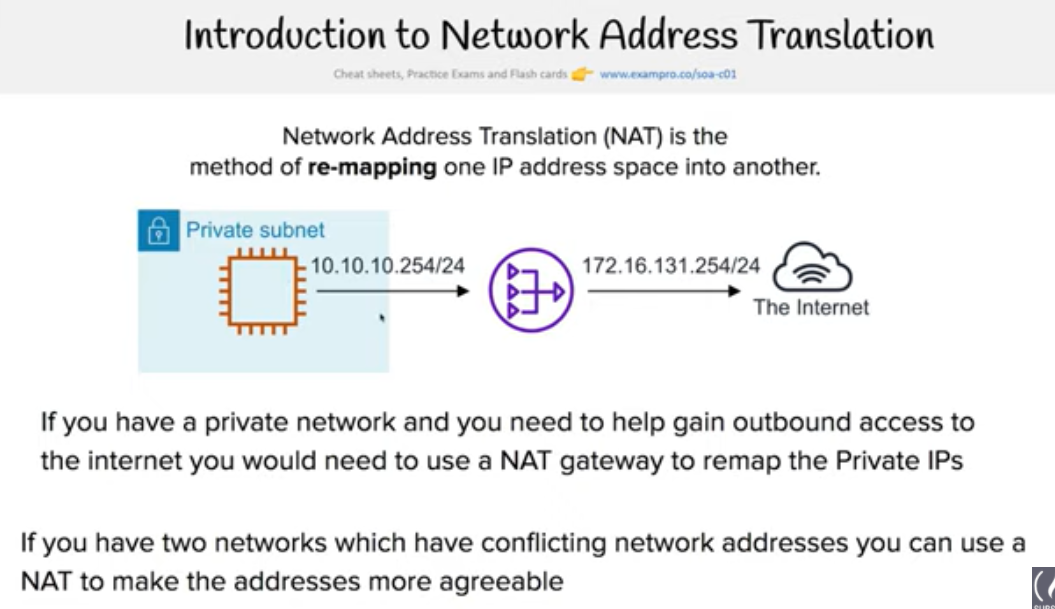
AWS Virtual Private Cloud – VPC

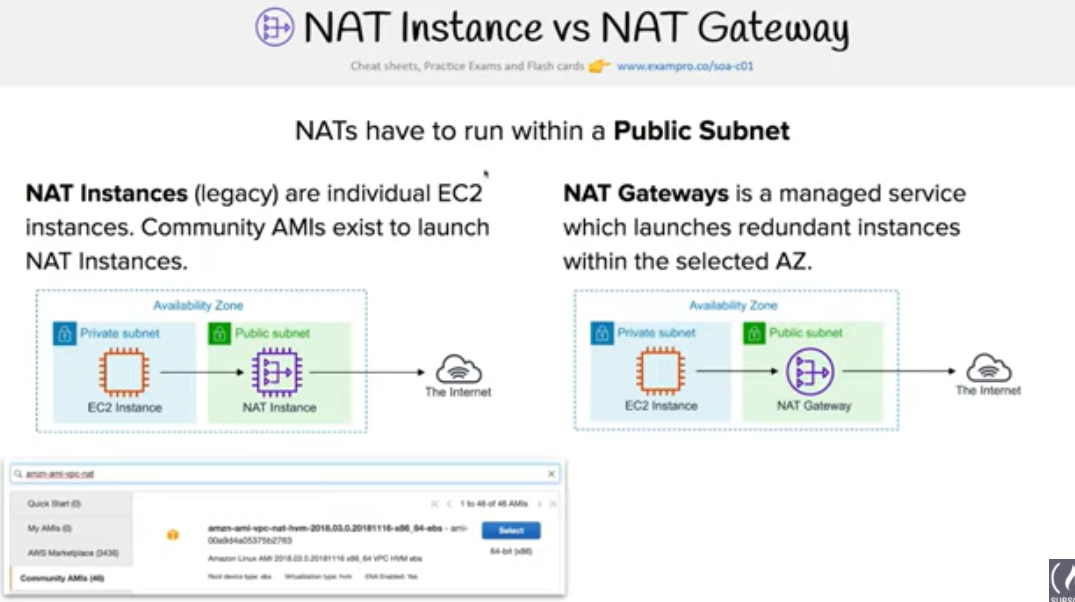
# NAT Overview



* **Network Address Translation (NAT) devices, launched in the public subnet, enables instances in a private subnet to connect to the Internet, but prevents the Internet from initiating connections with the instances.**
* **Instances in private subnets would need internet connection for performing software updates or trying to access external services**
* NAT device performs the function of both address translation and port address translation (PAT)
* NAT instance prevents instances to be directly exposed to the Internet and having to be launched in Public subnet and assignment of the Elastic IP address to all, which are limited.
* **NAT device routes the traffic, from the private subnet to the Internet, by replacing the source IP address with its address and for the response traffic it translates the address back to the instances’ private IP addresses.**
* **AWS allows NAT configuration in 2 ways**
  + **NAT Instance**
  + **NAT Gateway, managed service by AWS**

