Time Based Resource Allocation in Cloud Computing

[1]Ajesh Chopra, [2]Malti Rani, [2]Er. Tulika Mehta
[1,2]Dept. of CSE, Punjab institute of technology, Kapurthala, Punjab, India
[3]Entrepreneur and Researcher, Panjab University Chandigarh, UT, India

Abstract
With the advancement in technology and to stay abreast in the technological world, cloud computing emerged as a new paradigm in the business market. It is a utility based model which offers Pay Per use. The increasing cloud computing services offer great opportunities for clients to find the maximum service and finest pricing, which however raises new challenges on how to select the best service out of the huge group, now to find the most appropriate node for our completion of task which is at quite a larger scale is challenging. Cloud Analyst is a tool that helps developers to simulate large-scale Cloud applications with the purpose of understanding performance of such applications under various deployment configurations

Keywords
Cloud Computing, Infrastructure as a Service (Iaas), Resource Allocation, Virtual Machines

I. Introduction

Cloud computing which is the most trusted convenient delivery model of supplying information technology. Cloud computing can be seen as an innovation in different ways. Cloud computing seems to pose manageable visual challenges on a business level, both from an operational as well as from a strategic point of view. One fundamental advantage of the cloud paradigm is computation outsourcing, where the computational power of cloud customers is no longer limited by their resource-constraint devices. By outsourcing the workloads into the cloud, customers could enjoy the literally unlimited computing resources in a pay-per-use manner without committing any large capital outlays in the purchase of hardware and software and/or the not only computation is easy and economically however the bandwidth and storage, is even appropriate . Cloud computing is on demand service in which shared resources, information, software and other devices are provided according to the client requirement at specific time. Cloud computing is an evolving paradigm with changing definitions, it is defined as a virtual infrastructure which provides shared information and communication technology services, via an internet i.e. cloud. Cloud computing provides a computer user access to Information Technology (IT) services (i.e., applications, servers, data storage) without requiring an understanding of the technology or even ownership of the infrastructure.
The main cloud computing attributes are pay per use, elastic self provisioning through software, simple scalable services, virtualized physical resources. Models, such as cloud computing based on Virtual technologies enables the user to access storage resources and charge according to the resources access. Cloud computing platforms are based on utility model that enhances the reliability, scalability, performance and need based configurability and all these capabilities are provided at relatively low costs as compared to the dedicated infrastructures. Services like computation, software, data access and storage are provided to its user without its knowledge about physical location and configuration of the server which is providing the services. Cloud works on the principle of virtualization of resources with on-demand and pay-as–you go model policy.

When talking about a cloud computing system we divide it into two sections: the front end and the back end. They connect to each other through a network, usually the Internet. Fig 1 gives the pictorial representation of the various sections of the cloud computing system. The front end is the side the computer user, or client, sees. The back end is the "cloud" section of the system. The front end includes the client's computer (or computer network) and the application required to access the cloud computing system. It is also not all necessary that all cloud computing systems have the same user interface. Services like Web-based e-mail programs leverage existing. Some systems have unique applications that provide network access to clients. On the back end of the system are the various computers, servers and data storage systems that create the "cloud" of computing services.

Data center and cloud architectures continue to evolve to address the needs of large-scale multi-tenant data centers and clouds. These needs are centered around seven dimensions: scalability in computing, storage, and bandwidth, scalability in network services, efficiency in resource utilization, agility in service creation, cost efficiency, service reliability, and security.

Now analyzing computing system used under cloud ,it include practically any computer program which imagine, from data processing to video games. Usually, each application will have its own dedicated server. A central server administers the system, monitoring traffic and client demands to ensure everything runs smoothly. It follows a set of rules called protocols and uses a special kind of software called middleware. Middleware allows networked computers to communicate with each other. Most of the time, servers don't run at full capacity.

