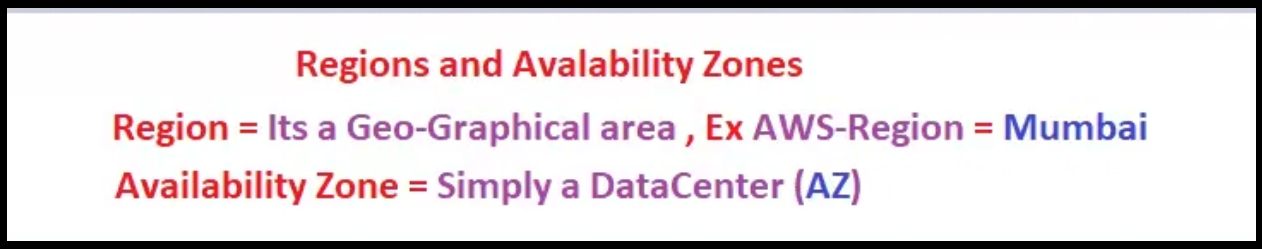
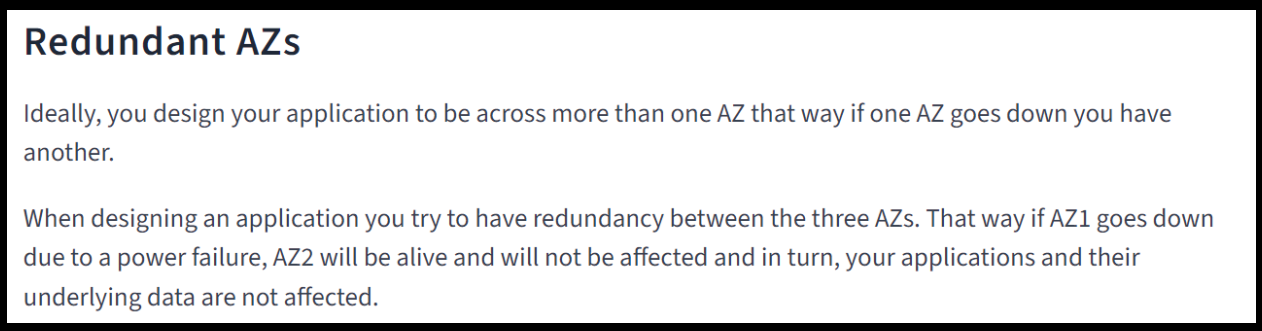
Region and AZs

Region is place where AWS has infrastructure!!
Region has multiple data centres!!
That data centre is called as AZs!!

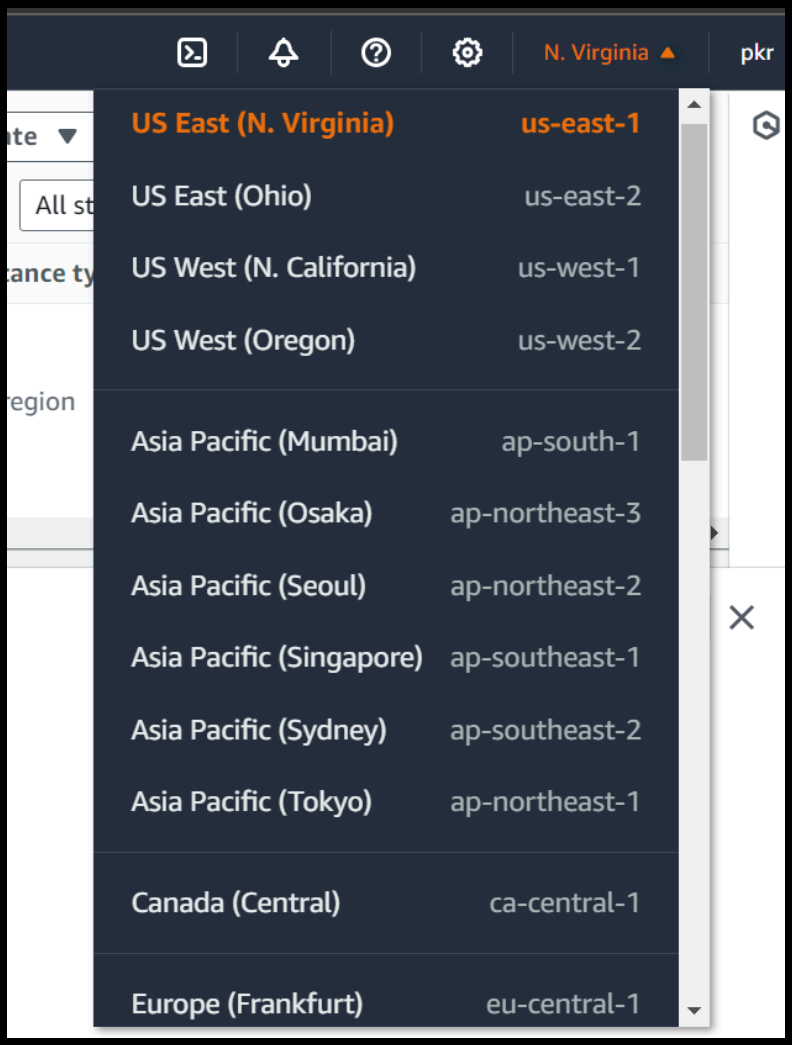
Data centre==AZs
So servers are present in AZs!!

