Review on Scheduling of Resources in Cloud Environment

Avanika M Chavada
Dept. of Computer Science and Engineering
Parul Institute of Technology
Vadodara, India

Upendra R Bhoi
Dept. of Computer Science and Engineering
Parul Institute of Technology
Vadodara, India

ABSTRACT
Cloud computing provides an environment, where a large pool of systems are connected in networks (i.e., private or public), to provide dynamically scalable infrastructure for storage of application, data and file. Resource allocation is a way to assign resources to the needed applications. Cloud service providers use lease manager for better resource allocation. Haizea is a lease manager that takes lease request and make scheduling decision. We describe cloud computing and its properties, different leases in Haizea and scheduling of resources in Haizea.

Keywords
Cloud computing, Resource Allocation, Haizea, OpenNebula, lease scheduling in Haizea

1. INTRODUCTION
In the advanced information and technology period, cloud computing is the upcoming generation framework. Cloud Provides services over the network, typically Internet.

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".[10]

The resources can be assigned and reassigned dynamically based on the requirement. Cloud fulfils unpredictable usage demands of consumers by scaling up and down the resources. The consumers use resources based on pay-per-use basis. Thus, resource allocation in cloud computing is important issue to provide qualitative service to the cloud consumer. [11]

Since a cloud consist of a huge amount of virtual machines and physical servers, efficient management of this virtual infrastructure becomes an important interest. Cloud service providers use lease managers for better resource allocation.

A Lease is defined as "a negotiated and renegotiable agreement between a resource provider and a resource consumer, where the former agrees to make a set of resources available to the latter, based on a set of lease terms presented by the resource consumer".[1]

1.1 Haizea
Haizea is an open-source VM-based lease management architecture. [1]Haizea uses leases to manage resources. It takes lease request from the user and makes scheduling decision that is at what time a set of virtual machines must start and stop. (See Figure 1)

In Haizea, the lease terms include the hardware resources (i.e., CPU, memory), software environments that are to be installed, and the availability period during which the resources must be available. For lease scheduling, Haizea maps leases to VMs. [12]

Haizea can be used in three modes: Open Nebula mode, unattended simulation mode and interactive simulation mode.

1) Unattended simulation mode: Takes a list of lease requests (from a trace file) and processes them in simulated time.

2) Interactive simulation mode: Enactment actions are simulated, but Haizea runs in "real time". Instead of having to provide a list of lease requests beforehand, Haizea's command-line interface is used to request leases interactively and query the status of the scheduler.

3) Open Nebula mode: Haizea is plugged into Open Nebula as backend scheduler.

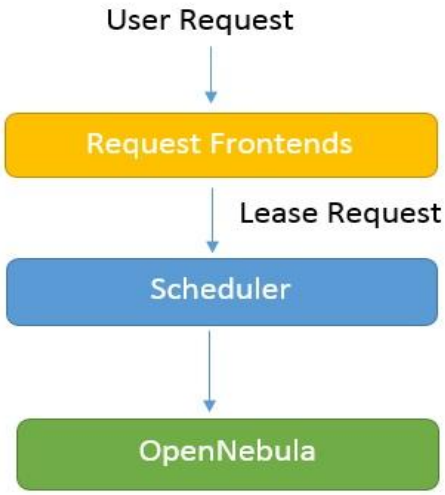

Fig 1: Haizea Architecture

There are three types of leases supported by Haizea: Immediate, Advance Reservation (AR) and Best-Effort (BE) lease. Immediate and AR leases have higher priority than BE leases.BE leases are pre-emptible whereas immediate and AR leases are non-pre-emptible.To fulfil request of immediate and AR leases, the resources allocated to BE lease are given to immediate and AR lease and BE lease is placed in a queue.

Haizea collects the data that can be analyzed offline (i.e., accepted/rejected leases, waiting time etc.) and this data is store in 'probes'. The scheduler keeps track of available

resources with the help of slot table. The table contains information like resource usage, the no. of nodes and the duration of a lease.

1.2 OpenNebula

The open source OpenNebula virtual infrastructure engine provides the functionality needed to deploy, monitor and control VMs on a pool of distributed physical resources.[2]OpenNebula can be integrated with different type of hypervisor (i.e., Xen or KVM) and infrastructure configurations.

OpenNebula is composed of three main components: The open nebula core -manages lifecycle of VM & provides management interface, the capacity manager - governs functionality provided by core and drivers - drivers for hypervisors supported by Open Nebula.

