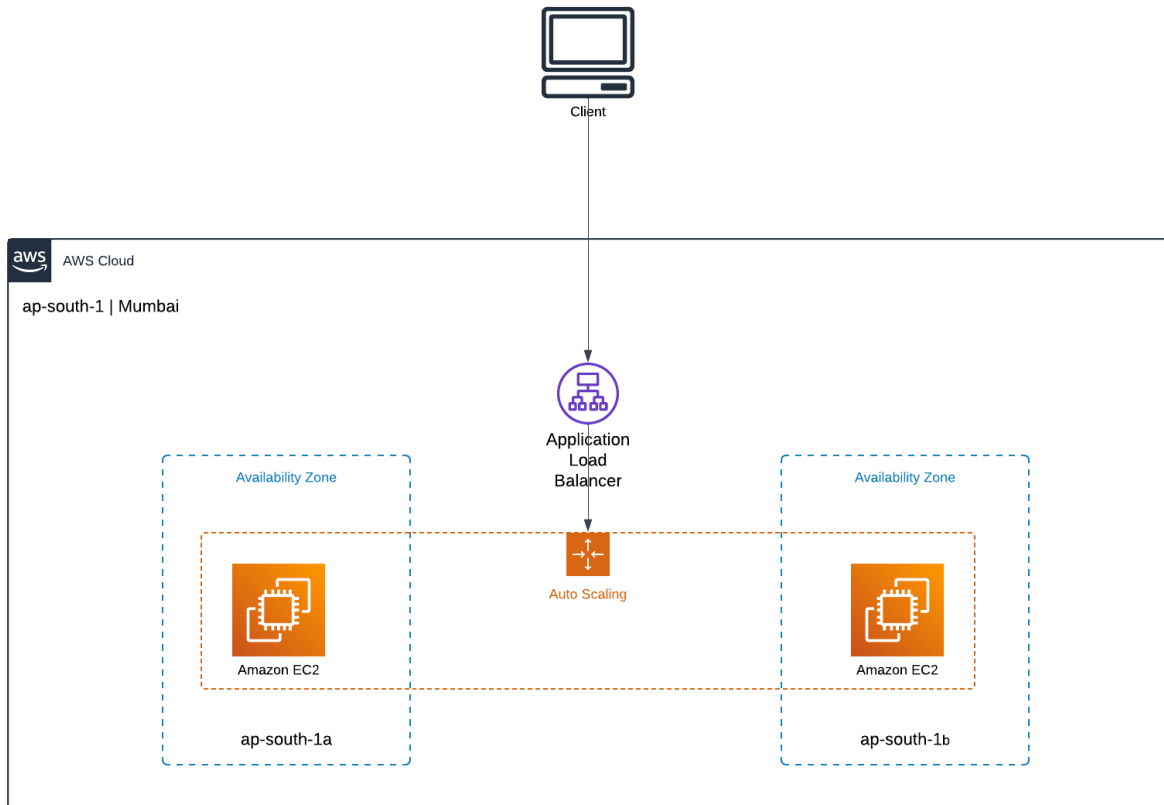


Project: Load Balancing with Auto Scaling in Multi-AZ

Key Takeaways:

- Deployed Application load balancer with target groups pointing to Auto scaling group in Multi-AZ
- configured security groups of ALB & ASG -EC2 to allow inbound from internet and ALB's Security Groups respectively.
- enabled auto scaling target tracking policy to auto scale out and in with appropriate min, max & desired no. of instances.

Diagram:



Workflow:

Security Groups

Security Groups (3)

Info

Actions

Export security groups to CSV

Create security group

Find resources by attribute or tag

<

1

>

<input type="checkbox"/>	Name	Security group ID	Security group name	VPC ID	Description
<input type="checkbox"/>	-	sg-04ef47015a39e5972	MyALBSecurityGroup	vpc-06f7381bda09fa98b	Allowing HTTP from Internet
<input type="checkbox"/>	-	sg-02b4b572fbaa45f06	MyASG_EC2SecurtiyGroup	vpc-06f7381bda09fa98b	Allowing HTTP/S from security group .

