AWS Cloud Computing

What is meant by cloud computing ?

Cloud computing is a delivery of compute services like servers , database, networking , storage, load balancer, security etc ..over the internet.

Cloud Providers

1. AWS - Amazon Web Services
2. Azure - Microsoft
3. GCP - Google cloud Platform

Benefits of Cloud Computing

1. Cost
2. Maintenance
3. Flexibility
4. Global scale
5. Security

Cloud service offerings

1. IAAS (Infrastructure as a service)
    a. We will launch the server and all maintenance activities are done by us
2. PAAS (Platform as a service)
    a. Servers are created by the cloud provider and maintenance of the resources is also taken care by the cloud provider .
    b. We will just use it.
3. SAAS (Software as a service)
    a. We consume these services over the internet as an end user.

Cloud Deployment Models

1. Public cloud
    a. All the resources are created and managed by the cloud provider .

2. Private Cloud
    a. All the resources like networking, storage, hardware are created and managed by us
3. Hybrid Cloud
    a. Combination of both Public cloud and Private cloud.

AWS Global Infrastructure

Region

It is a location where our applications or servers have to launch .

- Mumbai
- Singapore
- Sydney
- Ireland

Availability Zone

1. Each region is divided into different availability zones.
2. If one availability zone fails to connect , we will connect to the other availability zone.
3. Availability zones are nothing but data centers , we can have many availability zones in a single region .

AWS EC2 (Elastic Compute Cloud)

1. EC2 is a virtual server in the cloud.
2. We can run our applications on EC2.

Demo Launch Linux EC2 and Connect to EC2

1. AMI (Amazon Machine Image)
    a. It is a virtual machine for launching our Ec2 instances
2. Instance type
    a. Here we can pick Like CPU and Memory for our instance
3. Configure Instance Details
    a. Here we will select networking details like VPC and Subnets etc…
4. Storage
    a. This is the harddisk of our virtual machine
5. Tags
    a. Tags enables us to give a name tor servers
6. Security groups
    a. Security groups are virtual firewall to ec2 instances

How can we connect to the EC2 instance ?

1. We can connect directly from the aws management console.
2. We can connect by using ssh clients
    a. Putty
    b. Mobaxterm

How to Install Apache Server in Linux Ec2 ?

1. sudo su - It will converts ec2user into the root user
2. yum install httpd -y (install apache server)
3. vi /var/www/html/index.html - (To create webpage)
    a. It will open one page
    b. Click letter “i”
    c. Then give some data which you have to reflect in the web page
    d. Finally save it by press esc key in keyboard and (‘:’) give ‘wq’ and click enter
4. service httpd status (To check the status of the server)
5. service httpd start (To start the server)
6. chkconfig httpd on (Auto restart the server)

AWS free tier Limitations

1. Only t2 micro is offered in the free tier .
2. You will get 750 hours of ec2 running .
3. If ec2 stops billing also should stop.
4. 30gb of storage is allowed in the free tier.

Different Types of Ec2 Instances

1. General purpose
    a. It is used for internal applications like jenkins , gitlab etc ..
    b. We can use this server and run applications like
        1. Web applications
        2. Data base
        3. Virtual servers
2. Compute Optimized
    a. These instances are designed for the high cpu’s
3. GPU optimized
    a. It is used for high graphics like video rendering and audio rendering etc ..
4. Storage optimized
    a. These instances are designed for high storage capacity .

5. Accelerated computing
    a. Accelerated computing instances use hardware accelerators, or co-processors, to perform functions, such as floating point number calculations, graphics processing, or data pattern matching, more efficiently than is possible in software running on CPUs

EC2 Instance Purchasing Options

AWS offers different purchase options to optimize the cost, as a solutions architect we should be aware of purchase options to optimize the cost in the project.

1. On-demand Instances
2. Reserved Instances
3. Spot Instances
4. Dedicated Hosts
5. Scheduled Instances
6. Etc.

On-Demand Instances

1. Per hour billing is higher than other purchase options
2. You can launch it any time and terminate it anytime.
3. Billing stops when you stop/terminate an instance.
4. We use this when we do not have long term commitments
5. We use it for doing POC (Proof Of Concepts)
6. For example flipkart announces a big billion day for 7 days, we want additional capacity(ec2) for 7 days, on-demand is suitable.
7. By default what we use is on demand.

Reserved Instances

1. If we know we want instances for the long term, this is a good option.
2. We give long term commitment. Like 1 year or 3 years in return we get significant discounts
3. You can save upto 50% as a discounts
4. After you make a reservation you can’t cancel it. But you can sell it in the AWS marketplace.
5. Billing will not stop even if you stop all instances.
6. There are different payment options
    a. All upfront
    b. Partial upfront
    c. No upfront

Spot Instances

1. Spot instances are allocated from unused capacity from AWS datacenter
2. It offers upto 90% discounts
3. It has spot interruption, when spot interruption occurs, amazon takes the instance by giving 2 minutes notice.
4. 2 minutes notice helps to grab a log file on the instance before it is terminated.
5. We can't deploy applications totally depending on spot, we have to mix and match, spot, reserved/ondemand.
6. When spot interruption can occur
    a. When spot price goes high
    b. If underlying hardware fails
    c. If AWS is running out of capacity

Dedicated Hosts

1. Dedicated Hosts are used to bring your own license (BYOL).
2. If a customer is migrating to the cloud and he has licenses with him and those licenses should be used in the cloud. Then go with Dedicated Hosts.
3. AWS allocates a complete physical host, and all instances will be placed on this physical host.

Scheduled Instances

- This is similar to reserved, but this is used for batch jobs that run in specific schedules
- With Scheduled Reserved Instances, you can reserve capacity that is scheduled to recur daily, weekly, or monthly, with a specified start time and duration, for a one-year term

AMI (Amazon Machine Image)

It is a template for launching a virtual machine

We can share our AMI in two ways

- Cross Region

  Sharing of AMI internally from one region to another region

- Cross Account

  Sharing of AMI from one account to another account

Note : If we delete Original AMI all AMI's will be deleted automatically

EBS (Elastic Block Store)

1. It is nothing but hard disk or storage of the server.
2. We can attach one or more volumes to Ec2
3. Volume is specific to the availability zone (AZ), both instances and volumes should be in the same availability zone .
4. We can detach volume from one instance to another instance .

Demo creating Volume and attach to it

1. We have to create volume in the same availability zone .
2. We have to attach the volume to the ec2 instance.
3. ssh into the server.
    a. sudo su
    b. lsblk (to check the list of volumes in the server)
    c. mkfs.ext4 /dev/xvdf (creates a file system)
    d. mkdir /gokul (creates a directory)
    e. mount /dev/xvdf /gokul (it mounts the volume to the directory)
    f. Finally check once again with lsblk

How to automount Volume