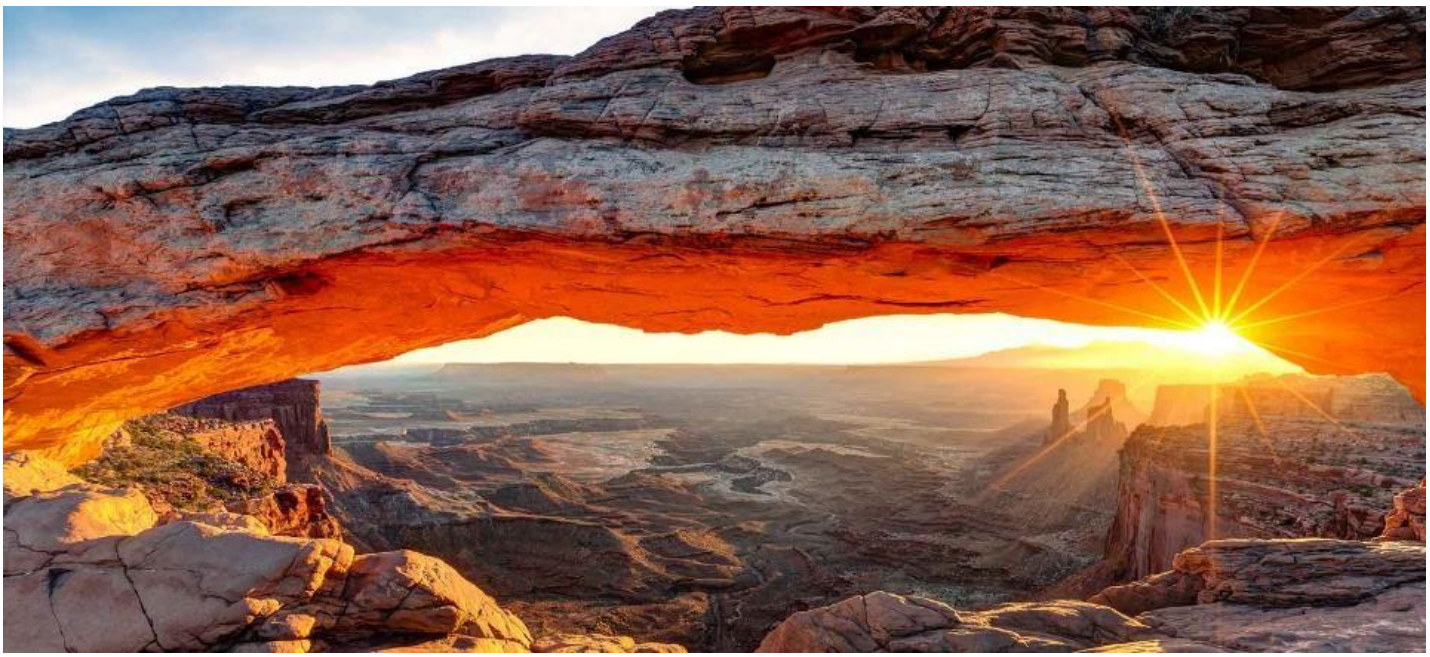


HPCC Systems® Enterprise Edition for the Amazon Marketplace

Setup Instructions

Accessible, Scalable, Proven Technology

Contact:

LNSSI AWS Engineering and Support
LexisNexis® Special Services Inc.
1150 18th Street, NW, Suite 250
Washington, D.C. 20036
Phone: 937-530-9004
Email: AWSMarketplaceSupport-LNSSI@lnssi.com
Homepage: http://lexisnexisspecialservices.com/awssupport/

January 23, 2018
Version 1.1

LexisNexis® SPECIAL SERVICES INC.

Table of Contents

Overview

This document is intended to be used by AWS Marketplace customers who have purchased LexisNexis® Special Services HPCC Systems® Enterprise Edition ("HPCC") software.

The following instructions assume that the user is familiar with the information outlined in the **"Amazon Elastic Compute Cloud User Guide for Linux Instances"**, http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-ug.pdf , and the PuTTY, PuTTYGen, and the PSCP utilities. To learn more about these utilities and how they are used to access AWS EC2 instances please refer to the section "**Connecting to Your Linux Instance from Windows Using PuTTY**" in the **"Amazon Elastic Compute Cloud User Guide for Linux Instances"** documentation.

Delivery Methods

Before creating an HPCC cluster you should decide the size of the cluster, the type of instances you need, and the type of work you wish to perform. The primary example videos include a cluster consisting of four Thor nodes and two Roxie nodes. There is also a video that demonstrates how to add instances to an existing cluster. The secondary videos demonstrate ways to verify the status of the system once it is running, and how to set up the IDE.

It is important to note that when using the multi-instance delivery methods two files are automatically provided on each instance:

- ips_mapping.csv
- private_ips.txt

These two files provide the list of public and private IP addresses for the instances that were provisioned at the time these files were created. It is recommended that you save a copy of these files since they are referenced during the HPCC configuration process and can help users keep a record of which HPCC services are running on the various instances. These files will also be useful if the user ever intends to add more instances to an existing configuration.

