

Resource Allocation and Scheduling Methods in Cloud- A Survey

AICY EALIYAS

*Department of Computer Sciences Technology
Karunya University
Coimbatore, India
aicyealiyas03@gmail.com*

DR. S. P. JENO LOVESUM

*Department of Computer Sciences Technology
Karunya University
Coimbatore, India
jenolovesum@karunya.edu*

Abstract— Cloud computing is an information technology paradigm, which provide services and resources through on-demand services over the internet. As there are many applications deployed in the cloud by users, proper scheduling is needed as various resources are consumed. If the application is not scheduled properly, it will cause load imbalance in the cloud data centers. In this paper, there will be proper utilization of the resources and a survey is done on the various scheduling and allocation methods for applications in cloud.

Keywords—cloud computing; resource allocation; efficiency;

I. INTRODUCTION

Cloud computing is a computing technology which consist of large numbers of servers to provide storage, services and resources. Whenever a user requires the services available in cloud, he/she can use the services by paying for it on timely basis. Now a day's more users use cloud because of the advantages of the cloud. Different application consumes various amount and type of resources. Allocation of resources is the major challenging issue in cloud environment. In resource allocation the available resources such as CPU, memory, network, I/O, storage are assigned according to the requirements and usage in an effective way. We need to assign resources in an effective and efficient way so that we can reduce the load in the cloud data centers and increase the resource utilization rate.

There are two types of cloud models present in cloud environment such as cloud deployment model and cloud service model.

A. Cloud deployment model

Cloud processing model can explained in a few ways. Based on qualities of these models they are grouped in to four different deployment models.

Public Cloud

Public cloud is a form of cloud computing, here the cloud service provider(CSP) provide resources like storage, virtual machines (VM), memory, CPU to public over the internet.

Public cloud offering pay-per usage model and sometimes it free cost. Eg: Amazon Elastic Computer Cloud (EC2), Sun Cloud, IBM's Blue Cloud.

Private Cloud

Private cloud is a type of cloud computing mostly private cloud is maintained for privacy and safety. It is used by selected users, company or organization instead of the general public cloud.

Eg: Amazon Elastic Computer Cloud (EC2)

Hybrid Cloud

Hybrid cloud integrates private cloud and public cloud to create hybrid cloud. In hybrid cloud sensitive data's keeps in private cloud and other data is stored in and accessible from public cloud.

Community Cloud

Several companies and administrations are using community cloud service. It can be shared to particular people or a group of people.

B. Cloud service model

Mainly three services offered by cloud computing such as Platform as Services (PaaS), Infrastructure as Services (IaaS), Software as Services(SaaS).

C.

This paper is organized as part II includes types of resources; part III presents survey on resource allocation methods in cloud, part IV includes comparative study of various resource allocation methods and section V presents conclusion.

II. TYPES OF RESOURCES

There are mainly four types of resources present in the cloud environment.

A. Computing resources

The main computing resources includes in cloud such as memory, processor, bandwidth, network, input/output devices. The resources are allocated and purchased according to user needs.

B. Networking resources

Several problems are encountered in networking in cloud such as traffic, bandwidth etc. This can be solved by using different protocols for enhancing quality of service of the cloud.

C. Storage resources

Scalability is an important criterion in cloud storage. It provides storage protection for data backup and archival and also provides disaster recovery option.

D. Power resources

Power resource is the system which deals with the usage of power. The energy used by the system for allocating the resources in cloud

III. LITERATURE SURVEY

Laing-Teh et al. [1] proposed a resource management technique in which with energy saving mechanism. The method of voltage scaling is used for adjusting the resources by examine the utilization of CPU in cloud. To achieve the energy-saving, bulky loaded works can be relocating to idle machine and voltage of lighter loading computer can be minimized to increase the utilization of resource and minimizes the consumption of energy. In DVFS technology, adjusted the frequency and system voltage with respect to the requirements of original CPU design into a varying working voltage without restarting the power supply. Energy consumption can be easily and effectively maximize when the CPU is working in lower voltage.

