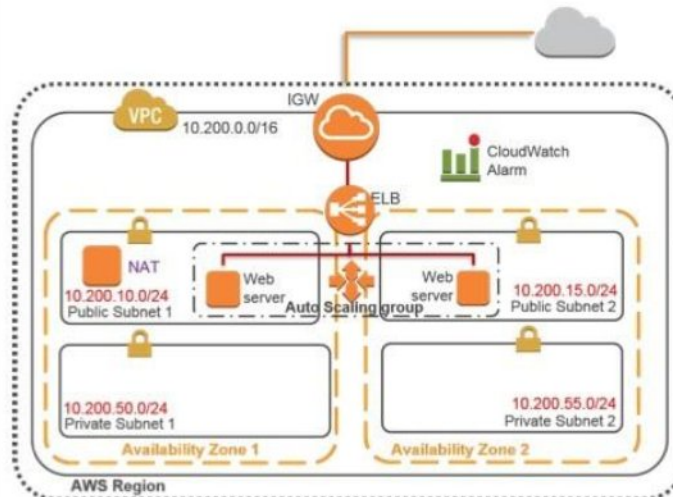


Lab 4: Working with Auto Scaling

Overview

In this lab, you will implement elasticity using Auto Scaling. Auto Scaling is not just adding and subtracting servers, it is also a mechanism to handle failures similar to the way that Load Balancing handles unresponsive servers. This lab demonstrates how to configure Auto Scaling to automatically launch and monitor Amazon EC2 instances and how to update an associated load balancer.



Auto Scaling allows you to automatically scale your Amazon EC2 capacity according to the conditions you define. With Auto Scaling, you can ensure that the number of Amazon EC2 instances you are using increases seamlessly during demand spikes to maintain performance, and decreases automatically during demand lulls to minimize costs. Auto Scaling is particularly well suited for applications that experience hourly, daily, or weekly variability in usage.

Objectives

After completing this lab, you will be able to:

- Create a launch configuration.

	<ul style="list-style-type: none">• Create a load balancer.• Create an Auto Scaling group.• Define Auto Scaling policies.
Pre-requisites	<p>This lab requires:</p> <ul style="list-style-type: none">• Access to a notebook computer with Wi-Fi running Microsoft Windows, Mac OS X, or Linux (Ubuntu, SuSE, or Red Hat).<ul style="list-style-type: none">◦ Note The qwikLABS lab environment is not accessible using an iPad or tablet device, but you can use these devices to access the student guide.• For Microsoft Windows users: Administrator access to the computer.• An Internet browser such as Chrome, Firefox, or Internet Explorer 9 (previous versions of Internet Explorer are not supported).• An SSH client, such as PuTTY.
Duration	This lab will require around 30 minutes to complete.

Task 1: Check the existing VPC and subnets

Overview

In this part of the lab, you will create a security group that will be used by your Auto Scaling group.

A security group acts as a virtual firewall that controls the traffic for one or more instances. When you launch an instance, you associate one or more security groups with the instance. To each security group you add rules that allow traffic to or from its associated instances. You can modify these rules for a security group at any time; the new rules are automatically applied to all instances that are associated with the security group. When you decide whether to allow traffic to reach an instance, all the rules from all the security groups that are associated with that instance are evaluated automatically.

Task 1-1: Check the existing VPC and subnets**Overview**

In a previous lab, you created a VPC and 2-4 subnets. This lab builds on top of what you have done so far. In this section of the lab, you'll check the properties of your environment's VPC.

Step	Instruction
1.1.1	In the AWS Management Console , click VPC .
1.1.2	In the VPC Dashboard pane, click Your VPCs .
1.1.3	Select the LabVPC check box, and make a note of the VPC ID (starting with vpc-) and the VPC CIDR range (e.g., 10.200.0.0/16).
1.1.4	In the navigation pane, click Subnets .
1.1.5	To view only the subnets that belong to the LabVPC , enter the VPC ID in the search bar.
1.1.6	You should be able to see that the subnets that have been created for you are identical to the subnets you had in the last lab.

Task 1-2: Create security groups

Overview	In this section of the lab, you will create security groups for your environment.
-----------------	---

Step	Instruction
1.2.1	On the Services menu, click EC2 .
1.2.2	In the EC2 Dashboard pane, click Security Groups .
1.2.3	Click Create Security Group , and then specify these settings: <ul style="list-style-type: none"> • Security group name: ELBSG • Description: My Lab ELB Security Group • VPC: LabVPC
1.2.4	Under Security group rules , on the Inbound tab, click Add Rule . Specify the following settings: <ul style="list-style-type: none"> • Type: HTTP • Source: Anywhere <p>Note If your browser does not provide a drop-down list for Source, select Custom IP and type 0.0.0.0/0</p>
1.2.5	Click Create .
1.2.6	Note the Group ID that starts with sg- for this new security group. Store this Group ID to a text file, so that you can retrieve it for use in later steps.
1.2.7	Click Create Security Group to create another security group. Specify the following settings: <ul style="list-style-type: none"> • Security group name: AppSG • Description: My App Security Group • VPC: LabVPC
1.2.8	Under Security group rules , on the Inbound tab, click Add Rule . <ul style="list-style-type: none"> • In the Type drop-down list, click HTTP. • In the Source drop-down list, select Custom IP, and enter the Group ID of your ELB SG from step 1.2.6.
1.2.9	Click Create .

