Effectively Scheduling User Request in Cloud Environment

Divya, Anil Arora, Aakash Gupta

Abstract— **Cloud environment is an environment where multiple users can be connected publically or privately to access the services and the products. When multiple users will request for the same Service and a product, then there is a need to schedule the request as well as to allocate the resources properly. Considering the situation of multiple clouds and multiple virtual machines , situation become worse when all the request are allocated to a single cloud which may lead to Overload condition. To avoid this, process migration can be done from one cloud to another cloud by releasing the process from one cloud and allocating it to another cloud and from one virtual machine to another .Clouds can be assigned on some priority basis. We are proposing an algorithm of Resource allocation in this paper.**

Index Terms— **Cloud computing, virtual machines, migration approach, deployment modeling**

I. INTRODUCTION

Cloud computingis a construct that allows you to access applications that actually reside at a location other than your computer or other internet-connected device. It has become one of the most talked about technologies in recent times and has got lots of attention from media as well as analysts because of the opportunities it is offering. The asset of cloud computing is that another company hosts your application (or the suite of applications, for that matter), that is they handle the costs of servers, manage the software updates, and—depending on how you craft your contract—you pay less for the service. Virtualization plays an important role in cloud computing. Virtualization is the creation of a virtual (rather than actual) version of something, such as a hardware platform, operating system (OS), storage device, or network resources. Virtualization and cloud computing are transforming IT. Cloud computing leverages Virtualization to enable a more scalable and elastic model for delivering IT services. Cloud computing systems generally falls into three course grain categories. Infrastructure as a service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS).
The NIST working definition summarizes cloud computing as [2]:
"A model for enabling 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." [3]

Divya, M.Tech Pursuing**,** G.I.E.T,Sonepat,Sonipat, Pin no.131001
Anil Arora, Assistant Professor, Management Coordinator G.I.E.T,Sonepat, Sonipat, Pin no.131001
Aakash Gupta, M.Tech Pursuing, G.I.E.T,Sonepat,Sonipat, Pin no.131001

Cloud computing means different to different people, its benefits are different to different people. To IT managers, it means to minimize capital-expenditure by outsourcing most of the hardware and software resources. To ISVs, it means to reach out to more users by offering a SaaS solution. To end users, it means to access an application from anywhere using any device. In cloud computing, the relationship between the user and machine are many-to-many. Many users can access an application that is served from many machines. Now, what was the reason of this evolution? What were the driving factors behind this? The reason for the evolution from PC-based application to Internet-based application was obvious. This happened because of the need of multiple users trying to access an application from their own machines [3].
The only way that it was possible was to have the application hosted on a central server and having separate client applications communicate to it. The evolution from internet-based applications to cloud computing, I think, is a bit more complex. There are several industry trends and user behaviors affecting this shift in the technology. We will get more into those in my next blog. Here, I touch upon what I believe is the biggest driving factor behind cloud computing [3].

1.1 Virtualization:

In computing, virtualization is the creation of a virtual (rather than actual) version of something, such as a hardware platform, operating system (OS), storage device, or network resources. While a physical computer in the classical sense is clearly a complete and actual machine, both subjectively (from the user's point of view) and objectively (from the hardware system administrator's point of view), a virtual machine is subjectively a complete machine (or very close), but objectively merely a set of files and running programs on an actual, physical machine (which the user need not necessarily be aware of) [3]

1.2 Cloud computing services delivery models:

Cloud computing systems generally falls into three course grain categories. Infrastructure as a service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS). Many companies are offering services [3].

1.2.1 Infrastructure as a Service (IAAS):

Infrastructure as a Service (IaaS) provisions hardware, software, and equipments to deliver software application environments with a resource usage-based pricing model. Infrastructure can scale up and down dynamically based on application resource needs. Typical examples are Amazon EC2 (Elastic Cloud Computing) Service and S3 (Simple Storage Service) where compute and storage infrastructures are open to public access with a utility pricing model. This

basically delivers virtual machine images to the IaaS provider, instead of programs, and the Machine can contain whatever the developer want.

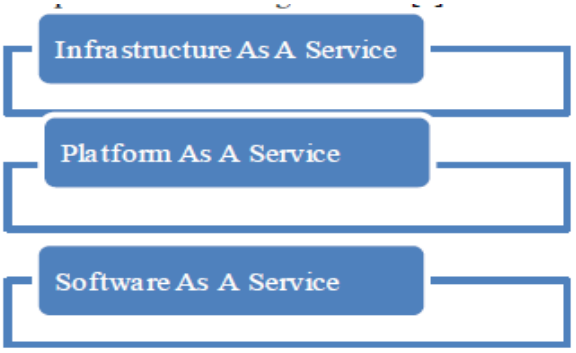

Figure 1: Service delivery models.

1.2.2 Platform As A Service (PAAS):

Platform as a Service (PaaS) offers a high-level integrated environment to build, test, and deploy custom applications. Generally, developers will need to accept some restrictions on the type of software they can write in exchange for built-in application scalability. An example is Google's App Engine, which enables users to build Web applications on the same scalable systems that power Google applications, Web application frameworks, Python Django (Google App Engine), Ruby on Rails (Heroku), Web hosting (Mosso), Proprietary (Azure, Force.com).

