



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 1, January 2020

Load Balancing Algorithms in Cloud Computing: Survey Paper

Nek Narayan Thakur¹, Dr. Satya Ranjan Patra²

Research Scholar, Dept. of Computer Science & Engineering, Bhopal Institute of Technology & Science, M.P,
India¹

Head/ Associate Professor, Dept. of Computer Science & Engineering, Bhopal Institute of Technology & Science,
M.P, India²

ABSTRACT: Cloud computing is Internet-based computing that offers on-demand "IT services as a service." Cloud Computing is a rapidly growing computational research and industry field. Task scheduling for scheduling tasks is required in Cloud Computing. Task scheduling in cloud means selecting the best option available for task execution. Task scheduling is primarily aimed at running user tasks within time limits and improving system reliability. But task scheduling without Load Balancing is not successful because some nodes are heavily loaded while others are idle or doing little work which leads to performance degradation. Cloud load balancing is the method of task transfer from a heavily loaded VM to an idle VM or less loaded VM. We're researching various Load Balancing Algorithms in Cloud Computing in this article.

KEYWORDS: Cloud Computing; Load Balancing; Load Balancing Algorithms

I. INTRODUCTION

1.1 Cloud Computing

Cloud computing is recognized as a hottest technology that will in the near future have an important effect on the IT sector. Cloud computing is a fast-growing field of technical science and industry in the present time. Cloud computing is Internet-based computing that offers on-demand "IT services as a service"[1]. Customers of cloud computing will not have to pay for the network and its installation[2]. Infrastructure as a service (IaaS) , Platform as a service (PaaS), Software as a service (SaaS) are known as cloud software models. Google App Engine, Gmail, Google Docs, Microsoft Windows Azure, and the Amazon Elastic Compute Cloud (EC2) are examples of cloud services.

1.2 Cloud Deployment Models [1]

Cloud Deployment models are classified as Public, Private, Hybrid and Community clouds.

- **Public cloud:** Public cloud services are accessible over the internet and are owned by a cloud provider. Public cloud service may be free or offered as pay-per usage. Many popular cloud services are public clouds including Amazon EC2, Google App Engine and Force.com.
- **Private cloud:** Private cloud is owned by a single organization or company. In this, services and infrastructures are maintained on a private network. Private cloud provides security.
- **Community cloud:** Several organizations jointly construct and share the same cloud infrastructure, their necessities and policies then such a cloud model is called as a Community cloud.
- **Hybrid cloud:** Hybrid cloud is an integrated cloud service in which private, community, or public clouds are collaborating together, offering the benefits of multiple deployment models.

1.3 Service Models of Cloud [2][3]

- **SaaS (Software-as-a-Service):** This model provides applications to a cloud's end user. In this model, users use the launched application on cloud infrastructure. Interfaces for these applications are browsers, and do not require



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 1, January 2020

installation. Gmail is the best example of this model.

- **PaaS (Platform-as-a-Service):** In this model, users hired platforms or operating systems and they can expand their required programs on it. The principal users of this layer are developers who develop and run a cloud application on a particular platform. Google App Engine is the example of this model.
- **IaaS (Infrastructure-as-a-service):** This model is related with a virtual machine and users can access to infrastructures with virtual machine.

