2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)

A Novel Approach to Address Interoperability Concern in Cloud Computing

Arunkumar.G and Neelanarayanan Venkataraman.

School of Computing Science and Engineering, VIT University, Chennai.India-600127

Abstract

For the reenactment of internet and cloud technologies, interoperability is an important touchstone. Most of the consumers are analyzing the appropriateness of cloud to employ themselves for their enterprise or personalized operations. Customers are self-satisfied at the inception, but expectation changes. Based on their business escalation it needs further adoption of modern cloud services the existing cloud provider fails to offer. Hence the user needs interoperability and portability to ship their assets from one cloud to other cloud. The complication faced by the customers in shifting their assets remains as a challenge to be addressed. In this paper we outline the existing cloud technologies, interoperability issues and possible solution to overcome the problems.

Keywords: Cloud Computing; Portability; Interoperability; Migration; Risks.

1. Introduction

Cloud Computing is a prototype that give users unlimited computing power that can be accessed from anywhere based on users convenience. User can simply include or free the resources like CPU, Memory, Storage and Bandwidth with request to the cloud provider. User can also make use of prefabricated services like Amazon Cognito, Elastic Beanstalk provided by the cloud provider to amplify smooth working of an application with less labor pool. Cloud Computing has four deployment models (public, private, community and hybrid) and three service models (SaaS, PaaS, IaaS) standardized by NIST [1] in September 2011. Public Cloud is common to all users. Users need to pay as per the Service Level Agreement (SLA) to continue using the cloud services. SLA is an electronic document that specifies how the cloud provider charges the user based on the selected configuration. The typical SLA also consists of maximum uptime and possible downtime. Amazon [2] has high uptime average of 99.95%. If the cloud provider failed to meet the SLA, the cloud provider pays credits to the user. The user can request

computing resources through the web interface or using APIs. Private Clouds like Eucalyptus [14], Openstack [15] and Cloud Foundry [16] are local implementation of cloud platform. It is specific to an organization, maintained and managed by the internal staffs. The primary choice of private cloud is due to its security and in-house management. Private cloud also yield cost reduction by consolidating the unused infrastructure and it enhances service quality. When the workload is distributed across public and private cloud, the blending is called as hybrid cloud. To emerge from increasing workloads in internal infrastructure, hybrid clouds are used. When private cloud infrastructures are outpaced with the increasing workloads, a part of loads are shifted to public cloud for processing. Community cloud is a model composed by a group of people to solve a problem for a common cause. The unused computing resource is donated between organizations.

Software as a Service (SaaS) is model where software is implemented using any of the four deployment models. The user's don't have access to data or the configuration and can only use the software hosted in the cloud. If the software is hosted in public cloud then it is termed as Public SaaS. Gmail is an example for Public SaaS. Platform as a Service (PaaS) like GAE [12] and Salesforce [13] provides the user with development platforms like java, python, php or custom platforms to build an application. Infrastructure as a Service (IaaS) provides the user with bare machine where the user can configure the operating system. Cloud is layered architecture consist of hardware, hypervisor, programming platform, application and cloud management interface. Virtualization can be achieved using the software component called hypervisor. Virtualization technology is used to reduce resource underutilization. Virtualization provides the rich feature like resource partitioning where multiple heterogeneous operating systems like Windows, Linux and Mac executes concurrently in same hardware, Migration and consolidation when system is under or over utilized, it is possible to move an executing virtual machine to another physical server without termination called as live migration. If an operating system is formatted to execute on top of the hypervisor it is called as virtual machine or VM. The running state of virtual machine is called as instance. In the virtual machine user can implement their business logic and can be accessed through the IP address. The organization of this paper as follows: Section 2 explains the need of virtualization in cloud computing. Section 3 proposes a taxonomy of cloud infrastructure tasks. Section 4 explains the cloud interoperability and portability issues from both customer and cloud administrator perspective, followed by conclusion and future work.

2. Virtualization

Virtualization is a paramount need for cloud computing. Virtualization can be achieved by hypervisor, and it is mainly of two types, Type 1 and Type 2 hypervisor. Type 1 hypervisor is also called as bare metal hypervisor. The virtualization technologies applied in the hypervisors are three types namely Full virtualization, Para Virtualization and Hardware Assisted Virtualization [3]. All the commercial virtualization products like KVM [17], xen [18], virtualbox [19] and VMware works in x86 architecture. The x86 architecture [4] has four rings where the ring 0 is responsible for executing privileged instruction, ring 1 and ring 2 are unused and in ring 3, the application programs executes. The operating system is always executed in highest privilege level (ring 0). The physical hardware called as host and virtual machine executes on top of the hypervisor and is called guest.

