ELASTIC COMPUTE CLOUD

By Vignesh S

EC2 - Elastic compute cloud

Amazon Elastic Compute Cloud (Amazon EC2) is a scalable cloud computing service offered by Amazon Web Services (AWS). It provides virtual servers, known as EC2 instances, that can be quickly provisioned and configured to run various applications and workloads in the cloud. Here are some key aspects of Amazon EC2

EC2 instances are virtual machines running in the cloud. They can be customized with different configurations, including CPU, memory, storage, and networking capacity. You have full control over the configuration and management of your instances.

Why EC2 ?

- Scalability and Flexibility: With EC2, you can easily scale the number of instances up or down to match your workload demands. This flexibility allows you to handle traffic spikes or scale resources as your needs change.
- Multiple Instance Types: EC2 offers a wide range of instance types optimized for different use cases, such as general-purpose computing, memory-intensive applications, high-performance computing, and storage-optimized workloads. Each instance type has specific characteristics and performance profiles.
- Security and Networking: EC2 integrates with Amazon Virtual Private Cloud (VPC), which provides a logically isolated network environment within the AWS cloud. You can control network access using security groups and define network subnets, IP addresses, and routing tables.

Why EC2 ?

- Storage Options: EC2 instances can be equipped with different storage options, including Amazon Elastic Block Store (EBS) for persistent block-level storage and Amazon Elastic File System (EFS) for scalable network file storage. You can also use Amazon S3 for object storage and other AWS storage services.
- Load Balancing and Auto Scaling: EC2 integrates with other AWS services like Elastic Load Balancing and Auto Scaling. Elastic Load Balancing distributes incoming traffic across multiple EC2 instances, while Auto Scaling automatically adjusts the number of instances based on predefined scaling policies.
- Pay-as-You-Go Pricing: Amazon EC2 follows a pay-as-you-go pricing model, where you pay only for the resources you consume. You can choose from different pricing options, including On-Demand Instances, Reserved Instances, and Spot Instances, depending on your budget and usage patterns.

Launching your First instance

- Sign in to the AWS Management Console: Go to the AWS website (https://aws.amazon.com) and sign in to the AWS Management Console using your AWS account credentials.
- Open the EC2 Dashboard: Once you're logged in, navigate to the EC2 service by selecting "Services" from the top navigation bar and then clicking on "EC2" under the "Compute" section.
- Launch an Instance: In the EC2 Dashboard, click on the "Launch Instance" button to begin the instance creation process.
- Choose an Amazon Machine Image (AMI): An AMI is a template for the root file system of the EC2 instance. Select the AMI that best suits your requirements, such as an Amazon Linux, Ubuntu, or Windows Server AMI.

Launching your First instance

- Choose an Instance Type: Select the desired instance type based on the CPU, memory, storage, and networking requirements of your application. Different instance families offer varying performance profiles and capabilities.
- Key Pair Selection: Select an existing key pair or create a new key pair. This key pair will be used to securely connect to your EC2 instance using SSH (for Linux) or RDP (for Windows).
- Configure Instance Details: Specify configuration details for your instance, such as the number of instances, network settings, subnet, security groups, and storage options (EBS volumes).
- Add Storage (Optional): Configure additional storage volumes if needed. You can attach EBS volumes to your instance for data storage.

Launching your First instance

- Configure Security Groups: Security groups control the inbound and outbound traffic to your instance. Define the necessary rules to allow access to the required ports and protocols.
- Launch Confirmation: After selecting the key pair, click on the "Launch Instances" button to launch the EC2 instance. You will see a confirmation message indicating that your instance is being launched.
- Access and Connect to Your Instance: Once the instance is successfully launched, you can access it using SSH or RDP. For Linux instances, you will connect via SSH using the private key associated with the key pair. For Windows instances, you will use the Remote Desktop Protocol (RDP) and the Administrator password associated with the instance.

RESERVATION

On-Demand Instances:

On-Demand Instances are the default pricing model in EC2. With On-Demand Instances, you pay for compute capacity by the hour or second without any upfront commitment or long-term contract. This model offers flexibility and allows you to launch instances whenever you need them, and terminate them when they are no longer required. On-Demand Instances are suitable for workloads with short-term or unpredictable usage patterns.

RESERVATION

Reserved Instances (RI):

Reserved Instances enable you to reserve capacity in advance for a specified term (1 or 3 years) in exchange for a significant discount compared to On-Demand pricing. Reserved Instances provide a cost-saving option for steady-state or predictable workloads.

Standard Reserved Instances:
These offer a substantial discount (up to 75%) compared to On-Demand pricing and provide capacity reservation in a specific Availability Zone.

RESERVATION

Convertible Reserved Instances:
These offer a smaller discount (up to 54%) but provide more flexibility. Convertible Reserved Instances can be modified to a different instance type within the same instance family, allowing you to adapt to changing workload requirements.

Scheduled Reserved Instances:
These allow you to reserve capacity for specific time windows on a recurring basis, which is useful for applications with predictable usage patterns.

RESERVATION

Spot Instances:

 Spot Instances are spare EC2 instances available at significantly discounted prices (up to 90% off On-Demand pricing). The pricing is determined by supply and demand dynamics in the Spot market, and the availability of instances can vary. Spot Instances are suitable for workloads that can tolerate interruptions and have flexible start and end times. If the Spot price exceeds your bid, your instances may be terminated with a two-minute notice. However, you can set up Spot Fleet or Spot Blocks to specify maximum prices or reserve capacity for a longer duration.

Dedicated Instances:

Dedicated Instances run on hardware dedicated to a single AWS account but are still virtual instances. They provide isolation at the hardware level, ensuring that your instances are not shared with instances from other accounts. This is useful for workloads with specific security and compliance requirements.

Dedicated Hosts:

 Dedicated Hosts provide physical servers dedicated exclusively to your AWS account. With Dedicated Hosts, you have full control over the underlying hardware, including instance placement and host affinity. Dedicated Hosts are beneficial for software licensing that requires dedicated hardware or for workloads that have specific regulatory or compliance requirements.

AMI

stands for Amazon Machine Image. It is a pre-configured template that contains the necessary information to launch an instance in Amazon Elastic Compute Cloud (EC2). An AMI includes the operating system, software, and configuration required to start an EC2 instance

Definition:

An AMI is a snapshot of a specific state of an EC2 instance, typically created by taking a snapshot of the root volume of the instance. It captures the entire file system, including the operating system, installed applications, data, and configuration settings.

Types of AMIs:

Public AMIs: AWS provides a variety of pre-built, publicly available AMIs that are maintained by AWS or other users. These AMIs cover a wide range of operating systems, software stacks, and application frameworks.

AWS Marketplace AMIs: The AWS Marketplace offers a vast selection of AMIs created by third-party vendors, including both free and paid options. These AMIs may include pre-configured software or specific application environments.

Custom AMIs: Users can create their own custom AMIs based on their existing EC2 instances or by importing their own virtual machine images. Custom AMIs enable users to capture their preferred configurations, applications, and security settings.