Easy Setup of Fast HPCC System on AWS

Timothy L Humphrey

6/17/2015
Last Updated: 02/01/2016

Table of Contents

Easy Setup of Fast HPCC System on AWS

1. Introduction

Configuring and deploying an HPCC System on AWS from your Windows computer is a two-step process when using the HPCC CloudFormation template and accompanying scripts found on github at https://github.com/tlhumphrey2/EasyFastHPCCoAWS. The two-step process is:

1. Copy the accompanying scripts and your ssh pem file to an S3 bucket.
2. Using CloudFormation on the AWS console to build a stack that does the rest.

When you get to section 3, "Using CloudFormation …", you need:

1. A placement group in the region where you deploy your HPCC System. Appendix B gives detailed instructions for making one.
2. An ssh key pair on your Windows machine. Appendix C gives detailed instructions on how to make the key pair on the AWS console and download it to your Windows machine.
3. Finally, if you aren't the administrator of your AWS account, then you may not have the permissions to build the stack that launches your HPCC System. So, ask your AWS administrator to follow the instructions in Appendix D, "Making IAM Super-Power-Group", to make the Super-Power-Group and add you to it.

1.1 Document Conventions

Throughout this document we have commands that you must enter at the DOS prompt of a DOS window. These commands are shown in this document like the following command.

```
dir
```

Also, throughout this document are screenshots that show current topic web pages. I don't always show the whole page. Often the right side of the web page has been chopped off. In addition, often you'll see a green arrow that points to something of interest. These green arrows look like:

2. Make an S3 Bucket and Copy Scripts and Ssh Pem File

Here is a summary of this process:

1. Make an S3 bucket to copy scripts (contents of AWSInstanceFiles directory) and ssh pem file.

2. Copy scripts (contents of AWSInstanceFiles directory) and ssh pem file into created S3 bucket.

You need s3cmd installed on your Windows machine. Installation instructions are given in Appendix A.

The following link shows the s3cmd usage commands: http://s3tools.org/usage

2.1 Make an S3 Bucket

First, you need an s3 bucket to copy the scripts and ssh pem file to. The following commands make an s3 bucket called BuildHPCCScripts.

```
python c:\s3cmd\s3cmd mb s3://BuildHPCCScripts
```

Next, in a DOS window, cd into the folder (AWSInstanceFiles) containing all scripts and your ssh pem file for the region you will deploy your HPCC System on AWS. **Don't forget to copy your ssh pem file into AWSInstanceFiles.**

2.2 Copy Scripts and Ssh Pem File to S3 Bucket

The following command copies everything in the current folder into the s3 bucket, BuildHPCCScripts.

```
python c:\s3cmd\s3cmd  --recursive put . s3://BuildHPCCScripts
```

You can check to see if the scripts and ssh pem file are in the s3 bucket using the following command.

```
python c:\s3cmd\s3cmd  ls s3://BuildHPCCScripts/*
```

3. Using CloudFormation on AWS Console to Configure and Deploy an HPCC System

First, before telling you how to use CloudFormation on the AWS console, here is how to use aws cli to start the process of setting up your HPCC System.

```
aws cloudformation create-stack --capabilities CAPABILITY_IAM \
  --stack-name riptide-hpcc  \
  --template-body  <path-to-your-cf-template> \
  ParameterKey=HPCCPlacementGroup,ParameterValue=<your-placement-group> \
  ParameterKey=HPCCPlatform,ParameterValue=HPCC-Platform-5.2.2-1 \
  ParameterKey=KeyPair,ParameterValue=<your-key-pair (pem file)> \
  ParameterKey=MasterInstanceType,ParameterValue=c3.2xlarge \
  ParameterKey=NumberOfRoxieNodes,ParameterValue=1 \
  ParameterKey=NumberOfSlaveInstances,ParameterValue=1 \
  ParameterKey=NumberOfSlavesPerNode,ParameterValue=2 \
  ParameterKey=RoxieInstanceType,ParameterValue=c3.2xlarge \
  ParameterKey=ScriptsS3BucketFolder,ParameterValue=s3://riptide-hpcc/ \
  ParameterKey=SlaveInstanceType,ParameterValue=c3.2xlarge \
  ParameterKey=UserNameAndPassword,ParameterValue=<your-username/your-password>
```

Here is a summary of this process:

1. Navigate to the CloudFormation (CF) web page on the AWS console.
2. Enter a Unique Name for the stack being created and give a path to the HPCC CF template.
3. Fill in the HPCC CF template parameters, e.g. name of pem file, placement group, etc.
4. Click the "Create" button to start the process of creating the HPCC System on AWS.

3.1 Navigate to CloudFormation Web Page

After logging into the AWS console (from your browser), you should see a web page that looks like Figure 1, below. Click on "Services",(pointed to by the green arrow in Figure 1).

Figure 1. Amazon Web Services
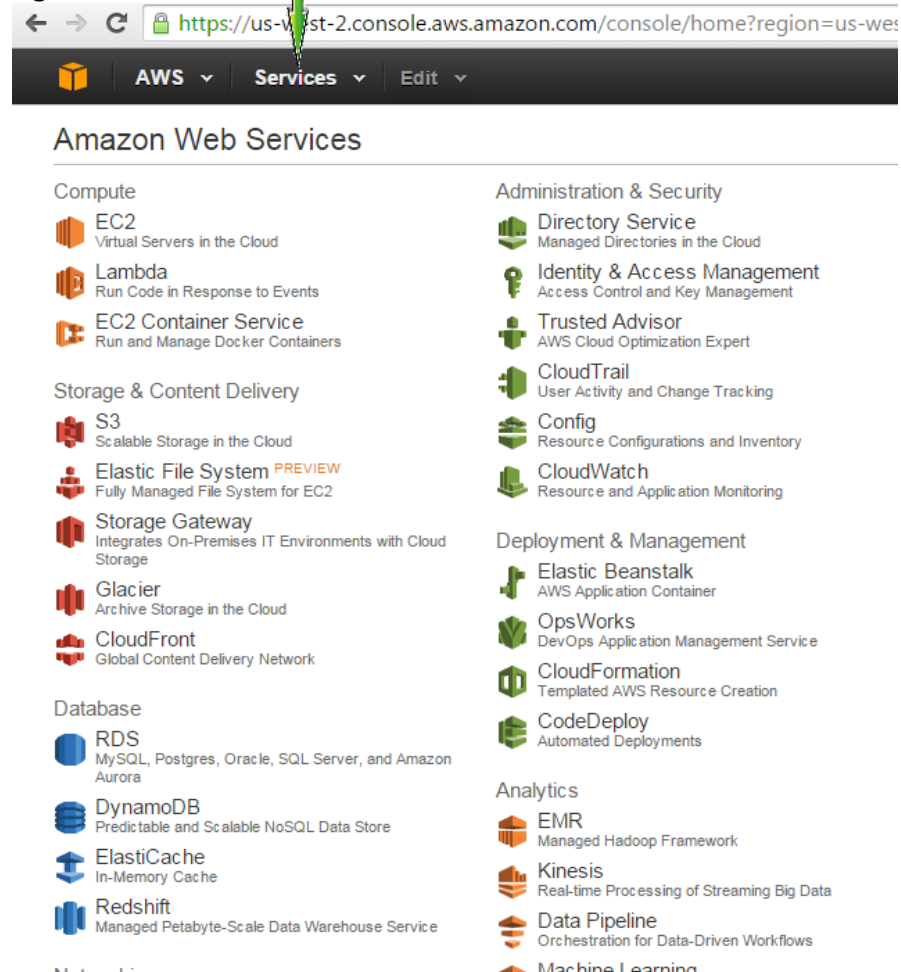
On the next web page, which is a list of all AWS services, click on CloudFormation (pointed to by the green arrow of Figure 2).

