# A Survey on Resource Management Algorithms in Cloud Computing

Reshma R. Nanadikar[1], A. S. Ghotkar[2], Anil Kumar Gupta[3]

[1]*PG Student, Department of Computer Engineering, PICT, Pune, Maharashtra, India*

[2]*Associate Professor, Department of Computer Engineering, PICT, Pune, Maharashtra, India*

[3]*CDAC, Pune, Maharashtra, India*

[1] reshmarn28@gmail.com

[2] aaghotkar@pict.edu

[3] anilg@cdac.in

***Abstract***

*Cloud computing is a modern concept which provides the services through the internet. Load balancing, Resource allocation and Workflow scheduling are the key concepts of the cloud computing. The aim of these algorithms in cloud computing is the cloud resources are accessible to end users in easy and convenient manner. In that Load balancing aims to avoid the situation of overload and underload of the nodes, Resource allocation aims to increase the performance improvement by allocating the proper resource to the user and Workflow scheduling aims to optimize the objectives like makespan, cost, energy consumption, reliability etc. Therefore, this paper focuses on some different important algorithms of load balancing, resource allocation and workflow scheduling. It represents the structured and comprehensive overview of existing algorithms.*

**Keywords**— *cloud computing, load balancing, resource allocation, workflow scheduling, algorithm*

## I. INTRODUCTION

Cloud computing is the modern technology which provides the services to clients at anytime and anywhere. Cloud computing uses the services like software development platforms, servers and storage. Software, over the internet, which referred as "cloud". Cloud computing is related to the virtualization.

Cloud computing consists of the following deployment models [27]:

Private Cloud: In this model an enterprise uses its own and controlled architecture and runs within its own data centre.

Public Cloud: In this model the third-party provider provides compute resources to the public over the internet.

Hybrid Cloud: In this model a mixture of on-premises, private cloud and third-party public cloud services with arrangement between the two platforms.

Cloud computing provides the following services [27]:

Software as a Service (Saas): In this service a third-party provider hosts application and makes them available to customers over the internet.

Platform as a Service (Paas): In this service a third-party provider hosts application development platform and tools on its own organizational structure and makes them available to customers over the internet.

Infrastructure as a Service (Iaas): In this model a third-party provider hosts server, storage and makes them available to customers over the internet.

The Resource Management consists of Resource allocation, Load balancing and Workflow scheduling which are the demanding issues as well as concern in the cloud computing. In load balancing the load gets balanced that is every node gets loaded and it avoids the situation of overload and underload of the nodes. Resource allocation technique, allocates the proper resource to the user as per the requirement which results in increased QoS. Workflow scheduling is required to select the proper suitable resources for task execution which affects the execution time and cost for the user. There is importance of every method in the Resource Management.

Resource allocation should avoid the following norms such as [24] Resource confliction, Scarcity of resources, Resource fragmentation, Over utilization, Under-utilization

Load balancing ensures [25] that load gets distributed uniformly on the nodes. It improves performance of the system. It satisfies the user expectations. It gives fast response to user requests. It gives the system stability. It reduces the carbon emission.

Workflow scheduling involves [26] the Resource provisioning and Task Scheduling.

This paper is further organized as Section II describes the literature work on different algorithms of Resource Allocation, Load Balancing and Workflow Scheduling. Section III contains the tables which give overview of the algorithms. Section IV is about the conclusion and future work.

## II. LITERATURE SURVEY

In this section we survey the existing techniques for the load balancing, resource allocation and workflow scheduling in cloud computing.

### *A. Resource Allocation*

Resource allocation is an integration of cloud provider

activities for utilization and allocation of abundant resources within a specific limit of cloud environment to meet the cloud application needs. The resource allocation techniques can be categorized into following types based on the different parameters.

**Cost based Resource Allocation:**

In the paper [1] proposed OCRP algorithm with stochastic programming model and applied decomposition algorithm. The multiple cloud service providers are considered in this paper. By using this algorithm, the author concludes that the cost gets reduced for resource allocation. The reservation plan of cloud gets considered in this paper to reduce the cost.

In the paper [2] proposed the Most Cost Effective Provider's Resources First (MCEPRF) algorithm. This algorithm is introduced for the co-operative cloud market. The resource allocation cost gets reduced but QoS goals are not fully satisfied.

