Dynamic Resource Prediction and Provisioning Strategy in Multi-Clouds

P.Rajeshwari[1], B.Namratha[2]
Anurag Group of Institutions

***Abstract*—** Cloud computing gradually becomes a ubiquitous choice for modern IT-solution of business activitiesand also for several scientific computing needs for research community. The diverse features of Cloud such as on-demand self-service, quality of service, pay-per-usage pricing, virtualization and elasticity make the cloud more popular in industries as well as research communities. Since modern Cloud systems operate in an open and dynamic world characterized by continuous changes, the development of efficient resource provisioning policies for Cloud-based services becomes increasingly challenging.

Multi-Cloud or inter-clouds or cloud-of-clouds is the use of multiple cloud computing services in a single heterogeneous architecture. There are number of reasons for deploying multi cloud architecture, like overcoming vendor lock-in issues, increasing flexibility through choice etc. It is similar to the use of best-of-breed applications from multiple developers on a personal computer. It is recognition of the fact that one provider cannot be everything for everyone.

Service provisioning is one of the important problem in cloud computing. The mapping of the cloud resources in forms of virtual machines (VMs) to fulfill the customer requests is very challenging task. There are several resource provisioning algorithms proposed in the literature. Best-Fit Heuristic approach is proposed for resource allocation in service clouds.

In this work we propose to extend the Best-Fit Heuristic approach for matching the customer requests (tasks) to the VMs of the Clouds to minimize the overall completion time of the requests in multi-clouds.

Keywords— Cloud computing, multi clouds, dynamic resource provisioning.

I. INTRODUCTION

Cloud provides services over the network. The standard definition can be described as follows:

"Cloud computing is a model for enabling 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[1]".

Cloud Service models can be categorized into:

1. Infrastructure as a service (IaaS): It provides access to fundamental resources such as physical machines, virtual machines, virtual storage, networks etc.
2. Platform as a service (PaaS): It provides the run time environment for applications, development and deployment tools.
3. Software as a service (SaaS): It allows using software applications as a service to end users.

There are many benefits of cloud computing some of them are flexibility, scalability, portability, lower computer cost, improved performance, reduced software costs, instant software updates etc.

II. MULTI TENANCY

Multi tenancy refers to a principle where a single instance of a software application serves multiple customers (tenants) i.e. vendor provides single version of software for all its customers. It is regarded as important attribute of cloud computing. Single instance of software is run on service provider's infrastructure.

Advantages are:

1) Lower costs through economies of scale
2) Shared infrastructure leads to lower costs.
3) Ongoing maintenance and updates.
4) Highly configurable
5) Efficient and sustainable scalability
6) Reduction in operational cost and complexity of product life cycle

III. MULTI CLOUDS

Multi cloud is a federation of multiple cloud providers. It follows two different approaches, first is the provider-centric in which one provider can request and provision the resources based on the agreement between the different cloud providers to satisfy the user requirements which also improves the provider's profit.

The advantage with this approach is, a federated and cooperated environment is created among the cloud providers which lead to their profits. The second approach is client-centric, in which the client has the option to choose the required resources from multiple cloud providers and can provision the resources. It helps the cloud users to avoid the vendor lock-in.

The resource allocation and provisioning by the multi tenants can be implemented with the help of creating virtual machines on physical servers. There are two strategies for resource allocation and provisioning: static and dynamic.

In the static approach, the cloud users can request for the required resources to allocate and provision whenever they run their application instead of provisioning in priory. And even the cloud providers can also allocate the resources on demand which can save the cost.

In dynamic resource allocation, it takes more time to restart/shutdown the Virtual machines for resource allocation/re-allocation and minimizing this downtime becomes very crucial.

The technology of streaming VM allows the customer to preview the VM before it is entirely ready. However, the VM is still not instantly available until enough proportion of it is ready. There will be some time before this technology can fully answer the need of instant provisioning. From the business perspective, a simple yet straightforward solution to this problem is to ask all the customers to provide future VM requests schedule so the cloud service provider can prepare all the VMs ahead of time. However, this is impossible for many reasons: (1) the customers have no obligation and usually unwilling to provide their schedule. (2) The customers themselves are unable to know when the computing resources are needed. (3) The constituents of customers are always changing. (4) The actual schedule may change at any time. Facing these technology limitations and business constraints, the only practical, achievable and effective solution to provide instant cloud is to predict the demands and prepare the VMs in advance.

Although multi-tenant systems help providers save cost by allocating multiple tenants to the same instance of an application, they incur huge reconfiguration costs. Cost and time spent on these reconfiguration activities can be reduced by re-constructing tenants from existing tenant configurations supported by service providers. Multi-tenant cloud- based systems also lack the facility of allowing clients to specify their requirements.

Specifically, in the Infrastructure-as-a-Service (IaaS) model there are more than fifty providers offering computing resources on subscription-basis using pay-as-you-go model. This presents a challenge in the selection of suitable cloud providers but at the same time, it opens an opportunity to create interconnected computing environments where the applications can provision resources from multiple cloud providers.