1.2.3 Software As A Service (SAAS):

User buys a Subscription to some software product, but some or all of the data and codes resides remotely. Delivers special-purpose software that is remotely accessible by consumers through the Internet with a usage-based pricing model. In this model, applications could run entirely on the network, with the user interface living on a thin client. Sales force is an industry leader in providing online CRM (Customer Relationship Management) Services. Live Mesh from Microsoft allows files and folders to be shared and synchronized across multiple devices [3].

1.3. CLOUD DEPLOYMENT MODEL

There are mainly four types of cloud modeling used in communication. These all are different from each other. Each has their own advantages and disadvantages. These are [9]:

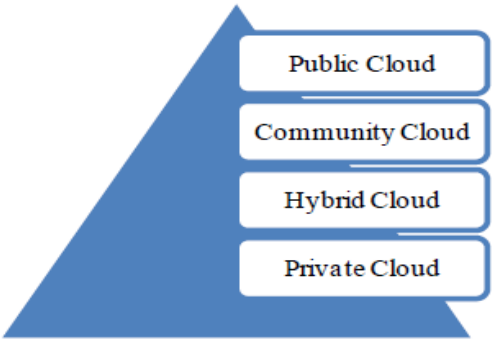

Figure 2: Deployment modeling

1.3.1 Public Cloud:

Public cloud realizes the key concept of sharing the services and infrastructure provided by an off-site, third-party service provider in a multi-tenant environment.

1.3.2 Community Cloud:

Community cloud is shared by several organizations and is supported by a specific community that has shared interests and concerns [9] (security, compliance, jurisdiction, etc.), whether managed internally or by a third-party and hosted internally or externally. The costs are spread over fewer users than a public cloud (but more than a private cloud), so only some of the cost savings potential of cloud computing are realized.

1.3.3 Hybrid Cloud:

Hybrid cloud is a composition of two or more clouds (private, community or public) that remain unique entities but are bound together, offering the benefits of multiple deployment models [9].

Hybrid clouds lack the flexibility, security and certainty of in-house applications. Hybrid cloud provides the flexibility of in house applications with the fault tolerance and scalability of cloud based services.

3.4 Private Cloud:

Private cloud entails sharing services and infra- structure provided by an organization or its specified service provider in a single-tenant environment. Enterprises' mission- critical and core-business applications are often kept in a private cloud [9].

II. LITERATURE REVIEW

In Year 2009, Kento Sato performed a work," A Model-Based Algorithm for Optimizing I/O Intensive Applications in Clouds using VM-Based Migration". Author propose a novel model-based I/O performance optimization algorithm for data-intensive applications running on a virtual cluster, which determines virtual machine(VM) migration strategies and the weighted edge represents a migration of a VM and time [4].

In Year 2010, Takahiro Hirofuchi performed a work," Enabling Instantaneous Relocation of Virtual Machines with a Lightweight VMM Extension". In this paper, Author proposes an advanced live migration mechanism enabling instantaneous relocation of VMs. To minimize the time needed for switching the execution host, memory pages are transferred after a VM resumes at a destination host. In addition, for memory intensive workloads, Presented migration mechanism moved all the states of a VM faster than existing migration technology [5]. In Year 2011, JiaRao performed a work," Self-Adaptive Provisioning of Virtualized Resources in Cloud Computing". In this paper, Author propose a distributed learning mechanism that facilitates self-adaptive virtual machines resource provisioning. Author treat cloud resource allocation as a distributed learning task, in which each VM being a highly autonomous agent submits resource requests according to its own benefit. The mechanism evaluates the requests and replies with feedbacks. Author develops a reinforcement learning algorithm with a highly efficient representation of experiences as the heart of the VM side learning engine [6].

In Year 2012, Michael Menzel performed a work," CloudGenius: Decision Support for Web Server Cloud Migration". Author presents a framework (called CloudGenius) which automates the decision-making process

based on a model and factors speci_cally for Web server migration to the Cloud. Cloud- Genius leverages a well known multi-criteria decision making technique, called Analytic Hierarchy Process, to automate the selection process based on a model, factors, and QoS parameters related to an application. Authors present an implementation of CloudGenius that has been validated through experiments [7]. In Year 2013, Christina Delimitrou performed a work," Paragon: QoS-Aware Scheduling for Heterogeneous Datacenters". Author present Paragon, an online and scalable DC scheduler that is heterogeneity and interference-aware. Paragon is derived from robust analytical methods and instead of profiling each application in detail; it leverages information the system already has about applications it has previously seen. It uses collaborative filtering techniques to quickly and accurately classify an unknown, incoming workload with respect to heterogeneity and interference in multiple shared resources, by identifying similarities to previously scheduled applications [8].

