**EC2 (ELASCTIC CLOUD COMPUTE)**

Amazon Elastic Cloud Compute (Amazon EC2)

Amazon EC2 is AWS primary web service that provides resizable compute capacity in the cloud.

Amazon EC2 allows you to acquire compute through the launching of virtual servers called instances. Instance is nothing but a Virtual Server.

**Instance Types (Hardware Configurations):**

The instance type defines the virtual hardware supporting an Amazon EC2 instance. There are many instance types available, based on the following dimensions:

• General purpose

• Compute Optimized (vCPUs)

• GPU Compute

• Memory Optimized

• Storage Optimized

**General Purpose**: General purpose instance family provides a balance of compute, memory, and network resources, and it is a good choice for many applications.

**Compute Optimized (vCPUs):** Compute Optimized instances are optimized for compute-intensive workloads and delivers high performance computing, batch processing.

**GPU Compute**: GPU Compute instances are next generation of general purpose GPU computing instances. We can use GPU instances for 30 visualizations, graphics-intensive remote workstation, 30 rendering, application streaming, video encoding, Machine/Deep learning, high performance computing and other server-side graphics workloads.

**Memory Optimized:** Memory Optimized category instances are most suitable for high performance databases, distributed memory caches, in-memory analytics, large-scale, enterprise-class, and In-memory applications.

**Storage Optimized:** Storage Optimized category instances are most suitable for low latency, very high random I/O performance, high sequential read throughput and provide high IOPS and NoSQL databases like Cassandra, MongoDB, Redis and In-memory databases.

|  |  |
| --- | --- |
| Compute optimized | For workloads requiring significant processing. |
| Memory optimized | For memory-intensive workloads. |
| Storage optimized | For workloads requiring high amounts of fast SSD Storage. |
| GPU-based instances | Intended for graphics and general-purpose GPU compute workloads |

**Instance launch pricing Options:**

• On-Demand Instances

• Reserved Instances

• Spotlnstances

**On-Demand Instances:**

The price per hour for each instance type published on the

AWS website represents the price for On-Demand Instances.

• On-Demand is most flexible pricing option, as it doesn't requires up-front commitment.

• We will have control over when the instance is launched and when it is terminated.

• Suitable for unpredictable workloads.

**Reserved Instances**: When purchasing a reserved instance we have to specify the instance type and Availability Zone for that Reserved Instance and achieve a lower effective hourly price for that instance for the duration of the reservation. You can select duration from 1 Yr to 3 yrs.

We have three payment options for Reserved Instances:

**All Upfront**: Pay for the entire reservation upfront. There is no monthly charge for the customer during the term.

**Partial Upfront**: Pay a portion of the reservation charge up front and the rest in monthly installments for the duration of the term.

**No Upfront**: Pay the entire reservation charge in monthly installments for the duration of the term.

• We can save up to 75 percent over on-demand hourly rate if we reserve instance through Reserved Option.

**Spot Instances**: For workloads that are not time critical and are tolerant of interruption, Spot Instances offer the greatest discount.

* We can specify the price they are willing to pay for a certain instance type.
* When the bid price is above the current Spot price, we'll get the requested instances.
* These instances will operate like all other Amazon EC2 instances; and the customer will only pay the Spot price for the hours that instance(s) run.

The instances will run until:

• Till we terminate them manually.

• The Spot price goes above our bid price.

• There is not enough unused capacity to meet the demand for Spot Instances.

• If Amazon EC2 needs to terminate a Spot Instance, the instance will receive a termination notice providing a two-minute warning prior to termination.

• If we terminate Instance manually we have to pay for Partial hours, if Amazon terminates we will not get charged for partial hours.

**Tenancy Options:**

**Shared Tenancy:** Shared tenancy is the default tenancy model for all Amazon EC2 instances. A single host machine may house instances from different customers. (One host may share with multiple customers).

**Dedicated Instances:** Dedicated Instances run on hardware that's dedicated to a single customer. As a customer runs more Dedicated Instances, more underlying hardware may be dedicated to their account.

**Dedicated Host:** An Amazon EC2 Dedicated Host is a physical server with Amazon EC2 instance capacity fully dedicated to a single customer's use. We will get complete control over which specific host runs an instance at launch.

**Placement Groups:** A placement group is a logical grouping of instances within a single Availability Zone.

**•** Placement groups enable applications to participate in a low-latency, 10 Gbps network.

• Recommended for applications that benefit from low network latency, high network throughput, or both.

• Only certain types of instances can be launched in a placement group.

• A placement group can't span multiple Availability Zones.

• The name you specify for a placement group must be unique within your AWS account.

