Session 1 : Introduction to Elastic Compute Cloud(EC2)

Elastic Compute Cloud : Amazon EC2 provides scalable computing capacity in the AWS Cloud

• We can use Amazon EC2 to launch as many or as few Virtual Servers as we need,configure security,Networking and manage Storage
• Amazon EC2 enables us to Scale Up or Scale Down the Instances Capacity
• Amazon EC2 has two Storage options i.e EBS & Instance Store
• Preconfigured templates are available known as Amazon Machine Images
• By default when we create an AWS account with amazon, our account is limited to a max of 20 instances per ec2 region with two default High I/O Instances

Types of EC2 Instances :

1.General Purpose : Balanced Memory and CPU
2.Compute Optimized : More CPU than Ram
3.Memory Optimized : More Ram
4.Accelerated Computing/GPU : Graphics Optimized
5.Storage Optimized : Low Latency
6.High Memory Optimized : High Ram,Nitro System
7.Previous Generations

Session 2 : General Purpose EC2 Instances

General Purpose Instances : These provide a balance of Compute,Memory and Networking Resources and can be used for a variety of Workloads

• 3Series in General Purpose Instances

1.A Series(Medium and Large) : A1
2.M Series(Large) : M4,M5,M5a,M5ad,M5d
3.T Series(Large,Medium,Small,Nano) : T2,T3,T3a.....T2 micro is eligible for Free Tier.....micro comes under nano type

• Available in Four Sizes : Nano,Small,Medium,Large

1.A Series :

a.A1 Instances : These are ideally suited for scale-out Workloads that are supported by ARM Ecosystem

• These instances are well suited for web server,Containerized microservices,Caching Fleets,Distributed Data Stores,Applications that require ARM Instruction Set

2.M Series :

a.M4 Instances : These Features a Custom Intel Xeon E5-2676 v3 Haswell processors optimized specifically for EC2
Capacities : VCPU : 2 to 40(MAX).....Ram : 8 to 160GB(MAX).....Instance Storage : EBS Only

b.M5,M5a,M5ad,M5d Instances : These Instances provide an ideal cloud infra,offering a balance of Compute,Memory,Networking Resources for a broad range of Applications
• Capacities : VCPU : 2 to 96(MAX).....Ram : 8 to 384GB(MAX).....Instance Storage : EBS and NVME SSD
• Used in Gaming Servers,Web Servers,Medium and Small Databases

3.T Series :

a.T2,T3,T3a Instances : Provides a Baseline level of CPU Performance with the ability to burst to a heigher level,when required by our Workload
• An unlimeted instances can sustain high cpu performance for any period of time and whenever required
• Capacities : VCPU : 2 to 8(MAX).....Ram : 0.5GB to 32GB
• Used for Websites and Web App,Code Repos,Developement,build,Test and for Microservices

Session 3 : Compute Optimized EC2 Instances

Compute Optimized Instances : Are ideal for compute-bound applications that benefit from high performance processors

• It only has C Series which has C4,C5,C5n

1.C Series

a.C4 Instances : These are optimized for compute intense workloads and deliver very cost effective and high performance at a low price per compute ratio
• Capacities : VCPU : 2 to 36....RAM : 3.75 to 65 GB.....Storage : EBS Only .....Network Bandwidth : 10GBPS
• Used for Web Server,Batch Processing,MMO Gaming,Video Encoding

b. C5 Instances : These are optimized for compute intense workloads and deliver very cost effective and high performance at a low price per compute ratio and are powered by NITRO SYSTEMS
• Capacities : VCPU : 2 to 72....RAM : 4 to 192 GB.....Storage : EBS & NVMe SSD .....Network Bandwidth : 25GBPS
• Used for High Performance Web Servers,Gaming,Video Encoding

• C5 support max 25 EBS Volumes and uses elastic network adapter and new EC2 Hypervisor(AWS Nitro System)

Session 4 : Memory Optimized EC2 Instances

Memory Optimized Instances : These are designed to deliver fast performance for workloads that process large data sets in memory

• These consists of 3 serieses....they are R,X,Z

1.R Series : R4,R5,R5a,R5ad,R5d

• High Performance,Relational(My Sql),Nosql(MangoDB,Cassandra) databases
• Distributed web scale cache stores that provide in-memory caching of key value type data
• Capacities : VCPU : 2 to 96.....RAM : 16 to 768GB....Instance Storage : EBS & NVMe SSD
• Used in Financial Services,Hadoop etc...