1.4 Advantages of Cloud Computing [2][4]

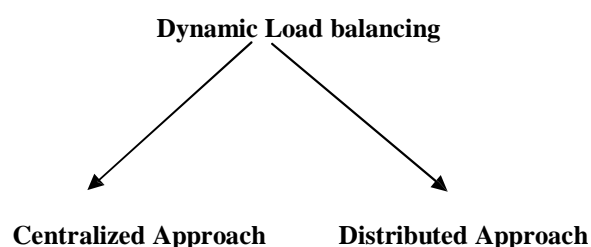
1. **Pay-As-You-Go:** Users only need to pay only for those resources which are used by them.
2. **Better Hardware Management:** It is easy for cloud service provider to handle the hardware easily because all computers run the same hardware.
3. **Disaster Management:** In case of disasters, an offsite backup is always useful. Keeping essential data backed up using cloud storage services is the need of the hour for most of the organizations. Also cloud storage services not only keep your data off site, but they also guarantee that they have systems in place for disaster recovery.
4. **Do more with less:** With Cloud Computing, companies can reduce the size of their own data centers or remove their data centre footprint altogether. The reduction of the numbers of servers, the software cost, and the number of staff can considerably reduce IT costs without impacting an organization's IT capabilities.
5. **Always-on availability:** Most cloud providers are very reliable in providing their services, with many maintaining 99.99% uptime. The connection is always on and as long as workers have an internet connection, they can get to the applications they need from almost anywhere. Some applications even work off-line.
6. **Flexible capacity:** Cloud is the flexible facility that can be turned up, down or off depending upon situation. For example, a sales promotion might be wildly popular, and capacity can be added rapidly to avoid crashing servers and losing sales. When the sale is more, capability can minimize to reduce costs.

II. LOAD BALANCING

2.1 Load Balancing: Load Balancing is one of the hottest issues in Cloud Computing. There can be various types of load like CPU load, delay or network load and memory capacity. Load Balancing in cloud is the process of distributing the work load among various nodes in a distributed system for optimal resource utilization, maximize throughput and job response time. Load Balancing guarantees that the entire the processor in the system or each node in the network does approximately do the equal amount of work at any time. [5].

There are basically two types of Load Balancing: [5]

- **Static Load Balancing:** This approach is mainly defined in the design or implementation of system. Static Load Balancing algorithm divides the traffic equally between all users. It uses only information about the average behavior of the system. These are much simpler and ignore the current state or the load of the node in the system and there is no need to regularly monitor the nodes for performance statistics.
- **Dynamic Load Balancing:** Dynamic Load Balancing algorithms allows the partition of work between nodes at run-time; they use current or load information while making allotment decisions. The advantage, of such approach is that in case if one or more nodes in the network fail, it will not slow down the total Load Balancing process; instead it will improve the system performance to a little extent.





International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 8, Issue 1, January 2020

These are described as follows:

- **Centralized Approach:** Only a single node is responsible for managing and sharing the load within the whole system.
- **Distributed Approach:** In Distributed Approach, no single node is responsible for making resource provisioning or Task scheduling decisions. Multiple domains observe the network to make precise Load Balancing decisions.

2.2 Metrics for Load Balancing: [6]

1. **Throughput:** It is used to calculate the all tasks whose execution has been completed. The performance of any system is enhanced if throughput is high.
2. **Fault Tolerance:** It means recovery from failure. The Load Balancing should be a superior fault tolerant technique.
3. **Migration time:** It is the time to transfer the jobs or resources from one node to other nodes. It should be minimized in order to improve the performance of the system.
4. **Response Time:** The time taken by a particular Load Balancing algorithm to response a task in a system. Response time should be minimized for better performance of a system.
5. **Scalability:** It is the capability of an algorithm to execute Load Balancing for any restricted number of nodes of a system. This metric should be enhanced for a good system.

III. LITERATURE REVIEW

Buyya et.al (2009) Author defines Cloud Computing in terms of its services to end users: “A Cloud is a market-oriented distributed computing system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resource(s) based on service level agreements established through negotiation between the service provider and consumers”[7].

L.Agostinho, G. Feliciano et.al (2011) It implements a cloud scheduler, based on Ant Colony Optimization (ACO) to allocate VMs to the physical resources belonging to a cloud. Experiments were carried out using Cloudsim and comparisons had been done with alternative cloud schedulers including–random a scheduler based on genetic algorithms. The scheduler allows fair assignment of VMs and delivers competitive performance with respect to the number of executed jobs per user [8].

Pandey S et.al (2010) Author describes that the Particle Swarm Optimization technique for Cloud Computing which is being inspired by the movement of birds or fishes in school. In PSO, each single solution is a “bird” in the search space which is called as “particle”. Fitness values of all the particles have evaluated by the fitness function to be optimized, and have velocities which express the flying of the particles. In each iteration, each particle is updated by following two “best” values pbest and gbest and these two values are need not be evaluated in Ant Colony Optimization algorithm. Moreover ACO is more applicable for problems that require crisps results and PSO is applicable for problems that are fuzzy in nature [9].

