Advances and Applications in Mathematical Sciences
Volume 18, Issue 8, June 2019, Pages 805-815

CATEGORIZATION AND TAXONOMY OF RESOURCE MANAGEMENT IN CLOUD COMPUTING

UDIT PRATAP SINGH and AMIT JAIN

University Institute of Computing
Chandigarh University
Mohali, India
Email: udit.singhmkp@gmail.com
amitjain.uic@cumail.in

Abstract

Cloud Computing is a service which is based on resource sharing model. Shared resource can be infrastructure, software and the information over the network. Cloud provides services as per the demand of client (individual or the organization). All the services are being provided to the user through internet. It uses the pay as you use service model. So as large number of users is using the resources at the same time, the main goal of service provider is to use the resources efficiently and to gain the maximum throughput from the system to maximize the profit so steps are need to be taken to optimize the system. The system can only be optimized if the available resources are managed properly. The other reason why management of resources is necessary is because the resources are limited and cloud providers have to provide all the requested resources. The study gives an elaborate idea about classification of cloud resources its locution is given and lastly a comparative study about different resource management algorithm has been provided.

I. Introduction

Cloud Computing is the current upcoming technology these days, it is really growing leap and bounds every day. It uses the distributed paradigms which have pool of computing resources. With it, one can make use of the power of computation instantly whenever required. Cloud model is based on sharing of resources, user can rent the services of cloud according to their need and can pay only for the services they have used. It is based on different layers (IaaS, PaaS, SaaS), companies use them differently according to the service they need.

2010 Mathematics Subject Classification: 68-XX.
Keywords: Cloud Computing, Scheduling, Resource Management, Throughput.
Received February 20, 2019; Accepted March 25, 2019

In cloud computing, different cloud buyers request assortment of services according to their progressively evolving necessities. So the major issue to the service provider is allotment of various finite resources to the service users. Some of the key features which make the resource management more complex in case of cloud computing are its support in heterogeneity of hardware and their capabilities, pay per use model and on-demand service model.

The content of paper is organized as under: The first section describes about the basics of cloud. The second section provide comparison of various research work their technique, algorithm their type and the issues. The third section provides classification of cloud resources and the fourth section describes taxonomy of resource management. At the last the paper shows the conclusion and future work.

II. Advantages of Cloud Computing

There are a few advantages of Cloud Computing which are portrayed beneath:

Open Access: Cloud specialist organization can be gotten to with the assistance of reasonable web association.

Enhanced economies of scale: On the client side, by diminish in the venture and running expenses and on the supplier side, by giving higher profitability in masterminding framework administrations with high survivability and adaptability.

Limit with respect to on-request foundation and computational control: Users can interest for the computational power, stockpiling also, other foundation as indicated by their need as per pay per utilize show.

Enhanced asset usage: Resources are used legitimately on the grounds that at whatever point clients needn't bother with an asset then they return it back to the cloud supplier. So the versatility and adaptability can be expanded.

Decreased data innovation (IT) framework needs: Distributed computing gives framework as-a-benefit on interest for client. So there is no lasting need to buy the foundation for the IT. The client can buy it from cloud supplier at

whatever point required. Asset pooling: The purchaser for the most part has no data about the region of the specialist organization. Along these lines the supplier serves various shoppers by appointing assets powerfully and practically. Control frameworks with conceptual approaches: There is no compelling reason to give superfluous insight about the center business part to the client.

Association's center around their center abilities: Non IT client can contact to IT specialist organization for their business movement needs.

III. Literature Survey

In a research paper, author focused on the pricing model in the dynamic environment. Authors talk that the policy of dynamic pricing should manage the number of allocated resources and number of requests (Mihailescu, M. and Teo, Y.M., [1]).

In another paper, two algorithms were proposed by author:

(1) Resource initialization

(2) Resource allocation and tested it using DVFS technique.

Author has tried to enhance the utilization of resources and reduce the power consumption. Algorithm was mainly focused for hybrid cloud (Jha, R. S. and Gupta, P. [2]).

Another author has proposed an algorithm named optimal cloud resource provisioning (OCRP) to reduce the cost of provisioning of resources. There were three phases in algorithm: reservation, expanding and on-demand phase and further each phase can have multiple stages. In this algorithm optimal pricing can be developed in future (Chaisiri, S. et al., [4]).

In another work, a researcher proposed a method known as CA_PROVISION in which resources are considered as liquid that can configured in depending upon user request. It is based on auction-based mechanism; the resource will be matched to users having highest bid capacity. This method was proved to be better than having a fix price of resources; algorithm was able to generate higher revenue to the service providers at peak times (Zaman, S. and Grosu, D., [5]).

