Achieving High Availability in Cloud through Live Migration

Abdul Wahab Khan[1] Shaikh Adnan Ahmed Usmani[2]

Abstract

Increasingly relying on the cloud for deployment and assessing critical applications and services for businesses makes its high availability as an extremely critical aspect. The paper evaluates virtualization based systems and techniques for the betterment of the overall resilience of a cloud environment. We have highlighted systems to perform monitoring, load balancing and dynamic allocation of resources, replication and live migration at backup sites and a number of pioneering approaches such as ghost VMs and Byzantine fault tolerance to ensure high availability. Moreover, hurdles and bottlenecks for the effectiveness and application of these systems are also identified. A real-world implementation of live migration is also presented with a concise discussion of the challenges faced during the setup and configuration phases.

Keywords: High Availability, Virtualization, Cloud Computing, Live Migration, Load Balancing

1 Introduction

Cloud computing, an emerging paradigm and committed to flourish accuracy & well-grounded delivery of computing and storage services is becoming an eminent choice for the enterprises. Over the years, organizations have shifted their core & critical business applications to the cloud and numerous users are becoming depended on online services. Considering modern scenario and increasingly growing traffic, it is crucial to maintain high availability of cloud services. The research to follow discusses the virtualization based techniques used to ensure high availability in the cloud and the limitations and problems associated with those.

Cloud Computing has flickered a mammoth amount of interest in the technology community. The on-demand nature of this paradigm makes it a perfect fit for today's deft business environment. With increasingly computing cost, cloud computing provides one excellent solution and a more obvious choice. Through improved consumption of resources and condensed administration and management costs, it offers substantial cost savings to enterprises. The entry of major industry players such as Microsoft, Google, and Amazon has accelerated the cloud's industry adoption rate as they introduce innovative features and drive costs down.

Virtualization is the core enabling technology behind this pattern. Multiplexing resources among applications and customers provide the elasticity to add new resources quickly and dynamically to a customer's resource pool. Greater efficiency is achieved by improving resource usages and noteworthy cost savings are realized by combining multiple servers into a single machine with numerous virtual machine instances.

Hypervisors, one of the most popular products of the research in virtualization technology and also called virtual machine managers, are programs that run directly on the bare-metal

[1]*PAF-Karachi Institute of Economics & Technology, Karachi, Pakistan / abdul.khan@pafkiet.edu.pk*
[2]*Universityof Karachi, Pakistan / adnan@szic.edu.pk*

hardware allowing multiple operating systems to share a single physical machine. Hypervisors also cope up the resources for the guest operating systems and ensure that they run in complete isolation from each other. When coupled with management tools and application programming interface, a hypervisor is an inclusive cloud platform that makes it convenient for users to setup and manage their own cloud.

Some of the renowned cloud platforms available today include commercial offerings such as Xen Cloud Platform [1] and VMWare ESX [2] as well as open source systems such as Eucalyptus [3] that are targeted for academic use. These platforms allow sharing of a single physical server by multiple virtual machines, failure in one node can result in disturbance of service to multiple applications and clients. Therefore, fiascos in the cloud have to be handled transparently to safeguard the cloud and its availability at all times.

Both Xen and VMWare have built-in mechanisms to warrant high availability. XenCenter includes a set of tools called Xen Essentials that comprises of high availability components. These include Xen Motion for performing live migrations of VMs and the Workload Balancer. VMWare HA (High Availability) is a utility that provides upbeat monitoring of servers and virtual machines, automatic detection of failures and swift restart and optimal placement of VMs after server failures [5]. Availability is ensured through dynamic adjustment of resource allocations, rapid restart or migration of VMs between hosts. Eucalyptus lacks in terms of high availability as its focus is further towards learning and research rather than usage for real-time applications.

2 Motivation

By 2012, 20% of the businesses will own no IT assets. The paradigm from in-house data centres to the cloud will result in businesses depending on the cloud for their precarious applications and services. Availability will be the key focus and concern for the organizations making this transition.