In full virtualization the host operating system executes on ring 0 where the guest operating system executes on any of the above rings depends on hypervisor. User application executes in ring 3. When the guest operating system need to execute privileged instructions the control passes to host operation system and the host executes in the hardware. The host operating system acts as broker between guest and the hardware. This is called as binary translation. In full virtualization the hypervisor has the ability to execute the guest operating system in less privileged levels.

Para virtualization or OS Assisted Virtualization the guest operating is system tricked by altering the kernel of the guest operating system to think that it is actually executing in native hardware. Guest operating system is altered in para virtualization. In Para virtualization the non virtualizable instructions of guest operating system are responded by the hypervisor itself.

Hardware Assisted Virtualization, the aid provided by the CPU hardware to run the hypervisor below ring 0. The guest operating system executes in ring 0 in native speed like traditional operating system. The hypervisor provide traps where the non virtualizable instruction always executes in hardware. The need of binary translation and Para virtualization is eliminated in hardware assisted virtualization.

Virtualization eliminated the need of hardware dependency. Operating system executes on top of the virtualization layer. Virtualization provides the user with resource partitioning. The same hardware can shared for the execution multiple operating system concurrently. Virtualization enables isolation of virtual machines. The execution of an operating system won't affect the execution of other VMs executing in the same hardware. Virtualization also improves resource utilization by migration and consolidation of VMs. When an infrastructure is over utilized the VMs can be migrated to other hardware without downtime is called as live migration. When an infrastructure is underutilized the VM is migrated to some other server to optimize the resource utilization. Virtualization provides encapsulation, the VM can be stored in a single file like .ovf, .vmdk or .img and portable across multiple cloud supports same virtualization technology. Based on a VMware-sponsored study [5] on 1000 employees it is reported that 72% in improved server utilization, 52% reduction in number of running servers, 48% in improved security, 42% in boosting availability and uptime, 40% in improved application and server management, 40% in improved data backup by adopting virtualization for the organization.

3. Infrastructure Cloud Tasks

All the infrastructure-cloud has compute, storage and network management in their environment. For these common tasks, single generalized API like JCloud [21], LibCloud [20], DeltaCloud [22] and Fog [23] etc. can be used. This API failed to cover cloud specific offerings. Cloud infrastructure tasks taxonomy is shown in figure 1.
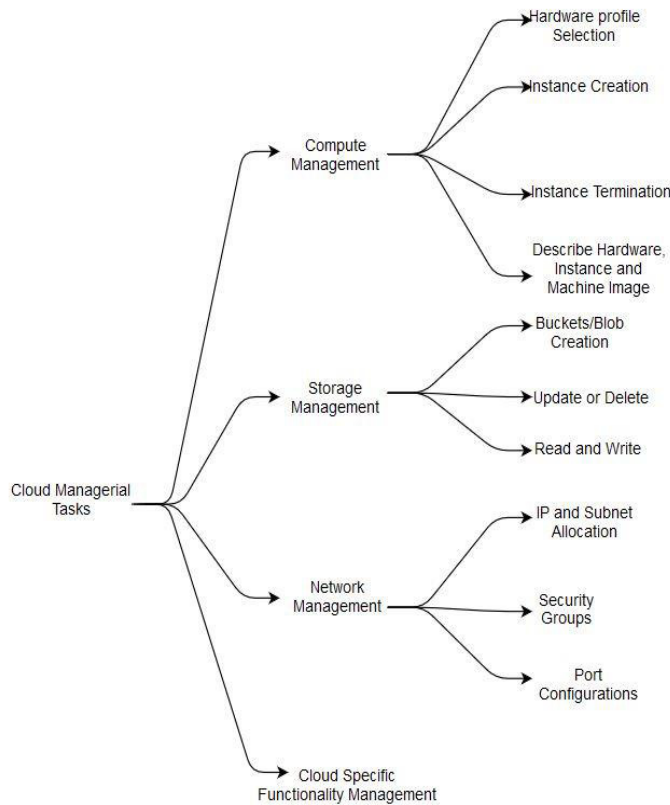

Fig. 1. Cloud Infrastructure Tasks

Compute management can be divided into Hardware profile selection, Instance creation, Instance termination and describe the resources. Hardware profile selection is the selection of computing resources based on the user needs. User can select their own requirements like cores of CPU, memory, network bandwidth and storage. This is also called as resource pooling. Inheriting required processing potential form the pool of resources. Instance creation is a process of selecting the operating system to run on the hardware profile. The operating system is preconfigured and ready to execute in the hardware. For the seamless execution, the operating system should match the hardware profile and underlying virtualization technology. Describing the resources will show the number of executing instances, hardware profile, type of operating system executing in the hardware and IP address associated with the instance. The basic storage tasks are bucket or blob creation, update or deletion of data, read and write data. The bucket or blob creation is process of engrave required storage capacity form the providers storage pool. The storage should be mounted to the running instance for user operation. This bucket or blob provides persistent storage.