## NAT device Configuration Key Points

* **needs to be launched in the Public Subnet**
* **needs to be associated with an Elastic IP address (or public IP address)**
* should have the **Source/Destination flag disabled** to route traffic from the instances in the private subnet to the Internet and send the response back
* should have a Security group associated that
  + allows Outbound Internet traffic from instances in the private subnet
  + disallows Inbound Internet traffic from everywhere
* Instances in the private subnet should have the Route table configured to direct all Internet traffic to the NAT device

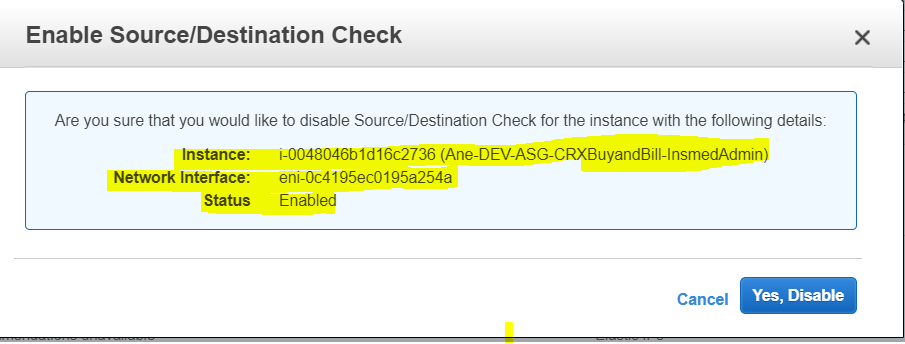


## NAT Gateway

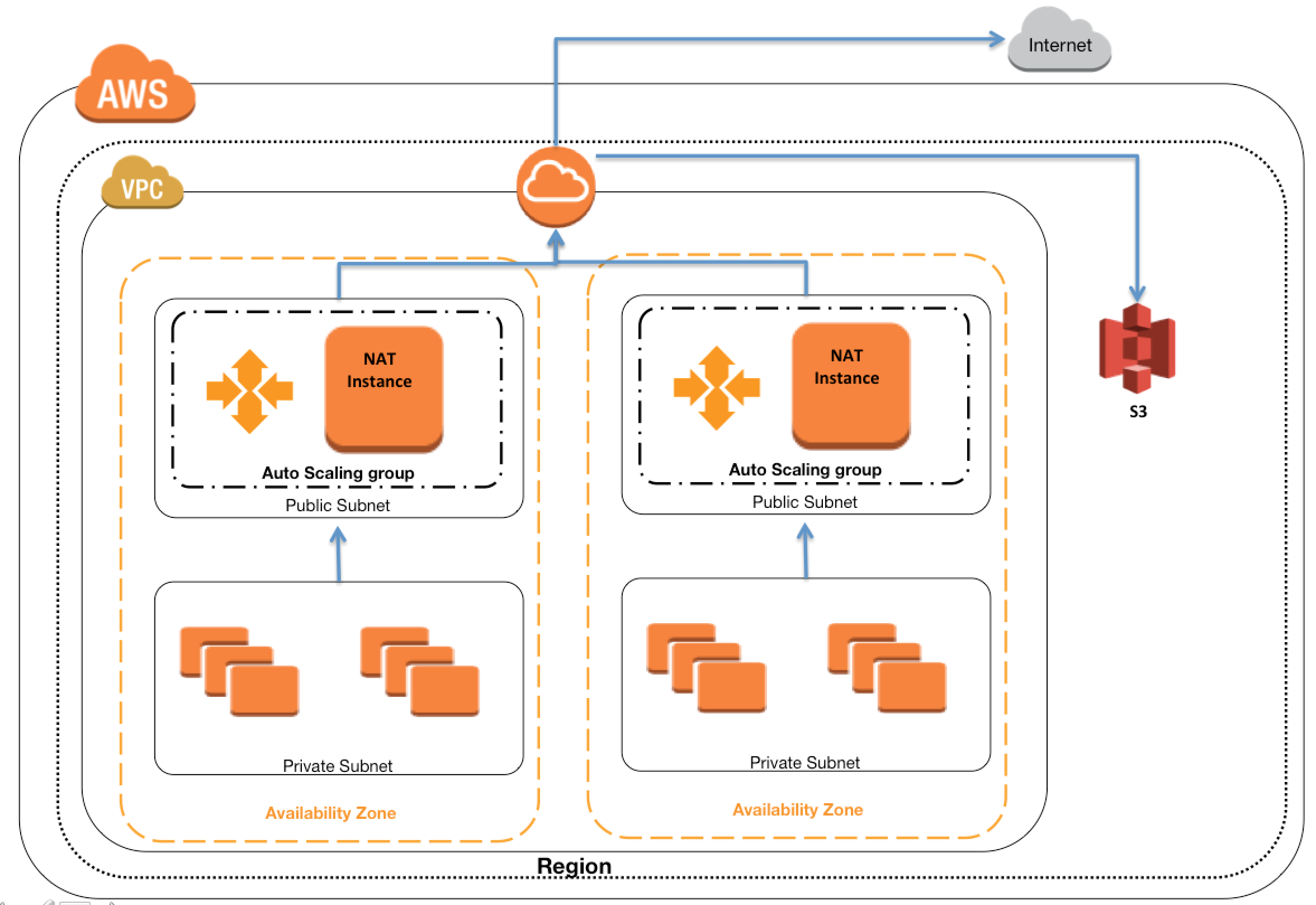
**NAT gateway is a AWS managed NAT service** that provides better availability, higher bandwidth, and requires less administrative effort.