Companies are trusting on the cloud to provide diversified services ranging from photo sharing, online storage and social networking to financial services for enterprise customers. These companies include both start-ups as well as large corporations with thousands of customers. Therefore any outage in the cloud will affect thousands of customers and impact would be magnificent. For many, the consequences will be devastating and might be realized in the form of financial losses, customer dissatisfaction, negative publicity and a number of other ways.

In order to safeguard service availability, generally customers sign Service Level Agreements (SLAs) with the cloud providers. Availability is one of the most critical elements pursued by the customers when selecting a service provider. The chief industry players are capitalizing in research and energies to develop new and upgraded techniques for growing availability of cloud services.

Despite numerous systems and techniques developed, many are still new and immature which in fact serve the paramount basis of this research. We analysed the various techniques and methods for ensuring reliable delivery of services in the cloud and tried to identify the matters and restrictions. As part of this research, we have also setup a small cloud environment to get a hands-on experience of live migration and the other technologies involved.

3 Research Contribution

Our research focuses on analysing the various monitoring, fault-tolerance and load balancing techniques used in practice to achieve high availability in the cloud with a special emphasis on Live Migration. An attempt is made to identify the gap in the research by highlighting the limitations of each of these techniques.

Supporting work includes the creation of a test cloud using the Xen Cloud Platform and the development of a management console. Via this console, users can manage the servers and virtual machines in the cloud. It also provides the ability to handle excess load by creating new instances of running VMs, dynamically change resource allocations and perform live migrations.

4 Literature Review

A Ganglia

- Just a Monitoring System. It provides scalable monitoring for distributed High Performance Computing (HPC) systems [10].
- Scalable, that is nodes can be added easily in a cluster without manual configuration.
- Every node of a cluster share heartbeat messages to each other using listen/announce protocol.
- Monitoring sys has 2 daemons.
  - gmond daemon: Runs on every node of a cluster.
  - gmetad daemon: Runs on each cluster and then data aggregation of each cluster.
- Uses XML for data representation. Support both built-in and application specific metrics.

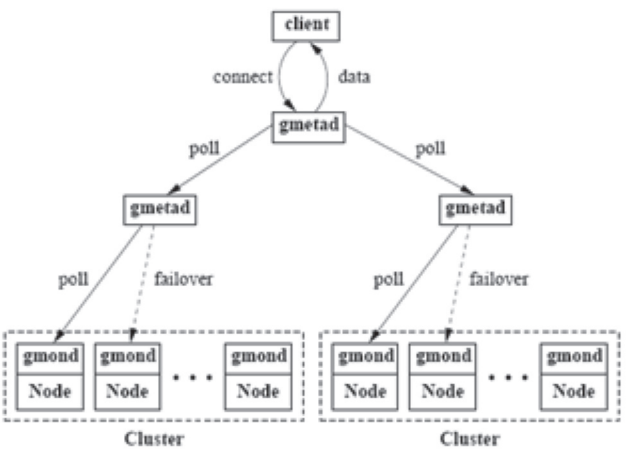

Figure1: Ganglia architecture

Figure 1 shows the architecture of Ganglia [10]. To federate multiple clusters, a tree of point to point connections is used. As each leaf node maintains monitoring information of its entire cluster, it logically represents a distinct cluster whereas each non-leaf node represents a set of clusters. At each point in the tree, child nodes are polled at regular intervals to aggregate data. Multiple nodes in a cluster are specified to handle failures for a leaf node.

Ganglia has a low per node overhead. This is chiefly due to new nodes discovery without any manual configuration thus eliminating noticeable management overhead. Through its multicast approach, scalability is automatically addressed. It provides support for both built-in and application specific metrics. Usage of XML for data representation and XDR for data transport enhances its extensibility, allowing integration with other information services.

Limitations. Although, Ganglia has been deployed in a number of real world environments and has proven its effectiveness in most cases, a few issues and constraints may limit its widespread adoption.

- Excessive network traffic for large of nodes.
- Can't run on WAN if WAN doesn't support multicast.
- Cost increases as nodes increases.
- Planet Lab tried implementation but 19.15 GB WAN bandwidth consumption makes the server down in a week [1].