**QoS based Resource Allocation:**

In the paper [3] proposed Multi-objective optimization algorithm that is cost effective and runtime friendly algorithm. In this paper, the proposed algorithm gives optimal choice for deploying the web application. Branch and bound technique is used for getting the optimal condition of resources. This is done on web applications with different scenarios. By using this algorithm author concludes that the cost gets reduced as well as QoS gets achieved.

**Auction based Resource Allocation:**

In the paper [4] proposed Modified Paddy Field Algorithm to get the best consumer and provider. In this auction the both consumer and provider get participated. This gives the high-quality computing resources. The important task done by this algorithm is to detect the optimal price and the auction winner. Hence the profit gets increased.

**Priority oriented Resource Allocation:**

In the paper [5] proposed k-means algorithm for classification of tasks in high, medium and low priority. In this the priority is considered for large amount of CPU that is which job requires the high amount of CPU comes first to serve or executed first. On the basis of current resource demand and its availability the cost is calculated.

In the paper [6] proposed a priority based scheduling algorithm which considers the fault tolerance. This algorithm is applicable to perform the pre-emption from low priority task to high priority task. Also in this paper the advanced reservation scheme is presented. This scheme is given to tasks which have high priority. This scheme is applicable for fault tolerance also.

**Agent oriented Resource Allocation:**

In the paper [7] proposed a new Agent based Automated Service Composition algorithm which consists of request processing and automated service composition phases. It reduces the cost of VM which are consumed by on-demand services. The author concludes that the algorithm gives right selection which satisfies the user requirements.

**Optimization oriented Resource Allocation:**

In the paper [8] proposed a Pareto based Fruit fly optimization algorithm which solves the problem of task scheduling and resource allocation. A search operator based on critical path, a reassigning operator based on a property and a random operator which are designed for implementation of smell-based search. The cost gets reduced for resource allocation by using this optimization algorithm.

In the paper [9] proposed Multi-objective Genetic algorithm for prediction of future resource requirements. Optimal solution for resource prediction is found under any stable and unstable utilization tendency. The author gets concluded that the algorithm increases the CPU and memory utilization and decreases the energy consumption.

### *B. Load Balancing*

Load balancing is one of the main concerns in cloud computing which reduces the response time, service time, job waiting time and efficiently utilizes the resources. We have classified the existing load balancing algorithms on the basis of following categories.

**Natural phenomena-based load balancing algorithms:**

In this category the idea is inspired by nature and biological behaviour around us.

In the paper [10] proposed a Cuckoo Optimization Algorithm for load balancing. The proposed method consists of 3 steps. Firstly, it is applied to detect the hosts which are over-utilized. Secondly, for migration from the over utilized host to the other host, one or more VMs are selected. At last, it uses for selecting VMs and for that minimum migration policy has been used.

In the paper [11] proposed a Genetic Algorithm for load balancing. It tries to minimize the completion time of given task set. This algorithm collects all the possible solutions and selects the optimal solution. By this algorithm author tries to minimize the traffic on server by eliminating the inappropriate distribution of the execution time.

**Agent based load balancing algorithms:**

In this category agent plays an important role. An agent is a part of software which works automatically and decide by itself what to do for satisfy the objectives.

In the paper [12] proposed an autonomous agent-based algorithm for cloud environment. For balancing the load, it uses the load agent, channel agent and migration agent. Information policy is controlled by the load agent. Transfer policy, selection policy and location policy are controlled by the channel and migration agent. Migration agents move other data centres and also communicate with the other load agent to get the present VM's status.

In the paper [13] proposed the collaborative agent-based problem solving technique for load balancing. It consists of VM agents, server manager agents, front end agents. Experiments in this paper proved that it balances the load efficiently through autonomous and dynamic collaboration. Also, it proves that it performs better for the centralized system.

**General load balancing algorithms:**

In this category the combination of traditional load balancing algorithms are considered.

In the paper [14] proposed the dynamic load balancing with

effective bin packing and VM configuration. It consists of 3 tiers that are Web tier, schedule tier and Resource allocation tier. Web tier accepts the user requests. Schedule tier schedules the jobs as per priority. Allocation of VM is done by the Resource allocation tier. Proposed work improves the throughput of the system.