IV. RELATED WORK:

Useful resource provisioning drawback is widely studied and active research area because of ever growing demand for multi-tenant and multi-cloud techniques. In [4] the authors discussed the possibilities and challenges for efficient parallel data processing in clouds making use of research task Nephele. Nephele is the primary information processing framework to explicitly make the most the dynamic useful resource allocation provided through at present's IaaS clouds for each, task scheduling and execution. Precise tasks of a processing job may also be assigned to distinct varieties of virtual machines which might be automatically instantiated and terminated for the duration of the job execution. ASAP (A Self-Adaptive Prediction System for Instant Cloud Resource Demand Provisioning) is proposed by Jiang et al. [5]. ASAP targets to extract high traits from VM provisioning request movement and notify the provisioning system to prepare VMs in advance. A solution based on the best-reply dynamics, which is compatible for a distributed implementation is proposed in [6]. They tested the effectiveness of this approach via performing numerical checks, given that multiple workloads and process configurations.

In [7], the authors developed Prediction-based resource measurement and provisioning procedures utilising Neural network and Linear Regression to meet upcoming resource needs. Experimental outcome display that the proposed process presents more adaptive resource management for functions hosted in cloud atmosphere, an foremost mechanism to obtain on-demand useful resource allocation within the cloud.

In [8], the authors proposed a virtual machine allocation strategy based on statistics. And The contribution of this paper include:(1) A load-resource model for estimating the resource demand of each virtual machine, (2) An algorithm for assigning virtual machines onto the hosts amongst useful resource pool in keeping with the resources requirement of virtual machine.

Experiments show that load-useful resource model can accurately estimate the useful resource requisites of virtual machines for physical host, and virtual machine assigning algorithm can achieve higher load balancing in comparison with the first-fit and best-fit algorithm.

In [9], the authors proposed a new approach for dynamic autonomous resource management in computing clouds. The main contribution of this work is two-fold. First, adopted a distributed architecture in which resource administration is decomposed into independent jobs, each of which is carried out by means of self sustaining Node agents which can be tightly coupled with the physical machines in a data centre. Second, the self sufficient Node retailers carry out configurations in parallel by way of multiple criteria decision analysis making use of the PROMETHEE procedure. Simulation results show that the proposed procedure is promising in phrases of scalability, feasibility and flexibility.

In [10], the authors formulated Demand for computing power and different resources as a resource allocation problem with multiplicity, the place computations that ought to be carried out at the same time are represented as tasks and a later task can reuse resources released via a prior task. The authors showed that discovering a minimized allocation is NP-complete. They offered an approximation algorithm with a proof of its approximation bound that may yield just about ultimate solutions in polynomial time. Enterprise users can make the most the option to slash the leasing cost and amortize the administration overhead (e.g., setting up VPNs or configuring a cluster). Cloud vendors could utilize the option to share their assets among a larger number of customers.

In [11], the authors proposed a dynamic resource allocation method based on the load of VMs on IaaS, abbreviated as DAIaS. This procedure allows for customers to dynamically add and/or delete one or more instances on the basis of the load and the conditions specified by the user. The authors implemented a prototype to evaluate the effectiveness and efficiency of DAIaS. Additionally, they performed an experiment to extract the prototype on a real cloud provider, particularly, Amazon EC2.

In [12], the authors formulate the dynamic VM provisioning and allocation problem for the auction-based model as an integer program considering multiple types of resources, then designed truthful greedy and most reliable mechanisms for the problem such that the cloud provider provisions VMs based on the requests of the profitable customers and determines their payments. The authors proven that the proposed mechanisms are fair, that is, the users do not have incentives to manipulate the system about their requested bundles of VM instances and their valuations and likewise performed extensive experiments making use of actual workload traces in an effort to investigate the performance of the proposed mechanisms. Proposed mechanisms achieve promising results in phrases of income for the cloud provider.

The problem of VM allocation can be divided in two: the first part is the admission of new requests for VM provisioning and placing the VMs on hosts, whereas the second part is the optimization of the current VM allocation. The first part can be seen as a bin packing problem with variable bin sizes and prices. [15] Solved it by applying a modification of the Best Fit Decreasing (BFD) algorithm. In this modification, the Modified Best Fit Decreasing (MBFD) algorithms, we sort all VMs in decreasing order of their current CPU utilizations, and allocate each and every VM to a host that supplies the least increase of energy consumption due to this allocation. This permits leveraging the heterogeneity of resources through picking the most energy-efficient nodes first. The complexity of the allocation part of the algorithm is n • m, the place n is the number of VMs that have to be allocated and m is the number of hosts.