B Sandpiper

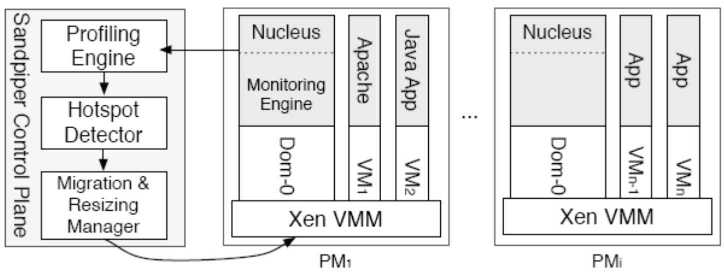

Figure 2: Sandpiper Architecture [12]

- Sandpiper goes a step further and in addition to monitoring the workloads, corrective actions are also possible [12].
- Nucleus runs on each physical host, for gathering network resource usage stats of the each host and then sends these stats to control plane.
- Control Plane uses automated resource management techniques and consists of:
  - Profiling Engine, whichconstructs resource usage profiles for each VM and host.
  - Hotspot Detector, which checks if the aggregate usage of processor, memory or network resource exceeds a threshold level.
  - Migration & Resizing Manager, which eliminates hotspots by
    - Assigning new resource shares to a VM

- Or migrating it to another physical host.
- Two Monitoring modes
  - Blackboxes suitable for environments where detailed peeking inside a VM is not possible. So on the basis of basics statsresource allocation is increased by a constant amount Δm if a hotspot is detected.
  - Graybox can use detailed OS and application level statistics. So accurate prediction of resource needs is possible.

Limitations

- For a large resource hungry VM, constant amount Δm in the black-box approach may not be very efficient.
- On other side of the picture, this constant amount may be quite large and result in over-allocation of resources.
- Migrating it to another physical host leave a VM unresponsive for a small period of time.
- VM on a single instance for their users, a migration may result in denial of service.

C Remus

basic aim is to provide a general purpose high availability service that offers a igh degree of fault tolerance without any modifications to software applications or the hardware

- It allows applications to switch on an alternate host within seconds of a failure [13].
- The Live Migration functionality is modified to replicate snapshots of running OS instances at frequencies as high as every 25ms. That is Remus is an extension of the ability of virtualization to migrate running VMs between hosts.

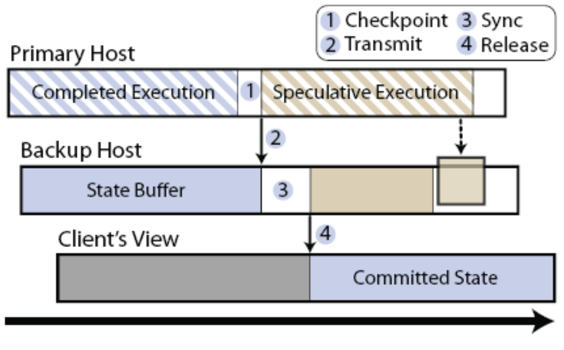

Figure 3: Asynchronous replication in Remus

- Asynchronous replication in Remus [15] shows the basic operation of Remus. At the checkpoint, the system state is transmitted to the backup host. And once the synchronization is complete, the output is released to the client.
- Speculative execution time allows the primary host to stay productive while its state is transferred asynchronously.
- Fast replicated performance is achieved by running the system tens of milliseconds in

the past.

- A single physical machine can serve as a backup host for multiple physical hosts.
- In comparison to commercial High Availability products that respond to a failure by rebooting the VM on another host, Remus provides a greater degree of protection. It allows recovery from failures on near live migration time frames leaving the VM running with the network connections intact. The state visible to the clients stays consistent and the disks are not corrupted.
- Failure detection mechanism is also integrated in the checkpoint process. When backup host stops responding, primary host assumes that backup has crashed. Similarly, when backup host stops receiving new checkpoints from primary host, it assumes a failure and resumes execution from the most recent checkpoint.

Limitations. Remus aims to reduce the cost associated with ensuring high availability but additional hardware is still required in the form of backup hosts.

- The paper suggests that a single backup host can be used in N to 1 configuration to serve a number of active hosts but in case of large data centres, the number of dedicated backup hosts might be quite large resulting in significant additional expense.
- During tests, it has been noticed that Remus does introduce significant network delay, especially for applications that have poor locality in memory writes. Therefore, for applications requiring low latency, Remus may not be the right choice.
- When VM is transferred over the network to the backup host, it may result in excessive network traffic.

