

NetBackup™ for HBase Administrator's Guide

UNIX, Windows, and Linux

Release 10.4



NetBackup™ for HBase Administrator's Guide

Last updated: 2024-03-27

Legal Notice

Copyright © 2024 Veritas Technologies LLC. All rights reserved.

Veritas, the Veritas Logo, Veritas Alta, and NetBackup are trademarks or registered trademarks of Veritas Technologies LLC or its affiliates in the U.S. and other countries. Other names may be trademarks of their respective owners.

This product may contain third-party software for which Veritas is required to provide attribution to the third party ("Third-party Programs"). Some of the Third-party Programs are available under open source or free software licenses. The License Agreement accompanying the Software does not alter any rights or obligations you may have under those open source or free software licenses. Refer to the Third-party Legal Notices document accompanying this Veritas product or available at:

<https://www.veritas.com/about/legal/license-agreements>

The product described in this document is distributed under licenses restricting its use, copying, distribution, and decompilation/reverse engineering. No part of this document may be reproduced in any form by any means without prior written authorization of Veritas Technologies LLC and its licensors, if any.

THE DOCUMENTATION IS PROVIDED "AS IS" AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID. Veritas Technologies LLC SHALL NOT BE LIABLE FOR INCIDENTAL OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH THE FURNISHING, PERFORMANCE, OR USE OF THIS DOCUMENTATION. THE INFORMATION CONTAINED IN THIS DOCUMENTATION IS SUBJECT TO CHANGE WITHOUT NOTICE.

The Licensed Software and Documentation are deemed to be commercial computer software as defined in FAR 12.212 and subject to restricted rights as defined in FAR Section 52.227-19 "Commercial Computer Software - Restricted Rights" and DFARS 227.7202, et seq. "Commercial Computer Software and Commercial Computer Software Documentation," as applicable, and any successor regulations, whether delivered by Veritas as on premises or hosted services. Any use, modification, reproduction release, performance, display or disclosure of the Licensed Software and Documentation by the U.S. Government shall be solely in accordance with the terms of this Agreement.

Veritas Technologies LLC
2625 Augustine Drive
Santa Clara, CA 95054

<http://www.veritas.com>

Technical Support

Technical Support maintains support centers globally. All support services will be delivered in accordance with your support agreement and the then-current enterprise technical support policies. For information about our support offerings and how to contact Technical Support, visit our website:

<https://www.veritas.com/support>

You can manage your Veritas account information at the following URL:

<https://my.veritas.com>

If you have questions regarding an existing support agreement, please email the support agreement administration team for your region as follows:

Worldwide (except Japan)

CustomerCare@veritas.com

Japan

CustomerCare_Japan@veritas.com

Documentation

Make sure that you have the current version of the documentation. Each document displays the date of the last update on page 2. The latest documentation is available on the Veritas website:

<https://sort.veritas.com/documents>

Documentation feedback

Your feedback is important to us. Suggest improvements or report errors or omissions to the documentation. Include the document title, document version, chapter title, and section title of the text on which you are reporting. Send feedback to:

NB.docs@veritas.com

You can also see documentation information or ask a question on the Veritas community site:

<http://www.veritas.com/community/>

Veritas Services and Operations Readiness Tools (SORT)

Veritas Services and Operations Readiness Tools (SORT) is a website that provides information and tools to automate and simplify certain time-consuming administrative tasks. Depending on the product, SORT helps you prepare for installations and upgrades, identify risks in your datacenters, and improve operational efficiency. To see what services and tools SORT provides for your product, see the data sheet:

https://sort.veritas.com/data/support/SORT_Data_Sheet.pdf

Contents

Chapter 1	Introduction	6
	Protecting HBase data using NetBackup	6
	Backing up HBase data	8
	Restoring HBase data	9
	NetBackup for HBase terminologies	10
	Limitations	11
Chapter 2	Deploying HBase plug-in for NetBackup	13
	About the HBase plug-in deployment	13
	Pre-requisites for installing the HBase plug-in	14
	Operating system and platform compatibility	14
	License for HBase plug-in for NetBackup	14
	Preparing the HBase cluster	14
	Best practices for deploying the HBase plug-in	15
	Post installation procedures	16
	Verifying the deployment of the HBase plug-in	16
Chapter 3	Configuring NetBackup for HBase	17
	About configuring NetBackup for HBase	17
	Managing backup hosts	18
	Adding a NetBackup client to the allowed list	19
	Configure a NetBackup Appliance as a backup host	20
	Adding HBase credentials in NetBackup	20
	Configuring the HBase plug-in using the HBase configuration file	22
	Configuring NetBackup for a highly-available HBase cluster	23
	Configuring communication between NetBackup and HBase clusters that have SSL enabled (HTTPS)	25
	Configuration for a HBase cluster that uses Kerberos	26
	Create a BigData policy for HBase clusters	26
	Disaster recovery of a HBase cluster	27

Chapter 4	Performing backups and restores of HBase	29
	About backing up a HBase cluster	29
	Prerequisites for running backup and restore operations for a HBase cluster with Kerberos authentication	30
	Backing up a HBase cluster	30
	Best practices for backing up a HBase cluster	31
	About restoring an HBase cluster	32
	Restoring HBase data on the same HBase cluster	32
	Restoring HBase data on an alternate HBase cluster	34
	Restoring truncated tables	37
	Best practices for restoring a HBase cluster	38
Chapter 5	Troubleshooting	39
	About NetBackup for HBase debug logging	39
	Backup fails with error 6609	40
	Backup fails with error 6601	40
	Backup fails with error 6623	41
	Restore fails with error 2850	41
	Backup fails with error 20	41
	Backup fails with error 112	41
	Backup operation fails with error 6654	42
	NetBackup configuration and certificate files do not persist after the container-based NetBackup appliance restarts	42
	Configuration file is not recovered after a disaster recovery	43

Introduction

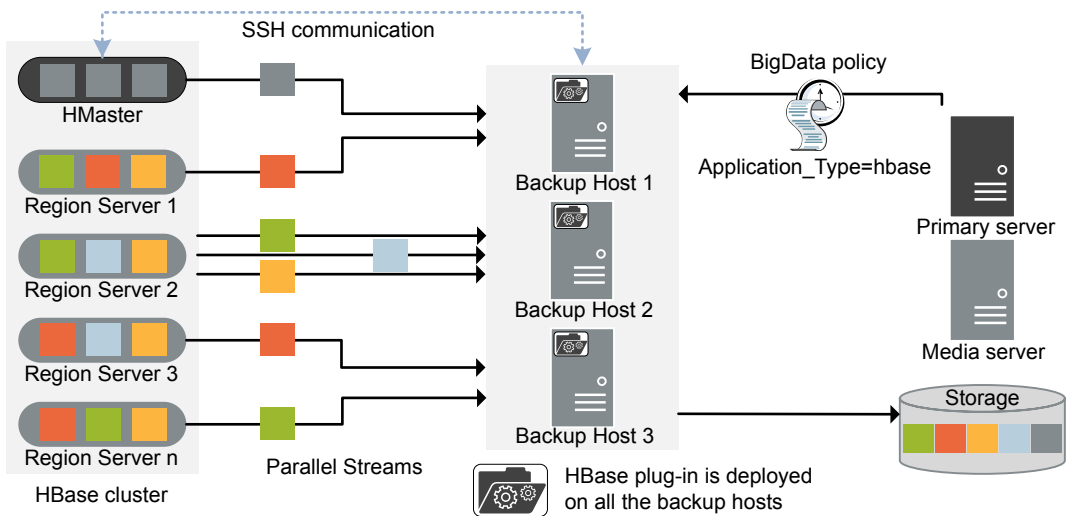
This chapter includes the following topics:

- [Protecting HBase data using NetBackup](#)
- [Backing up HBase data](#)
- [Restoring HBase data](#)
- [NetBackup for HBase terminologies](#)
- [Limitations](#)

Protecting HBase data using NetBackup

Using the NetBackup Parallel Streaming Framework (PSF), HBase data can now be protected using NetBackup.

The following diagram provides an overview of how NetBackup protects HBase data.

