Enterprise Computing Assignment #1

Due on June 21, 2021

Basanta Joshi, Ph.D.

Dipak Poudel

Ashlesh Pandey

PUL074BEX007

Contents

List of Figures

List of Abbreviations

AMI Amazon Machine Image.

AWS Amazon Web Services.

EBS Elastic Block Store.

EC2 Elastic Compute Cloud.

GCE Google Cloud Engine.

GCP Google Cloud Platform.

IaaS Infrastructure as a Service.

IAM Identity and Access Management.

IT Information Technology.

PaaS Platform as a Service.

SaaS Software as a Service.

VPC Virtual Private Cloud.

1 Cloud Computing

The on-demand delivery of Information Technology (IT) resources over the shared internet is known as cloud computing. It is a concept that allows outsourcing of different IT resources such as storage, processing capabilities, or even software via the network that uses the pay-as-you-go model. Advancements in cloud computing have allowed users to shift away from buying and maintaining the data centers, servers, and other resources to accessing such resources as per their requirement from cloud vendors such as Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP). Different sized organizations in the industry are turning their heads towards cloud computing for use cases that range from disaster recovery, data backup, virtual desktop environments to big data analysis and software development.

1.1 Types of Cloud Computing Model

1.1.1 Infrastructure as a Service (IaaS)

Infrastructure as a Service (IaaS) cloud computing model provides computing infrastructure such as servers, network equipment, storage, operating systems through a technology called virtualization. This allows clients that buy IaaS on-demand services complete control over the bought infrastructure such that they don't have to think about maintaining the actual physical hardware. It is the most flexible model as clients have complete control over their infrastructure while still employing the pay-as-yo-go model. Amazon Web Services (AWS), Google Cloud Engine (GCE) are some common examples.

1.1.2 Platform as a Service (PaaS)

Platform as a Service (PaaS) cloud computing model provides platform for software creation over the internet. This allows the client company to let its developers focus on software devel-

opment such that they don't have to focus on the underlying operating system, or any infrastructure. It also provides different kinds of services that are helpful for development and testing of applications. AWS Elastic Beanstalk, Heroku are some common examples.

1.1.3 Software as a Service (SaaS)

Software as a Service (SaaS) cloud computing model provides cloud applications managed by third-party distributors over the internet. Mostly small companies and short-term projects that are looking for instant IT solutions use SaaS. Dropbox, Google Workspace, and GoToMeeting are some common examples.

1.2 Benefits of Cloud Computing

1.2.1 Variable Expense

Companies that employ their services using cloud services trade the overhead cost of buying and maintaining infrastructures with variable expense with the pay-as-you-go model.

1.2.2 On-demand Scaling of Resources

Clients can avoid under or over estimating the capacity they need and move towards a better scaled resources using auto scaling services.

1.2.3 Increased Speed of Deployment

Since the IT resources required are available at the push of a button, the deployment process is greatly sped up.

1.2.4 Economies of Scale

With the increase in number of users, the proportional cost saving using the pay-as-you-go model drastically increases.

1.2.5 Global Reach without much Hassle

Clients can deploy their applications to multiple AWS regions without much hassle. This improves their global reach and they can have servers in different geographical regions up and running in few minutes. This ensures low latency for the application hence boosting consumer experience.

1.2.6 No Maintenance Cost

Clients can focus on the delivery of their applications rather than maintaining the infrastructure that are essential. This greatly reduces their maintenance expenditure.

2 AWS Global Infrastructure

In order to deliver highly reliable, scalable cloud services over a secure connection, Amazon has segregated it's global infrastructures with 81 availability zones in 25 actual global regions around the globe.

2.1 AWS Regions

These are physical geographical locations where data replication across regions depends on the clients need. Currently there are 25 AWS regions that themselves house one or more availability zones. Each region is completely independent of one another and pricing vary in these regions.

2.2 Availability Zones

These are isolated AWS infrastructures inside an AWS region that are connected via high bandwidth private networks. Currently there are 81 availability zones that themselves house one or more data centers. The availability zones ensure isolation from natural calamities. 21 new availability zones in 7 AWS regions have been announced recently.

2.3 Data Centers

These are physical data centers that store the actual data. Clients can't select data centers instead can select availability zones to deploy their application.

