PGP in Cloud Computing

Cloud Computing on AWS

Points to Remember

- **Regions** are geographical locations
- **Availability Zones** – Individual Data Centres physically separated from one another but connected internally within a Region.
  Users pick up a certain availability zone as per the **user proximity** so that the network latency is minimum.
- **Launch**ing an instance is like starting a Virtual Machine in a specific region
- **Amazon Machine Image (AMI)** is a bundle which includes the OS and OPTIONALLY can also contain other apps like Ruby, Python, Java
- **My AMI** is uploading your own VM image and importing it on AWS
  Instance name – 't, m, c, r' denotes the class of machines
  't1, m2, c1, c4' represent the generation
  Amazon gives options to set up caps for both upper and lower billing amounts.
- **Spot Instances** are unpredictable. You bid and get them but as the price rises and crosses you limit; they will be reclaimed.
  It is therefore recommended to bid for a higher price, you will still be charged as per the current price.
- **On demand instances** have a fixed price.
- **Reserved instances** are significantly cheaper but the number of instances and the period for which they are taken remains unaltered.

- **Private Cloud** – is when we create a cloud from ground up
  OR
  We take a public cloud and create a private network on it (VPC)
  Subnets ROUGHLY maps on to an Aavailbity Zone (AZ).
  By default, there is one subnet created for each AZ
  But we can customize it by making multiple subnets in an AZ for resiliency and availability.

EC2 Instance Launch

- **Terminating** an instance is releasing the VM and all storage associated with that back to AWS.
  Whereas, **stopping** an instance is releasing the VM but retaining the storage.
  We can restart a stopped instance but NOT a terminated one.
  Instances can be allocated

PGP in Cloud Computing

Shared	Dedicated	Dedicated Hosts
Run a VM on a machine	non virtualized instance	multiple VMs can be launched
Along with the VMs from other customers	Just a single instance	Multiple VMs from the same customer

- **T2 unlimited** option allows you to go beyond 100% capacity using the CPU credits that were collected when the CPU was underutilized.
- **Under advance details >> User data**
  Custom scripts can be placed which will run with ROOT privileges when the instance is launched.
- Provisioned I/O instances are expensive. They allow to choose at what rate do you want to read and write.
- If **delete on termination** is unchecked, disk remains even after an instance is terminated and data is retained.
- Security Groups are like firewall rules.
- To ssh into an ec2 instance
  ssh –I <pem file name> <username>@<public ip>
- Launching a similar instance will create an instance with same settings but not the same storage.
- To launch an exact copy of an instance with same storage as well, create a customized AMI and use it to launch.
- To launch a Windows instance >> RDP port must be open
- **Initialization checks**
  Instance is reachable (connected to the network)
  OS is able to accept traffic (pingable)

Load Balancing

- **Load Balancer (LB)** cannot span across multiple regions but they must spam across multiple AZ.
- LB sends traffic to the **Target Group (TG)** which is a basket of EC2 instances.
- Instances can be added to a TG manually or by an autoscaling group depending on the infrastructure.
- A VPC can be linked (via VPN) to an on premise network and create a hybrid cloud. The LB in the VPC can be linked to the TG of IPs in the on – premise network.
- **Session affinity** is covered in TG for customers looking to recreate their on premise datacenter in cloud.
  Autoscaling group (ASG) keeps monitoring the TG
- If cumulative CPU utilization > threshold, add instances
- If cumulative CPU utilization < threshold, remove instances
- Min and Max specification in ASG denotes the min and max number of instances that the ASG can add to a TG on its behalf.

PGP in Cloud Computing

- ASG tries to launch instances in multiple Availability Zones if multiple subnets are provided.
- If an autoscale instance is manually terminated, another instance comes up automatically whereas, if a manually launched instance goes down, nothing happens unless the CPU utilization crosses the threshold limits.
- **Amazon Machine Image** is metadata sitting on top of a snapshot of boot volume and making it bootable.
- When you create an AMI, either set up the API with your app designed to start automatically with the EC2 OR use bootstrap script as a part of the EC2 instance startup.

