01 CloudFormation

July 5, 2020

1 Cloud Formations

Cloud formation lets you create and update the things you have in AWS without having to click around on the console or write fragile scripts

With CloudFormation, you define your AWS resources as a YAML script. Then you point CloudFormation to your AWS account, and it creates all the resources you defined.

If you make a change that has already been made, it wont don anything. If you update a part, it will only update that part

Rule of thumb is to allow CloudFormation to deal with all the AWS things that are either static or change very rarely; things including

- VPC configuration
- security groups
- load balancers
- deployment pipelines
- IAM roles

Not for CloudFormation - DynamoDB tables - Kinesis streams - Auto Scaling settings - S3 buckets

2 Diagrams

When you do not actually have the physical serve, as is the case in cloud computing, you have to make diagrams

When your creating the infrastucture, you cant just start working on it, you need to show your team, etc.

Implementing a diagram means to take it and implement it into a script and deploy it into production

3 Lucid Charts

A good diagraming tool program, where you can make a free account

3.1 AWS Container

The first item is an AWS container, it just captures everything in your account

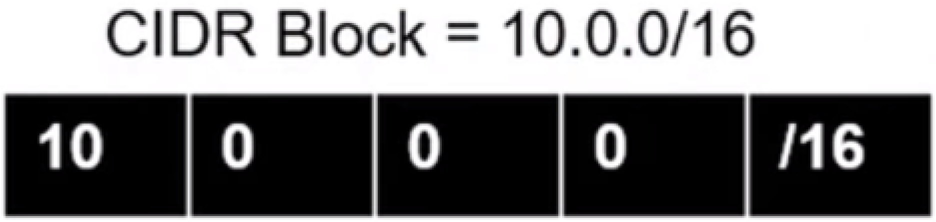
- Users are not inside of your AWS account

3.2 Avalibility Zones

- the ability zone is where your data centers are
- never have only one data center because it can be a single point of faliure
- when they say design diagram with high avaliablity means to have many data centers

4 Virtual Private Cloud

A virtual private cloud is a pool of networked cloud resources. It can span more than one avability zone



CIDR Block = 10.0.0/16

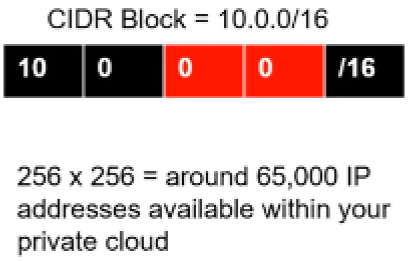
10	0	0	0	/16

The main attribute of the VPC is to block a set of IP adresses or the adress space. This implies the number of avaliable IP adresses that you will have to deploy resources within this private network such as window server, linux servers, databases.

Each number in the $ 10.0.0/16 $ is called an octed because each number has 8 bits in binary

$ 10.0.0/16 $ means reserve the top 16 bits, two octets or first two numbers ot be the fixed part of the network

This is same as how your area code for the phone number $ 718 $ is the same for large group of phone numbers



CIDR Block = 10.0.0/16

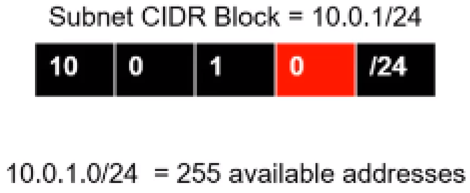
10	0	0	0	/16

256 x 256 = around 65,000 IP addresses available within your private cloud

This leaves you with the last two mumbers, two octives, or 16 bits avaliable for your IP adreses

4.1 Subnets

- create sepreation between resources
- Block or Allow access to/form groups of resources
- provide services to a specified set of resources


Subnet CIDR Block = 10.0.1/24
10
0
1
0
/24
10.0.1.0/24 = 255 available addresses

If our VPC is $ 10.0.0.0/16 $, we could create a subnet called $ 10.0.1.0/24 $. This means that that the the top three bits, totaling $ 24 $ bits are constant. This is with $ 8 $ bits for adresses or $ 255 $ avaliable adresses

