

A Resource Scheduling Approach: Throughput Optimization in Private Cloud

Ms. Snehaben Patel¹ Mr. Narendra Limbad² Mr. Sazid Mahammad³

^{1,2}Department of Information Technology

^{1,2}L.J. Institute of Engineering & Technology, Ahmedabad ³ISRO, Ahmedabad

Abstract— A cloud computing is a fast growing area in computing research and industry today. It enables on-demand network access to a shared pool of configurable computing resources such as servers, storage and applications. Three main services provided by the cloud are Infrastructure as a Service (IaaS), Software as a Service (SaaS) and Platform as a Service (PaaS). Private Cloud is restricted to a particular organization and data security risks are less compared to the public cloud. It is very difficult to execute multiple jobs on multiple systems rather than on single system simultaneously. Resource Scheduling refers to the set of policies to control the order of work to be performed by a computer system. It also manages availability of CPU time, processing time and good scheduling policy gives maximum utilization of resource. Round Robin, Equally Spread Current Execution and Throttled are existing algorithms. The objective of this paper is to propose efficient and enhanced scheduling algorithm that can maintain multiple VMs dynamically based on user requirement. Proposed algorithm is applied to optimize the throughput for processing of satellite images. Multiple Jobs are executed simultaneously considering system parameters like process time and CPU time in private cloud.

Key words: Cloud computing; Open Nebula; Resource Scheduling; INSAT-3D

I. INTRODUCTION

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of computing resources. Cloud computing is a concept to provide online IT resources and IT services on demand using pay per use model [14]. Cloud computing is the Internet-based computing and this type of computing enables to shared resources, software's and information can be provided to computers and other devices on demand. With a cloud app, you just open a browser, log in, customize the app, and start using it. Amazon was the founder of the clouds and offer cloud services to public like Amazon Elastic Compute Cloud (EC2), Simple Storage Service (S3). Cloud computing is not a total new concept; it is originated from the large-scale Grid environment. Cloud computing provides all the features of grid computing, software as a service and utility computing. It utilizes the concept of virtualization. Figure 1 shows the three main types of service models and four common deployment models.

II. BACKGROUND THEORY

There are many types of satellites like Astronomy satellites, Atmospheric Studies satellites, Communications satellites, Navigation satellites, Reconnaissance satellites, Remote Sensing satellites, Search and Rescue satellites, Space Exploration satellites, Weather satellites etc. INSAT-3D is a

meteorological, data relay and satellite aided search and rescue satellite developed by the Indian Space Research Organization. The satellite has many new technology elements like star sensor, micro stepping Solar Array Drive Assembly (SADA) to reduce the spacecraft disturbances and Bus Management Unit (BMU) for control and telecom and telemetry function.

A. Resource Scheduling

Resource scheduling is the process of mapping tasks to available resources on the basis of tasks' characteristics and requirements. It is an essential aspect in efficacious working of cloud as many task parameters need to be considered for proper scheduling. Resources scheduling strategy is the key technology in cloud computing. The available resources should be utilized efficiently without affecting the service parameters of cloud.

Resource Scheduling process in cloud can be generalized into three stages.

- 1) Resource discovering and filtering – Data Centre Broker discovers the presence of resources in the network and collects status information related to them.
- 2) Resource selection - Resource is selected based on certain parameters of task and resource. At this stage, decisions are made.
- 3) Task submissions - Task is submitted to resource selected[1].

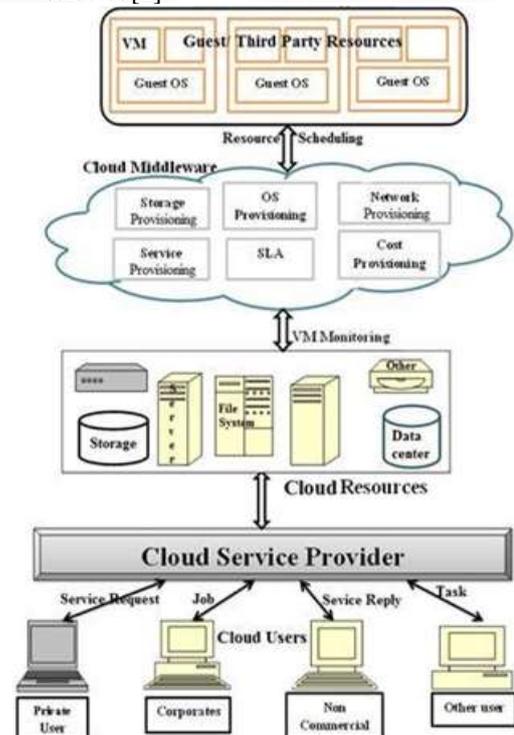


Fig. 1: Resource Scheduling In Cloud Computing Environment[2]

Figure 1 shows whole process of resource scheduling. Resource scheduling is done between cloud middleware and third party resources.