Manual Launch - "New VPC Cluster" Set Up and Configuration

Provisioning

To create a new HPCC cluster you will need to fill in the "Stack name" and "PermitSSH" fields on the first page. For the "PermitSSH" field it is possible, but not advisable, to enter "0.0.0.0/0". This will allow any client to **SSH** to your instances. It is suggested instead to use a service such as http://whatismyipaddress.com to determine your IP address and use your IP address and the x.x.x.x/32 CIDR pattern. The "EBS", "InstanceType", and "NumberOfInstances" will vary depending upon your needs. The "KeyPair" will need to coincide with the private key that you wish to use to **SSH** to the instances after they are created and running.

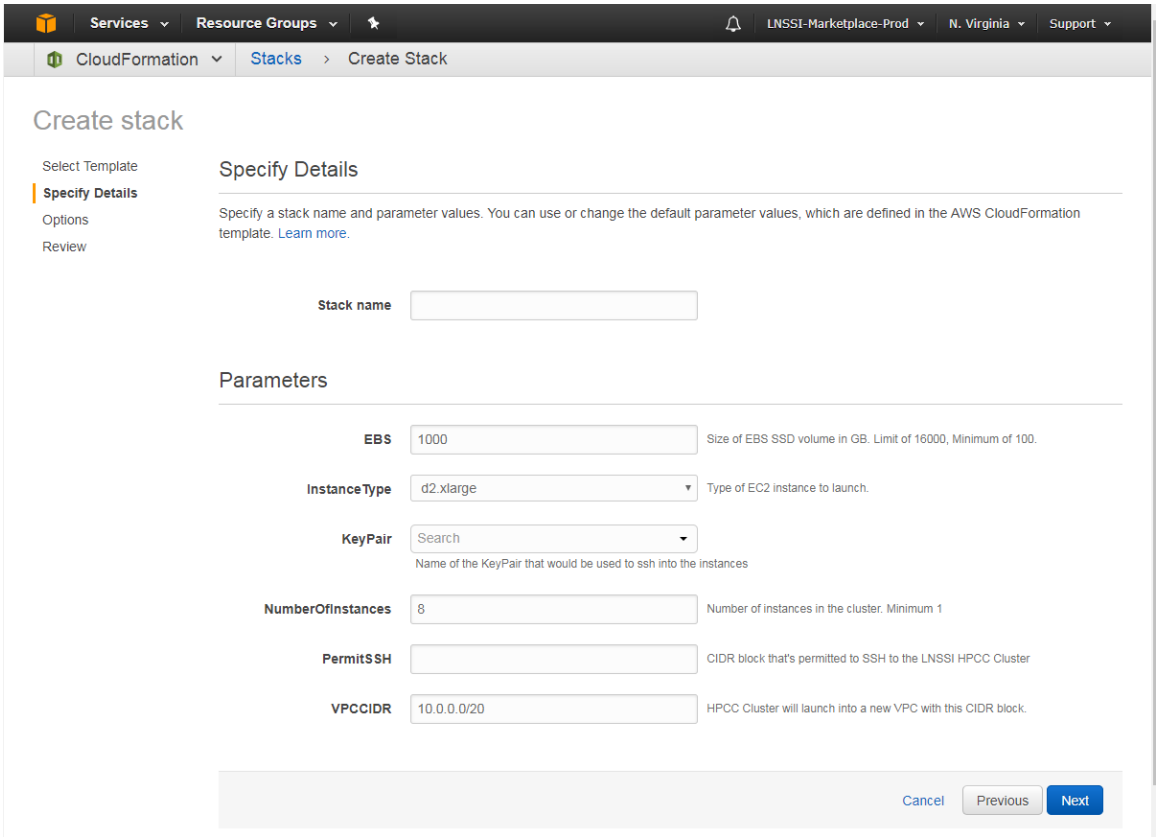
A Tag is required to be entered on the second page. This Name and Value will be used to identify the Security Group that will be created with the instances.

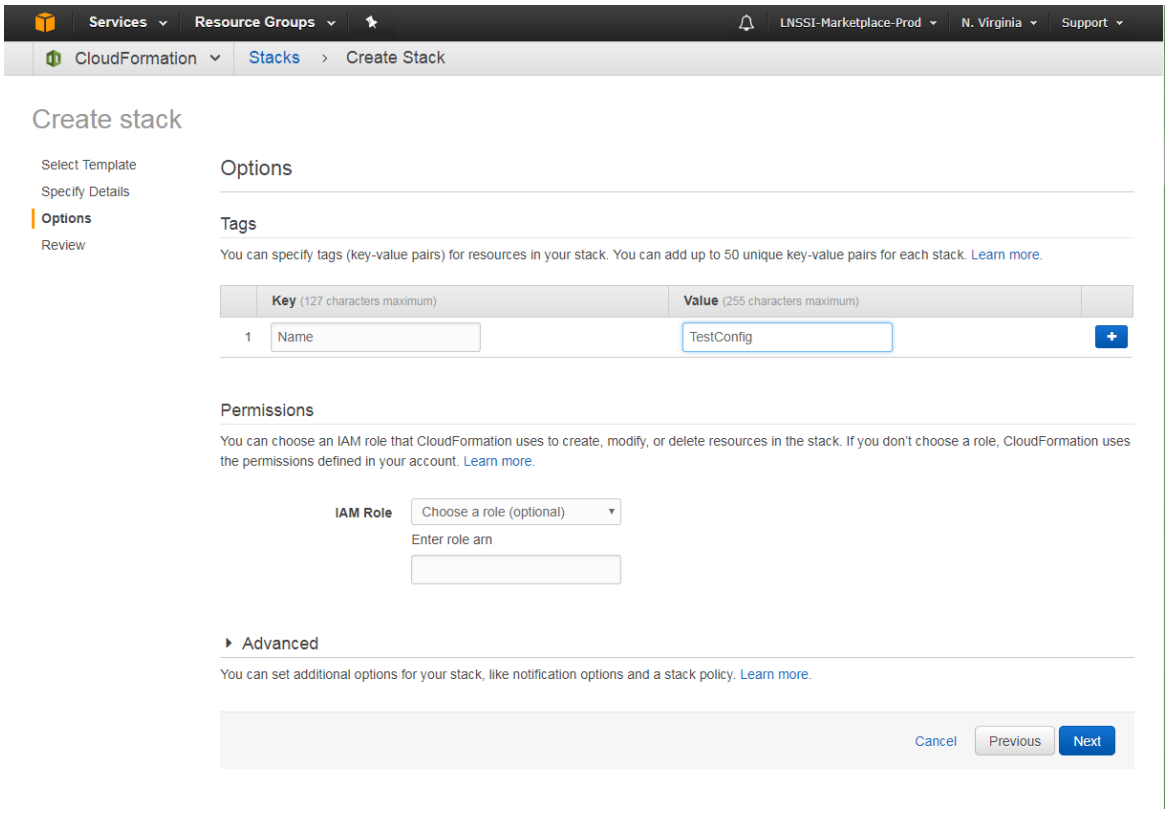
In some situations you may need to fill in an additional field, "ClusterInstanceProfile". The value required will be an Instance Profile ARN with the following format:

arn:aws:iam::<Your 12 digit AWS Account Number>:instance-profile/<A role you have created>

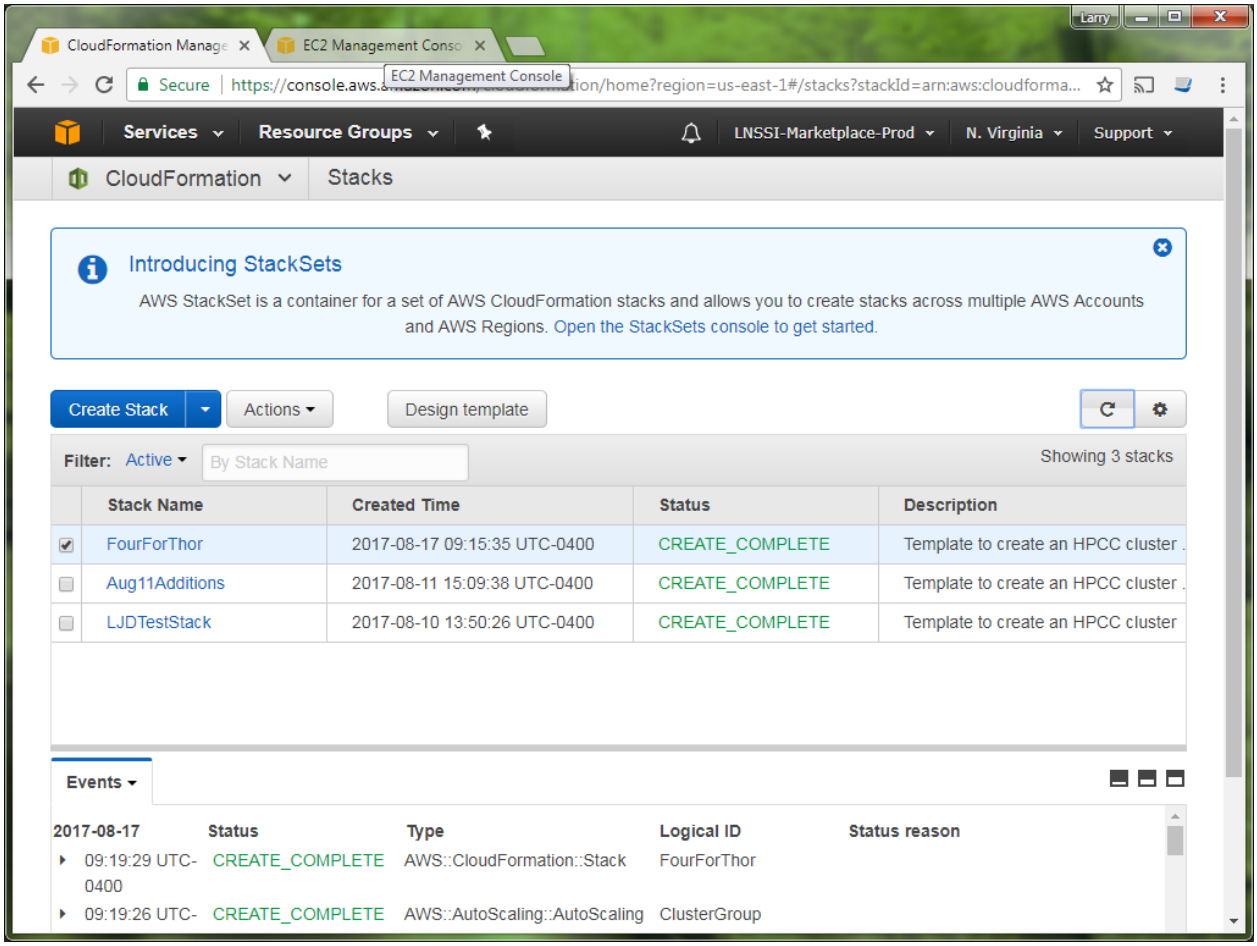
For example:

arn:aws:iam::123456789012:instance-profile/IC_CFT_HPCC_Role





It is recommended that you wait until the Status displays "CREATE_COMPLETE" before attempting any of additional steps. See the following example:



The following video demonstrates the steps just described:

http://lexisnexisspecialservices.com/Content/Video/Cldf_ProvisionInstances.mp4

Copy Your Private Key to an Instance

Copy a private key, .pem file, to one of the instances, after it has been launched.