Task 2: Creating a Load Balancer

Overview

Before you create an Auto Scaling group, you first need a load balancer. This load balancer will send requests to your Amazon EC2 instances, dynamically distributing them across Amazon EC2 instances as the Auto Scaling group increases and decreases in size.

Task 2-1: Create a load balancer

Overview	In this section of the lab you will create an Elastic Load Balancing load balancer.
-----------------	---

Step	Instruction
2.1.1	In the EC2 Dashboard pane, click Load Balancers .
2.1.2	Click Create Load Balancer .
2.1.3	Specify these settings: <ul style="list-style-type: none"> • Load Balancer name: MyLB • Create LB Inside: LabVPC Leave the remaining settings with their default values.
2.1.4	In Select Subnets , under Available Subnets , select PublicSubnet1 and PublicSubnet2 by clicking the plus signs on the left side of their rows.
2.1.5	Click Next: Assign Security Groups .
2.1.6	On the Assign Security Groups page, select ELBSG from the existing security group list and clear the default check box. The only check box that should be selected is the one for ELBSG .
2.1.7	Click Next: Configure Security Settings .
2.1.8	Click Next: Configure Health Check .
2.1.9	On the Configure Health Check page, specify these settings: <ul style="list-style-type: none"> • Ping Path: /index.php • Response Timeout: 10 seconds • Health Check Interval: 15 seconds • Unhealthy Threshold: 5 • Healthy Threshold: 2 Leave the remaining settings with their default values.
2.1.10	Click Next: Add EC2 Instances .

2.1.11	Do not select any of the instances. Click Next: Add Tags .
2.1.12	Do not specify any tags. Click Review and Create .
2.1.13	Click Create .
2.1.14	Click Close .

Task 3: Creating a Launch Configuration

Overview

Your first step in creating an Auto Scaling group is to generate a launch configuration. A launch configuration specifies details such as the AMI to be used when launching new instances, the instance type, and the configuration scripts.

Task 3-1: Create a launch configuration

Overview	In this section of the lab, you will create a launch configuration that will be used to create your Auto Scaling group.
-----------------	---

Step	Instruction
3.1.1	In the navigation pane, click Launch Configurations .
3.1.2	Click Create Auto Scaling group .
3.1.3	Click Create launch configuration .
3.1.4	In the row for Amazon Linux AMI... (HVM) , click Select .
3.1.5	Select the t2.small instance type.
3.1.6	Click Next: Configure details .
3.1.7	In the Name box, type LabLC
3.1.8	For Monitoring , select the Enable CloudWatch detailed monitoring option.
3.1.9	Expand the Advanced Details section. Copy the following script from the Command Reference File and paste it into the User data box: <pre>#!/bin/bash yum -y update yum -y install httpd php chkconfig httpd on wget https://d2lrzjb0vjvp5.cloudfront.net/architecting/v4.6/lab-4-working-with-autoscaling/static/phpapp.zip unzip phpapp.zip -d /var/www/html/ service httpd start yum -y install stress sleep 90 stress --cpu 8 --io 8 --vm 6 --hdd 8 -t 400</pre>
3.1.10	Leave the remaining settings in their default values and click Next: Add Storage .
3.1.11	Click Next: Configure Security Group .
3.1.12	For Assign a security group , click the Select an existing security group option and then select AppSG .

3.1.13	<p>Click Review.</p> <p>In the warning message, click Continue.</p> <p>Note In this lab, you will not be logging in to the instance via SSH; therefore, that port is not opened in the AppSG security group.</p>
3.1.14	Review the settings and then click Create launch configuration .
3.1.15	When prompted, accept the key pair generated by qwikLABS.
3.1.16	Select the acknowledgement check box, and then click Create launch configuration .