3 Amazon Elastic Compute Cloud (EC2)

Elastic Compute Cloud (EC2) is a service available on the Amazon console dashboard that allows clients to create virtual machines, generally known as EC2 instances. In terms of cloud computing types, an EC2 can be regarded as an IaaS since the client can choose the AMI, and capacity of servers. The said EC2 instances allow clients to host virtual machines on which they can deploy applications in somewhat a similar fashion as an on-premise environment. Few use-cases where EC2 instances are used are as application, web, game, media, mail, database servers. EC2 instances can be launched in the different possible availability zones. The breakdown for Amazon EC2 is as,

- **Elastic :** Since number of instances can be auto-scaled.
- **Compute :** Since most use-cases of EC2 include the use of compute resources.
- **Cloud :** Since EC2 instances are available as cloud service.

3.1 Choose an Amazon Machine Image

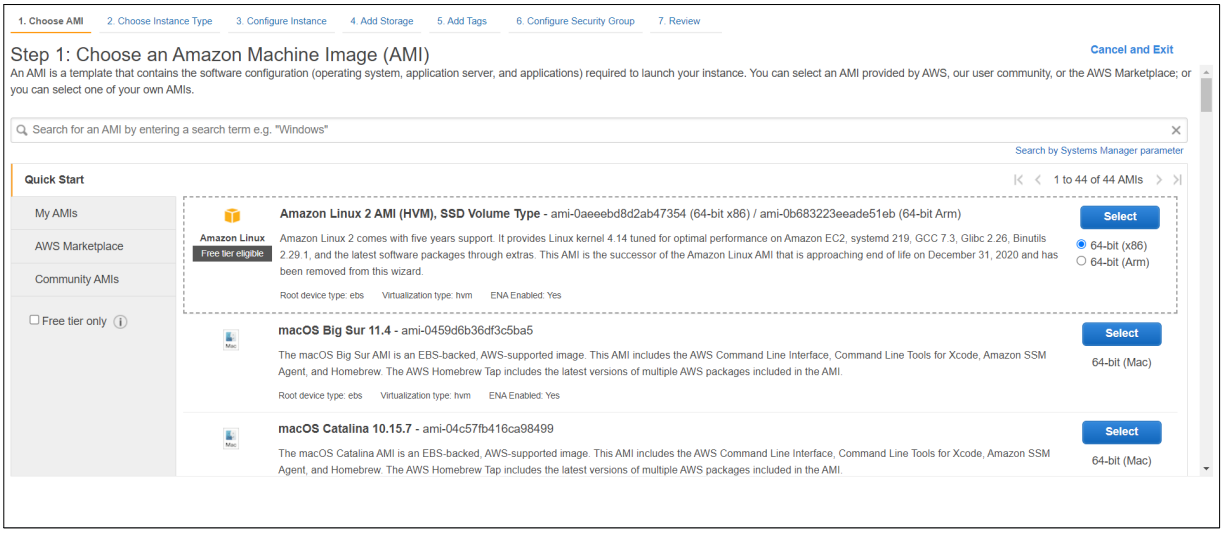

Figure 1: Choose an Amazon Machine Image

This is the first step in creation of an EC2 instance. This step involves selection of an AMI based on which the desired EC2 instance is created. Amazon Machine Image (AMI) is an image that contains the configurations such as operating system, required application servers. AWS allows selection of AMI provided by AWS, or the ones created by the vast user community or some available on the AWS marketplace. For this example, Amazon Linux is selected as the AMI.