Resource monitoring is an important part of resource management under the cloud computing environment which provides a better reference for resource allocation, task scheduling and load balancing.

In Figure 2 is the ongoing monitoring process done by the service provider. This process is done to optimize the resources available. If there is underutilized computing power, the service provider will perform a load balancing exercise that will result a better utilization among the available resources [2].

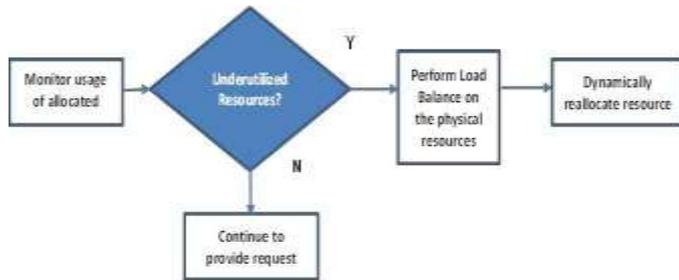


Fig. 2: Resource Monitoring With Load Balancing Approach^[2]

B. Load Balancing

Load balancing is the technology to distribute workload across multiple computers or a computer cluster, central processing units, disk drives, RAM, or other resources, to achieve optimal resource utilization, maximize throughput, minimize response time, avoid overload, and minimize application down time[3]. Figure 3 shows process of load balancing.

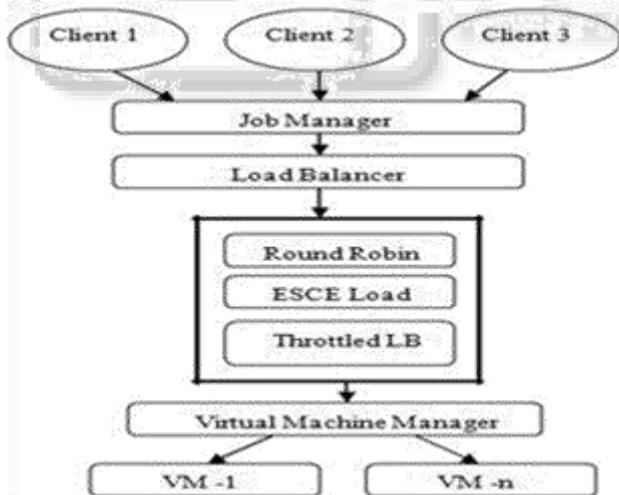


Fig. 3: Load balancing Algorithms Execution^[6]

1) Goals of Load Balancing

The goals of load balancing are as below:

- Improve the system performance
- Provide backup in case of system failure.
- Maintain system stability
- Provide future modification in system

2) Types of Load Balancing Algorithm

Depending on who initiated the process, load balancing algorithms can be of three categories as given in.

- 1) Sender Initiated: If the load balancing algorithm is initiated by the sender, it is known as Sender initiated load balancing algorithm.
- 2) Receiver Initiated: If the load balancing algorithm is initiated by the receiver, it is known as Receiver initiated load balancing algorithm.
- 3) Symmetric: It is the combination of both sender initiated and receiver initiated. If the load balancing algorithm is initiated by the receiver and sender, it is known as Symmetric load balancing algorithm.

Depending on the current state of the system, load balancing algorithms can be divided into 2 categories as given in.

- 1) Static Load Balancing: It doesn't depend on the current state of the system. Prior knowledge of the system is required.
- 2) Dynamic Load Balancing :Decisions on load balancing are based on current state of the system. No prior knowledge is required. So it is better than static approach[3].

C. Load Balancing Algorithms

1) Round Robin Algorithm

It passes each new connection request to the next server in queue, distributing connections evenly across the array of machines being load balanced. It could be more effective if the equipment that we are load balancing is roughly equal in processing speed, connection speed, and/or memory. It will not check whether the server is heavy loaded or not, it will directly assign the request whenever its turn comes so that is the reason, some server are heavy loaded while some are lightly loaded[5].