Kumar Nishant et.al (2012) Author describes that an Ant Colony Optimization has been connected from the point of view of cloud with the primary point of Load Balancing of nodes. This modified algorithm has an edge over the first approach in which every ant constructs their own particular individual result set and it is later on assembled into a complete arrangement. Further, as it is realized that a cloud is the gathering of numerous nodes, which can strengthen different sorts of application that is utilized by the customers on a premise of pay for each utilization. In proposed approach the ants constantly update a single result set rather than updating their own result set. As cloud support various types of application that is used by the clients on a basis of pay peruse. Therefore, such a system should function smoothly and should have algorithms that can continue the proper system functioning even at peak usage hours [10].

Mishra Ratan et.al (2012) Author shows that, the Ant Colony Optimization technique which is a Load Balancing technique is based on Swarm Intelligence. The motive of this paper is to produce an effective Load Balancing algorithm using Ant Colony Optimization technique to enhance or decrease different performance parameters like CPU load, Memory capacity, Delay or network load for the clouds of different sizes. They have also explained the working of a load balancer that how it works and what the various phases of the load balancer. In this paper, a heuristic



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 1, January 2020

algorithm is proposed which is dependent on Ant Colony Optimization. [11]

Fatemeh Rastkhadiy et.al (2016) Objective of this study is to achieve well balanced load across virtual machines and deliver to the minimum makespan and an algorithm is evaluated for different conditions and datasets. It requires two lists one containing the tasks that have been sent by the user, and the other list contains all user VMs that are already allocated to a physical resource and scheduler works dynamically. This approach works well for heterogeneous Cloud Computing systems. The algorithm can handle all conditions, and outperforms FCFS and ACO algorithms in Cloud Computing environment. [12]

L.D. Dhinesh Babu et.al (2013) Author showed the detailed information about the Honey Bee algorithm i.e. it includes the various types of flow of data in the algorithm and the other important information about implementing it and also comparing it with some already existing methods of Load Balancing. This algorithm adjusts the load, as well as additionally mulls over the needs of tasks that have been expelled from intensely stacked Virtual Machines. The tasks evacuated from these VMs are dealt with as honey bees, which are the data updater all inclusive. This algorithm likewise considers the needs of the tasks. Honey bee conduct propelled Load Balancing moves forward the general throughput of handling and need based adjusting Concentrates on diminishing the measure of time an undertaking needs to attend to a line of the VM. In this manner, it decreases the reaction of time of VMs [13].

Kruekaew B. et.al (2014) Author presents that ABC algorithm is suitable for cloud registering environment in light of the fact that the algorithm has the capacity viably use the expanded framework assets and reduce makespan. Under the situation of expanding or diminishing the quantity of servers, the heap adjusting algorithm ought to be finished by utilizing ABC_LJF (Longest Job First) algorithm as a part of request to keep up framework dependability and booking and to keep the framework crash. The results are evaluated by using the Cloudsim toolkit and the Artificial Bee Colony Algorithm (ABC) is compared with the many other existing approaches and it is proved that when Honey Bee algorithm is used in the area of artificial intelligence it have performed well when compared. [14]

Mohd Nadhir Ab Wahab et.al (2015) This shows SI algorithms that include Genetic Algorithms (GA), Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Differential Evolution (DE), Artificial Bee Colony (ABC). Ant Colony Optimization (ACO) is a Meta heuristic approach and is inspired by the foraging behavior of real ants. ABC performs quite well compared with numerous other approaches. It is inspired by the intelligent behavior of real honey bees in finding food sources, known as nectar, and the sharing of information about that food source among other bees in the nest [15].