3.2 Choose an Instance Type

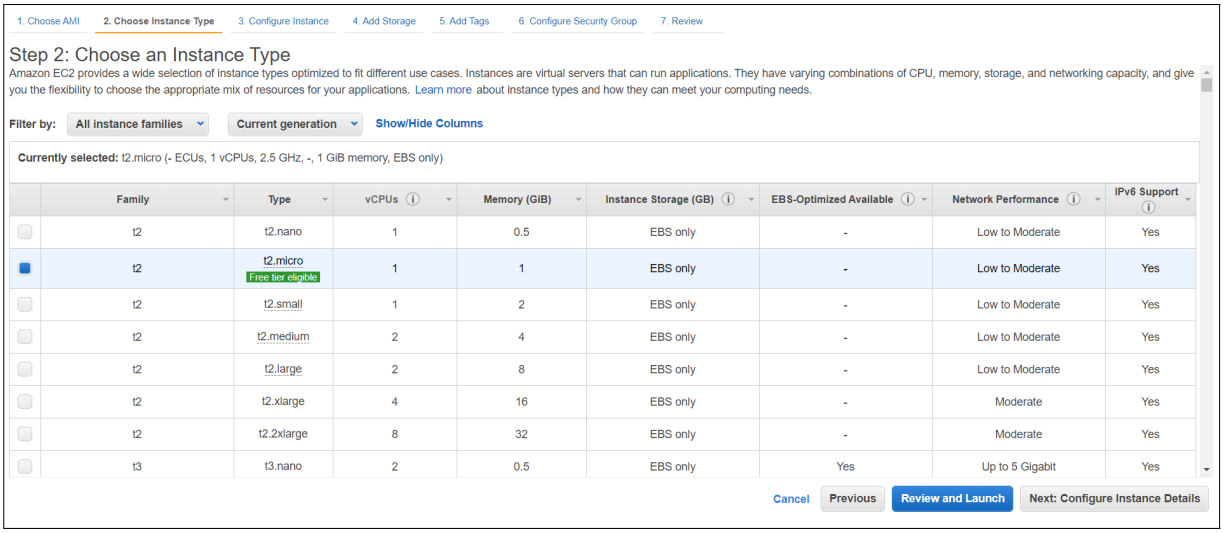

Figure 2: Choose an Instance Type

This is the second step in creation of an EC2 instance. This step involves selection of an EC2 types based on which the desired combination of CPU, storage, network performance, memory other resources are set. Elastic Compute Cloud (EC2) instance types provide different optimal combinations out of which *t2.micro*, a free tier type is selected for this example. The *t2.micro* instance type provides 1 vCPU, 1 GiB memory, low to moderate network performance and EBS storage support.

3.3 Configure Instance Details

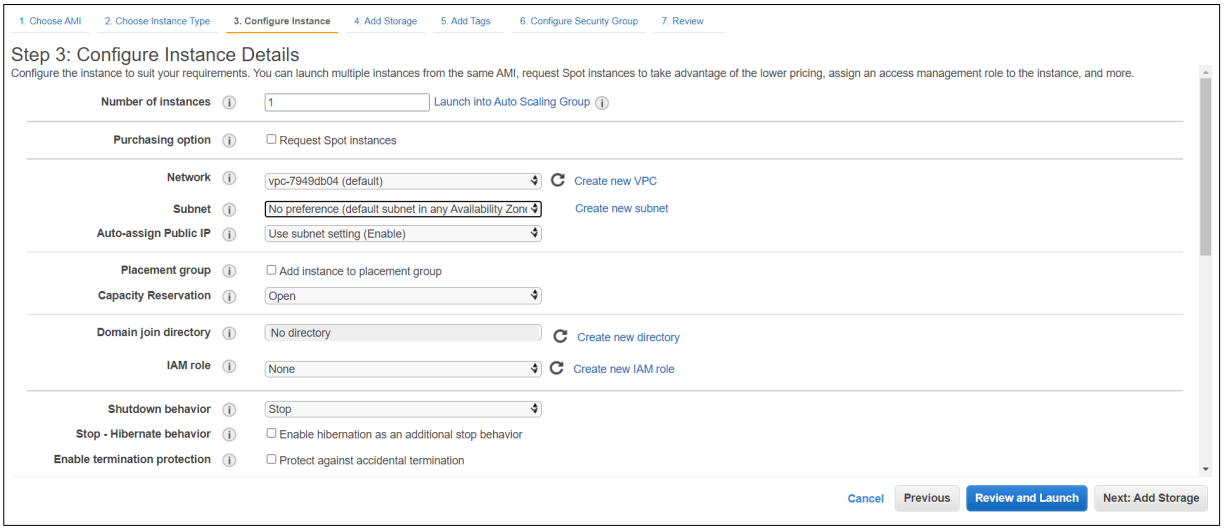

Figure 3: Configure Instance Details

This is the third step in creation of an EC2 instance. This step involves configuration of the EC2 instance details. Configuration options such as number of instances under the same AMI, different purchasing options, network settings for VPC creation and setting subnets, IAM roles, behavior on shutdown and hibernation can be set using this step of instance launch.