In the video example the private key file, MarketplaceTestKeyPar.pem, will be referenced and was created earlier using PuTTYGen. The "MarketPlaceNode" is a PuTTY "Saved Session" that was created and saved earlier using the information outlined in the **"Amazon Elastic Compute Cloud User Guide for Linux Instances"**, documentation, http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-ug.pdf . In a later step the private key file is used to distribute new **SSH** keys to all of the other instances using a script that has been provided.

http://lexisnexisspecialservices.com/Content/Video/CopyPemFile.mp4

Other utilities such as WinSCP can be used to accomplish the same thing.

Install Keys on Instances

SSH to the same instance that the private key was copied to. All of the remaining steps will be performed from this instance or from a browser. Make a note of which instance this in case you decide to add more instances to this cluster in the future. Use the **ips_mapping.csv** file found on the instance

to help track which HPCC services are assigned to the different instances and to record any notes that you might have.

In this step a script must be run that will generate new key pairs for HPCC and install those key pairs to the instances listed in the **private_ips.txt** file. The new key pair is used by HPCC to allow communication between the instances in the cluster. Example:

sudo ./install.sh private_ips.txt [*Your private key*].pem

Note, it is recommended that the private key file is deleted once this script has finished executing.

Configure Instances and Start the HPCC Platform

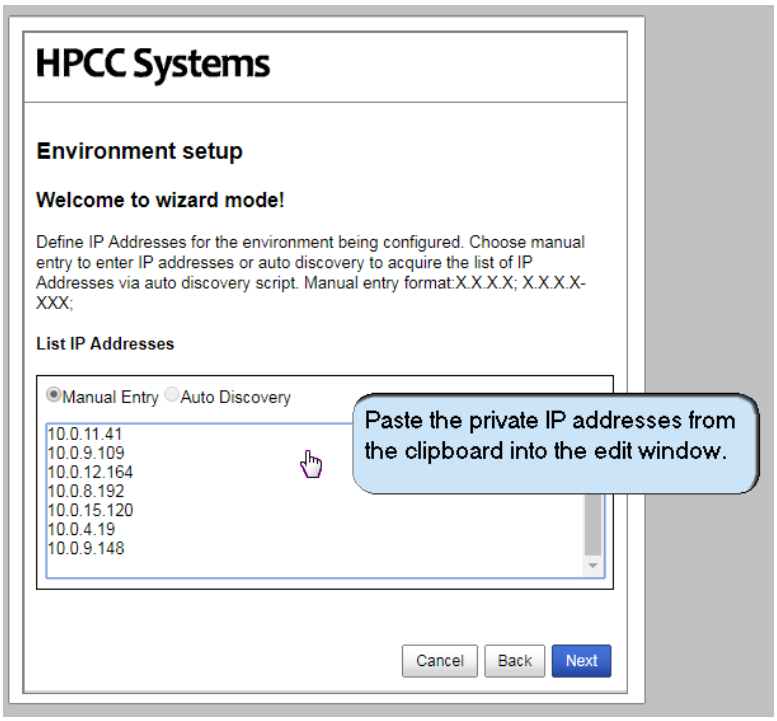
There are a number of steps that must now be performed in order to start the HPCC platform and to ensure all of the required processes are running. All of the steps outlined in this section are captured on video.

1. Start the HPCC Configuration Manager. Use the following command to start the HPCC Configuration Manager:

   sudo /opt/HPCCSystems/sbin/configmgr

2. Use a browser to access the running Configuration Manager utility, **//http:<public IP of the Configuration Manager instance>:8015**

3. Add the private IP addresses of all of the instances to the Configuration Manager by copying the contents of the **private_ips.txt** file into the edit window.



4. Define a configuration file, environment.xml, where your modifications will be stored.

5. Set parameters for the "ESP – myesp" service:

a. HTTPS passphrase – “asdf4321” is the required passphrase for this release of the product.

b. ESP Service Binding protocol from HTTP to HTTPS in two places.

6. If you include a Roxie cluster in your configuration you will need to set the UDP “roxieMulticastEnabled” parameter to “false”. This parameter can be found under the UDP tab within the “Roxie Cluster – myroxie” parameters.

7. Take note of the private IP address of the ECL Watch service identified within the Configuration Manager. You will use this value to find the associated public IP address in the ips_mapping.csv file that was provided with the instance. Using the public IP address you will be able to access the ECL Watch utility so that you can examine the status of various HPCC services, the distributed file system, and ECL workunits.