2.X Series : X1,X1e

• Well suited for High Performance Database,Memory intensive enterprise applications,Relational database Workload,SAP HANA, Electronic Design Automation
• Capacities : VCPU : 4 to 128.....RAM : 122 to 3904GB....Instance Storage : NVMe SSD

3.Z Series : Z1d

• High Frequency Z1d deliver a sustained all core frequency of upto 4.0GHZ,the faster of any cloud Instances
• AWS Nitro System,Xeon Processor upto 1.8TB of instance Storage
• Capacities : VCPU : 2 to 48.....RAM : 16 to 384GB....Instance Storage : NVMe SSD
• Use cases are Electronic design automation,certain database workloads with high per core licensing costs

Session 5 : Storage Optimized EC2 Instances

Storage Optimized Instances :These are designed for workloads that require high sequential read write access to very large data sets on local storage
• These are optimized to deliver tens of thousands of low latency,random I/O operations per second(IOPS) to application
• These consists of 3 Serieses...they are D,H,I

1.D Series : D2

• Well suited for massive parallel processing(MPP) data warehouse,MAP reduce and hadoop distributed computing,Log or data processing app
• Capacities : VCPU :4 to 36.....RAM :30.5 to 244GB.....Storage : SSD

2.H Series : H1

• This family features 16TB of HDD based local storage,high disk throughput and balance
• Well suited for app requiring sequential access to large amounts of data on direct attached instance storage
• Applications that require high throughput access to large quantities of data
• Capacities : VCPU :8 to 64.....RAM :32 to 256GB.....Storage : HDD

3.I Series : I3 and I3en

• Well suited for high frequency online transaction processing system(OLTP),Relational Databases(No Sql,Distributed file system,data warehousing application)
• Capacities : VCPU :2 to 96.....RAM :16 to 768GB.....Storage : NVMe SSD.....Networking Performance : 25 to 100 GBPS.....Sequential Throughput : Read-16Gb/-s.....Write-6.4GB/s(I3).....Write-8GB/s(I3en)

Session 6 : Accelerated Computing EC2 Instances

Accelerated Computing Instance : These use hardware accelerators or co-processors to perform some functions such as floating point number calculations,graphics processing or data pattern, matching more efficiently than is possible in software running on CPU's

• These consists of 3 Serieses...they are P,G,F

1.F Series : F1

• These offer customizable hardware acceleration with field programmable gate arrays(FPGA)
• Each FPGA contains 2.5 million logic elements and 6800 DSP engines
• Designed to accelerate computationally intensive alogorithms,such as data flow or highly parallel operations
• F1 provides local NVMe SSD storage
• Capacities : VCPU : 8to 64.....RAM : 122 to 976GB.....FPGA : 1 to 8.....Storage : NVMe SSD
• Used in Financial Analytics,genomics research,real time video recording & big data research

2.P Series : P2 & P3

• It uses NVIDIA Tesla GPU's,provide high bandwidth networking,upto 32Gb of memory per GPU,which makes them Ideal for deep learning & computational fluid dynamics
• P2 Instances : VCPU : 4 to 64,GPU : 1 to 16, RAM : 61 to 732GB,GPU RAm : 12 to 192GB,Network B/W : 25GBPS
• P3 Instances : VCPU : 8 to 96,GPU : 1 to 8, RAM : 61 to 768GB,GPU RAm : 12 to 192GB,Network B/W : 25GBPS...Storage : SSD & EBS
• Used in Machine learning,databases,sesimic analysis,genomics,molecular modeling,AI,Deep Learning
• P3 supports CUDA9 & OpenCL API
• P2 supports CUDA9 and Open CL 1.2

3.G Series : G2 & G3

• Optimized for Graphics Intensive Applications
• Well suited for app like 3D Visualization
• G3 instances use NVIDIA Tesla m60 GPU and provide a cost effective high performance platform for graphics application
• Capacities : VCPU : 4 to 64,GPU : 1 to 4, RAM : 30.5 to 488GB,GPU RAM : 8 to 32GB,Network B/W : 25GBPS
• Used in Video creation services,3D visualization,Streaming Graphics intensive applications

