
amazon EC2

❖ ELASTIC COMPUTE CLOUD (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.
- **Virtual servers** can run different **OS'S**, but most commonly run a flavor of **Linux, MacOS & Windows**.

EC2 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, hibernate, 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
- Virtual networks you can create that are logically isolated from the rest of the AWS Cloud, and that you can optionally connect to your own network, known as **virtual private clouds (VPCs)**

➢ EC2 PURCHASING OPTIONS:

- You can use the following options to optimize your costs for Amazon EC2:

ON-DEMAND:
- Pay, by the second, for the instances that you launch.

SAVING PLANS:
- Reduce costs by making a commitment to a consistent amount of usage, in USD per hour, for a term of 1 or 3 years.

RESERVED INSTANCES:
- Reduce costs by making a commitment to a consistent instance configuration, including instance type and Region, for a term of 1 or 3 years.

SCHEDULED INSTANCES:
- Purchase instances that are always available on the specified recurring Schedule, for a one-year term.

SPOT INSTANCES:
- It fluctuates based on the supply and demand of available unused EC2 capacity.

DEDICATED HOSTS:
- Pay for a physical host that is fully dedicated to running your instances, and bring your existing per-socket, per-core, or per-VM software licenses to reduce costs.

DEDICATED INSTANCES:
- Pay, by the hour, for instances that run on single-tenant hardware.

CAPACITY RESERVATION:
- Reserve capacity for your EC2 instances in a specific AZ for any duration.

➢ EC2 INSTANCE TYPES:

- EC2 provides a wide selection of instance types optimized to fit different use cases.
- Instance types comprise varying combinations of CPU, memory, storage, and networking capacity and give you the flexibility to choose the appropriate mix of resources for your applications.
- Each instance type includes one or more instance sizes, allowing you to scale your resources to the requirements of your target workload.

GENERAL PURPOSE:

- These are providing a balance of compute, memory and networking resources, and can be used for a variety of diverse workloads. These instances are ideal for applications that use these resources in equal proportions such as web servers and code repositories.

COMPUTE OPTIMIZED:

- These are ideal for compute bound applications that benefit from high performance processors.
- This family are well suited for batch processing workloads, media transcoding, high performance web servers, high performance computing (HPC), scientific modeling, dedicated gaming servers and ad server engines, machine learning inference and other compute intensive applications.

MEMORY OPTIMIZED:

- These instances are designed to deliver fast performance for workloads that process large data sets in memory.

ACCELERATED COMPUTING:

- Accelerated computing instances use hardware accelerators, or co-processors, to perform functions, such as floating-point number calculations, graphics processing, or data pattern matching, more efficiently than is possible in software running on CPUs.

STORAGE OPTIMIZED:

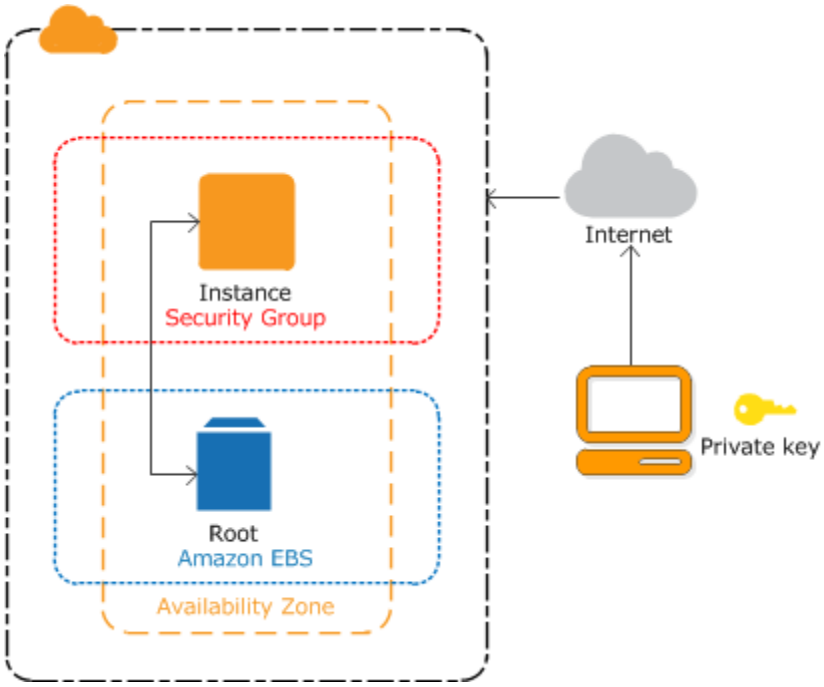
- Storage optimized instances are designed for workloads that require high, sequential read and write access to very large data sets on local storage.
- They are optimized to deliver tens of thousands of low-latency, random I/O operations per second (IOPS) to applications.