EC2 > Security Groups > sg-02b4b572fbaa45f06 - MyASG_EC2SecurtiyGroup

sg-02b4b572fbaa45f06 - MyASG_EC2SecurtiyGroup

Actions

Details

Security group name

MyASG_EC2SecurtiyGroup

Owner

445567072214

Security group ID

sg-02b4b572fbaa45f06

Inbound rules count

2 Permission entries

Description

Allowing HTTP/S from security group of ALB

Outbound rules count

1 Permission entry

VPC ID

[vpc-06f7381bda09fa98b](#)

Inbound rules

Outbound rules

Sharing - new

VPC associations - new

Tags

Inbound rules (2)

Manage tags

Edit inbound rules

Search

<

1

>

<input type="checkbox"/>	Name	Security group rule...	IP version	Type	Protocol	Port range
<input type="checkbox"/>	-	sgr-008d69e727174c1...	-	HTTP	TCP	80
<input type="checkbox"/>	-	sgr-086217bf614097e...	-	HTTPS	TCP	443

EC2 > Security Groups > sg-04ef47015a39e5972 - MyALBSecurityGroup

sg-04ef47015a39e5972 - MyALBSecurityGroup

Actions

Details

Security group name

MyALBSecurityGroup

Owner

445567072214

Security group ID

sg-04ef47015a39e5972

Inbound rules count

2 Permission entries

Description

Allowing HTTP from Internet

Outbound rules count

1 Permission entry

VPC ID

[vpc-06f7381bda09fa98b](#)

Inbound rules

Outbound rules

Sharing - new

VPC associations - new

Tags

Inbound rules (2)

Manage tags

Edit inbound rules

Search

<

1

>

<input type="checkbox"/>	Name	Security group rule...	IP version	Type	Protocol	Port range
<input type="checkbox"/>	-	sgr-0136d948792fbf0e8	IPv4	HTTPS	TCP	443
<input type="checkbox"/>	-	sgr-0108dca8762bf599b	IPv4	HTTP	TCP	80

Application Load Balancer:

[EC2](#) > [Load balancers](#) > MyALB

MyALB

Actions

▼ Details

Load balancer type Application	Status Active	VPC vpc-06f7381bda09fa98b	Load balancer IP address type IPv4
Scheme Internet-facing	Hosted zone ZP97RAFLXTNZK	Availability Zones subnet-028e3fc02f6011af7 ap-south-1a (aps1-az1) subnet-0ad66cecb03e84487 ap-south-1b (aps1-az3)	Date created November 14, 2024, 12:52 (UTC+05:30)
Load balancer ARN arn:aws:elasticloadbalancing:ap-south-1:445567072214:loadbalancer/app/MyALB/1957ef8ac31f7901		DNS name info MyALB-1264122458.ap-south-1.elb.amazonaws.com (A Record)	

Listeners and rules | Network mapping | Resource map - new | **Security** | Monitoring | Integrations | Attributes | Capacity - new | Tags

Security groups (1)

Edit

A security group is a set of firewall rules that control the traffic to your load balancer.

Security Group ID	Name	Description
sg-04ef47015a39e5972	MyALBSecurity...	Allowing HTTP from Internet

Launch Template:

[EC2](#) > ... > MyLaunchTemplate-v1

MyLaunchTemplate-v1 (lt-0d118dc13cc41c7b8)

Actions

Delete template

Launch template details

Launch template ID lt-0d118dc13cc41c7b8	Launch template name MyLaunchTemplate-v1	Default version 1	Owner arn:aws:iam::445567072214:user/sahal
--	---	--------------------------------------	---

Details | Versions | Template tags

Launch template version details

Actions

Delete template version

Version 1 (Default)	Description A Demo template	Date created 2024-11-14T07:19:24.000Z	Created by arn:aws:iam::445567072214:user/sahal
--	--	--	--