III. PROPOSED ALGORITHM

There are different service providers that provide Cloud Services to different users for different business services. We are proposing a middle layer Architecture called Intermediate Layer. In this concept we are placing a layer between the users and the web services. The middle layer will accept the user requests and also monitor the cloud servers for the available load over the services. The middle layer will perform the cloud allocation sequentially and if the service allocation is not possible for a specific cloud, it will perform the migration of process from one cloud to other.

Objectives:

Create an Intermediate Architecture that will accept the user request and monitor the cloud servers for their capabilities.

Scheduling of the users requests is performed to identify the order of allocation of the processes.

Performing the effective resource allocation under defined parameters and the cloud server capabilities.

Define a dynamic approach to perform the process migration from one cloud to other.

Analysis of the work using graph under different parameters

Research Design:

The proposed system is middle layer architecture to perform the cloud allocation in case of under load and overload conditions. The over load conditions will be handled by using the concepts of process migration. The middle layer will exist between the clouds and the clients. As the request will be performed by the user this request will be accepted by the middle layer and the analysis of the cloud servers is performed by this middle layer. The middle layer is responsible for three main tasks:

1. Scheduling the user requests
2. Monitor the cloud servers for its capabilities and to perform the process allocation
3. Process Migration in overload conditions

Algorithm:

1. Input the M number of Clouds with L number of Virtual Machines associated with each cloud.
2. Define the available memory and load for each virtual machine.
3. Assign the priority to each cloud.
4. Input N number of user process request with some parameters specifications like arrival time, process time, required memory etc.
5. Arrange the process requests in order of memory requirement
6. For i=1 to N
7. {
8. Identify the priority Cloud and Associated VM having AvailableMemory>RequiredMemory(i)
9. Perform the initial allocation of process to that particular VM and the Cloud
10. }
11. For i=1 to N
12. {
13. Identify the Free Time slot on priority cloud to perform the allocation. As the free slot identify, record the start time, process time, turnaround time and the deadline of the process.
14. }
15. For i=1 to N
16. {
17. Iffinishtime(process(i))>Deadline(Process(i))
18. {
19. Print “Migration Required”
20. Identify the next high priority cloud that having the free memory and the time slot and perform the migration of the process to that particular cloud and the virtual machine.
21. }
22. }

IV. CONCLUSIONS AND FUTURE WORK

In this paper, a resource allocation scheme on multiple clouds in both the under load and the over load conditions are take. As the request is performed by the user, certain parameters are defined with each user request, these parameters includes the arrival time, process time, deadline and the input output requirement of the processes. The cloud environment taken in this work is the public cloud environment with multiple clouds. Each cloud is here defined with some virtual machines. To perform the effective allocation, we have assigned some priority to each cloud. The virtual machines are here to perform the actual allocation. These are defined with certain limits in terms of memory, load etc.

In our future work, in case of over load condition, the migration of the processes is performed from one cloud to other. The Future enhancement of the work is possible in the following directions:

1. The presented work is defined the overload conditions in terms of deadline as well as the
memory limit of the clouds. In future some other parameters can also be taken to decide the migration condition.

2. The presented work is defined for the public cloud environment, but in future, the work can be extended to private and the hybrid cloud environment.

REFERENCES

[1] Shigeru Imai," Elastic Scalable Cloud Computing Using Application-Level Migration", 2012 IEEE/ACM Fifth International Conference on Utility and Cloud Computing 978-0-7695-4862-3/12 © 2012 IEEE.

[2] BalajiViswanathan," Rapid Adjustment and Adoption to MIaaS Clouds", Middleware 2012 Industry Track, December 3-7, 2012, Montreal, Quebec, Canada.ACM 978-1-4503-1613-2/12/12.

[3] HadiGoudarzi," SLA-based Optimization of Power and Migration Cost in Cloud Computing", 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing 978-0-7695-4691-9/12© 2012 IEEE.

[4] Kento Sato," A Model-Based Algorithm for Optimizing I/O Intensive Applications in Clouds using VM-Based Migration", 9th IEEE/ACM International Symposium on Cluster Computing and the Grid 978-0-7695-3622-4/09© 2009 IEEE.

[5] Takahiro Hirofuchi," Enabling Instantaneous Relocation of Virtual Machines with a Lightweight VMM Extension", 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing 978-0-7695-4039-9/10© 2010 IEEE.

[6] JiaRao," Self-Adaptive Provisioning of Virtualized Resources in Cloud Computing", SIGMETRICS'11, June 7–11, 2011, San Jose, California, USA.ACM 978-1-4503-0262-3/11/06.

[7] Michael Menzel," CloudGenius: Decision Support for Web Server Cloud Migration", WWW 2012, April 16–20, 2012, Lyon, France. ACM 978-1-4503-1229-5/12/04.

[8] Christina Delimitrou," Paragon: QoS-Aware Scheduling for Heterogeneous Datacenters", ASPLOS'13, March 16–20, 2013, Houston, Texas, USA.ACM 978-1-4503-1870-9/13/03.

[9] HadiGoudarzi," SLA-based Optimization of Power and Migration Cost in Cloud Computing", 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing 978-0-7695-4691-9/12© 2012 IEEE