International Journal of Computer Sciences and Engineering
Vol.**11**, Issue **2**, pp.**12-17**, February **2023**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

ISSN: 2347-2693 (E)

Research Paper

Analysis of Various Performance Measurement Parameters (PMP) of Load Balancing Algorithms Used in Public Cloud Environment

Mangal Nath Tiwari[1*], Nagesh Salimath[2], Vijay Anand Sullare[3]

[1,2,3]Dept of Computer Science, Madhyanchal Professional University, Bhopal, India

**Corresponding Author: mangal.tiwari81@gmail.com*

Received: 23/Jan/2023; **Accepted**: 04/Feb/2023; **Published**: 28/Feb/2023. **DOI**: https://doi.org/10.26438/ijcse/v11i2.1217

Abstract: Cloud computing is on demand and hosted solution mechanism on the internet which provides the facilities to the individual and industry to use needed infrastructure and application via Internet. It reduces infrastructure setup cost of resources and offers cost efficiency, scalability, availability, reliability and flexibility. Cloud computing offers a competitive edge over your competitors, in this way cloud services helps the users to access the latest applications any time without spending time and money on installing and mainlining hardware and software resources [1]. Load balancing is a technique in which load balancers are responsible to manage and maintain the workload on the back end servers, It also offer the mechanism to distribute the workload across the servers evenly in such a way that no any server is found under loaded or overloaded. Load balancing improves application responsiveness and overall system performance. In this paper, an attempt has been made to study a few of the existing performance measurement matrices of load balancers that distribute the workload across the back end servers and also analyse the various performance metrics that affect the load balancing process. The load balancer applies several load balancing algorithms to determine the appropriate resources. However, it faces several problems while distributing the load across the available resources. The main objective and motivation of this paper is to do a comparative study of different types of existing performance measurement matrices that are used by load balancer to manage the workload distribution task in cloud computing environment.

Keywords: Cloud computing, performance measurement matrices, load balancing techniques, load balancer, public cloud, and cloud services.

1. Introduction

Cloud computing is to host digital materials on the server and provide services for the users, which provides clients using the Internet services. Through cloud computing, the user can store, access and share any type of digital content as per their need on pay per use basis policy. This service provides facilities to the user to use hardware or software provisions, maintenance and upfront very easily without any investment [5]. Cloud computing models provide information about any type of digital content without any restriction and concern for the user by the computer, which the user can get hosted services. Thus, by using this technology user can focus on their business police rather than spending lots of time and money to needed to manage their business process [7]. Here the cloud service provider also ensures what, how and which digital materials users use as per their need and requirement and which materials should be restricted and maintained without any concern.

Cloud Computer Technology is a distribution system of resources and services earlier on the data centre server, which provides various essential and useful structures. With the help of its basic structure, the service can be made separately by different people and dividing the service keeping in minds the boundaries of physical hardware [3]. These servers can be effectively managed by data centres, and hence any use this technology without any physical boundary I.e. Services offered by Cloud Service Providers can be used from Anywhere, Anytime and heterogeneous digital devices [8]. Since, users use cloud services as per their requirement with the help of internet. These users access the preserved data on more than one cloud at the same time, so the server faces heavy traffic servers. Therefore, it is mandatory to manage and provide traffic on the server, in this situation, the only option as a solution to this type of problem is the only option load balanced technology[2]. Load balancing is an important role of a lodged balancer in balanced technology, which is responsible for traffic management. The load helps us full transferring the request of one server to another to enhance the efficiency of the server and to maintain the application, so that a healthy result can be achieved.

2. Related Work

There has been some work done in load balancing performance analysis and various load balancing algorithms in cloud environments some of which we will consider in this section.

A comparative study of two round robin and throttled virtual machine load balancing algorithms is proposed in [26]. In this study, round robin and throttled virtual machine load balancing policies are used with optimized response time service broker policy and the overall response time, datacenter hourly average processing time, Datacenter request servicing time is simulated by adjusting the parameters to observe according to the response time. Region, user base, hourly response time and total cost which have a significant impact on performance. According to the simulation results, the combination of the proposed strategy of throttle and optimized response time service broker policy has better performance than the round robin load balancing algorithm in a heterogeneous cloud computing environment.

In [27] the authors present a review of some load balancing algorithms in cloud computing to identify qualitative components for simulation and to analyze the execution time of load balancing algorithms. In this study, the simulation process is executed for three load balancing algorithms: round robin, central queuing and randomized with million instructions per second VM versus a MIPS host.

