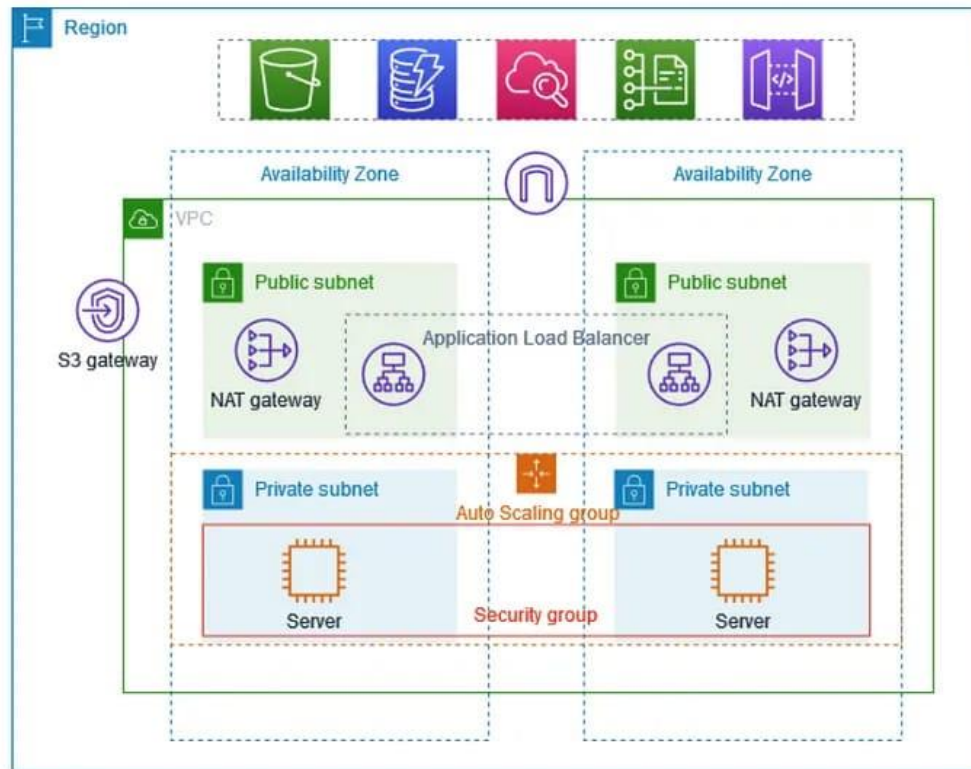


Our requirement is to create VPC and two private instances with auto scaling to show work of two web servers as shown in below figure



There are some steps to complete this project

STEP 1:

1. Go to aws console and sign-in to aws account
2. Choose the vpc through search bar

Note: If we choose (vpc only) option then we have to create components step by step manually

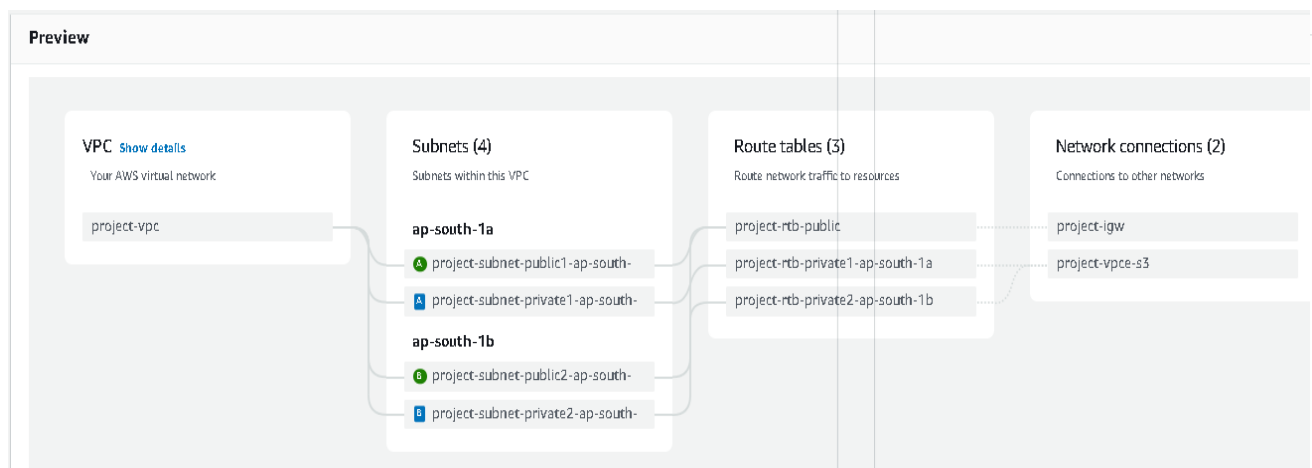
STEP 2:

1. Create one vpc
2. Create four subnets as two public subnets and two private subnets for private subnets choose two different availability zones

3. Create one public route table and two private route tables
4. Update the subnet associations as (public subnets to public routes) and (private subnets to private routes)
5. Create one internet gateway and attach to public route tables
6. Create one NAT gateway and attach to private route tables

Note: If we choose (vpc and more) option and select the components as our requirements then it will create the connections automatically

1. You can see all the component connections in preview
2. As shown in below figure



STEP 3:

1. Go to EC2 service from vpc
2. Select the launch templates
3. Create the launch templates and give the name for launch template
4. Choose the instance type as free tier as shown in below figure

▼ Instance type
Info | Get advice
Advanced

t2.micro
Free tier eligible

Family: t2 1 vCPU 1 GiB Memory Current generation: true
On-Demand Linux base pricing: 0.0124 USD per Hour
On-Demand Windows base pricing: 0.017 USD per Hour
On-Demand RHEL base pricing: 0.0268 USD per Hour
On-Demand SUSE base pricing: 0.0124 USD per Hour

☐ All generations
[Compare instance types](#)

Additional costs apply for AMIs with pre-installed software

5. Select the created key pair or generate a new key pair
6. Update the network settings by clicking on edit option
 - No need to change subnet
 - Click on create security group
 - enter the security group name
 - enter the description it's your choice
 - select the created vpc for the project
 - update the inbound rules
 - one is ssh and another is http port 80
 - launch the templates
 - do this as shown in below figure

Network settings [Info](#)

Subnet [Info](#)

Don't include in launch template [Create new subnet](#)

When you specify a subnet, a network interface is automatically added to your template.

Firewall (security group) [Info](#)

A security group is a set of firewall rules that control the traffic for your instances. Add rules to allow specific traffic to reach your instances.

☐ Select existing security group ☒ Create security group

Security group name - required

project-sg

This security group will be added to all network interfaces. This name can't be edited after the security group is created. Max length is 255 characters. Valid characters: a-z, A-Z, 0-9, space, and _ (underscore). * = 255

Description - required [Info](#)

allow all traffic

VPC [Info](#)

vpc-04c036a96dddb46f (project-vpc1-vpc) 10.0.0.0/16

Inbound Security Group Rules

Security group rule 1 [TCP, 22, 0.0.0.0/0] [Remove](#)

Type	Protocol	Port range
ssh	TCP	22

Source type: Anywhere

Source: 0.0.0.0/0

Description - optional: e.g. SSH for admin desktop

Security group rule 2 [TCP, 80, 0.0.0.0/0] [Remove](#)

Type	Protocol	Port range
HTTP	TCP	80

Source type: Anywhere

Source: 0.0.0.0/0

Description - optional: e.g. SSH for admin desktop

STEP 4:

1. Go to auto scalling group
2. Create the auto scalling group
3. Give the name for auto scalling group
4. Select the created launch templats
5. Choose instance type requirement
 - Choose VCPUs minimum and maximum
 - (2 is minimum) and (3 is maximum)
 - Choose memory minimum and maximum
 - (4 is minimum) and (8 is maximum)
 - Select as shown in below figure

Instance type requirements [Info](#) Reset to launch template

You can keep the same Instance attributes or Instance type from your launch template, or you can choose to override the launch template by specifying different Instance attributes or manually adding Instance types.

