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 6: Creating a Batch Processing Cluster

Lab Overview

Introduction

In this lab, you will use the AWS Management Console to build a basic batch processing cluster.

You will:

- Launch and configure an Amazon EC2 instance that will serve as the template for future worker nodes in your batch processing cluster.
- Create an Amazon Machine Image (AMI) from that instance.
- Use Amazon SQS to create task queues for passing messages to your instances.
- Launch an Auto Scaling group of instances based on your AMI.
- Schedule work via your task queue.
- Observe the output queue.

The worker nodes in your cluster have a simple job: to convert some number of individual images into a single montage image. A worker node will download images from a list that you provide and will then stitch them into a composite montage using the ImageMagick tool. While this is not the most CPU-intensive job, it does require some cycles; the larger the size and number of images you provide for each job, the more work each node will have.

For this lab, you will provide a newline-delimited list of image URLs. An Amazon EC2 worker node will download each image and produce output such as:



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	40 minutes
Total:	**41 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:

- Bootstrap an EC2 instance using user data.
- Create an AMI from a running instance.
- Use the AWS Management Console to create an Amazon SQS queue.
- Create an Auto Scaling group with scaling policies based on an Amazon SQS queue.
- Use the AWS Management Console to pass messages to, and read messages from, an Amazon SQS queue.

© 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 IAM Role

Introduction

Your batch processing nodes will communicate with Amazon SQS to receive processing instructions and will then store results in Amazon S3. Start by creating an IAM role that grants access to both Amazon SQS and Amazon S3. This role will be assigned to your Amazon EC2 instance.

Task 6.1: Create a role

To create an IAM role:

Step	Action
6.1.1	In the AWS Management Console, on the **Services** menu, click **IAM**.
6.1.2	In the navigation pane, click **Roles**.
6.1.3	Click **Create New Role**.
6.1.4	In the **Role Name** box, type **BatchProcessing**
6.1.5	Click **Next Step**.
6.1.6	On the Select Role Type page, ensure **AWS Service Roles** is selected.
6.1.7	In the AWS Service Roles menu, locate **Amazon EC2** and click **Select.**
6.1.8	On the **Set Permissions** page, ensure that **Select Policy Template** is selected.
6.1.9	In the **Select Policy Template** menu, locate the **Amazon SQS Full Access** policy and click **Select**. A policy document should be displayed for you to review.
6.1.10	Click **Next Step**.
6.1.11	Click **Create Role**.
6.1.12	Click the name of the newly created **BatchProcessing** role to display the details view.
6.1.13	To add an additional role policy, click **Attach Role Policy**.
6.1.14	Locate **Amazon S3 Full Access**, and click **Select**.
6.1.15	Click **Apply Policy**.

© 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 Master Amazon EC2 Instance

Introduction

You will now launch an Amazon EC2 instance with a configuration script that loads ImageMagick and the batch processing software. This master instance will then be used to create an Amazon Machine Image (AMI).

Task 6.2: Launch an instance

To create a master instance:

Step	Action
6.2.1	On the **Services** menu, click **EC2**.
6.2.2	In the navigation pane, click **Instances**.
6.2.3	Click **Launch Instance**.
6.2.4	On the **Choose an Amazon Machine Image (AMI)** page, click **Select** for **Amazon Linux AMI... (HVM)**.
6.2.5	On the **Choose an Instance Type** page, leave the default and click **Next: Configure Instance Details**.
6.2.6	On the **Step 3: Configure Instance Details** page, use the following values: • **Network: Lab VPC** • **Subnet:** Select either **Public Subnet 1** or **Public Subnet 2** • **Auto-assign Public IP: Enable** • **IAM Role: BatchProcessing**
6.2.7	Click **Advanced Details.**
6.2.8	Next, you will supply a user data field to the Amazon EC2 instance so that it is configured to load ImageMagick and the batch processing software. In the **User data** section: • Verify that the **As text** option is selected. • Using the command reference file that you downloaded earlier, copy the user data text for this procedure into the **User data** box.
6.2.9	Click **Next: Add Storage**.
6.2.10	Accept the default and click **Next: Tag Instance**.
6.2.11	In the **Value** box, type **Master**
6.2.12	Click **Next: Configure Security Group**.
6.2.13	For **Assign a security group**, the option **Create a new security group** should be selected. Specify the following settings: • In the **Security group name** box, type **BatchProcessing** • In the **Description** box, type **Batch Processing Security Group**

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 Master Amazon EC2 Instance, Continued

Task 6.2: Launch an instance, continued

Step	Action
6.2.14	To accept the default traffic rule settings, click **Review and Launch.** **Note:** For the purposes of this lab, you are creating a security group that is open to the world via SSH. This is because a more specific security group may not work on the network where you are performing this lab. In practice, you would specify a security group locked down to a specific public IP range allocated to your organization.
6.2.15	Review the settings, and then click **Launch**.
6.2.16	To accept the key pair generated by qwikLABS, select the acknowledgement check box.
6.2.17	Click **Launch Instances**.
6.2.18	Click **View Instances**. Wait until the Instance State of the **Master** Instance changes to *running* and the Status Checks shows *2/2 checks passed*.

