

A Review on Techniques of Resource Allocation in Cloud Computing

Sagar S. Brahmshatriya

Department of Computer Science and Engineering
Parul Institute of Technology
Vadodara, India

Jignesh Prajapati

Department of Computer Science and Engineering
Parul Institute of Technology
Vadodara, India

ABSTRACT

Cloud computing is the new era of technology, it uses virtualization and automation to provide user with self-service computing products that provide infinite scale at low cost. It serves user in many ways such as Infrastructure as a Service (IaaS), Software as a Service (SaaS), Platform as a Service (PaaS) etc. While serving as an IaaS it is inevitable that resources must be provided in an efficient way in order to meet the Service Level Agreement (SLA). This paper mainly addresses challenges, performance issues and techniques for resource allocation in cloud computing.

Keywords

Cloud Computing, Resource Allocation, Social Cloud, Virtualization

1. INTRODUCTION

From the ever growing usage of Information Technology (IT) it is easily predictable that one day computing will be the 5th utility after water, electricity, gas and telephony [1]. One research area that has attracted the attention of scientific community recently is the cloud computing.

As defined by NIST, “Cloud computing is a model that enables ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction”.[12]

We can consider cloud computing today as the electric grid was, 100 years ago. 100 years ago if you wanted to build widgets or power machines in a factory, you needed to build your own electric power station. Which would be incredibly expensive to both create and maintain the dynamo that is required for building up the power station. It doesn't take a huge leap of the imagination to drop hair loss between this dynamo and your data center. The computing resources that you consume in your data center come from limited number of suppliers and are very costly to build and maintain and there is always a chance that your system administrator could put things together the wrong way and break everything.

As electric grid change the face of a world, when companies could start plugging into electric power to drive things like blenders, televisions, mobile devices so on and so forth the world change overnight and it sparked the hundred year of innovation the lights that world has never seen. A lot of people hope and think that cloud computing is going to change the tale of the planet in exactly the same way where we can plugin to the computing resources like we plugin to a socket of our wall [2].

2. MOTIVATION

In cloud computing requirement of resources are dynamic and it is the job of cloud computing to meet the ever changing requirements of the cloud user but resources in cloud are always limited so it is very difficult for cloud providers to provide all the demanded services in time. For the cloud provider perspective it is inevitable that resources must be provided in an efficient way in order to meet the SLA.

The main goal of resource allocation resource allocation techniques at cloud provider perspective is to maximize the profit of providers and at cloud consumer perspective is to minimize the cost for the consumer.

3. LITERATURE REVIEW AND RELATED WORK

A.Meera and S.Swamynathan [3] proposed an approach for allocating resources based on the analysed data that is being analysed by a monitoring agent. The monitoring agent will collect the resources usage information that is currently being used by a virtual machine and will display it in a dashboard. Statistical report that is being displayed on a dashboard provides an information for cloud administrator for better optimization of resources. At this point this approach suffers from the centralized coordination because agents are not autonomous so future work can be focused on autonomous agents.

Weiwei Lin, James Z. Wang, Chen Liang, and Deyu Qi [4] proposed a threshold based dynamic approach for allocation of resources at an application level. This approach allocates resources dynamically among the cloud computing application based on their load change and uses the threshold method to optimize the decision of resource allocation. This work can be extended by concentrating on the selection of threshold value.

K C Gouda, Radhika T V, and Akshatha M [5] proposed a priority based resource allocation approach with minimum wastage and a maximum profit. Priority is being considered in terms of different parameters like time, cost, number of processor requests etc. priority can be used for better resource allocation in cloud environment.

Ewnetu Bayuh Lakew, Francisco Hernandez-Rodriguez, Lei Xu, and Erik Elmroth [6] proposed a fully distributed approach in order to manage resource allocation for services running across multiple clusters. In this approach services running across clusters can compete dynamically for resources by maintaining the global cloud level allocation limit. Global polls is being triggered by services in order to compete for resources. Global polls is being triggered only when the allocation balance of a cluster decreases below the threshold. Even though it is a fully distributed approach it still suffers from the number of global polls.

Kyle Chard, Simon Caton, Kris Bubendorfer, Omer Rana [7] proposed an approach of a social cloud for sharing the resources on the base of relationship, trust and risk, and policies with the coordination of social market. A Social cloud is resource and service sharing framework utilizing relationships established between members of a social network. Social market is being used in order to regulate the resource sharing within the group.

Hao-peng CHEN and Shao-chong LI [8] proposed an approach for managing performance of the cloud. In this approach authors had introduced queuing based model for managing the performance of cloud. In this model, web applications are modelled as queues and the virtual machines are modelled as service centers. Queuing theory is being applied in order to dynamically create and remove virtual machine.

Diptangshu Pandit, Matangini Chattopadhyay, and Nabendu Chaki [9] proposed an efficient resource allocation algorithm with the use of simulated annealing. In this approach authors had introduced the concept of bin, soft computing and simulated annealing. In this approach, problem of resource allocation is being solved with the help of bin packing problem. In this approach temperature is being considered as a control parameter but no formal procedure of selecting the temperature has been described in this approach.

Chu-Fu Wang, Wen-Yi, Hung, and Chen-Shun Yang [10] proposed an Energy Conserving Resource Allocation Scheme with Prediction (ECRASP). In this scheme arriving jobs trends can be predicted i.e. dense or sparse. In order to predict authors have used the concept of Exponential smoothing prediction.