☒ **Specify instance attributes**
Provide your compute requirements. We fulfill your desired capacity with matching Instance types based on your allocation strategy selection.

☐ **Manually add instance types**
Add one or more Instance types. Any of the Instance types may be launched to fulfill your desired capacity based on your allocation strategy selection.

Required instance attributes
Enter your compute requirements in virtual CPUs (vCPUs) and memory.

vCPUs
Enter the minimum and maximum number of vCPUs per Instance.

minimum maximum

☐ No minimum ☐ No maximum

Memory (GiB)
Enter the minimum and maximum GiBs of memory per Instance.

minimum maximum

☐ No minimum ☐ No maximum

Additional instance attributes - optional
Add Instance attributes to further limit which Instance types may be used to fulfill your desired capacity.

Add attribute

► **Preview matching instance types (31)**
This list includes all the Instance types that match your compute requirements. Amazon EC2 may provision from any of these Instance types. The exact Instance types that are used to fulfill your desired capacity depend on the allocation strategy you choose and available capacity.

6. Select created vpc
7. Select 2 private availability zones as shown in below figure

Network [Info](#)

For most applications, you can use multiple Availability Zones and let EC2 Auto Scaling balance your instances across the zones. The default VPC and default subnets are suitable for getting started quickly.

VPC

Choose the VPC that defines the virtual network for your Auto Scaling group.

vpc-04c4936a86dddb46f (project-vpc1-vpc)
10.0.0.0/16



[Create a VPC](#)

Availability Zones and subnets

Define which Availability Zones and subnets your Auto Scaling group can use in the chosen VPC.

Select Availability Zones and subnets



ap-south-1a | subnet-0242f562295a21bcf (project-
vpc1-subnet-private1-ap-south-1a) X
10.0.128.0/20

ap-south-1b | subnet-0b6eeaaacddfe49a5 (project-
vpc1-subnet-private2-ap-south-1b) X
10.0.144.0/20

[Create a subnet](#)

8. No need to select load balancer leave as no load balancer
9. Choose desired capacity and scaling option as shown in below figure

Group size [Info](#)

Set the initial size of the Auto Scaling group. After creating the group, you can change its size to meet demand, either manually or by using automatic scaling.

Desired capacity type

Choose the unit of measurement for the desired capacity value. vCPUs and Memory(GiB) are only supported for mixed instances groups configured with a set of instance attributes.

Units (number of instances)

Desired capacity

Specify your group size.

2

Scaling [Info](#)

You can resize your Auto Scaling group manually or automatically to meet changes in demand.

Scaling limits

Set limits on how much your desired capacity can be increased or decreased.

Min desired capacity

2

Equal or less than desired capacity

Max desired capacity

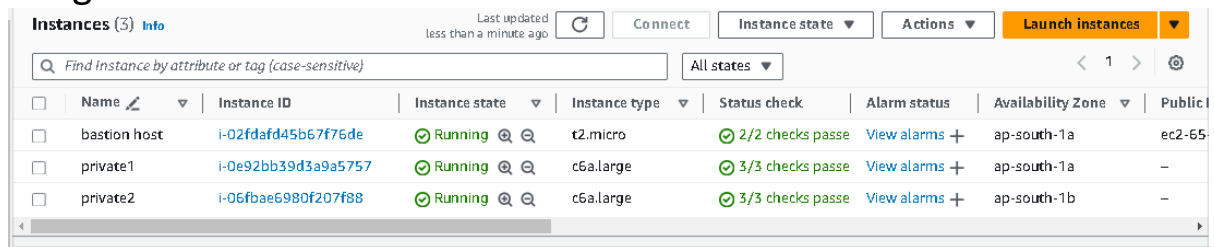
3

Equal or greater than desired capacity

10. Auto scaling group is created

STEP 5:

1. Go to instance
2. Check the instance their 2 private instances are created from auto scalling group
3. Give the name for 2 private instances as private 1 and private 2 to avoid confusions
4. Create one public instance named as bastion host as shown in figure



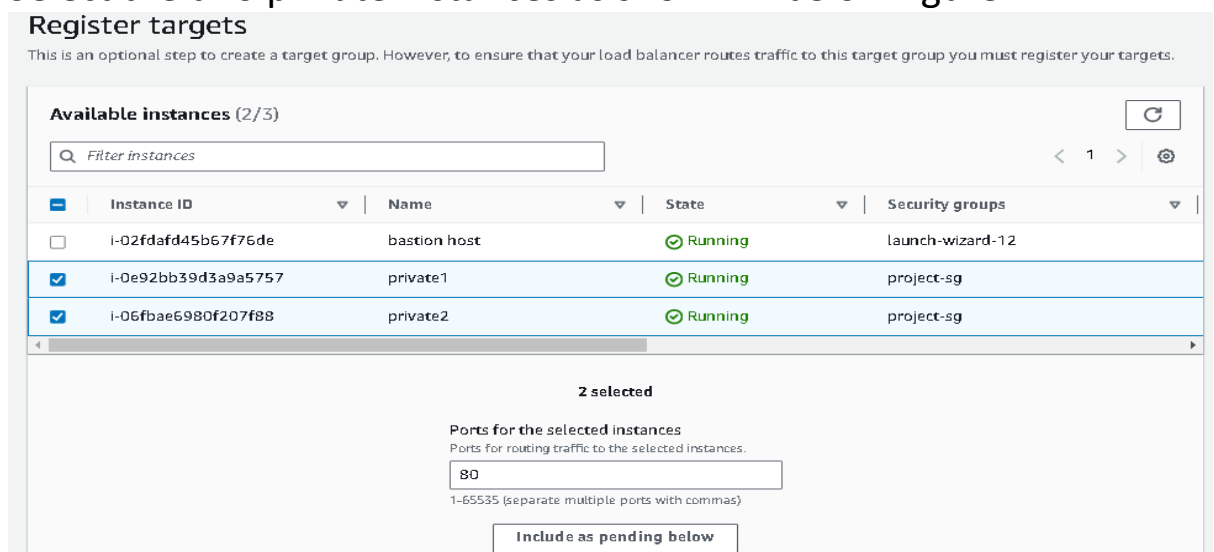
The screenshot shows the AWS EC2 Instances console. At the top, there are buttons for 'Connect', 'Instance state', 'Actions', and 'Launch instances'. Below these is a search bar and a table of instances. The table has columns for Name, Instance ID, Instance state, Instance type, Status check, Alarm status, Availability Zone, and Public. Three instances are listed: 'bastion host' (Instance ID: i-02fdafd45b67f76de, Instance type: t2.micro, Status check: 2/2 checks passed), 'private1' (Instance ID: i-0e92bb39d3a9a5757, Instance type: c6a.large, Status check: 3/3 checks passed), and 'private2' (Instance ID: i-06fbae6980f207f88, Instance type: c6a.large, Status check: 3/3 checks passed).

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public
bastion host	i-02fdafd45b67f76de	Running	t2.micro	2/2 checks passed	View alarms +	ap-south-1a	ec2-65
private1	i-0e92bb39d3a9a5757	Running	c6a.large	3/3 checks passed	View alarms +	ap-south-1a	-
private2	i-06fbae6980f207f88	Running	c6a.large	3/3 checks passed	View alarms +	ap-south-1b	-

5. From that public instance we can connect the private through ssh client
6. completing the terminal work for two private instances

STEP 6:

1. Go to target groups
2. Create the target group
3. Give the name for target group
4. Select the two private instances as shown in below figure



The screenshot shows the 'Register targets' page in the AWS EC2 console. It includes a search bar for 'Filter instances' and a table of available instances. Two instances are selected: 'private1' (Instance ID: i-0e92bb39d3a9a5757) and 'private2' (Instance ID: i-06fbae6980f207f88). Below the table, it says '2 selected' and 'Ports for the selected instances'. A text input field contains the port number '80'. At the bottom, there is a button labeled 'Include as pending below'.