© 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.

Connecting to Your Instance

Introduction

You will use your **Master** instance as the basis for an AMI. Before doing so, you will connect to the instance via SSH to ensure that the necessary files were loaded via the User data script.

Task 6.3: Download a key pair

To download the key pair file generated by qwikLABS:

Step	Action
6.3.1	Return to the qwikLABS web page and click the **Download PEM/PPK** drop-down list. • Windows users: click **Download PPK**. • Mac/Linux users: click **Download PEM**.
6.3.2	Save the file to the directory of your choice.

© 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.

Task 6.4: Connecting to your instance with PuTTY (Windows)

Note: This section is for **Windows** users only. **Mac** or **Linux** users, skip to Task 6.5.

To connect to the newly created Master instance:

Step	Action
6.4.1	Launch **PuTTY**.
6.4.2	Copy the **Public IP** of your Master instance from the EC2 Dashboard and paste it in the **Host Name** box.
6.4.3	In the **Category** list, expand **SSH** and then click **Auth**.
6.4.4	In the **Private key file for authentication** box, browse to the **PPK** file that you downloaded in the previous step.
6.4.5	Click **Open** to initiate the connection, and log in as **`ec2-user`**.
6.4.6	Run the **`ls`** command to view the contents of your user directory. If you see entries named **`image_processor.py`**, **`MessageProducer.py`** and **`jobs`**, your instance is correctly configured.
6.4.7	To end your SSH session, type **exit**
6.4.8	Proceed to **Task 6.6: Create an AMI from a running instance.**

© 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.

Task 6.5: Connecting to your instance using a shortcut

Note: This section is for **Mac** and **Linux** users only.

To connect to your Amazon EC2 instance using PEM:

Step	Action
6.5.1	Copy the **Public DNS** value of your **Master** instance from the EC2 Dashboard.
6.5.2	Run the following commands in Terminal to connect to the Amazon EC2 instance. **chmod 600** *<location-of-pem>* **ssh -i** *<location-of-pem>* **ec2-user@***<host-IP>* Substitute the Public DNS value for *<host-IP>*.
6.5.3	Run the **ls** command to view the contents of your user directory. If you see entries named **image_processor.py**, **MessageProducer.py** and **jobs**, your instance is correctly configured.
6.5.4	To end your SSH session, type **exit**
6.5.5	Proceed to **Task 6.6: Create an AMI from a running instance.**

© 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 AMI from Your Batch Processing Instance

Introduction

In this section, you will use the AWS Management Console to create an AMI from the running **Master** instance. By creating an AMI from the instance you configured, you can launch many more identically configured instances quickly and easily.

Task 6.6: Create an AMI from a running instance

Create an AMI named **Worker Image** from the **Master** instance that is running. You learned how to create an AMI in Lab 3. Follow the steps below.

To connect to your Amazon EC2 instance using PuTTY:

Step	Action
6.6.1	Return to the **EC2 Dashboard**.
6.6.2	In the navigation pane, click **Instances**.
6.6.3	If it is not already selected, select **Master**.
6.6.4	In the **Actions** drop-down list, point to **Image** and then click **Create Image**.
6.6.5	In the **Create Image** dialog box, specify the following settings: • In the **Image name** box, type **Worker Image** • In the **Image description** box, type **Batch Processing worker** Leave the rest as the default.
6.6.6	Click **Create Image**.
6.6.7	On the confirmation screen, click the **View pending image** link.

© 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 Two Amazon SQS Task Queues

Introduction

In this section, you will use the AWS Management Console to create two Amazon Simple Queue Service (SQS) queues to hold input and output tasks. You will then send a message to your input queue. You will eventually dispatch work via the input queue and view the results provided by your worker nodes in the output queue.

Task 6.7: Create queues

To create queues named **input** and **output** in Amazon SQS:

Step	Action
6.7.1	On the **Services** menu, click **SQS**.
6.7.2	Click **Create New Queue**.
6.7.3	In the **Create New Queue** dialog box, specify the following settings: • In the **Queue Name** box, type **input** • Change the **Default Visibility Timeout** to **90** seconds. Leave the remaining settings in their default values.
6.7.4	Click **Create Queue**.
6.7.5	Repeat these steps to create another queue named **output** with a default visibility timeout value of **90** seconds.
6.7.6	Select your **input** queue. Ensure that the **output** queue is not selected.
6.7.7	In the **Queue Actions** drop-down list, click **Send a Message**.
6.7.8	Copy the list of image URLs from the lab's provided Command Reference File and paste it in the **Enter the text of a message you want to send** box.
6.7.9	Click **Send Message**.
6.7.10	Click **Close**.

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