Identity and Access Management

- For MFA authentication, first install one of the virtual MFA apps, scan the QR code >> generates 2 authentication codes >> activate the virtual MFA
- First create an MFA in root account before starting user creation.
- AWS Organisations is where one account becomes the parent account and others the child account. All bills are passed on from child accounts to parent account.
- Steps to add a user
  - ❏ Specify the access type (CLI/AWS console)
  - ❏ Add permissions/access policies
  - ❏ Download .csv for the secret access key and access key ID

Storage

Elastic Block Store(EBS)	Instance Store	Elastic File System(EFS)	Simple Storage Service(S3)
block storage	block storage	file storage	Object storage
Retained if delete on terminate protection is enabled	Lost when an instance is terminated (ephemeral)	unmounted on instance termination	No effect of instance termination
cannot be shared or used across AZ	cannot be shared by multiple instances	can be shared by multiple instances across AZ	can be used by multiples instances across AZ

- **EBS** is great for app data, db files.

  https://aws.amazon.com/ebs/

  you need to pay for the entire EBS you take whether or not in use.

  One volume can be attached only to one EC2 within the same Availability Zone.It needs to be **mounted** on EC2 instance before it becomes usable.

PGP in Cloud Computing

- **Instance Store** is the temporary storage that lives only as long the the EC2 is running.

  It is great for temporary storage, caching, NAT, etc

- **EFS** can be used for reading and writing data but it is used predominantly for reading since there may be synchronization issues while writing.It can be shared by mounting it on multiple EC2 instances across multiple AZ.

  https://aws.amazon.com/efs/

- **Placement of functionality** rule states that use the resources from the cloud provider for what it is designed for, although it may appear that the resource might be used for other use cases, it may not be that efficient.
- In order to mount an EFS on a cluster >> Write a bootstrap script when an instance is created >> Scripts should create a mountpoint and automatically mount the EFS on the new EC2 instance that is added.
- **Snapshot** can be

  Image you can boot from

  Image of your data volume

- Before creating a snapshot, the instance is stopped and started after the snapshot is taken
- Snapshots can be used to **migrate** data from one volume to another within or across regions.
- Snapshots can act as **backups**
- **UNMOUNTING** a disk means the disk is still attached to the instance but not mounted to any place so nobody can use it

  whereas

  ATTACH is to link it to the EC2 instance, then mount it to make it usable.

- To delete a volume

  Unmount volume >> Detach volume >> Delete volume

S3

- **S3** cannot be mounted on EC2.
- **S3** can trigger events whenever files are modified,added,removed. These events can be routed to Lambda or SQS.
- **S3** has managed services like running SQL queries on csv (Athena).
- There is no limit to the number of folders you can create in S3.

  Max file size for any file is 5TB.

- S3 stores 10 copies of data for durability.
- In S3, read after writes for new objects created are strongly consistent BUT for overwrites there is an **eventual** consistency.

PGP in Cloud Computing

- S3 has the ability to **host static sites**. All content can be thrown into a bucket in S3. Specify index.html and error page, S3 gives you a DNS. The DNS address is used to access the website.CDN sitting in front of it can make the website available to the world with caching in place.
- **Storage Classes**
  - ❏ Standard - durability of 99.9999999
  - ❏ Standard IA - for data that is infrequently accessed, durability and cost lesser than standard
  - ❏ Reduced Redundancy - lesser available,durable,low performing data storage at cheaper rates

  Storage class and permissions can be changed at any point of time.

- **CORS (Cross-Origin Resource Sharing)** is a way by which the client web apps are loaded from one domain to interact with resources in a different domain.

  https://en.wikipedia.org/wiki/Cross-origin_resource_sharing

- To do **replication** across regions, versioning of S3 bucket needs to be enabled.
- After setting up replication, only the new objects created will be replicated and not the existing ones.
- **MUTEX (Mutual Exclusion)** is a locking mechanism which allows multiple instances to access a file but not simultaneously. When one instance is writing to a file on EFS, a mutex with a key is locked; other files need to wait to write to it unless the mutex is unlocked.

  https://en.wikipedia.org/wiki/Mutual_exclusion

- **Use cases of S3**

  Use Case 1

  someone uploads a pdf

  a lambda function is invoked

  it extracts all the text from pdf and feeds it into a search engine

  the pdf becomes searchable/make a DB entry

  Use Case 2

  Upload a medical transcript in S3 and that whole image can go into a lambda function

  text can be extracted using the Optical Character Recognition

  the prescription can be mailed to the doctor and the patient

  Networking

- **Virtual Private Cloud (VPC)** is the network that encompasses all Availability Zones in a given region.

  https://aws.amazon.com/vpc/

PGP in Cloud Computing

- **Subnets** belong to VPC and are sprinkled across AZ.
- One AZ can have one,more than one or NO subnet.
- AWS creates a default VPC and one subnet per AZ when you create an account in any region.
- **Peering** is making one VPC talk to another. It is a one to one peering and NOT transitive peering.Peering is possible across regions.