Figure 1-1 Architectural overview

The diagram contains the following information:

- The data is backed up in parallel streams wherein the Region servers stream data blocks simultaneously to multiple backup hosts. The job processing is accelerated due to multiple backup hosts and parallel streams.
- The communication between the HBase cluster and the NetBackup is enabled using the NetBackup plug-in for HBase. The plug-in is installed with the NetBackup installation.
- For NetBackup communication, you need to configure a BigData policy and add the related backup hosts.
- You can configure a NetBackup media server, client, or primary server as a backup host. Depending on the number of Region servers, you can add or remove backup hosts. You can scale up your environment easily by adding more backup hosts.
- The communication between the Hmaster and the backup hosts happens over SSH.
- The NetBackup Parallel Streaming Framework enables agentless backup where the backup and restore operations run on the backup hosts. There is no agent footprint on the cluster nodes. HBase cluster upgrades or maintenance do not affect NetBackup.

For more information:

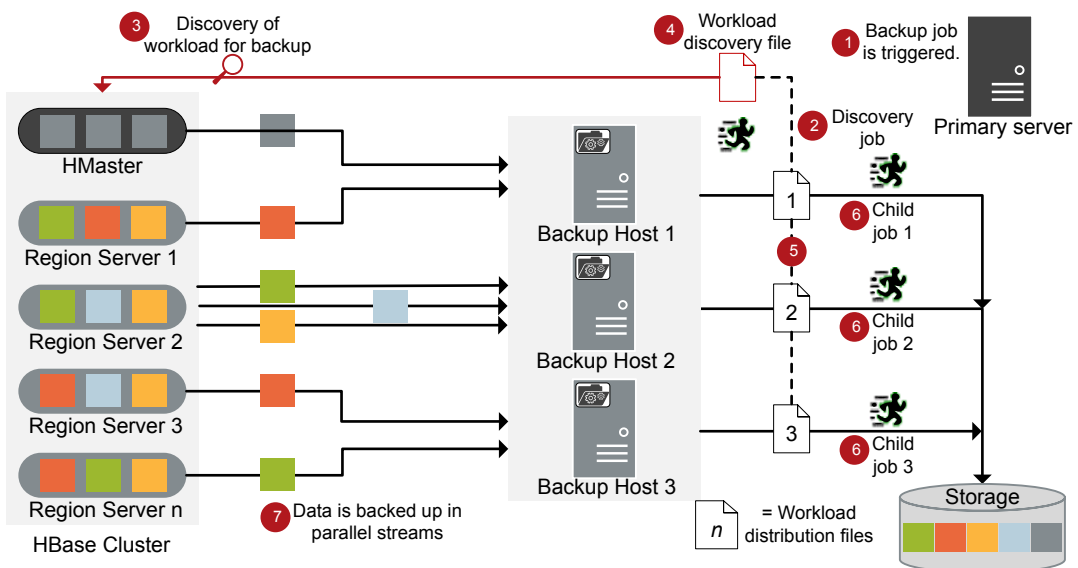
- See “[Backing up HBase data](#)” on page 8.
- See “[Restoring HBase data](#)” on page 9.
- See “[Limitations](#)” on page 11.
- See “[NetBackup for HBase terminologies](#)” on page 10.
- For information about the NetBackup Parallel Streaming Framework (PSF) refer to the *NetBackup Administrator's Guide, Volume I*.

Backing up HBase data

HBase data is backed up in parallel streams wherein HBase Region servers stream data blocks simultaneously to multiple backup hosts.

The following diagram provides an overview of the backup flow:

Figure 1-2 Backup flow



As illustrated in the following diagram:

1. A scheduled backup job is triggered from the primary server.
2. Backup job for HBase data is a compound job. When the backup job is triggered, first a discovery job is run.
3. During discovery, the first backup host connects with the HMaster and performs a discovery to get details of data that needs to be backed up.

4. A workload discovery file is created on the backup host. The workload discovery file contains the details of the data that needs to be backed up from the different Region servers.
5. The backup host uses the workload discovery file and decides how the workload is distributed amongst the backup hosts. Workload distribution files are created for each backup host.
6. Individual child jobs are executed for each backup host. As specified in the workload distribution files, data is backed up.
7. Data blocks are streamed simultaneously from different Region servers to multiple backup hosts.