The 255 subnets can be created by doing $ 10.0.1.1 $ or $ 10.0.1.2 $ or $ 10.0.1.255 $

The goal of subnets is to use the IP adresses as our key for routing traffic

So for example your database, you want that in a private subnet because you dont want to expose traffic to it

4.2 Subnet Key Points

- A subnet is a subset of the overall VPC network and it only exists in a single availability zone, unlike its parent network the VPC
- A subnet contains resources and can be assigned access rights that apply to all resources within that subnet
- Subnets can be public or private. Public subnets are accessible to external users. Private subnets are only accessed internally by other resources within your cloud container

4.3 Key Points

- VPCs provide you with private IP adresses for your networking resources
- Subnets are smaller subnets of your available IP adress space
- The $ /00 $ at the end is the number of bits, from left to right that are fixed
- Subnets help with routing and services to specific groups of resources
- Create subnets and VPCs with future expansion in mind

4.4 Use IP adresses for routing traffic

- Use IP adresses as the “keys” for routing traffic. We can route traffic to stay within the VPC, or within a prticular subnet, for security reasons
- For example, a database or any sensitive data will be placed in a private subnet. A public server, like a web server, can be placed in a public subnet. Routing rules applied to a subnet allow us to define access to all resources placed inside the subnet

4.5 Internet

If you have a VPC, you need to add an internet gateway

If you had only an internal facing tool that you dont want to connect to the internet, you would use a direct connection or a VPN

The direct connection is basically a physicall line from you to the server

4.5.1 Network Adress Translation (NAT)

Network Adress Translation (NAT) Gateway provides outbound-ony internet gateway for private services to access the internet.

This keeps the private service protected from inbound connections, but allows it to connect to the internet in order to perform functions such as downloading software updates

The NAT gateway serves as an intermediary to take a private resource request, connect to the internet, and then relay the response back to the private resource without exposing that private resource's IP adress to the public

Place NAT gateways inside the public subnets and not the private subnets. NAT gateways need to be in the public subnet so that they can communicate with the public internet, and handle requests from resources that are in a private subnet

Basically, your private subnets cannot conenct directly to the internet, thus they need to use a NAT to connect for them. They then simply use the NAT

5 Autoscaling Groups

An autoscaling group manages multiple instances of the same resource (for example, servers), based on need.

For instance, when there is a lot of internet traffic to a site, the autoscaling group can start more servers and vice versa

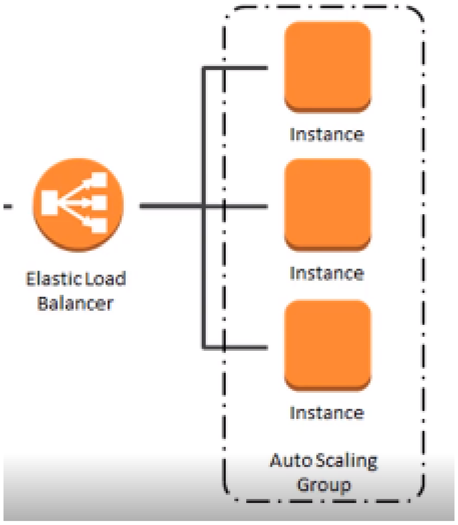
Autoscaling: - Need more than one subnet - It can be used for both high availability and elasticity - Elasticity is the ability to expand and contract your resources to meet demand

5.1 Best Pratices

- It is reccommended that an autoscaling group spans more than one availability zone, for reliabilit
- If we set the autoscaling group to run one resource, it will run that one resource in one of the availability zones
- If there is a failure of that resource, the autoscaling group will shut it down in that availability zone and start that same resource in the other availability zone

6 Load Balancers

- A service designed to distribute work requests meant for a target group
- A target group is a collection of servers providing a common service
- As requests come in, the load balancer will spread the requests evenly across its target group
- load balancer can be combined with an autoscaling group
- load balancer can perform health checks
- without load balancers there would have to be two different urls to access each EC2 instance

