HORTONWORKS HDP WITH NEUTRINO CONFIGURATION AND BEST PRACTICES GUIDE

ABSTRACT

This white paper describes the configuration and best practices for using Hortonworks HDP with EMC Neutrino.

June, 2016

REDEFINE

EMC WHITE

EMC²

To learn more about how EMC products, services, and solutions can help solve your business and IT challenges, contact your local representative or authorized reseller, visit www.emc.com, or explore and compare products in the EMC Store

Part Number HXXXXX <required, see Part numbers below for more info>

TABLE OF CONTENTS

PRODUCT VERSIONS

This document applies to following product versions.

Table 1 PRODUCT VERSIONS

PRODUCT	VERSION
Hortonworks HDP	2.4.2.0-258
Ambari	2.2.2.0
Hortonworks Cloudbreak	1.2.2
CentOS	7.2
Neutrino	1.0.0.0

INSTALLATION OPTIONS

The following installation options are available.

1. Hortonworks Cloudbreak - This can be used to completely automate the installation of Hortonworks HDP on an OpenStack environment such as Neutrino.

2. Manual OS provisioning using the OpenStack Horizon UI.

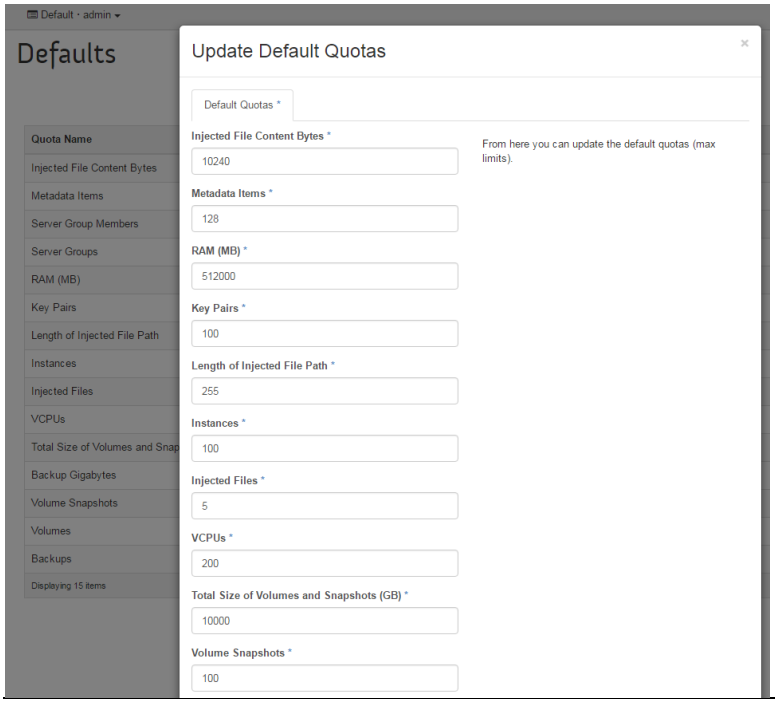
INSTALLATION USING HORTONWORKS CLOUDBREAK

CONFIGURE NEUTRINO

This document assumes that the Neutrino system has already been installed. The Cloud Compute (OpenStack) services must already be deployed on the desired number of nodes.

