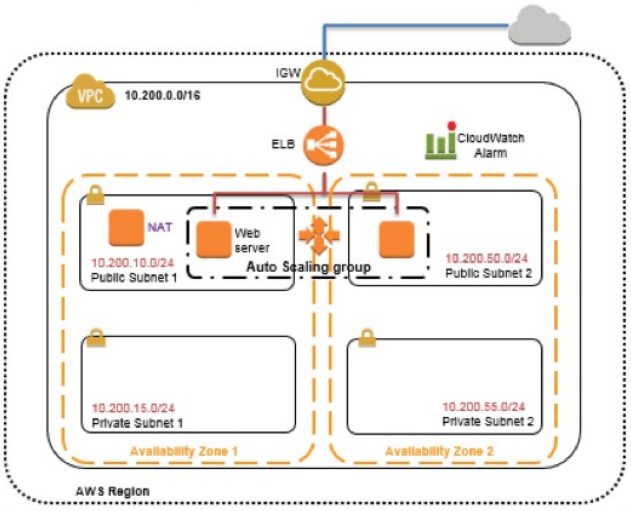
PRINTED BY: gleung@datapipe.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Lab 4: Working with Auto Scaling

Lab Overview

Introduction

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.

Continued on next page

© 2015 Amazon Web Services, Inc. and its affiliates. All rights reserved.

PRINTED BY: gleung@datapipe.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Lab Overview, Continued

Allotted time

The allotted time for this lab is listed as follows:

Component	Time
Overview	1 minute
Lab	30 minutes
Total:	**31 minutes**

Prerequisites

This lab requires:

- Access to a notebook computer with Wi-Fi on a Microsoft Windows, Mac OS X, or Linux (Ubuntu, SuSE, or Red Hat) system
- The qwikLABS lab environment is not accessible on an iPad or tablet device, but you may use these devices to access the student guide (PDF)
- For Microsoft Windows users: Administrator access to the computer
- An Internet browser such as Chrome, Firefox, or IE9 (previous versions of Internet Explorer are not supported)
- An SSH client such as PuTTY

Objectives

After completing this lab, you will be able to:

- Create a launch configuration.
- Create a load balancer.
- Create an Auto Scaling group.
- Define Auto Scaling policies.

© 2015 Amazon Web Services, Inc. and its affiliates. All rights reserved.

PRINTED BY: gleung@datapipe.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Creating a Security Group

Introduction

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 4.1: Check the existing VPC and subnets

In a previous lab, you created a VPC and two subnets. This lab builds on top of what you have done so far. To check the properties of those existing subnets:

Step	Action
4.1.1	In the AWS Management Console, on the **Services** menu, click **VPC**.
4.1.2	In the navigation pane, click **Your VPCs**.
4.1.3	Select **Lab VPC**, and make a note of the **VPC ID** (starting with *vpc-*).
4.1.4	In the navigation pane, click **Subnets**.
4.1.5	To view only the subnets that belong to the **Lab VPC**, enter the **VPC ID** from step 4.1.3 in the search bar.
4.1.6	Make a note of those subnets' CIDR ranges.

Continued on next page

© 2015 Amazon Web Services, Inc. and its affiliates. All rights reserved.

PRINTED BY: gleung@datapipe.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Creating a Security Group, Continued

Task 4.2: Create security groups

To create a security group:

Step	Action
4.2.1	On the **Services** menu, click **EC2**.
4.2.2	In the navigation pane, click **Security Groups**.
4.2.3	Click **Create Security Group**, and then specify these settings: • In the **Security group name** box, type **ELB SG** • In the **Description** box, type **My Lab ELB Security Group** • For **VPC**, select **Lab VPC (10.200.0.0/16)**.
4.2.4	Under **Security group rules**, click **Add Rule**. • In the **Type** drop-down list, click **HTTP**. • In the **Source** drop-down list, click **Anywhere**. **Note:** If your browser does not provide a drop-down menu for **Source,** type **0.0.0.0/0**
4.2.5	Click **Create**.
4.2.6	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.
4.2.7	Click **Create Security Group** to create another security group. • In the **Security group name** box, type **App SG** • In the **Description** box, type **My App Security Group** • For **VPC**, click **Lab VPC (10.200.0.0/16)**.
4.2.8	Under **Security group rules**, click **Add Rule**. • On the **Type** drop-down menu, click **HTTP**. • For **Source**, select **Custom IP**, and enter the **Group ID** of your **ELB SG** from step 4.2.6.
4.2.9	Click **Create**.