Figure 2. All AWS Services
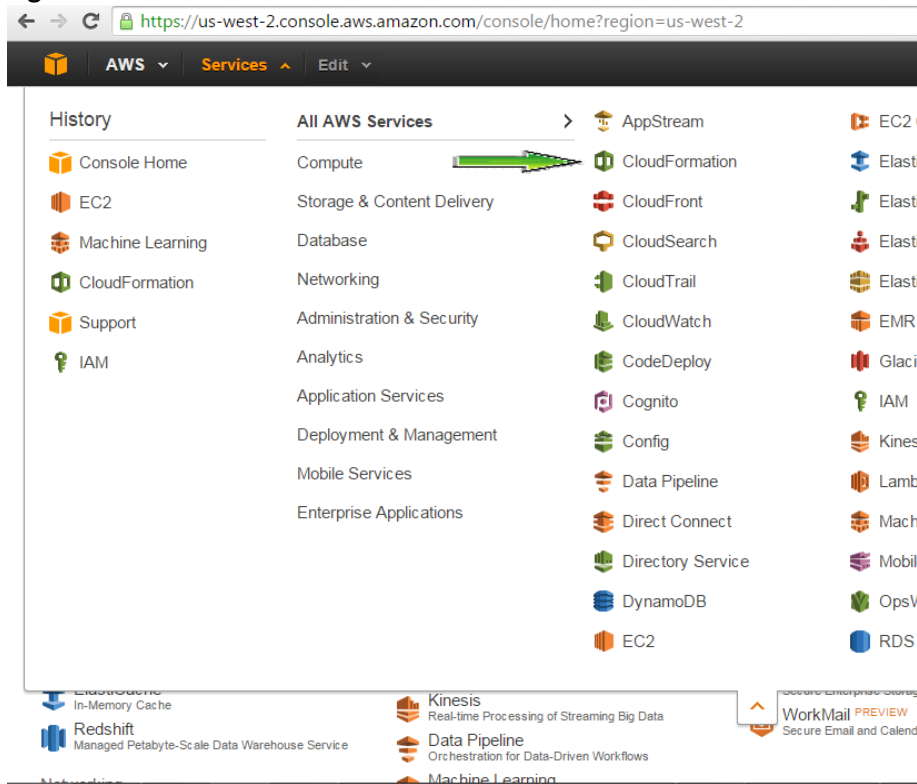
Figure 3 shows the first page of the CloudFormation section. Click on "Create Stack" to begin the process of configuring and deploying an HPCC System to AWS. ("Create Stack" is pointed to by the green arrow of Figure 3).

Figure 3. First CloudFormation Web Page
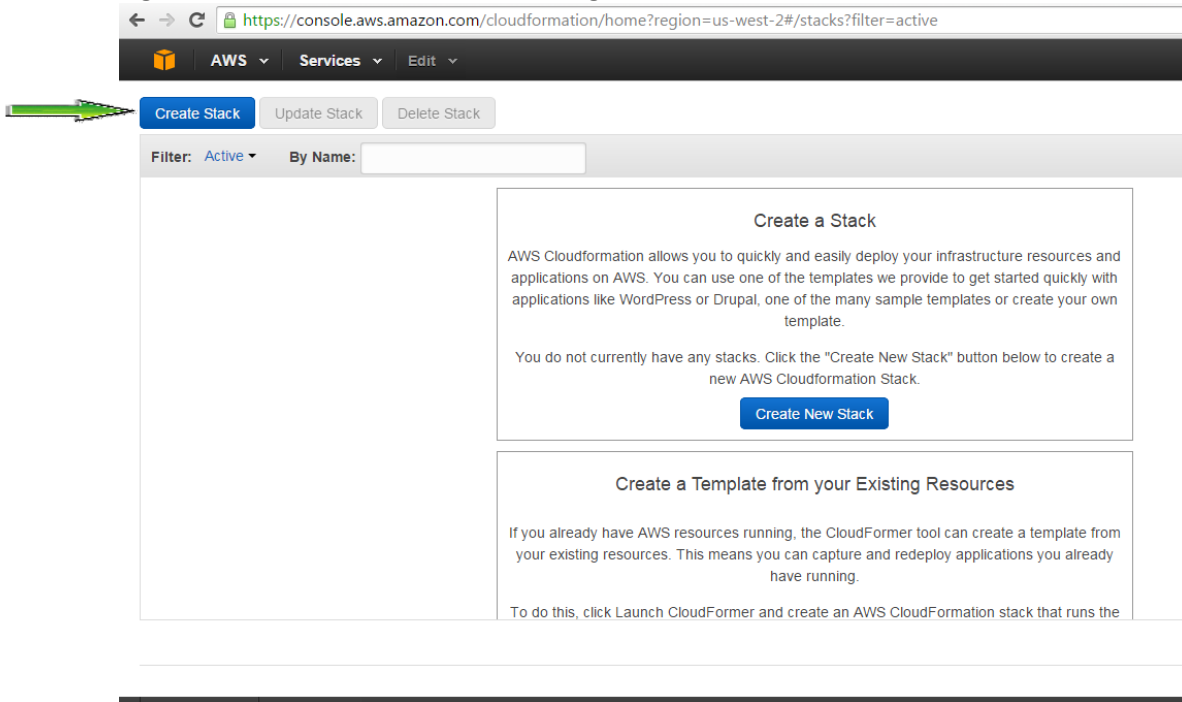
3.2 Enter a Unique Stack Name and Path to HPCC CF Template

Figure 4 shows the next web page you will see. On this page, enter a unique stack name in the textbox labeled "Name" (pointed to by the top green arrow of Figure 4). Then click on the "Choose File" button (pointed to by the bottom green arrow of Figure 4).

⚠️	The stack name MUST be unique – different than any previous stack names.

Figure 4. Select Template Web Page

When you click on "Choose File", you get an open-file dialog box, like that of Figure 5, below.

Figure 5. File Open Dialog Box
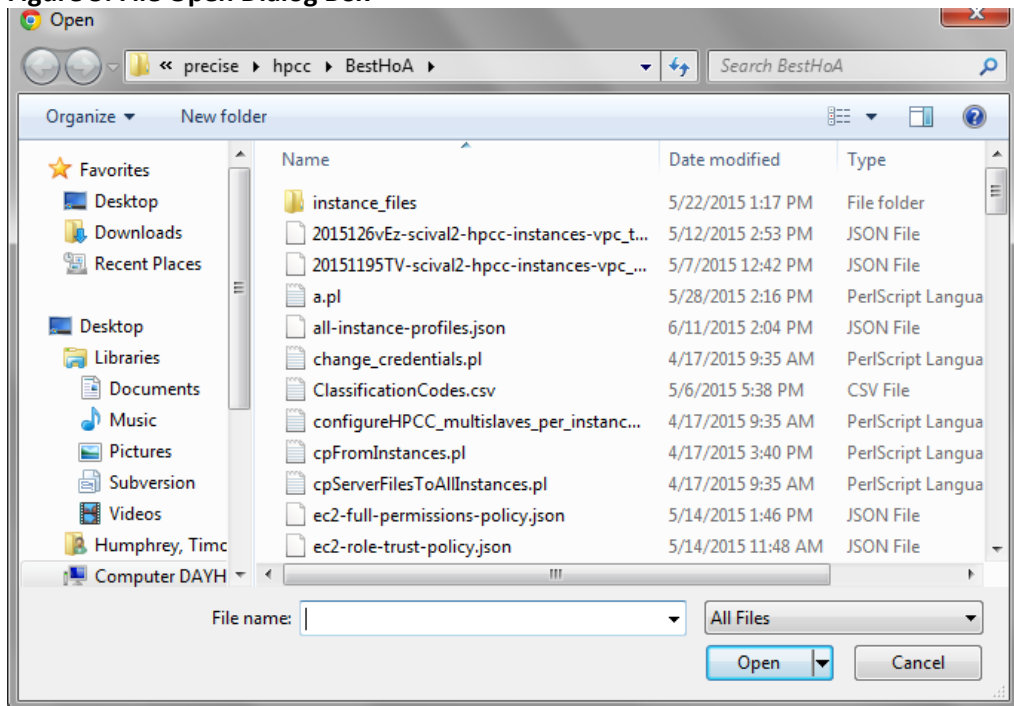
Use the open-file dialog box to find the HPCC CF template on your Windows computer and then click on open. The name of the HPCC CF template is: MyHPCCCloudFormationTemplate.json.

After entering a unique name for the stack and choosing the HPCC CF template, the web page should look like Figure 6, below (note. The following screenshot is after I scrolled down enough to see the "Next" button). Now, click on the "Next" button (pointed to by the green arrow in Figure 6, below).