The compound backup job is not completed until all the child jobs are completed. After the child jobs are completed, NetBackup cleans all the snapshots from the HMaster. Only after the cleanup activity is completed, the compound backup job is completed.

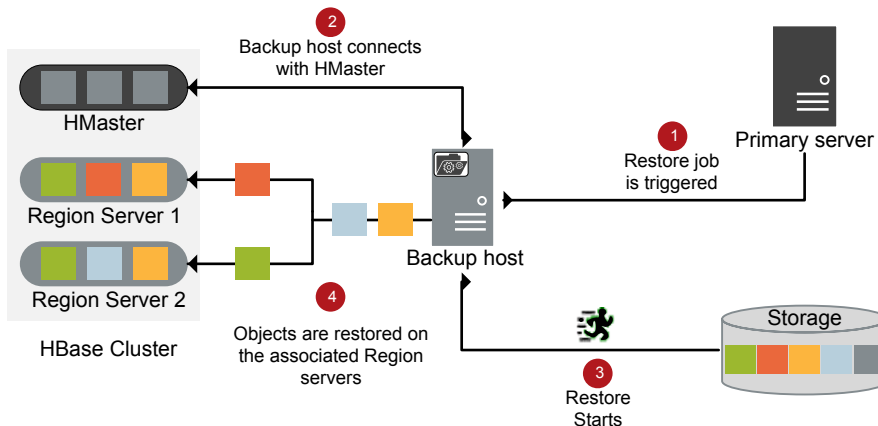
See [“About backing up a HBase cluster”](#) on page 29.

Restoring HBase data

For restore only one backup host is used.

The following diagram provides an overview of the restore flow.

Figure 1-3 Restore flow



As illustrated in the diagram:

1. The restore job is triggered from the primary server.

2. The backup host connects with the HMaster. Backup host is also the destination client.
 3. The actual data restore from the storage media starts.
 4. The data blocks are restored on the Region servers.
- See [“About restoring an HBase cluster”](#) on page 32.

NetBackup for HBase terminologies

The following table defines the terms you will come across when using NetBackup for protecting HBase cluster.

Table 1-1 NetBackup terminologies

Terminology	Definition
Compound job	A backup job for HBase data is a compound job. ■ The backup job runs a discovery job for getting information of the data to be backed up. ■ Child jobs are created for each backup host that performs the actual data transfer. ■ After the backup is complete, the job cleans up the snapshots on the HMaster and is then marked complete.
Discovery job	When a backup job is executed, first a discovery job is created. The discovery job communicates with the HMaster and gathers information of the block that needs to be backed up and the associated Region servers. At the end of the discovery, the job populates a workload discovery file that NetBackup then uses to distribute the workload amongst the backup hosts.
Child job	For backup, a separate child job is created for each backup host to transfer data to the storage media. A child job can transfer data blocks from multiple Region servers.
Workload discovery file	During discovery, when the backup host communicates with the HMaster, a workload discovery file is created. The file contains information about the data blocks to be backed up and the associated Region servers.
Workload distribution file	After the discovery is complete, NetBackup creates a workload distribution file for each backup host. These files contain information of the data that is transferred by the respective backup host.

Table 1-1 NetBackup terminologies (*continued*)

Terminology	Definition
Parallel streams	The NetBackup parallel streaming framework allows data blocks from multiple Region servers to be backed up using multiple backup hosts simultaneously.
Backup host	The backup host acts as a proxy client. All the backup and restore operations are executed through the backup host. You can configure media servers, clients, or a primary server as a backup host. The backup host is also used as destination client during restores.
BigData policy	The BigData policy is introduced to: ■ Specify the application type. ■ Allow backing up distributed multi-node environments. ■ Associate backup hosts. ■ Perform workload distribution.
Application server	HMaster is referred to as a application server in NetBackup.
Primary HMaster	In a high-availability scenario, you need to specify one HMaster with the BigData policy and with the <code>tpconfig</code> command. This HMaster is referred as the primary HMaster.
Fail-over HMaster	In a high-availability scenario, the HMaster other than the primary HMaster that are updated in the <code>hbase.conf</code> file are referred as fail-over HMaster.