© 2015 Amazon Web Services, Inc. and its affiliates. All rights reserved.

PRINTED BY: gleung@datapipe.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Creating a Load Balancer

Introduction

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 4.3: Create a load balancer

To create a load balancer:

Step	Action
4.3.1	In the navigation pane, click **Load Balancers**.
4.3.2	Click **Create Load Balancer**.
4.3.3	Specify these settings: • In the **Load Balancer name** box, type **My-LB** • For **Create LB Inside**, select **Lab VPC**. Leave the remaining settings with their default values.
4.3.4	Click **Continue**.
4.3.5	On the **Configure Health Check** page, specify these settings: • For **Ping Path**, delete the default value and type **/index.php** • Change the **Response Timeout** value to **10** seconds. • Change the **Health Check Interval** value to **15** seconds. • For **Unhealthy Threshold**, select **5**. • For **Healthy Threshold**, select **2**. Leave the remaining settings in their default values.
4.3.6	Click **Continue**.
4.3.7	On the **Select Subnets** page, in the **Available Subnets**, select **Private Subnet 1** and **Private Subnet 2**.
4.3.8	Click **Continue**.
4.3.9	On the **Assign Security Groups** page, select **ELB SG** from the existing security group list and clear the **default** check box.
4.3.10	Click **Continue**.
4.3.11	On the **Add Instances to Load Balancer** page, click **Continue**. **Note:** Do not select any of the instances.
4.3.12	Click **Continue** without specifying any tag.
4.3.13	Click **Create**.
4.3.14	Click **Close**.

© 2015 Amazon Web Services, Inc. and its affiliates. All rights reserved.

PRINTED BY: gleung@datapipe.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Creating a Launch Configuration

Introduction

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 4.4: Create a launch configuration

To create a launch configuration:

Step	Action
4.4.1	In the navigation pane, click **Launch Configurations**.
4.4.2	Click **Create Auto Scaling group**.
4.4.3	Click **Create launch configuration**.
4.4.4	Click **Select** for **Amazon Linux AMI... (HVM)**.
4.4.5	Select **t2.small**.
4.4.6	Click **Next: Configure details**.
4.4.7	In the **Name** box, type **Lab LC**
4.4.8	Select **Enable CloudWatch detailed monitoring.**
4.4.9	Expand the **Advanced Details** section and then add the following to the **User data** box: **#!/bin/bash** **yum -y update** **yum -y install httpd-2.2.27 php-5.3.29** **chkconfig httpd on** **wget https://us-west-2-aws-training.s3.amazonaws.com/awsu-ilt/architecting/v4.2/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 4 --io 4 --vm 2 --hdd 4 -t 400** **sleep 300** **stress --cpu 6 --io 6 --vm 4 -hdd 6 -t 400** **sleep 300** **stress --cpu 8 --io 8 --vm 6 --hdd 8 -t 400** **Note:** For your convenience, a Command reference text file that contains a text version of this script is available on the qwikLABS web page. Copy and paste the script from that text file.
4.4.10	Leave the remaining settings in their default values and click **Next: Add Storage**.
4.4.11	Click **Next: Configure Security Group**.
4.4.12	Click the **Select an existing security group** option and then select **App SG**.

Continued on next page

© 2015 Amazon Web Services, Inc. and its affiliates. All rights reserved.

PRINTED BY: gleung@datapipe.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Creating a Launch Configuration, Continued

Task 4.4: Create a launch configuration, continued

Step	Action
4.4.13	Click **Review**. In the warning message, click **Continue**. **Note:** In this lab, you will not be logging in to the instance via SSH; therefore, that port is not opened in the **App SG** security group.
4.4.14	Review the settings and then click **Create launch configuration**.
4.4.15	When prompted, accept the key pair generated by qwikLABS.
4.4.16	Select the acknowledgement check box, and then click **Create launch configuration**.

© 2015 Amazon Web Services, Inc. and its affiliates. All rights reserved.