We have mumbai and hyderabad region!!



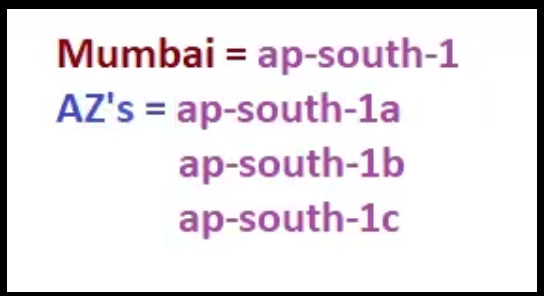


Every region has a code like mumbai has ap-south-1



See the codes above!! These are region!!
Ap→asia pacific

Mumbai has 3 AZs

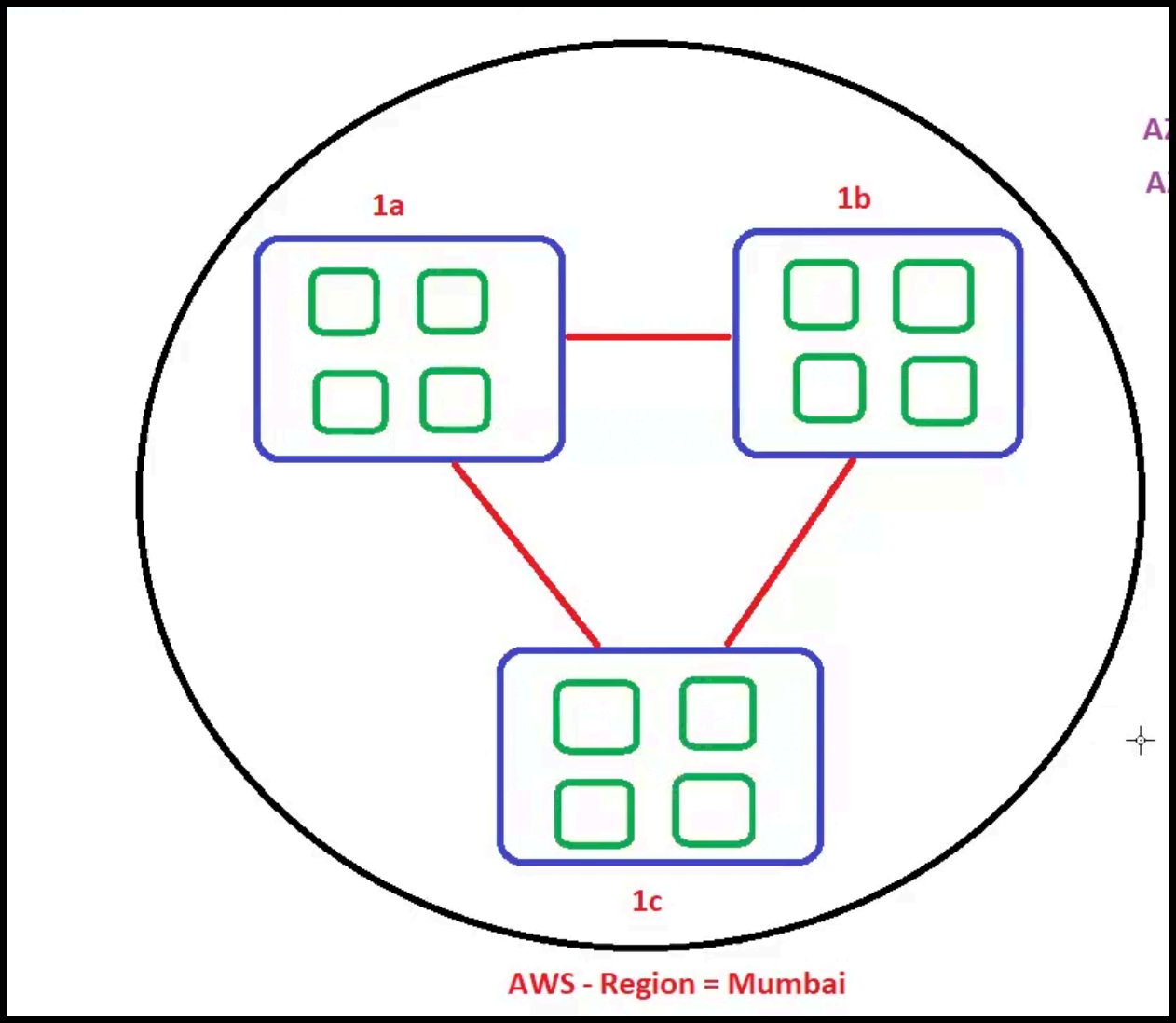


To get Azs code just put a ,b,c at the end!!

Regions and Azs are managed byAWS!!

Now we launch 10 ec2 instances. Will we put all in one AZ or at multiple AZs?
If we put at 1 Azs,if that one Azs is down then we will not get response!!

So we distribute!!
But all AZs are connected by internet !! so can launch few in one Az and other is some other Az
Suppose we are talking about mumbai region!! So 3 Azs



Red line tells interconnected AZs
Green boxes are ec2 instances!!

AWS has Global Infrastructure