Instance details | Storage | Resource tags | Network interfaces | **Advanced details**

AMI ID ami-08bf489a05e916bbd	Instance type t2.micro	Availability Zone -	Key pair name -
Security groups -	Security group IDs sg-02b4b572fbaa45f06		

EC2 User data Script:

```
#!/bin/bash
# Use this for your user data (script from top to bottom)
# install httpd (Linux 2 version)
yum update -y
yum install -y httpd
systemctl start httpd
systemctl enable httpd
echo "<h1>Sahal's : Instance hostname: $(hostname -f)</h1>" > /var/www/html/index.html
```

Auto Scaling Groups:

1. naming ASG & choosing launch template (already created)

Auto Scaling groups > Create Auto Scaling group

Choose launch template [Info](#)

Specify a launch template that contains settings common to all EC2 instances that are launched by this Auto Scaling group.

launch template

instance launch options

optional
the advanced options

optional
the group size and

optional
configurations

optional

Name

Auto Scaling group name
Enter a name to identify the group.

my-asg

Must be unique to this account in the current Region and no more than 255 characters.

Launch template [Info](#)

Launch template
Choose a launch template that contains the instance-level settings, such as the Amazon Machine Image (AMI), instance type, key pair, and security groups.

my-launch-template

[Create a launch template](#)

Version

Default (1)

2. configuring network settings. Here selecting only two AZ of Mumbai Region

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-06f7381bda09fa98b (default-vpc)
172.31.0.0/16 Default

[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-028e3fc02f6011af7
172.31.32.0/20 Default

ap-south-1b | subnet-0ad66cecb03e84487
172.31.0.0/20 Default

[Create a subnet](#)

Availability Zone distribution - new
Auto Scaling automatically balances instances across Availability Zones. If launch failures occur in a zone, select a strategy.

☒ **Balanced best effort**
If launches fail in one Availability Zone, Auto Scaling will attempt to launch in another healthy Availability Zone.

☐ **Balanced only**
If launches fail in one Availability Zone, Auto Scaling will continue to attempt to launch in the unhealthy Availability Zone to preserve balanced distribution.

Cancel Skip to review Previous **Next**

3. creating & attaching ALB | although we can manually create load balancer and create target groups with ASG

Create Auto Scaling group

Configure advanced options - *optional* [Info](#)

Use a load balancer to distribute network traffic across multiple servers. Enable service-to-service communications with VPC Lattice. Shift resources away from impaired Availability Zones with zonal shift. You can also customize health check replacements and monitoring.

Load balancing [Info](#)

Use the options below to attach your Auto Scaling group to an existing load balancer, or to a new load balancer that you define.

☐ No load balancer
Traffic to your Auto Scaling group will not be fronted by a load balancer.

☐ Attach to an existing load balancer
Choose from your existing load balancers.

☒ Attach to a new load balancer
Quickly create a basic load balancer to attach to your Auto Scaling group.

Attach to a new load balancer

Define a new load balancer to create for attachment to this Auto Scaling group.

Load balancer type
Choose from the load balancer types offered below. Type selection cannot be changed after the load balancer is created. If you need a different type of load balancer than those offered here, [visit the Load Balancing console](#). [?](#)

☒ Application Load Balancer
HTTP, HTTPS

☐ Network Load Balancer
TCP, UDP, TLS

Load balancer name
Name cannot be changed after the load balancer is created.

my-alb

4. enabling ALB health checks (to auto terminate unhealthy instances)

Health checks

Health checks increase availability by replacing unhealthy instances. When you use multiple health checks, all are evaluated, and if at least one fails, instance replacement occurs.

EC2 health checks

[?](#) Always enabled

Additional health check types - *optional* [Info](#)

☒ Turn on Elastic Load Balancing health checks **Recommended**
Elastic Load Balancing monitors whether instances are available to handle requests. When it reports an unhealthy instance, EC2 Auto Scaling can replace it on its next periodic check.

