Amazon EC2 – Associate

Private vs Public IP (IPv4)

- Networking has two sorts of IPs, IPv4 and IPv6:

    - IPv4: `1.160.10.240`
    - IPv6: `3ffe:1900:4545:3:200:f8ff:fe21:67cf`

- In this course, we will only be using IPv4.

- IPv4 is still the most common format used online.

- IPv6 is newer and solves problems for the Internet of Things (IoT).

- IPv4 allows for **3.7 billion** different addresses in the public space.

- IPv4 format: `[0-255].[0-255].[0-255].[0-255]`

Private vs Public IP (IPv4) Example

This example illustrates the communication between public and private networks using IPv4:

- There is a **Web Server** with a public IP: `79.216.59.75`
- There is another **Server** with a public IP: `211.139.37.43`

Two companies have their own private networks connected to the Internet:

Company A
- Internet Gateway (public IP): `149.140.72.10`
- Private Network: `192.168.0.1/22`

Company B
- Internet Gateway (public IP): `253.144.139.205`
- Private Network: `192.168.0.1/22`

The Web Server, Server, and the Internet Gateways of both companies communicate over the **World Wide Web (WWW)**.

Each company uses a private IP range internally and connects externally through a public IP via their Internet Gateway.

Private vs Public IP (IPv4) Fundamental Differences

Public IP
- A Public IP means the machine can be identified on the internet (WWW).
- It must be unique across the whole web (no two machines can have the same public IP).
- It can be geo-located easily.

Private IP
- A Private IP means the machine can only be identified on a private network.
- The IP must be unique across the private network.
- Two different private networks (e.g., two companies) can have the same IPs.

- Machines connect to the WWW using a NAT + Internet Gateway (acting as a proxy).
- Only a specified range of IPs can be used as private IPs.

Elastic IPs

- When you stop and then start an EC2 instance, it can change its public IP.
- If you need to have a fixed public IP for your instance, you need an Elastic IP.
- An Elastic IP is a public IPv4 IP that you own as long as you don't delete it.
- You can attach it to one instance at a time.

Elastic IP

- With an Elastic IP address, you can mask the failure of an instance or software by rapidly remapping the address to another instance in your account.

- You can only have 5 Elastic IPs in your account (you can ask AWS to increase that limit).

- Overall, **try to avoid using Elastic IPs**:

  - They often reflect poor architectural decisions.
  - Instead, use a random public IP and register a DNS name to it.
  - Or, use a Load Balancer and do not use a public IP.

Private vs Public IP (IPv4) in AWS EC2 – Hands On

- By default, your EC2 machine comes with:

  - A private IP for the internal AWS network.
  - A public IP for the WWW.

- When we are doing SSH into our EC2 machines:

  - We can't use a private IP, because we are not in the same network.
  - We can only use the public IP.

- If your machine is stopped and then started, **the public IP can change**.

Placement Groups

- Sometimes you want control over the EC2 instance placement strategy.

- That strategy can be defined using **placement groups**.

- When you create a placement group, you specify one of the following strategies:

  - **Cluster**
    Clusters instances into a low-latency group in a single Availability Zone.

  - **Spread**
    Spreads instances across underlying hardware (maximum 7 instances per group per Availability Zone).

  - **Partition**
    Spreads instances across many different partitions (using different sets of racks) within an AZ.

Scales to hundreds of EC2 instances per group.

Recommended for distributed systems like **Hadoop**, **Cassandra**, **Kafka**.

Placement Groups – Cluster

A Cluster Placement Group places EC2 instances close together within the same Availability Zone to achieve low-latency, high-throughput networking.

- **Pros**:

  - Great network performance (10 Gbps bandwidth between instances with Enhanced Networking enabled – recommended).

- **Cons**:

  - If the Availability Zone (AZ) fails, **all instances in the group fail together**.