Session 7 : High Memory EC2 Instances and Previous Generation EC2 Instances

High Memory Instance : These are purpose built to run large-in-memory databases & including production developments of SAP HANA in the cloud
• These instances are bare metal instances and do not run on a hypervisor
• Only available under dedicated host purchasing category(For Min 3 Years term)
• OS directly on the Hardware

Features :

• Latest Intel Gen Intel Xeon Pentium 8176M Processor
• 6,9,12TB of instance memory,the largest of any EC2 instances
• Powered by the AWS nitro system,a combination of dedicated hardware & Lightweight hypervisor
• Bare metal performance with direct access to host hardware
• EBS optimized by default at no additional cost

• This consists of U series,,,,,they are U-6tb.metal,U-8tb.metal,U-12tb.metal
• Network performance is 25Gb/s and dedicated EBS bandwidth-14GBPS
• Each instance offer 448 logical processors

Previous Gen Instances : T1,M1,C1,CC2,M2,CR1,CG1,i2,HS1,M3,C3 and R3
• These are not deleted instances and we can still purchase these instances

Session 8 : EC2 Purchasing Options

EC2 Instances Purchasing Options :

1.On-demand :
2.Dedicated Instances :
3.Schedule Instances :
4.Reserved Instances(RI) : Standard RI,Convertible RI,Scheduled RI

5.Dedicated Host :
6.Spot Instances :

• There are three ways to pay for EC2 Instance i.e On-Demand,Reserved Instance and Spot Instances
• Dedicated host and Dedicated instances costs are calculated as per On-Demand instance costs and Scheduled instances are billed as per reserved instance costs
• We can also pay for dedicated host which provides us with EC2 instance capacity on physical server dedicated for our use

On-Demand Instances :

• These are virtual servers that run in AWS or AWS Relational Database Server(RDS) and are purchased at a fixed rate per hour
• AWS recommends using these instances for applications with short term irregular workloads that cannot be uninterrupted
• They are also suitable for use during testing and development of apps on EC2
• With these we can only pay for EC2 Instances we use
• The use of these instances free from the cost and complexities of planning,purchasing and maintaining hardware and transforms what are commonly large fixed costs into much smaller variable cost
• Pricing is per instance-hour consumed for each instance,from the time an instance is launched until it is terminated or stopped
• Each partial instance consumed will be billed per second for linux instances and as a full hour for all other instance types

Dedicated Instances :

• Dedicated instances are run in a vpc on hardware that is dedicated to a single customer
• Our dedicated instances are physically isolated at the host hardware level from instances that belong to other AWS Account
• These instances may share hardware with other instances from the same aws account that are not dedicated instnaces
• Pay for dedicated instances is based on on-demand and save upto 70% by purchasing reserved instances and upto 90% by purchasing spot instances when compared to Dedicated instances

Dedicated Host :

• An amazon EC2 dedicated host is a physical server with ec2 instances capacity fully dedicated for our use
• Dedicated hosts can help us address compliance requirements and reduce costs by allowing us to use our existing server bound software requirements
• Pay for a physical host fully dedicated to running our instances and bring our existing per-socket,per-core,per-vm software license to reduce cost

Spot Instances :

• These let us take advantage of unused ec2 capacity in the aws cloud.These are available upto 90% discount compared to on-demand prices
• We can use these for various test & development workloads
• we also have option to hibernate,stop or terminate our spot instances when ec2 reclaims the capacity back with two minutes of notice
• These get interrupted when actually ec2 capacity requirement increases(On-demand and reserved instances) and amazon reclaims the space with 2 min notification.
• These can also get interrupted when the spot price raises above our choosen max spot price

Schedule Instances :

• These enables us to purchase capacity reservation that recur on a daily,weekly or monthly basis

with a specified start time and duration,for a one-yaer term
• We reserve the capacity in advance so that we know it is available when we need it
• We pay for the time that the instances are scheduled,even if we do not use them
• Schedule instances are a good choice for workloads that do not run continuously but do run on a regular basis
• Purchase instances that are always available on the specified recurring schedule,for a one-year term
• Eg: We can use these instances for an application that runs during business hours or for batch processing that run at the end of the week

Reserved Instances :