2) Dynamic Round Robin Algorithm

Dynamic Round Robin algorithm weights assignments is based on continuous monitoring of the servers and is therefore constantly changing. Distributions of connections is done on the basis of server performance analysis such as the current number of instances i.e. connection per node or the response time of a fastest node.

3) Equally Spread Current Execution (ESCE) Algorithm

A load balancer monitors the jobs which are asked for execution. The balancer looks over the queue frequently for new jobs and then allots them to the list of free virtual server. The balance also maintains the list of task allotted to virtual servers, which helps them to identify that which virtual machines are free and need to be allotted with new jobs. It works on equally spreading the load on different virtual machine[10].

4) Throttled Algorithm

Throttled algorithm is completely based on virtual machine. Here the client first requests the load balancer to check the right virtual machine which access that load easily and perform the operations which is given by the client. The client first requests the load balancer to find a suitable Virtual Machine to perform the required operation[7][9].

Algorithm Name	Parameters Used	Type Of Algorithm	Merits	Demerits
Round Robin ^[12]	Completion time	Static	Every process get equal weight age so no process will go under starvation.	Most of the time processor remains idle
Equally Spread Current Execution ^[13]	Response time,	Dynamic	Maximize throughput	Less priorities process need to wait too long.
Throttled Load Balancer ^[7]	Communication cost, Load Movement Factor	Dynamic	High Load Movement Factor	High communication

Table 1: Comparison of Load Balancing Algorithms

III. PROPOSED ALGORITHM

There are various algorithms like Round robin algorithm, Equally Spread Current Execution (ESCE) algorithm, throttled algorithm for resource scheduling. But various issues in the different algorithm are still some research to be done to optimize throughput and improve efficiency of the algorithms.

In ESCE technique load balancer makes attempt to preserve equal load to all the virtual machines connected with the data center. In this technique load balancer gets number of virtual machines by maintaining an index table and queue of number of requests currently assigned to the virtual machine. The balancer continuously scans the queue and the index list of virtual machines. If there is a VM available that can handle request of the client, the VM is allocated to that request. If on the other hand there is a VM that is free and there is another VM that needs to be freed of the load, then the balancer distributes some of the tasks of that VM to the free one so as to reduce the overhead of the former VM. Similarly Throttled algorithm maintains an index list of all virtual machines as well as their states (Available or Busy). Whenever a request comes to the data center the load balancer scans the index list from top until the first available VM is found or the index list is scanned fully. If a match is found on the basis of size and availability of the VM, then the load balancer accepts the request of client and allocates that VM to the client. Though there is no VM available that matches the criteria then the load balancer returns -1 and the request is queued.

Proposed algorithm maintains index list of VM allocation status as well as list to count the allocated request. First of all allocated request list is compared with the index list of VMs. If VMs index list if greater than allocated request list it means that VMs are available to take request else request has been queued until VM is been available. Proposed algorithm optimize throughput with maximum resource utilization in private cloud using different parameters like process time and CPU time. Figure 4 shows flowchart of proposed algorithm.

