of the cloud system and also gather relevant cloud services. The indexing controller can be considered a benchmark for QoS by gathering information and performing comparisons. After the latter, it analyzes the ranking parameters using neural networks to rank cloud services, and then develops an autonomous cloud crawler [14]. Machine learning play a very important in various domains of life like medical, smart energy, wireless communication, Business Intelligence, smart city etc. [15–20].

4 Proposed Methodology

The proposed model consists of three main modules: a cloud infrastructure, a QoS ranking module, and a standard benchmark cloud. The model is outlined in Figs. 4 and 5. Due to the heterogeneous structure of cloud computing, there is no standard parameter to determine the

ranking of the cloud [16]. Standardization of cloud services based on the different QoS parameters and performances in various regions is a big challenge.

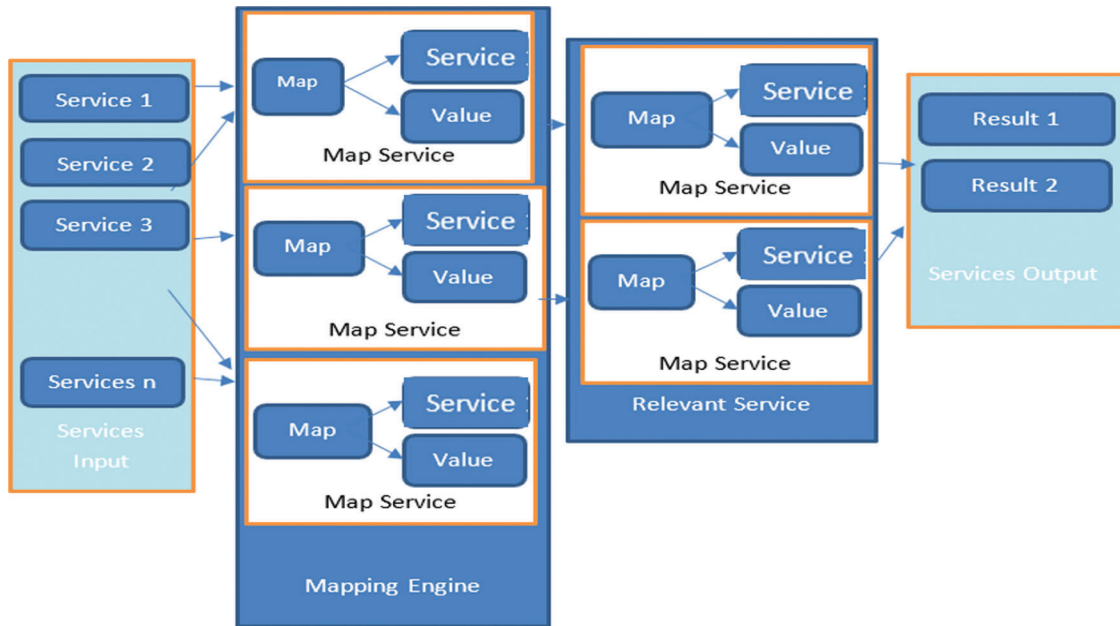


Figure 4: Cloud service mapping

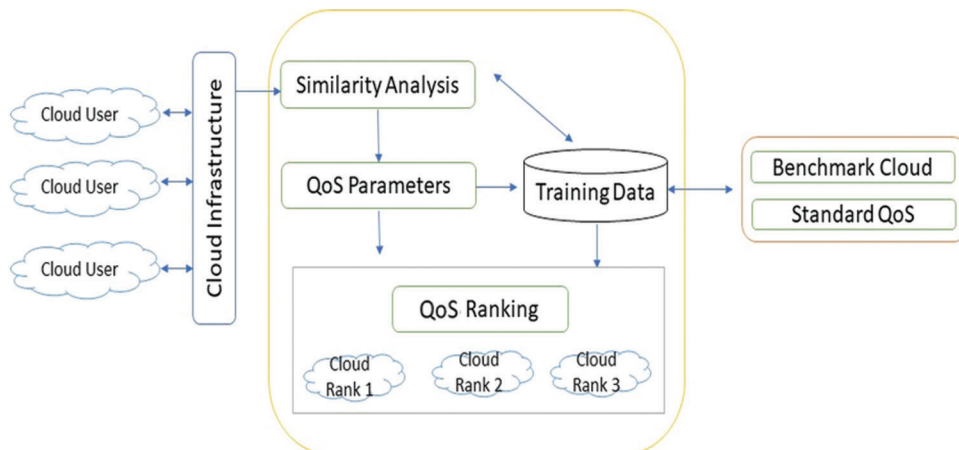


Figure 5: Proposed cloud ranking framework for measuring multiple QoS parameters

4.1 Cloud Infrastructure

1. The cloud services are accessed through web services over the Internet. Cloud service providers offer versatile services such as computing services, storage services, and content delivery networks. A single service can be offered in multiple regions. The main module

of the proposed model is the QoS ranking module, which is further divided into three sub-modules:

- Similarity analysis module
 - QoS parameter module
 - Training data module
2. Similarity analysis can be used to identify cloud services of the same type according to the service model, e.g., to differentiate between software-as-a-service or platform-as-a-service. QoS parameters can be used to rank the cloud services according to their respective service category. The QoS parameters used are