HPC OPTIMIZED:

- High performance computing (HPC) instances are purpose built to offer the best price performance for running HPC workloads at scale on AWS.
- HPC instances are ideal for applications that benefit from high-performance processors such as large, complex simulations and deep learning workloads.

➢ EC2 INSTANCE OVERVIEW:

- The instance is an Amazon EBS-backed instance (meaning that the root volume is an EBS volume).
- You can think of an Availability Zone as an isolated data center.



REGION:

- AWS has the concept of a Region, which is a physical location around the world where we cluster data centers.
- Each AWS Region consists of multiple, isolated, and physically separate AZs within a geographic area.

AVAILABILITY ZONES (AZ'S):

- It consists of one or more discrete data centers, each with redundant power, networking and connectivity, housed in separate facilities.
- These AZ's offer you the ability to operate AWS services.

INSTANCE:

- Virtual computing environments, known as **instances**

SECURITY GROUPS (SG'S):

- Security Groups are act as a virtual firewall for controlling traffic at the **instance level**.
- It contains a set of rules that filter traffic coming **Inbound** and **Outbound** of an EC2 instance.
- Security Groups are support only **allow rules**.
- **Stateful:** If an inbound request passes, then the outbound request will pass as well.

KEY-PAIR:

- Used to securely communicate with EC2 instances and prevent unauthorized access.
- EC2 uses public key cryptography to **Encrypt** and **Decrypt** login information.
- **Public key** cryptography uses to encrypt a piece of data, the recipient uses the **private key** to decrypt the data
- The **Public** and **Private** keys are known as a **Key Pair**.
    - **Public key:** Used to encrypt data and stored on the EC2 instance
    - **Private key:** Used to decrypt data and kept by the user

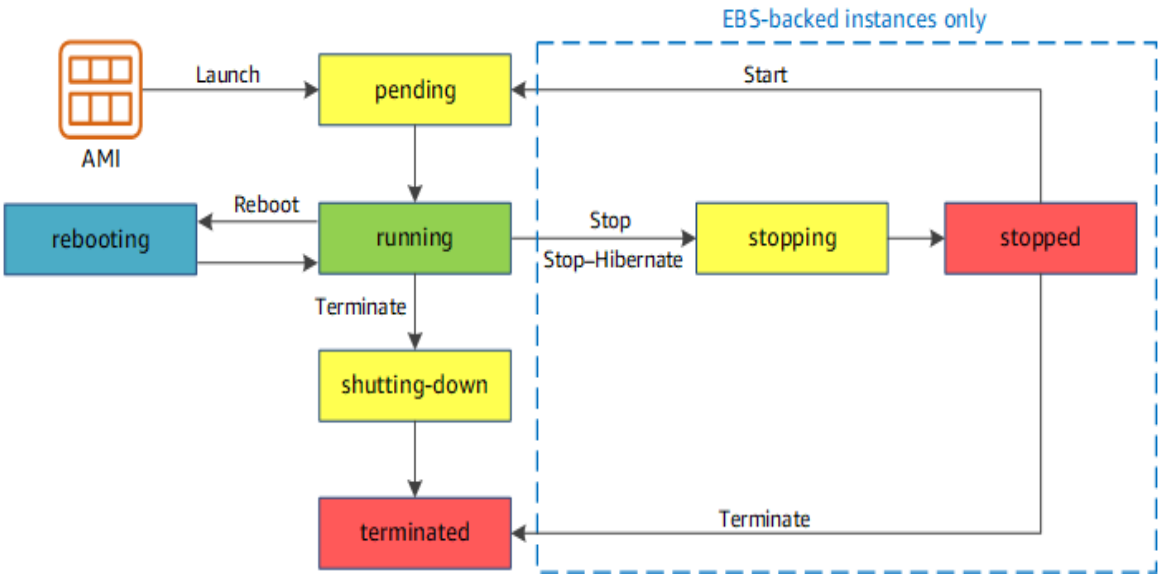
ROOT AMAZON ELASTIC BLOCK STORE (EBS):

- The root Amazon Elastic Block Store (EBS) volume is the primary volume attached to an Amazon Elastic Compute Cloud (EC2) instance when it launches:
- The root EBS volume **contains the operating system** and **boot files** that allow the instance to start and run.
- The root **EBS volume is persistent**, meaning that it persists even if you stop the instance. If you restart the instance, the volume is automatically remounted, restoring the instance state and any stored data.

➢ INSTANCE LIFECYCLE:

- Amazon EC2 instance transitions through different states from the moment you launch it through to its termination.
- The following illustration represents the transitions between instance states.

NOTE: You can't stop and start an instance store-backed instance.



- The instance lifecycle of an Amazon EC2 instance is the path it takes from launch to termination, and includes several different states:

Launch: The instance enters a pending state after launch.

Running: The instance is running and ready for use.

Stopped: The instance is shut down and cannot be used. The instance can be started at any time.

Shutting-down: The instance is preparing to be terminated.

Terminated: The instance is no longer in use and has been terminated.