In the paper [15] proposed the Hybrid scheduling algorithm which is of Divide-and-conquer and Throttled algorithm. This algorithm schedules the incoming requests to request handler or VMs which depend on the availability of load of machine. According to the author, proposed work reduces the execution time.

**Application based load balancing algorithm:**

In this category application related algorithms are considered. In the paper [16] proposed a load balancing algorithm in mobile cloud computing which is based on Max-Min Ant system. Experiment results shows that the proposed algorithm increases the performance effectively in mobile cloud.

Table I gives the overview of the Load Balancing Algorithms.

### *C. Workflow Scheduling*

Workflow scheduling is the process to map tasks and order these tasks on appropriate resources for maintenance of execution flow of work. It also satisfies the performance objectives. Workflow scheduling algorithms are classified by Nature of algorithms. The nature of algorithms depend on the optimal solution which get.

**Single Objective:**

In this we get only single optimal solution that is the best optimal solution for defining the objective. We can find out the resultant solution which is better than the other solutions. In the paper [17] proposed Bandwidth Aware Task Scheduling (BATS) algorithm which is a single objective workflow scheduling algorithm for parallel task scheduling type applications. Author used the non-linear programming model for optimization of tasks allocation which is based on their requirements.

In the paper [18] proposed Enhanced Intelligent Water Drops (EIWD) algorithm which is single objective workflow scheduling algorithm for optimization of makespan. It is applicable for job-shop scheduling. It minimizes the makespan of the workflow.

**Multi Objective:**

In this we cannot get a direct proper solution for defining the objective. There are many optimal solutions available, but for evaluation of performance only one solution is to be selected. In the paper [19] proposed the admission control and scheduling algorithm. It is a SLS based virtual machine management algorithm which is for heterogeneous workloads. Authors used ANN for future workload management. It gives better utilization of resources and minimizes SLA violations.

In the paper [20] proposed Maximum Effective Reduction (MER) algorithm which is resource efficient workflow scheduling. This is for the reduction of resources but makespan gets increased slightly. MER can give the optimal solution for makespan increment and resource decrement. By this algorithm energy also gets consumed.

In the paper [21] proposed Bi-level advanced reservation strategy in which global scheduling and local scheduling is performed respectively. In this paper author gets used the approaches for global scheduling are Critical Path Extraction algorithm and DAG partitioning algorithm. It is applicable for the evaluating an environment. Local scheduling uses Multi-criteria advanced reservation algorithm which is applicable for mapping of task on resources.

**Bi-criteria:**

In the paper [22] proposed a provisioning and scheduling algorithm which based on PSO technique. Author used this method for minimizing the total execution cost. Author has considered other parameters like performance variation and VM boot time. It also handles the QoS requirements and basic principles of Iaas cloud like pay-as-you-model, dynamicity, elasticity of resources etc.

In the paper [23] proposed the Enhanced Max-Min algorithm. It works of assignment of a task with average execution time to minimum completion time of resource. It also handles the makespan and load balancing.

Table II gives the overview of the Workflow Scheduling Algorithms.

## III. TABLES

**TABLE I. OVERVIEW OF LOAD BALANCING ALGORITHMS**

| Ref. No. | Category of Algorithm | Objective | Advantages | Disadvantages |
|---|---|---|---|---|
| [10] | Natural Phenomena based | Reduction of Energy consumption<br>Increment the Resource Utilization | Simple for detecting over / under utilized hosts<br>Uses Minimum Migration Time MMT | Not clear with live / dead VM migration<br>Increase response time due to VM migration |
| [11] | Natural Phenomena based | Minimization of Response time | Improves system performance<br>Improves resource utilization<br>Reducing time span of jobs | Not consider job priorities<br>Low throughput |

| | | | | |
|---|---|---|---|---|
| [12] | Agent based | Maximization of Resource Utilization<br>Reduction of Response time | Increment the resource utilization<br>Reduction of Migration cost<br>Reduction of waiting time of tasks in queue | Do not have self-destroying timer<br>It waits till message come from parent |
| [13] | Agent based | Reduction of Service time | Reduces the response time<br>Improves the Response time | Time consuming<br>Heavy computation involves |
| [14] | General | Job waiting time reduction | Throughput improvement<br>Increment of resource utilization<br>Reduction of waiting time of job | Energy saving is not considered |
| [15] | General | Reduction of response time and processing time | Improve response time and processing time<br>Accurate job assignment | Problem of energy consumption and carbon emission<br>Complicated to find out the actual processing power |
| [16] | Application based | QoS improvement | Improve the performance<br>Improves scheduling profit<br>Reduction of energy consumption | Dynamic resource requirement for application is not considered |