Figure 6. Filled-in Select Template Web Page

3.3 Enter HPCC CF Template Parameters

After clicking on the "Next" button, the next web page you see should look like Figure 7, below.

Figure 7. HPCC Stack Parameters Web Page

First, notice that the ScriptsS3BucketFolder has been set to the name of the bucket where we placed the scripts and ssh pem, i.e., s3://BuildHPCCScripts (see Section 2.2, above). And, KeyPair has been set to the name of the ssh pem file that was put into the above mentioned s3 bucket; that is tlh_keys_us_west_2 (notice that '.pem' is omitted).

The other non-empty text boxes have been set to the recommended settings for an HPCC System that runs fast on AWS, except for NumberOfRoxieNodes, which has been set to "0".

HPCCPlacementGroup is the only text box that doesn't have anything in it. Enter the name of the placement group you want to use for the current region (Appendix B shows you how to setup a placement group using the AWS console).

> The current AWS region is found in the top right-hand corner of the AWS console display. It is a dropdown menu. So, you can select a different AWS region. My current AWS region was "Oregon", us-west-2.

For the purpose of demonstrating the use of the HPCC CF template, I will change some of the recommended values so the HPCC System that is configured and deployed doesn't cost me much money. The following screenshot is what Figure 7 looks like after making my changes.

Figure 8. Filled-in HPCC Stack Parameters Web Page

Notice I changed NumberOfSlavesPerNode from 12 to 2, the NumberOfSlaveInstances from 7 to 1 and the instance types for the Master and Slaves, I changed to c3.2xlarge. Also, notice in Figure 8, above, that I've set HPCCPlacementGroup to "pg-tlh-best", which is the name of the placement group I created for the us-west-2 region, which is where this stack will be created.

3.4 Start Creation Process

Once you have the parameters the way you want them, you click on "Next" (where green arrow points in Figure 8, above). The next web page looks like Figure 9, below.

Figure 9. Stack Options Web Page

There is nothing to do on this web page. So, just click the "Next" button to continue to the next web page which looks like Figure 10, below.

Figure 10. Review Parameters Web Page

This page shows all your settings, so you can verify they are what you want before starting the stack building process.

The next screenshot shows the same web page as above, except I've scrolled down so you can see the "Create" button. You click on the "Create" button to start the stack building process.

But, before you do, you must click on the acknowledge checkbox (pointed to by the green arrow of Figure 11, below) to indicate that you know that the template comes from a trusted source. Why? Because the template creates AWS resources that you have to pay for and also the template gives permission to the instances created to access AWS resources such as S3 buckets.

Figure 11. Bottom of Review Parameters Web Page

The next web page you see should look like Figure 12, below.

It is divided into 2 parts. The top part shows the stacks (in this case only one). And the bottom part shows the progress of the stack build process for the stack with the checkbox "checked" (where the green arrow points in Figure 12).

Figure 12. Show Stacks Web Page
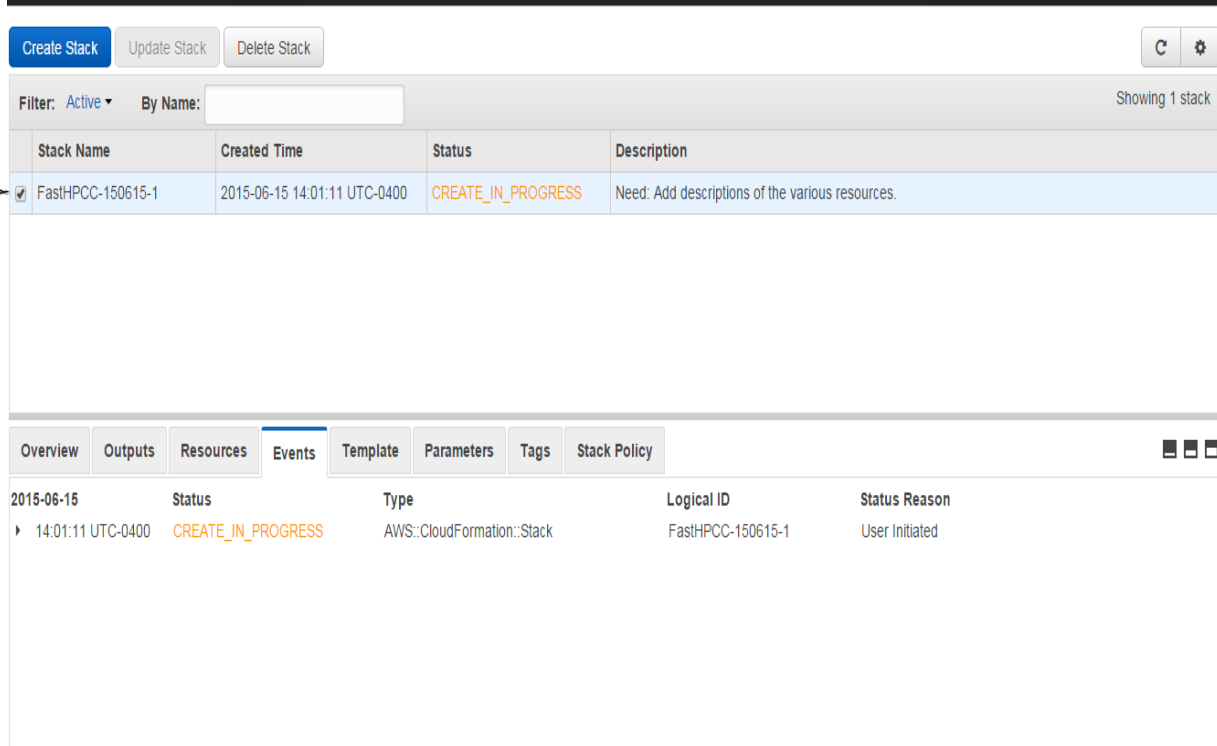
If you refresh the web page, you will see an update of the progress of the stack build process (like Figure 13, below).
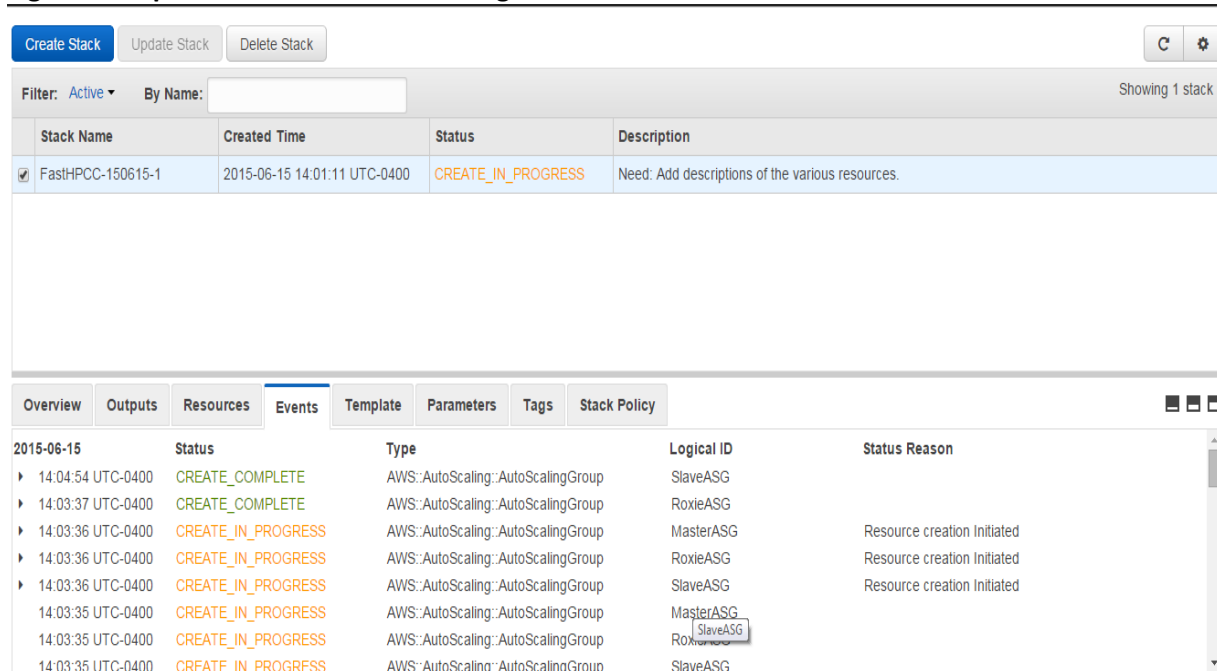
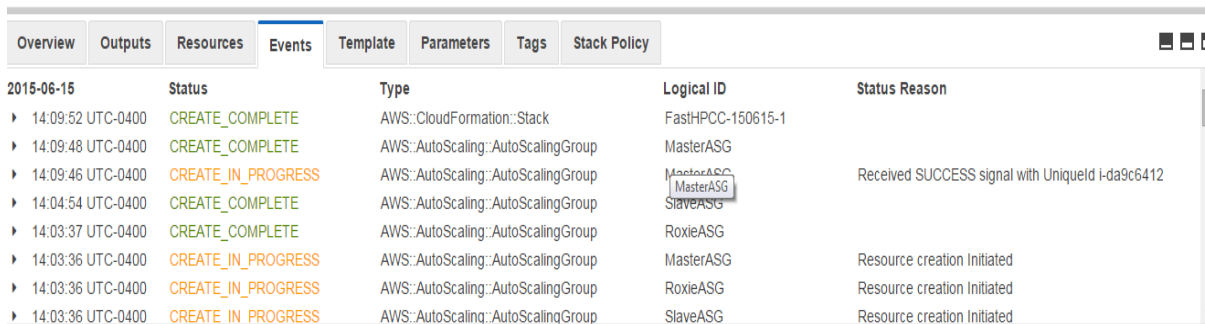**Figure 13. Updated Show Stacks Web Page**