Further, one of the authors has integrated resource discovery and resource monitoring tool and has proposed a model for resource monitoring architecture (Aguado, G.J., [7]).

In another paper author has adapted heuristic methodology to solve the problem, method has been proposed to map the request made by the user's application for the resources in shared environment connecting various resources over network (Papagianni, C., [8]).

In a research paper, author says that virtual machines can be considered as an operating system which is statically loaded and balanced. Another author also explained that by using Dynamic Resource Allocation (DRA) can increase the host machine availability. The algorithm aims to maintain the similar load on CPU by determining the load on each machine. OpenNebula core has been developed to support this, it make decision on placement of VM's and scheduling of resources (Yang, C.T., [9]).

One of the authors focused is on high priority of a job which means that low priority jobs can't delay the processing of high priority jobs. A dynamic VM allocation model has been presented by the author which is capable of dynamically configuring the jobs, which state that even if low priority job is being executed by the machine and high priority job arrives then the low priority job has to be preempted once the execution of high priority job gets complete then the low priority which is preempted will be executed (Saraswathi, A. T. et al., [10]).

Further, the author proposed a tier-based resource allocation algorithm *T*-BICA(Tier-centric Business Impact and Cost Analysis), in this resource allocation is done tier-wise after monitoring of each tier whereas in traditional approach allocation is done without monitoring each tier (Khasnabish, J. N., et al. [11]).

Also, an author discussed auction based method was proposed which was market-driven in this first the users are identified and identification is based on payment capacities. Method consisted of three phases a) Pre-auction b) Market-driven open auction c) Preference-driven payment (Kumar, N. and Saxena, S., [12]).

Further, an author proposed a modified approach of Round Robin algorithm. Approach explains that the initial start time of each request will be same as of the first request, which later is modified when first request gets

to end. On adding a new request, algorithm will check number of requests in the ready queue and will calculate the average sum of time of each request and the time of new request will also be added. Algorithm will result in reduction of turnaround time and waiting time of process is also reduced (Pradhan, P. et al., [15]).

Further, an author proposed an algorithm in which multiple SLA parameters were used. Preemptable task execution method was used in which scheduler will divide request into multiple tasks, in which two scheduling algorithms were used cloud list scheduling (CLS) and cloud min-min scheduling (CMMS), after this list based on task priority was made and resources will be allocated based on the list formed (Pawar, C.S. et.al., [16]).

The following table shows different research work their technique, algorithm their type and the issues.

Sr. No.	**Author (Year)**	**Techniques / Algorithms**	**Tools and/or workload used**	**Type**	**Future Work and/or Gaps in Existing**
1	Mihailescu, M. and Teo, Y.M., 2010	Dynamic Resource Pricing	Planet Lab	Resource Pricing	Scalability is a issue.
2	R. S. and Gupta. P., 2015	Power and Load Aware VM Allocation Policy .	CloudSim	Resource Allocation / Reallocation	Experiments were done on simulators, not on real environments
3	Chaisiri, S. et.al., 2012	Optimal Cloud resource provisioning	Stimulation	Cost Reduction for resource provisioning	Scenario reduction techniques, optimal pricing scheme for providers.
4	Zaman, S. and Grosu, D., 2013	Combinatorial auction-based for dynamic VM	Stimulation with real workloads	Resource Provisioning	Combination of two combinatorial auction based mechanism.
5	Aguado, G.J., 2016	IaaSMon	Nagios / OpenStack	Resource Monitoring	Integration of both tools
6	Papagianni, C., 2013	Optimal Allocation of Virtual Resources using Mixed Integer Programming (MIP)	Simulator for Controlling Virtual Infrastructures (CVI-Sim) Resource Allocation / Reallocation Implementation of proposed Framework.	Resource Allocation / Reallocation	Implementation of proposed framework.

7	Yang, C.T., 2011	Dynamic Resource Allocation	HPCC	Resource Allocation	Dynamically change cluster capacity.
8	Saraswathi AT. et.al., 2015	Dynamic Resource Allocation based on Priority	CloudSim	Reduce cost for resource provisioning	Extend in real time cloud environment with economy based preemption
9	Khasnabish, J.N. et.al., 2015	Tier centric resource allocation	AWS EC2 public cloud	Monitor tier level resource utilization	Modify algorithm to reduce cost factor.
10	Kumar, N. and Saxena, S., 2015	Market driven auction with preference based	CloudSim	Auction based Resource Allocation	Energy efficient scheduling strategy.
11	Pradhan, P. et.al., 2016	Modified Round Robin	MATLAB	Optimal Scheduling method with time quantum	Modify the time quantum further.
12	Pawar, C.S. et.al., 2013	Multiple SLA parameters	CloudSim	Dynamic Resource Allocation	Improve the algorithm further for resource contention

IV. Classification of Cloud Resources

Cloud computing is can be said as utility service as it provides services on demand of user as per their requirement. In cloud, services can be called as the resources. Classification of cloud resources is can be done in different ways such as physical and logical resources or can be stated as hardware and software resources. Now we will classify the resources of cloud computing:

Compute Resources: Many physical resources are needed to be combined together to form the compute resources. The computational capacity of clouds environment comprises of multiple elements like processors, memory element, network element and local I/O.