A comparative study of three distributed load balancing algorithms for cloud computing scenarios has been proposed in [28]. In this study three representative algorithms were chosen for comparing performance evaluation. The first was directly based on naturally occurring phenomenon, honey bee foraging, the second sought to engineer a desired global outcome from biased random sampling, while the third used system rewiring which is called Active Clustering.

A comparative study of three distributed load balancing algorithms for cloud computing scenarios is proposed in [28]. Three representative algorithms were chosen to compare the performance evaluation in this study. The first was based directly on a naturally occurring phenomenon, honey bee foraging, the second sought to engineer a desired global outcome from biased random sampling, while the third used systems rewiring called active clustering.

In [29] the authors discuss the performance comparison of various load balancing algorithms of virtual machines and policies in cloud computing. Four well-known load balancing algorithms are considered in this study. The performance of round robin, throttled, execution load, and first-come-first-served load balancing algorithms is analyzed based on average response time, average datacenter request service time, and total cost. According to the Cloud Analyst simulator, simulation results show that round robin has the best integration performance.

3. Public cloud and its benefits

Public Cloud Model: A public Cloud uses the tradition scenario of Cloud computing with the capability to use the resources from anywhere in the world with internet. The public Clouds can be used in a pay per use basis, meaning that just the resources that are being used will be paid by transaction fees [9]. Today the public Cloud is a very popular technology for huge amount of data storage purposes because data stored on it can be backed up and accessible from any location [6].

3.1 ***Benefits of Public Cloud:***

- Presently many IT organizations are engaged in using public cloud without any concern for physical resources and expansion of IT infrastructure. There are many benefits to using a public cloud infrastructure, which can help save money, time, and cost. Some of the important benefits of using public cloud computing is as follows:

- ***Maintenance cost is very less:*** Since cloud computing is more cost-effective to manage and difficult to maintain, the cost of maintaining IT equipment in public cloud-based services is also managed by the cloud service provider. Other costs can be avoided by the public cloud and maintenance is also less.

- ***Equipment purchase cost is less:*** In public cloud service, we do not have to worry about buying different types of equipment as per the requirement, because the user buys the equipment only when he needs it. Using public cloud–based applications is often less expensive than managing different software packages and other related IT equipment [5].

- ***It provides hassle-free Infrastructure management:*** Lets the user concentrate on their business functions and operations rather than purchasing, managing and maintaining the complete hardware and software. Our cloud service providers are responsible for all cloud managed tasks and do it for us.

- ***Public Cloud provides high scalability to its users***: Users of the public cloud use a service with high scalability, allowing them to scale up or scale down their resources as needed.

- ***Cost effectiveness:*** Users of a public cloud pay only for the resources they use. Users do not need to make any other form of payment.

4. Load Balancing Techniques

Load balancing is fully distributed and software defined solution that user traffic or workload across the multiple server. Load balancing technique ensure that neither any server sits idle (under load) nor it will be overloaded means each server handling user request has equal workload to process. Application responsiveness, maximize application availability are the main objective of load balancing. The process of managing and distributing the workload across

multiple servers evenly is the responsibility of Load Balancers [7]. Modern applications cannot run without load balancers in cloud computing environment today.

Load balancer also perform health check-up of the servers to know that whether they are able to handle user request or not, it means load balancer check the load status of each server before distribution of workload. If required the load balancer remover unhealthy server until they are restored [23].

Following figure depicts the process of work load distribution by load balancers on the cloud environment. As Load balancers takes the incoming request by the end users through internet and distribute them across the available application servers on public cloud.