	During the stack creation process, you may get an error message that contains the words "not authorized to perform:iam:CreateRole", which means you don't have the proper permissions to create the stack. So, ask your AWS administrator to follow the instructions in Appendix D, "Making IAM Super-Power-Group", to make the Super-Power-Group and add you to it.

The stack build process is done (meaning the HPCC System is setup) when "CREATE_COMPLETE" is the Status of the stack building process (where the green arrow points in Figure 14, below).

Figure 14. Finished Stack Creation

When the stack creation process has completed go to the "Instances" page on the AWS console to get the IP address of the Master, which is needed to access ECL Watch and to setup the ECL IDE.
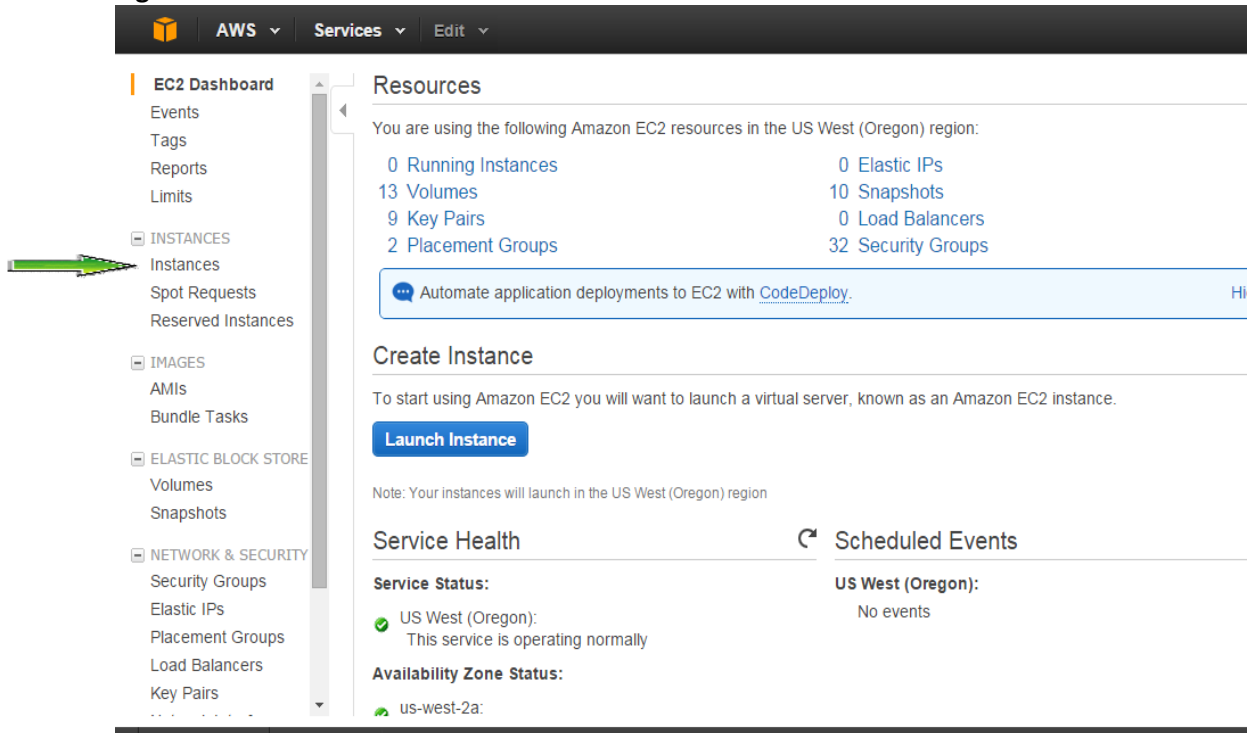
	Once you have finished with the deployed HPCC System, delete the stack so all resources are terminated and AWS charges won't continue.

Getting Public IP of Master

From the EC2 Dashboard, click on "Instances" (pointed to by the green arrow of Figure 15, below).

Figure 15. EC2 Dashboard with Instances Marked

Next, you will see the Instance page which should look like Figure 16, below. The Master instance has been selected by clicking on the button just to the left of the instance's name (pointed to by the top-left green arrow in Figure 16). And because this instance is selected, its Public IP address is shown where the bottom-right green arrow points. You can swipe across this Public IP address and save it. Then, paste it into your browser's address bar and append ":8010" to access ECL Watch.

Figure 16. Instances Page with Master Marked

Appendix A. Install S3cmd on Your Windows Machine

The following link gives detailed instructions for installing s3cmd:

http://tecadmin.net/setup-s3cmd-in-windows/

S3cmd needs both python and gpg for windows. This web site shows how to install both as well as how to install and configure s3cmd.

Figure 17. How to Setup S3cmd on Windows
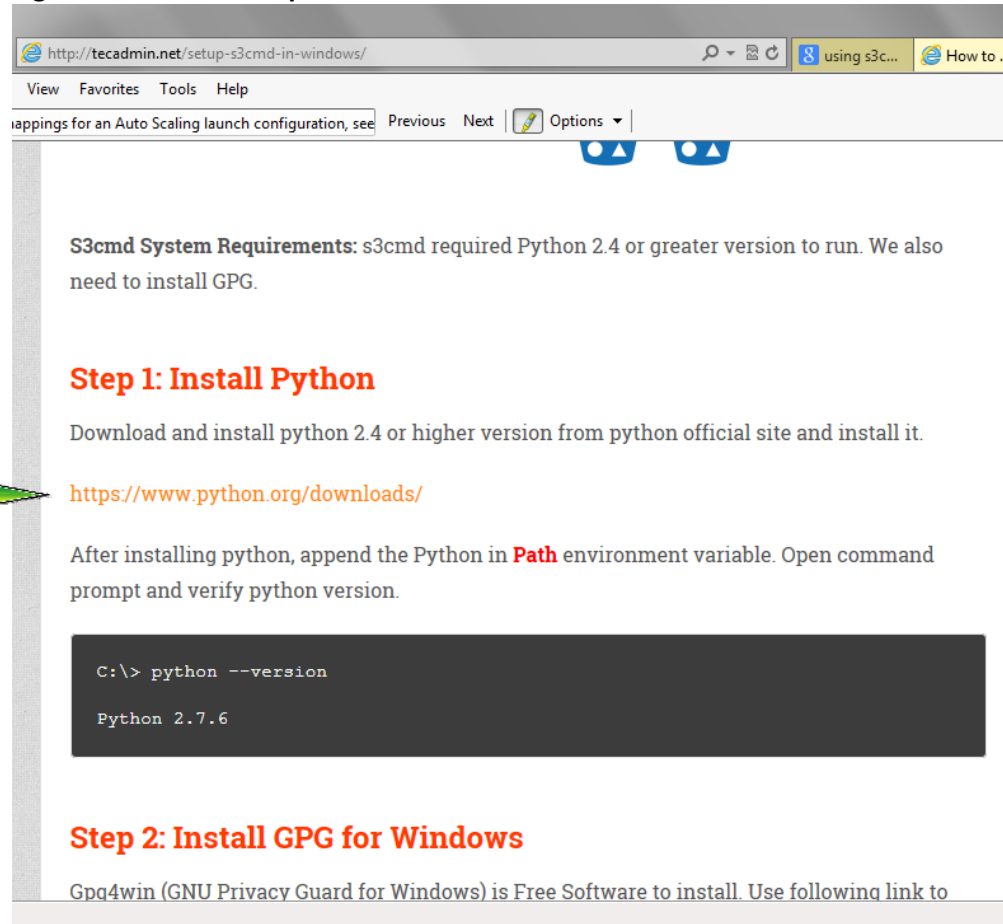