- Security feathers
 - Availability of cloud services
 - Downtime
 - Outage
 - Response time
 - Price
 - Trust.

Cloud services are compared with the standard cloud benchmark.

4.2 Cloud Feathers Extractor

All of the cloud services are listed and given to the input layer. First, the services are selected, that is, the cloud services, storage services, and computation services. Next, the input data are given to the next layer for further extraction of the QoS parameter. Further sampling is done based on QoS parameters.

5 Results and Discussion

Three types of data, namely computing, storage, and content delivery network, are used for training purposes.

In Fig. 6, three different types of cloud services are used to evaluate the performance of the proposed cloud ranking framework: compute services (CS), storage services (SS), and a content delivery network (CDN). A total of 524 cloud servers are assessed as shown in Tab. 1.

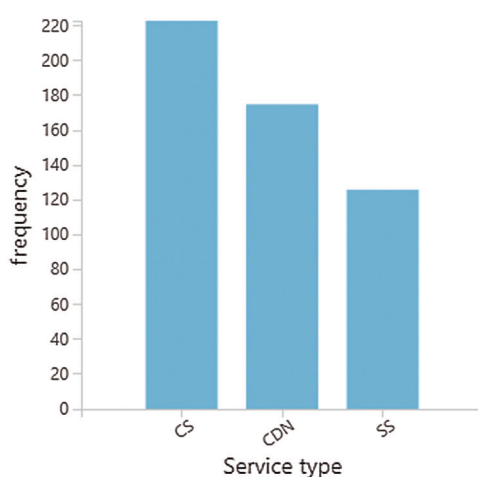


Figure 6: Frequency of different cloud servers with service type used in the proposed cloud ranking model

Table 1: Dataset of cloud services

Service type	Number of cloud servers
CDN	175
CS	223
SS	126

Fig. 7 shows the number of outages per cloud server.

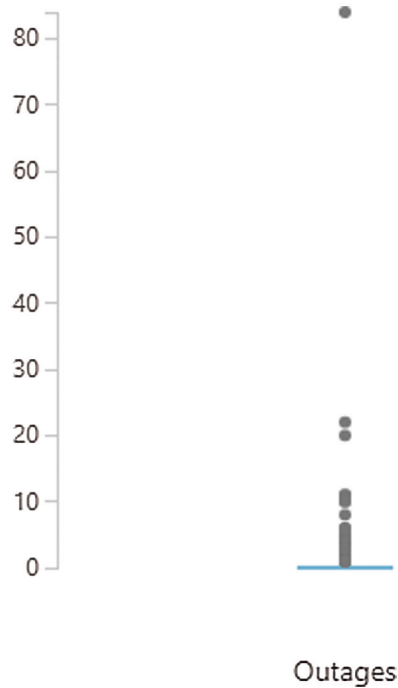
**Figure 7:** Cloud server outage frequency

Fig. 8 shows different cloud servers divided into two classes: good server and bad server. We assign the cloud servers into these classes, we used some criterion as the following: IF (AND ([@ [Downtime in Mint]] < 20, [@[30 Day Availability]] = 100%, [@Outages] = 0), “Good Server,” “Bad Server”). To be classified as a good server, the threshold for downtime must be less than 20 min, 30 d of availability must equal 100%, and the server must not be out of service to be considered; otherwise, the server is categorized as a bad server.