D Live Migration across Wide Area Networks (WANs)

The various techniques discussed above enable Live VM migration in a LAN environment making data centre management easier and non-disruptive. In a WAN environment, performing a live migration can yield similar benefits.

- It can be especially useful in scenarios where the data centre as a whole becomes unavailable due to catastrophic events or data centre wide maintenance operations. In such situations, migrating services to another data centre may be the only viable solution.
- Unfortunately, performing live migration across WAN involves a number of complexities and challenges. The cooperative context aware approach proposes a strategy where the server, network and storage cooperate and coordinate in such a way that they are aware of the service context to achieve seamless migration over WAN [14].
- The IP address mobility problem is addressed by using tunnelling technology to establish connectivity between data centres over WAN on a pre-emptive basis.
- Storage replication is performed in two steps. Initial transfer of data takes place in asynchronous mode. Afterwards, synchronous replication is used to ensure consistency and prevent data loss during migration process.

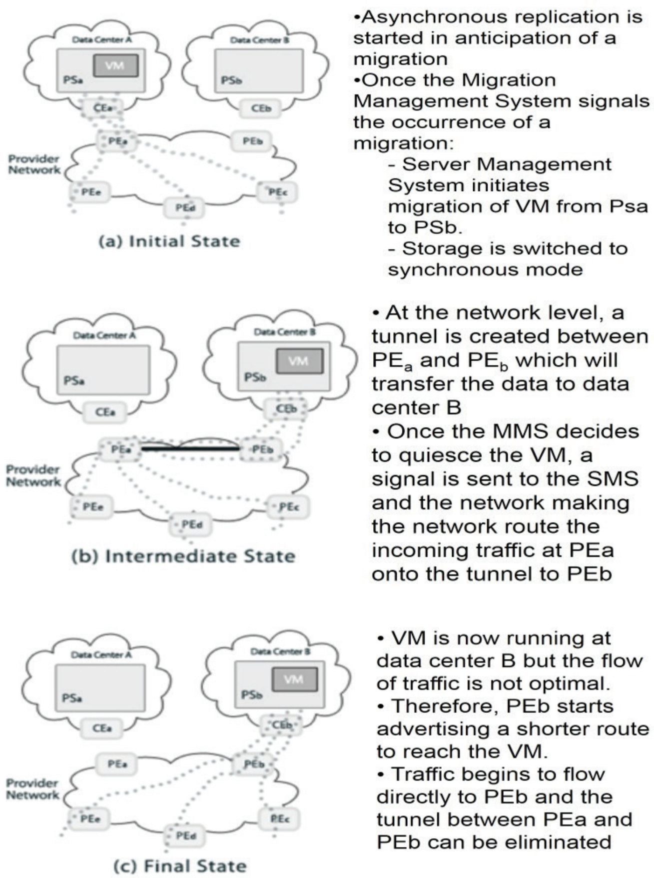

Figure 4: Live Server Migration across Wide Area Network [10]

Limitations

- Migrating VMs over WAN can be a very slow process. In comparison to live migrations over LAN, total migration time will be exponentially high. For VMs with large amount of state information, the technique may not be feasible. Especially in case of complete data centre outages requiring every single VM to be migrated to the backup site. Using this approach will require large amount of network bandwidth and an unreasonably long time frame.
- Using asynchronous replication on a continuous basis for transferring state information over WAN to the backup site might be a more feasible option. In case of an outage or failure, only the pending state changes will be transferred to the backup site decreasing the total time for migration. However, this approach will result in increased performance overheads under normal operation.

E Cloud Net and Cloud to Cloud Migration

- Uses Virtual Private Networks to connect multiple clouds forming Virtual Private Clouds [15].
- Through this technique, cloud to cloud migration can become as simple as performing a living migration over the local network.

