

Chapter 5 : Elastic Compute Cloud (EC2)

Rahul Daware

June 3, 2018

Contents

1 EC2 Options	3
1.1 On Demand Instances	3
1.2 Reserved Instances	3
1.3 Spot Instances	3
1.4 Dedicated Hosts	3
2 EC2 Instance Types	4
3 Lab Summary	4
4 Security Group Lab	4
5 Volumes and Snapshots	5
6 Snapshots of Root Device Volumes	5
7 Volumes vs Snapshots	5
8 AMI Types (EBS vs Instance Store)	5
9 EBS vs Instance Store - Exam Tips	6
10 Elastic Load Balancers	6
11 CloudWatch	6
12 Placement Group	6
12.1 Clustered Placement Group	6
12.2 Spread Placement Group	6
12.3 Points to remember	7
13 EFS - Elastic File System	7
14 AWS Lambda	7
14.1 How is Lambda priced?	7
14.2 Why is Lambda cool?	8
14.3 Lambda - Exam Tips	8

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides computer capacity in the cloud. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change.

Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you will actually use. Amazon EC2 provides developers the tools to build failure resilient applications and isolate themselves from common failure scenarios.

1 EC2 Options

- On Demand - allow you to pay a fixed rate by the hour (or by the second) with no commitment
- Reserved - Provide you with a capacity reservation, and offer a significant discount on the hourly charge for an instance. 1 Year or 3 Year terms
- Spot - Enables you to bid whatever price you want for instance capacity, providing for even greater savings if your applications have flexible start and end times
- Dedicated Hosts - Physical EC2 server dedicated for your use. Dedicated hosts can help you reduce costs by allowing you to use your existing server-bound software licenses

1.1 On Demand Instances

- Users that want the low cost and flexibility of Amazon EC2 without any upfront payment or long term commitment
- Applications with short term, spiky, or unpredictable workloads that cannot be interrupted
- Applications being developed or tested on Amazon EC2 for the first time

1.2 Reserved Instances

- Applications with steady state or predictable usage
- Applications that require reserved capacity
- Users able to make upfront payments to reduce their total computing costs even further
 - Standard Reserved Instances (Up to 75% off on demand)
 - Convertible Reserved Instances (Up to 54% off on demand) capability to change the attributes of the RI as long as the exchange results in the creation of reserved instances of equal or greater value
 - Scheduled Reserved Instances are available to launch within the time windows you reserve. This option allows you to match your capacity reservation to a predictable recurring schedule that only requires a fraction of a day, week, or a month.

1.3 Spot Instances

- Applications that have flexible start and end times
- Application that are only feasible at very low compute prices
- Users with urgent computing needs for large amounts of additional capacity

1.4 Dedicated Hosts

- Useful for regulatory requirements that may not support multi tenant virtualization
- Great for licensing which does not support multi tenancy or cloud deployments
- Can be purchased on Demand (hourly)
- Can be purchased as a reservation for up to 70% off on the on demand price

2 EC2 Instance Types

Family	Specialty	Use Case
D2	Dense Storage	Fileservers/Data Warehousing/Hadoop
R4	Memory Optimized	Memory Intensive Apps/DBs
M4	General Purpose	Application Servers
C4	Compute Optimized	CPU Intensive Apps/DBs
G2	Graphics Intensive	Video Encoding/3D Application Streaming
I2	High Speed Storage	NoSQL DBs, Data Warehousing etc
F1	Field Programmable Gate Array	Hardware acceleration for your code
T2	Low Cost, General Purpose	Web Servers, Small DBs
P2	Graphics/General purpose GPU	Machine Learning, Bit Coin Mining etc
X1	Memory Optimized	SAP HAN, Apache Spark, etc

- D - Density
- R - RAM
- M - Main Choice for General Purpose Apps
- C - Compute
- G - Graphics
- I - IOPS
- F - FPGA
- T - Cheap General purpose (think T2 Micro)
- P - Graphics (think P1s)
- X - Extreme Memory

3 Lab Summary

- Termination Protection is turned off by default, you must turn it on.
- On an EBS backed instance, the default action is for the root EBS volume to be deleted when the instance is terminated.
- EBS root volumes of your default AMIs cannot be encrypted. You can also use a third party tool to encrypt the root volume, or this can be done when creating AMIs in the AWS console or using the API
- Additional volumes can be encrypted.

4 Security Group Lab

- All inbound traffic is blocked by default
- All outbound traffic is allowed
- Changes to security groups take effect immediately
- You can have any number of EC2 instances within a security group
- Security groups are stateful. If you create an inbound rule allowing traffic in, that traffic is automatically allowed back out again
- You can have multiple security groups attached to EC2 instances
- You cannot block specific IP addresses using security groups, instead use Network Access Control Lists

5 Volumes and Snapshots

- Volumes exist on EBS - Virtual Hard Disk
- Snapshots exist on S3
- Snapshots are point in time copies of Volumes
- Snapshots are incremental - this means that only the blocks that have changes your last snapshot are moved to S3
- If this is your first snapshot, then it will take some time to create
- To create a snapshot for Amazon EBS volumes that serve as root devices, you should stop the instance before taking the snapshot. However, you can take a snap while the instance is running
- You can create AMIs from EBS-backed instances and snapshots
- You can change EBS volume sizes on the fly, including changing the size and storage type
- Volumes will always be in the same availability zone as the EC2 instance
- To move an EC2 volume from one AZ/region to another, take a snap or an image of it, then copy it to the new AZ/region
- Snapshots of encrypted volumes are encrypted automatically
- Volumes restored from encrypted snapshots are encrypted automatically
- You can share snapshots, but only if they are unencrypted. These snapshots can be shared with other AWS accounts or made public