To install python, click on the link pointed to by the green arrow in Figure 15, above. The next page you see looks like Figure 18, below – for downloading python.

You don't want the latest Python. You want Python 2.7.6 which is in the list of other versions shown at the bottom of the page (you may have to scroll down to see the list and them scroll the list to find version 2.7.6).

Figure 18. Download Python
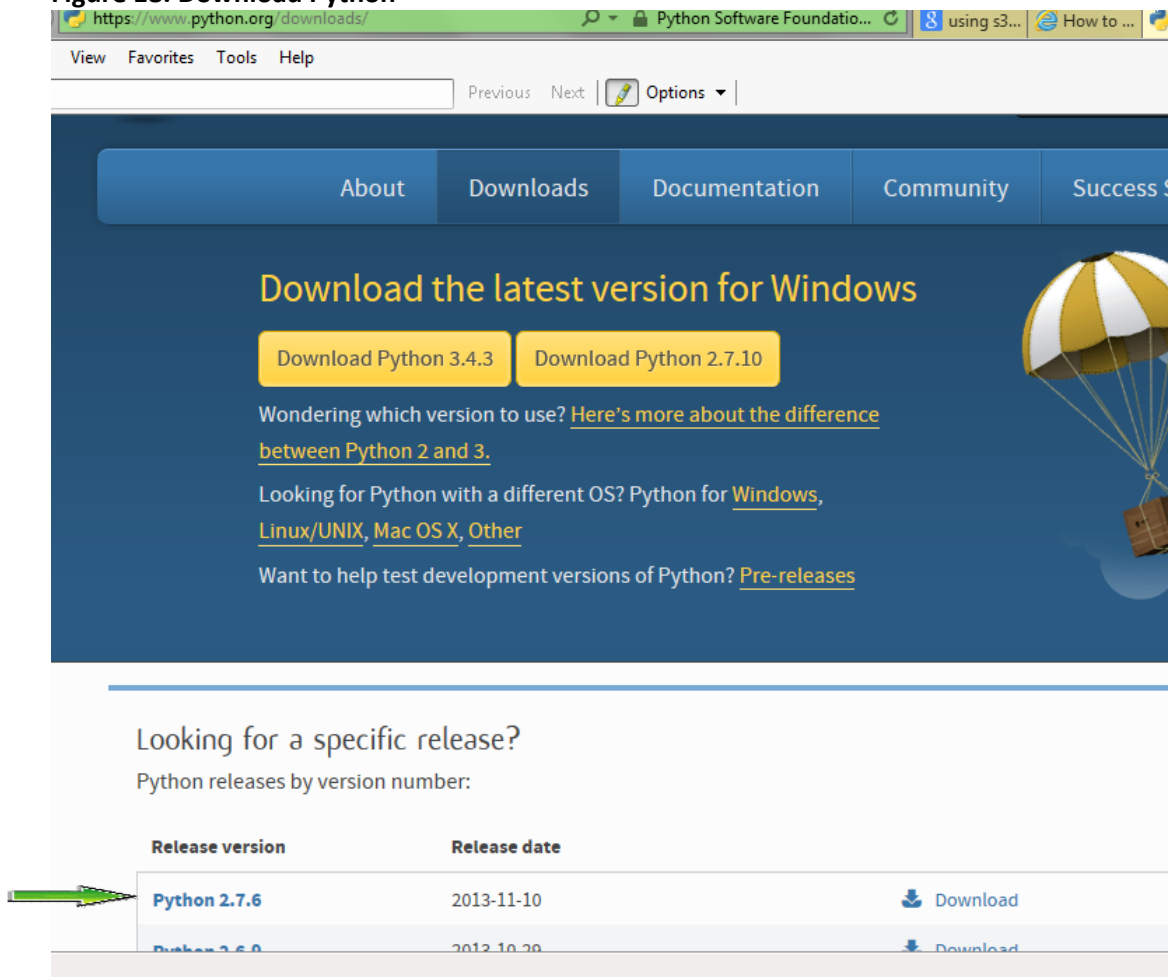
Installing pgp4win

No additional instructions are needed for the installation of pgp4win. Just following the instructions on the "How to Install S3cmd on Windows" web page.

	Remember the path where pgp4win is installed. Later, it is needed for the configuration of s3cmd.

Installing s3cmd

No additional instructions are needed for the installation of s3cmd. Just following the instructions on the "How to Install S3cmd on Windows" web page.

Configuring s3cmd

Once, s3cmd is install, do the following command to configure it. This command creates a the file, .s3cfg, in your home directory.

```
Python c:\s3cmd\s3cmd --configure
```

The above command will prompt you for information. You will use the defaults supplied accept for those prompts given in the following table.

Table 1. S3cmd Configuration Parameters

Prompt	Your Response
Access key	Enter your AWS access key
Secret key	Enter your AWS secret key
Encryption password	Just hit enter, i.e. don't provide a password
Path to GPG program	Enter the path where gpg was installed (my is C:\Program Files (x86)\GNU\GnuPG\pub\gpg.exe)
Test access with supplied credentials	Enter yes (this assures you didn't make any mistakes when entering the above parameters)
Save settings	Enter yes

Appendix B. Making a Placement Group

The EC2 instances of the deployed HPCC System are in the same placement group. This assures the instances are close together and therefore communication is faster.

The following gives detailed instructions for making a placement group in the region your HPCC System is deployed.