• These provide a significant discount upto 70% comapred to on-demand pricing and provide capacity reservation when used in a specific availability zone
• Reserved instances give us the option to reserve a DB instances for a one or three years term an in turn receive a significant discount comapred on demand instances pricing
• 3 Types of RI....Standard,Convertible and Scheduled RI

• *Standard RI :* These provide the most significant discount upto 75% off on-demand and are best suited for steady state usage
• *Convertible RI :* These provide a discount upto 54% and the capability to change the attribute of RI as long as the exchange results in the creation of reserved instances of greater or equal values
• *Scheduled RI :* These are available to launch within in the time window we reserve....Same as Scheduled Instancs
• We cannot transfer a convertible or standard RI from one region to another region
• We can change the config of convertible RI from ec2 management console or get reserved instance management quota API
• There is no extra charge for converting from one config to another config but we need to pay the cost as per the changed config

Session 9 : EC2 Access,Status Check and EC2 Meta Data

EC2 Access Data :

• To access instances,we need a key and key pair name
• We can download the private key only once
• The public key is saved by aws to match it to the key pair name and private key when we try to login to the instance
• Without key pair we cannot access isntances via RDP or SSH(Linux)
• There is a 20 ec2 instances soft limit per region ,and we can submit requst to aws to increase limit

EC2 Status Check :

• By default aws ec2 instances performs automated status checks every 1 min
• This is done on every running ec2 instances to identify any H/W or software issues
• Status check is built into the aws ec2 instance
• They cannot be configured,deleted or disabled
• EC2 services can send its metric data to aws cloudwatch every 5 min (enabled by default)
• Enabled detailed monitoring is chargeable and sends metrics in every 1 min
• We can not charged for ec2 instances if they are stopped but attached ebs volumes get charged

When we stoppes an ebs backed ec2 instance :

• Instances perform a shutdown
• state changes from running to stopping
• ebs volumes remain attached to the instance

• any data cached in ram or instance store volume is gone
• instances retain its private ipv4 address and any ipv6 address
• instances releases its public ipv4 address back to aws pool
• Instances retain its elastic ip addresses

EC2 Terminate :

• When we terminate a running instance the instance state changes from running to shutting down and then to terminated
• During the shutting down and terminated states,we do not incur charges
• By default ebs root devices volumes are deleted automatically when the ec2 instances are terminated
• Any additional (non boot/boot) volumes attached to the instances by default,persist after the instances is terminated
• We can modify both behaviours by modifying the 'delete on termination' attribute of any ebs volumes during instances launch or while running
• Enable ec2 termination protection against accidental termination

Ec2 Metadata :

• This is instance data that we can use to configure or manage the instance
• Eg : ipv4 addr,ipv6 addr,dns hostname,AMI-Id,Instance id,instance type,local hostname,public keys,security groups
• Metadata can be only viewed from within the instance itself i.e we need to login to the instance
• Metadata is not protected by encryption,anyone that has access to the instance can view this data
• To view instance metadata use GET http://169.254.169.254/latest/metadata

Instances User Data :

• Data supplied by the user at instance launch in the form of a script to be executed during the instance boot
• User data is limited to 16kb
• We can change user dat by stopping ec2 first
• User data is not encrypted

EC2 Bare Metal Instances :

• Non virtualized environment
• Operating Systems runs directly on hardware
• Suitable for licensing restricted tier 1 business critical application
• i3 metal,i5 metal,r5metal,z1d metal,u-6tb1.metal

Elastic Block Storage : EBS backed instance

• We can easily replicate between availability zones with snapshots etc..
• EBS volumes attached at launch are deleted when instance terminate
• EBS volumes attached to a running instance are not deleted when instance is terminated but are deattached with data interact
• EBS is network attached storage

Instance Storage : Instance backed storage

• Physically attach to the host server
• Data not lost when os is rebooted
• Data is lost when underlying drive fails,instance is stopped or terminated
• We can not attach or detach to another instance
• Do not rely on for valuable long term data

LABS :

Session 10 : Creating Windows Server in AWS EC2

Session 11 : Install Webserver IIS and Creating Webpage in Win Server

Session 12 : Attaching extra volumes in existing Win Machine

Session 13 : Creating Linux Machine AWS EC2

Session 14 : Retriving Metadata of Linux Machine