Region is a place where AWS has its Infrastructure

A Region contains multiple DataCenters

A Region contains multiple AZ's

Servers = Instances

Servers / Instances are placed in AZ's

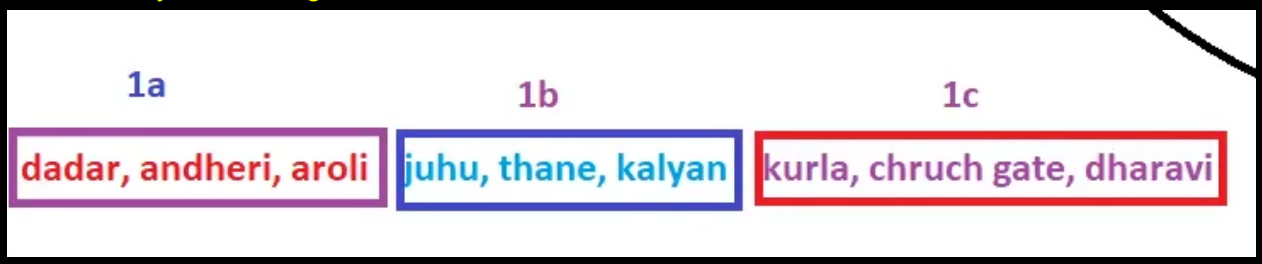
AZ's are sync with each other (network), not data

If required we can share the data to other AZ's in same Region.

Best practise is to distribute the EC2 instances across multiple AZ's

By default data will not be sync across multiple AZs you need to sync them !!
Till now ,there is no such time when Az goes down!!

1a is not a data center its a group of data center!!
So thats why AZ never goes down!



A ec2 instance in 1a wil not be replicated to 1b or 1c!!

AZ's are sync with each other (network), not data

If required we can share the data to other AZ's in same Region.