Ismael solis Moreno et al. in [2] introduced approaches to allocation of workload it will efficiently improves the energy in cloud datacentres. To minimizing the workload and efficiently increasing the energy in cloud datacenters through reallocate various kinds of workloads based up on the interference level. Based on the resource patterns grouping workloads and selecting the servers. It will minimize the performance interference and maximizes the energy efficiency.

Amr Alaska's et al. in [3] proposed an algorithm prediction-based resource allocation algorithm (PBRA). Resource reservation cost will be reduced in this method and much resource is reserved in cloud. Algorithm is proposed to minimize the issues of taking decision for resource allocation. The cost is reduced by determining reservation time and the quantity of reserved resources in the cloud.

Dongsung Kim Press et al. in [4] proposed a priority scheduling for monitoring the virtual machines. Tasks are scheduled in accordance with utilization of I/O and priorities of task. Virtual machine monitor is used for the implementation and evaluation of the suggested algorithm. Hence I/O events response time is enhanced. The proposed technique is attained by send guest level information to virtual machine monitors. However this method has some disadvantages. Firstly guest kernel modification is required for th

is approach. It is very difficult to apply this approach in operating system like windows. Secondly, the information is not secure in the guest level Guest level.

Wei ming et al. in [5] implemented a Ant Colony Optimization algorithm based on polymorphism, used to improve the QOS of cloud and to interchange node information list dynamically. Transfer task to Cloud Control Queen by Master at the time of the user submits the task. After that, accordingly to previous functions, the algorithm is organized into several parts. Average completion time and local optima is minimized in this method.

Amit Agarwal et al. in [6] implemented a priority algorithm for execute the task efficiently later it compare with Round Robin and FCFS Scheduling. This algorithm contributes greater performance than other scheduling algorithm undertakes testing in cloud Sim toolkit. It is seen that the proposed algorithm is better than FCFS and round robin algorithm.

Brototi Mondala et al. in [7] introduced a load balancing technique. For allocating the task to the servers or virtual machines, Stochastic Hill climbing optimization technique is used. Cloud Analyst examines the performance of the technique. A stochastic climbing used to balance load in cloud environment. This methodology is compared with two methodologies First Come First Serve and round robin.

Xiao-long et al. in [8] implemented an algorithm based on Pareto-fruity fly optimization to solve the issues in resource allocation as well as task scheduling problem in cloud computing environment. PFOA was implemented to solve the issues in multi objective resource allocation and task scheduling.

Wahid Hussain et al. in [9] implemented agent based architecture for resource allocation by defining the roles at each level in dynamic environment. This architecture has different strategies related to resource allocation such as performance scalability, reliability and security related problems in distributed environment. Broker agent, service providers, consumer agent, expert agent, monitoring agents and network administrator are included in this proposed architecture. Moreover, monitoring agent includes RHA, AA, RAA, ENA, CAA and GLAA agents. In system architecture, the role and duties of each and every agent is briefly explained. Efficiency, reliability, content analysis, scalability are considered in the proposed system.

Zhiyuan Xu et al. in [10] presented efficient resource allocation framework and it is called Deep Reinforcement Learning (DRL). It minimizes the total power consumption and also ensures the demand of each wireless user is satisfied. It defines the action space, state space and reward function for DRL agent, at the same time by applying a Deep Neural Network to approximate the action-value function as well as

formulate the resource allocation problem as a convex optimization problem.

IV. COMPARATIVE STUDY OF VARIOUS RESOURCE ALLOCATION METHODS

Resource allocation is an important factor in cloud computing. We need to allocate resources in an efficient and effective manner. Table 1 shows the comparative study of the various resource allocation techniques. The comparative study is done for finding the best resource allocation method for efficiently and effectively allocating the resources in order to balance the load in the cloud data centers.