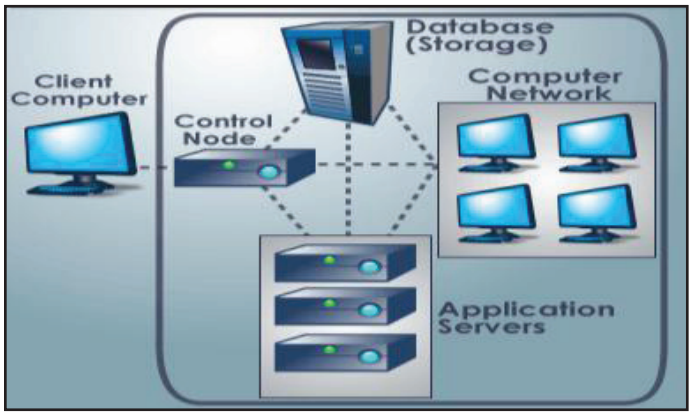

Fig. 1: Cloud Computing System [14]

That means there's unused processing power going to waste. It's possible to fool a physical server into thinking it's actually multiple servers, each running with its own independent operating system. The technique is called server virtualization. By maximizing the output of individual servers, server virtualization reduces the need for more physical machines. If a cloud computing company has a lot of clients, there's likely to be a high demand for a lot of storage space. Some companies require hundreds of digital storage devices. Cloud computing systems need at least twice the number of

storage devices it requires to keep all its clients' information stored. That's because these devices, like all computers, occasionally break down. A cloud computing system must make a copy of all its clients' information and store it on other devices. The copies enable the central server to access backup machines to retrieve data that otherwise would be unreachable. Making copies of data as a backup is called redundancy.

II. Importance of Allocation of Resources

A Resource allocation [13] is a subject that has been addressed in many computing areas, such as operating systems, grid computing, and datacenter management. A Resource Allocation System (RAS) in Cloud Computing can be seen as any mechanism that aims to guarantee that the applications' requirements are attended to correctly by the provider's infrastructure. Along with this guarantee to the developer, resource allocation mechanisms should also consider the current status of each resource in the Cloud environment, in order to apply algorithms to better allocate physical and/or virtual resources to developers' applications, thus minimizing the operational cost of the cloud environment.

Allocation of resources is an important component of cloud computing. Its efficiency will directly influence the performance of the whole cloud environment. It requires the type and amount of resources needed by each application in order to complete a user job.

Cloud resources can be seen as any resource (physical or virtual) that developers may request from the Cloud. For example, developers can have network requirements, such as bandwidth and delay, and computational requirements, such as CPU, memory and storage. Generally, resources are located in a datacenter that is shared by multiple clients, and should be dynamically assigned and adjusted according to demand. It is important to note that the clients and developers may see those finite resources as unlimited and the tool that will make this possible is the RAS. The RAS should deal with these unpredictable requests in an elastic and transparent way.

Cloud computing has its own features, an optimal RAS should avoid the following criteria as follows:

- Resource contention arises when two applications try to access the same resource at the same time.
- Scarcity of resource arises when there are limited resources and the demand for resources is high.
- Resource fragmentation arises when the resources are isolated. There would be enough resources but cannot allocate it to the needed application due to fragmentation into small entities.
- Over provisioning arises when the application gets surplus resources than the demanded one.
- Under provisioning of resources occurs when the application is assigned with fewer numbers of resources than it demanded

The hardware and software resources are allocated to the cloud applications on-demand basis. For scalable computing, Virtual Machines are rented [1].

Keeping in view the analysis of software and hardware paradigms it has been seen that correct knowledge of the nodes ie the resource nodes is the prime importance .
Also the applications would merely be applicable on per usage basis .

III. Advantages and Limitations

Taking a comparative look at the advantages and limitations of resource allocation in cloud.

A. Advantages

Major advantage of resource allocation is that user neither has to install software nor hardware to access the applications, to develop the application and to host the application over the internet. Also there is no limitation of place and medium. We can reach our applications and data anywhere in the world, on any system. Cloud providers can share their resources over the internet during resource scarcity.

B. Limitations

Since users rent resources from remote servers for their purpose, they don't have control over their resources.
Migration problem occurs, when the users wants to switch to some other provider for the better storage of their data.

IV. Projected Work

The dynamic resource allocation in cloud computing has attracted attention of the research community in the last few years. It is one of the most challenging problems in the resource management problems. Many researchers around the world have come up with new ways of facing this challenge. In [9] authors propose a model and a utility function for location-aware dynamic resource allocation.

A comprehensive comparison of resource allocation policies is covered in [10]. Hua's paper [12] proposed an ant colony optimization algorithm for resource allocation, in which all the characteristics in cloud are considered. It has been compared with genetic algorithm and annealing algorithm, proving that it is suitable for computing resource search an allocation in cloud computing environment.