Task 4: Create an Auto Scaling group

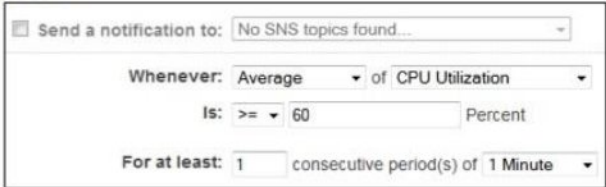
Overview

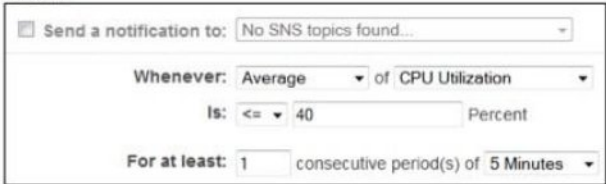
The launch configuration controls *which* instances are launched and how they are configured, and the Auto Scaling group controls *when* instances are launched or terminated and what criteria trigger an auto scaling action.

Task 4-1: Create an Auto Scaling group

Overview

After creating your **Launch Configuration** in the last procedure, you should be automatically redirected to the **Create an Auto Scaling Group** page, where you will now create your Auto Scaling group.

Step	Instruction
4.1.1	<p>On the Create Auto Scaling Group page, specify the following settings:</p> <ul style="list-style-type: none"> • Group name: LabASG • Group size: Start with 2 instances • Network: LabVPC • Click in the Subnet box, and then select PrivateSubnet1 and PrivateSubnet2.
4.1.2	<p>Expand the Advanced Details section:</p> <p>For Load Balancing, select the Receive traffic from Elastic Load Balancer(s) check box, and then select MyLB in the box (the load balancer that you created earlier).</p>
4.1.3	Click Next: Configure scaling policies .
4.1.4	Select the Use scaling policies to adjust the capacity of this group option, and change the policy to "Scale between 2 and 6 instances."
4.1.5	Under Increase Group Size , click the Add new alarm link for the Execute policy when section.
4.1.6	<p>In the Create Alarm dialog box:</p> <ul style="list-style-type: none"> • Clear the Send a notification to check box. • In the Percent box, type 60 • Change consecutive period(s) of 5 Minutes to 1 Minute. <p>Result</p> 

4.1.7	Click Create Alarm .
4.1.8	<p>In the Increase Group Size section, do the following:</p> <ul style="list-style-type: none"> Click Create a simple scaling policy. Change the Take the action value to "Add 2 instances." Change the And then wait value to "90 seconds before allowing another scaling activity."
4.1.9	Under Decrease Group Size , click the Add new alarm link for the Execute policy when section.
4.1.10	<p>In the Create Alarm dialog box:</p> <ul style="list-style-type: none"> Clear the Send a notification to check box. Change ">=" to "<=". In the Percent box, type 40 <p>Result</p> 
4.1.11	<p>Click Create Alarm.</p> <p>Note Ensure that the two alarms do not have identical names. In rare circumstances, the names created by the tool will be identical and cause conflicts. If they are identical, click "Edit" next to one of the policies and give it a unique name instead.</p>
4.1.12	<p>Still in the Decrease Group Size section, do the following:</p> <ul style="list-style-type: none"> Click Create a simple scaling policy. Change the Take the action value to "Remove 2 instances." Change the And then wait value to "90 seconds before allowing another scaling activity."
4.1.13	Click Review .
4.1.14	Click Create Auto Scaling group .
4.1.15	Click Close .

4.1.16	<p>With the newly created LabASG auto scaling group selected, click the Activity History tab. The Auto Scaling group should launch two instances because the group size was set to 2.</p> <p>From the EC2 Dashboard, click Instances. Wait until the instances are fully started and the Status Checks shows <i>2/2 checks passed</i>.</p> <p>In Step 4.4.9, you defined the launch configuration with user data that emulates resource consumption:</p> <pre>... yum -y install stress sleep 90 stress --cpu 4 --io 4 --vm 2 --hdd 4 -t 400</pre> <p>Refresh your Activity History tab regularly, and within about 5-7 minutes your Auto Scaling Group should add two more instances in response to the high stress being placed on the instances' CPUs. Within another 3 minutes, two more instances should be started. Finally, 3 minutes after that, your Auto Scaling Group should detect that the stress functions have completely timed out and it will subsequently terminate your two oldest instances.</p>
--------	---

Ending your lab

Overview

When you are finished with your lab, terminate the lab environment using the following steps.

Step	Instruction
1	To log out of the AWS Management Console, from the menu, click awsstudent @ [YourAccountNumber] and choose Sign out (where [YourAccountNumber] is the AWS account generated by qwikLABS).
2	Close any active SSH client sessions or remote desktop sessions.
3	Click the End Lab button on the qwikLABS lab details page in your browser.
4	When prompted for confirmation, click OK .
5	<p>For My Rating, rate the lab (using the applicable number of stars), optionally type a Comment, and click Submit.</p> <p>Note: The number of stars indicates the following: 1 star = very dissatisfied, 2 stars = dissatisfied, 3 stars = neutral, 4 stars = satisfied, and 5 stars = very satisfied. Also, you may close the dialog if you do not wish to provide feedback.</p>