The basic network management tasks are IP and subnet allocation, Security groups and port configurations. The user can configure the IP address of the running instances. The IP can be either dynamic or static. If the user likes to construct a cluster in cloud environment, all the virtual pc should be in same subnet for secured and efficient internal communication. Port configuration enables security to the running instance. The user can block all the unused port to ensure security to the running instances. User can associate the port configuration to the security group. All instance linked to the specific security group follows same port configuration. Some cloud providers gives the user with prefabricated services like Hadoop as a service [25] and database as a service. That functionality can only be accessed by the cloud vendor specific application programming interface. It is hard to construct a generalized API for such needs.

4. Cloud Interoperability and Portability Issues

The ability to make the components to work in independent cloud with minimal or null user effort is called interoperability [6]. The federation of multiple cloud is called as Sky Computing [8].Cloud has higher adoption risks summed upto 84%. This adoption risks includes security, lock-in, control, legal and service. Paul Bannerman [10] states that interoperability is an aspect of lock-in. Reducing issues in interoperability also reduces lock-in which has the risk rate of 22.9%. Approaches to solve interoperability issues [7] are integrated approach: where all the cloud providers are agreed to follow same standard, Unified approach: where the cloud provider has the equivalence model to map other standards, federated approach: agreements between the cloud vendors to share the working semantics. Porting user asserts to different cloud is a burdensome challenge. Interoperability and portability can be address in storage, network, application, platform, virtual appliance and management. In cloud computing environment, interoperability issues arise in various levels.

4.1. Application Interoperability

The traditional application development is different form cloud application development in various aspects. In traditional application development the infrastructure is physically present where as in cloud computing it is provided as a service to the user. To reduce user hardship most cloud providers offer the user various specialized cloud service like Amazons elastic cache [11], cloud front and cloud formation. In order to take the advantage of those provider offerings the application must be designed using specific cloud provider API. Most IaaS providers offer storage, compute and network services in common. To address the interoperability need between compute, storage and network various independent API are ready to use but these APIs failed to cover cloud specific offerings. Application reengineering is needed to port an application to various clouds. The application reengineering scenarios are given below

- To make use of cloud specific services, application needs to be reengineered when it is ported to cloud environment form a traditional infrastructure.
- Same services cannot be offered by all the cloud providers. When porting an application from one cloud to different cloud, application needs reengineering.

- When porting application form IaaS cloud provider to PaaS cloud provider, application needs reengineering.

4.2. Platform Interoperability

Platform interoperability can be achieved by two methods they are platform source portability and machine image portability [9]. Platform source portability is applicable when a user implemented the business logic in the platform which is custom made and coupled with the hardware. To port an application, due to coupling and customization, the platform should also port in other cloud. Platform source portability, overcomes such difficulty. The platform is ported and compiled to reform the bonding between the hardware and the application implemented.

Machine Image is an assortment of operating system, platform, application and configuration. This machine image is portable across multiple infrastructure clouds following same virtualization technology. This virtual appliance can be formatted into various cloud specifications is called machine image portability. KVM and Xen are hypervisors uses .img format, VMware uses .vmdk and .ovf image formats. .Factors to consider for platform interoperability as a cloud administrator are

- Choose a generic programming platform which has less blend with the hardware.
- Should consider the hardware profile matches the programming platform implementation because in cloud everything is a provided as a service, the actual hardware is behind virtualization layer.
- Consider the compatibility and applicability for porting a virtual appliance or machine image to independent cloud.

4.3. Storage Interoperability

The machine images or the virtual appliance are not responsible for storing the user data. The user data are saved in storage device called buckets and they are mounted to the running instance for memory. From the provider perspective in storage to employ protection and performance all the cloud providers are applying RAID technologies in their storage infrastructure. To improve resource utilization and reduce running capital, the same physical hard disk is partitioned using RAID levels and shared with multiple users. When the logical partition has reached the threshold, a new partition is added from new physical hard disk. This in turns extends the storage capacity of the running instance.

Further distribution of data happens inside those partitions called tearing. Hence it is really hard to find where the data actually resides. User has no control over the data at storage level. User should use the front end or API to access or store the data. In very minor cases, where the data is uploaded from single point uses data replication and local mirroring to reduce the effect of vendor lock-in. It is hard to incorporate mirroring when data is transferred to cloud from multiple source point. Storage Networking Industry Association (SNIA) proposed a unified solution called Cloud Storage Management Interface (CMDA) which is capable of self-provisioning, administering and accessing data elements in multiple cloud using Restful web services. The current version is CDMI 1.1 published in August 2014. CDMI [24] supports Transport security TLS, Access control, Integrity and Security capabilities.