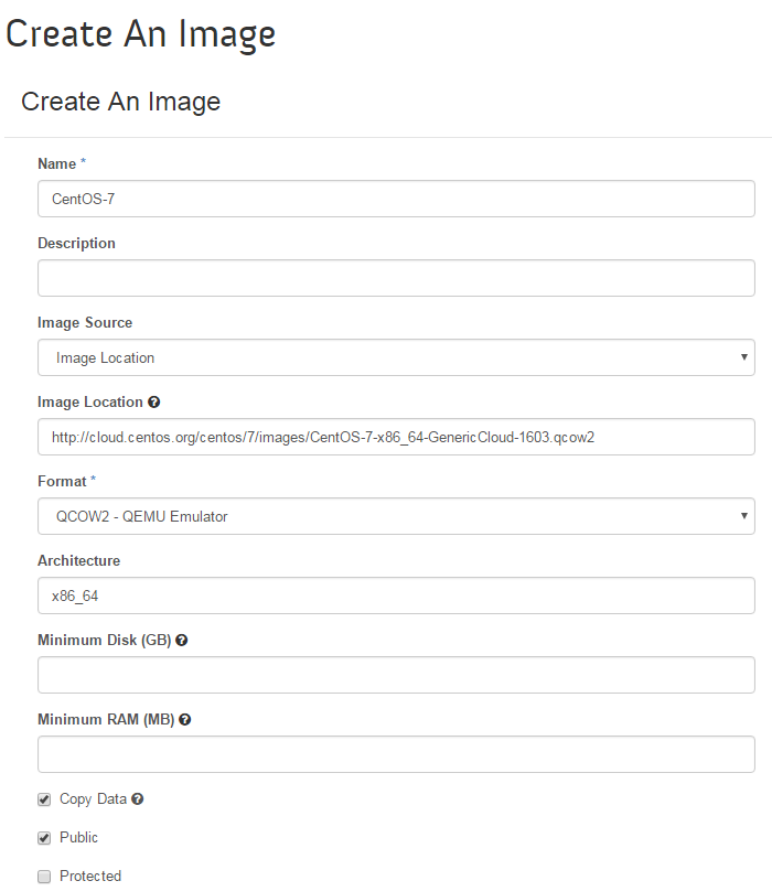
When deploying a Hadoop cluster in OpenStack, you may find that the default set of quotas are too small. To change the defaults, open the Horizon UI and navigate to System -> Defaults.

In particular, increase the RAM, Instances, VCPUs, and Total Size of Volumes and Snapshots as shown in the screen shot below.



You will eventually need to deploy virtual machines with CentOS. To add images, follow these steps:

1. In the OpenStack Horizon UI, navigate to System -> Images -> Create Image.
2. Assign a name and use one of the following URLs, depending on the desired CentOS version.
   http://cloud.centos.org/centos/7/images/CentOS-7-x86_64-GenericCloud-1603.qcow2
   http://cloud.centos.org/centos/6.5/images/CentOS-6-x86_64-GenericCloud-20140929_01.qcow2
3. Click Create Image.



DEPLOY HORTONWORKS CLOUDBREAK

Follow the Cloudbreak documentation at http://sequenceiq.com/cloudbreak-docs/. Be aware of the suggestions below.

When installing Cloudbreak Deployer, use option #1 since the OpenStack image for option #2 it is not available.

If Neutrino is installed with a self-signed SSL certificate, then follow these steps to configure Cloudbreak to trust the SSL certificate.