3.4 Add Storage

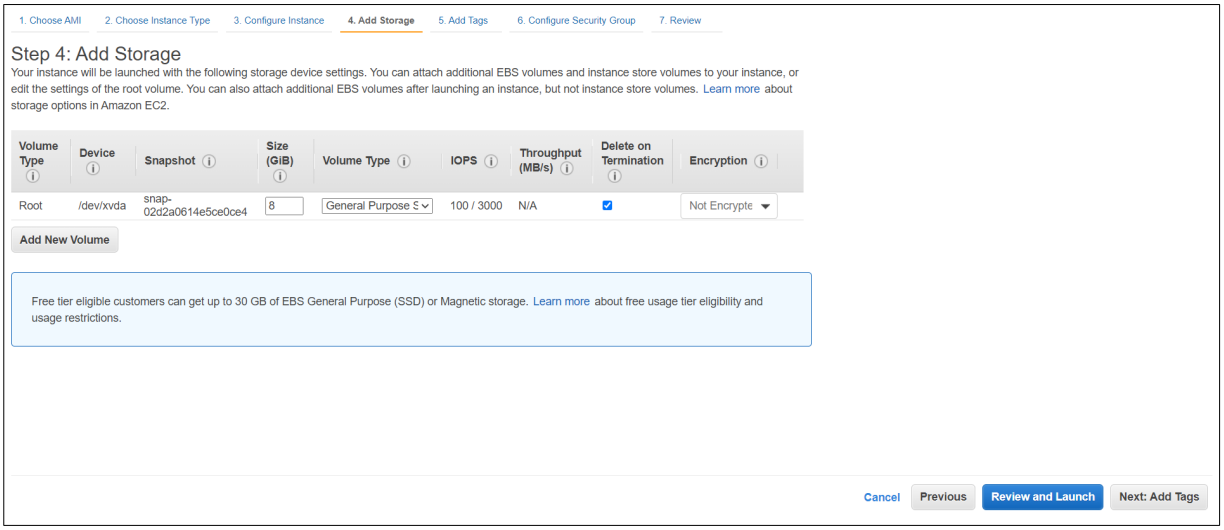

Figure 4: Add Storage

This is the fourth step in creation of an EC2 instance. This step involves configuration of the storage settings that determine the EBS volumes used by the launched instance. Instance store volumes are also selected and attached in this step, which is a must since instance store volumes can't be attached after launching the instance. Root directory can also be edited in this step while launching an EC2 instance.

3.5 Add Tags

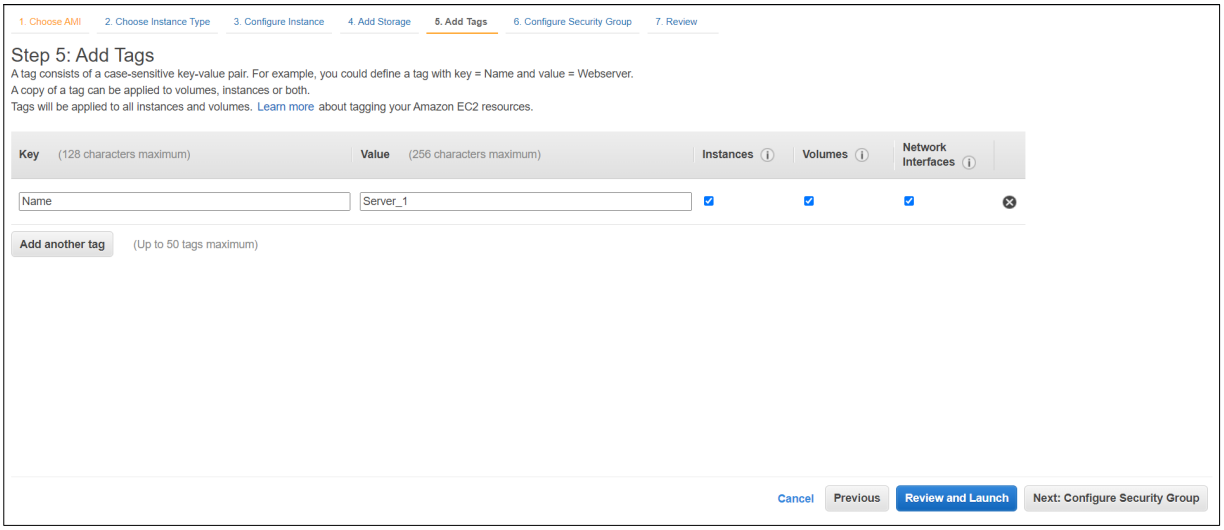

Figure 5: Add Tags

This is the fifth step in creation of an EC2 instance. This step involves adding or editing tags for the instance. Tags are nothing but key-value identifer pair used to filter the launched instances. For this example, a *Name* tag with the value *Server_1* is created. This tag is associated to all instance, volume and the used network interfaces.