• AWS recommend homogenous instances within placement groups.

**•** You can't merge placement groups.

• You can't move an existing instance into a placement group.

**Amazon Machine Images (AMis):**

The Amazon Machine Image (AMI) defines the initial software that will be on an instance when it is launched.

• The Operating System (OS) and its configuration.

• . The initial state of any patches.

• Application or system software.

All AMis are based on x86 OSs, either Linux or Windows.

We can launch instances from four options

1. Published by AWS

2. AWS Marketplace

3. Generated from existing Instance

4. Uploaded Virtual Servers

**Accessing an Instance**:

We can access our Instances by Using Public DNS, Public IP address and Elastic IP addresses.

**Public DNS**: When we launch instance, we will get one Public DNS associated for that instance.

* Public DNS will generate automatically. We can't specify.
* We can found this information in Instance description.
* We cannot transfer this Public DNS to another instance.
* We will get public DNS when the instance is in running state.

**Public IP:**

* When we launch instance, we will get one Public IP address also.
* AWS will allocate this address, no option to select specific IP.
* This is unique on the Internet.

**Elastic IP:**

* An Elastic IP address is a static IPv4 address designed for dynamic cloud computing. An Elastic IP address is associated with your AWS account.
* To use an EIP address, we have to generate one to our AWS account, and then associate it with your instance or a network interface.
* We can disassociate an EIP address from a resource, and reassociate it with a different resource.
* A disassociated EIP address remains allocated to your account until you manually release it.

**Steps to get EIP Address**:

1. Login to A WS account and navigate to Amazon EC2 console.
2. In the navigation pane, choose Elastic IPs.
3. Choose Allocate new address.
4. Select Allocate. Close the confirmation screen.

**Enhanced networking**: Reduces the impact of virtualization on network performance by enabling a capability called Single Root I/O Virtualization (SRIOV). This results in more Packets per Second, lower latency, and less jitter.

**Instance launch process:**

1. Login to Your AWS Account, Select and switch to the required Region and find EC2 under Compute Section.
2. Select the Launch instance option and it will launch an instance launch wizard.
3. We have Windows and Linux operating systems available in Quick start option

Along with the Quick Start option, you can also spin up your instances using the AWS Marketplace and the Community AMis section. Both these options contains list of customized AMis that have been created by either third-party companies or by developers and can be used for a variety of purposes.

1. Choose an instance type:

In the next step, we have to select the instance type as per our requirements. You can filter instances according to their families.

We can use the general purpose t2.micro instance type, which is comes under the free tier eligibility and configuration is 1 vCPU and 1 GB of RAM.

1. **Configure instance details**:

**Number of instances**: You can specify how many instances the wizard should launch using this field. By default, the value is always set to one single instance.

**Purchasing option**: We can this instance under spot instances request. For now let's leave this option.

**Network**: Select the default Virtual Private Cloud (VPC) network that is displayed in the dropdown list. We can even go ahead and create a new VPC network for this instance, but we will leave and will see VPC in later chapters.

**Subnet**: select the Subnet in which you wish to deploy your new instance.

You can either choose to have AWS select and deploy your instance in a particular subnet from an available list or you can select a particular choice of subnet on your own.

**Auto-assign Public IP**: Each instance that you launch will be assigned a Public IP. We are going to use this public IP to connect to our Instance over Internet.

**IAM role**: You can additionally select a particular IAM role to be associated with your instance.

**Shutdown behavior**: This option allows us to select whether the instance should stop or be terminated when issued a shutdown request. In this case, we have opted for the instance to stop when it is issued a shutdown command.

**Enable termination protection:** Select this option in case you wish to protect your instance against accidental deletions. It adds additional step for instance termination. If, we enable this option, we need to manually Disable to terminate the instance.

**Monitoring**: By default, A WS will monitor few basic parameters about your instance for free, but if you wish to have an in-depth insight into your instance's performance, then select the Enable Cloud Watch detailed monitoring option. But you'll get charged for detailed monitoring.

**Tenancy:** We can choose to run our instances on physical servers fully dedicated for your use. The use of host tenancy will request to launch instances onto dedicated hosts.

**Bootstrapping** (User-data): We can configure instances and install applications programmatically when an instance is launched. The process of providing code to be run on an instance at launch is called bootstrapping.

On Linux instances this can be shell script, and on Windows instances this can be a batch style script or a Power Shell script.

**Step 6: Add Storage**

We can add EBS volumes to our instances. To add new volumes, simply click on the Add New Volume button. This will provide with 0ptions to provide the size of the new volume along with its mount points. There is an 8 GB volume already attached to our instance. This is the t2.micro instance's root volume.