hadoop-bigtop Roll: Users Guide

1.5.0 Edition

hadoop-bigtop Roll: Users Guide :

1.5.0 Edition

Published May 27 2022

Table of Contents

List of Tables

Preface

The hadoop-bigtop Roll installs and configures hadoop.

- Cluster Planning: To understand the default value of some attributes, this roll roll assumes the following minimal partition scheme by default:

```
/        (reformatted at kickstart)
/var     (reformatted at kickstart)
/scratch (a partition to host local dirs to be shared in cluster)
/hadoop  (a partition to be used by hadoop)
```

  To change the default partition/dir used by hadoop, set changes in rocks attrs BIGTOP_HadoopDataDir and/or BIGTOP_HadoopData. For example

```
###global
#/opt/rocks/bin/rocks set attr BIGTOP_HadoopDataDir value="/bigdata"
### or host specific
#/opt/rocks/bin/rocks set host attr compute-0-0 BIGTOP_HadoopDataDir value="/bigdata"
#/opt/rocks/bin/rocks set host attr compute-0-0 BIGTOP_HadoopData value="/bigdata/mydata"
```

Please visit the hadoop-bigtop site[1] to learn more about their release and the individual software components.

## Notes

1. http://bigtop.apache.org

Chapter 1. Installing the hadoop-bigtop Roll

1.1. Adding the Roll

The hadoop-bigtop roll can be installed during the Frontend installation step of your cluster (see On a New Server) or added to a running frontend (see Install on Running System).

1.2. On a New Server

The hadoop-bigtop roll is added to a Frontend installation in exactly the same manner as the required HPC Roll. Specifically, after the HPC Roll is added the installer will once again ask if you have a Roll (see below). Select 'Yes' and insert the hadoop-bigtop roll.
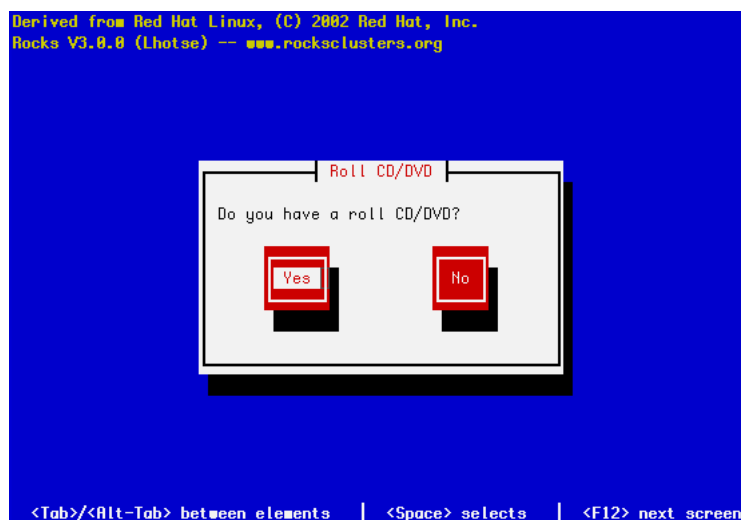
During the installation process, there will be a screen related to the hadoop-bigtop roll. As shown below.

To choose the default hadoop primary and secondary servers.

If not sure what to choose, leave it as is, these default can be changed later any time.

1.3. Install on Running System

The hadoop-bigtop roll can be installed on a running system. The following assumes that roll is available in .iso form and is called "hadoop-bigtop.iso".

```
$ su - root
# rocks add roll hadoop-bigtop.iso
# rocks enable roll hadoop-bigtop
# cd /export/rocks/install
# rocks create distro
# rocks run roll hadoop-bigtop | bash
# rocks add attr BIGTOP_HadoopNameNode value=compute-0-0
# rocks add attr BIGTOP_HadoopSecondaryNameNode value=compute-0-1
# rocks add attr BIGTOP_HadoopDataDir value="file:///hadoop"
# rocks add attr BIGTOP_HadoopData value="file:///hadoop/data"
# rocks add attr BIGTOP_HadoopCheckPointDirs value="file:///hadoop"
# rocks add attr BIGTOP_HadoopNamenodeNameDir value="file:///hadoop/dfs/name"
# rocks add attr BIGTOP_HadoopCheckPointPeriod value=600
# rocks add attr BIGTOP_HadoopUpdateFstab value=1
# rocks add attr BIGTOP_HadoopReplicationDefault value=2
```

