Performance Analysis of Resource Provisioning for Cloud Computing Frameworks

Sanjeev Kumar Pippal, Rahul Kumar Dubey, Arjun Malik, Nishant Singh, Prateek Jain, Rajan Kumar Bharti and Dharmender Singh Kushwaha

1-7 MNNIT Allahabad, Allahabad, India

Email: dsk@mnnit.ac.in

Abstract—Resource provisioning is key to good performance of any cloud based system. The performance of a cloud computing framework depends a lot on how the dynamic and elastic resource provisioning is. Ad hoc cloud enable existing infrastructure as cloud compliant and the available resources in the system are utilized non-intrusively. In ad hoc clouds where resource is scarce and only voluntarily available, it becomes extremely significant the way resource provisioning is performed. In this work we have analyzed provisioning through virtual machines v/s provisioning through physical machines and effect of different parameters on data transfer between virtual machines (VMs). We also conclude with study of various scenarios in which to use virtualization techniques for ad hoc cloud.

Index Terms—Resource provisioning, Cloud, Ad hoc Cloud, Virtualization.

I. INTRODUCTION

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. Essential characteristics of cloud computing are on demand self-service, broad network access, resource pooling, rapid elasticity, measured services. Three service models are Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS).

An Ad hoc cloud [1,2] as shown in Fig. 1 derives data and cloud services from fixed cloud as well from the host cloud infrastructure, the connectivity to them is provided using an ad hoc link (V-SAT). The S, P and V nodes in the ad hoc data center have the following descriptors:

- **S node**: Super-node which is permanent node at remote location with ad hoc connectivity with the fixed cloud to facilitate cloud formation at remote site
- **P nodes**: Persistent-node which are organization’s hosting cloud and data services.
- **V nodes**: Volunteer-nodes which are other participating nodes within an organization.

The S nodes promote the stake holders to establish their own collaborative dispersed data center. The P nodes within a data center provide reliability and availability through replication of services and data. The V nodes

© Elsevier, 2014
can voluntarily cache data and provide availability and performance in the absence of persistent node and large number of requests. The nodes participating in data center can be heterogeneous in terms of computing resources, database technology. All nodes participating in the data center are logically hierarchically organized and communication between them is encrypted with key shared and provided by hierarchically common parent node.

![An Ad hoc cloud scenario](image)

Resource Provisioning: It is important and plays a vital role in the performance of a cloud computing framework. Cloud provisioning is the allocation of cloud provider’s resources to the customers. When a cloud provider accepts a request from a customer, it must create appropriate number of Virtual Machines (VMs) and allocate resources to support them. There are three types of provisioning namely:

Advance Provisioning: In this customer contracts the provider for services and the provider prepares the appropriate resources in advance of start of service.

Dynamic provisioning: In this provider allocates more resources as they are needed and removes them when they are not. The customer is billed on a pay-per-use basis.

User-self Provisioning or cloud-self-service: In this customer purchases resource from the cloud provider through a web form, creating a customer account and paying for resources with online payment.

Virtualization: The term virtualization broadly describes the separation of a resource or request for a service from the underlying physical delivery of that service. Virtualization techniques can be applied to IT infrastructure layers-including networks, storage, laptop or server hardware, operating system and applications. It provides a layer of abstraction between computing, storage and networking hardware, and the applications running on it. Virtualization enables today’s x86 computer to run multiple operating systems and applications, making our infrastructure simple and more efficient and seamlessly integrate-able. Applications get deployed faster, performance and availability increase and operations become automated.
VMware Virtualizations: VMware’s virtual machine approach creates a uniform hardware image implemented in software on which operating systems and applications run. VMware is designed from the ground up to ensure compatibility with customer’s existing software infrastructure investments. This includes not just operating systems, but also software for management, high availability, clustering, and replication and so on.

II. RELATED WORK

This section presents the related work done. In [3] the authors have proposed resource allocation strategies that are relevant to automatic resource management. They have provided valuable insights into the performance of alternative resource allocation strategies and job scheduling disciplines for a cloud computing infrastructure. The service level agreement is based on response time distribution, which is more relevant than the mean -response time with respect to the performance requirement of interactive applications. It discusses an efficient and effective algorithm to determine the allocation strategy that result in smallest number of server required. The paper focuses only on the case of two job classes and does not consider complex scenarios. Also the paper considers the processing of interactive jobs only. In [4] the author proposes challenges of design of cloud computing and identify important research directions in this increasingly important area. The paper covers its essential concepts, architectural design, prominent characteristics, key technologies as well as research directions. This paper also discusses key challenges in cloud computing like automatic resource provisioning, power management and security management. In [5] the author discusses about the Virtual machine provisioning for a computer system. This paper gives detailed methodologies of how provision through VMs is done. It also discussed the idea of remote provisioning. It also discussed about the Computer generated VM provisioning requests as well as about how to provision a network of computers. It also discussed about implementing a method for doing business by charging users a fee for each VM provisioned and/or deployed, based on number and type of components included in a deployed VM, as a function of time of actual or available use, or any combination of these or other possible fee structures.