* **A NAT gateway supports bursts of up to 10 Gbps of bandwidth.**
* **For over 10 Gbps bursts requirement, the workload can be distributed by splitting the resources into multiple subnets**, and creating a NAT gateway in each subnet
* **NAT gateway is associated with One Elastic IP address which cannot be disassociated after it’s creation.**
* **Each NAT gateway is created in a specific Availability Zone and implemented with redundancy in that zone.**
* **A NAT gateway supports the following protocols: TCP, UDP, and ICMP.**
* NAT gateway cannot be associated a security group. Security can be configured for the instances in the private subnets to control the traffic
* Network ACL can be used to control the traffic to and from the subnet. NACL applies to the NAT gateway’s traffic, which uses ports 1024-65535
* NAT gateway when created receives an elastic network interface that’s automatically assigned a private IP address from the IP address range of the subnet.
* Attributes of this network interface cannot be modified
* NAT gateway cannot send traffic over VPC endpoints, VPN connections, AWS Direct Connect, or VPC peering connections. Private subnet’s route table should be modified to route the traffic directly to these devices.

## NAT Instance

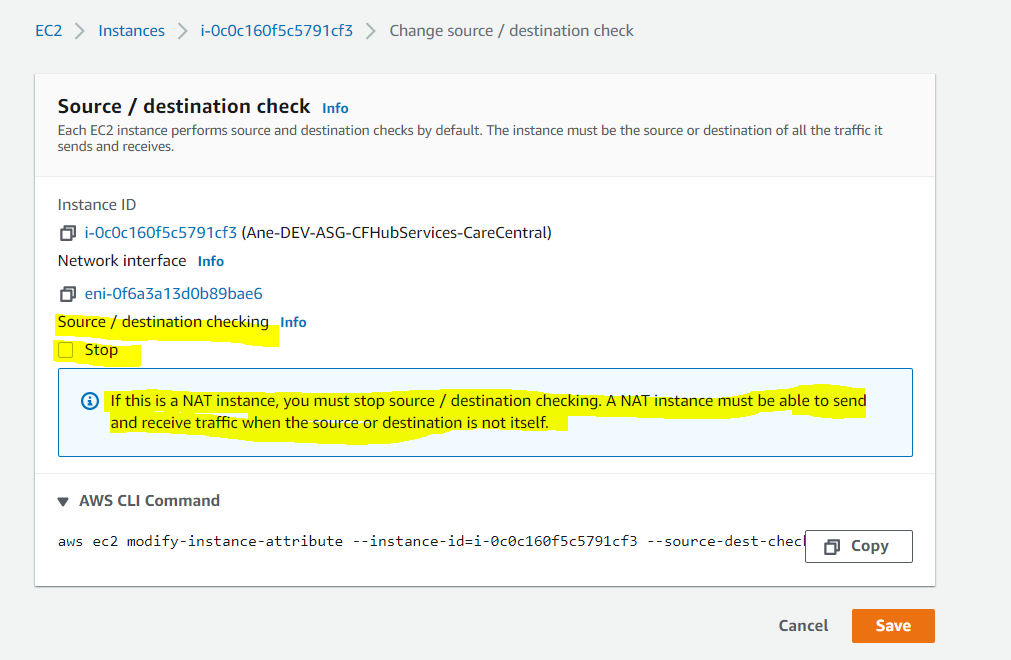
* **NAT instance can be created by using Amazon Linux AMIs configured to route traffic to Internet.**
* **They do not provide the same availability and bandwidth and need to configured as per the application needs.**
* NAT instances must have security groups associated with Inbound traffic enabled from private subnets and Outbound traffic enabled to the Internet
* **NAT instances should have the Source Destination Check attribute disabled, as it is neither the source nor the destination for the traffic and merely acts as a gateway**
* 