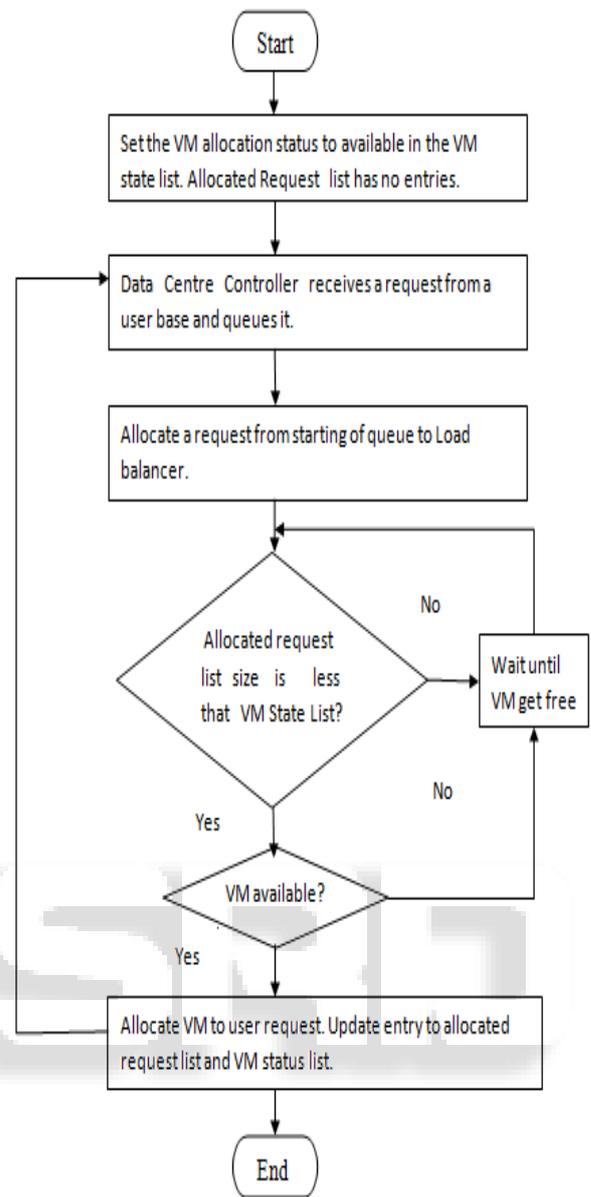


Fig. 4: Flowchart of Proposed System

A. Steps of Proposed Algorithm

Steps of proposed algorithm are as below.

- 1) Step 1: Initialize all of the VMs allocation status to AVAILABLE in the VM state list.
- 2) Step 2: Initialize the allocated request with no entries.
- 3) Step 3: Data Center Controller receives a new request.
- 4) Step 4: Data Center Controller queries new Load Balancer for next allocation.
- 5) Step 5: If allocated request list size < VM state list size, Then Allocate the VM. Else wait for the VM to get free.
- 6) Step 6: When the VM finish processing the request and the Data Center Controller receives the cloudlet response, it notices the load balancer of the VM deallocation.
- 7) Step 7: The load balancer update the status of VM in VMs state list and allocated request list.

IV. SET UP CLOUD ENVIRONMENT AND RESULTS

There are many cloud computing platforms like open nebula, open stack, nimbus, Amazon EC2, eucalyptus.

A. Open Nebula Architecture

Open Nebula is a cloud computing platform for managing heterogeneous distributed data center infrastructures. The Open Nebula platform manages a data center's virtual infrastructure to build private, public and hybrid implementations of infrastructure as a service. Figure 5 shows the architecture of Open Nebula cloud environment. It is open source.

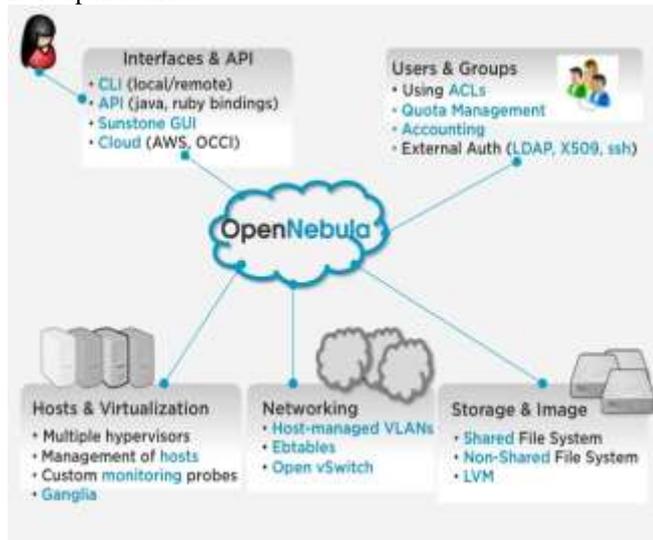


Fig. 5: Open Nebula Architecture

Main components of Open Nebula are as under:

- Interfaces & APIs: Interfaces to manage physical and virtual resources.
- User and Groups: Supports accounts, groups, authentication, and authorized mechanisms. Access Control List for fine grain permission granting.
- Hosts and Virtualization: supports Xen, KVM, and VMware.
- Networking: supports for VLANs and Open vSwitch.
- Storage and Images: supports both non-shared and shared file systems.