8. Save the configuration file before stopping Configuration Manager.

9. Within the PuTTY session, stop the Configuration Manager using Ctrl-C.

10. Make a backup copy of the original configuration file, **/etc/HPCCSystems/environment.xml**.

    sudo cp /etc/HPCCSystems/environment.xml /etc/HPCCSystems/environment.xml.bak

11. Copy the new configuration file, **/etc/HPCCSystems/source/environment.xml**, which was just generated to the **/etc/HPCCSystems** working directory.

    sudo cp /etc/HPCCSystems/source/environment.xml /etc/HPCCSystems/environment.xml

12. Use the **hpcc-push.sh** script to distribute the **environment.xml** file the other instances that will be part of the HPCC Platform. This will be necessary before the platform is started.

    sudo /opt/HPCCSystems/sbin/hpcc-push.sh –s /etc/HPCCSystems/environment.xml –t /etc/HPCCSystems/environment.xml

13. Start the platform using the **hpcc-run.sh** script:

    sudo /opt/HPCCSystems/sbin/hpcc-run.sh –a hpcc-init start

14. Use ECL Watch to verify the status of the HPCC services. **https://<Public IP of ESP instance>:18010** .The necessary steps to verify the status of the services is also documented in the **HPCCCertification-6.2.0-1.pdf** that is provided with the AMI.

The following videos demonstrate the steps just outlined:

http://lexisnexisspecialservices.com/Content/Video/Cldf_InstallAndCfgMgr.mp4

http://lexisnexisspecialservices.com/Content/Video/Cldf_CfgMgrSetUp.mp4

http://lexisnexisspecialservices.com/Content/Video/Cldf_StartHPCC.mp4

Plugin and Playground Tests

The AMI includes a number of plugins:

- Java
- Python
- MySQL
- JavaScript V8
- R

A script has been provided that the user can run that will test the plugins once the HPCC platform is operational.

The following video shows how to test the installed plugins, along with a simple test that can be run within the ECL Watch Playground.

http://lexisnexisspecialservices.com/Content/Video/Cldf_PluginTest.mp4

Setting Up the ECL IDE

A directory on the AMI, **/home/centos/clienttools**, contains the Windows install programs for the HPCC command line client tools, and the HPCC ECL IDE. You will need to copy the ECL IDE install program to your Windows workstation and run it before trying to duplicate the steps in the video. Information regarding the documentation for command line client tools can be found at the end of this document.

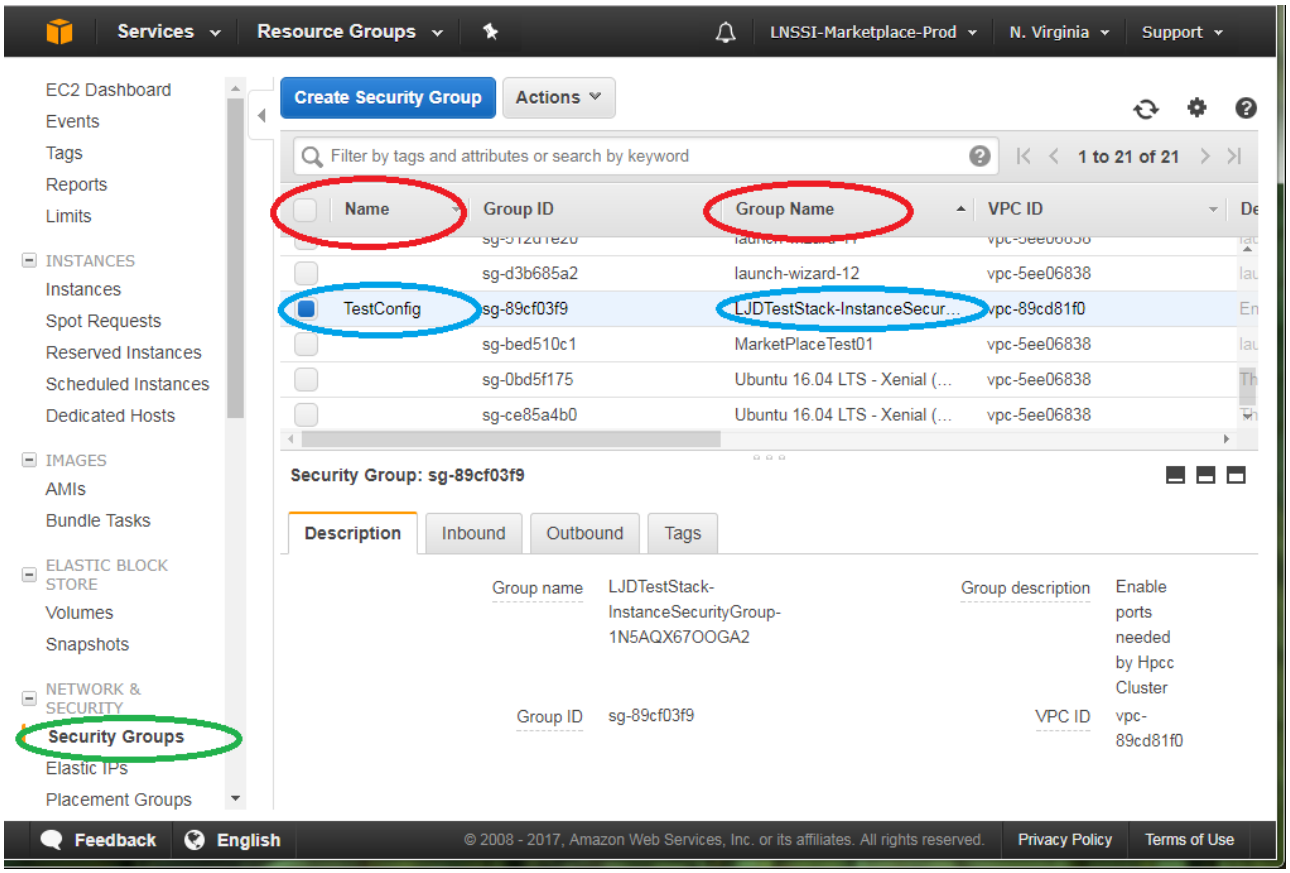
The ECL IDE is the development environment used by software engineers to write ECL code to help standardize data, develop complex algorithms, and develop Roxie queries. The following video will help walk you through the necessary steps that will allow you to use the IDE.

