Cloud Computing Assignment - 2

Name: Akash Kotnala

Roll Number:- 18CS01001

Q1) Write a comparative study of pre-copy and post-copy based virtual machine migration.

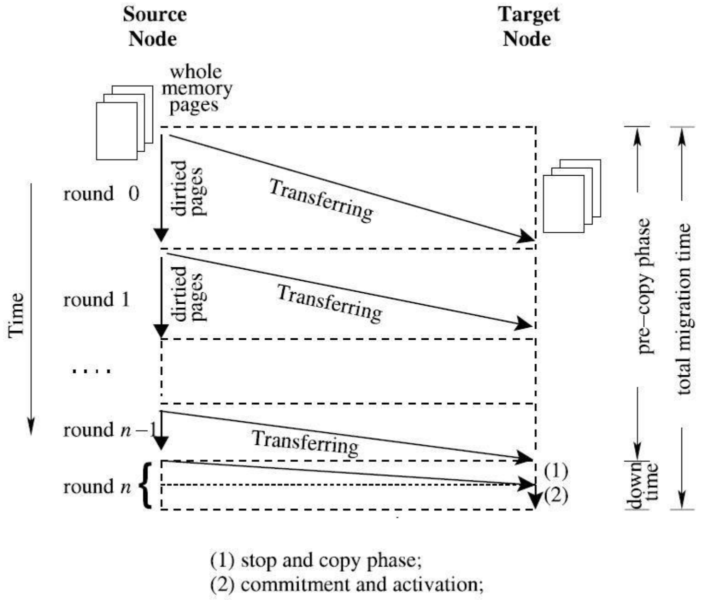
There are two methods used for live migration. They are:-

1) PRE-COPY 2) POST-COPY

Precopy approach for virtual machine migration:-

In precopy migration, in the first iteration, the hypervisor copies all the memory pages from the source physical machine to the destination machine. In the next iterations, the dirty pages of the source machine are copied and transferred to the destination machine until the rate of pages being dirtied reaches a certain low value. After this value is reached, the VM is turned off at the source machine and after the transfer of last dirty pages is completed, the VM is started on the destination machine.

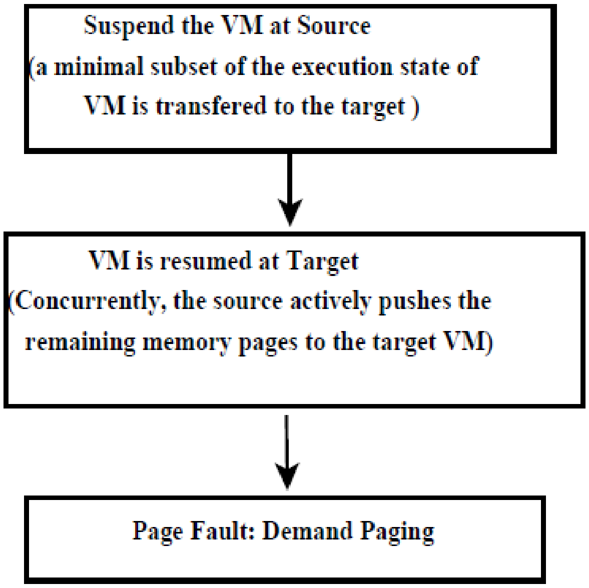
Pre-copy approach transfer diagram:-



Postcopy approach for virtual machine migration:-

In this method, initially the VM is suspended on the source node and then a minimal subset of the execution state of the VM which includes CPU state, registers and non-pageable memory is transferred to the destination machine. The VM is then started at the destination machine. At the same time, the source machine actively pushes the remaining memory pages of the VM to the target machine. This is known as ‘pre paging’. On the other hand, if the running VM at the destination node tries to access a page that has not yet been transferred, a page fault occurs. These page faults are trapped at the destination and are redirected to the source machine, which then sends the faulted pages. This is known as Demand Paging. An ideal pre-paging scheme would mask the majority of page-faults, although its performance depends upon the memory access pattern of the VM’s workload.