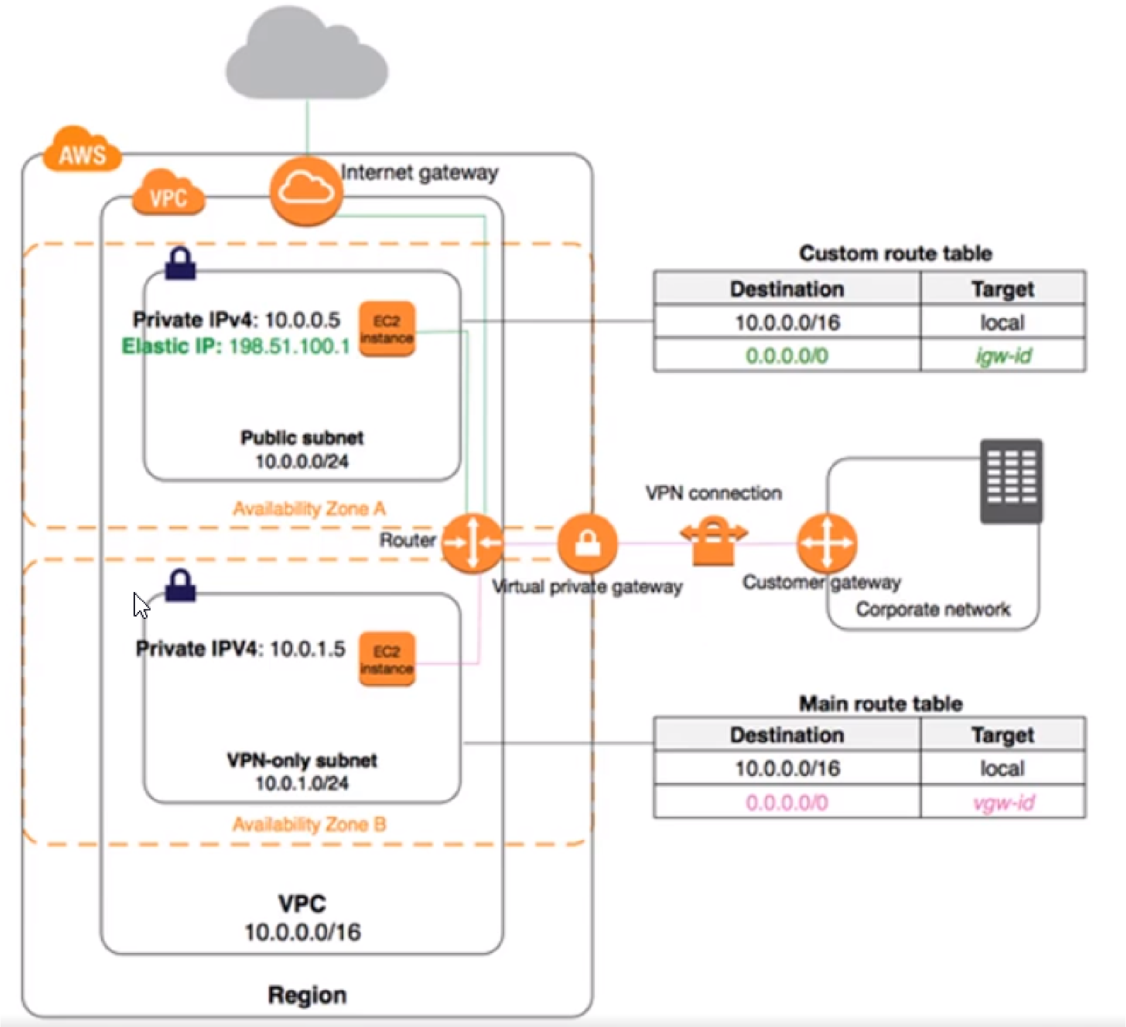


7 Security Group

- Security groups manage traffic at the server level (the resource level)
- Security groups arent for managing higher level groups such as subnets, VPC or user accounts
- The same security group can be assigned to multople resources that require the same security access settings defined by the security group

8 Routing Table

- Route tables allow the routing of traffic to or away from your network
- A set of entries or rules associated with one or more of your subnets inside of your VPC
- These rules allow or deny traffic to/from the address ranges that you can specify
- Rules can be as open as the entire world or restricted to a single IP adress