- The architecture consists of two controllers: Cloud Manager and Network Manager that automate the management of resources between the two clouds.
- Cloud Manager handles the creation of new virtual machines. It is also responsible for managing the performance within the VPC. By partitioning physical routers into slices with independent control planes for each, virtual or logical routers are created and used to configure the Customer Edge routers in the cloud.
- The Network Manager creates VPNs and provides resource provisioning capabilities. These VPNs eliminate the need for endhost configurations and have lower overheads.
- CloudNet ties networks across WAN into a single LAN. As a result, VM migrations just take place as on a local area network.

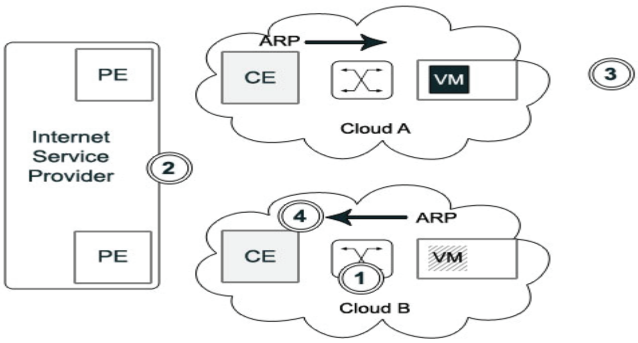

Figure 5: WAN migration using CloudNet [15]

- CloudNet first initializes a virtual LAN (VLAN) endpoint for the destination cloud. A VPN is created to link together the source and destination VLANs after which it is possible to migrate a VM between the two sites. Once the VM is transferred to the other cloud, it transmits an ARP message. This message is used by the local switch to map the VM's MAC address and its switch port. It is also forwarded to the VM's original site where the old switch will replace its MAC address mapping with the new entry. From this point onwards, the data will be forwarded to the VM's new location.

Limitations

- Network delays is quite high. Transferring the VM state over WAN can be a very slow process making it less viable for performing load balancing.

F Ghost Virtual Machines

The concept of Ghost VMs takes high availability to a new level by providing almost instantaneous resource allocations [16].

- Ghost VMs are pre-deployed spare VMs maintained on each physical server that remain in active state but do not service client requests as they are detached from the internet. These ghost VMs can be activated as and when required by making them visible to the content switch that handles client requests. These VMs consume minimal resources on their host machines in idle state.
- Although they remain hidden from the internet, the ghost VMs can communicate with each other using the host system's second network card or through the L2 functionality of the switch. When another VM is required for an application, a compatible ghost VM

can be activated by reconfiguring the switch. The process is extremely fast and has been shown to be completed within as little as 2 seconds.

- Decision Manager Component: Manages the virtual machines and performs the resource management function.This component performs its resource management function in three stages. It decreases the capacity for applications with over-allocation of resources. Next, it increases the resource allocations for applications that are under-allocated. Finally, it ensures that ghost VMs will be available to applications that are likely to see an increase in demand.
- Ghost Manager Component: Allocates ghost VMs to applications and ensures availability of spare ghost VMs

Limitations

- In a public cloud with a large number of heterogeneous applications running, it may not be possible to maintain a ghost VM for every different application.
- If application is overloaded, the ghost manager searches for compatible ghost VM. In case, if a compatible VM is not found, the ghost manager stops the running application and restarts the application. This is a highly undesirable scenario as the time taken to start the application will be much longer.

G ZZ: Cheap Byzantine Fault Tolerance Using Virtualization

- Fault tolerance is a highly desirable trait for systems requiring high availability. Byzantine Fault Tolerance (BFT) is a technique that uses replication to minimize the impact of faults. BFT is highly effective in dealing with faults but its high cost prevents its widespread adoption. It requires at least 2f+1 execution replicas to tolerate f Byzantine faults.
- ZZ is a system that uses virtualization to decrease the cost associated with BFT to half [13]. In order to tolerate 2 faults, a typical BFT system will require 2f+1=5 replicas. ZZ can decrease the number of replicas to f+1=3 resulting in significant cost savings and the usage of fewer resources to process non-faulty requests. The basic concept behind ZZ is to use additional replicas only in the event of a failure. Virtualization can enable these additional replicas to be activated on demand.
- In order to decrease the cost of implementing BFT, ZZ uses 3 mechanisms. First, it uses virtualization to enable fast replica start-ups and to multiplex a small pool of free servers across recovery replicas for a large number of applications. Second, it employs a recovery protocol that allows additional replicas to begin processing requests through a state transfer mechanism that fetches state on demand. Finally, ZZ performs incremental check-pointing by exploiting the snapshot feature of modern file systems without requiring modifications to the application source code.
- In a typical data centre, N independent applications will be running at a given time. During normal operation, f+1 replicas will be run as separate virtual machines. To run these replicas, a small pool of free servers can be multiplexed. The system makes the assumption that not all N applications will experience a fault simultaneously. Therefore we can use a smaller pool of f+1 replicas instead of the standard 2f+1.