B. Results

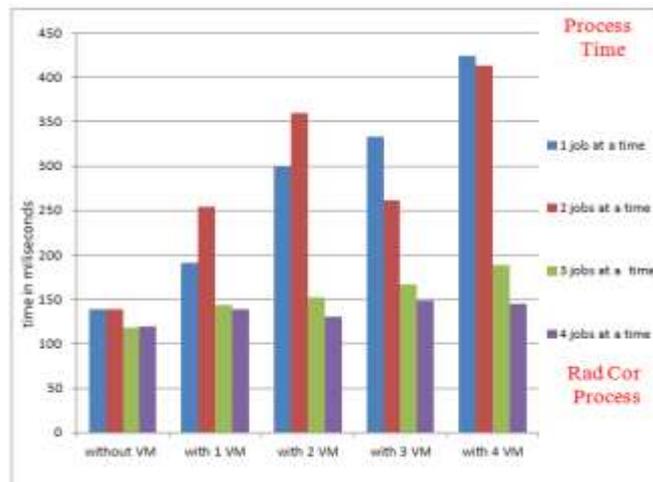


Fig. 6: Process Time of Rad cor Process

We have performed proposed algorithm on INSAT 3D image processing. Each image contains 1 GB size. Image has .raw extension. Rad Cor and ServoCor are processes of imager in INSAT 3D satellite. Minimum process time gives better throughput and performance of the system and minimum resource utilization.

Figure 6 shows process time in process of image processing enabling different types of configuration like single system, with 1 VM, with 2 VMs, with 3 VMs and with 4 VMs with multiple jobs simultaneously. Process time is more in 1 job & 2 jobs at a time in all cases. But process time is less in 4 jobs at a time in all cases.

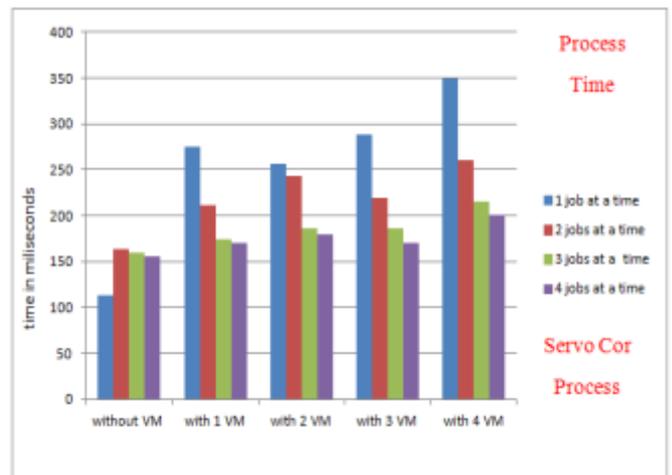


Fig. 7: Process Time of Servo Cor Process

Figure 7 shows process time in ServoCor process of image processing in all cases same as radcor process. Process time is more in 1 job & 2 jobs at a time in all different cases. Process time is decreased in 4 jobs at a time compared to others in all configuration cases.

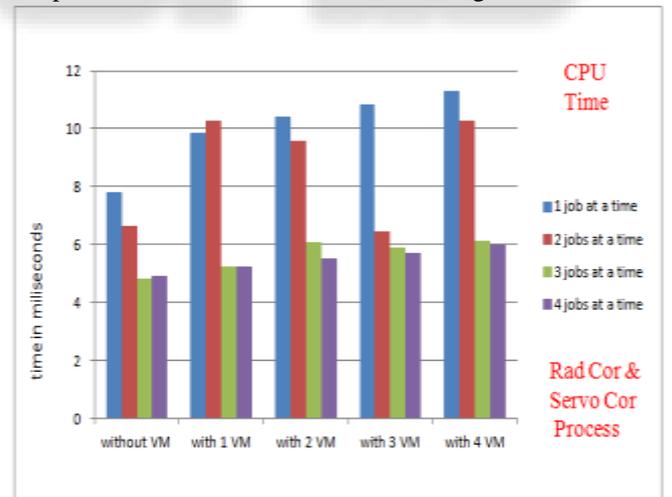


Fig. 8: CPU Time Of Rad Cor & Servo Cor Process

Figure 8 shows CPU time of radcor and servocor process of image processing with multiple jobs as described in figure 6. In all cases shown in figure, CPU time is more in 1 job & 2 jobs at a time. In all other cases CPU time of 3 jobs at a time is more than 4 jobs at a time.