In the proposed cloud prediction model shown in Fig. 9, we have used four perceptron layers. Layer 1 is the input layer and has 321 units or nodes. Layer 4 has 2 units and serves as the output layer. Layer 2 and 3 each have 50 units and perform rectification.

Fig. 10 presents the ROC as a performance measurement for the classification problem. The ROC is a probability curve that shows the region of convergence of the proposed ranking model, which represents the degree or measure of separability.

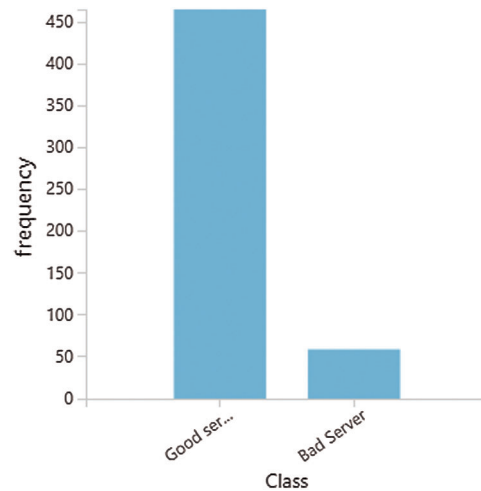


Figure 8: Different classes of cloud servers

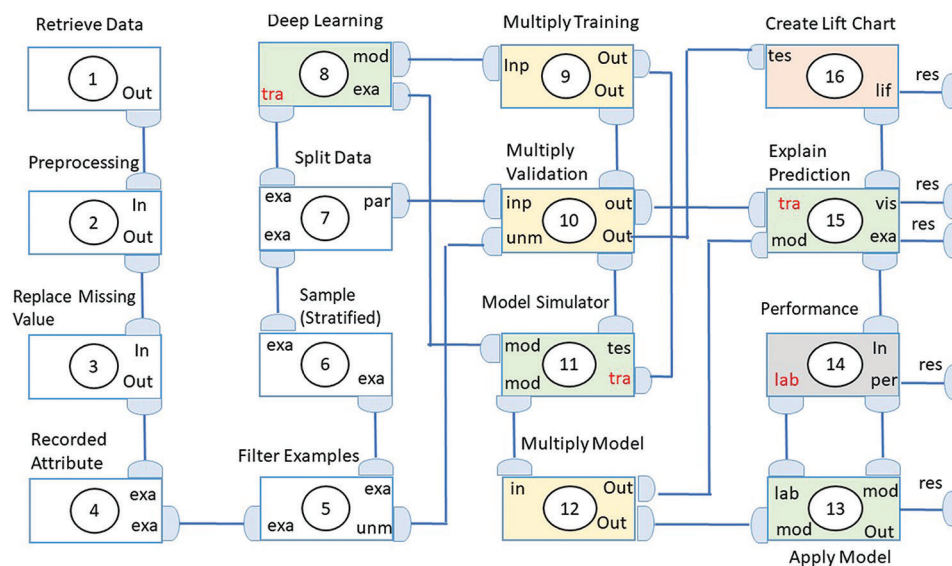


Figure 9: Workflow of the proposed cloud ranking model

We conducted precision analysis of the proposed cloud ranking model. The precision for predicting bad servers is 85.71% and that for predicting good servers is 100%. Additionally, the recall for bad servers is 100% and that for good servers is 97.85%. The overall accuracy of the proposed system is 98.10%, and its classification error is 1.9%. The results are presented in Fig. 11.

Fig. 12 shows the life chart of the proposed cloud ranking model concerning population and target achieved at 80% of the population; it shows 100% confidence in the prediction. The life chart indicates the machine learning performance for good and bad servers. We can also observe that the predictions and can become more useful by optimizing the ratio of good and bad servers.