Post-copy approach state diagram:-


Suspend the VM at Source
(a minimal subset of the execution state of VM is transfered to the target)
VM is resumed at Target
(Concurrently, the source actively pushes the remaining memory pages to the target VM)
Page Fault: Demand Paging

Post-copy vs Pre-copy approach:-

1. In pre-copy approach, the VM retains an up-to-date state on the source machine during migration, whereas in the post copy approach, the state of the virtual machine is distributed over both source and destination machines. If the destination fails during migration, pre copy method can recover the VM, whereas it’s not possible in the post copy method.

2. In the pre-copy approach the same page may be transferred multiple times if the page is getting dirtied repeatedly at the source machine during migration. However, in the post-copy approach, each page is transferred exactly once.

3. The downtime and migration times are also different for both these approaches. In most of the cases, downtime is lesser for the pre-copy approach than post-copy approach and migration time is lesser in post copy approach than pre copy approach. This is because, the total of pages transferred is more in case of pre-copy approach (as pages keep getting dirtied) and as a result migration time is higher for pre copy approach. The downtime is higher for post copy approach because after the machine is turned off, the minimal processor state is copied and transferred and this process usually takes more time. So post-copy approach generally has more downtime than pre-copy approach.

Q2) Write an experimental report on Xen Virtual Machine Migration service and related performance and security challenges.

Xen is a type-1 hypervisor, providing services that allow multiple computer operating systems to execute on the same computer hardware concurrently. Xen uses para-virtualization architecture and required the OS to be modified to run on Xen. Xen can also migrate virtual machines between different physical machines running Xen. The following are the steps to be followed for migration of VM using Xen:-

1) Migration with shared storage:-

1. The persistent storage for the VMs must be shared, so that both Dom0s can see the same disk, at the same location.

2. Then the command **xl migrate <domain> <host>** is ran to migrate the VM.

2) Migration without shared storage:-

- First, nbd-client is installed on all Dom0s (apt-get install nbd-client)
- Then nbd-server is installed on the Dom0 of the machine that contains the storage using the command: “apt-get install nbd-server”.
- The following information is added to ‘/etc/nbd-server/config’ file:

```
[<name>]
        exportname = <path to disk to share>
        port = 9000
```

- Then “nbd-client 127.0.0.1 9000 /dev/nbd1” is ran on the same machine
- Next, “nbd-client <hostname of machine with storage> 9000 /dev/nbd1” is ran on the machine we are migrating to.
- Finally, ‘/etc/xen/foo.cfg’ file is modified to use ‘/dev/nbd1’ as its disk, on the host backing the storage.
- Start the VM
- Run “xl migrate <Domain> <new host>”

Security Challenges Xen VM Migration Service:-

The major security challenges faced by Xen VM migration are classified into 3 types. They are:-

1. Control Plane challenges
2. Data Plane challenges
3. Migration Module challenges

Control Plane Challenges:-

Control Plane refers to the communication mechanisms employed by the VMM to initiate and manage live virtual machine migrations. These communication mechanisms must be authenticated. In addition, the protocols used in the control plane must be protected against spoofing and replay attacks. A lack of proper access control may allow an attacker to arbitrarily initiate VM migrations.

The major control plane challenges are:-

1) **Incoming Migration Control:-** By initiating unauthorized incoming migrations, an attacker may live-migrate a guest VMs onto the attacker's machine and gain full control over guest VMs and thus compromising the guest VM.
2) **Outgoing Migration Control:-** By initiating outgoing migrations, an attacker may migrate a large number of guest VMs to a legitimate victim VMM, overloading it and causing disruptions or a denial of service.
3) **False Resource Advertising:-** In an environment where live-migrations are initiated automatically to distribute load across a number of servers, an attacker may falsely advertise available resources via the control plane. The attacker may influence the control plane to migrate a VM to a compromised VMM by pretending to have a large number of spare CPU cycles.

Data Plane Challenges:-

To protect the virtual machine’s state, the data plane across which VM migrations occur must be secured and protected against snooping and tampering. In some scenarios man in the middle attacks can occur.