6 Snapshots of Root Device Volumes

To create a snapshot for Amazon EBS volumes that serve as root devices, you should stop the instance before taking the snapshot

7 Volumes vs Snapshots

- Snapshots of encrypted volumes are encrypted automatically
- Volumes restored from encrypted snapshots are encrypted automatically
- You can share snapshots, but only if they are encrypted. These snapshots can be shared with other AWS accounts or made public

8 AMI Types (EBS vs Instance Store)

You can select your AMI based on:

- Region (see regions and availability zones)
- Operating System
- Architecture (32 bit or 64 bit)
- Launch permissions
- Storage of root device (Root Device Volume) - Instance Store (Ephemeral Storage) or EBS Backed Volumes

All AMIs are categorized as either backed by Amazon EBS or backed by instance store. For EBS volumes, the root device for an instance launched from the AMI is an Amazon EBS volume created from an Amazon EBS snapshot. For instance store volumes, the root device for an instance launched from the AMI is an instance store volume created from a template stored in Amazon S3.

9 EBS vs Instance Store - Exam Tips

- Instance store volumes are sometimes called ephemeral storage
- Instance store volumes cannot be stopped. If the underlying host fails, you will lose your data
- EBS backed instances can be stopped. You will not lose the data on this instance if it is stopped
- You can reboot both, you will not lose your data
- By default, both root volumes will be deleted on termination, however with EBS volumes, you can tell AWS to keep the root device volume.

10 Elastic Load Balancers

- Instances monitored by ELB are reported as in service or out of service
- Health checks check the instance health by talking to it
- Have their own DNS name. You are never given an IP address
- Read the ELB FAQ for classic load balancers

11 CloudWatch

Standard monitoring is always 5 mins. Detailed monitoring is 1 min.

What can I do with CloudWatch?

- Dashboards - Creates awesome dashboards to see what is happening with our AWS environments
- Alarms - Allows you to set Alarms that notify you when particular thresholds are hit
- Events - CloudWatch events helps you respond to state changes in your AWS resources
- Logs - CloudWatch logs helps you to aggregate, monitor, and store logs

12 Placement Group

Two Types of placement groups :

- Clustered Placement Group
- Spread Placement Group

12.1 Clustered Placement Group

A clustered placement group is a grouping of instances within a single availability zone. Placement groups are recommended for applications that need low network latency, high network throughput, or both. Only certain instances can be launched in to a clustered placement group.

12.2 Spread Placement Group

A spread placement group is a group of instances that are placed on distinct underlying hardware. Spread placement groups are recommended for applications that have a small number of critical instances that should be kept separate from each other.

12.3 Points to remember

- A clustered placement group can't span multiple availability zones.
- A spread placement group can
- The name you specify for a placement group must be unique within your AWS account
- Only certain types of instances can be launched in a placement group (Compute optimized, GPU, Memory optimized, Storage Optimized)
- AWS recommend homogeneous instances within placement groups
- You cannot merge placement groups
- You cannot move an existing instance into a placement group. You can create an AMI from your existing instance, and then launch a new instance from the AMI into a placement group

13 EFS - Elastic File System

Amazon Elastic File System (EFS) is a file storage service for Amazon Elastic Compute cloud (Amazon EC2) instances. Amazon EFS is easy to use and provides a simple interface that allows you to create and configure file systems quickly and easily. With Amazon EFS, storage capacity is elastic, growing and shrinking automatically as you add and remove files, so your application have the storage they need, when they need it.

Features:

- Supports the network file system version 4 (NFSv4) protocol
- You only pay for the storage you use (no pre-provisioning required)
- Can scale up to the petabytes
- Can support thousands of concurrent NFS connections
- Data is stored across multiple AZs within a region
- Read after write latency

14 AWS Lambda

AWS Lambda is a compute service where you can upload your code and create a lambda function. AWS lambda take care of provisioning and managing the servers that use to run the code. You don't have to worry about the operating systems, patching, scaling, etc. You can use Lambda in the following ways. As an event driven compute service where AWS lambda runs your code in response to events. These events could be changes to data in an Amazon S3 bucket or an Amazon dynamoDB table. As a compute service to run your code in response to HTTP requests using Amazon aPI gatwat or API calles made using AWS SDKs.

14.1 How is Lambda priced?

- Number of requests - First 1 million requests are free. \$0.20 per 1 million requests thereafter.
- Duration - Duration is calculated from the time your code begins executing until it returns or otherwise terminates, roynoded up to the nearest 100ms. The price depends upon the amount of memory you allocate to your function. You are charged \$0.00001667 for every GB-second used. Function cannot execute for more than 5 mins

14.2 Why is Lambda cool?

- No servers
- Continuous scaling
- super super super cheap

14.3 Lambda - Exam Tips

- Lambda scales out automatically
- Lambda functions are independent, 1 event = 1 function
- Lambda is serverless
- Know what services are serverless
- Lambda functions can trigger other lambda functions, 1 event can = x functions if functions trigger other functions
- Architectures can get extremely complicated, AWS X-ray allows you to debug what is happening
- Lambda can do things globally, you can use to back up S3 buckets to other S3 buckets etc
- Know your triggers