**TABLE II. OVERVIEW OF WORKFLOW SCHEDULING ALGORITHMS**

| Ref. No | Name of Algorithm | Nature of Algorithm | Objective Constraints | Advantages |
|---|---|---|---|---|
| [17] | Bandwidth aware Task Scheduling (BATS) | Single Objective | Bandwidth | Ideal for the scheduling tasks which consider bandwidth as a constraint |
| [18] | Enhance Intelligent Water Drops (EIWD) | Single Objective | Makespan | Minimizes the makespan |
| [19] | Admission Control and Scheduling Algorithm | Multi Objective | Cost, QoS parameters | Gives better utilization of resources<br>Minimize the SLA violations |
| [20] | Maximum Effective Reduction (MER) | Multi Objective | Energy, Makespan | Minimum energy consumption<br>Reduction in resource demand |
| [21] | Bi-level Advanced Reservation Strategy | Multi Objective | Time, Cost | Meets QoS requirements<br>High reliability |
| [22] | Provisioning and Scheduling Algorithm | Bi-criteria | Cost, CPU Time, Makespan | Minimize the total cost |
| [23] | Enhanced Max-Min Algorithm | Bi-criteria | Memory, Makespan | Handles the makespan and load balancing |

## IV. CONCLUSION AND FUTURE WORK

In this paper we have surveyed the literature in the area of load balancing, resource allocation and workflow scheduling. It gives an overall idea about the algorithms in resource management. We have seen that the different resource parameters are considered for different algorithms. We have also found that the advantages and objectives of the algorithms. The survey helps for analyzing the gap between what is already done and what is required to be done. As future works, we suggest that 1) To study and analyze the more techniques. 2) Also evaluate and compare the techniques based on new parameters.

## REFERENCES

[1] Sivadon Chaisiri, Bu-Sung Lee, Dusit Niyato, "Optimization of Resource Provisioning Cost in Cloud Computing", IEEE Transactions on Services Computing, Volume: 5, April-June 2012.

[2] K Hemant Kumar Reddy, Geetika Mudali and Diptendu Sinha Roy, "A novel coordinated resource provisioning approach for cooperative cloud market", Journal of Cloud Computing Advances, Systems and Applications, 31st March,2017.

[3] Seyedehmehrnaz Mireslami, Logan Rakai, Behrouz Homayoun Far, Mea Wang, "Simultaneous Cost and QoS Optimization for Cloud Resource Allocation", IEEE Transactions on Network and Service Management, Volume: 14, Issue:3, Sept. 2017.

[4] B. Neethu, K. R. Remesh Babu, "Dynamic Resource Allocation in Market Oriented Cloud Using Auction Method", International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE), Sept. 2016.
[5] Dilip Kumar S. M., Naidila Sadashiv, R. S. Goudar, "Priority based resource allocation and demand based pricing model in peer-to-peer clouds", International Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 2014.
[6] Shubhakankshi Goutam, Arun Kumar Yadav, "Preemptable priority based dynamic resource allocation in cloud computing with fault tolerance", International Conference on Communication Networks (ICCN), Nov. 2015.
[7] Aarti Singh, Dimple Juneja, Manisha Malhotra, "A novel agent based autonomous and service composition framework for cost optimization of resource provisioning in cloud computing", Journal of King Saud University - Computer and Information Sciences, Volume 29, Issue 1, January 2017.
[8] Xiao-long Zheng, Ling Wang, "A Pareto based fruit fly optimization algorithm for task scheduling and resource allocation in cloud computing environment", IEEE Congress on Evolutionary Computation (CEC), July 2016.
[9] Fan-Hsun Tseng, Xiaofei Wang, Li-Der Chou, Han-Chieh Chao, Victor C. M. Leung, "Dynamic Resource Prediction and Allocation for Cloud Data Center Using the Multi objective Genetic Algorithm", IEEE Systems Journal, Volume: 12, Issue: 2, June 2018.
[10] Moona Yakhchi, Seyed Mohssen Ghafari, Shahpar Yakhchi, Mahdi Fazeli, Ahmad Patooghi, "Proposing a Load Balancing Method Based on Cuckoo Optimization Algorithm for Energy Management in Cloud Computing Infrastructures", 6th International Conference on Modeling, Simulation, and Applied Optimization (ICMSAO), May 2015.
[11] Kousik Dasgupta, Brototi Mandal, Paramartha Dutta, Jyotsna Kumar Mandal, Santanu Dam, "A Genetic Algorithm (GA) based Load Balancing Strategy for Cloud Computing", International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA), Volume 10, 2013.
[12] S. Keshvadi, B. Faghih, "A multi-agent based load balancing system in IaaS cloud environment", International Robotics & Automation Journal, Oct. 2016.
[13] A. Singha, D. Juneja, M. Malhotra, "Autonomous Agent Based Load-balancing algorithm in Cloud Computing", International Conference on Advanced Computing Technologies and Applications (ICACTA), Volume 45, 2015.
[14] D. Komarasamy, V. Muthuswamy, "A novel approach for dynamic load balancing with eff ective Bin packing and VM reconfiguration in cloud", Indian Journal of Science and Technology, Vol 9, March 2016.
[15] Nguyen Khac Chien, Nguyen Hong Son, Ho Dac Lo, "Load Balancing Algorithm Based on Estimating Finish Time of Services in Cloud Computing", International Conference on Advanced Commutation Technology (ICACT), 2016.
[16] X. Wei, J. Fan, T. Wang, Q. Wang, "Efficient application scheduling in mobile cloud computing based on MAX–MIN ant system", Soft Computing, Volume 20, Issue 7, July 2016.
[17] Weiwei Lin, Chen Liang, James Z. Wang, Rajkumar Buyya, "Bandwidth - aware divisible task scheduling for cloud computing", Journal of Software Practice and Experience, Vol. 44, Feb. 2014.
[18] S. H. Niu, S. K. Ong, A. Y. C. Nee, "An improved Intelligent Water Drops algorithm for achieving optimal job-shop scheduling solutions", ournal International Journal of Production Research, Volume 50, Issue 15, 2012.
[19] Saurabh Kumar Garg, Adel Nadjaran Toosi, Srinivasa K.Gopalaiyengar, Rajkumar Buyya, "SLA-based virtual machine management for heterogeneous workloads in a cloud datacenter", Journal of Network and Computer Applications, Volume 45, October 2014.
[20] Young Choon Lee, Hyuck Han, Albert Y. Zomaya, Mazin Yousif, "Resource-efficient workflow scheduling in clouds", Knowledge-Based Systems, Volume 80, May 2015.
[21] Sahar Adabi, Ali Movaghar, Amir Masoud Rahmani, "Bi-level fuzzy based advanced reservation of Cloud workflow applications on distributed Grid resources", The Journal of Supercomputing, Volume 67, Issue 1, January 2014.
[22] Maria Alejandra Rodriguez, Rajkumar Buyya, "Deadline Based Resource Provisioning and Scheduling Algorithm for Scientific Workflows on Clouds", IEEE Transactions on Cloud Computing, Volume: 2, Issue: 2, April-June 2014.
[23] Upendra Bhoi, Purvi N. Ramanuj, "Enhanced Max-min Task Scheduling Algorithm in Cloud Computing", International Journal of Application or Innovation in Engineering & Management, Volume 2, Issue 4, April 2013.
[24] Lijin P, Dr. S. Siva Sathya, K. S Guruprakash, "A Survey on Resource Allocation Techniques in Cloud Computing", International Journal of Engineering & Technology, 2018.
[25] Sajjan R.S, Biradar Rekha Yashwantra, "Load Balancing and its Algorithms in Cloud Computing: A Survey", International Journal of Computer Sciences and Engineering, Volume-5, Issue-1, Jan. 2017.
[26] Fuhui Wu, Qingbo Wu, Yusong Tan, "Workflow scheduling in cloud: a survey", Journal of Supercomputing, May 2015.
[27] https://searchcloudcomputing.techtarget.com/definition/cloud-computing