Best practise is to distribute the EC2 instances across multiple AZ's

very very less chances that 1AZ goes down

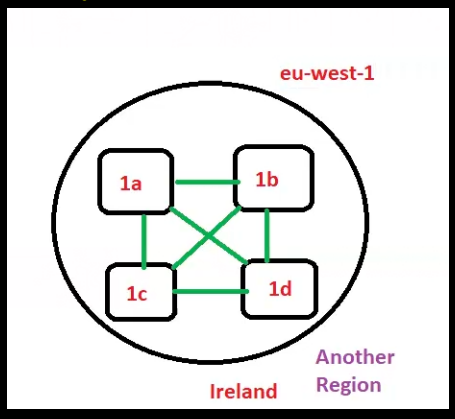
1a or 1b or 1c = group of DC's

1 AZ is a group of DC's

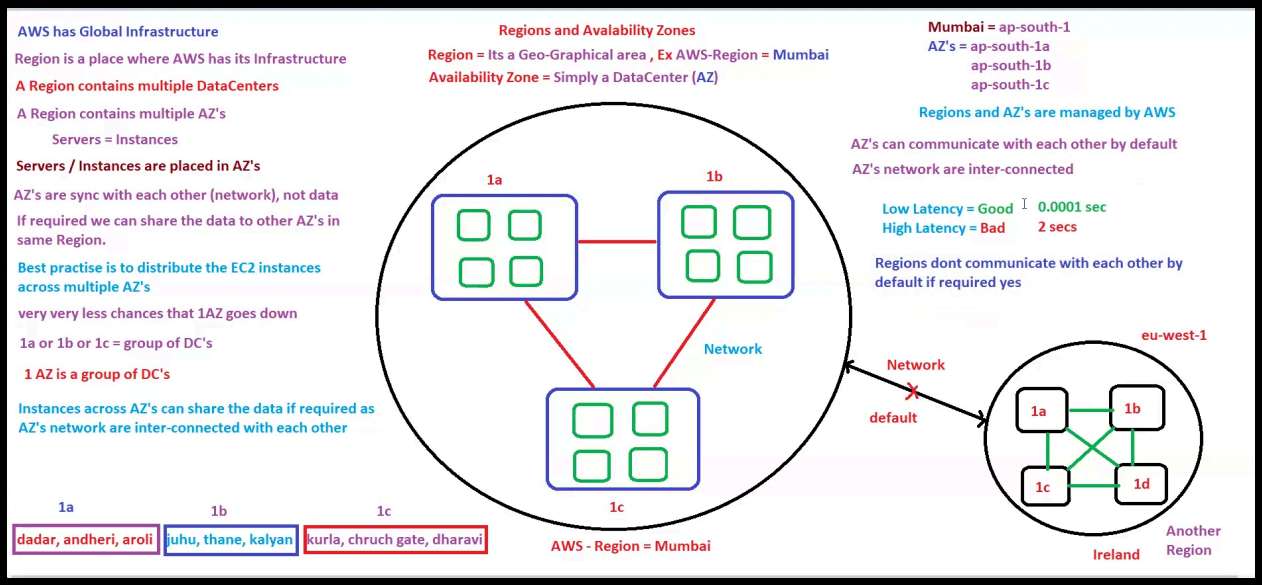
Instances across AZ's can share the data if required as AZ's network are inter-connected with each other

Latency
1.high latency→ not wanted
2.low latency→we want (time need to load a website) (very low response time)

If customer from India set server in India only not in germany as germany servers will have high latency!!



Now we have another region!!
Now Two Azs in a region can communicate with each other
But two regions cant communicate with each other!!
Mumbai region cant communicate with ireland by default!If required they can communicate!



What is EC2?

