**Creating an Auto Scaling Group and Application Load Balancer in AWS**

**About this Hands-on Lab**

In this hands-on lab scenario, you’re a cloud network engineer working for an organization that sells products online. You’re gearing up for the annual sale that provides a 50% discount on all items. This sale drives a ton of traffic and revenue. Your job is to ensure the website doesn’t go down and is able to handle every request efficiently. In this lab, you will integrate two powerful AWS services: Elastic Load Balancers and Auto Scaling groups. Specifically, you will create an Auto Scaling group of EC2 instances operating as web servers and then configure an Application Load Balancer to load balance between the instances inside that Auto Scaling group. After everything is set up, you’ll simulate stress tests on the EC2 instances to confirm the Auto Scaling group works as expected.

**Learning Objectives**

Successfully complete this lab by achieving the following learning objectives:

**Create an Application Load Balancer**

1. Navigate to EC2 > Load Balancers.
2. Click Create Load Balancer.
3. Click the Create button under the Application Load Balancer and set the following values after choosing your region:

Name: HOLALB

Scheme: internet-facing

IP address type: ipv4

Select the VPC.

Add two Availability Zones to your ALB.

Load Balancer Protocol: HTTP

Port: 80

1. Configure Security Groups.

Select to Create a new security group for your ALB, and set the following values:

Name: ALBSG

Description: ALBSG

**Inbound Rules:**

Add two rules to allow standard HTTP traffic from 0.0.0.0/0 and ::/0 (IPV6).

**Outbound Rules:**

The default value allows standard HTTP traffic from 0.0.0.0/0 , so leave as it.

Add a rule to allow all traffic from ::/0 (IPV6.

On the “Create Application Load Balancer” screen, refresh the “**Security Groups** “ drop down and select the security group you just created.

Ensure only one Security Group is selected for your ALB.

1. Configure **Listeners and routing** and enter the following values:

Protocol: HTTP

Port: 80

Select “Create target group” and enter the following values:.

***Step 1  
Specify group details***

Name: ALBTG

Target type: Instances

Protocol: HTTP

Port: 80

Leave the VPC as default , Protocol version as HTTP1 and Health check path as default path.

1. Expand Advanced health check settings, and reduce the healthy and unhealthy threshold checks down to 2.

This means the load balancer can respond faster and instances come into service and vice versa.

Add a Tag :

Key: Name

Value:ALBTG

***Step 2  
Register targets***

1. Click Next: Register Targets.
2. Click Create Target Group.
3. Back in the “Listeners and routing” section of the “Create Application Load Balancer” screen, refresh the “Default action“ drop down and select the target group you just created.

Add a Tag :

Key: Name

Value: HOLALB

1. Click Create Target Group.

**Make a note of the DNS name associated with the load balancer and open in a new browser tab. You should see a 503 error since we don’t have any operational EC2 instances associated with the load balancer.**

**Create a Launch Template**

Create an SSH key pair that the EC2 instances will use to control access.

We will use the key pair to connect to the EC2 instances in order to perform load testing:

1. Navigate to EC2 > Network & Security > Key Pairs.
2. Click Create Key Pair.
3. Key pair type -> RSA
4. Private key file format ->.pem
5. Add a Tag :

Key: Name

Value: ALBKP

1. Name the key pair ALBKP, and download the file to your local machine in a publicly accessible directory like C:\Users\<username> (for Windows machines.Do the equivalent for Mac). Check that the extension of your file is .pem after download.

Create a security group for EC2 instances. This security group will allow us to SSH into the instances and it will allow the ALB to point front end connections to the instances over port 80.

1. Navigate to EC2 > Network & Security > Security Groups.
2. Click Create Security Group.
3. The name and description are EC2WEBSG.
4. Set the VPC to the default VPC.
5. Add two **Inbound** rules allowing SSH from 0.0.0.0/0 and ::/0 (IPV6).
6. Add a rule allowing HTTP from the Security Group ID of the security group for the ALB created in the previously.
7. Add a Tag :

Key: Name

Value: EC2WEBSG

8.Create the security group.

Create a launch template that will be used by the Auto Scaling group. The launch template defines what the instances are and how they are created.

1. Navigate to EC2 > Instances > Launch Templates.
2. Create a new template, and call it HOLLT for the name and description. Check “Auto Scaling guidance”.
3. Add a Template Tag :

Key: Name

Value: HOLLT

1. For "AMI", and pick the Amazon Linux 2 AMI (64-bit x86).
2. Set the instance type as t2.micro.
3. Select the key pair you created earlier.
4. The network type is VPC.
5. Select the EC2WEBSG security group you created earlier.
6. Storage should automatically be populated with a volume, so leave that as default and don’t add anything to the network section.
7. Expand Advanced Details, and paste the user data (provided in the ***Assignment*** folder) in the box.

Note: These are commands to install a web server and download website content.

1. Click Create Launch Template.
2. Click Close.

**Create an Auto Scaling Group**

**Note:** Make sure the load balancer is ready at this point.

1. EC2 > Auto Scaling > Auto Scaling Groups
2. Click Create Auto Scaling group.
3. Call the group HOLASG.
4. Select Launch Template, and choose the template you just created.
5. Click Next.
6. Select Adhere to Launch Template.
7. Pick the default VPC , and select the subnets belonging to the same AZ as the Load Balancer.
8. Click Next.
9. Check Attach to an existing load balancer.
10. Select target group ALBTG.
11. Leave the default for Health checks.
12. Select Enable group metrics collection with CloudWatch.
13. Click Next.
14. For Group Size, enter the following values:

* Desired Capacity: 2
* Minimum Capacity: 2
* Maximum Capacity: 6

1. For Scaling Policies, select Target tracking scaling policy and enter the following values:

* Scaling Policy Name: Target Tracking Policy
* Metric type: Average CPU utilization
* Target value: 30
* Instances need: 300

1. Click Next.
2. No need to Add Notifications. Click Next.
3. Add a Tag :

Key: Name

Value: HOLASG

1. Click Next and Review
2. Click Create Auto Scaling Group.

**Attempt to Connect to Website**

Copy the DNS name associated with the load balancer and open in a new browser tab.

Try to access the simple website.

**Test Horizontal Scaling**

1. Connect to one of the EC2 instances via SSH.
2. Copy the connection string and in your command prompt navigate to the directory where you stored the key pair file.
3. If you will use an SSH client on a macOS or Linux computer to connect to your Linux instance, use the following command to set the permissions of your private key file so that only you can read it.Run chmod on the .pem key pair file before running the connection string.

chmod 400 my-key-pair.pem

400 protects it by making it read only and only for the owner.

If you do not set these permissions, then you cannot connect to your instance using this key pair.

1. Install the stress test application (using the commands provided in the Assignment folder).
2. Run the stress test on the EC2 instance (using the command provided in the Assignment folder).
3. After a few minutes, watch the number of instances increase. It enacts the scale-out policy.
4. After a few minutes, stop the stress test or wait till it times out after 300 seconds. It enacts the scale-in policy.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*