Dervis Karaboga et.al (2007) Author describes the comparison results on the performance of the Artificial Bee Colony algorithm for constrained optimization problems It has been concluded that the ABC algorithm can be proficiently used for solving constrained optimization problems [16]

Soumya Banerjee et.al (2009) In this author presents an initial heuristic algorithm to apply Modified Ant Colony Optimization approach for the diversified service allocation and scheduling mechanism in cloud paradigm and modification supports to minimize the make span of the Cloud Computing based services and probability of servicing the request also has been converged using the modified scheduling [17]

Mohit et.al (2016) Author presents that Load Balancing is vital for efficient operations in distributed environments. Load Balancing is a way of proper utilization of resources. Many algorithms were suggested to provide various approaches for the client's request to cloud nodes. There algorithms are used to increase the performance of cloud with the aim of user satisfaction.[18]

Anju Baby et.al (2013) Author describes that Load Balancing is one of the important and challenging thing in Cloud Computing. So it is necessary to balance the load among different virtual machine in order to increase the efficiency. So in this paper, Honey Bee Foraging algorithm is compared by various algorithms. Honey bee behavior inspired Load Balancing algorithm have more advantages over other algorithms.[19]

Harshada Raut et.al (2015) Author presents that Ant Colony Optimization algorithm with travelling salesman problem is one of the well-known and broadly studied problems in discrete or combinatorial optimization and asks for the shortest round trip of minimal total cost visiting each given node. When server gets overloaded this method gives the solution for finding the minimum distance from one node to neighbouring nodes. Here, we used temporary memory to store the previous executed requests; it reduces the time if the same request has already been executed on the same node. [20]



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 1, January 2020

IV. LOAD BALANCING ALGORITHMS

4.1 Ant Colony Optimization

Ant Colony Optimization is basic foraging behaviour of an ant that encouraged those to find the optimal shortest path from their nest to food introduced by Dory go and Gambardella. ACO algorithm is a random search algorithm, like other evolutionary algorithms. Ant Colony Optimization (ACO) is one of the bio inspired algorithm inspired from the behaviour of real world ants in foraging the food and return to their colony. Ants while searching for food initially start searching randomly by laying a chemical substance called *Pheromone* and upon finding the path to the food, they return to colony via shortest path based on the pheromone intensity. The remaining ants follow the path based on the intensity of the pheromone i.e. ants follow the shorter one as the density of trail pheromone is more. However, as time passes the pheromone intensity starts evaporating, the evaporation of pheromone provides an advantage of finding optimal solution. In Cloud Analyst tool, ant is considered as cloudlet and food as virtual machine. The ants in our proposed algorithm will constantly initiate from the Head node. These ants traverse the width and length of the network in such a way that they know about the location of under loaded or overloaded nodes in the computing network. These ants while traversing will update the pheromone table, which will keep track of the resources utilization by each node. The main task of ants in the algorithm is to redistribute work among the nodes in the entire system. The ants traverse throughout the cloud network along with selecting the nodes for their next step based on the pheromone. [11][21]

4.2 Artificial Bee Colony

Artificial Bee Colony (ABC) algorithm was proposed by Karaboga [16], [22]. This method is inspired by the foraging behavior of honey bees. In this model, three kinds of honey bees are used to search food sources, which include scout bees search for food source randomly, employed bees search around the food source and share food information with the onlooker bees, and then the onlooker bees compute the fitness and select the best food source. After scout bees find the food source and return to the hive, they compare the quality of food source and go to the dance floor to perform a dance known as “waggle dance”. The waggle dance is the communication of bees to share the information about road of the food source, distance from the hive, and the nectar amount of the food source. While sharing information, bees evaluate the nectar quality and energy waste and after sharing the information on the dance floor, onlooker bees select the best food source and then scout bees will return to the food source to bring nectar back to the hive.