Chapter 2. Using the hadoop-bigtop Roll

2.1. Using the hadoop-bigtop Roll

Tutorials for installing hadoop-bigtop and its components are available in section 2.2 'Installing and configuring Hadoop' and configuration examples in section 2.3 'Hadoop Configuration Examples'

2.2. Installing and configuring Hadoop

To install hadoop in a node, first is needed to set rocks attr BIGTOP_HADOOP to true, then run the command shown below (or reinstall).

After installation run the rocks command to configure it. Below is an example for setting a datanode.

```
$ su - root
##
## Set flag to hadoop installation (once)
##
#  rocks set host attr compute-0-2 BIGTOP_HADOOP True
##
## Run rocks commands to install and configure hadoop
##
#  rocks sync host bigtop hadoop install compute-0-2
#  rocks sync host bigtop hadoop config compute-0-2
##
## Set services on boot
##
#  rocks run host compute-0-2 command='systemctl enable hadoop-hdfs-datanode'
##
## Start hadoop service
##
#  rocks run host compute-0-2 command='systemctl start hadoop-hdfs-datanode'
```

Example of installation by forcing node reinstall

```
$ su - root
##
## Set flags to hadoop installation (not needed if already set)
##
#  rocks set host attr compute-0-2 BIGTOP_HADOOP True
#  rocks set host attr compute-0-2 BIGTOP_Hadoop_StartDaemonDataNode True
##
## Reinstall
##
#  rocks set host boot compute-0-2 action=install
#  ssh compute-0-2 "shutdown -r now"
```

2.3. Hadoop Configuration Examples

The following are short examples of how to customize Hadoop using Rocks commands.

- Change default Hadoop Node Name on all compute Appliances:

```
rocks set appliance attr compute BIGTOP_HadoopNameNode value=hadoop-0-0
rocks sync host bigtop hadoop config compute
```

- Change default Hadoop Secondary Name on all compute Appliances:

```
rocks set appliance attr compute BIGTOP_HadoopSecondaryNameNode value=hadoop-0-1
rocks sync host bigtop hadoop config compute
```

- Change default Hadoop Data Dir on all compute Appliances, for example two data disks on compute nodes:

```
rocks set appliance attr compute BIGTOP_HadoopData
        value="file:///hadoop/data,file:///hadoop2/data"
rocks sync host bigtop hadoop config compute
```

2.4. Customizing Hadoop

By default Hadoop is NOT installed on Rocks clusters. To install on Rocks **compute** appliances set attr
BIGTOP_HADOOP.

```
##
## turn on Bigtop Hadoop installation for compute nodes at kickstart.
##
#/opt/rocks/bin/rocks add appliance attr compute BIGTOP_HADOOP value=true
```

Table 2-1. Attributes Used in hadoop bigtop Roll for Hadoop customization

Attibute Name	Description
BIGTOP_HadoopNameNode	Configure Hadoop NodeName server name used on any particular Appliance or Host installation. Check for a line like this: HADOOP_NAMENODE=compute-0-0 in /etc/sysconfig/hadoop Default: compute-0-0
BIGTOP_HadoopSecondaryNameNode	Configure Hadoop SecondaryName server name used on any particular Appliance or Host installation. Check for a line like this: HADOOP_SECONDARY_NAMENODE=compute-0-1 in /etc/sysconfig/hadoop Default: compute-0-1
BIGTOP_HadoopDataDir	Configure Hadoop base data dir used on any particular Appliance or Host installation. Check for a line like this: HADOOP_DATADIR=file:///hadoop in /etc/sysconfig/hadoop Default: file:///hadoop

Attibute Name	Description
BIGTOP_HadoopData	Configure Hadoop data dir used on any particular Appliance or Host installation. Check for a line like this: HADOOP_DATA=file:///hadoop/data in /etc/sysconfig/hadoop Default: file:///hadoop/data
BIGTOP_HadoopCheckPointDirs	Configure Hadoop check point dirs used on any particular Appliance or Host installation. Check for a line like this: HADOOP_CHECKPOINT_DIRS=file:///hadoop in /etc/sysconfig/hadoop Default: file:///hadoop
BIGTOP_HadoopNamenodeNameDir	Configure Hadoop Name node dirs used on any particular Appliance or Host installation. Check for a line like this: HADOOP_NAMENODE_NAME_DIR=file:///hadoop in /etc/sysconfig/hadoop Default: file:///hadoop
BIGTOP_HadoopCheckPointPeriod	Configure Hadoop check point period used on any particular Appliance or Host installation. Check for a line like this: HADOOP_CHECKPOINT_PERIOD=600 in /etc/sysconfig/hadoop Default: 600
BIGTOP_HadoopUpdateFstab	Configure Hadoop for update fstab used on any particular Appliance or Host installation. Check for a line like this: HADOOP_UPDATE_FSTAB=1 in /etc/sysconfig/hadoop Default: 1
BIGTOP_Hadoop_StartDaemonDataNode	Configure to start Hadoop DataNode daemon in a particular Appliance or Host after installation Default: no default
BIGTOP_Hadoop_StartDaemonNameNode	Configure to start Hadoop NameNode daemon in a particular Appliance or Host after installation Default: no default
BIGTOP_Hadoop_StartDaemonSecondaryNameNode	Configure to start Hadoop SecondaryNameNode daemon in a particular Appliance or Host after installation Default: no default

Chapter 3. Command Reference

3.1. report

3.1.1. report host bigtop hadoop config

rocks report host bigtop hadoop config {host} [ConfigFile=*string*]

Output the BIGTOP Hadoop Local Configuration Uses Rocks Attributes: BIGTOP_HadoopNameNode, BIGTOP_HadoopDataDir, BIGTOP_HadoopData, BIGTOP_HadoopSecondaryNameNode, BIGTOP_HadoopCheckPointDirs, BIGTOP_HadoopCheckPointPeriod, BIGTOP_HadoopReplicationDefault, BIGTOP_HadoopUpdateFstab, ganglia_address, Kickstart_PrivateDNSDomain

arguments

host

> One host name.

parameters

[ConfigFile=*string*]

> Defaults to: /etc/sysconfig/hadoop

examples

$ rocks report host bigtop hadoop config compute-0-0

> Create the BIGTOP Hadoop Configuration for compute-0-0

$ rocks report host bigtop hadoop config compute-0-0 ConfigFile=/etc/sysconfig/hadooptest

> Create the BIGTOP Hadoop Configuration for compute-0-0 on file /etc/sysconfig/hadooptest

3.1.2. report host bigtop hadoop install

rocks report host bigtop hadoop install {host}

Output the BIGTOP hadoop wrapper install script Uses Rocks Attributes:
BIGTOP_HADOOP,rocks_version_major,
BIGTOP_Hadoop_StartDaemonDataNode,BIGTOP_Hadoop_StartDaemonNameNode,
BIGTOP_Hadoop_StartDaemonSecondaryNameNode

arguments

host

> One host name.

examples

$ rocks report host bigtop hadoop install hadoop-0-0

> Create wrapper script to install BIGTOP hadoop for hadoop-0-0

3.2. sync

3.2.1. sync host bigtop hadoop config

rocks sync host bigtop hadoop config [test=*bool*]

Reconfigure BIGTOP Hadoop on the named hosts.

parameters

[test=*bool*]

> If want to test output set this parameter. Default is no.

examples

rocks sync host bigtop hadoop config compute-0-0

> Rewrite /etc/sysconfig/hadoop and /root/HadoopConfigurator then run the HadoopConfigurator on host compute-0-0

rocks sync host bigtop hadoop config test=yes compute-0-0

> Show the bash script that will run to rewrite /etc/sysconfig/hadoop on host compute-0-0

3.2.2. sync host bigtop hadoop install

rocks sync host bigtop hadoop install [test=*bool*]

Install hadoop on the named hosts.

parameters

[test=*bool*]

> If want to test output set this parameter. Default is no.

examples

rocks sync host bigtop hadoop install node-0-0

> add hdfs user, add hadoop group, and install hadoop rpms on host node-0-0 if attr BIGTOP_HADOOP is set on host.

rocks sync host bigtop hadoop install node-0-0 test=yes

Show the bash script that will run to install hadoop on host node-0-0

Chapter 4. Copyrights

4.1. Third Party Copyrights and Licenses

This product includes software developed by the apache software foundation.