In [6] the authors have identified several challenges from the Cloud computing adoption perspective. Also, it highlighted the cloud interoperability issue that deserves substantial further research and development. The author articulated the relationships amongst Cloud computing, Service-Oriented Computing, and Grid computing. It analyzed a few challenges on the way towards adopting Cloud computing. The interoperability issue was highlighted and a number of solutions are discussed thereafter for different cloud service deployment models. In [7] the authors have provided a practical solution to the problem of driving dynamic resource provisioning in shared centers in face of the complication of applications both in scale and architecture by laying out a theoretical framework. In this work authors have analyzed the effectiveness of the regression method with different number of classes. The results for the performance evaluation show that performance indexes obtained can be effectively used as the basis for autonomic resource provisioning. In [8] the authors propose a middleware framework that avoids a priori capacity analysis and allows such private cloud owners to provision resources automatically such that a specified grade of service is maintained. The proposed framework performs dynamic resource provisioning that also leads to a reduction of operational cost. Additional resources are acquired during high traffic periods and released during low traffic periods such that the desired grade of service is always maintained. The paper describes the architecture of the framework and the experience gained from a prototype implementation including a preliminary analysis of its performance. In [9] the author demonstrates that use of a virtualization layer between the service and the physical infrastructure extends the advantages of VM platforms to distributed infrastructures and introduces new capabilities as well. The paper discusses the use of two different cluster-based services (a computing cluster and a web server cluster) on top of a virtualized infrastructure, with the capacity of integrating external cloud resources and takes into account communication overheads. On the basis of performance results it very well establishes the advantages of introducing a virtualization layer. In [10] the authors have studied the components in Map-Reduce processing and build a cost function that explicitly models the relationship between the amount of data, the available system resources (Map and Reduce slots), and the complexity of the Reduce function for the target Map-Reduce program. This paper solves the problem of optimizing resource provisioning to minimize the financial cost for a specific job.

In [11] the author states how we securely bridged the semantic gap into the operating system semantics. The presented solution enables the novel security services for fast changing cloud environments where customers run a variety of guest operating systems, which need to be monitored closely and quarantined promptly in
case of compromise. This paper does not discuss about agent injection from security VMs into guest VMs through the introspection interface. In [12] the authors highlight the limitations existing in developing advanced resource provisioning and scheduling mechanisms. This work investigates how the precision of availability information affects resource provisioning in multiple site environments. Performance evaluation is conducted considering both multiple scheduling policies in resource providers and multiple provisioning policies in brokers, while varying the precision of availability information. They also demonstrate that multiple resource partition policies improve the slowdown of Grid jobs. Resource provisioning: Singh et al. [13, 14] present a provisioning model where Grid sites provide information on the time slots over which sets of resources are available. The sites offer their resources to the Grid in return for payments, thus they present a cost structure consisting of fixed and variable costs over the resources provided. The main goal is to find a subset of the aggregated resource availability, termed as resource plan, such that both allocation costs and application make-span are minimized. Our work is different in the sense that we investigate multiple approaches to obtain availability information and how reliable this information can be in multiple site environments.

III. PROPOSED WORK

In proposed work we have analyzed provisioning resources like CPU, Memory through virtual machines v/s provisioning through physical machines and evaluated the effect of different parameters on data transfer between virtual machines (VMs). The following steps are used to do the analysis:
1. Communications between virtual machines and transfer of data between VMs through STDIN and STDOUT.
2. Creation of super-table (student’s record) by exchange of data between VMs.
3. Analysis and comparisons of data transfer time in following scenarios:
   4. Between two Virtual Machines (VMs) on same host machine.
   5. Between two physical machines.
   6. Between two Virtual Machines on different hosts.
4. Comparisons have been done by plotting graph for all three above scenarios.
5. Analysis of effect of RAM size of virtual machines on data transfer time between VMs by varying the percentage of RAM allotted to Virtual machines.
6. Analysis of performance of Virtual Machines (VMs) by varying the RAM size of one virtual machine and keeping the RAM size of other VM constant.
7. Analysis of how percentage RAM allocation to VMs affects the performance of host machines.