Passive Snooping:-

By monitoring the migration transit path, an attacker can extract information from the memory of the migrating VM such as passwords, keys, application data, and other sensitive information. This is known as passive snooping.

Active Manipulation:-

An attacker may manipulate the memory of a VM as it is migrated across the network. Such a man-in-the-middle attack may result in a complete and secret compromise of the guest OS. This is one of the most severe attacks.

Migration Module Challenges:-

The module that implements live migration functionality in a VMM must also be resilient to attacks. Common software vulnerabilities such as stack, heap, and integer overflows can be exploited by a remote attacker to bring down the VMM.

Security of a VMM's migration module is important because:

1) The VMM controls all the guest operating systems running in it and hence the severity of a VMM vulnerability is much greater than most normal software.
2) If VMM gets compromised through its migration module then the integrity of any guest VM running within the VMM and the integrity of VMs migrated to that VMM in the future will also be compromised.

Xensploit:-

It is a tool developed to perform man-in-the-middle attacks. It is based on the ‘fragroute’ framework. It manipulates the VM as it traverses through the internet. Xensploit can manipulate the VMWare and Xen migrations.

Using Xensploit 3 kinds of attacks were performed. They are:-

1) Simple Memory Manipulation
2) Sshd Authentication Manipulation
3) Xen migration module

Simple Memory Manipulation:-

Using Xensploit, a simple manipulation during the live migration of a Xen VM is performed. In Xen terminology, a host VMM is known as a dom0 domain while guest VMs are known as domU domains.
Testing involves the following three machines:-

1) The source dom0
2) The destination dom0
3) A malicious node running Xensploit.

A new guest domU, the domain to be migrated, is started within the source dom0. Inside domU, a test process that simply prints a "Hello World" string to the terminal each second is executed. The DomU on source is live migrated on to destination. As the memory pages of the running guest OS are transmitted over the network and passed through the malicious node running Xensploit, the “Hello World” string is replaced by a custom string. The attack was successful and after migration, the migrated VM printed the custom string instead of printing “Hello World”.

SSHD Authentication manipulation:-

Xensploit is now used to manipulate the memory of the Secure Shell daemon (sshd) process of a guest VM during a live migration.
The Testbed has four machines:

1) The source VMM
2) The destination VMM
3) A management node to manage the VMMs and initiate the migration
4) The malicious node running Xensploit.

Before initiating the migration, the sshd process was configured to only allow authentication of the type ‘PubkeyAuthentication’. An attempt to ssh to the guest OS running within the source VMM was failed as the public key was not in the root user's ‘.ssh/authorized_keys’ file, and hence the access was denied. The live migration is then initiated and the man-in-the-middle attack was performed using ‘xensploit’. The in-memory object code of the sshd process, originating from ‘user_key_allowed2’ function in ‘auth2-pubkey.c’, is manipulated during migration to successfully authenticate any incoming ssh logins. After xensploit's attack, an attempt to ssh to the VM succeeds due to the manipulated sshd process. Hence this attack also succeeded and result is that any attempt to SSH succeeds.

Migration Module:-

By exploring the Xen source code, multiple issues which fall into the migration module class of live migration threats were discovered. The vulnerabilities are present in Xen's VMM migration routines, specifically the code in ‘xc_domain restore.c’, which is used to restore an incoming migration to operational state. One vulnerability exploits an integer signed-ness issue resulting in a stack overflow, and another involves a malloc () integer overflow resulting in a potential heap overflow. These two issues allow a remote attacker to achieve privileged code execution and completely compromise the Xen VMM and host machine. These are some of the security challenges present in Xen virtual machine migration service.

Q3) Write a report on Google Cloud Platform and discuss on application development, usability and performance challenges.

Google Cloud Platform:-

Google Cloud Platform, offered by Google, is a suite of cloud computing services that runs on the same infrastructure that Google uses internally for its end-user products, such as Google Search, Gmail, Google Drive, and YouTube. Alongside a set of management tools, it provides a series of modular cloud services including computing, data storage, data analytics and machine learning. Registration requires a credit card or bank account details.