TABLE I. COMPARATIVE STUDY

Sl. no	Author & Year	Techniques Used	Advantages	Disadvantages
1	Liang-tech lee et al. 2013	Dynamic voltage frequency scaling	Increased resource utilization. Reduced energy consumption.	Memory migration is complex.
2	Jie xu et al. 2013	Cloud management system (CMS) architecture with an interference aware allocation module	Less performance interference. Increased energy efficiency.	Less performance.
3	Amar alsaad et al. 2015	Prediction based resource allocation algorithm	Less cost for resource reservation	Multiple cloud providers are involved.
4	Hwanju et al. 2008	Priority based scheduling	Response time is very less.	Guest-level information untrusted
5	Wei ming et al. 2014	Ant colony optimization	Increased efficiency of resource allocation	Less Stability
6	Brototi mondal	Generalized priority	Efficient execution	High execution

	et al. 2012	algorithm	of task	time.
7	Amit agarwal et al. 2014	Stochastic hill climbing	Effective resource utilization	Optimization problem
8	Ling wang et al. 2016	Pareto based fruit fly optimization algorithm	Efficient resource allocation and task scheduling	High transferring time
9	Wahid et al. 2016	Agent based architecture	Scalable reliable resource allocation	Bad customer satisfaction
10	Zhiyuan et al. 2017	Deep reinforcement learning (DRL) based framework	Less power consumption	Does not satisfy the user demand.

V. CONCLUSION

Cloud computing is a computing technology which consist of large numbers of servers to provide storage, services and resources. As there are many applications deployed in the cloud by users, proper scheduling is needed as various resources are consumed. If the application is not scheduled properly, it will cause load imbalance in the cloud data centers. Cloud computing provides pool of resources and services and one of the challenging problem is resource allocation. In this paper, there will be proper utilization of the resources and a survey is done on the various scheduling allocation methods for applications in cloud.

REFERENCES

- [1] L.T. Lee , K.Y. Liu , H.Y. Huang , A dynamic resource management with energy saving mechanism for supporting cloud computing, Int. J.Grid Distrib. Comput.6 (1) (2013) 67–76.
- [2] I.S. Moreno , R. Yang , J. Xu , Improved energy-efficiency in cloud datacenters with interference-aware VM placement, Proceedings of the IEEE Eleventh International Symposium on Autonomous Decentralized Systems,2013, pp. 1–8 .
- [3] A. Alasaad , K. Shafee , H.M. Behairy , Innovative schemes for resource allocation in the cloud for media streaming applications, IEEE Trans. Parallel Distrib. Syst. 26 (4) (2015)
- [4] D. Kim , H. Kim , M. Jeon , Guest-aware priority-based virtual machine Scheduling for highly consolidated server, Lect. Notes Comput. Sci. 5168 (2008) 285–294 .

- [5] W. Ming , Z. Chunyan , Q. Feng , Resources allocation method on cloud computing, in: 2014 International Conference on Service Sciences (ICSS), 2014, pp. 199–201.
- [6] B. Mondal , K. Dasgupta , P. Dutta , Load balancing in cloud computing Using stochastic hill climbing-a soft computing approach, Procedia Technol. 4 (11) (2012) 783C789.
- [7] A. Agarwal , S. Jain , Efficient optimal algorithm of task scheduling in cloud computing environment, Int.J.comput.trendstechnol. 9 (7) (2014) 344–349 .
- [8] Xiao-long Zheng, Ling Wang, A Pareto based Fruit Fly Optimization Algorithm for Task Scheduling and Resource Allocation in Cloud Computing Environment,Department of Automation, Tsinghua University Beijing 100084, China,978-1-5090-0623-6/16/2016 IEEE
- [9] Wahid Hussain, Rao Muzamal Liaqat, Nazar Abbas Saqib, A Secure Agent Based Architecture for Resource Allocation in Cloud Computing.
- [10] Zhiyuan Xu, Yanzhi Wang, Jian Tang, Jing Wang, and Mustafa Cenk Gursoy, A Deep Reinforcement Learning based Framework for Power-Efficient Resource Allocation in Cloud RANs , IEEE ICC 2017 Next Generation Networking and Internet Symposium,978-1-4673-8999-0/17/2017 IEEE