The Test-bed Setup includes Processor: Intel(R) Core™ i5 CPU M480 @2.67GHz, Installed Memory (RAM): 4.00 GB, System type: 64-bit operating system, Host Operating system: Windows 7 Home Premium. Virtual machine: VMware Workstation v7 and Guest operating system: Ubuntu 11.10.

IV. RESULTS

In Table 1 Comparisons of Data transfer rate between physical machines and virtual machines is shown. The below results as shown in Fig. 2 shows that the time of data transfer between two machines either virtual or physical increases linearly with the increase in size of data being transferred. It is also observed that the minimum data transfer time was taken in case of two VMs on same host.

In Table 2 The effect of RAM size on data transfer rate between two VMs where data size transferred =3kb and available RAM is 4 GB.

<table>
<thead>
<tr>
<th>S. No</th>
<th>Data Size (Kbs)</th>
<th>Time (in Sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Two VMs on same Host(25% RAM size)</td>
<td>Two Physical Machines</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.024</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.028</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.036</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.040</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.048</td>
</tr>
</tbody>
</table>
Fig 2. Data transfer Vs Data transfer in case of two VM on same host, two VM on different host and two physical machine is plotted

<table>
<thead>
<tr>
<th>S. No</th>
<th>VM1 (in %)</th>
<th>VM2 (in %)</th>
<th>Time (in sec)</th>
<th>Host Machine (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.25</td>
<td>6.25</td>
<td>0.024</td>
<td>87.5</td>
</tr>
<tr>
<td>2</td>
<td>12.5</td>
<td>12.5</td>
<td>0.020</td>
<td>75.0</td>
</tr>
<tr>
<td>3</td>
<td>25.0</td>
<td>25.0</td>
<td>0.016</td>
<td>50.0</td>
</tr>
<tr>
<td>4</td>
<td>37.5</td>
<td>37.5</td>
<td>0.028</td>
<td>25.0</td>
</tr>
<tr>
<td>5</td>
<td>50.0</td>
<td>50.0</td>
<td>0.040</td>
<td>00.0</td>
</tr>
</tbody>
</table>

Figure 3. Time of Data transfer vs RAM size%

<table>
<thead>
<tr>
<th>S. No</th>
<th>VM1 (in %)</th>
<th>Vm2 (in %)</th>
<th>Time (sec)</th>
<th>Host Machine (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.5</td>
<td>12.5</td>
<td>6.641</td>
<td>75.0</td>
</tr>
<tr>
<td>2</td>
<td>12.5</td>
<td>25.0</td>
<td>6.542</td>
<td>62.5</td>
</tr>
<tr>
<td>3</td>
<td>12.5</td>
<td>37.5</td>
<td>6.306</td>
<td>50.0</td>
</tr>
<tr>
<td>4</td>
<td>12.5</td>
<td>50.0</td>
<td>6.157</td>
<td>37.5</td>
</tr>
<tr>
<td>5</td>
<td>12.5</td>
<td>62.5</td>
<td>6.408</td>
<td>25.0</td>
</tr>
<tr>
<td>6</td>
<td>12.5</td>
<td>75.0</td>
<td>6.592</td>
<td>12.5</td>
</tr>
</tbody>
</table>
In figure 3 above graph shows that the best performance for a virtual machine can be achieved if the total allocated RAM to virtual machine is 50% of total available RAM size.

In Table 3 we have analyzed the effect of RAM size on process execution time for VM on a host with another constantly running VM having 12.5% fixed RAM for Matrix multiplication of two matrices of dimension 1000 by 1000.

As shown in figure 4. The Process execution time decreases with the increase in RAM size for a virtual machine (VM2) up to certain percentage of RAM allocated but it increases after that. In our case the best performance was observed when the total RAM allocated to VMs is 62.5 % (VM1: 12.5 % and VM2:50 %) of total RAM.

V. CONCLUSIONS

The performance of a cloud computing framework depends a lot on how the dynamic and elastic resource provisioning is. In this work we have analyzed provisioning through virtual machines v/s provisioning through physical machines and effect of different parameters on data transfer between virtual machines (VMs) and found that the time of data transfer between two machines either virtual or physical increases linearly with the increase in size of data being transferred. It is also observed that the minimum data transfer time was taken in case of two VMs on same host and best performance for a virtual machine can be achieved if the total allocated RAM to virtual machine is 50% of total available RAM size also Process execution time decreases with the increase in RAM size for a virtual machine (VM2) up to certain percentage of RAM allocated but it increases after that. Further these empirical observations can be incorporated into resource provisioning mechanism of ad hoc and public clouds where efficient resource provisioning of cloud resources are very critical.

REFERENCES