GCP provides Infrastructure as a service (IaaS), Platform as a service (PaaS), and serverless computing environments. Google App Engine, a platform for developing and hosting web applications in Google-managed data centers, was the first cloud computing service from google. Since the announcement of App Engine, Google has added multiple cloud services to the platform. Google Cloud Platform is a part of Google Cloud, which includes the Google Cloud Platform public cloud infrastructure, as well as Google Workspace (G Suite), enterprise versions of Android and Chrome OS, and application programming interfaces (APIs) for machine learning and enterprise mapping services. Google lists over 100 products under the Google Cloud brand.

Some of the key services are given below:-

Compute Services:-

- App Engine - Platform as a Service to deploy Java, PHP, Node.js, Python, C#, .Net, Ruby and Go applications.
- Compute Engine - Infrastructure as a Service to run Microsoft Windows and Linux virtual machines.
- Google Kubernetes Engine (GKE) or GKE on-prem offered as part of Anthos platform - Containers as a Service based on Kubernetes.
- Cloud Functions - Functions as a Service to run event-driven code written in Node.js, Java, Python, or Go.
- Cloud Run - Compute execution environment based on Knative. Offered as Cloud Run (fully managed) or as Cloud Run for Anthos. Currently supports GCP, AWS and VMware management.

Storage & Databases Services:-

- Cloud Storage - Object storage with integrated edge caching to store unstructured data.
- Cloud SQL - Database as a Service based on MySQL, PostgreSQL and Microsoft SQL Server.

- Cloud Bigtable - Managed NoSQL database service.
- Cloud Spanner - Horizontally scalable, strongly consistent, relational database service.
- Cloud Datastore - NoSQL database for web and mobile applications.
- Persistent Disk - Block storage for Compute Engine virtual machines.