V. CONCLUSION

Scheduling is one of the most important tasks in cloud computing environment. Scheduling manages availability of

CPU memory, processing time and good scheduling policy gives maximum utilization of resource. Analysis of various scheduling algorithm which efficiently schedules the computational tasks in cloud environment is given. Proposed algorithm is implemented and performed on INSAT-3D image processing with execution of multiple jobs simultaneously in Radcor and Servocor process with different cases like in stand-alone system, with 1 VM, 2 VM, 3 VM and 4 VM in real private cloud environment(Open Nebula). Jobs are performed on various parameters like process time and CPU time. Image size is 1 GB each. Process and CPU time is decreased at execution of multiple jobs simultaneously in with 3 VM and 4 VM.

REFERENCES

- [1] Dr Ajay Jangra and Tushar Saini, “ Scheduling Optimization In Cloud Computing”, International Journal Of Advanced Research In Computer Science and Software Engineering, April 2013, ISSN: 2277-128X,Page(s): 62-65
- [2] Mohd Hairy Mohamaddiah, Azizol Abdullah, Shamala Subramaniam, and Masnida Hussin , “A Survey On Resource Allocation And Monitoring In Cloud Computing” International Journal Of Machine Learning and Computing, 2014 ,ISSN:2010-3700,Page(s):31-38
- [3] Amandeep Kaur Sidhu, Supriya Kinger, “A Sophisticate Approach For Job Scheduling In Cloud Server”, International Journal Of Computer Trends And Technology, July,2013,ISSN:2231-2803,Page(s): 2055-2058
- [4] Peter Sempolinski, Douglas Thain, “A Comparison And Critique Of Eucalyptus, Open Nebula And Nimbus”, Cloud Computing Technology And Science (Cloud Com),DOI 10.1109/ Cloud Com. 2010, Page(s):417-426
- [5] Ms.NITIKA, “Comparative Analysis Of Load Balancing Algorithms In Cloud Computing ”, International Journal Of Engineering And Science, 2011,ISSN:2319-1813,Page(s): 34-38
- [6] Dr Hemant Mahalle,Parag R Kaveri, Dr. Vinay Chavan“ Load Balancing On Cloud Data Centers”, International Journal Of Advanced Research In Computer Science & Software Engineering, January,2013,ISSN:2277-128X Page(s):1-4
- [7] Shridhar G.Domanal and G.Ram Mohana Reddy “Load Balancing In Cloud Computing Using Modified Throttled Algorithm”, IEEE International Conference on Cloud Computing in Emerging Markets, Oct, 2013,Page(s): 1-5
- [8] Hamid Shoja ,Hossein Nahid, Reza Azizi “ A Comparative Survey On Load Balancing Algorithms In Cloud Computing”, International Conference On Computing, Communication And Networking Technologies, July,2014 , Page(s):1-5
- [9] Veerawali Behal, Anil Kumar, “Cloud Computing: Performance Analysis Of Load Balancing Algorithms In Cloud Heterogeneous Environment”, 5th International Conference Confluence The Next Generation Information Technology Summit, Sept,2014, Page(s): 200-205
- [10]Subasish Mohapatra, K.Smruiti Rekha, Subhadarshini Mohanty, “ A Comparison Of Four Popular Heuristics For Load Balancing Of Virtual Machines In Cloud Computing”, International Journal Of Computer Applications, April 2013, ISSN:0975-8887,Page(s):33-38
- [11]George Chang, Shan Malhotra, Paul Wolgast, “Leveraging The Cloud For Robust And Efficient Lunar Image Processing”,IEEE, 2011, Page(s):1-8
- [12]Pooja Samal,Pranati Mishra ,“Analysis Of Variants In Round Robin Algorithms For Load Balancing In Cloud Computing” International Journal Of Computer Science And Information Technologies, 2013,ISSN:0975-9646,Page(s): 416-419
- [13]Akshay Daryapurkar, Mrs. V.M. Deshmukh, “Efficient Load Balancing Algorithm n Cloud Environment ” International Journal Of Computer Science And Applications,2013, ISSN:0974-1011,Page(s):308-312
- [14]<http://www.nist.gov/itl/csd/cloud-102511.cfm>
- [15]<http://opennebula.org>