Figure 19 is the top right corner of the home page of the AWS console. It and other AWS console pages have the region where instances will be launched in the top right-hand corner of the page (where the green arrow is pointing). If you click on the region, you get a dropdown menu containing all the regions. And, you can pick from the dropdown list another region.
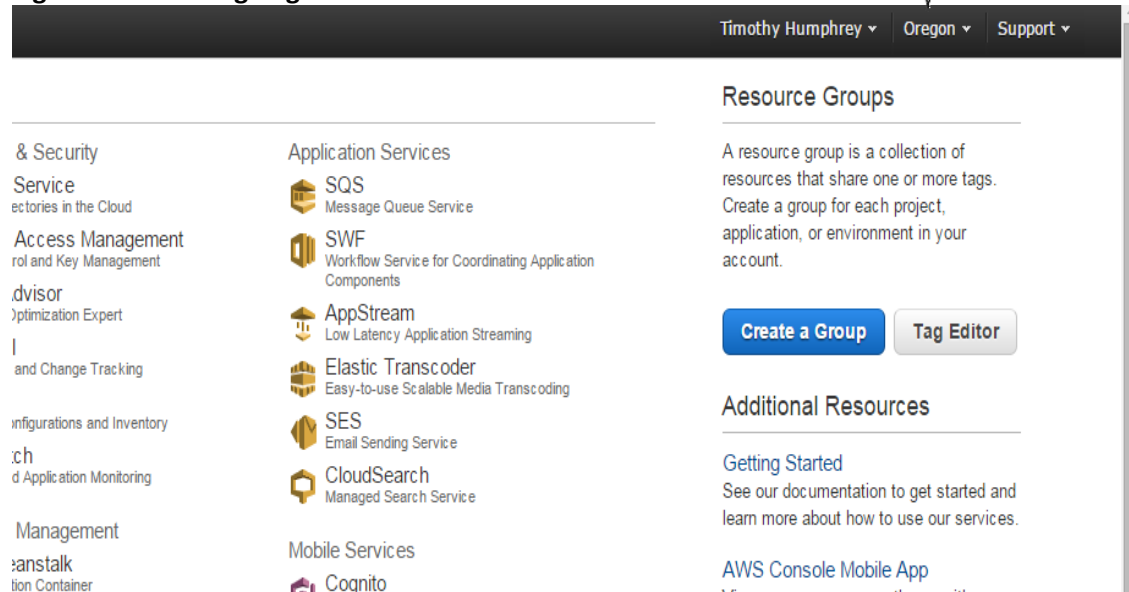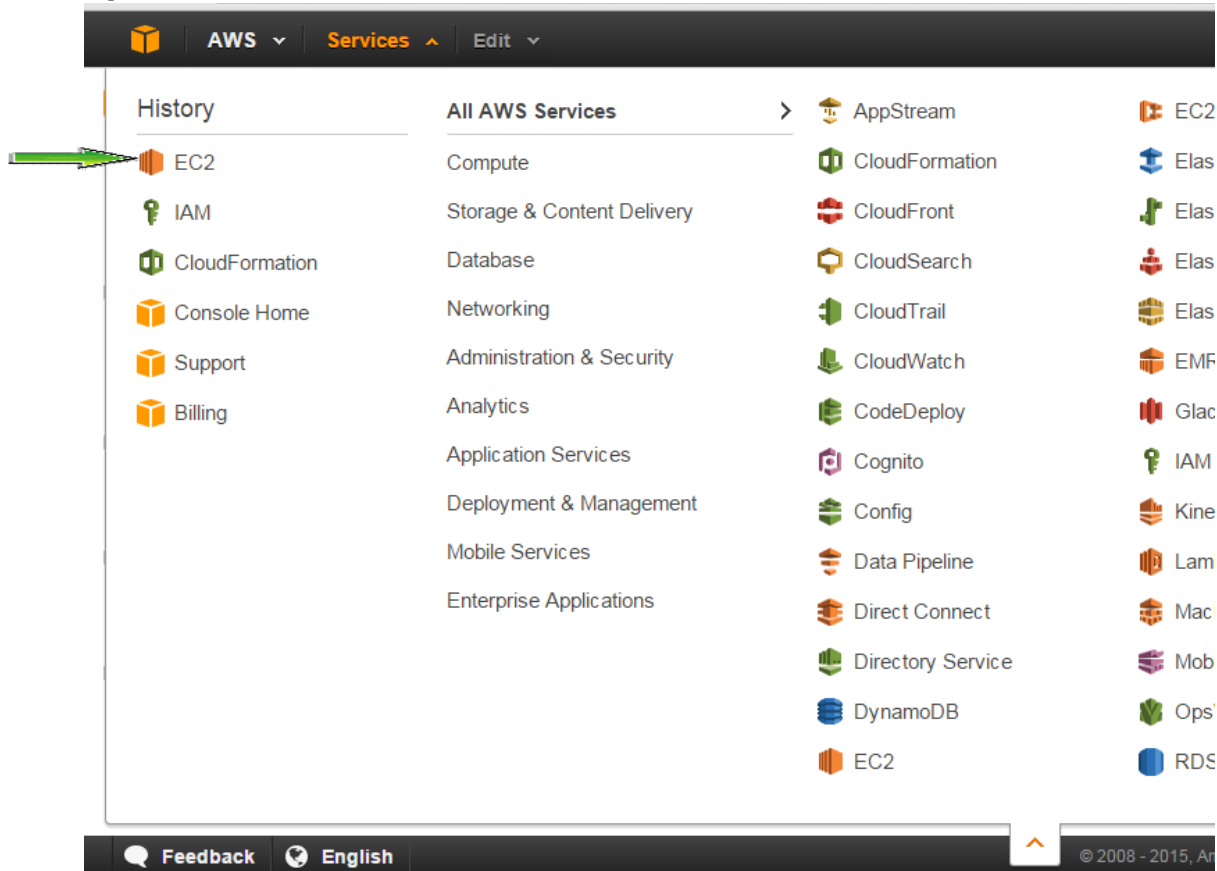
Figure 19. Selecting Region

Figure 20, below, shows the AWS Services page which you get when you click on "Services" in the top left-hand corner of the AWS console home page. Click on EC2 pointed at by the green arrow of Figure18.

Figure 20. AWS Services

Next you will see the EC2 Dashboard page, Figure 21, below. Click on "Placement Groups", pointed to by the green arrow of Figure 21.

Figure 21. EC2 Dashboard

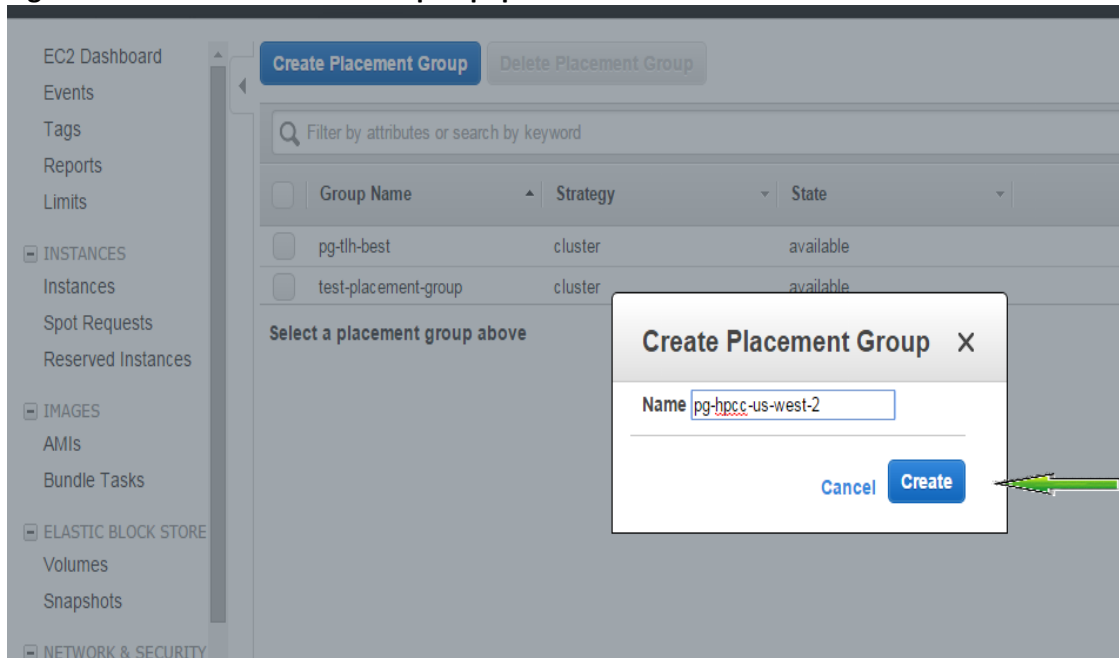Next, you will see Figure 22, Placement Groups. Click on "Create Placement Group", pointed to by the green arrow of Figure 22.