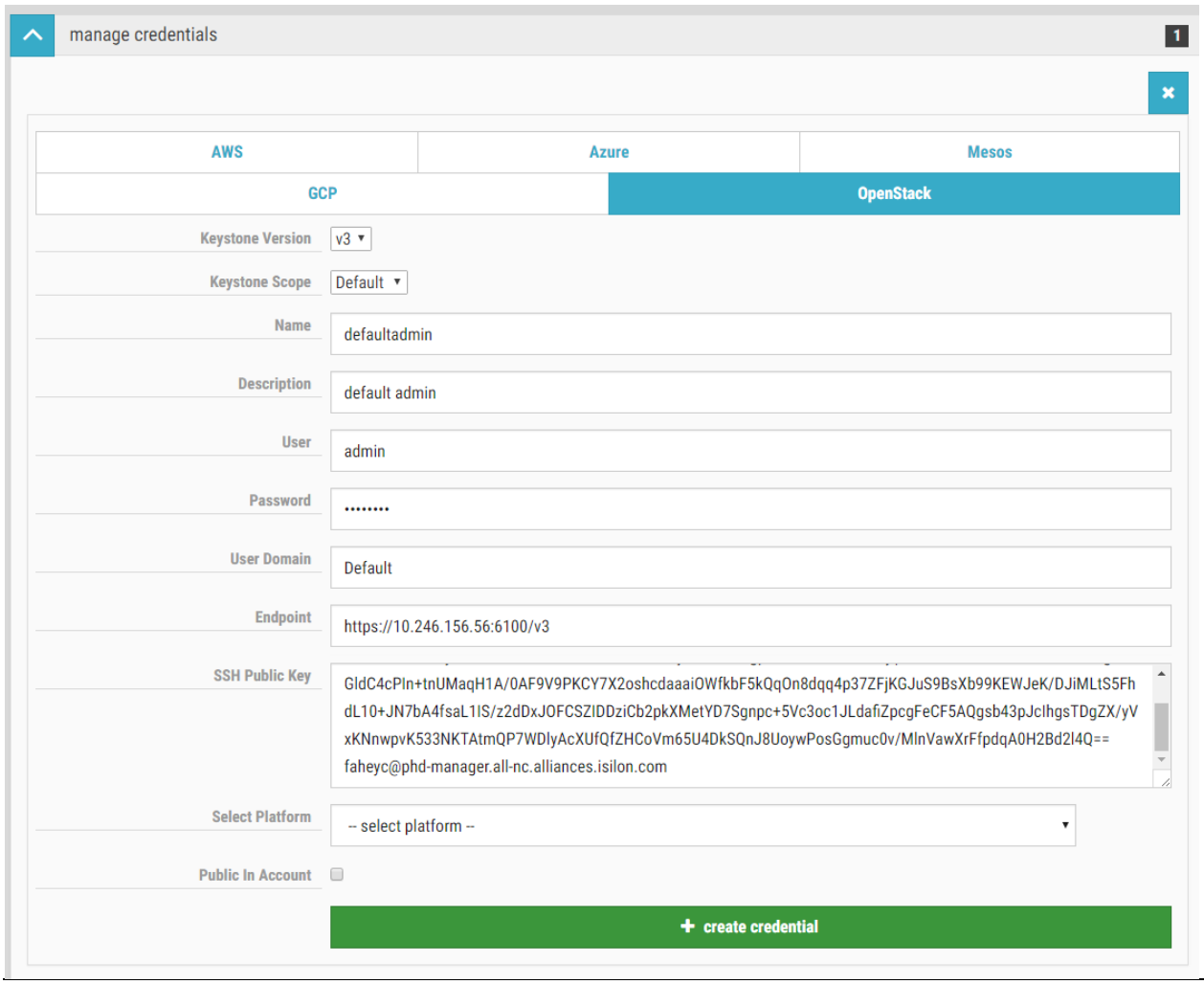
1. SSH as root to the Cloudbreak Deployer host.
2. View the SSL certificate. Replace *neutrino* with the FQDN or IP address of your Neutrino system.
   [root@cloudbreak-deployer ~]# echo -n | openssl s_client -showcerts -connect *neutrino*:6100
3. Install the certificate as a trusted certificate in Cloudbreak:
   [root@cloudbreak-deployer ~]# docker exec -it cbreak_cloudbreak_1 bash
   root@4b9559b3bdee:/# cat - > /usr/local/share/ca-certificates/neutrino.crt
   -----BEGIN CERTIFICATE-----
   MIIDBjCCAe6gAwIBAgIJAIRhrq4eoPu4MA0GCSqGSIb3DQEBCwUAMDgxEDAOBgNV
   …