Networking Resources: In this cloud environment, several resources are organized in to a cluster and each resource is inter-connected through high speed network which is built using InfiniBand technologies.

Storage Resources: These resources manage customer's data and make it available over network, cloud provide the elasticity of storage resources as user can upscale or downscale its amount of space on lease according to the need which is difficult in traditional database.

V. Taxonomy on Resource Management

In Cloud Computing, resource management means to use all the resources efficiently and effectively and provide the quality service to the cloud user. The cloud taxonomy is divided into two phases:

- Initial Resource Assignment
- Periodic Resource Optimization

Phase 1

Initial Resource Assignment

In this required resources are requested by the application, the application perform the request on behalf of cloud user. The whole process work in sequence of steps:

(1) **Request Identification:** This is the initial stage; request is identified by the service provider.

(2) **Resource Gathering:** In this available resource are listed out, this stage is also known as resource formation.

(3) **Resource Brokering:** This progression is arrangement of assets with cloud purchasers to ensure that they are accessible according to necessity.

(4) **Resource Discovery:** This progression will consistently group different assets according to the necessities of cloud purchasers.

(5) **Resource Selection:** This progression is to pick finest resource from the resources pool for requirements given by cloud purchasers.

(6) **Resource Mapping:** This progression will outline assets with physical assets (like hub and ports) given by cloud suppliers.

(7) **Resource Allocation:** Its principle objective is to fulfill cloud consumers need with most optimized resource which can satisfy their needs and income generation for cloud suppliers.

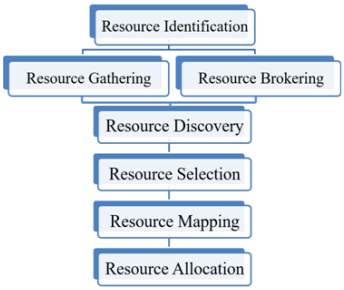

Figure 1. Phases of Initial Resource Assignment.

Initial Resource Assignment Algorithms

(1) **First Come First Served:** It is non-preemptive algorithm for resource scheduling. In this a list is maintained for all incoming tasks and resources are assigned to the task in the same sequence they have arrived.

(2) **Shortest Job First:** It is also a non-preemptive algorithm for resource scheduling. In this the available resources are assigned to incoming request for the processing according to CPU burst value, the task with the smallest CPU burst value will get assigned with the required resources. The main issue here is to calculate the value for the time period for which an algorithm will run.

(3) **Round Robin:** This algorithm does not consider the priority and the length of the process the thing which is taken in consideration for the assignment of the cloud resources is time intervals.

(4) **Priority Based:** In this priority is assigned to each process. Process get the assignment of the resources according to the priority, in this task with higher priority are executed first and the task with least priority is executed at the last (Khasnabish, J.N. et.al., [11]).

Phase 2

Periodic Resource Optimization

It is second phase; in this the resource management is being done. It is categorized in different phases:

(1) **Global Scheduling of Virtualized Resources:** In Cloud Computing VM scheduling algorithms are utilized to plan the VM solicitations to the Physical Machines (PM) of the specific Data Center (DC) according to the

necessity satisfied with the asked resources. Scheduling is a balancing situation in which procedures are scheduled according to the given necessities and utilized algorithms.

(2) **Resource Allocation:** It is one of the most challenging jobs of cloud computing. In cloud computing resource allocation is taken into consideration at many points like data center management and grid computing. Resource allocation of cloud resources means that the available resources are efficiently being divided between users and cloud.

(3) **Resource Utilization Estimation:** The requirement for resource use never stops as long resources are limited contrasted with the expanding request on computers. Resources are pooled to serve multiple purchasers utilizing a multi-inhabitant model, with various physical and virtual resources powerfully allotted and reassigned by user's requests.

(4) **Local Scheduling of Virtualized Resources:** The main aim of scheduling in case of virtual machines is to use the resources efficiently that is to maximize the utilization and minimize the execution time, these objectives can only be achieved when the whole load is distributed equally between all the processors.

(5) **Application Scaling and Provisioning:** Provisioning of cloud resources means assignment of the cloud provider's resources available, to the cloud users according to their requirement. Dynamic provisioning is done and customers are charges according to pay-per-use. The customer buys resources using web portal.