- A ratio "r" is used to determine the number of backup servers required for fault-mode execution and denotes the ratio of the time for ZZ to recover and replace a faulty replica to the mean time to failure for the application. Usually, this ratio is less than one as the recovery time is in seconds and the mean time to failure can be in days or more.
- Virtualization enables replicas to be started instantaneously in the event of a failure. The replicas can be maintained as VMs in paused state which requires no CPU consumption and negligible memory. On detecting a failure, the VMs can be un-paused within milliseconds.
- ZZ is based on the Xen platform and has been shown to recover a replica with 400MB of disk state in less than 4 seconds in test results whereas the existing approaches required 60 seconds. It incurs minimal overhead during normal operation.

Limitations

- In a public cloud with a large number of different applications running in parallel, the probability and the number of failures simultaneously occurring is difficult to determine. Therefore, the calculation of the ratio "r" may prove to quite difficult as it requires knowledge about the applications and their mean time to failures.

5 Development Of Live Migration Implementation

In order to support our research, we started off with a small project that aimed to setup a cloud computing environment. The purpose was to get hands on experience of live migration using one of the most popular cloud platforms called Xen. The research also includes the development of a small application that will enable users to manage their cloud by letting them perform basic operations as well as load balancing through live migrations and dynamic resource provisioning. Although due to a number of constraints and challenges, the load balancing part could not be completed.

A cloud environment was setup using Xen Cloud Platform. The purpose is:

- Hands on experience of setting up and configuring a cloud environment
- Perform Live migration and observe its efficacy with regards to ensuring high availability
- Provide high availability and service levels based on the defined QoS parameters

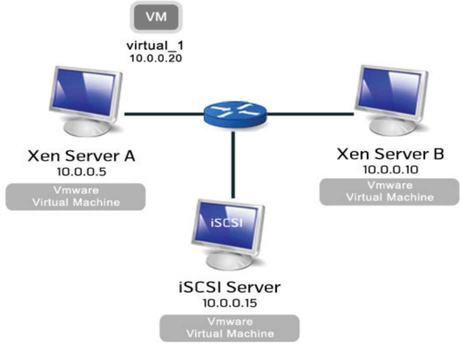

Figure 6: Test Cloud Architecture

Figure 6 shows the basic architecture of our test cloud. There are two physical hosts XenServerA and XenServerB. A separate iSCSI Server is used to provide the shared storage. The three servers are connected with each other through a router. In order to perform live migration, the virtual machine to be migrated has to be reside on a shared storage device so that it can be accessible from both hosts during the migration process. Due to this reason, virtual_1 uses ISCSI Server as its storage device. Note that due to the unavailability of dedicated hardware, we are running these servers as virtual machines on VMWare Workstation 6.5 using two physical hosts [18].

Table 1: Server Configurations

Machine hosting XenServer A (PM1)	Machine Hosting XenServer B (PM2)
AMD Athlon X2 @ 1.9GHz	Intel Core 2 Duo E6500 @ 2.66GHz
2.5GB of RAM	3GB of RAM
120 GB HDD	320GB HDD
VMWare Workstation 6.5	VMWare Workstation 6.5

Table 1 shows the configuration of the physical hosts PM1 and PM2. In addition to running XenServerA, PM1 also hosts the virtual machine for Open Filer iSCSI server [15].

A Management Interface

A software application has been developed to provide the management interface for the test cloud. The application makes use of XenAPI provided as part of the Xen Server Software Development Kit [19]. XenAPI exposes functions and parameters that allow remote configuration and management of a Xen Server and the associated guest VMs. The application enables users to connect to a specific server in the cloud and manage the virtual machines residing on that server. Detailed information of the virtual machines is displayed including their current state (started, paused, halted) and the number of processors assigned. Through the management interface, VMs can be started, paused, resumed, suspended or stopped. The user can start new instances of the same virtual machine by using the clone operation.Dynamic resource allocation can be performed by changing the amount of RAM allocated to each VM or migrating it to another physical host available in the pool.

B Migrating a VM

From the management interface, the user can perform live migration of a VM between two servers. Live migration is considered to be a non-disruptive operation that allows a running VM to be migrated to another host with unnoticeable downtime and all the network connections intact.

For live migration to take place, the VM should use a shared storage that is accessible to both source and destination servers. In my testing, I have installed the guest VM on XenServerA using iSCSI Server as the storage repository. The VM was pinged for network connectivity while it was being migrated to XenServerB. Packet loss was observed for a very brief moment approximating to 5 sec during which the final step of migration was being performed and the VM was active on XenServerB immediately afterwards. The total time taken by the migration process was around 4mins. Both the downtime and the overall migration time are quite high. Professional Xen implementations have reported downtimes in milliseconds and total migration times less than 2 minutes. In my testing, I have used an extremely lightweight VM with small memory footprint. Therefore the performance is way below the standards. This poor performance can be easily attributed to the hardware used. Unlike commercial Xen implementations that use multi core multi-processor systems with high end SAN devices and Gigabit Ethernet, the test environment relied on notebook computers with limited system RAM and 100Mbps Ethernet. The performance was further degraded due to multiple layers of virtualization as the XenServers were also running as virtual machines.

C Challenges/Constraints

- Unavailability of Highly Specs Hardware: is a serious constraint. As a solution, VMWare workstation can be used.
- Installation of First Xen Server: is a difficult and time taking task. After a number of failed attempts and guidance from forums, wewere able to get Xen running as a virtual machine.
- Guest VM installation: We need a guest VM with either Windows OS with HVM or Linux OS in paravirtualized mode. Windows OS with HVM was not available so we went with Linux OS and used Lightweight VM Xen SDK due to performance issue.
- Installation of Second Xen Server: is required once again and two Xen Servers cannot run on a single machine. As a solution, second physical host is needed.
- Unavailability of shared storage device: As a solution, we came across an open source storage management appliance Open Filer.
- Creation of a Resource Pool (Tightly coupled collection of servers): VMs can be migrated between servers belonging to the same resource pool. Live migration is supported between homogenous platforms, otherwise, we need to join resource pool forcefully.

6 Conclusion

This research explores the systems and techniques that can make the cloud more reliable and highly available. Our work focuses on the automated techniques that exploit the benefits of virtualization and technologies such as live migration to provide fast and efficient fault tolerance and failure prevention mechanisms. We have analysed monitoring systems such as Ganglia and Sandpiper that gather resource usage statistics to detect hotspots and perform load balancing. Sandpiper's graybox monitoring mode enables it to proactively adjust resource allocations and prevent hotspots from occurring. The scalable nature of these systems makes them ideal for usage in a cloud environment.Replication is one of the basic techniques for achieving high availability. Remus provides availability as a service at the virtualization layer by using replication in an innovative way.Live migration enables running VMs to be migrated with minimal downtimes between physical hosts in a LAN environment. Some other techniques and systems that we have discussed include Ghost VMs and ZZ. Ghost VMs provide the benefits of live migration without the performance overheads and with decreased downtimes. Based on Xen platform, ZZ is a system that uses virtualization to make the popular Byzantine Fault Tolerance (BFT) technique more cost effective and viable for usage in the cloud. Although, these techniques can prove to be extremely beneficial in improving the availability of a distributed environment, a number of challenges and constraints limit their effectiveness. The research briefly discusses these challenges.In the end, we have presented the details of our live migration implementation. Using Xen Cloud platform, we have setup a cloud environment and configured it to support live migrations. Moreover, a management console has been developed to facilitate administration of the cloud. We have also provided details of the challenges faced during setup and configuration.

7 Recommendations For Future Work

There are a number of opportunities to further extend this research. Additional work can be performed in the areas of fault tolerance and replication. Multiple systems and techniques can be combined to take a holistic approach towards availability in the cloud. For example, the ghost VM approach can be incorporated in sandpiper to further optimize the load balancing component, reducing downtime and performance overheads at the same time. Increased network traffic and performance overhead is a major drawback of the load balancing and monitoring techniques discussed above. Further research can be focused on devising techniques to lower these overheads. A monitoring system such as Ganglia can be integrated in the system allowing cloud administrator to get real time access to usage statistics. The addition of automated load balancing techniques can also prove to be equally useful.

References

[1] Xen Cloud Platform XAPI: Open source software to build private and public clouds

[2] VMWare ESX Server

[3] D. Nurmi, R. Wolski, C. Grzegorczyk, G. Obertelli, S. Soman, L. Youseff, D. Zagorodnov, “The Eucalyptus Open-source Cloud-computing System”, in the Proceedings of the 9th IEEE/ACM International Symposium on Cluster Computing and the Grid.

[4] Siddharth Jain, Rakesh Kumar, Anamika, Sunil Kumar Jangir, “A Comparative Study for Cloud Computing Platform on Open Source Software”. An International Journal of Engineering & Technology (AIJET) Vol. 1, No. 2 (December, 2014)

[5] Makhdoom Muhammad Naeem, *Hidayatullah Mahar Furqan Memon, Muhammad Siddique, Abdul Rauf, “AN OVERVIEW OF VIRTUALIZATION & CLOUD COMPUTING”. Sci. Int.(Lahore),28(4),3799-3803 ,2016 ISSN 1013-5316;CODEN: SINTE 8

[6] Rakesh Kumar , Sakshi Gupta, “Open Source Infrastructure for Cloud Computing Platform Using Eucalyptus”. Global Journal of Computers & Technology Vol. 1, No. 2, December 24, 2014 www.gpcpublishing.com ISSN: 2394-501X

[7] CitrixXen Center

[8] VMWare High Availability

[9] Gartner, "Gartner Highlights Key Predictions for IT Organizations and Users in 2010 and Beyond", STAMFORD, Conn., January 13, 2010.

[10] M. L. Massie, B. N. Chun, and D. E. Culler, "The Ganglia Distributed Monitoring System: Design, Implementation, and Experience", 15th June 2014.

[11] Chandramouli Reddy , Suchithra R "Virtual Machine Migration in Cloud Data Centers for Resource management". International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 5 Issue 09 September 2016 Page No.18029-18034

[12] T. Wood, "Improving Data Center Resource Management, Deployment, and Availability with Virtualization", University of Massachusetts Amherst, September 2011.

[12] Mukil Kesavan, Irfan Ahmad, Orran Krieger, Ravi Soundararajan, Ada Gavrilovska, and Karsten Schwan(2013). "Practical Compute Capacity Management for Virtualized Data Centers". IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 1, NO. 1, JANUARY-JUNE 2013

[13] B. Cully, G. Lefebvre, D. Meyer, M. Feeley, N. Hutchinson, and A. Warfield, "Remus: High Availability via Asynchronous Virtual Machine Replication", Dept. of Computer Sciences, University of British Columbia.

[14] K.K. Ramakrishnan, P. Shenoy, J. V. Merwe, "Live Data Center Migration across WANs: A Robust Cooperative Context Aware Approach",AT&T Labs – Research / University of Massachusetts.

[15] T. Wood, P. Shenoy, A. Gerber, K. K. Ramakrishnan, J. V. Merwe, "The Case for Enterprise-Ready Virtual Private Clouds",AT&T Labs – Research / University of Massachusetts.

[16] W. Zhang, H. Qian, C. E. Wills, M. Rabinovich, "Agile Resource Management in a Virtualized Data Center",Worcester Polytechnic Institute, Worcester.

[17] T. Wood, R. Singh, A. Venkataramani, and P. Shenoy, "ZZ: Cheap Practical BFT using Virtualization", University of Massachusetts Amherst, 2009.

[18] VMWare Workstation Workstation for windows

[19] OpenFiler Storage Management Appliance

[20] Xen Software Development Kit