Chapter 5 : Elastic Compute Cloud (EC2)

Rahul Daware

May 29, 2018

Contents

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 an boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change.

Aamzon 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 yo 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 u 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 Intensize 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 Pics)
- 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 tfor 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 encrpted 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