-----END CERTIFICATE-----
root@4b9559b3bdee:/# update-ca-certificates

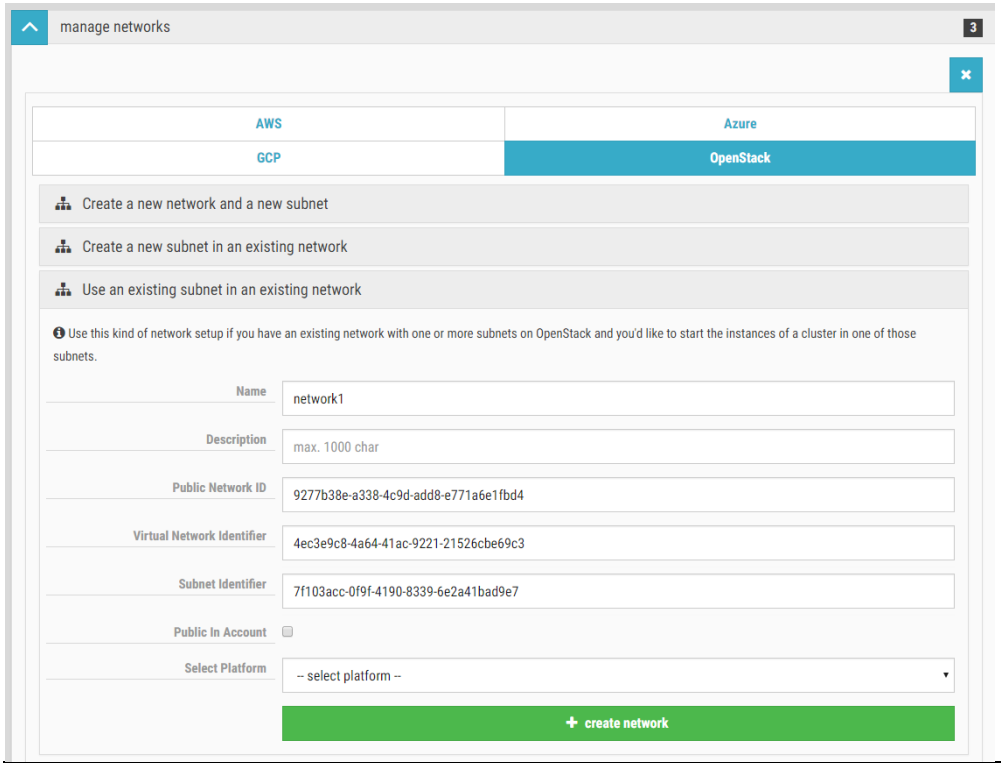
4. Restart Cloudbreak:
   [root@cloudbreak-deployer cloudbreak-deployment]# docker restart cbreak_cloudbreak_1

5. Test the certificate:
   curl https://*neutrino*:6100/

When defining credentials for Cloudbreak to use to connect to OpenStack, the default admin credential works reliably without authorization errors. The diagram below shows the settings that are known to work.



When defining networks for Cloudbreak, the following is known to work. Refer to the following table in order to determine the correct values of the IDs.



CLOUDBREAK FIELD	SOURCE
Public Network ID	Networks -> public -> ID
Virtual Network Identifier	Networks -> default_network -> ID
Subnet Identifier	Networks -> default_network -> Subnets -> (…) -> ID

In Neutrino version 1.0, you must create VMs without swap files. Attempting to create VMs with swap files will result in failures during VM creation. Create flavors in OpenStack without swap files and create corresponding templates in Cloudbreak.

DEPLOYING A HADOOP CLUSTER USING CLOUDBREAK

Follow the Cloudbreak documentation at http://sequenceiq.com/cloudbreak-docs/. Be aware of the suggestions below.

After a Hadoop cluster has been deployed, you may wish to safely reduce the HDFS replication setting (dfs.replication) from 3 to 2. With a value of 2, HDFS will maintain 2 replicas of each HDFS block and ScaleIO will maintain 2 copies of each underlying disk block, for a total of 4 copies of each logical file byte.

INSTALLATION USING OPENSTACK HORIZON

CONFIGURE NEUTRINO

Refer to the corresponding section in **Installation Using Hortonworks Cloudbreak** for configuring Neutrino.

PROVISION VIRTUAL MACHINES

1. Using the OpenStack Horizon UI, create VMs for each of the master and worker nodes as desired. For Neutrino 1.0, you must ensure that the flavor uses zero swap.

2. In your DNS infrastructure, create a DNS zone named "openstacklocal" and a corresponding reverse zone for the internal IP addresses.

3. Create DNS A and PTR records for the internal IP addresses. Ensure that forward and reverse DNS works properly.

4. For each worker, create a single large volume and attach it to the instance. These volumes will be used by HDFS and temporary local storage.

5. Partition, format, and mount the attached volume as /grid/0 on each virtual machine. To help automate this task on many nodes, see https://github.com/claudiofahey/devops-scripts.

INSTALL HORTONWORKS HDP USING AMBARI

Refer to the latest Ambari installation documentation available at http://docs.hortonworks.com/.

After a Hadoop cluster has been deployed, you may wish to safely reduce the HDFS replication setting (dfs.replication) from 3 to 2. With a value of 2, HDFS will maintain 2 replicas of each HDFS block and ScaleIO will maintain 2 copies of each underlying disk block, for a total of 4 copies of each logical file byte.

TESTING THE HADOOP INSTALLATION

To ensure that all required Hadoop components are functional, it is recommended that the functional and performance tests as documented in the related validation brief (**EMC Technical Validation of Hortonworks HDP with Neutrino**) be performed.