Limitations

Review the following limitations before you deploy the HBase plug-in:

- Only RHEL and SUSE platforms are supported for HBase clusters and backup hosts.
- HBase plug-in does not capture Extended Attributes (xattrs) or Access Control Lists (ACLs) of an object during backup and hence these are not set on the restored files or folders.
- For a highly available HBase cluster, if a failover happens during a backup or restore operation, the job fails.
- You can cancel a backup and restore job manually while the discovery job for a backup operation is in progress. However, in this case the snapshot entry does not get removed from the HBase web graphical user interface (GUI).

- Backup of read-only tables is not supported.
- You need to add tables one-by-one in the backup selection when you create the backup policy.
- Backup and restore operations are not supported with Kerberos authentication if `NB_FIPS_MODE` is enabled at the `bp.conf`.

Note: To perform backup with Kerberos authentication, deploy a new backup host with `NB_FIPS_MODE=0` or disabled.

- If the CRL expires during the backup of an HTTPS-based Hadoop cluster, the backup runs partially.
- If you have multiple CRL-based Hadoop clusters, ensure that you add different backup hosts for every cluster.

Deploying HBase plug-in for NetBackup

This chapter includes the following topics:

- [About the HBase plug-in deployment](#)
- [Pre-requisites for installing the HBase plug-in](#)
- [Operating system and platform compatibility](#)
- [License for HBase plug-in for NetBackup](#)
- [Preparing the HBase cluster](#)
- [Best practices for deploying the HBase plug-in](#)
- [Post installation procedures](#)
- [Verifying the deployment of the HBase plug-in](#)

About the HBase plug-in deployment

The HBase plug-in is installed with NetBackup. Review the following topics to complete the deployment.

Table 2-1 HBase plug-in deployment

Task	Reference
Pre-requisites and requirements	See “Pre-requisites for installing the HBase plug-in” on page 14.
Preparing the HBase cluster	See “Preparing the HBase cluster” on page 14.

Table 2-1 HBase plug-in deployment (*continued*)

Task	Reference
Best practices	See “Best practices for deploying the HBase plug-in” on page 15.
Verifying the deployment	See “Verifying the deployment of the HBase plug-in ” on page 16.
Configuring	See “About configuring NetBackup for HBase” on page 17.

Pre-requisites for installing the HBase plug-in

Ensure that the following pre-requisites are met before you install the HBase plug-in:

- See [“Operating system and platform compatibility”](#) on page 14.
- See [“License for HBase plug-in for NetBackup”](#) on page 14.

Operating system and platform compatibility

With this release, RHEL and SUSE platforms are supported for HBase clusters and NetBackup backup hosts.

For more information, see the [NetBackup Master Compatibility List](#).

License for HBase plug-in for NetBackup

Backup and restore operations using the HBase plug-in for NetBackup, require the Application and Database pack license or NetBackup Platform Base – Big Data Workload Edition license.

More information is available on how to add licenses.

See the [NetBackup Administrator’s Guide, Volume I](#)

Preparing the HBase cluster

Perform the following tasks to prepare the HBase cluster for NetBackup:

- Update firewall settings (port 50070 by default) so that the backup hosts can communicate with the HBase cluster.
- Add the entries of all the HMaster and region servers to the `/etc/hosts` file on all the backup hosts. You must add the hostname in FQDN format.
Or

Add the appropriate DNS entries in the `/etc/resolve.conf` file.

- Add the entries of all the backup hosts to `/etc/hosts` file on the HMaster and region servers.
- Ensure that HBase service is enabled on the HBase cluster.
- HMaster user should be able to do SSH.
- Ensure that jdk package installed. Also ensure that, Java path is set and compatible with HBase version.
- Ensure that jps command is working on HMaster. For more details, refer the HBase documentation.
- Set the following environment variables for HMaster in the `.bashrc` file for all users configured under `tpconfig` of primary server.

```
■ export JAVA_HOME= PATH_OF_JAVA_DIR
■ export HADOOP_HOME=PATH_OF_HDFS_DIR
■ export HADOOP_MAPRED_HOME=$HADOOP_HOME
■ export HADOOP_COMMON_HOME=$HADOOP_HOME
■ export HADOOP_HDFS_HOME=$HADOOP_HOME
■ export YARN_HOME=$HADOOP_HOME
■ export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
■ export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
■ export HADOOP_INSTALL=$HADOOP_HOME
■ export HADOOP_OPTS="$HADOOP_OPTS
-Djava.library.path=$HADOOP_HOME/lib/native"
■ export HBASE_HOME=PATH OF HBASE DIR
■ PATH=$PATH:$HBASE_HOME/bin:$JAVA_HOME/bin
■ export CLASSPATH=$CLASSPATH:/usr/local/hadoop/hbase/lib/*
```