(6) **Workload Management:** Work load is can be defined as the amount of task that can be given to the system to process. So, workload management is the process of distributing the load equally between the processors, so that all the resources can be utilized efficiently, along with the utilization of several algorithm being used [Jennings, B.and Stadler, R., 2014].

VI. Conclusion

Optimization of resources is the most important part of cloud computing. Various algorithms from different researchers and their various phases along with the research issues have been discussed. Resources of cloud have been categorized in compute resources, network resources and storage resources.

Resource management is one of the complex thing, as it is long process which involves identification of request that what kind of resources are required to fulfil the requirement of customer, as per the requirement. The resources are assigned to user application, which involve lot of process like broking and discovering of resources, global scheduling of virtual resources at every tier. Further, it is required to get the current utilization of resources along with the procedure to find how much resources are free for allocation. The whole process is mainly focused to get the maximum utilization of all the available resources and to generate maximum profit for the service provider. Lot of work has been done and many algorithms have been designed but still lot of modification can be done in algorithms or new algorithms can be written.

References

[1] M. Mihailescu and Y. T. Teo, Dynamic resource pricing on federated clouds. In Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, pages 513-517 (2010), IEEE Computer Society.

[2] R. S. Jha and P. Gupta, Power & load aware resource allocation policy for hybrid cloud, Procedia Computer Science 78 (2016), 350-357.

[3] S. Mustafa, B. Nazir, A. Hayat and S. A. Madani, Resource management in cloud computing: Taxonomy, prospects, and challenges, Computers & Electrical Engineering, 47 (2015), 186-203.

[4] S. Chaisiri, B.S. Lee and D. Niyato, Optimization of resource provisioning cost in cloud computing, IEEE Transactions on Services Computing, Vol. 5, No. 2, (2012).

[5] S. Zaman and D. Grosu, A combinatorial auctionbased mechanism for dynamic VM provisioning and allocation in clouds, IEEE Transactions on Cloud Computing, Vol. 1, No. 2, (2013).

[6] V. Goyal and Amit Jain, Role of SOA & Cloud Computing in IT Industry, International Journal of Engineering Research & Technology (IJERT), Vol. 1, Issue 6, (2012).

[7] J. G. Aguado, J. M. A. Calero and W. D. Villanueva, Iaasmon: Monitoring architecture for public cloud computing data centers, Journal of Grid Computing, (2016), pages 1-15.

[8] C. Papagianni, A. Leivadeas, S. Papavassiliou, V. Maglaris, C. Cervello Pastor and A. Monje, On the optimal allocation of virtual resources in cloud computing networks. IEEE Transactions on Computers 62(6), (2013), 1060-1071.

[9] C. T. Yang, H. Y. Cheng and K. L. Huang, A dynamic resource allocation model for virtual machine management on cloud, Springer-Verlag Berlin Heidelberg, CCIS 261, (2011), pp. 581-590.

[10] A. T. Saraswathi, Y. R. A. Kalaashri and S. Padmavathi, Dynamic resource allocation scheme in cloud computing, Procedia Computer Science 47 (2015), 30-36.

[11] J. N. Khasnabish, M. F. Mithani and S. Rao, Tier-Centric resource allocation in multi-tier cloud systems, IEEE Transactions on Cloud Computing, (2015), in press, Citation information: DOI 10.1109/TCC.2015.2424888.

[12] N. Kumar and S. Saxena, A Preference-based resource allocation in cloud computing systems, 3rd International Conference on Recent Trends in Computing 2015 (ICRTC2015), Published by Elsevier B.V., Procedia Computer Science 57 (2015), 104-111.

[13] S. Parikh and P. Narendra, AB-INITIO (initial) resource assignment in cloud computing: An experimental survey, International Journal of Creative Research Thoughts 5 (4) (2017), 9.

[14] B. Jennings and R. Stadler, Resource management in clouds: survey and research challenges, Journal of Network and Systems Management (2014), doi: 23. 10.1007/s10922-014-9307-7.

[15] P. Pradhan, P. K. Behera and B. N. B. Ray, Modified Round Robin algorithm for resource allocation in cloud computing, International Conference on Computational Modeling and Security, Published by Elsevier B. V, Procedia Computer Science 85 (2016), 878-890.

[16] C. S. Pawar and R. B. Wagh, Priority Based dynamic resource allocation in cloud computing, (ISCOS), International Symposium on Cloud and Services Computing, IEEE (2013).

[17] M. Dabbagh, Toward energy-efficient cloud computing: Prediction, consolidation, and over commitment, IEEE Network 29.2 (2015), 56-61.