  https://docs.aws.amazon.com/AmazonVPC/latest/PeeringGuide/Welcome.html

- VPC can connect to on-premise network via
  - ❏ Virtual Private Network (VPN)
  - ❏ Direct Connect
- In **Direct Connect**, a dedicated cable is laid from on-premise network to Amazon direct connect center.

  https://aws.amazon.com/directconnect/

- Components in a VPC

  IGW (Internet Gateway)

  Router

  Route Tables

  NACL (Network Access Control List)

  Security Groups

  Resources + Managed Services

- A default route table and internet gateway are created when we create a **VPC**
- To connect a private subnet only to a public subnet and not to an IGW:

  1) Allow public subnet to access private subnet through security group.

  2) Take a **NAT instance** in public subnet so that the private subnet can access internet

  3) Make the entry of the NAT instance in the default route table of the VPC

- The EC2 instance we first connect in order to further ssh into the private EC2 instance is called **Jump Box or Bastion host**.

  https://en.wikipedia.org/wiki/Bastion_host

- **Network Address Translation (NAT)** is an instance that has an ability to provide tunneling.It's like a broker sitting in between a private EC2 instance and an internet gateway.

  https://en.wikipedia.org/wiki/Network_address_translation

- **Route 53** is a DNS service.

PGP in Cloud Computing

Routing logic can be one of the following :

1) Redirect the traffic to primary first, if not available then to secondary.
2) Weighted : x% of traffic goes to a particular region and y% to another region
3) Latency : sends traffic in order to achieve minimum latency.
4) Geolocation
5) Simple : There is only one route and all traffic is sent to it
6) Multioption : Consider a combination of two options (eg latency and geolocation)

- **CIDR** - Classless Inter-Domain Routing

  defines a range of available IP-addresses

  eg. 10.0.1.0/24 will have 2^{32-24} = 1024 IP addresses

  they will be 10.0.1.[0-255] = 256 addresses

  10.0.2.[0-255] = 256 addresses

  10.0.3.[0-255] = 256 addresses

  10.0.4.[0-255] = 256 addresses

  1024 addresses

- The IP addresses that are not usable are

  10.0.1.0 - Network address

  10.0.1.1 - VPC routing itself

  10.0.1.2 - Reserved by AWS

  10.0.1.3 - Reserved by AWS

  10.0.1.255 - Broadcast

- One IGW cannot be attached to multiple VPC.
- One VPC cannot be attached to more than one IGW
- **Security Groups** are stateful. They are bidirectional.Rules set for either inbound or outbound imply on both.
- **NACL(Network Access Control)** is stateless. Inbound and outbound rules need to be explicitly defined.

  https://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_Introduction.html

- NAT instance has to be managed by yourself whereas a NAT gateway is a managed service.
- NACL is for granular policies.
- In ACL, rules are implemented in ascending order or rule number
- For peering, route tables should have entries of the peers.

PGP in Cloud Computing

- The cloud provider manages certain services like database, caching, etc. and offer it as a pay as you go model.
- These services are extremely important from a business perspective because it helps business stitch together solutions quickly.
- Firewalls isolate any unauthorized traffic reaching the business applications.
- **Web Application Firewall(WAF)** will intercept each and every request that comes and makes it go through a cascading set of rules to find out any kind of unauthorized access or hack attempts.
  https://en.wikipedia.org/wiki/Web_application_firewall
- WAF rules can be **dynamic** i.e. can be modified and applied in real time.
- Pricing:
  - ❏ There is no free quota
  - ❏ $5 per ACL per month
  - ❏ $1 per rule per ACL per month
  - ❏ $0.60 per million web requests
- Requests coming to ELB are first sent to WAF so as to decide whether it will be allowed or refused to go further.
- WAF checks the following
  ACL >> Rules >> Conditions
- **Conditions** for WAF:
  - ❏ Cross site scripting – prevents users from other domains to enter your websites
  - ❏ Unauthorized SQL injection
  - ❏ Bad Bots
  - ❏ Scanner – scans and probes for unauthorized access
  - ❏ Http flooding
  - ❏ IP address whitelist/blacklist
  - ❏ Attack Protection – can lead to service outage or a significantly large elasticity problem increasing the costs.
- Filters are always **ORed** and conditions are always **ANDed**.