☐ Turn on VPC Lattice health checks
VPC Lattice can monitor whether instances are available to handle requests. If it considers a target as failed a health check, EC2 Auto Scaling replaces it after its next periodic check.

☐ Turn on Amazon EBS health checks
EBS monitors whether an instance's root volume or attached volume stalls. When it reports an unhealthy volume, EC2 Auto Scaling can replace the instance on its next periodic health check.

Health check grace period [Info](#)
This time period delays the first health check until your instances finish initializing. It doesn't prevent an instance from terminating when placed into a non-running state.

300 seconds

5. ALB listeners & routing & target group details

Network mapping

Your new load balancer will be created using the same VPC and Availability Zone selections as your Auto Scaling group. You can select different subnets and add subnets from additional Availability Zones.

VPC
vpc-06f7381bda09fa98b [🔗](#) default-vpc

Availability Zones and subnets

You must select a single subnet for each Availability Zone enabled. Only public subnets are available for selection to support DNS resolution.

☒ ap-south-1a subnet-028e3fc02f6011af7 ▼

☒ ap-south-1b subnet-0ad66cecb03e84487 ▼

☐ ap-south-1c Select a subnet ▼

Listeners and routing

If you require secure listeners, or multiple listeners, you can configure them from the [Load Balancing console](#) [🔗](#) after your load balancer is created.

Protocol HTTP

Port 80

Default routing (forward to) Create a target group ▼

New target group name
An instance target group with default settings will be created.

my-alb-tg

Tags - optional
Consider adding tags to your load balancer. Tags enable you to categorize your AWS resources so you can more easily manage them.

Add tag

6. Group capacity & scaling policy

Configure group size and scaling - optional [Info](#)

Define your group's desired capacity and scaling limits. You can optionally add automatic scaling to adjust the size of your group.

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.

1

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 1
Equal or less than desired

Max desired capacity 2
Equal or greater than desired

Automatic scaling - optional

Choose whether to use a target tracking policy [Info](#)
You can set up other metric-based scaling policies and scheduled scaling after creating your Auto Scaling group.

☐ No scaling policies
Your Auto Scaling group will remain at its initial size and will not dynamically resize to meet demand.

☒ Target tracking scaling policy
Choose a CloudWatch metric and target value and let the scaling policy adjust the desired capacity in proportion to the metric's value.

Scaling policy name
Target Tracking Policy

Metric type [Info](#)
Monitored metric that determines if resource utilization is too low or high. If using EC2 metrics, consider enabling detailed monitoring for better scaling performance.

Average CPU utilization ▼

Target value
30

Instance warmup [Info](#)
300 seconds

☐ Disable scale in to create only a scale-out policy

7. Reviewing

Step 1: Choose launch template

Edit

Group details

Auto Scaling group name

MyASG

Launch template

Launch template

MyLaunchTemplate-v1

lt-0d118dc13cc41c7b8

Version

Default

Description

A Demo template

Step 2: Choose instance launch options

Edit

Network

VPC

vpc-06f7381bda09fa98b

Availability Zones and subnets

Availability Zone	Subnet	Subnet CIDR range
ap-south-1a	subnet-028e3fc02f6011af7	172.31.32.0/20
ap-south-1b	subnet-0ad66cecb03e84487	172.31.0.0/20

Availability Zone distribution

Balanced best effort

Step 3: Configure advanced options

Edit

Load balancing

Load balancer 1

Name	Type	Target group
MyALB	Application/HTTP	my-alb-tg

Step 4: Configure group size and scaling policies

Edit

Group size

Desired capacity	Desired capacity type
1	Units (number of instances)

Scaling

Minimum desired capacity	Maximum desired capacity	
1	2	
Target tracking policy	Scaling policy name	Execute policy when
Policy type	Target Tracking Policy	As required to maintain Average CPU utilization at 30
Target tracking scaling		
Take the action	Instances need	Scale in
Add or remove capacity units as required	300 seconds to warm up before including in metric	Enabled

Output:

