Enabling the Hortonworks Data Platform on Red Hat Storage

An installation and configuration guide for on-premise deployment

Version 0.92
June 2013

Table of Contents

I can't update the TOC in Libre office... it needs to be updated.

1 Executive Summary

1.1 Introduction

Apache Hadoop is an open source project from the Apache Software Foundation that provides a distributed computing framework for processing large datasets across clusters of computers.

Hadoop provides a pluggable FileSystem architecture which supports a Hadoop FileSystem plugin for a given File System. Hadoop ships with a plugin and File System known as the Hadoop Distributed File System (HDFS). Red Hat Storage (RHS) is also a Distributed File System providing a Hadoop File System plugin that enables it for use with Hadoop.

Red Hat Storage (RHS) provides a multi-datacenter, petabyte scale, POSIX compliant distributed file system with no single point of failure. RHS features “in-place data” meaning that data stored in RHS does not first need to be extracted to another set of storage nodes in order to run Hadoop analytics. This allows a single storage cluster to serve as both a general purpose file system, and as a specialized distributed, scalable, highly available analytics cluster. RHS also provides the capability to NFS mount data volumes and to expose the data through both a file system and a SWIFT block store interface. For more information on Red Hat Storage, please see http://www.redhat.com/promo/liberate/

1.2 Objectives

Hortonworks is a Red Hat Partner that provides a Hadoop distribution known as the Hortonworks Data Platform (HDP). This document describes how to configure Red Hat Storage for use with HDP.

2 The Hortonworks Data Platform

Hortonworks Data Platform (HDP) is a 100% open source data management platform for Apache Hadoop. Built and packaged by the core architects, builders and operators of Hadoop, HDP includes all of the necessary components to manage a cluster at scale and uncover business insights from existing and new big data sources.

Unlike many Hadoop distributions, HDP includes a powerful management console known as Ambari. Ambari assists with the deployment, monitoring, and management of Hadoop clusters. Through the open source community, including several Red Hat team members, Ambari has been enhanced to support Hadoop Compatible File Systems (HCFS) of which RHS is an example.

Hortonworks and Red Hat are committed to contributing all code, patches & bug fixes back to Apache Hadoop. HDP is 100% complete, open source and free software available to the community without proprietary license. Our goal is to ensure your success with Hadoop and to provide stable, reliable, well-tested software as part of our commitment to you and the community.

The figure below displays how the HDP platform interacts with the Red Hat Storage software.

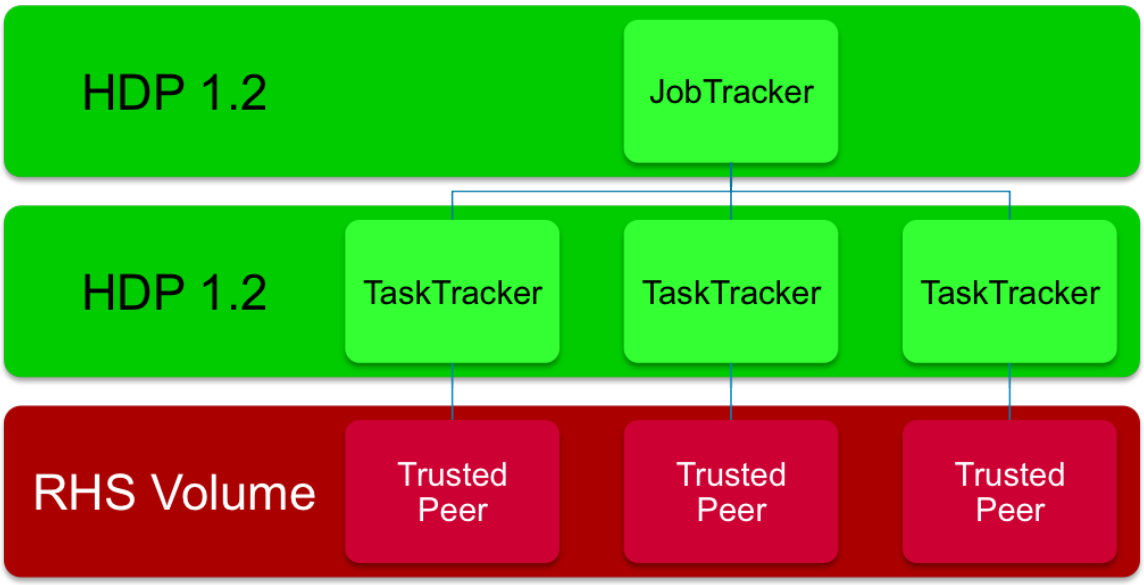

Figure 1 – Distribution of Red Hat and Partner Software

3 Versions

This solution has been tested and validated with Red Hat Storage (RHS) 2.0.5 and HDP 1.2.3.1.

RHS 2.0.5 can be obtained from: https://rhn.redhat.com/rhn/software/channel/downloads/Download.do?cid=14689. Click the "ISO" link to download. If we decide to use RHS 2.0.4, which is not customer downloadable then the SA will need to get it from here: http://download.devel.redhat.com/devel/candidate-trees/RHS-2.0-u4/2.0/RHS/x86_64/iso/

HDP 1.2.2 can be obtained from: http://hortonworks.com/thankyou-hdp122/

However, NOTE: HDP 1.2 is downloaded to each storage node by the Ambari install wizard, so there is no need to do this download separately.

The current RHS-Hadoop plugin, version 0.0.5, supports the HDFS APIs invoked by Hadoop jobs such as MapReduce. It is bundled in the RHS tarball and is also available here: http://23.23.239.119/archiva/browse/org.apache.hadoop.fs.glusterfs/glusterfs-hadoop/0.0.5

4 Pre-Deployment / Pre-Requisites

This section contains information about deployment considerations you may want to evaluate before you actually start installing the software.

Hardware Recommendations

In order to properly make use of all the available resources on a server it is important to have a balanced server design to avoid resource bottlenecks. For performance, servers should have at least 12 data disks to increase the distribution of I/O requests across more than one disk. For storage capacity, it is more cost effective to use large form factor disks for your data disks. A 10GbE network reduces the likelihood of your Hadoop applications becoming network bound. The amount of work a server can do in Hadoop is determined by how many Map and Reduce slots it has available to it. Faster processors with lots of cores coupled with a lot of memory allow one to configure a lot of slots per server. However, if performance at a commodity price point is your goal, we recommend the configuration below:

Component	Configuration
CPU	2 socket 2.4 GHz + 6 core processors
Memory	48GB -64GB RAM
Disks	2 x 250 GB Drives for Mirrored OS, RHS and HDP runtimes

	12 x 2 or 3TB SATA Drives for Data in RAID 6 Configuration 1 x RAID Controller
Network	2 x 10GbE NICs

Disk Subsystem configurations

Steve: Do we need to explain and add information about mapred.local.dir requiring an XFS LVM on each data disk. Explain about RHS requiring RAID 6 Configuration. Explain about how to do a horizontal stripe to cater to both. Explain that section 2.1 covers how to manage this if your I/O system and RHS is already in use and can't be changed.

5 Deployment Scenarios

Keep in mind that, prior to installing RHS, you may want to configure your I/O subsystem to implement the horizontal striping recommendation in Section 4.

Section 5 is divided into the two most common deployment scenarios, namely:

- Section 5.1: Red Hat Storage is already installed
- Section 5.2: Red Hat Storage is not already installed.

In both cases Ambari is the management tool that handles the deployment, configuration and monitoring of Hadoop jobs. You need to decide whether you want:

1) dedicated servers for the JobTracker and for Ambari that are not part of the RHS trusted storage pool.
2) To collocate the JobTracker and Ambari on the same server, and this server is not part of the RHS trusted storage pool.
3) To collocate the JobTracker and Ambari on the same server, and this server is one of the data nodes (task tracker nodes) in the RHS trusted storage pool.

Our recommendation is option 1. This decision affects which node you choose to install Ambari on when following Section 5.x.

After RHS has been installed, or updated, follow the Ambari instructions below.

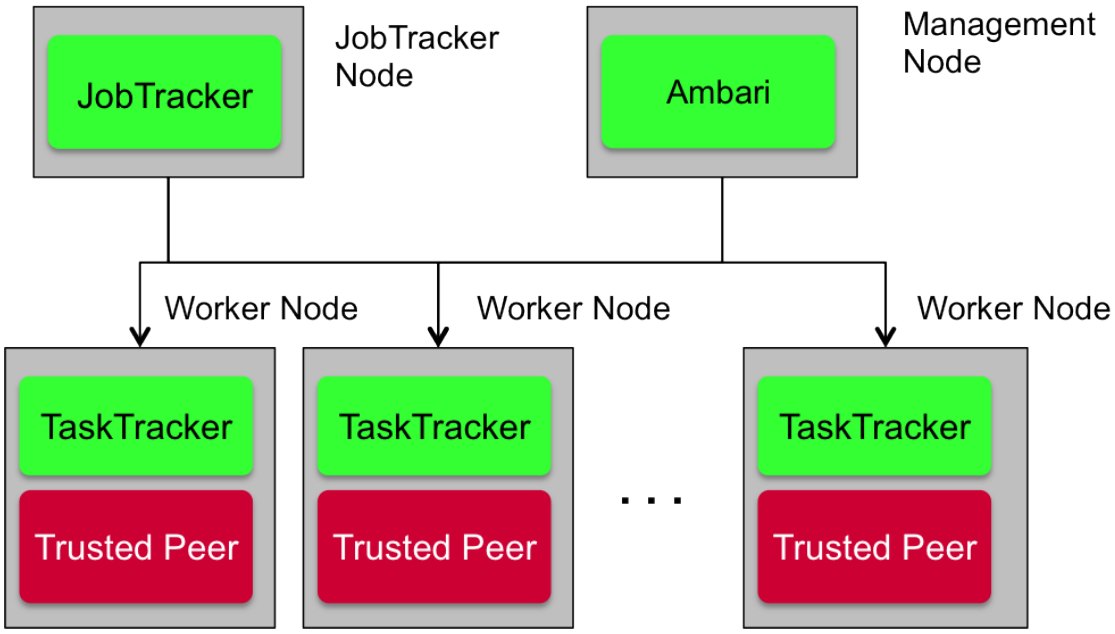

Figure 2 – Deployment Topology

5.1 Red Hat Storage is already installed

(Please skip to section 5.2 since we are not fully ready for "existing RHS storage" big data installations.)
The assumption is that you have a Distributed Replicated 2 RHS Volume running RHS 2.1 update 4. If you don't have that you will need to install the update using the following procedure:

Questions for us: 1) where does an EA customer get RHS 2.0.4 if they're on an earlier version? From our custom ISO, which has (hopefully) the FUSE patches. 2) if EA customer is already on 2.0.4 how do they get the FUSE patches? From our custom ISO? Do we provide instructions for installing the

Install HDP on the same machines/nodes as your existing RHS cluster. HDP 1.2.3.1 can be obtained from http://hortonworks.com/download/

The enhanced Ambari management console is provided by Red Hat via the custom RHS 2.0.4-XX ISO. Extract just the Ambari RPMs to /tmp:
- *mkdir /mnt/iso*
- *mkdir /tmp/ambari*
- *mount -o loop* <pathTo/>*RHS-2.0.4-XX.iso /mnt/iso*
- *cp /mnt/iso/ambari-agent-XX.rpm /tmp/ambari*
- *cp /mnt/iso/ambari-server-XX.rpm /tmp/ambari*

Use the following Ambari instructions to install the JobTracker and TaskTracker processes across your cluster.

Steve?: Add Ambari instructions for MapReduce(only? What about Pig and Hive, etc.) Deployment and HCFS (RHS) configuration. Make sure to reference and apply previous section where the user had to make a decision about where to deploy Ambari and the JobTracker. Carefully cover that the TaskTrackers must be deployed on the RHS Trusted peers for the intended volume.

Now proceed to section 5.3 to configure Red Hat Storage

5.2 Red Hat Storage is not already installed

The tarball provided as part of the early adopter program contains:

* the “install.sh” shell script which is the main script driving the configuration of RHS for Hadoop workloads,

* a README.txt file which should be read first and contains an overview of the install process, assumptions, and brief instructions,

* the “hosts.example” text file which should be read and contains more information about the “hosts” file that the person doing the RHS-HDP installation needs to create,

* the data/ sub-directory which contains the files to be copied to each storage node. These files include the following:

- the current version of the Gluster/Hadoop plugin JAR file which enables the Hadoop APIs on top of RHS,
- an enhanced version of Apache Ambari which supports RHS. Note: using a Red Hat specific version of Ambari may not be needed in future RHS-HDP installations,
- the RHEL 6.2 kernel FUSE patch RPMs which are required and automatically installed by the scripts,
- the companion “prep_node.sh” install script which is invoked by “install.sh”.

This tarball can be download to any directory on any of the storage nodes or to a separate server/laptop with password-less ssh access to each storage node. The “install-from” server must be set up for password-less ssh to all storage nodes as the root user.

5.2.1 RHS 2.0 Installation

Ensure that RHS 2.0.5 has been installed correctly on each data node. The RHS install guide can be found here: https://access.redhat.com/site/documentation/Red_Hat_Storage/ . Scroll down to the “Red Hat Storage 2.0” section and click the PDF link for “2.0 Update 4 and Update 5 Release Notes”. The RHS 2.0.5 ISO is available here: https://rhn.redhat.com/rhn/software/channel/downloads/Download.do?cid=14689. Click the “ISO” link and download the file. The RHS 2.0.4 ISO is only available from a Red Hat SA. SA note, see: http://download.devel.redhat.com/devel/candidate-trees/RHS-2.0-u4/2.0/RHS/x86_64/iso/.

RHS documentation, including release notes and installation instructions is available here: https://access.redhat.com/site/documentation/Red_Hat_Storage/. Scroll down to “Red Hat Storage 2.0” where

administration and install documents can be found in various formats. The “Red Hat Storage 2.0 Installation Guide” describes the prerequisites and provides step-by-instructions to install Red Hat Storage. It is available in HTML (https://access.redhat.com/site/documentation/en-US/Red_Hat_Storage/2.0/html/Installation_Guide/index.html) and is also available as a PDF.

5.2.2 RHS Install Changes for Hadoop Environments

RHS 2.0 Installation Guide at:
https://access.redhat.com/site/documentation/en-US/Red_Hat_Storage/2.0/html/Installation_Guide/index.html
Exceptions:

* 4.1.2 – set up static IP addresses, not DHCP.

* 4.1.4 – chose a “custom layout” to create a dedicated storage partition. You want to have already set up a RAID 6 device for this partition.

RHS 2.0 Administration Guide at:
https://access.redhat.com/site/documentation/en-US/Red_Hat_Storage/2.0/html/Installation_Guide/index.html
Exceptions:

* skip all of section 7 -- don't create a trusted storage pool.

* skip all of section 8 – don't create volumes.

* the RHS-HDP installation configures volumes for optimal performance with Hadoop workloads.

* the rest of the above guide can be read but not acted upon.

5.2.3 Instructions for Installing RHS-HDP

1) a “data” partition for cluster storage should be created on each storage node as a RAID 6 device. This device file is the only required parameter to the install.sh script

2) extract the contents of the supplied rhs-hdp-install tarball and cd to the rhs-hdp-install-<version> directory

3) read the README.txt file and then create the required “hosts” file either in the same directory that install.sh will be run from, or if not, then be sure to specify the “–-hosts <path>” option when running install.sh

4) execute install.sh supplying the above RAID 6 storage device path and any additional options. install.sh –-help summarizes the available options

5) the install.sh output is displayed to the terminal window and written to the /var/log/RHS-install.log log file on both the “install-from” node and on each storage node. There is no “quiet” option yet

6) install the Ambari server RPM on the designated Ambari server node.

7) from the Ambari server begin the Ambari install process: http://<ambari-server-host>:8080

NOTE: install.sh installs the Ambari agent RPM on each data node; therefore, when running the Ambari install wizard be sure to select “manual install of agent”. However, the agent is not started; therefore, the agent must be started on each server node after the Ambari install wizard completes, using the *ambari-agent start* command.

5.4 Ambari Installation

The steps below consist of the standard Hortonworks instructions located at http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/index.html augmented by RHS specific instructions or explanations.

It is assumed that RHS 2.0.5 has been installed on each data node in the cluster. The management node ("admin" node) does not require RHS. Any RHEL 6.2+ release will work for the management node. If RHS has not been installed then follow the RHS Installation Guide before doing the next steps.

The reader should open the above URL, and after reading each Hortonworks Data Platform (HDP) installation step, switch back here to see what steps may differ, if any, when using Red Hat Storage (RHS). No actions should be taken until after the RHS specific instructions have been read and understood. If there is not a corresponding RHS section then the existing HDP documentation is sufficient. The entire HDP installation guide must be read. Each HDP install step URL will be listed first, followed by RHS specific installation notes or comments, if applicable.

1. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ambari-chap1-1.html -- "Understand the Basics":
RHS 2.0.5 provides a Hadoop Compatible File System (HCFS) in place of using the Hadoop Distributed File System (HDFS). Some of the services shown on this page may not be fully implemented for early adopter customers, but MapReduce, Ganglia, and Nagios are available. Other services will be added as needed on a case-by-case basis.

2. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ambari-chap1-2.html -- "Meet Minimum System Requirements":
  - 2.2. Note: RHS 2.0.5 is built on top of 64 bit RHEL 6.2.

3. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ambari-chap1-3.html -- “Decide on Deployment Type”:
Early adopter customers are expected to deploy RHS-HDP on bare metal (not VMs). The Red Hat Systems Architect (SA) has already helped the customer size their cluster.

4. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ambari-chap1-4.html -- “Collect Information”:
RHS does not use a NameNode or secondary NameNode. For the early adopter/POC program the only services initially supported are MapReduce, Nagios, and Ganglia. Hive, HCatalog, Zookeeper, HBase, Pig, Sqoop, and Oozie are not yet fully tested and may not work correctly if selected.

5. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ambari-chap1-5-1.html -- “Check Existing Installs”:
Use the "RHEL/CentOS v6" column.

6. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ambari-chap1-5-2.html -- “Set Up Password-less SSH”:
Passwordless SSH is required from the Amabari Management server to each storage node; however, it is not required for RHS Hadoop workloads.

7. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ambari-chap1-6.html -- “Optional: Configure the Local Repositories”:
It is important to not yum install from a repo that gets Ambari bits from hortonworks.com in order to not remove the Red Hat specific changes to the Ambari Management console. The Red Hat enhanced version of Ambari will be installed by the install.sh script.

8. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ambari-chap2-1.html -- “Set Up the Bits”:
Use the "RHEL and CentOS 6" row. Note that the HDP RPM downloaded here also contains Ambari, but this version of Ambari is being replaced by the version provided in the RHS-HDP install.

9. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ambari-chap2-4.html -- “Optional: Set Up LDAP or Active Directory Authentication”:
Skip this step.

10. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ambari-chap2-2a.html -- “Optional: Change the Ambari Server Port”:
Skip this step.

11. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ambari-chap2-3.html -- “Start the Ambari Server”:
Note: It's not mentioned above but the *ambar-server reset* command erases the cluster you've defined and lets you start over.

12. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ambari-chap3-3.html -- “Install Options”:
Currently, even though all of the host names are defined in the local “hosts” file, these hosts need to be repeated in the text box in this screen. We hope to reduce redundant steps in future versions of the RHS-HDP install. install.sh will copy the Ambari agent RPMs to each storage node; therefore, “perform manual registration” should be selected.

13. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ch03s05.html -- “Choose Services”:

Select HCFS, for Hadoop Compatible File System, of which RHS is an example. For now, all services other than Nagios and Ganglia should not be selected.

14. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ambari-chap3-5.html -- “Assign Masters”:
With HCFS selected in the previous screen, there will be no Namenode or Secondary Namenode hosts to assign, due to the fact that Hadoop on RHS does not need these nodes.

15. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ambari-chap3-6.html -- “Assign Slaves and Clients”:
Select the JobTracker node to be separate from the storage cluster or it can be one of the data nodes.

16. http://incubator.apache.org/ambari/1.2.3/installing-hadoop-using-ambari/content/ambari-chap3-7.html-- “Customize Services”:
This is where RHS-specific configuration values are set. See section 5.6 below for details.

5.5 Special Considerations

RHS for Hadoop has a temporary requirement that all jobs must be run as the same user and we recommend "mapred" as this user. This means that the user id for all Hadoop jobs needs to be change to “mapred” (or whatever has been chosen as the common user). In addition, the steps below must be followed in order to allow the mapred user access to file locality information tracked by RHS.

- Follow the steps below on the **JobTracker** node:
1) set the fs.glusterfs.automount property in the core-site.xml file to false,
2) set the fs.glusterfs.getfattrcmd property in core-site.xml to "sudo getfattr -m . -n trusted.glusterfs.pathinfo" (no

quotes),

Note: /mnt/glusterfs needs 0774 or 0770 permissions and should be owned by the "hadoop" group (or whichever group your tracker + map/reduce jobs run under). E.g.,

chmod 770 /mnt/glusterfs

chown mapred:hadoop /mnt/glusterfs

Note: above assumes the RHS mount directory is /mnt/glusterfs and the RHS volume name is HadoopVol, which are the recommended names.

4) create a file called /etc/sudoers.d/gluster with permissions 0440 and add the following line to it:

mapred ALL= NOPASSWD: /usr/bin/getfattr

5) make sure the fs.glusterfs.mount property value matches the mount point in step 3) above.

5.6 Confirm core-site configuration

- core-site.xml settings (**per node**):

Below are the necessary core-site.xml file settings for using RHS for Hadoop workloads. Some of these settings are done by Ambari but others need to be done manually. All settings should be verified:

PropertyName --> Property Value
---
fs.glusterfs.impl --> org.apache.hadoop.fs.glusterfs.GlusterFileSystem
fs.default.name --> glusterfs://<nodeHostname>:9000
fs.glusterfs.server --> <nodeHostname> note: fs.glusterfs.server and fs.default.name should be set to the hostname of the machine where the core-site.xml file resides and have the same node name value
fs.glusterfs.volname --> HadoopVol (suggested volume name)
fs.glusterfs.mount --> /mnt/glusterfs (suggested mount name)
fs.glusterfs.automount --> false

fs.glusterfs.getfattrcmd --> sudo getfattr -m . -n
trusted.glusterfs.pathinfo

5.7 Additional Information

Additional information on how to deploy and configure Red Hat Storage is available at the following link: https://access.redhat.com/products/Red_Hat_Storage/

6 Configuring Red Hat Storage for use with Hadoop

Configure the settings described in this section on all nodes in the cluster. These additional steps may be automated in future releases of the installation package.

1) t*uned-adm profile rhs_high_throughput*

Verify that the tuned script worked and that the I/O scheduler was successfully set to use the deadline scheduler and that the RAID 6 device read ahead buffer was successfully set to 64kb.

Examples below:

 # cat /sys/block/sdc/queue/scheduler
noop anticipatory [deadline] cfq

 # cat /sys/block/dm-2/queue/read_ahead_kb
65536

2) Set the read ahead buffer to 4MB for the JBOD Disks. Need to be done for the Hadoop Disks:

echo 4096 > /sys/block/sdb/queue/read_ahead_kb