V. PROPOSED ALGORITHM

In multi cloud environment, we need to test the proposed Best-Fit Heuristic approach for matching the customer requests (tasks) to the VMs of the Clouds to minimize the overall completion time of the requests by assuming c number of multi clouds, in each cloud of cj, assuming i number of hosts in each data center, and allocating optimized time for resource allocation and provisioning to VMs based on the price using the proposed Best Fit Heuristic approach for multi clouds.

A. Algorithm:

Procedure VM allocation (VM_j, $Host_i$, $cloud_c$) //number of VMs, physical hosts and clouds
for each cloud c to n
add service tenants to VM based on matching configuration
sort VMs in decreasing order based on price
for all VM_j to p do
for all Host i to q do
if VM_j fits in $Host_i$ then
Calculate the remaining CPU capacity of host after VM has been added
else
No processing element left in hosts or failed to be allocated
end if
end for
Start a new host in data center, and allocate remaining VM into a new Host
end for
Start a new data centre in next cloud, and allocate new host and allocate remaining VM based on price.
end for
end procedure

VI. CONCLUSION

In this paper we discussed about various existing dynamic resource allocation strategies in multi-tenant, single-cloud environment. The existing Best-fit heuristic algorithm for multi-tenant single cloud approach is modified for multi-cloud environment. In this regard, we proposed a new algorithm which dynamically allocates and provisions the resources required for the VMs in advance for multi-cloud.

VII. REFERENCES:

[1]. P. Mell and T. Grance,"The nist definition of cloud computing", NIST Special Publication 800-145, 2011.
[2]. M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica and M. Zaharia, "A view of cloud computing", Communications of the ACM, Volume 53, Issue 4,pp. 50-58, April 2010.
[3]. N. M. Calcavecchia, O. Biran, E. Halad and Y. Moatti, "VM Placement Strategies for Cloud Scenarios", In: Proceeding of the 5th IEEE International Conference on Cloud Computing (CLOUD), 2012.
[4]. Warneke D, Kao O. Exploiting dynamic resource allocation for efficient parallel data processing in the cloud. IEEETransactions on Parallel and Distributed Systems 2011; 22(6):985–997
[5]. Jiang Y, Perng CS, Li T, Chang R. ASAP: a self-adaptive prediction system for instant cloud resource demandprovisioning. In: Proceeding of the 11th International Conference on Data Mining (ICDM), pp. 1104–1109, 2011.
[6]. Mauro Passacantando, Danilo Ardagna, Anna Saviz, "Service Provisioning Problem in Cloud and multi-Cloud Systems", INFORMS Journal on Computing, volume 28, pp. 265-277, 2016.
[7]. Islam S, Keung J, Lee K, Liu A. Empirical prediction models for adaptive resource provisioning in the Cloud. FGCSD. F. Specht,"Probabilistic neural networks", Neural Networks 3, pp.109118, 1990.Elsevier, 2012; 28(1):155–162.
[8]. Zhang Z, Xiao L, Li Y, Ruan L. AVM-based resource management method using statistics. In: Proceeding of the International Conference on Parallel and Distributed Systems, 2012.C. Cortes and V. Vapnik,"Support-Vector Networks", MachineLearning, Volume 20, Issue 3, pp 273-297, 1995.
[9]. Yazir YO, Matthews C, Farahbod R, Neville S, Guitouni A, Ganti S, Coady Y. Dynamic resource allocation in computing clouds using distributed multiple criteria decision analysis. In: Proceeding of IEEE 3rd International Conference on Cloud Computing, pp. 91–98, 2010.J. R. Quinlan,"C4.5: Programs for Machine Learning", Morgan Kauffman Publishers, 1993.
[10]. Chang N, Ren J, Viswanathan R. Optimal resource allocation in clouds. In: Proceeding of the 3rd IEEE International Conference on Cloud Computing, pp. 418–425, 2010.G. Shmueli,"Practical Time Series Forecasting: A Hands-on Guide", 2nd ed., ISBN-13: 978-1468053456, 2011.
[11]. Inomata A, Morikawa T, Ikebe M, Rahman M. Proposal and evaluation of dynamic resource allocation method based on the load of VMs on IaaS. In: Proceeding of 4th IFIP International Conference on New Technologies, Mobility and Security, 2011.
[12]. Nejad MM, Mashayekhy L, Grosu D. A family of truthful greedy mechanisms for dynamic virtual machine provisioning and allocation in clouds. In: Proceeding of the IEEE 6th International Conference on Cloud Computing.pp. 188–195. 2013.
[13]. Espadas J, Molina A, Jimnez G, Molina M, Ramrez R, Concha D. A tenant-based resource allocation model for scaling Software-as-a-Service applications over Cloud computing infrastructures. Future Generation Computer Systems 2013; 29(1):273–286.
[14]. L.Ramachandran, N.C.Narendra and K. Ponnalagu, "Dynamic provisioning in multi-tenant service clouds", Service Oriented Computing and Applications, 2012.
[15]. Verma, Manish, et al. "Dynamic resource demand prediction and allocation in multi-tenant service clouds." Concurrency and Computation: Practice and Experience (2016).