What Is Amazon EC2?

Amazon Elastic Compute Cloud (Amazon EC2) provides scalable computing capacity in the Amazon Web Services (AWS) cloud. Using Amazon EC2 eliminates your need to invest in hardware up front, so you can develop and deploy applications faster. You can use Amazon EC2 to launch as many or as few virtual servers as you need, configure security and networking, and manage storage. Amazon EC2 enables you to scale up or down to handle changes in requirements or spikes in popularity, reducing your need to forecast traffic

Features of Amazon EC2

Amazon EC2 provides the following features:

• Virtual computing environments, known as *instances*

• Preconfigured templates for your instances, known as *Amazon Machine Images (AMIs)*, that package the bits you need for your server (including the operating system and additional software)

• Various configurations of CPU, memory, storage, and networking capacity for your instances, known as *instance types*

• Secure login information for your instances using *key pairs* (AWS stores the public key, and you store the private key in a secure place)

• Storage volumes for temporary data that's deleted when you stop or terminate your instance, known as *instance store volumes*

• Persistent storage volumes for your data using Amazon Elastic Block Store (Amazon EBS), known as *Amazon EBS volumes*

• Multiple physical locations for your resources, such as instances and Amazon EBS volumes, known as *regions* and *Availability Zones*

• A firewall that enables you to specify the protocols, ports, and source IP ranges that can reach your instances using *security groups*

• Static IPv4 addresses for dynamic cloud computing, known as *Elastic IP addresses*

• Metadata, known as *tags*, that you can create and assign to your Amazon EC2 resources

Amazon Machine Images and Instances

An *Amazon Machine Image (AMI)* is a template that contains a software configuration (for example, an operating system, an application server, and applications). From an AMI, you launch *instances*, which are copies of the AMI running as virtual servers in the cloud.

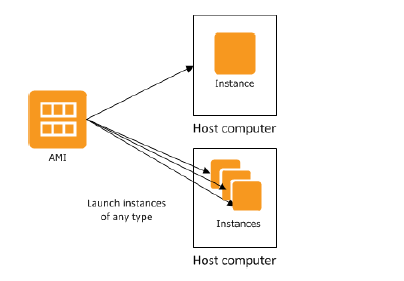
Amazon publishes many AMIs that contain common software configurations for public use. In addition, members of the AWS developer community have published their own custom AMIs. You can also create your own custom AMI or AMIs; doing so enables you to quickly and easily start new instances that have everything you need.

For example, if your application is a website or web service, your AMI could include a web server, the associated static content, and the code for the dynamic pages. As a result, after you launch an instance from this AMI, your web server starts, and your application is ready to accept requests.

You can launch different types of instances from a single AMI. An *instance type* essentially determines the hardware of the host computer used for your instance.

Each instance type offers different compute and memory facilities. Select an instance type based on the amount of memory and computing power that you need for the applications or software that you plan to run on the instance.

You can also launch multiple instances from an AMI, as shown in the following figure.

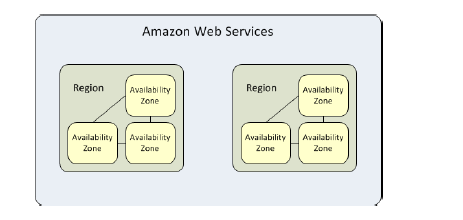


Your Windows instances keep running until you stop or terminate them, or until they fail. If an instance fails, you can launch a new one from the AMI.

Regions and Availability Zones

Amazon has data centers in different areas of the world (for example, North America, Europe, and Asia). Correspondingly, Amazon EC2 is available to use in different *regions*. By launching instances in separate regions, you can design your application to be closer to specific customers or to meet legal or other requirements. Prices for Amazon EC2 usage vary by region (for more information about pricing by region, see Amazon EC2 Pricing).

Each region contains multiple distinct locations called *Availability Zones*. Each Availability Zone is engineered to be isolated from failures in other Availability Zones, and to provide inexpensive, lowlatency network connectivity to other zones in the same region. By launching instances in separate Availability Zones, you can protect your applications from the failure of a single location.