Haizea provides scheduling brain to OpenNebula.

2. LITERATURE REVIEW AND RELATED WORK

Initially clouds use immediate provisioning model, where virtualized resources are allocated at the time they are requested. This allocation was done without the possibility of requesting resources at a specific time and requests are placed in FIFO manner.

B. Sotomayor, R.S. Montero, I.M. Llorente, I. Foster [3] introduce capacity provisioning model which uses resource leases as a fundamental provisioning abstraction. According to this architecture, advance reservation leases are scheduled first and best-effort leases are placed in a queue and served in FIFO manner. They showed that OpenNebula and Haizea can be integrated together to provide a VM management solution to support all the various lease types. Lease suspension by Haizea schedule the leases such that there would not be any overlay in the Virtual Machine actions. There will be higher accuracy of virtual machine when assigned specific period of time needed is assigned.

B. Sotomayor, K.Keahey, I. Foster[4] introduce a lease management architecture that allows resource consumers to request resource leases with semantics that encompass three things-hardware resource, software environments, and availability. According to this architecture, advance reservation leases are scheduled first and best-effort leases are placed in a queue. The scheduler keeps track of available resources with the help of slot table. The table contains information like resource usage, the no. of nodes and the duration of a lease. The authors described a virtual resource model, and a set of scheduling strategies that combine both batch job platforms as deadline sensitive. The overhead of deployment and management of virtual machine for the batch jobs are very large. And this two factors varies most. The virtual model separates resource utilization of the overhead of Virtual Machine deployment from resources available to VM itself. Thus, this type of scheduling allocates the resource slots of overhead resources to the virtual machine slots in equal proportion.

Nathani A., Chaudhary.S and Somani G.[5] proposed a policy based resource allocation method in IaaS cloud. The general system of Haizea considers only a single slot to schedule a lease. This existing system not only considers a single slot but also tries to put in several slots together. For this, swapping and backfilling techniques are used. In swapping, the order of lease reschedule is decided. For that, two leases are swapped if the first lease request less resources than the second lease and both the leases can be scheduled within its deadline. Thus, swapping will arrange the leases in decreasing order of required number of nodes. Backfilling is used to utilize idle resources. It will first go through BE leases queue. Then it checks for the leases that can be given idle resources. It uses several time slots instead of single time slot to satisfy a lease. The leases are splitted if required, to scheduling its parts into different time slots found.

M.K.Nivodhini, Dr.K.Kousalya, Dr.S.Malliga [6] implemented a queue for immediate, advance reservation and deadline sensitive leases. The starvation of resources may occur if the lease gets suspended (i.e., in case of BE leases) for more number of times because it has lower priority and also pre-emptible. Immediate and Advance Reservation leases will get rejected if the leases are not schedulable at that particular requested time. This will increase the lease request rejection rate. For immediate, AR and DLS leases, the next free available slots are intimated to the user. Based upon the user's wish, the leases will be scheduled. This will increase the resource request acceptance rate and the system utilization rate will also increase. Thus, by implementing waiting queue for the immediate and advance-reservation lease, resource request acceptance rate and the system utilization increases.

Pratick C., Somani G. [7] showed that all the leases are static in nature. Once resources are allocated to a lease, they cannot be altered throughout the lifetime of the lease. This contradicts with the concept and implementation of a pure on-demand elastic cloud where resources of a lease are incessantly monitored against the utilization and changed based on the requirements. They introduce a new class of lease: Dynamic lease to accommodate resource changes. In dynamic lease, two sub leases i.e., increase lease and decrease lease are given to alter resources.

Heba Kurdi, Ebtesam Aloboud, Sarah Alhassan, Ebtehal T. Alotaibib[8] showed one of Haizea disadvantages that is BE leases will be preempted whenever their resources are required by AR or Immediate leases. Thus, when the system has a more number of AR or immediate lease requests, BE leases will wait for a long time, or even forever, resulting in the starvation problem. To solve this problem, some policies suggest to reject AR lease requests if BE leases have been waiting for a long time. They propose anti-starvation algorithm that uses a combination of aging and negotiation methods. An aging module counts the number of times each BE lease has been suspended. This technique stops an AR lease from suspending a BE lease as its aging counter reaches to a threshold. The priority of the BE lease also increases as its aging counter increases. A negotiating module is implemented, so that AR lease would not be rejected directly if requested resources are not available. A chance is given to the consumer, in this case, to choose between rejecting his/her AR lease and converting it to a BE lease.