Custom route table

Destination	Target
10.0.0.0/16	local
0.0.0.0/0	igw-id

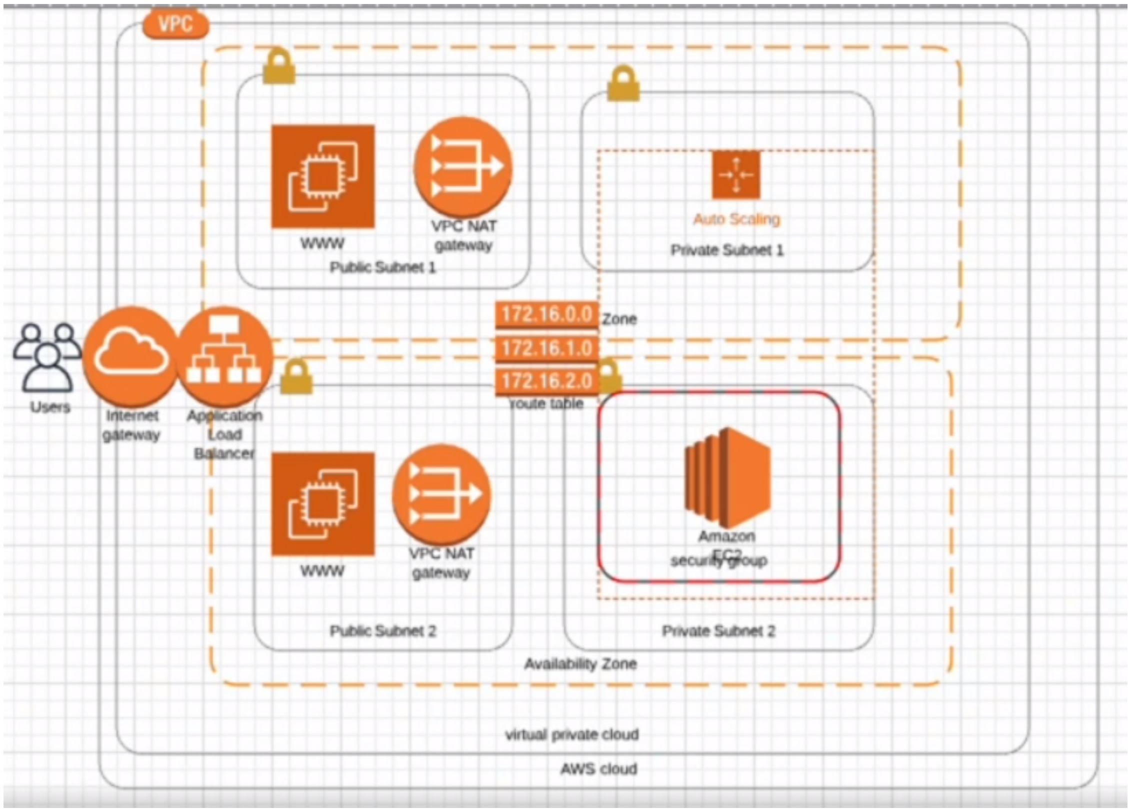
Main route table

Destination	Target
10.0.0.0/16	local
0.0.0.0/0	vgw-id

- What makes the route of the private network private, is that its target is the `vgw-id` which poitns to the virtual private gateway
- note that the `local` on the tables means that the same subnets in the VPN can talk to each other
- the `igw-id` refers to the internet gateway
- follow the pink and green lines

9 S3

- An S3 bucket is a public service for users to upload or download files
- Place the S3 service outside of your VPC
- This is because S3 is an service and thus you require internet traffic to reach


VPC
WWW
VPC NAT gateway
Public Subnet 1
Auto Scaling
Private Subnet 1
172.16.0.0
Zone
172.16.1.0
172.16.2.0
route table
Users
Internet gateway
Application Load Balancer
WWW
VPC NAT gateway
Amazon EC2
security group
Public Subnet 2
Private Subnet 2
Availability Zone
virtual private cloud
AWS cloud