PRINTED BY: gleung@datapipe.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Creating an Auto Scaling Group

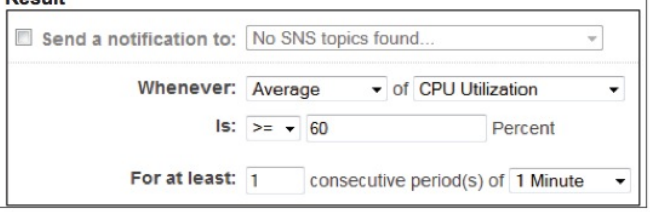
Introduction

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 4.5: Create an Auto Scaling group

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

To create an Auto Scaling group:

Step	Action
4.5.1	On the **Create Auto Scaling Group** page, specify the following settings: • In the **Group name** box, type **Lab-ASG** • In the **Group size** box, type **2** • For **Network**, select **10.200.0.0/16 | Lab VPC**. • For **Subnet**, select **Private Subnet 1** and **Private Subnet 2**.
4.5.2	Expand the **Advanced Details** section: • For **Load Balancing**, select the **Receive traffic from Elastic Load Balancer(s)** check box, and then select **My-LB** (the load balancer that you created earlier).
4.5.3	Click **Next: Configure scaling policies**.
4.5.4	Click the **Use scaling policies to adjust the capacity of this group** option, and change the policy to "Scale between **2** and **6** instances."
4.5.5	Under **Increase Group Size**, click the **Add new alarm** link for the **Execute policy when** section.
4.5.6	In the **Create Alarm** dialog box: • Clear the **Send a notification to** check box. • In the **Percent** box, type **60** • Change **consecutive period(s) of** 5 Minutes to **1 Minute**. **Result** 

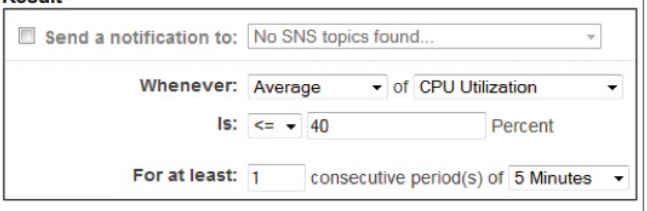
Continued on next page

© 2015 Amazon Web Services, Inc. and its affiliates. All rights reserved.

PRINTED BY: gleung@datapipe.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Creating an Auto Scaling Group, Continued

Task 4.5: Create an Auto Scaling group, continued

Step	Action
4.5.7	Click **Create Alarm**.
4.5.8	In the **Increase Group Size** section, specify the following settings: • Change the **Take the action** value to "Add **2** instances." • Change the **And then wait** value to "**90** seconds before allowing another scaling activity."
4.5.9	Under **Decrease Group Size**, click the **Add new alarm** link for the **Execute policy when** section.
4.5.10	In the **Create Alarm** dialog box: • Clear the **Send a notification to** check box. • Change ">=" to "<=". • In the **Percent** box, type **40** **Result** 
4.5.11	Click **Create Alarm**.
4.5.12	Still in the **Decrease Group Size** section, modify the following: • Change the **Take the action** value to "Remove **2** instances." • Change the **And then wait** value to "**90** seconds before allowing another scaling activity."
4.5.13	Click **Review**. **Note:** You will not configure notifications in this lab.
4.5.14	Click **Create Auto Scaling group.**
4.5.15	Click **Close**.

Continued on next page

© 2015 Amazon Web Services, Inc. and its affiliates. All rights reserved.

PRINTED BY: gleung@datapipe.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Creating an Auto Scaling Group, Continued

Task 4.5: Create an Auto Scaling group, continued

Step	Action
4.5.16	With the newly created Lab-ASG auto scaling group selected, click the **Scaling History** tab. The Auto Scaling group should launch two instances because the group size was set to 2. From the EC2 Dashboard, click **Instances**. Wait until the instances are fully started and the **Status Checks** shows **2/2 checks passed**. In Step 4.4.9, you defined the launch configuration with user data that emulates resource consumption: … yum -y install stress sleep 90 stress --cpu 4 --io 4 --vm 2 --hdd 4 -t 400 sleep 300 stress --cpu 6 --io 6 --vm 4 -hdd 6 -t 400 sleep 300 stress --cpu 8 --io 8 --vm 6 --hdd 8 -t 400 Refresh your Scaling History tab regularly, and within about 5 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.

Ending your lab

When you are done, follow the instructions in Appendix B to end your lab.

Achievements

Having completed this lab, you are able to:

- Create a launch configuration.
- Create a load balancer.
- Create an Auto Scaling Group.
- Define Auto Scaling policies.

Your results are verified by your ability to complete the lab in the allotted time. When you have completed the lab, your instructor will facilitate a discussion of your results.

© 2015 Amazon Web Services, Inc. and its affiliates. All rights reserved.