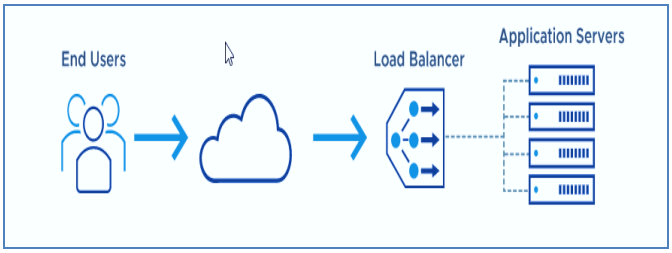

Figure 1: Load Balancing Technique

5. Need and Benefits of Load Balancing

Load balancing is fully distributed and software defined solution that user traffic or workload across the multiple server. Load balancing technique ensures application's availability, scalability, security, and performance. In cloud computing environment, millions of users request come across the servers via internet, it is needed to manage the entire user request properly. Hence the term Load Balancing is used to manage and make proper distribution of this user traffic of request [22]. Fair distribution of work load among server pool is possible with the load balancing technique in order to obtain highest system performance. The main goal of load balancing technique is to maintain the workload to each server in the server farm such that all backend servers neither overloaded nor under loaded. It means each server has equal workload at any point of execution time of application. Load balancing also aims to minimize the response time maximize the throughput for tasks execution and improve resource utilization, which enhances system performance at lower cost. Load balancing also aims to provide scalability and flexibility for those applications whose size may increase in future and requires more resources as well as to provide priority to jobs that need instant execution as compared with other jobs [23]. Other objectives of load balancing are reducing energy consumption and carbon emission, avoiding bottlenecks, resource provisioning. There is a need for proper workload mapping and load balancing techniques that consider different matrices. Following are the main reason that why we need Load Balancing for Cloud Computing.

1. To improve service availability on the network.
2. To increase the resource utilization.
3. To minimize the waiting time of the services and task execution.
4. To maximize the overall performance of services.
5. To build a system which efficiently tolerate fault occurred during service delivery.
6. To build a scalable and cost effective system.

Following figure shows the benefits of load balancing techniques for cloud computing environment. Some main benefits are high service availability, High Resource Utilization, High Reliability, High Scalability, and Fault Tolerant System.

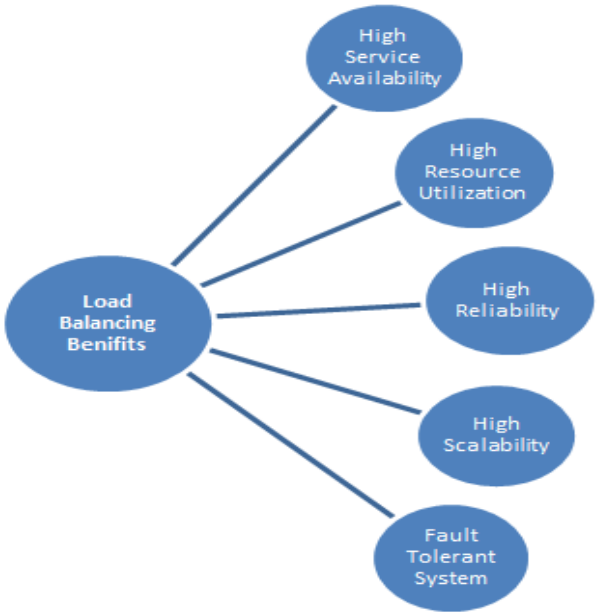

Figure 2: Benefits of Load Balancing techniques for public cloud.

6. Need to measure the performance of Load Balancers

To ensure no single server is exhausted and clients can get uninterrupted service, load balancers perform the task of load distribution evenly across the multiple servers with optical performance of application. It also provides network security to application data of the client requested for the on demand services [25]. Following are the major issues for which it is essential to measure the performance of load servers for cloud computing environment.

- Recognize the barriers, if any and fix them in the system.
- Offer unparalleled client experience.
- Prevent bottleneck in the backend.
- Maintain and optimize system health.
- Boost efficiency and accuracy.

7. Various Performance Measurement Parameters of LB Algorithms

Load balancing is just like a traffic management by a traffic policeman. In the same way with cloud computing scenario, a load balancer act as a "traffic cop" which is configured in front of server and manage the all incoming service request across all the servers in a manner that maximizes the speed,

system performance, high resource utilization and high service availability. Load balancers also ensure that no one server is under loaded or idle and likewise no one server is overloaded, which cloud degrade the system performance [11, 12].

To measure or test the performance of any algorithm which is used by Load Balancers for balancing or managing the Load across the servers in the network, there are several following measurement matrices or testing parameters has been studied by many researchers. Some of these parameters are as follows.