- **Use cases**:

  - Big Data jobs that need to complete fast.
  - Applications requiring **extremely low latency** and **high network throughput**.

Placement Groups – Spread

A Spread Placement Group distributes EC2 instances across multiple hardware and Availability Zones (AZs) to minimize correlated failures.

- **Pros**:

  - Can span across multiple Availability Zones (AZs).
  - Reduced risk of simultaneous failure.
  - EC2 instances are placed on **different physical hardware**.

- **Cons**:

  - Limited to **7 instances per AZ** per placement group.

- **Use cases**:

  - Applications that need to **maximize high availability**.
  - **Critical applications** where each instance must be isolated from failure of others.

Placement Groups – Partition

A Partition Placement Group divides instances into logical partitions to reduce correlated failures and provide isolation at the rack level.

- **Limits**:

  - Up to **7 partitions per Availability Zone (AZ)**.
  - Can **span across multiple AZs** in the same region.
  - Can scale to **hundreds of EC2 instances**.

- **Behavior**:

  - Instances in a partition do **not share racks** with instances in other partitions.
  - A **partition failure** can affect many EC2 instances, but does **not impact other partitions**.

- EC2 instances receive partition information as **metadata**.

- **Use cases**:

    - Distributed systems like **HDFS**, **HBase**, **Cassandra**, **Kafka**.

Elastic Network Interfaces (ENI)

- Logical component in a VPC that represents a **virtual network card**.

- An ENI can have the following attributes:

    - Primary private IPv4
    - One or more secondary IPv4 addresses
    - One Elastic IP (IPv4) per private IPv4
    - One Public IPv4
    - One or more security groups
    - A MAC address

- ENIs can be:

    - Created independently
    - Attached/detached **on the fly** to EC2 instances (useful for **failover**)

- Bound to a specific **Availability Zone (AZ)**

- Example:

    - `eth0` – primary ENI: `192.168.0.31`
    - `eth1` – secondary ENI: `192.168.0.42` (can be moved to another EC2)

EC2 Hibernate

- We know we can **stop** and **terminate** EC2 instances:

    - **Stop**: the data on disk (EBS) is kept intact and available at the next start.
    - **Terminate**: any EBS volumes (root) set up to be destroyed are lost.

- On instance **start**, the following happens:

    - **First start**: the OS boots and the EC2 **User Data** script is run.
    - **Subsequent starts**: only the OS boots up.
    - Then, the application starts, caches get warmed up – and that takes time!

EC2 Hibernate

Introducing EC2 Hibernate
- The **in-memory (RAM) state is preserved**.
- The instance **boots much faster** (the OS is not stopped/restarted).
- Under the hood:
    - The RAM state is saved to a file in the **root EBS volume**.
    - The **root EBS volume must be encrypted**.

Use cases
- Long-running processing.

- Saving the RAM state between sessions.
- Services that require **time to initialize** (avoiding cold start delays).

Summary (Compared to Stop/Start)

Action	RAM State	Boot Time	EBS Data	OS Restarted	Root Volume Encrypted
Stop	Lost	Normal	Kept	Yes	No
Terminate	Lost	N/A	Lost	Yes	No
Hibernate	Preserved	Much faster	Kept	**No**	**Yes**

EC2 Hibernate – Good to Know

- **Supported Instance Families**: C3, C4, C5, I3, M3, M4, R3, R4, T2, T3, ...

- **Instance RAM Size**: must be less than **150 GB**

- **Instance Size**: not supported for **bare metal instances**

- **AMI**: supported OS include Amazon Linux 2, Linux AMI, Ubuntu, RHEL, CentOS, Windows

- **Root Volume**:

  - Must be **EBS**
  - Must be **encrypted**
  - Must be **large**
  - **Instance store** is not supported

- Available for:

  - **On-Demand**
  - **Reserved**
  - **Spot** Instances

- **Limitation**:
  An instance can **NOT** be hibernated for more than **60 days**