### High Availability NAT Instance

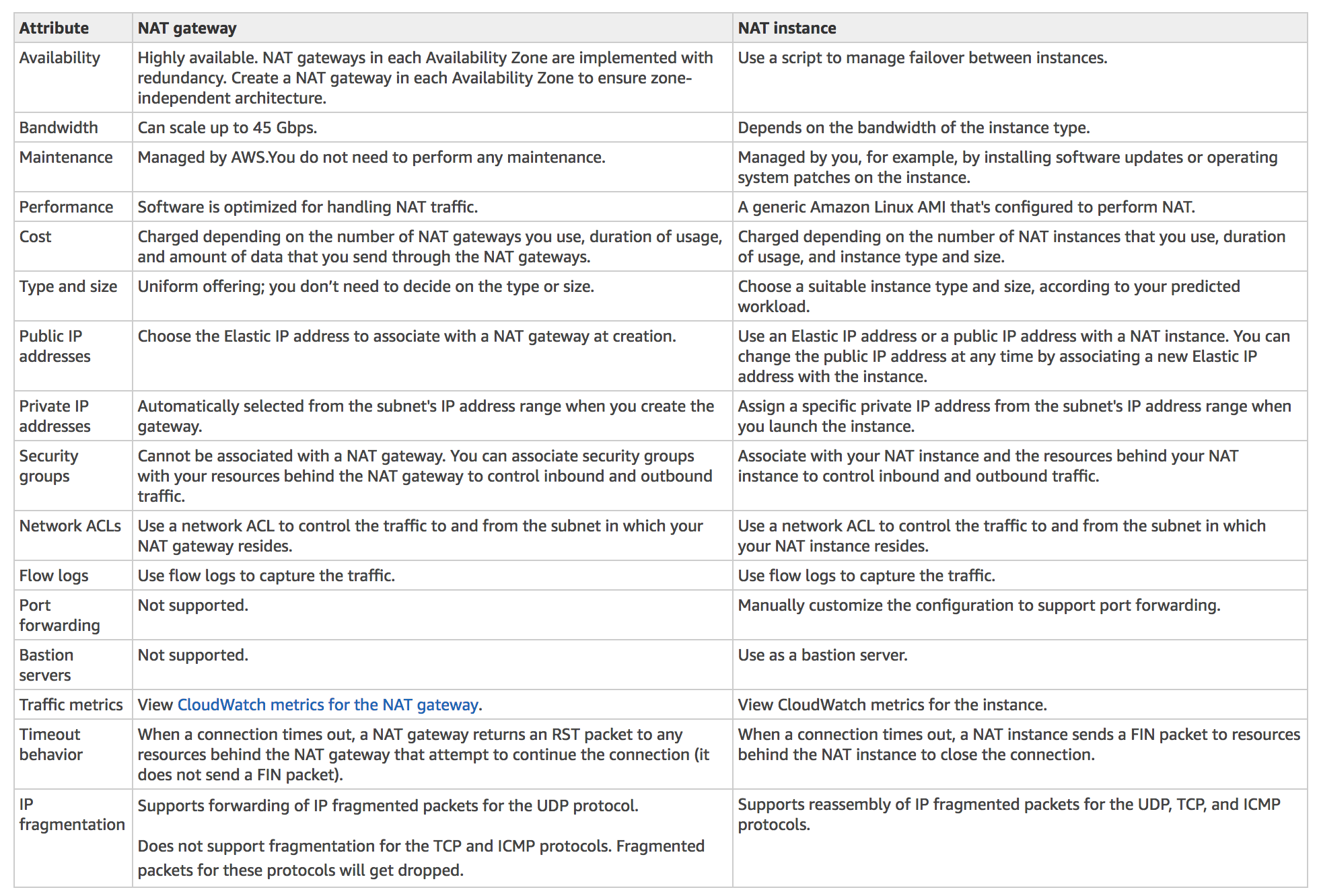


* Create One NAT instance per Availability Zone
* Configure all Private subnet route tables to the same zone NAT instance
* Use Auto Scaling for NAT availability
* Use Auto Scaling group per NAT instance with min and max size set of 1. So if NAT instances fail, Auto Scaling will automatically launch an replacement instance
* NAT instance is highly available with limited downtime
* Let Auto Scaling monitor health and availability of the NAT instance
* Bootstrap scripts with the NAT instance to update the Route tables programmatically
* Keep a close watch on the Network Metrics and scale vertically the NAT instance type to the one with high network performance

### Disabling Source/Destination checks

* Each EC2 instance performs source/destination checks, by default, and the instance must be the source or destination of any traffic it sends or receives
* **However, as the NAT instance acts as a router between the Internet and the instances in the private subnet it must be able to send and receive traffic when the source or destination is not itself.**
* 
* **Therefore, the source/destination checks on the NAT instance should be disabled**

## NAT Gateway & Instance Comparison



## AWS Certification Exam Practice Questions

1. After launching an instance that you intend to serve as a NAT (Network Address Translation) device in a public subnet you modify your route tables to have the NAT device be the target of internet bound traffic of your private subnet. When you try and make an outbound connection to the Internet from an instance in the private subnet, you are not successful. Which of the following steps could resolve the issue?
   1. Attaching a second Elastic Network interface (ENI) to the NAT instance, and placing it in the private subnet
   2. Attaching an Elastic IP address to the instance in the private subnet
   3. Attaching a second Elastic Network Interface (ENI) to the instance in the private subnet, and placing it in the public subnet
   4. **Disabling the Source/Destination Check attribute on the NAT instance**
2. You manually launch a NAT AMI in a public subnet. The network is properly configured. Security groups and network access control lists are property configured. Instances in a private subnet can access the NAT. The NAT can access the Internet. However, private instances cannot access the Internet. What additional step is required to allow access from the private instances?