4.3 Shortest Response Time First

The idea of this algorithm is straight forward. In this every process is assigned a priority to be run. In this equivalent priority processes are scheduled in FCFS order. The (SJF) algorithm is a particular case of general priority Scheduling algorithm. In SJF algorithm is priority is the opposite of the next CPU burst. It means, if longer the CPU burst then lower the priority. The SJF strategy selects the job with the shortest (expected) processing time first. It works when all the jobs are available at an exact time. In this algorithm shorter jobs are executed earlier than long jobs. In SJF, it is very important to know or estimate the processing time of each job which is major problem of SJF. [5]

4.4 Throttled Load Balancing Algorithm

It is mainly based on virtual machine. In this algorithm firstly client request the load balancer to check the right virtual machine which access that load simply and perform the operation that is given by the client. [5]

4.5 Round Robin Load Balancing

The algorithm [23] is proposed called Round Robin, which uses the time slicing operation. This algorithm works in the round manner in which every node has given a time slice and has to wait for their turn. The time is divided into interval is fixed to each node in which nodes have to complete their task. It uses random selection method where first node is selected randomly and jobs are allocated to other nodes in a Round Robin manner. The key point of this algorithm is that it yields no starvation and gives a faster reaction in case of equal workload distribution among processes. But as different processes have different processing times, therefore at any time some nodes may be heavily loaded while others remain idle and underutilized.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 1, January 2020

4.6 Min-Min Load Balancing

Initially there is a task set which is not assigned to any of node [24] for all the available nodes the minimum completion time is considered. On finding the minimum time the task having the completion time minimum is chosen and assign to the respective node. The execution time of all other tasks available in that machine is updated and the task gets discarded from the available task set. Once all the tasks have been assigned to proper machine this process is repeated. The advantages of this algorithm are that it is an easy and fast algorithm yields better performance. The algorithm works superior in the situation where small tasks are more in number than larger tasks. Though, the main disadvantage is it assigns the smaller task first as, smaller tasks will get executed first, while the larger tasks keeps on in the waiting period ,which will finally results in poor machine use. It leads to starvation and it is less fault tolerant and Less Scalable.

4.7 Max-Min Load Balancing

Max-Min load balancing algorithm [25] is correlated to the Min-Min algorithm excluding the following: First it finds the minimum execution times, then the maximum value is selected which is the maximum time between all tasks on the resources. Then according to the maximum time, the task is programmed on the corresponding machine. The execution time for all other tasks is updated on that machine and the assigned task is separated from the list of tasks that are to be assigned to the machines. This algorithm has advantages over Min-Min algorithm where smaller tasks are in high numbers as compared to that of larger ones. For e.g. if in a task set only a single larger task is presented then Max-Min algorithm runs smaller tasks parallel along with larger task. This algorithm gives high throughput and optimal resource utilization. The algorithm suffers from starvation where the tasks having the greatest completion time be executed first while the tasks having the minimum completion time have left over.

V. CONCLUSION

Load Balancing is the one of the important aspects in the Cloud Computing which focuses on distributing the load over the available virtual machines. In this paper various studies and techniques related to efficient Load Balancing have been discussed. In the "Autonomous Agent Approach" the authors have used Ant Colony Optimization to find less overloaded machine.

ACKNOWLEDGEMENT

Nek Narayan Thakur would like to thank thesis guide Dr. Satya Ranjan Patra for his great effort and instructive comments in this paper work. Lastly, I wish to thank to all those who helped me during the lifetime of my MTEC H research work.

REFERENCES

1. Tharam Dillon, Chen Wu, Elizabeth. Chang, "Cloud Computing: Issues and Challenges", 24th IEEE International Conference on Advanced Information Networking and Applications, 2010.
2. Yashpalsinh Jadeja, Kirit Modi, "Cloud Computing-Concepts, Architecture and Challenges", International Conference on Computing, Electronics and Electrical Technologies, 2012.
3. Niloofar Khanghahi and Reza Ravanmehr, "Cloud Computing Performance Evaluation: Issues and Challenges", International Journal on Cloud Computing Services and Architecture, Vol. 3, 2013.
4. Kiranjot Kaur, Anjandeep Kaur Rai, "A Comparative Analysis: Grid, Cluster and Cloud Computing", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, 2014.
5. Amandeep, Vandana Yadav, Faz Mohammad, "Different Strategies for Load Balancing in Cloud Computing Environment: a critical Study", International Journal of Scientific Research Engineering & Technology ,Vol. 3, 2014.
6. Rajwinder Kaur, Pawan Luthra, "Load Balancing in Cloud Computing", Conference on Recent Trends in Information, Telecommunication and Computing, ITC, 2014.
7. R. Buyya, C.S Yeo, S. Venugopal, J.Broberg, and I. Brandic, "Cloud Computing and Emerging IT Platforms: Vision, hype, and reality for delivering computing as the 5th utility", Future Generation Computer System, 2009.
8. L. Agostinho, G. Feliciano, L. Olivi, E. Cardozo, E. Guimaraes, "A Bio-inspired Approach to Provisioning of Virtual Resources in Federated Clouds", Ninth International Conference on Dependable, Autonomic and Secure Computing (DASC), DASC 11, IEEE Computer Society, Washington, DC, USA, 2011.
9. Pandey S, Wu L, Guru S, Buyya R. "A particle swarm optimization-based heuristic for scheduling workflow applications in Cloud Computing



ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 1, January 2020

- environments”, International conference on advanced information networking and applications, 2010.
10. Kumar Nishant, Sharma Pratik, Krishna Vishal, “Load Balancing of Nodes in Cloud Using Ant Colony Optimization”, 14th IEEE Conference on Modelling and Simulation, 2012.
 11. Ratan Mishra and Anant Jaiswal, “Ant colony Optimization: A Solution of Load balancing in Cloud”, International Journal of Web & Semantic Technology, Vol. 3, 2012.
 12. Fatemeh Rastkhadiv, Kamran Zamanifar, “Task Scheduling Based on Load Balancing Using Artificial Bee Colony in Cloud Computing Environment”, International Journal of Advanced Biotechnology and Research, Vol. 7, 2016.
 13. L.D. Dhinesh Babu, Krishna P. Venkata, “Honey bee behavior inspired load balancing of tasks in cloud computing environments”, 2013.
 14. Kruekaew B. and Kimpan W., “Virtual Machine Scheduling Management on cloud computing using Artificial Bee Colony”, International Multi Conference of Engineers and Computer Scientists, 2014.
 15. Mohd Nahir Ab Wahab, Samia Nefti-Meziani and Adham Atyabi, “A Comprehensive Review of Swarm Optimization Algorithms”, PLOS ONE, 2015.
 16. Dervis Karaboga and Bahriye Basturk, “Artificial Bee Colony (ABC) Optimization Algorithm for Solving Constrained Optimization problems”, Springer-Verlag Berlin Heidelberg IFSA, 2007.
 17. Soumya Banerjee, Indrajit Mukherjee, and P.K. Mahanti, “Cloud Computing initiative using modified ACO framework”, World Academy of Science, Engineering and Technology, Vol.3, 2009.
 18. Mohit, Jitender Kumar “A Survey of Existing Load Balancing Algorithms in Cloud Computing”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 6, 2016.
 19. Anju Baby, Jenolovesum, “A Survey on Honey Bee Inspired Load Balancing of tasks in Cloud Computing”, International Journal of Engineering Research & Technology, 2013.
 20. Harshada Raut, Kumud Wasnik, “Load Balancing in Cloud Computing using Ant Colony Optimization” International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, 2015.
 21. Ping Duan, Yong AI, “Research on an Improved Ant Colony Optimization Algorithm and its Application”, International Journal of Hybrid Information Technology, Vol. 9, 2016.
 22. D.Karaboga, and B. Basturk, “On the performance of Artificial Bee Colony (ABC) algorithm”, in Applied Soft Computing, 2008.
 23. Nusrat Pasha, Dr. Amit Agarwal and Dr.Ravi Rastogi, “Round Robin Approach for VM Load Balancing Algorithm in Cloud Computing Environment”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, 2014.
 24. T. Kokilavani and Dr. D.I. George Amalarethinam, “Load Balanced Min-Min Algorithm for Static Meta-Task Scheduling in Grid Computing”, International Journal of Computer Applications, Vol. 20–No.2, 2011.
 25. Xiaofang Li, Yingchi Mao, Xianjian Xiao, Yanbin Zhuang, “An Improved max -min task scheduling algorithm for elastic cloud”, IEEE proceedings of International Symposium on Computer, Consumer and Control, 2014.