

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

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](#)

Copyright © 2016 EMC Corporation. All Rights Reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

The information in this publication is provided "as is." EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

VMware and <insert other VMware marks in alphabetical order; remove sentence if no VMware marks needed. Remove highlight and brackets> are registered trademarks or trademarks of VMware, Inc. in the United States and/or other jurisdictions. All other trademarks used herein are the property of their respective owners.

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

TABLE OF CONTENTS

PRODUCT VERSIONS4

INSTALLATION OPTIONS.....4

INSTALLATION USING HORTONWORKS CLOUDBREAK4

 CONFIGURE NEUTRINO..... 4

 DEPLOY HORTONWORKS CLOUDBREAK 5

 DEPLOYING A HADOOP CLUSTER USING CLOUDBREAK..... 7

INSTALLATION USING OPENSTACK HORIZON7

 CONFIGURE NEUTRINO..... 7

 PROVISION VIRTUAL MACHINES 7

 INSTALL HORTONWORKS HDP USING AMBARI 8

TESTING THE HADOOP INSTALLATION8

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.

The screenshot shows the 'Update Default Quotas' form in the OpenStack Horizon UI. On the left is a sidebar with a 'Defaults' section and a list of quota categories: Injected File Content Bytes, Metadata Items, Server Group Members, Server Groups, RAM (MB), Key Pairs, Length of Injected File Path, Instances, Injected Files, VCPUs, Total Size of Volumes and Snapshots, Backup Gigabytes, Volume Snapshots, Volumes, Backups, and a note 'Displaying 15 items'. The main area contains a form with the following fields and values:

- Default Quotas *
- Injected File Content Bytes *: 10240
- Metadata Items *: 128
- RAM (MB) *: 512000
- Key Pairs *: 100
- Length of Injected File Path *: 255
- Instances *: 100
- Injected Files *: 5
- VCPUs *: 200
- Total Size of Volumes and Snapshots (GB) *: 10000
- Volume Snapshots *: 100

A note on the right side of the form states: 'From here you can update the default quotas (max limits).'

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.

Create An Image

Create An Image

Name *	<input type="text" value="CentOS-7"/>
Description	<input type="text"/>
Image Source	<input type="text" value="Image Location"/>
Image Location ?	<input type="text" value="http://cloud.centos.org/centos/7/images/CentOS-7-x86_64-GenericCloud-1603.qcow2"/>
Format *	<input type="text" value="QCOW2 - QEMU Emulator"/>
Architecture	<input type="text" value="x86_64"/>
Minimum Disk (GB) ?	<input type="text"/>
Minimum RAM (MB) ?	<input type="text"/>
<input checked="" type="checkbox"/> Copy Data ?	
<input checked="" type="checkbox"/> Public	
<input type="checkbox"/> Protected	

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.

manage credentials

1

X

AWS Azure Mesos

GCP OpenStack

Keystone Version v3

Keystone Scope Default

Name defaultadmin

Description default admin

User admin

Password

User Domain Default

Endpoint https://10.246.156.56:6100/v3

SSH Public Key GldC4cPln+tnUMaqH1A/0AF9V9PKCY7X2oshcdaaiOWfkbF5kQqOn8dqq4p37ZFjKGJuS9BsXb99KEWJeK/DJIMLtS5Fh
dL10+JN7bA4fsaL1IS/z2dDxJOFCsZIDDziCb2pkXMetYD7Sgnpc+5Vc3oc1JLdafiZpcgFeCF5AQgsb43pJclhgsTDgZX/yV
xKNnwpvK533NKTAtmQP7WDlyAcXufQZHCvM65U4DkSQnJ8UoywPosGgmuc0v/MlnVawXrFpdqA0H2Bd2l4Q==
faheyc@phd-manager.all-nc.alliances.isilon.com

Select Platform -- select platform --

Public In Account ☐

+ create credential

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.

manage networks
3

AWS

Azure

GCP

OpenStack

Create a new network and a new subnet

Create a new subnet in an existing network

Use an existing subnet in an existing network

Use this kind of network setup if you have an existing network with one or more subnets on OpenStack and you'd like to start the instances of a cluster in one of those subnets.

Name

network1

Description

max. 1000 char

Public Network ID

9277b38e-a338-4c9d-add8-e771a6e1fbd4

Virtual Network Identifier

4ec3e9c8-4a64-41ac-9221-21526cbe69c3

Subnet Identifier

7f103acc-0f9f-4190-8339-6e2a41bad9e7

Public in Account

☐

Select Platform

-- select platform --

+ create network

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

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