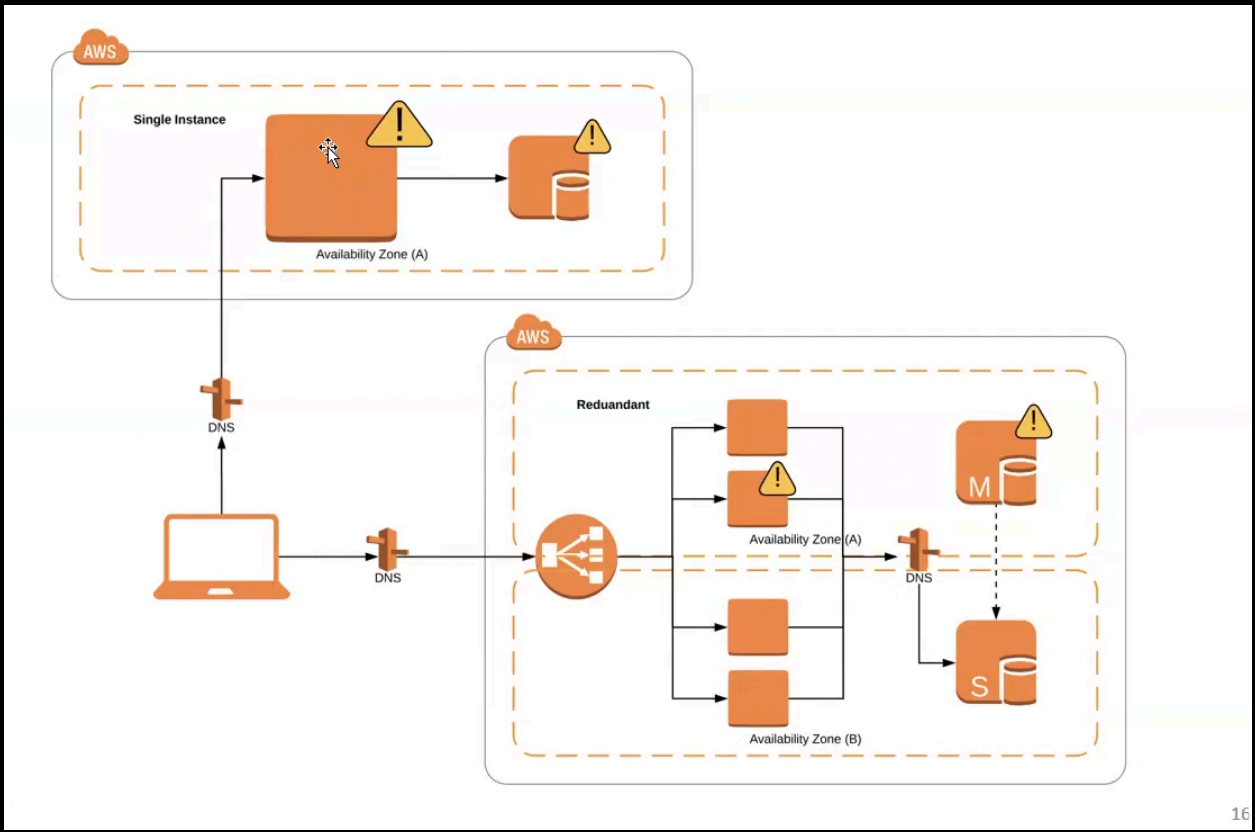
This is a web service that provides a re-sizable compute in the cloud.

AWS EC2 reduces the time required to obtain and boot new server instances to minutes allowing you to quickly scale capacity, both up and down, as your computing requirements change.

In the old world if you need an app server or DB server, you would first need to talk to your developer, decide the size and number of cores you need, and then talk to your procurement team, and by the time the server is in your data center it could take 2+ months.

Now with EC2, you have this with a click of a button. This also means from a startup perspective you do not have the upfront cost of buying all the hardware you need.