1. **Response Time:** This is the time slice taken by an algorithm to respond to a request. Therefore, for better performance, efficient and stable working of the system, It ought to be ***minimum*** response time. An algorithm which takes minimum response time to react for its request is considered as the best load balancing algorithm.
2. **Throughput:** How much work has been done? The answer of this question can be obtained from throughput. Actually by this measurement parameter, we can measure the total number of request has been successfully completed per unit time. For better performance and effective optimization, value of throughput parameter ought to be ***maximized***. It means maximum throughput leads to increase the proficiency of the system. In other word we can say higher throughput means higher efficiency of the load balancer.
3. **Scalability:** Scalability is the ability to accommodate (increase or decrease) IT resources as needed to meet the changing demand. For Improving efficiency and performance of the system the value of scalability for any load balancing algorithm ought to ***high***. It also improves overall availability of IT resources and increase throughput of the system.
4. **Resource Utilization:** Response time is the time algorithms take to respond to a request. Resource utilization is a major issue and important performance measurement parameter that measures how effectively each resource is utilized against their availability or capacity.
   Using this metric, resource managers can evaluate and continuously monitor the responsibility of each resource available and occupied by the algorithm, and also analyse if they are overworked or underworked. In short resource utilization is the measure to analyses the usage of resources needed for the system to work properly and efficiently. This matric ought to be ***optimal*** for better and efficient system performance.
5. **Migration Time:** During the execution time of the algorithm, it is needed to shift or transfer a request from one node to another node; the time needed to reallocate the resources from one node to another is called the migration time. For better system performance and efficiency of load balancer, the value of the migration time ought to be ***minimized.***
6. **Fault Tolerance:** Fault or failure leads to any system break down or shut down. A fault may occur as network fault, physical fault, process fault, service expiry fault, these fault leads to degradation of system performance. Fault tolerance is the capacity of an algorithm to work consistently regardless of any failure (network, physical, process, service expiry fault) in the system.
   Fault tolerance enables the system to serve the request even some of the components are not working [19]. This is another important parameter that helps us understand how well the load balancer performs at the time of load management.
7. **Reliability:** Reliability is yet another measurement matric of measuring performance of any load balancer to work better for the system. We have to measure reliability to check the efficiency and performance of the system. High reliability leads to better performance of any load balancer [24].

8. Results and Discussion

Following table shows the result of comparative analysis of various existing performance measurement matrices or parameters of different surveyed load balancing algorithms for public cloud computing environment.

Following table represent the standard value of each measurement matric for the better system performance. Any load balancing algorithm has following matrices value is said to be best algorithm. This table also shows that what value should have any measurement matrices of any load balancing algorithm to obtain better performance of the system. As the value of response time and throughout measurement matrices ought to be minimum and maximum respectively to enhance the performance of work distribution task.

Table 1. Standard values of Performance measurement matrices

SN	Measurement Matrices	Efficient Value
1.	Response Time	Minimum
2.	Throughput	Maximum
3.	Scalability	High
4.	Resource Utilization	Optimal
5.	Migration Time	Minimum
6.	Fault Tolerance	High
7.	Reliability	High

Following table depicts the comparison of different surveyed load balancing algorithm under the different necessary performance measurement matrices or parameters.

Table 2. Standard values of Performance measurement matrices

Measurement Parameter	**Honey bee**	**Biased Random**	**Active Clustering**	**Join Idle Queue**	**Min-Min**
Throughput	No	No	No	No	Yes
Overhead	No	Yes	Yes	Yes	Yes
Fault Tolerant	No	No	No	No	No
Migration Time	No	No	Yes	No	No
Response Time	No	No	No	Yes	Yes

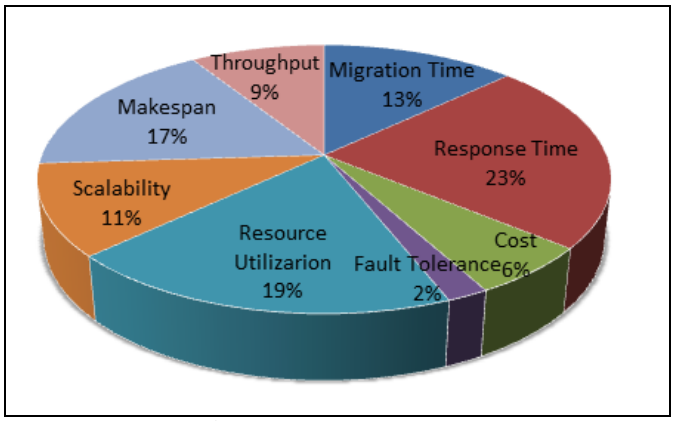

Figure 3. Percentage of load-balancing matrices in existing techniques. [17]

.