http://lexisnexisspecialservices.com/Content/Video/Cldf_SetUpIDEandBuildData.mp4

Manual Launch - "Existing VPC Cluster" Set Up and Configuration

Provisioning

Before adding new instances to an existing HPCC cluster you will need to know the Name and Group Name of the Security Group associated with the existing HPCC cluster that you wish to add instances to. See the following as an example:



You will need this information to fill out the "SecurityGroup", "VPC", and "VPCSubnet" fields of the AWS CloudFormation form. All other fields are similar to those described earlier the **"Manual Launch - "New VPC Cluster" Set Up and Configuration"** section of this document. You can choose the number and type of instances you now wish to add to the existing cluster. Refer to the following video for more details:

http://lexisnexisspecialservices.com/Content/Video/Cldf_AddToExistingVPC.mp4

Set Up and Configuration

It is recommended that the user be familiar with the **"HPCC Configuration Manager"** documentation before modifying their existing environment.xml file. This document does not cover all of the details and features associated with the Configuration Manager service.

The following outlines the necessary steps to add the new instances to an existing HPCC cluster:

1. **SSH** into the instance where you first ran the Configuration Manager utility. Refer back to the **"Manual Launch - "New VPC Cluster" Set Up and Configuration"** section of this document and what was performed earlier, if necessary. This instance could also actually be any of the current instances that are part of the existing HPCC configuration, but not be one of the new instances that were recently created since it will not have the most current environment.xml file.

2. Stop all of the HPCC services:

   ```
   sudo /opt/HPCCSystems/sbin/hpcc-run.sh -a hpcc-init stopall
   ```

3. Check the status of the services to ensure that they have all stopped:

   ```
   sudo /opt/HPCCSystems/sbin/hpcc-run.sh -a hpcc-init status
   ```

4. Make a backup copy of the **private_ips.txt** file on this instance.

5. Using a text editor you will need to add the IP addresses from one of the **private_ips.txt** file from one of the new instances to the **private_ips.txt** file on this instance. When this is complete the **private_ips.txt** file on this instance should have all of the IP addresses for all of the instances in the cluster.

6. Copy a private key, .pem file to this instance. See the **"Copy Your Private Key to an Instance"** section that appears earlier in this document.

7. Execute the script that will generate new key pairs for HPCC configuration and install those new key pairs to the instances listed in the **private_ips.txt** file. The new key pair is used by HPCC to allow communication between the instances in the cluster.

   sudo ./install.sh private_ips.txt [*Your private key*].pem

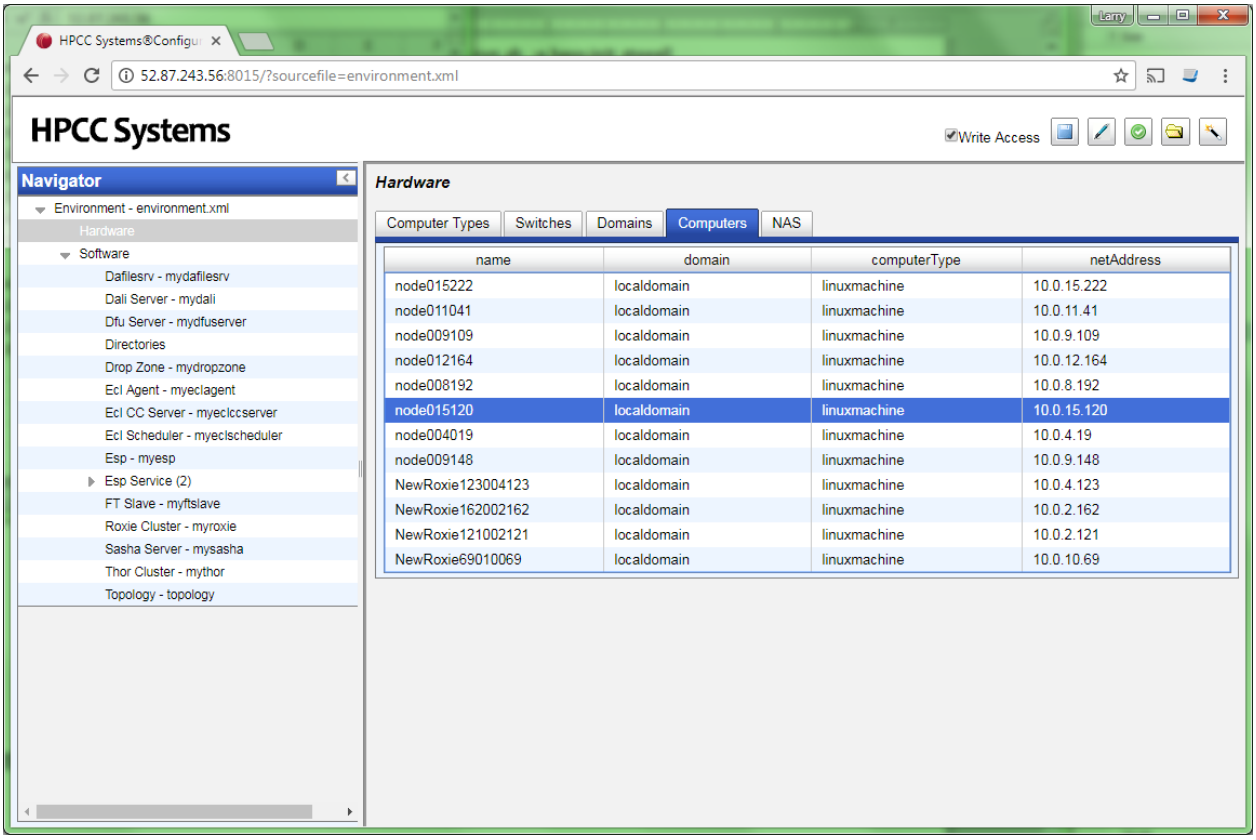
8. Start the Configuration Manager service:

   sudo /opt/HPCCSystems/sbin/configmgr

9. From a browser, use the **"Advanced View"** option to open the environment.xml file.

10. Enable "Write Access"

11. Add the new instances to the Hardware section of Configuration Manager being sure to use the private IP addresses of the new instances. Example:



12. Using the Configuration Manager assign these new instances to HPCC services such as Thor or Roxie. Refer to the **"HPCC Configuration Manager"** documentation for more details.

13. Save the configuration.

14. Stop the Configuration Manager service by entering Ctrl-C in the **SSH** session.

15. Make a backup copy of the original configuration file, **/etc/HPCCSystems/environment.xml**.

    sudo cp /etc/HPCCSystems/environment.xml /etc/HPCCSystems/environment.xml.bak

16. Copy the updated configuration file, **/etc/HPCCSystems/source/environment.xml**, to the location where the HPCCSystems' scripts will reference it.

    sudo cp /etc/HPCCSystems/source/environment.xml /etc/HPCCSystems/environment.xml

17. Use the **hpcc-push.sh** script to distribute the **environment.xml** file to the other instances that will be part of the HPCC Platform. This will be necessary before the platform is started.

    sudo /opt/HPCCSystems/sbin/hpcc-push.sh –s /etc/HPCCSystems/environment.xml –t /etc/HPCCSystems/environment.xml

18. Start the platform using the **hpcc-run.sh** script:

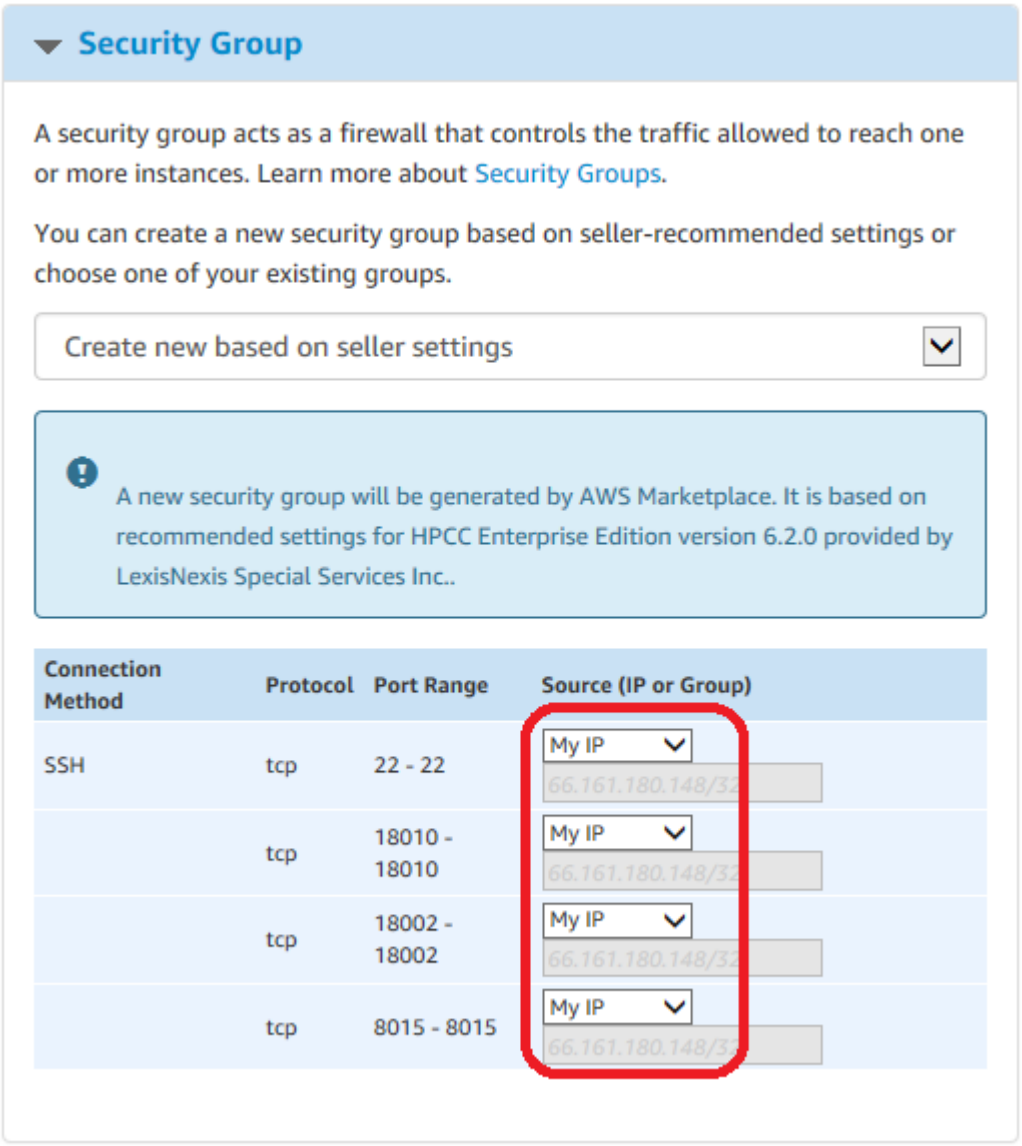
    sudo /opt/HPCCSystems/sbin/hpcc-run.sh –a hpcc-init start