Register targets

This is an optional step to create a target group. However, to ensure that your load balancer routes traffic to this target group you must register your targets.

Available instances (2/3)

Instance ID	Name	State	Security groups
i-02fdafd45b67f76de	bastion host	Running	launch-wizard-12
i-0e92bb39d3a9a5757	private1	Running	project-sg
i-06fbae6980f207f88	private2	Running	project-sg

2 selected

Ports for the selected instances

Ports for routing traffic to the selected instances.

80

1-65535 (separate multiple ports with commas)

Include as pending below

5. Click on include as pending below
6. Click on create the target group then target group is created

STEP 7:

1. Go to load balancer
2. Click on create load balancer
3. Give the name for load balancer
4. Update network mapping
 - Choose created vpc
 - Choose two availability zones as shown in figure

Network mapping [Info](#)

The load balancer routes traffic to targets in the selected subnets, and in accordance with your IP address settings.

VPC [Info](#)

The load balancer will exist and scale within the selected VPC. The selected VPC is also where the load balancer targets must be hosted unless routing to Lambda or on-premises targets, or using VPC peering. To confirm the VPC for your targets, view [target groups](#). For a new VPC, [create a VPC](#).

project-vpc1-vpc
vpc-04c4936a86dddb46f
IPv4 VPC CIDR: 10.0.0.0/16

Mappings [Info](#)

Select at least two Availability Zones and one subnet per zone. The load balancer routes traffic to targets in these Availability Zones only. Availability Zones that are not supported by the load balancer or the VPC are not available for selection.

Availability Zones

☒ ap-south-1a (aps1-a21)

Subnet

subnet-0242f562295a21bcb
IPv4 subnet CIDR: 10.0.128.0/20

project-vpc1-subnet-private1-ap-south-1a

IPv4 address

Assigned by AWS

☒ ap-south-1b (aps1-a23)

Subnet

subnet-0b6eeaaacddfe49a5
IPv4 subnet CIDR: 10.0.144.0/20

project-vpc1-subnet-private2-ap-south-1b

The selected subnet does not have a route to an Internet gateway. This means that your load balancer will not receive Internet traffic. You can proceed with this selection; however, for Internet traffic to reach your load balancer, you must update the subnet's route table in the [VPC console](#).

5. Update the created security group as shown in below figure

Security groups [Info](#)

A security group is a set of firewall rules that control the traffic to your load balancer. Select an existing security group, or you can [create a new security group](#).

Security groups

Select up to 5 security groups

project-sg
sg-05a11c9b26e48448c VPC: vpc-04c4936a86dddb46f

6. Click on create load balancer then load balancer is created
7. Copy the load balancer DNS name to access the two web servers

Load balancers (1)

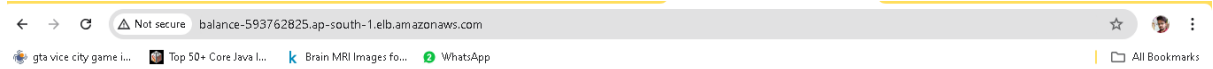
Elastic Load Balancing scales your load balancer capacity automatically in response to changes in incoming traffic.

Filter load balancers

	Name	DNS name	State	VPC ID	Availability Zones	Type
<input type="checkbox"/>	balance	balance-593762825.ap-s...	Active	vpc-0c5f4cad6ad8ae...	2 Availability Zones	application

STEP 8:

1. Go to Google page or new web page
2. Paste the copied DNS name on URL path
3. You can see the two created web servers as shown in figure



This is web server 1

4. Just re-refresh the page then you see second server



This is web server 2

That's it, our requirement is done.