3.6 Configure Security Groups

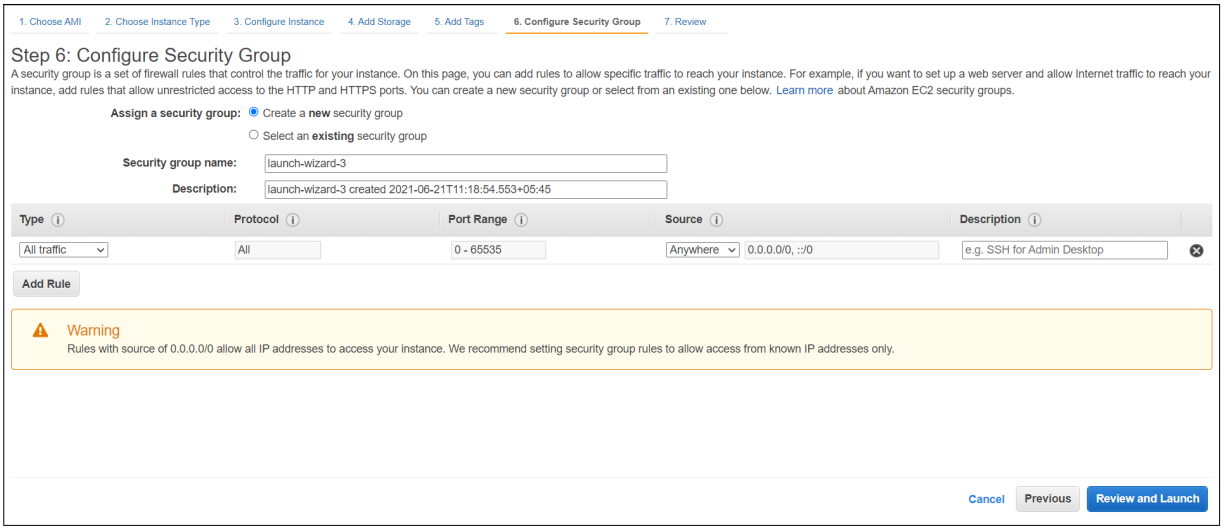

Figure 6: Configure Security Groups

This is the sixth step in creation of an EC2 instance. This step involves configuration of security groups that are essentially act as firewall to control the inbound traffic to the launched instances. The rules can be set for different traffic types, to follow various communication protocols, use the required port ranges accordingly and source address. For this example, all traffic is allowed for requests coming from any address on the internet. This is obviously not the recommended setting since security will be compromised.

3.7 Review and Launch

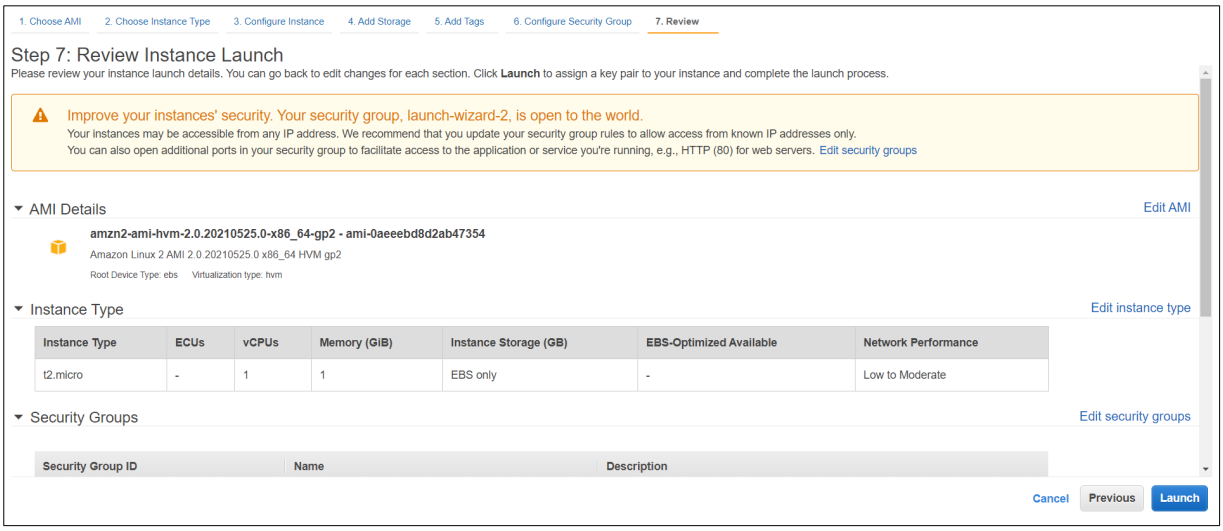

Figure 7: Review and Launch

This is the seventh and final step in creation of an EC2 instance. This step involves reviewing the set configurations in step 1 to 6 and launching the instance. While launching the instance, clients are recommended to use an existing key pair or generate a new one, that is an encrypted key used for remote access. Once launched the instance will take some time to clear the different status checks.