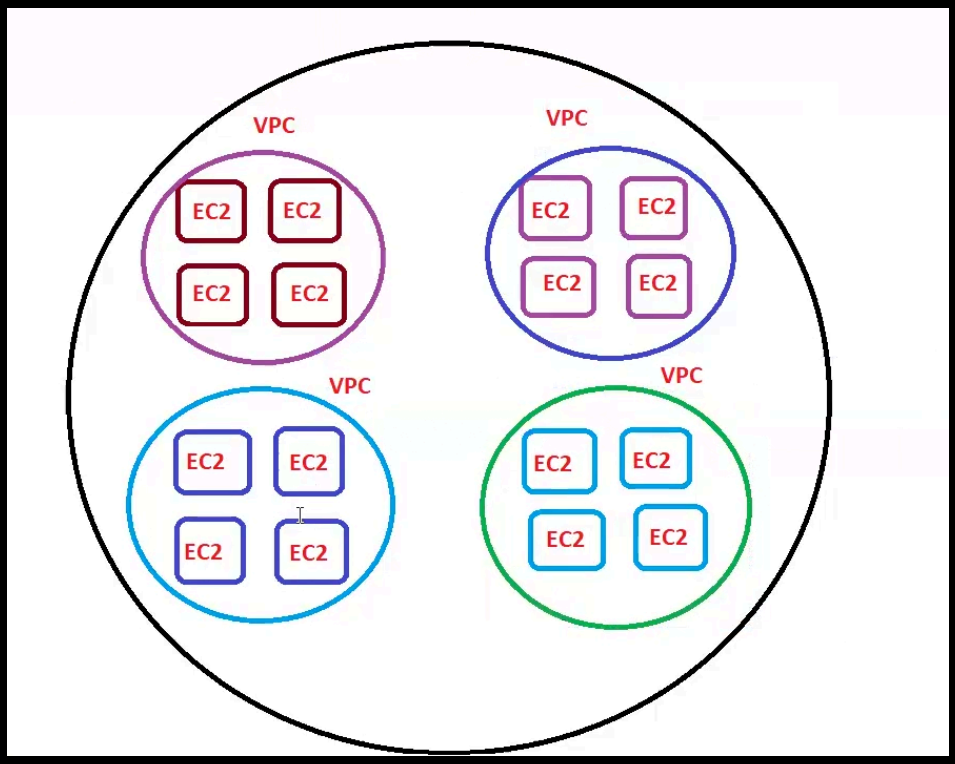
EC2 changes the economics of computing by allowing you to pay only for the capacity that you actually use. You also have several tools at the disposal of the developers to build the applications to be resilient and isolate them from failure scenarios.



Orange boxes are ec2 instances!! Exclamation mark tells its down!
Top one is having 1 ece so no redundancy so not highly available!

In down load balancer distributes traffic across multiple Az , 2 machines in Azs so yes highly available !!
Load balancer is regional device!! It can distribute traffic across multiple ec2-instances (servers) across Multiple Azs in region

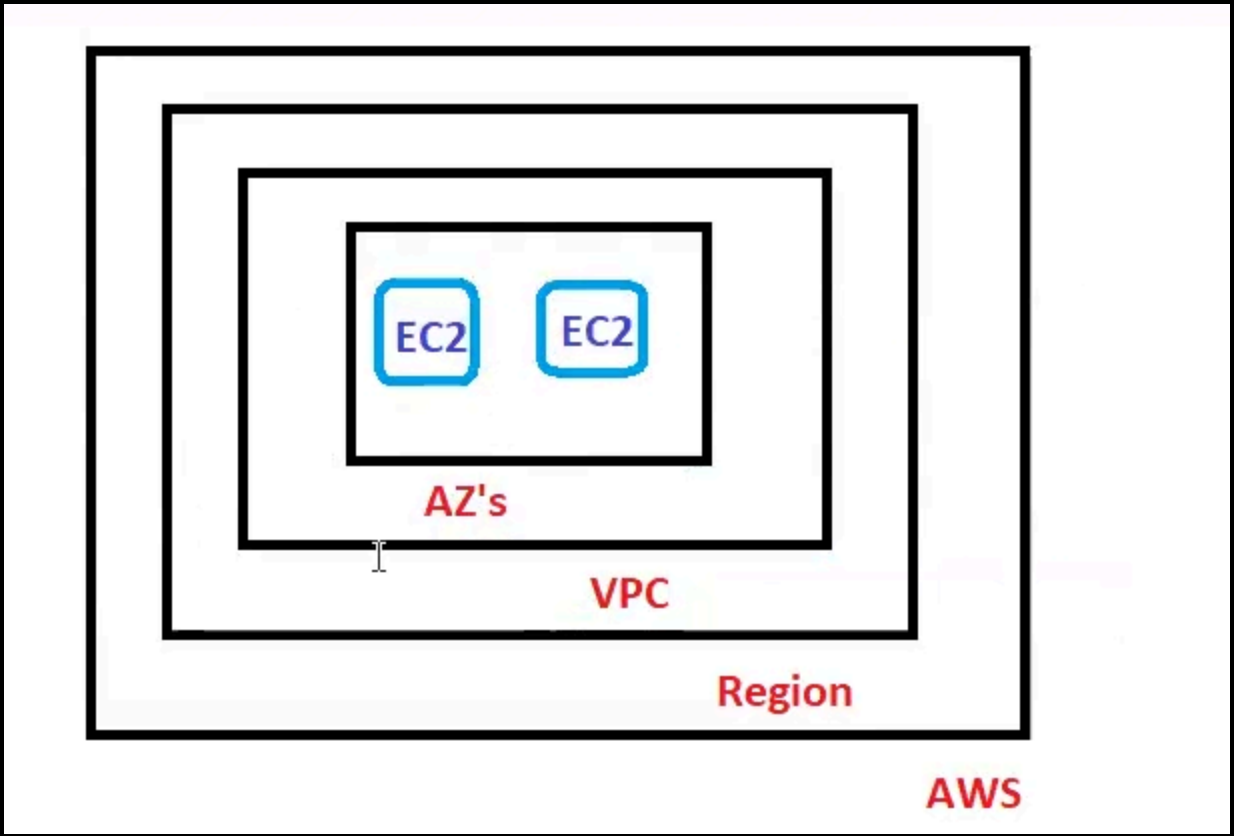
A,B,C,D initialises some servers
Now we dont want different people Ec2 instance to communicate with each other
So we need a private cloud so for that in which all ec2 instance of person can communicate but outside people cant!!
So we use VPC(virtual private cloud) for that!!