6. Conclusion and Future Scope

This research paper is to study and analyse commonly used load balancing algorithms and their performance measurement metrics in Cloud Computing environment. Because load balancing is essential to optimal performance of cloud computing systems, understanding performance metrics and optimizing these metrics helps ensure that the system performs to the best of its potential. It is proposed that in the future we may envision algorithms that combine existing load balancing algorithms such as weighted round robin with task prioritization to come up with an efficient load balancing strategy. Further work needs to be done to develop such load balancing algorithm that meets the standard value of its performance measurement matrices. This work is also a great motivation to present some efficient and optimal load balancing algorithms that will satisfy all the performance measurement matrices to meet up the better and highest system performance.

Conflict of Interest

All authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this paper. We have no conflicts of interest including any financial, personal or other relationships with other people or organizations that could inappropriately influence to disclose. All authors declare that they have no conflicts of interest.

Funding Source

We do not have received any funding, supporting grants from any sources to complete this work. Funding source for this work is: none

Authors' Contributions

Each and every author contributed their effort evenly to complete this work. Author-1 researched literature and conceived the study. Author-2 involved in figure and facts designing, development, data analysis. Author-3 wrote the first section of the manuscript and has the responsibility of data manipulation work presented in all data tables. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Acknowledgements

I would like to express my gratitude towards Dr. Atul Kumar Dwivedi and Dr. Pravesh Kumar Dwivedi, Dr. Sandeep Kumar Sharma, Associate Professor, Department of Computer Science and Engineering, Tata Institute of Technology, Sidhi, MP, India, for his expert guidance, constructive suggestions and timely feedback to complete this work.

References

[1]. Dimple Juneja, Atul Garg, "*Collective Intelligence based framework for load balancing of web servers*", IJICT, Vol 3 No.1, 2012.

[2]. Jayant Adhikari, Sulbha Patil, "*Load Balancing the Essential Factor in Cloud Computing*", International Journal of Engineering Research and Technology. ISSN: 2278- 0181 Vol.1 No.10, December 2012.

[3]. Mayank, Atul, "*Comparative Survey of Load Balancing Algorithms in Cloud Computing Environment*", IJDCC, Vol.1, No.2, 2013.

[4]. Tushar, Jignesh, "*A Survey on Various Load Balancing Techniques and Challenges in Cloud Computing*", IJSTR, Vol.2, No.11, 2013.

[5]. Tiwari M., Gautam K., and Katare K., "*Analysis of Public Cloud Load Balancing using Partitioning Method and Game Theory*", International Journal of Advanced Research in Computer Science and Software Engineering, 4(2): pp. 807-812, (2014).

[6]. A.K. Singh et al, "*Scheduling with load balancing in Cloud computing*", *International Journal of scientific engineering and research*, vol.2, no. 1, pp. 38-43, Jan-2014.

[7]. Tiwari M. et al., "*Perspective study of Public Cloud: A highly scalable Cloud deployment model*", International Journal of Science & Engineering Development Research, ISSN: 2455-2631, Vol.7, Issue 9, page no.848 - 853, September-2022.

[8]. Singh,A., Sahu, S., Gautam, K.,&Tiwari, M., *"Private Cloud Scheduling with SJF, Bound Waiting, Priority and Load Balancing*", International Journal of Advanced Research in Computer Science and Software Engineering,4(1), 367-372, 2014

[9]. Armbrust, M. et al., "*Above the Clouds: A Berkeley View of Cloud computing*", Technical Report No. UCB/ EECS-2009-28, University of California, Berkley, 2009.

[10]. Ananya Jain, 2021, "*Study on Load Balancing Algorithms and Performance Metrics in Cloud Computing*", International Research Journal of Engineering and Technology (IRJET), Volume 08, Issue 07, July 2021.

[11]. Marc, Anthony, "*Strategies for Dynamic Load Balancing on Highly Parallel Computers*", IEEE Trans. on Parallel and Distributed Systems, Vol.4, No.9, 1993.

[12]. Amandeep Kaur Sidhu and Supriya Kinger, "*Analysis of Load Balancing Techniques in Cloud Computing*", International Journal of Computers & Technology, Vol.4 No. 2, March-April, 2013.

[13]. Meenakshi, "*Comparative Study of Load Balancing Algorithms in Cloud Computing*", Environment, International Journal of Engineering Research and Technology (IJERT), Volume 02, Issue 10, October 2013.

[14]. Amit Kumar Dhingra, Dr Dinesh Rai, "*Study on the Performance Analysis Tools and Techniques of Cloud Computing*", Journal of Critical Reviews, Volume 7, Issue 15, 2020.