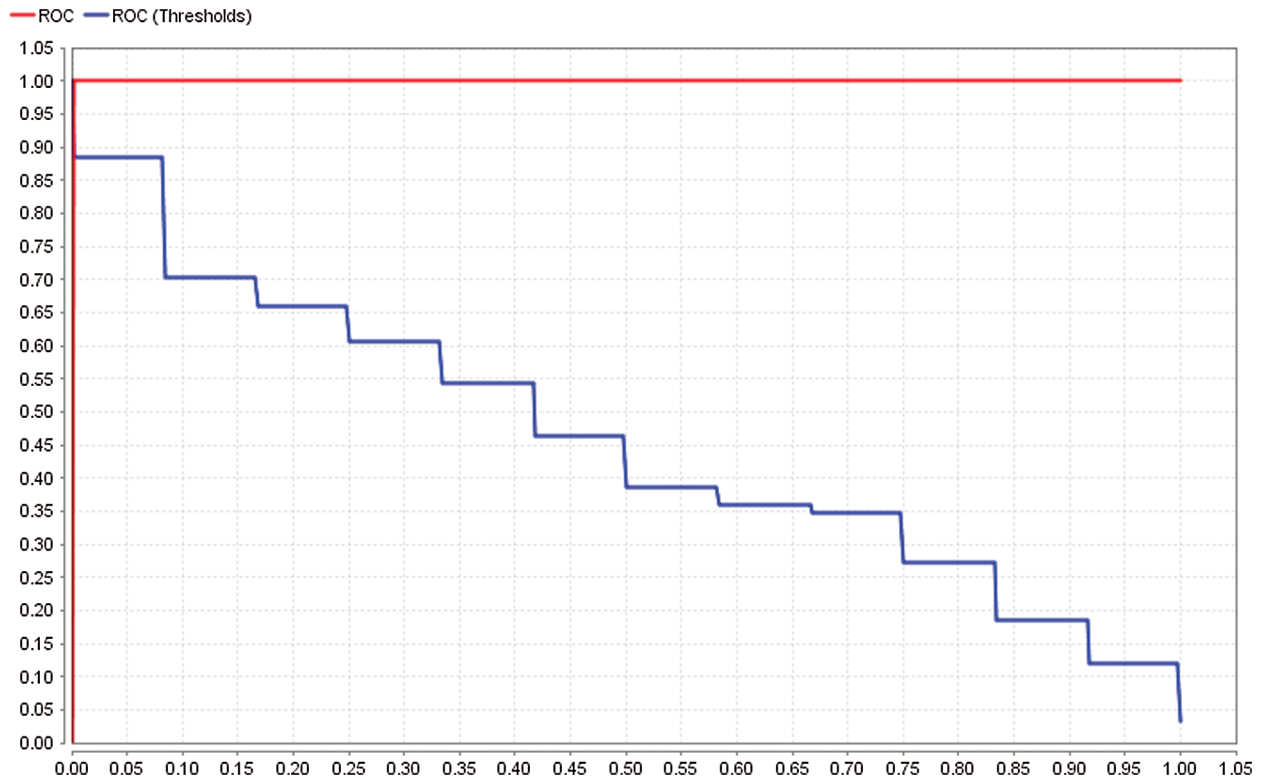


Figure 10: Proposed cloud ranking model ROC threshold

accuracy: 98.10%

	true Bad Server	true Good server	class precision
pred. Bad Server	12	2	85.71%
pred. Good server	0	91	100.00%
class recall	100.00%	97.85%	

