More EC2 Topics

Private vs Public vs Elastic IP

Networking have two types of internet process address. IPv4, IPv6.

A public IP address allows internet to reach the host using internet. They are unique on the internet.

A private network has its own set of private IP address that the normal internet cannot access. They are private. The network is exposed only via an internet gateway IP address which allows communication to the outside world. However, the hosts that are inside the private network are allowed to communicate to each other using their respective private IP addresses.

Two different private network can have the same private IPs that is perfectly fine.

Elastic IPS

When you stop and start EC2 instance, the public IP is changed. If you need a fixed public IP then you will need Elastic IP. You attach it to one of the EC2 instance and as long as you don't delete it you will own a IPv4 IP address for your EC2 instance.

So even if you stop and restart your EC2 instance, it will have the same elastic IP.

You can only have 5 elastic IP account, avoid them if possible. Use random public IP and use DNS name to it.

Placement groups

Use placement groups when you want to control over how EC2 is placed in the AWS infrastructure. You don't get to directly access the hardware and placement of EC2 instance but you can make suggestions to AWS.

You can specify three strategies:

- Cluster: Clusters instance into a low-latency group in single availability zone
- Spread: Spread instances across underlying hardware (max 7 instance per group per AZ), use this if you have critical applications
- Partition: Similar to spread, but they are spread across many different partition (which underneath under different sets of racks) within an AZ. Let you scale to 100s of EC2 per

group.

Cluster

EC2 are in the same rack, same hardware, and in the same availability zone.

Gives you super low latency between instances. 10Gbps!

Cons: If the rack fails, all instances fails at the same time. Big risk.

Recommended Big data job that need to be completed fast, application that need extremely low latency and high network throughput between instances.

Spread

All EC2 will be located on different hardware. You can spread it across different availability zone.

Reduce the risk of simultaneous failure since EC2 instances are on different physical hardware.

But you are limited to 7 instances per availability zone per placement group that you have created.

Recommended good for app that need high availability and failures needs to be isolated from each other

Partition

Up to 7 partition in one availability zone. Each partition can contain many EC2 instances. Each partition correspond to a server racket.

Up to 100s of EC2 instances.

A partition failure won't affect the other partitions.

Recommended for big data applications.

Elastic Network Interface

Logical component in a VPC that represents a virtual network card. **They give EC2 instance access to internet and to the private network. They also give the EC2 the public IP and private IP.**

ENI can have attributes:

- Primary private IPv4, one or more secondary IPv4
- One elastic IP per private IPv4. Or one public IP per private IPv4.
- One or more security groups attached to ENI and also a MAC address attach to it

ENI can be created independently and attach on the fly (move them to another EC2) instances.

They are bounded to a specific availability zone.

Good for failover purposes, just attach the ENI which is associated with a private IP to another EC2 instances to cover the failure of the failed instances.

In addition, when you terminate a EC2 instance, the default ENI that's associated will also be deleted. But if you create your own ENI and attach to it they will remain after EC2 instance is deleted. You have more control over the private IP with your own ENI.

EC2 Hibernate

A new state for EC2. The in-memory state is preserved. Instance booting is much faster since OS is not stopped. Whatever is in RAM will be stored to a file in root EBS volume.

When your instance is started again the RAM is restored from the file so in effect, **it is like your instance was never stopped**. It was just restoring the state.

Recommended for long running process, saving RAM state, or for services that takes long time to initialize.

Hibernate is supported for many family, there is a limitation for RAM size. The root volume must be EBS, encrypted.

Available for on-demand, reserved, and spot instances.

You cannot hibernate more than a period of time, 60 days as of recording.

Revision #2
Created 2023-02-09 19:57:32 UTC by Tamarine
Updated 2023-02-09 21:24:55 UTC by Tamarine