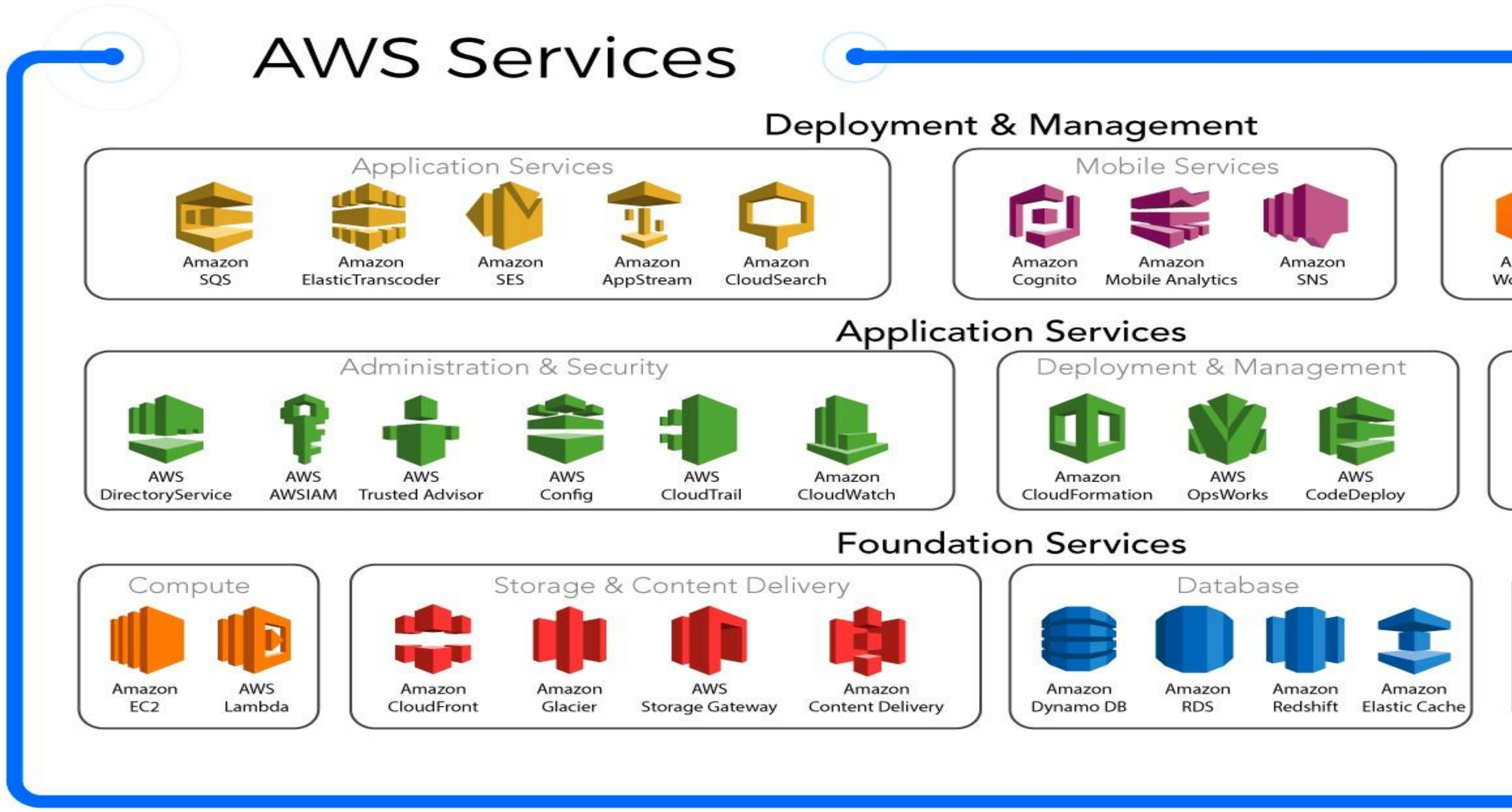
Networking Services:-

- VPC - Virtual Private Cloud for managing the software defined network of cloud resources.
- Cloud Load Balancing - Software-defined, managed service for load balancing the traffic.
- Cloud Armor - Web application firewall to protect workloads from DDoS attacks.
- Cloud CDN - Content Delivery Network based on Google's globally distributed edge points of presence.

Q4) Write a report on Amazon Web Services Cloud

Amazon Web Services:-

Amazon Web Services (also known as “AWS”) is an evolving cloud computing platform provided by Amazon that includes a mixture of infrastructure as a service (IaaS), platform as a service (PaaS) and packaged software as a service (SaaS) offerings. It provides on-demand cloud computing services and APIs to individuals, companies, and governments, on a metered pay-as-you-go basis. These cloud computing services provide a variety of basic abstract technical infrastructure and distributed computing building blocks and tools.



Services offered by AWS

Currently, AWS consists of over 200 products and services including computing, storage, networking, database, analytics, application services, deployment, management, machine learning, mobile, developer tools and tools for the Internet of Things. The most popular include Amazon Elastic Compute Cloud (EC2), Amazon Simple Storage Service (Amazon S3), Amazon Connect, and AWS Lambda (a serverless function enabling serverless ETL). Most of these services are not exposed directly to end users, but instead offer functionality through APIs for developers to use in their applications. These services can be accessed over HTTP, using the REST architectural style (and SOAP protocol) for older APIs and exclusively JSON for newer ones.

The most popular services offered by AWS are:-

1) Amazon EC2 (provides virtual servers in the cloud)
2) Amazon Simple Storage Service (S3) (scalable storage in the cloud)
3) Amazon DynamoDB (managed NoSQL database)
4) Amazon Aurora (high performance managed relational database)
5) Amazon RDS (managed relational database service for MySQL, PostgreSQL, Oracle, SQL Server, and MariaDB)

1) Amazon EC2 (IaaS):-

Amazon Elastic Compute Cloud (EC2) is a part of Amazon Web Services, that allows users to rent virtual computers on which to run their own computer applications. EC2 encourages scalable deployment of applications by providing a web service through which a user can boot an Amazon Machine Image (AMI) to configure a virtual machine, called as an "instance", containing any desired software. A user can create, launch, and terminate server-instances as needed, paying by the second for active servers. Amazon's auto-scaling feature of EC2 allows it to automatically adapt computing capacity to site traffic. The schedule-based (e.g. time-of-the-day) and rule-based (e.g. CPU utilization thresholds) auto scaling mechanisms are easy to use and efficient for simple applications. EC2 provides users with control over the geographical location of instances that allows for latency optimization and high levels of redundancy. In November 2010, Amazon switched its own retail website platform to EC2.