   1. Enable Source/Destination Check on the private Instances.
   2. Enable Source/Destination Check on the NAT instance.
   3. Disable Source/Destination Check on the private instances
   4. **Disable Source/Destination Check on the NAT instance**
3. A user has created a VPC with public and private subnets. The VPC has CIDR 20.0.0.0/16. The private subnet uses CIDR 20.0.1.0/24 and the public subnet uses CIDR 20.0.0.0/24. The user is planning to host a web server in the public subnet (port 80. and a DB server in the private subnet (port 3306.. The user is configuring a security group of the NAT instance. Which of the below mentioned entries is **not** required for the NAT security group?
   1. For Inbound allow Source: 20.0.1.0/24 on port 80
   2. For Outbound allow Destination: 0.0.0.0/0 on port 80
   3. **For Inbound allow Source: 20.0.0.0/24 on port 80**(Refer [NATSG](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_NAT_Instance.html#NATSG))
   4. For Outbound allow Destination: 0.0.0.0/0 on port 443
4. A web company is looking to implement an external payment service into their highly available application deployed in a VPC. Their application EC2 instances are behind a public facing ELB. Auto scaling is used to add additional instances as traffic increases. Under normal load the application runs 2 instances in the Auto Scaling group but at peak it can scale 3x in size. The application instances need to communicate with the payment service over the Internet, which requires whitelisting of all public IP addresses used to communicate with it. A maximum of 4 whitelisting IP addresses are allowed at a time and can be added through an API. How should they architect their solution?
   1. **Route payment requests through two NAT instances setup for High Availability and whitelist the Elastic IP addresses attached to the NAT instances**
   2. Whitelist the VPC Internet Gateway Public IP and route payment requests through the Internet Gateway. (Internet gateway is only to route traffic)
   3. Whitelist the ELB IP addresses and route payment requests from the Application servers through the ELB. (ELB does not have a fixed IP address)
   4. Automatically assign public IP addresses to the application instances in the Auto Scaling group and run a script on boot that adds each instances public IP address to the payment validation whitelist API. (would exceed the allowed 4 IP addresses)