The binary version of this product includes software developed by the apache software foundation under the apache license http://www.apache.org/licenses/LICENSE-2.0.txt.

The software contained in this distribution is released under the academic license agreement which requires to acknowledge the use of the software that results in any published work.

4.2. Apache v2.0

```
                        Apache License
                  Version 2.0, January 2004
                http://www.apache.org/licenses/

   TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION

1. Definitions.
    "License" shall mean the terms and conditions for use, reproduction,
    and distribution as defined by Sections 1 through 9 of this document.

    "Licensor" shall mean the copyright owner or entity authorized by
    the copyright owner that is granting the License.

    "Legal Entity" shall mean the union of the acting entity and all
    other entities that control, are controlled by, or are under common
    control with that entity. For the purposes of this definition,
    "control" means (i) the power, direct or indirect, to cause the
    direction or management of such entity, whether by contract or
    otherwise, or (ii) ownership of fifty percent (50%) or more of the
    outstanding shares, or (iii) beneficial ownership of such entity.

    "You" (or "Your") shall mean an individual or Legal Entity
    exercising permissions granted by this License.

    "Source" form shall mean the preferred form for making modifications,
    including but not limited to software source code, documentation
    source, and configuration files.

    "Object" form shall mean any form resulting from mechanical
    transformation or translation of a Source form, including but
    not limited to compiled object code, generated documentation,
    and conversions to other media types.

    "Work" shall mean the work of authorship, whether in Source or
    Object form, made available under the License, as indicated by a
    copyright notice that is included in or attached to the work
    (an example is provided in the Appendix below).
```

"Derivative Works" shall mean any work, whether in Source or Object
form, that is based on (or derived from) the Work and for which the
editorial revisions, annotations, elaborations, or other modifications
represent, as a whole, an original work of authorship. For the purposes
of this License, Derivative Works shall not include works that remain
separable from, or merely link (or bind by name) to the interfaces of,
the Work and Derivative Works thereof.

"Contribution" shall mean any work of authorship, including
the original version of the Work and any modifications or additions
to that Work or Derivative Works thereof, that is intentionally
submitted to Licensor for inclusion in the Work by the copyright owner
or by an individual or Legal Entity authorized to submit on behalf of
the copyright owner. For the purposes of this definition, "submitted"
means any form of electronic, verbal, or written communication sent
to the Licensor or its representatives, including but not limited to
communication on electronic mailing lists, source code control systems,
and issue tracking systems that are managed by, or on behalf of, the
Licensor for the purpose of discussing and improving the Work, but
excluding communication that is conspicuously marked or otherwise
designated in writing by the copyright owner as "Not a Contribution."

"Contributor" shall mean Licensor and any individual or Legal Entity
on behalf of whom a Contribution has been received by Licensor and
subsequently incorporated within the Work.

2. Grant of Copyright License. Subject to the terms and conditions of
   this License, each Contributor hereby grants to You a perpetual,
   worldwide, non-exclusive, no-charge, royalty-free, irrevocable
   copyright license to reproduce, prepare Derivative Works of,
   publicly display, publicly perform, sublicense, and distribute the
   Work and such Derivative Works in Source or Object form.

3. Grant of Patent License. Subject to the terms and conditions of
   this License, each Contributor hereby grants to You a perpetual,
   worldwide, non-exclusive, no-charge, royalty-free, irrevocable
   (except as stated in this section) patent license to make, have made,
   use, offer to sell, sell, import, and otherwise transfer the Work,
   where such license applies only to those patent claims licensable
   by such Contributor that are necessarily infringed by their
   Contribution(s) alone or by combination of their Contribution(s)
   with the Work to which such Contribution(s) was submitted. If You
   institute patent litigation against any entity (including a
   cross-claim or counterclaim in a lawsuit) alleging that the Work
   or a Contribution incorporated within the Work constitutes direct
   or contributory patent infringement, then any patent licenses
   granted to You under this License for that Work shall terminate
   as of the date such litigation is filed.