This paper is not intended to address any specific resource allocation strategy, but to provide a review of some of the existing resource allocation techniques. Not many papers which analyses various resource allocation strategies are available as cloud computing being a recent technology. The literature survey focuses on resource allocation strategies and its impacts on cloud users and cloud providers. It is believed that this survey would greatly benefit the cloud users.

The Resource Allocation Strategy is all about integrating cloud provider activities for utilizing and allocating scarce resources within the limit of cloud environment so as to meet the needs of the cloud application.

V. Conclusion

In cloud paradigm, an effective resource allocation strategy is required for achieving user satisfaction and maximizing the profit for cloud service providers. This paper summarizes the classification of RAS and its impacts in cloud system. Some of the strategies discussed above mainly focus on CPU, memory resources but are lacking in some factors.

References

[1] Dorian Minarolli, Bernd Freisleben,"Uitlity–based Resource Allocations for virtual machines in cloud computing", IEEE, 2011.

[2] Gunho Lee, Niraj Tolia,"Parthasarathy Ranganathan, Randy H.Katz,"Topology aware resource allocation for data-intensive workloads", ACM SIGCOMM Computer Communication Review, 41(1), pp. 120-124, 2011.
[3] Li Li, Niu Ben,"Particle swarm optimization [M]", Beijing: Metallurgical Industry Press, 2009.
[4] Daniel Warneke, Odej Kao,"Exploiting dynamic resource allocation for efficient parallel data processing in the cloud", IEEE Transactions On Parallel And Distributed Systems, 2011.
[5] Valkenhoef G, Famchurn S D, Vytelingum P, et al. "Continuous Double Auctions with Execution Uncertainty[C]", Proc. of Workshop on Trading Agent Design and Analysis. Pasadena, California, USA [s.n.], 2009.
[6] Irwin D, Chase J S, Grit L et al.,"Sharing networked resources with brokered leases [C]", Proceedings of the USENIX Technical Conference, Boston, MA, USA, 2006, pp. 299-212.
[7] Padala P, Shin K G, Zhu Xiao Yun et al.,"Adaptive control of virtualized resources in utility computing environments[C]", Proceedings of the 2nd ACMSIGOPS/EuroSys European Conference on Computer Systems 2007", Lisbon, Portugal, 2007, pp. 289-302.
[8] Valkenhoef G, Famchurn S D, Vytelingum P, et al. "Continuous Double Auctions with Execution Uncertainty[C]", Proc. of Workshop on Trading Agent Design and Analysis. Pasadena, California, USA [s.n.], 2009.
[9] Irwin D, Chase J S, Grit L et al.,"Sharing networked resources with brokered leases[C]", Proceedings of the USENIX Technical Conference, Boston, MA, USA, 2006, pp. 299-212.
[10] Padala P, Shin K G, Zhu Xiao Yun et al.,"Adaptive control of virtualized resources in utility computing environments[C]", Proceedings of the 2nd ACMSIGOPS/EuroSys European Conference on Computer Systems 2007. Lisbon, Portugal, 2007, pp. 289-302.
[11] Cheng Shiwei, Pan Yu.,"Credibility-based dynamic resource distribution strategy under cloud computing environment [J]. Computer Engineering, 2011, 6(37), pp. 45-48.
[12] Hua Xiayu, Zheng Jun, Hu Wenxin,"Ant colony optimization algorithm for computing resource allocation based on cloud computing environment [J]", Journal of East China Normal University (Natural Science). 2010.
[13] Glauco Estácio Gonçalves, Patrícia Takako Endo, Thiago Damasceno,"Resource allocation in clouds: Concepts", Tools and Research Challenges.
[14] CloudComputing architecture: [Online] Available: http://blog.infizeal.com/2011/12/cloud -computing-architecture.htm

Er. Ajesh Chopra graduated with B tech in computer science engineering from PTU Jalandhar. Presently he is pursuing M.TECH in CSE from PTU main campus. He is NSS volunteer from last three years. He has also attended G20 youth forum in May 2014, held in germany.
He is also nominated for Indira Gandhi national award for the year 2014 – 15

Er. Tulika Mehta is an Electronics Engineering professional having worked previously with Telecom Industry. She is presently a young entrepreneur also associated to Panjab University contributing in research and development activities in the field of wireless communications. She is presently working as a regional head for her organization Fortune Professional consultants Pvt. Ltd. In this organization she is also into research and management. She is ME in ECE MBA in HR both from Panjab university Chandigarh. She has published several technical and management research papers.
She is active participant in organizing events and conference in India and abroad.