Storage

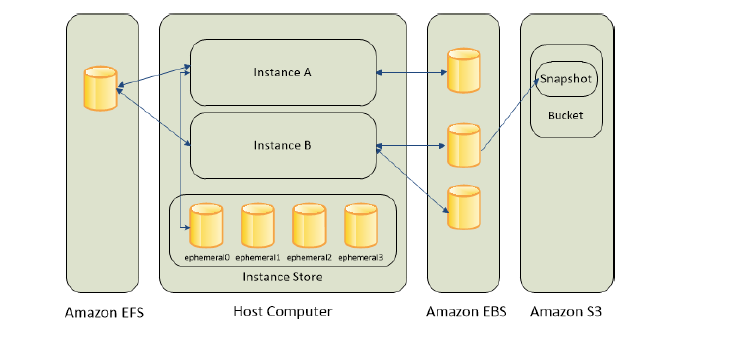
When using Amazon EC2, you may have data that you need to store. Amazon EC2 offers the following storage options:

• Amazon Elastic Block Store (Amazon EBS)

• Amazon EC2 Instance Store (p. 843)

• Amazon Simple Storage Service (Amazon S3)

The following figure shows the relationship between these types of storage.



Amazon EBS Volumes

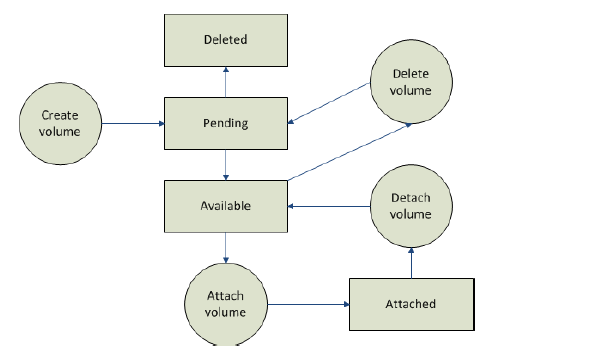
Amazon EBS volumes are the recommended storage option for the majority of use cases. Amazon EBS provides your instances with persistent, block-level storage.

Amazon EBS volumes are essentially hard disks that you can attach to a running instance. Amazon EBS is especially suited for applications that require a database, a file system, or access to raw block-level storage.

As illustrated in the previous figure, you can attach multiple volumes to an instance. Also, to keep a backup copy of your data, you can create a *snapshot* of an EBS volume, which is stored in Amazon S3.

You can create a new Amazon EBS volume from a snapshot, and attach it to another instance. You can also detach a volume from an instance and attach it to a different instance.

The following figure illustrates the life cycle of an EBS volume.



Instance Store

All instance types, with the exception of Micro instances, offer *instance store*, which provides your instances with temporary, block-level storage. This is storage that is physically attached to the host computer.

The data on an instance store volume doesn't persist when the associated instance is stopped or terminated.

Instance store is an option for inexpensive temporary storage. You can use instance store volumes if you don't require data persistence.

Amazon S3

Amazon S3 is storage for the Internet. It provides a simple web service interface that enables you to store and retrieve any amount of data from anywhere on the web.

AWS Identity and Access Management

AWS Identity and Access Management (IAM) enables you to do the following:

• Create users and groups under your AWS account

• Assign unique security credentials to each user under your AWS account

• Control each user's permissions to perform tasks using AWS resources

• Allow the users in another AWS account to share your AWS resources

• Create roles for your AWS account and define the users or services that can assume them

• Use existing identities for your enterprise to grant permissions to perform tasks using AWS resources

By using IAM with Amazon EC2, you can control whether users in your organization can perform a task using specific Amazon EC2 API actions and whether they can use specific AWS resources.

Key Pair

AWS uses public-key cryptography to secure the login information for your instance. You specify the name of the key pair when you launch your instance, then provide the private key to obtain the administrator password for your Windows instance so you can log in using RDP.

If you haven't created a key pair already, you can create one using the Amazon EC2 console. Note that if you plan to launch instances in multiple regions, you'll need to create a key pair in each region.

Security Group

Security groups act as a firewall for associated instances, controlling both inbound and outbound traffic at the instance level. You must add rules to a security group that enable you to connect to your instance from your IP address using RDP.

You can also add rules that allow inbound and outbound HTTP and HTTPS access from anywhere.

Note that if you plan to launch instances in multiple regions, you'll need to create a security group in each region.