4. Redistribution. You may reproduce and distribute copies of the
   Work or Derivative Works thereof in any medium, with or without
   modifications, and in Source or Object form, provided that You
   meet the following conditions:

   (a) You must give any other recipients of the Work or
       Derivative Works a copy of this License; and

```
(b) You must cause any modified files to carry prominent notices
    stating that You changed the files; and

(c) You must retain, in the Source form of any Derivative Works
    that You distribute, all copyright, patent, trademark, and
    attribution notices from the Source form of the Work,
    excluding those notices that do not pertain to any part of
    the Derivative Works; and

(d) If the Work includes a "NOTICE" text file as part of its
    distribution, then any Derivative Works that You distribute must
    include a readable copy of the attribution notices contained
    within such NOTICE file, excluding those notices that do not
    pertain to any part of the Derivative Works, in at least one
    of the following places: within a NOTICE text file distributed
    as part of the Derivative Works; within the Source form or
    documentation, if provided along with the Derivative Works; or,
    within a display generated by the Derivative Works, if and
    wherever such third-party notices normally appear. The contents
    of the NOTICE file are for informational purposes only and
    do not modify the License. You may add Your own attribution
    notices within Derivative Works that You distribute, alongside
    or as an addendum to the NOTICE text from the Work, provided
    that such additional attribution notices cannot be construed
    as modifying the License.

You may add Your own copyright statement to Your modifications and
may provide additional or different license terms and conditions
for use, reproduction, or distribution of Your modifications, or
for any such Derivative Works as a whole, provided Your use,
reproduction, and distribution of the Work otherwise complies with
the conditions stated in this License.
```

5. Submission of Contributions. Unless You explicitly state otherwise, any Contribution intentionally submitted for inclusion in the Work by You to the Licensor shall be under the terms and conditions of this License, without any additional terms or conditions. Notwithstanding the above, nothing herein shall supersede or modify the terms of any separate license agreement you may have executed with Licensor regarding such Contributions.

6. Trademarks. This License does not grant permission to use the trade names, trademarks, service marks, or product names of the Licensor, except as required for reasonable and customary use in describing the origin of the Work and reproducing the content of the NOTICE file.

7. Disclaimer of Warranty. Unless required by applicable law or agreed to in writing, Licensor provides the Work (and each Contributor provides its Contributions) on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied, including, without limitation, any warranties or conditions of TITLE, NON-INFRINGEMENT, MERCHANTABILITY, or FITNESS FOR A PARTICULAR PURPOSE. You are solely responsible for determining the appropriateness of using or redistributing the Work and assume any risks associated with Your exercise of permissions under this License.

8. Limitation of Liability. In no event and under no legal theory,

whether in tort (including negligence), contract, or otherwise,
unless required by applicable law (such as deliberate and grossly
negligent acts) or agreed to in writing, shall any Contributor be
liable to You for damages, including any direct, indirect, special,
incidental, or consequential damages of any character arising as a
result of this License or out of the use or inability to use the
Work (including but not limited to damages for loss of goodwill,
work stoppage, computer failure or malfunction, or any and all
other commercial damages or losses), even if such Contributor
has been advised of the possibility of such damages.

9. Accepting Warranty or Additional Liability. While redistributing
   the Work or Derivative Works thereof, You may choose to offer,
   and charge a fee for, acceptance of support, warranty, indemnity,
   or other liability obligations and/or rights consistent with this
   License. However, in accepting such obligations, You may act only
   on Your own behalf and on Your sole responsibility, not on behalf
   of any other Contributor, and only if You agree to indemnify,
   defend, and hold each Contributor harmless for any liability
   incurred by, or claims asserted against, such Contributor by reason
   of your accepting any such warranty or additional liability.

END OF TERMS AND CONDITIONS