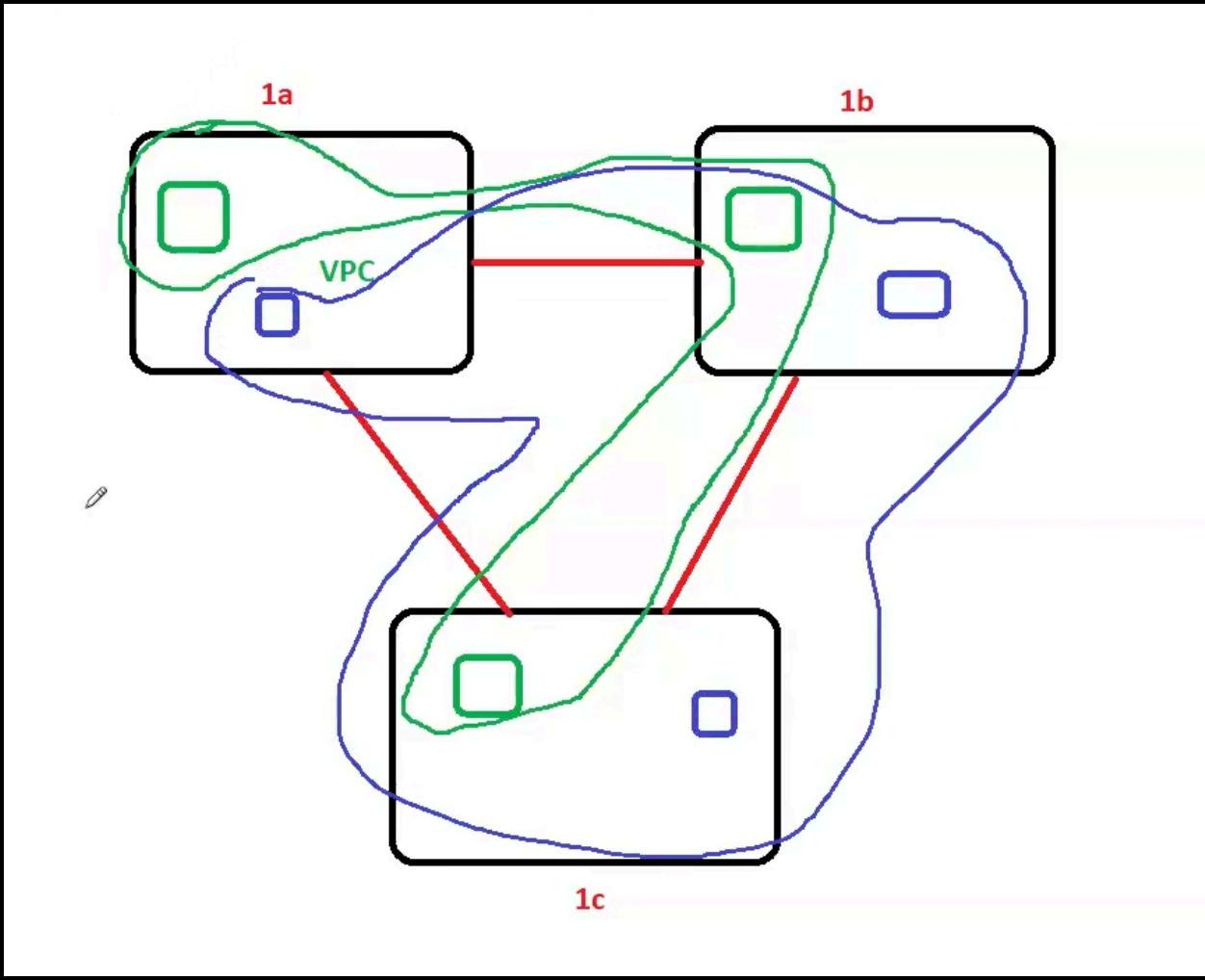
Inner circles are VPC!!
Two Vpcs cant communicate with each other by default!
But if required can communicate!