- To diminish the uncertainty about data fatality, consider having an in-house backup solution with active sync.
- Should consider the change in data semantics when porting to heterogeneous cloud with different storage architecture.
- Should consider the bulk data movement between clouds and associated risks like transformation, integrity, metadata handling, data loss and corruption.

4.4. Management and Configuration Interoperability

All cloud providers aim to provide the user with increased uptime of their virtual servers. To attain such uptime, the physical system should be managed and configured to withstand any unanticipated failure. For the cloud provider the changes are concerned with the optimization of internal infrastructure to obtain maximum possible uptime. The infrastructure includes the physical server, Networking components, virtualization management and

incorporating technological updates. For the cloud consumer the formulation of web server, application server and database server configuration is a tiresome course of implementation. User should inspect the feasibility of their enterprise application to appropriate cloud delivery models. The networking configuration should match the ported virtual appliance for the seamless execution. To configure the virtual server user can also use service orchestration management tool like juju, chef, puppet, salt etc.

5. Conclusion and Future Work

In this paper we reviewed interoperability issues in cloud computing and also compared cloud deployment models with commercial offerings. We also have analyzed and formed an infrastructure cloud management taxonomy and mapped with the existing API's using the taxonomy. The best practices to eliminate possible interoperability problem are outlined. Addressing interoperability issues also reduces the adoption risks.

As for the future work, we planned to devise a working mechanism through mapping between models using private cloud. The mechanism has a cloud broker to handle portability request from the user and itself capable of migrating asset.

References

1. National Institute of Standards and Technology (2011), The NIST Definition of Cloud Computing, Online:http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf
2. Amazon Web Service (2013), Amazon EC2 Service Level Agreement, Online: http://aws.amazon.com/ec2/sla/
3. VMware, Understanding Full Virtualization, Paravirtualization, and Hardware Assist, Online: www.vmware.com/files/pdf/VMware paravirtualization.pdf
4. John Fisher Ogden (2006), Hardware Support for Efficient Virtualization, Online: http://cseweb.ucsd.edu/$\sim$jfisherogden/hardwareVirt.pdf
5. VMWare, Beyond Cost Savings :Four Compelling Reasons to Virtualize Your IT Environment, Online: http://www.technologent.com/wp-content/uploads/2013/10/VMware_vSOM_CallCampaign_102313_BeyondCostSavingsWP.pdf
6. Zhizhong Zhang, Chuan Wu, David W.L. Cheung, A survey on cloud interoperability: taxonomies, standards, and practice , ACM SIGMETRICS Performance Evaluation Review, Volume 40 Issue 4, March 2013
7. Istva n Mezga and Ursula Rauschecker, The challenge of networked enterprises for cloud computing interoperability, Elsevier, Computers in Industry, Volume 65, Issue 4, May 2014, Pages 657–674
8. André Monteiro, Joaquim Sousa Pinto, Cláudio Teixeira, Tiago Batista (2011), Sky Computing Exploring the aggregated Cloud resources – Part I, IEEE, Information Systems and Technologies (CISTI), 6th Iberian Conference.
9. The Open Group (2013), Cloud Computing Portability and Interoperability, ISBN:1-937218-30-0.
10. P.L. Bannerman, Cloud computing adoption risks: state of play, in 17th Asia Pacific Software Engineering Conference Cloud Workshop, 2010, pp. 10–16.
11. Amazon Web Services, http://aws.amazon.com/, November 2014.
12. Google App Engine, https://developers.google.com/appengine/, November 2014.
13. Salesforce, http://www.salesforce.com/, November 2014.
14. Eucalyptus, https://www.eucalyptus.com/, November 2014.
15. Openstack, http://www.openstack.org/, November 2014.
16. Cloud Foundry, http://cloudfoundry.org/index.html, November 2014.
17. Kernel Based Virtual Machine, http://www.linux-kvm.org/page/MainPage, November 2014.
18. Xen Hypervisor, http://xen.org/, November 2014.
19. VirtualBox, https://www.virtualbox.org/, November 2014.
20. Libcloud, https://libcloud.apache.org/, November 2014.
21. Jclouds, https://jclouds.apache.org/, November 2014.
22. Deltacloud, https://deltacloud.apache.org/, November 2014.
23. Fog, http://fog.io/, November 2014.
24. Cloud Data Management Interface (CDMI), http://www.snia.org/cdmi, November 2014.
25. Xplenty,Hadoop platform as a Service, https://www.xplenty.com, November 2014.