[15]. Sushil Kumar, Deepak Singh Rana, "*Various Dynamic Load Balancing Algorithms in Cloud Environment: A Survey*", International Journal of Computer Applications, Volume 129 – No.6, pp-0975 – 8887, November 2015.

[16]. Versha Thakur, Sanjay Kumar, "*Load Balancing Approaches: Recent Computing Trends*", International Journal of Computer Applications, Volume 131, No.14, pp-0975–8887 December2015.

[17]. Pawan Kumar and Rakesh Kumar, "*Issues and Challenges of*

Load Balancing Techniques in Cloud Computing: A Survey". ACM Computing. Survey. 51, 6, Article 120, February 2019.
[18]. Rajesh and Jeykrishan, "*A Survey on Load Balancing in Cloud Computing Environment*", IJARCEE, Vol.2, No.12, 2013.
[19]. Gokhroo, M.K.; Govil, M.C.; Pilli, E.S., "Detecting and mitigating faults in cloud computing environment", *In Proceedings of the International Conference on Computational Intelligence & Communication Technology (CICT)*, Ghaziabad, India, 9–10 February 2017.
[20]. Bushan and Urmila, "*Pros and Cons of Load Balancing Algorithms for Cloud Computing*", International Conference on Information Systems and Networks, 2014.
[21]. Ritika and Harjot, "*Load Balancing Techniques using Mobile Agents*", IPASJ-IIJCS, Vol.2, No.12, 2014.
[22]. Salimath, Nagesh and Kavitha, Dr. C. and Sheetlani, Dr. Jitendra, "Analysis of Resource Management and Security Management in Cloud Computing Environment", Proceedings of the Second International Conference on Emerging Trends in Science & Technologies for Engineering Systems, ICETSE, 2019.
[23]. Nagesh Salimath & C. Kavitha, "*Secure data compression and error detecting codes for networks and cloud storage*", International Journal of Research and Analytical Reviews (IJRAR), Vol.6, Issue 2, page no. 452 - 458, May-2019.
[24]. Nagesh Salimath & C. Kavitha, "*Review of error detecting codes for networks and cloud storage*", International Journal of Emerging Technologies and Innovative Research, ISSN: 2349-5162, Vol.6, Issue 5, page no. pp 370-375, May-2019.
[25]. Ray, S. and A. De Sarkar, "*Execution analysis of load balancing algorithms in cloud computing environment". International Journal on Cloud Computing: Services and Architecture,* IJCCSA, 2(5), 2012.
[26]. Randles, M., D. Lamb, and A. Taleb-Bendiab. "A comparative study into distributed load balancing algorithms for cloud computing". Advanced Information Networking and Applications Workshops (WAINA), IEEE 24th International Conference on. 2010. IEEE.
[27]. Mohapatra, S., K. Smruti Rekha, and S. Mohanty, "*A Comparison of Four Popular Heuristics for Load Balancing of Virtual Machines in Cloud Computing*". International Journal of Computer Applications, 68(6): pp. 33-38, 2013.

AUTHORS PROFILE

Mangal Nath Tiwari received his Master of Philosophy Degree in Computer Science from APS University Rewa, MP, India and pursuing Ph.D in Computer Science from Madhyanchal Professional University, Bhopal, India. Currently working as Professor and Dean, Faculty of Science and IT, Tata Institute of Technology, Sidhi, MP. His research interests include Cloud Computing, Algorithm and Design, Distributed Computing, Machine Learning and Artificial intelligence, Data Science. He has authored more than 6 research papers in reputed international journals and conference proceedings.

Dr. Nagesh Salimath, He is a dynamic person with experience of more than 9+ years in both industry and academic. He has completed his Ph.D. in Computer Science from Sri Satya Sai University and Medical Sciences, Sehore, MP, India. His area of Interest is on Cloud Computing, Machine Learning, Internet of Things (IoT), Wireless Sensor Network, Security and Big Data.
He has authored and published 4 books and has 3 patents. He has presented and published 17 research papers in different International as IEEE, Springer, Scopus, and UGC care journals. He has delivered many keynotes as resource person on Cloud Computing, Networking, and Software Testing and Role of ICT in NAAC. He is committed member, reviewer and session chairperson for many international conferences. He is single point contact (SPoC) for national program technology enhanced learning (NPTEL), IIT Madras and also he is coordination member of spoken tutorial IIT Bombay.