You went to sulabh sauchalaya its public but in that we have private compartment thats like it!!

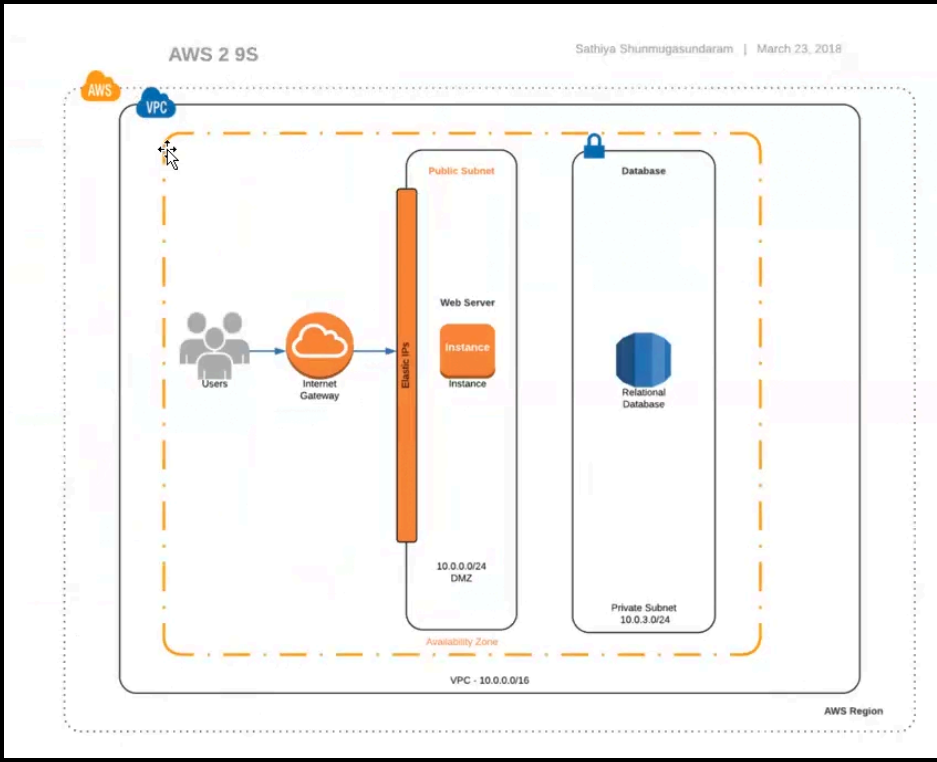
VPC is regional !! Vpc you can create for region!! At max you can have 5 VPC's!



VPC is like private cloud ab poora aws thdi private kr doge islie we put limit of 5 max VPC's!



See above green server are in one ec2 instance and purple one in another!! See vpc expands across Az's



This is not Highly available as only one ec2 instance no redundancy !