Figure 22. Placement Groups
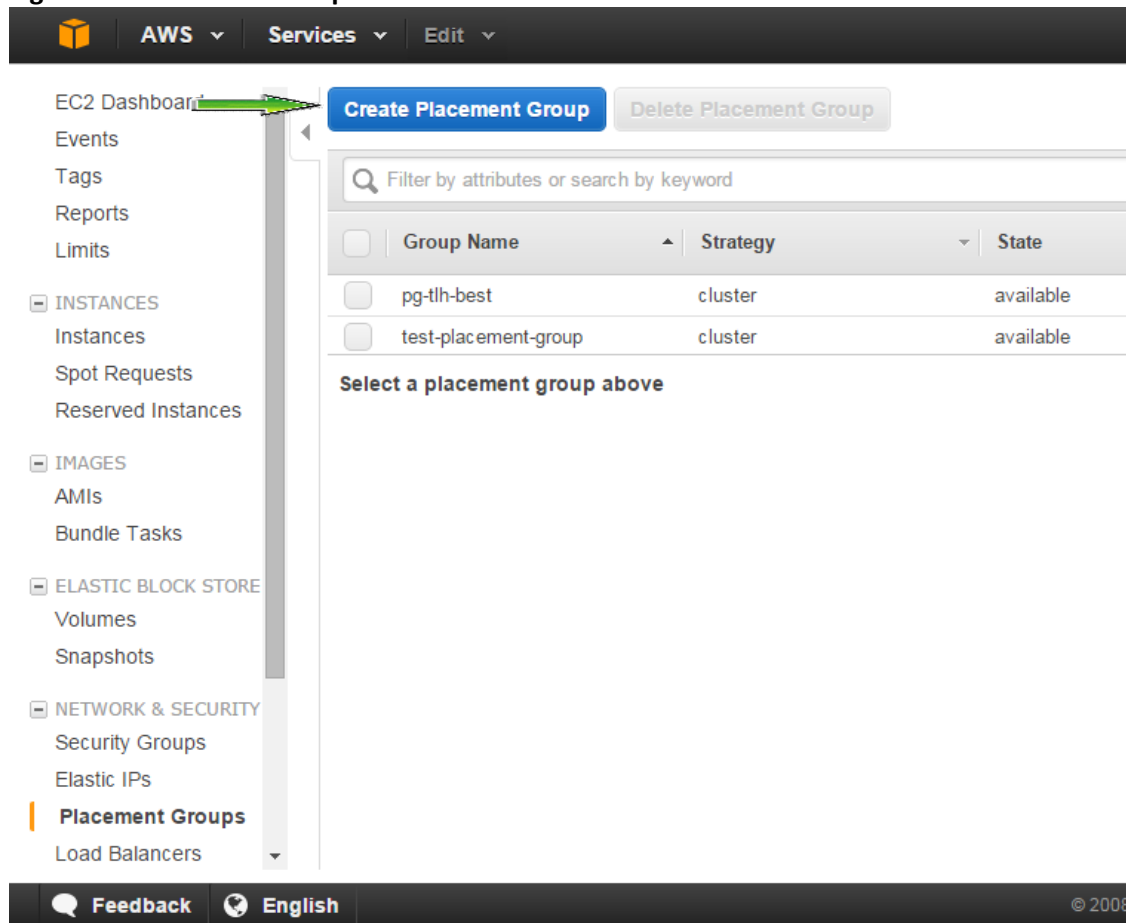
This causes a popup as shown in Figure 23, below. Enter the name you want the placement group to have in the textbox named "Name" (In Figure 23, I have entered the name "pg-hpcc-us-west-2". After entering the name, click on "Create" pointed to by the green arrow of Figure 23. This causes the placement group to be created.

Figure 23. Create Placement Group Popup

Next, as shown in Figure 24, you will see the newly created placement group in the list of placement groups (see where green arrow is pointing in Figure 24).

Figure 24. Newly Created Placement Group
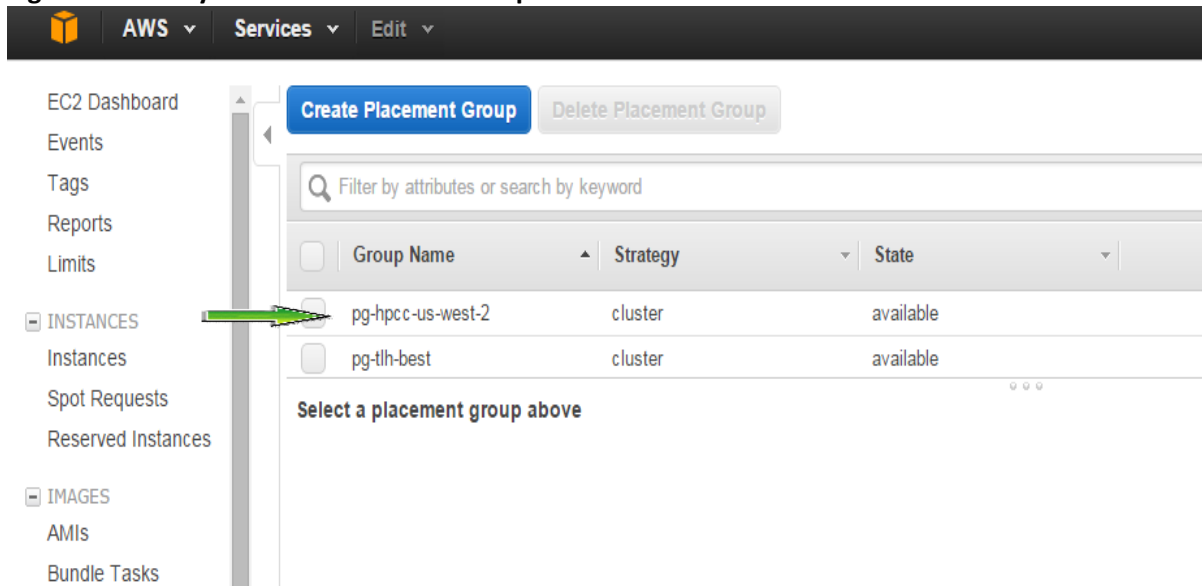
Appendix C. Make a Ssh Pem File

An ssh key pair is used by the nodes of your deployed HPCC System to access other nodes of the system. The following gives details on how to create an ssh key pair and have it pem file downloaded to your Windows computer.

First from the EC2 Dashboard, click on Key Pair (pointed to by the green arrow of Figure 25).

Figure 25. EC2 Dashboard with KeyPair Marked
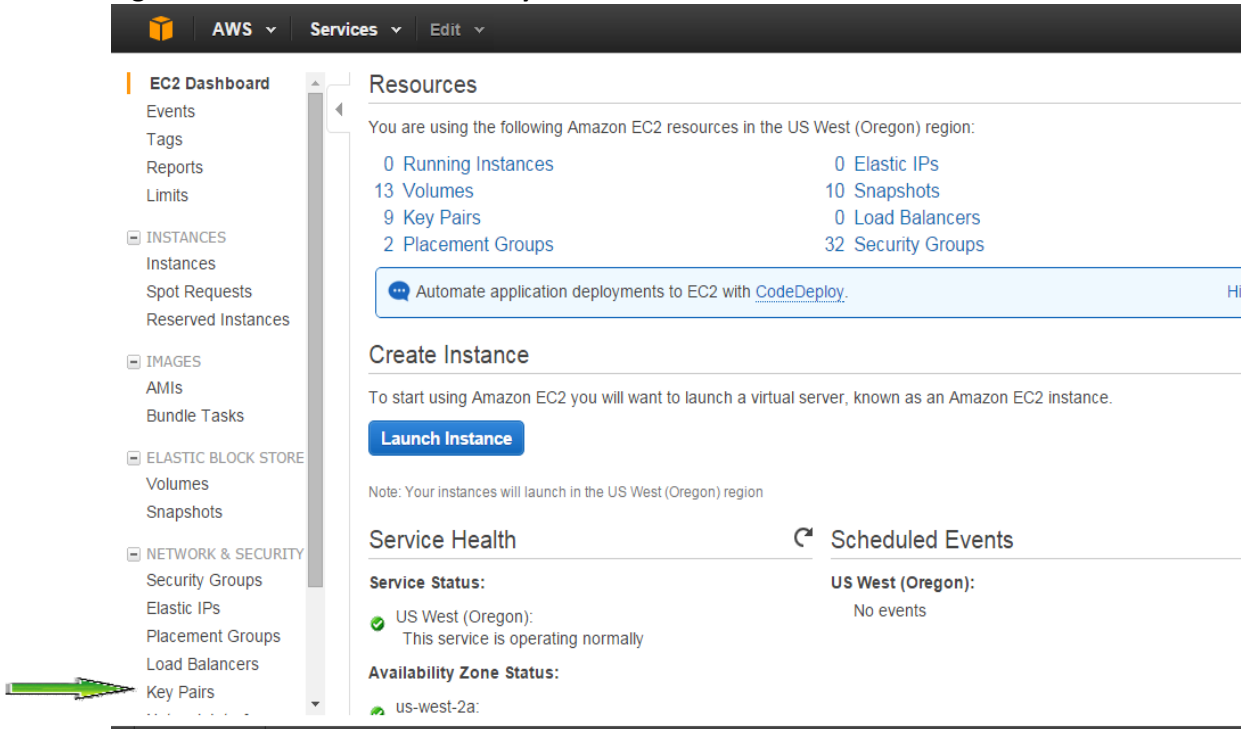
Next, you will see the page where you can create a new key pair. It should look like Figure 26, below. Click on "Create Key Pair" (pointed to by the green arrow of Figure 26).