2) Amazon S3 (Storage-as-a-Service):-

Amazon Simple Storage Service (Amazon S3) is an object storage service offering industry-leading scalability, data availability, security, and performance. Customers of all sizes and industries can store and protect any amount of data for virtually any use case, such as data lakes, cloud-native applications, and mobile apps. With cost-effective storage classes and easy-to-use management features, we can optimize costs, organize data, and configure fine-tuned access controls to meet specific business, organizational, and compliance requirements.

3) Amazon DynamoDB (Database):-

Amazon DynamoDB is a fully managed, serverless, key-value NoSQL database designed to run high-performance applications at any scale. DynamoDB offers built-in security, continuous backups, automated multi-region replication, in-memory caching, and data export tools. It is a fast, flexible NoSQL database service with single-digit millisecond performance at any scale. Working diagram of DDB:-

4) Amazon Aurora (Database):-

Amazon Aurora is a MySQL and PostgreSQL-compatible relational database built for the cloud that combines the performance and availability of traditional enterprise databases with the simplicity and cost-effectiveness of open source databases. Amazon Aurora is fully managed by Amazon Relational Database Service (RDS), which automates time-consuming administration tasks like hardware provisioning, database setup, patching, and backups. Amazon Aurora features a distributed, fault-tolerant, self-healing storage system that auto-scales up to 128TB per database instance. It delivers high performance and availability with up to 15 low-latency read replicas, point-in-time recovery, continuous backup to Amazon S3, and replication across three Availability Zones.

5) Amazon RDS:-

Amazon Relational Database Service (Amazon RDS) is a distributed relational database service by AWS. It is a web service running in the cloud designed to simplify the setup, operation, and scaling of a relational database for use in applications. Administration processes like patching the database software, backing up databases and enabling point-in-time recovery are managed automatically. Scaling storage and compute resources can be performed by a single API call to the AWS control plane on-demand. Amazon RDS is available on several database instance types and provides us with six familiar database engines to choose from, including Amazon Aurora, PostgreSQL, MySQL, MariaDB, Oracle Database, and SQL Server. We can use the AWS Database Migration Service to easily migrate or replicate your existing databases to Amazon RDS.

6) AWS Lambda (Compute Service):-

AWS Lambda is a serverless, event-driven compute service that lets you run code for virtually any type of application or backend service without provisioning or managing servers. You can trigger Lambda from over 200 AWS services and software as a service (SaaS) applications, and only pay for what you use.

7) Amazon VPC:-

Amazon Virtual Private Cloud (Amazon VPC) gives us full control over our virtual networking environment, including resource placement, connectivity, and security. In the AWS console, we can define how our VPCs communicate with each other across accounts, Availability Zones, or AWS Regions. In the example below, network traffic is being shared between two VPCs within each Region.

8) Amazon CloudWatch:-

Amazon CloudWatch is a monitoring and observability service built for DevOps engineers, developers, site reliability engineers (SREs), IT managers, and product owners. CloudWatch provides us with data and actionable insights to monitor our applications, respond to system-wide performance changes, and optimize resource utilization. CloudWatch collects monitoring and operational data in the form of logs, metrics, and events. We get a unified view of operational health and gain complete visibility of your AWS resources, applications, and services running on AWS and on-premises. We can use CloudWatch to detect anomalous behavior in your environments, set alarms, visualize logs and metrics side by side, take automated actions, troubleshoot issues, and discover insights to keep your applications running smoothly. Amazon CloudWatch also provides real-time monitoring to Amazon's EC2 customers on their resource utilization such as CPU, disk, network and replica lag for RDS Database replicas. These are some of the most popular services provided by AWS Cloud.