19. Use ECL Watch to verify the status of the HPCC services. **https://<Public IP of ESP instance>:18010** .The necessary steps to verify the status of the services is also documented in the **HPCCCertification-6.2.0-1.pdf** that is provided with the AMI.

"1-Click Launch" Set Up and Configuration

Provisioning

When provisioning a single instance using 1-Click the user will need to set the network rules in the Security Group accordingly by setting the **"Source (IP or Group)"** to "My IP". For example:



The following video demonstrates this process:

http://lexisnexisspecialservices.com/Content/Video/Cldf_OneClick.mp4

Set Up

A single instance configuration requires fewer steps than what is required for a multi-instance configuration. There are no new steps so becoming familiar with the steps associated with a multi-instance configuration will be helpful when reviewing the following outline:

1. Copy your private key, .pem file, to your instance as it is required when executing the **install.sh** script.

2. Create a file on the instance that will contain the private IP address of the instance. This file is required input for the **install.sh** script. There are multiple ways to getting the private IP address, but the easiest way it via the EC2 Dashboard. Once you have logged into the instance via **SSH** you can use the "echo" command to create the file. For example:

   echo 172.31.24.32 > ips.txt

   where 172.31.24.32 is the private IP address of the instance.

3. Execute the **install.sh** script:

   sudo ./install.sh ips.txt [*Your private key*].pem

4. Start the HPCC Configuration Manager using the following command:

   sudo /opt/HPCCSystems/sbin/configmgr

5. At this point the user should be able to follow the steps outlined in the "Manual Launch - New VPC Cluster" section of this document related to configuring the HPCC services.

6. Save the original environment.xml file:

   sudo cp /etc/HPCCSystems/environment.xml /etc/HPCCSystems/environment.xml.bak

7. Copy the new environment.xml file to the HPCC working directory:

   sudo cp /etc/HPCCSystems/source/environment.xml /etc/HPCCSystems/environment.xml

8. Use the hpcc-push.sh script to the other instances that will be part of the HPCC Platform. This will be necessary before the platform is started.

   sudo /opt/HPCCSystems/sbin/hpcc-push.sh –s /etc/HPCCSystems/environment.xml –t /etc/HPCCSystems/environment.xml

9. Start the platform using the **hpcc-run.sh** script:

   sudo /opt/HPCCSystems/sbin/hpcc-run.sh –a hpcc-init start

10. Use ECL Watch to verify the status of the HPCC services. **https://<Public IP of ESP instance>:18010** .The necessary steps to verify the status of the services is also documented in the **HPCCCertification-6.2.0-1.pdf** that is provided with the AMI.

Appendix A: Instance Mapping Table

The following table provides some guidance as to the mapping between AWS instances and the three primary HPCC services: Thor, Roxie, and Admin.

AWS Instance Type	HPCC Service
m3.medium	All
m3.large	All
m3.xlarge	All
m3.2xlarge	All
r3.large	Roxie or Admin
r3.xlarge	Roxie or Admin
r3.2xlarge	Roxie or Admin
r3.4xlarge	Roxie or Admin
r3.8xlarge	Roxie or Admin
d2.xlarge	Thor
d2.2xlarge	Thor
d2.4xlarge	Thor
d2.8xlarge	Thor
m4.large	All
m4.xlarge	All
m4.2xlarge	All
m4.4xlarge	All
m4.10xlarge	All
m4.16xlarge	All
x1.16xlarge	All
x1.32xlarge	All
r4.large	Roxie or Admin
r4.xlarge	Roxie or Admin
r4.2xlarge	Roxie or Admin
r4.4xlarge	Roxie or Admin
r4.8xlarge	Roxie or Admin
r4.16xlarge	Roxie or Admin

Appendix B: Additional Documentation

HPCC Configuration Manger - http://lexisnexisspecialservices.com/wp-content/uploads/2017/03/UsingConfigManager-6.2.0-1.pdf

HPCC Client Tools – http://lexisnexisspecialservices.com/wp-content/uploads/2017/03/HPCCClientTools-6.2.0-1.pdf

ECL IDE and HPCC Client Tools – http://lexisnexisspecialservices.com/wp-content/uploads/2017/03/TheECLIDEandHPCCClientTools-6.2.0-1.pdf

Installing and Running the HPCC Platform - http://lexisnexisspecialservices.com/wp-content/uploads/2017/03/Installing_and_RunningTheHPCCPlatform-6.2.0-1.pdf

Complete HPCC Preflight and Certification - http://lexisnexisspecialservices.com/wp-content/uploads/2017/03/HPCCCertification-6.2.0-1.pdf

If you wish to know more about the HPCC platform please visit http://www.HPCCSystems.com.