Figure 11: Proposed cloud ranking model precision accuracy

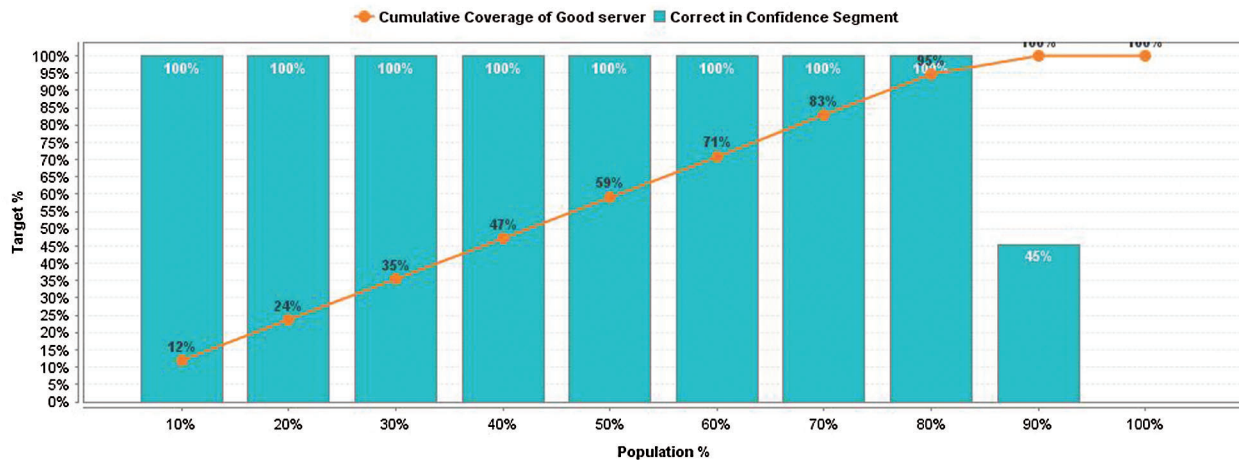


Figure 12: Proposed cloud ranking model life chart

While predicting cloud ranking in a hyperconverged cloud ecosystem, the outages factor is contradicted at a high rate as shown in Fig. 13 by red color bars. The 30-d-availability values are in support of good services and the value of the cloud service region is good.



Figure 13: Contribution of factors to the class “good server” in the proposed cloud ranking model

Fig. 14 shows the contradiction regarding the class “bad server” while prediction of cloud ranking in the hyperconverged cloud ecosystem is achieved by measuring outages that are contradicted at a high rate.

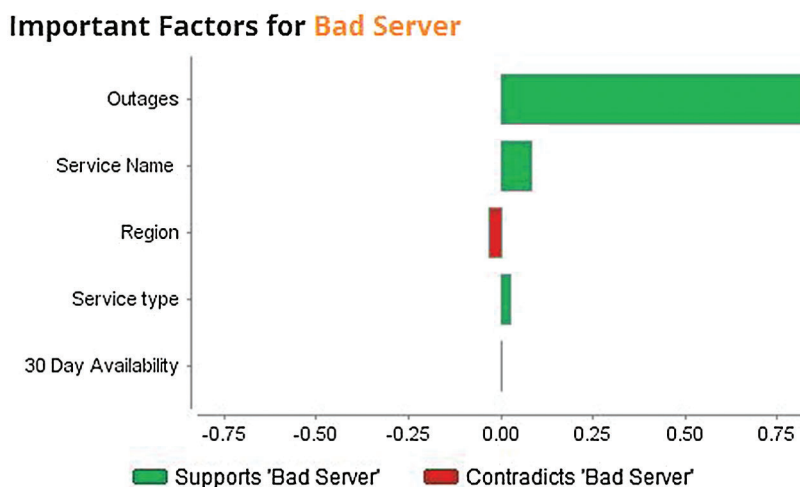


Figure 14: Contribution of factors to the class bad server

6 Conclusion

Modern computing requires a considerable increase in computational power for performing large-scale experiments and analytics to deal with hyperconverged infrastructure environments.

We have tested cloud performance and classified servers as good or bad. The ranking process combines performance data from multiple machines, user experience, workload, and availability analysis of good servers. Simulation results showed promising 98% accuracy of the proposed cloud ranking model based on QoS parameters.

Acknowledgement: Thanks to our families and colleagues who supported us morally.