Parvathy S. Pillai, and Shrisha Rao [11] proposed a mechanism for machines on the cloud in order to reduce wastage of IaaS resources. Mechanism is based on the principles of coalition formation and the uncertainty game theory.

4. SUMMARY

Table 1. Summary of resource allocation techniques

Year	Author	Approach	Future Work
2013	A.Meera and S.Swamynathan	Allocating resources with the help of monitoring agent	Mobile agents that take automatic decisions based on resource usage can be introduced as a future work
2011	Weiwei Lin, James Z. Wang, Chen Liang, and Deyu Qi	Threshold based dynamic resource allocation technique	Some future work can be done on the selection of the value of threshold

2013	K C Gouda, Radhika T V, and Akshatha M	Allocating resources by assigning priority to the requests	Work can be done in order to make this approach dynamic
2012	Ewnetu Bayuh Lakew, Francisco Hernandez-Rodriguez, Lei Xu, and Erik Elmroth	Dynamic allocation of resources in clustered environment with the use of Global Polls	Future work can be done in order to reduce global polls and enforcement of the concept of distributed quota
2011	Kyle Chard, Simon Caton, Kris Bubendorfer, Omer Rana	Concept of social cloud and resource allocation with the help of social cloud	As a future work network traffic that is being generated in social cloud can be reduced
2010	Hao-peng CHEN and Shao-chong LI	The concept of queuing based model for managing the performance of cloud	As a future work deploy it in a commercial environment in order to do more precise measurement and analysis
2014	Diptangshu Pandit, Matangini Chattopadhyay, and Nabendu Chaki	Algorithm for allocating resource using simulated annealing	As a future work formal procedure for selecting temperature can be described
2014	Chu-Fu Wang, Wen-Yi, Hung, and Chen-Shun Yang	Introduced an Energy Conserving Resource Allocation Scheme with Prediction	This scheme helps in reducing unnecessary power consumption and can make adequate decision on

			when to shut down and start up a physical machine
2014	Parvathy S. Pillai, and Shrisha Rao	Mechanism based on coalition formation and uncertainty game theory	With the use of this approach high resource utilization and high request satisfaction can be achieved

5. CONCLUSION AND DISCUSSION

Cloud computing is the new era of technology that one day will be the 5th utility after water, electricity, gas and telephony. In order to be successful cloud computing must provide a quality services according to the service level agreement. Due to the limited amount of resources availability it is very important for cloud providers to efficient utilization and management of the resources. So in this paper we have discussed problem and challenges that is addressed with currently available resource allocation techniques and mainly the ultimate goal of resource allocation is to maximize the profit for cloud providers and to minimize the cost for cloud consumers.

6. REFERENCES

- [1] Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg, Ivona Brandic, “Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility”, *Future Generation Computer Systems* Volume 25, Issue 6, June 2009, Pages 599–616
- [2] Nicholas Caar, “The Big Switch”
- [3] A.Meera, S.Swamynathan, “Agent based Resource Monitoring system in IaaS Cloud Environment”, *International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA)*, 2013
- [4] Weiwei Lin, James Z. Wang, Chen Liang, and Deyu Qi, “A Threshold-based Dynamic Resource Allocation Scheme for Cloud Computing”, *Procedia Engineering* volume 23, 2011, Pages 695–703
- [5] K C Gouda, Radhika T V, Akshatha M, “Priority based resource allocation model for cloud computing”, *International Journal of Science, Engineering and Technology Research (IJSETR)* Volume 2, Issue 1, January 2013
- [6] Ewnetu Bayuh Lakew, Francisco Hernandez-Rodriguez, Lei Xu and Erik Elmroth, “Management of Distributed Resource Allocations in Multi-cluster Environments”, *Performance Computing and Communications Conference (IPCCC)*, IEEE 31st International, 2012
- [7] Kyle Chard, Simon Caton, Kris Bubendorfer, Omer Rana, “Social Cloud Computing: A Vision for Socially Motivated Resource Sharing”, *Services Computing*, IEEE Transactions on (Volume:5 , Issue: 4), 23 June 2011, Pages 551-563
- [8] Hao-peng CHEN, Shao-chong LI, “A Queuing-based Model for Performance Management on Cloud”, *Advanced Information Management and Service (IMS)*, 2010 6th International conference, Nov. 30 2010-Dec. 2 2010, Pages 83-88
- [9] Diptangshu Pandit, Matangini Chattopadhyay, and Nabendu Chaki, “Resource Allocation in Cloud using Simulated Annealing”, *Applications and Innovations in Mobile Computing (AIMoC)*, Feb. 27 2014 – March 1 2014, Pages 21-27
- [10] Chu-Fu Wang, Wen-Yi, Hung, and Chen-Shun Yang, “A Prediction Based Energy Conserving Resources Allocation Scheme for Cloud Computing”, *IEEE International Conference on Granular Computing (GrC)*, 22-24 Oct. 2014, Pages 320-324
- [11] Parvathy S. Pillai, and Shrisha Rao, “Resource Allocation in Cloud Computing Using the Uncertainty Principle of Game Theory”, *System Journal IEEE*, 9 May 2014, Pages 1-12
- [12] Peter Mell, Timothy Grance, “The NIST Definition of Cloud Computing”, *National Institute of Standards and Technology Special Publication 800-145*, September 2011