Vivek Shrivastava, D. S. Bhilare [9] introduce a policy to maintain order of execution of users' tasks. The authors have proposed an algorithm that prioritize consumers based on CRI (Consumer Rating Index).By using this policy, a large number of leases can be served efficiently. Also the consumer's request caa be accepted or rejected based on the CRI score. The authors have also shown a way to distinguish consumer if their CRI score is same.

3. SUMMARY

Table 1. Summary of lease scheduling concepts & techniques

Year	Author	Approach	Future Work
2008	B.Sotomayor, R.S. Montero, I.M. Llorente, I. Foster	Introduce capacity leasing model using OpenNebula engine	Explore workloads with more combination of leases & policies to resolve conflicts in resource requirements between leases
2008	B.Sotomayor, K. Keahey, and I. Foster	Introduce design of lease management architecture	Explore policies to accept & reject leases based on criteria rather than availability of resources
2011	Nathani A., Chaudhary.S and Somani G	Applies two concepts of swapping & backfilling to utilize idle resources by finding multiple slots	Admission control policy for best-effort leases can be developed depending on the size of queue
2013	M.K.Nivodhini,Dr.K.Kousalya, Dr.S.Malliga	Implement queue for Immediate, AR & BE leases	Future work can be done in order to minimize resource request rejection rate
2013	Pratick C., Somani G.	Introduce concept of dynamic leases with auto-negotiantion	New leases can be developed to automatically alter allocated resources for better utilization
2014	Heba Kurdia, Ebtesam Aloboudb, Sarah Alhassanb, Ebtehal T. Alotaibib	Introduce policy to handle starvation of BE leases & minimize request rejection rate of AR leases	Future work can be done to handle starvation of BE leases in new ways
2014	Vivek Shrivastava, D. S. Bhilare	Introduce Consumer Rating Index and an algorithm to prioritize consumers	Future work can be done by taking CRI as primary key and amount of resources required as secondary key

4. CONCLUSION

In cloud deployment, Lease management is very important issue. Lease management can be done through Haizea lease manager. Haizea as backend with Open Nebula virtual infrastructure manager provides best scheduling of different types of leases. More combination of leases are also possible to add in Haizea as it provides facility to write own policies. Moreover, scheduling algorithms can be modify and also new concepts can be added in Haizea.

5. REFERENCES

[1] Haizea, web site. [Online]. Available:http://haizea.cs.uchicago.edu/ [accessed Sep. 8, 2014]

[2] OpenNebula Pro, OpenNebulaPro White Paper, [Online].https://support.opennebula.pro/attachments/token/coiuzlpxct7oyvq/?name=OpenNebulaPro_White_Paper_Rev20110126.pdf

[3] B. Sotomayor, R.S. Montero, I.M. Llorente, I. Foster. Capacity leasing in cloud systems using the OpenNebula engine. In: Cloud Computing and Applications; 2008, pp. 1–5

[4] B. Sotomayor, K. Keahey, and I. Foster. Combining Batch Execution and Leasing Using Virtual Machines. 17th International Symposium on High Performance Distributed Computing (HPDC'08 :), ACM, Boston Massachusetts 2008; 87:96

[5] Nathani A., Chaudhary.S and Somani G.(2011), 'Policy based resource allocation in IaaS cloud', Future Generation Computer Systems, pp.94-103

[6] M.K.Nivodhini, Dr.K.Kousalya, Dr.S.Malliga, 'Algorithms to Improve Scheduling Techniques in IaaS Cloud', IEEE-2013

[7] Pratick C., Somani G., 'Dynamic Resource Allocation using Auto-negotiation in Haizea',IEEE-2013

[8] Heba Kurdia, Ebtesam Aloboudb, Sarah Alhassanb, Ebtehal T. Alotaibib, 'An Algorithm for Handling Starvation and Resource Rejection in Public Clouds', The 9th International Conference on Future Networks and Communications (FNC-2014).

[9] Vivek Shrivastava, D. S. Bhilare, "CRI: A Novel Rating Based Leasing Policy and Algorithm for Efficient Resource Management in IaaS Clouds", (IJCSIT) Vol. 5 (3) , 2014, 4226-4230

[10] The NIST Definition of Cloud Computing, Peter Mell Timothy Grance NIST Special Publication 800-145

[11] 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

[12] B. Sotomayor, "The Haizea Manual"