Funding Statement : The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] G. G. Castañé, H. Xiong, D. Dong and J. P. Morrison, "An ontology for heterogeneous resources management interoperability and HPC in the cloud," *Future Generation Computer Systems*, vol. 88, pp. 373–384, 2018.
- [2] B. L. M. Bello and M. Aritsugi, "A transparent approach to performance analysis and comparison of infrastructure as a service providers," *Computers and Electrical Engineering*, vol. 69, pp. 317–333, 2018.
- [3] K. Iqbal, M. A. Khan, S. Abbas, Z. Hasan and A. Fatima, "Intelligent transportation system for smart-cities using mamdani fuzzy inference system," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 2, pp. 94–105, 2018.
- [4] M. Farid, R. Latip, M. Hussin and N. A. W. A. Hamid, "Scheduling scientific workflow using a multi-objective algorithm with fuzzy resource utilization in a multi-cloud environment," *IEEE Access*, vol. 8, pp. 24309–24322, 2020.
- [5] R. Li, Q. Zheng, X. Li and Z. Yan, "Multi-objective optimization for rebalancing virtual machine placement," *Future Generation Computer Systems*, vol. 105, pp. 824–842, 2020.
- [6] S. E. Zahra, M. A. Khan, M. N. Ali and S. Abbas, "Standardization of cloud security using mamdani fuzzifier," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 3, pp. 292–297, 2018.
- [7] N. S. Naz, S. Abbas, M. A. Khan, B. Abid and N. Tariq, "Efficient load balancing in cloud computing using multi-layered mamdani fuzzy inference expert system," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 3, pp. 569–577, 2019.
- [8] S. Xuan and K. Dohyeun, "Performance analysis of IoT services based on clouds for context data acquisition," in *IEEE 3rd Int. Conf. on Big Data, Cloud Computing, and Data Science Engineering*, San-Francisco, Chicago, Cham. Springer, pp. 81–92, 2018.
- [9] H. Haddadpajouh, A. Dehghantanha and R. Khayami, "A deep recurrent neural network based approach for internet of things malware threat hunting," *Future Generation Computer Systems*, vol. 9, no. 3, pp. 315–321, 2018.
- [10] K. P. N. Jayasena, L. Li and Q. Xie, "Multi-modal multimedia big data analyzing architecture and resource allocation on cloud platform," *Neurocomputing*, vol. 253, pp. 135–143, 2017.
- [11] C. Melo, "Availability models for hyper-converged cloud computing infrastructures," in *12th Annu. IEEE Int. Systems Conference*, Vancouver, British Columbia, Canada, pp. 1–7, 2018.
- [12] S. S. Wagle and P. Bouvry, "Cloud service providers ranking based on service delivery and consumer experience," in *IEEE 4th Int. Conf. on Cloud Networking*, Niagara Falls, Ontario, Canada, pp. 209–212, 2015.
- [13] L. Wang, Y. Ma, J. Yan, V. Chang and A. Y. Zomaya, "Pipscloud: High-performance cloud computing for remote sensing big data management and processing," *Future Generation Computer Systems*, vol. 78, pp. 353–368, 2018.
- [14] T. H. Noor, Q. Z. Sheng, A. H. H. Ngu and S. Dustdar, "Analysis of web-scale cloud services," *IEEE Internet Computing*, vol. 18, no. 4, pp. 55–61, 2014.

- [15] A. Haider, M. A. Khan, A. Rehman, M. U. Rahman and H. S. Kim, "A real-time sequential deep extreme learning machine cybersecurity intrusion detection system," *Computers Materials & Continua*, vol. 66, no. 2, pp. 1785–1798, 2021.
- [16] M. A. Khan, W. U. H. Abidi, M. A. A. Ghamdi, S. H. Almotiri, S. Saqib *et al.*, "Forecast the influenza pandemic using machine learning," *Computers Materials & Continua*, vol. 66, no. 1, pp. 331–357, 2021.
- [17] M. A. Khan, S. Abbas, A. Rehman, Y. Saeed, A. Zeb *et al.*, "A machine learning approach for blockchain-based smart home networks security," *IEEE Network, Early Access*, vol. 35, no. 1, pp. 1–8, 2021.
- [18] M. W. Nadeem, M. A. A. Ghamdi, M. Hussain, M. A. Khan, K. M. Khan *et al.*, "Brain tumor analysis empowered with deep learning: A review, taxonomy, and future challenges," *Brain Sciences*, vol. 10, no. 2, pp. 118–135, 2020.
- [19] M. A. Khan, S. Saqib, T. Alyas, A. U. Rehman, Y. Saeed *et al.*, "Effective demand forecasting model using business intelligence empowered with machine learning," *IEEE Access*, vol. 8, pp. 116013–116023, 2020.
- [20] M. A. Khan, S. Abbas, K. M. Khan, M. A. A. Ghamdi and A. Rehman, "Intelligent forecasting model of covid-19 novel coronavirus outbreak empowered with deep extreme learning machine," *Computers Materials & Continua*, vol. 64, no. 3, pp. 1329–1342, 2020.