Figure 26. Create Key Pair
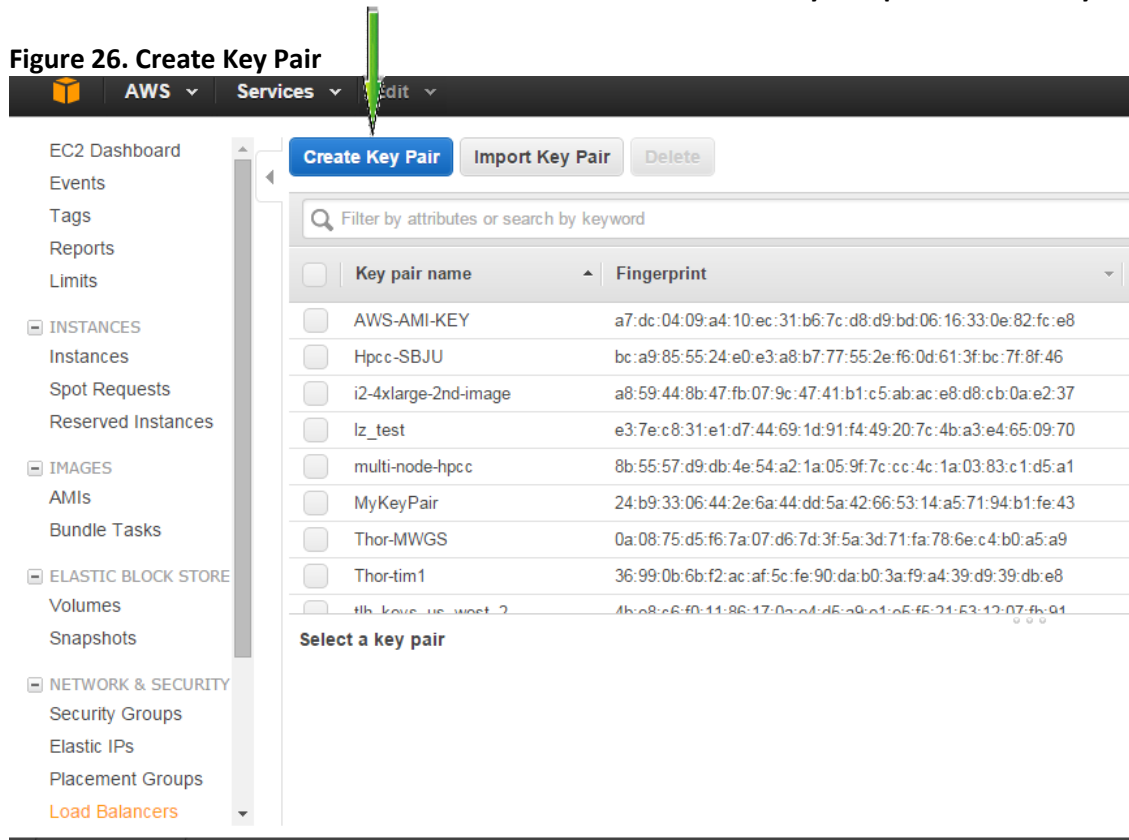
Next, you will see a popup with the title "Create Key Pair". It should look like Figure 27. Enter the name of your new key pair and then click on "Create" (pointed to by the green arrow of Figure 27). You will notice that in Figure 27, I entered the name "us-west-2-keypair".

Figure 27. Create Key Pair Popup
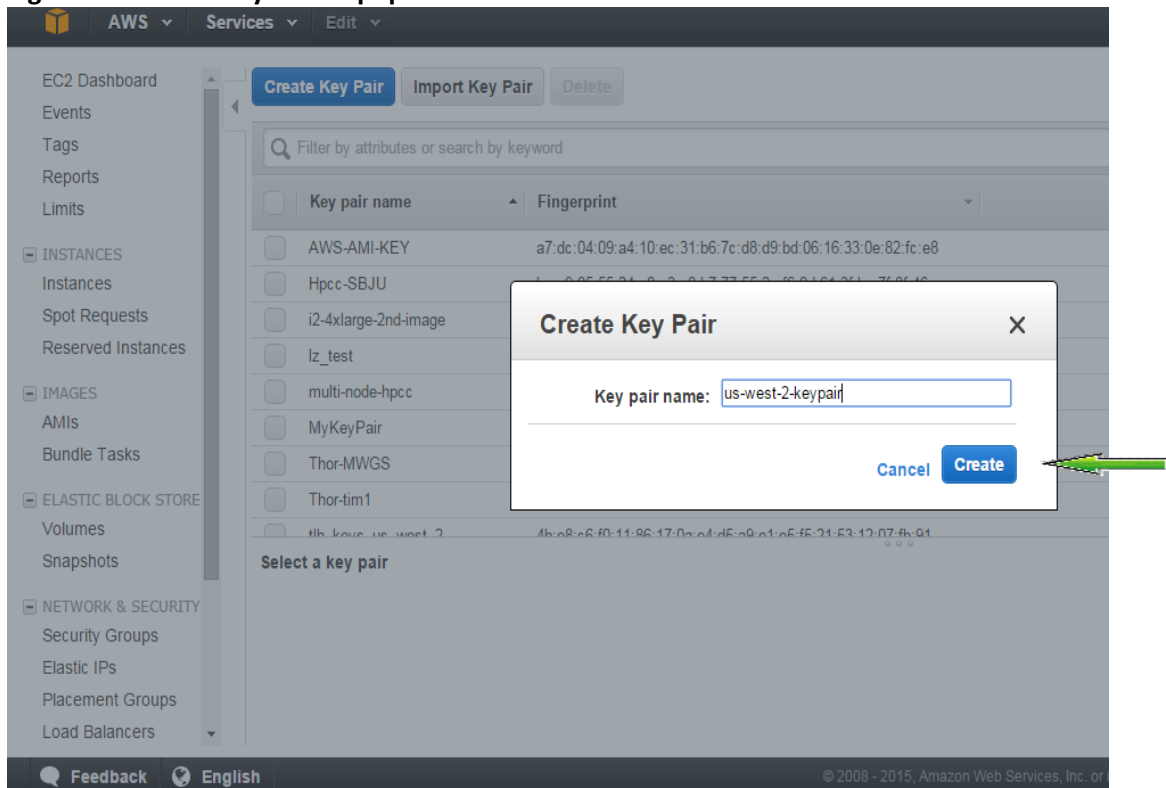
After clicking on the "Create" button, the next think you will see is a page that looks like Figure 28, below. And, you should get an indication that the pem file for the newly created key pair was downloaded to your Windows computer (pointed to by the green arrow of Figure 28, below).

Figure 28. Create Key Pair with Pem File Downloaded

Appendix D. Making IAM Super-Power-Group

This must be done by the administrator of your AWS account. And, it only needs to be done if you get an error during stack creation saying you are "not authorized to perform:iam:CreateRole".

Frist the administrator of your AWS account will create a new IAM group which I call Super-Power-Group using the policy in the file super-user-iam-policy.json. Then, they will place you in that group.

The following AWS cli command creates the Super-Power-Group (This can also be done on the AWS console using the instructions at http://docs.aws.amazon.com/IAM/latest/UserGuide/Using_CreatingAndListingGroups.html).

```
aws iam put-group-policy --group-name Super-Power-Group --policy-document file://super-user-iam-
policy.json  --policy-name super-power-policy
```

To add a user to the Super-Power-Group:

1. Sign into the AWS console and open the IAM console, i.e. select IAM from Services.
2. Open the Users section by clicking on "Users" in the menu on the left.
3. Select the user you want to add to the group.
4. From the User Actions select Add User to Group
5. Select the group and click on Add to Groups (bottom right corner).