Best practices for deploying the HBase plug-in

Consider the following when you deploy HBase plug-in and configure NetBackup for HBase:

- Use consistent conventions for hostnames of backup hosts, media servers, and primary server. For example, if you are using the hostname as **HBase.veritas.com** (FQDN format) use the same everywhere.

- Add the entries of all the HMaster and region servers to the `/etc/hosts` file on all the backup hosts. You must add the hostname in FQDN format.
Or
Add the appropriate DNS entries in the `/etc/resolve.conf` file.
- Always specify the HMaster and region servers in FQDN format.
- Ping all the nodes (use FQDN) from the backup hosts.

Post installation procedures

Complete the post-installation procedures:

See [“Verifying the deployment of the HBase plug-in ”](#) on page 16.

See [“Configuration for a HBase cluster that uses Kerberos”](#) on page 26.

See [“Configuring NetBackup for a highly-available HBase cluster”](#) on page 23.

Verifying the deployment of the HBase plug-in

After you install the HBase plug-in, the following files are deployed:

- `/usr/opensv/lib/psf-plugins/hbase/libaapipgnhbase.so`

Configuring NetBackup for HBase

This chapter includes the following topics:

- [About configuring NetBackup for HBase](#)
- [Managing backup hosts](#)
- [Adding HBase credentials in NetBackup](#)
- [Configuring the HBase plug-in using the HBase configuration file](#)
- [Configuration for a HBase cluster that uses Kerberos](#)
- [Create a BigData policy for HBase clusters](#)
- [Disaster recovery of a HBase cluster](#)

About configuring NetBackup for HBase

Table 3-1 Configuring NetBackup for HBase

Task	Reference
Adding backup hosts	See “Managing backup hosts” on page 18. If you want to use NetBackup client as a backup host, you need to whitelist the NetBackup client on the primary server. See “Adding a NetBackup client to the allowed list” on page 19.
Adding HBase credentials in NetBackup	See “Adding HBase credentials in NetBackup” on page 20.

Table 3-1 Configuring NetBackup for HBase (*continued*)

Task	Reference
Configuring the HBase plug-in using the HBase configuration file	See “Configuring the HBase plug-in using the HBase configuration file” on page 22. See “Configuring NetBackup for a highly-available HBase cluster” on page 23.
Configuring the backup hosts for HBase clusters that use Kerberos	See “Configuration for a HBase cluster that uses Kerberos” on page 26.
Configuring NetBackup policies for HBase plug-in	See “Create a BigData policy for HBase clusters” on page 26.

Managing backup hosts

A backup host acts as a proxy client that hosts all the backup and restore operations for HBase clusters. For the HBase plug-in for NetBackup, the backup host performs all the backup and restore operations without the need to install a separate agent on the HBase cluster.

The backup host must be a Linux computer. NetBackup supports only RHEL and SUSE platforms as a backup host.

The backup host can be a NetBackup client or a media server or a primary server. NetBackup recommends that you have a media server as a backup host.

Consider the following before adding a backup host:

- For backup operations, you can add one or more backup hosts.
- For restore operations, you can add only one backup host.
- A primary, media, or client can perform the role of a backup host.
- HBase plug-in for NetBackup is deployed on all the backup hosts.
- When you use multiple backup hosts, make sure that all backup hosts can communicate with the media server.

You can add a backup host when you configure the BigData policy. More information is available on how to create a policy.

See [“Create a BigData policy for HBase clusters”](#) on page 26.

Add a backup host

To add a backup host

- 1 Open the policy that you want to edit.
- 2 In the **Backup selections** tab, click **Add**.
- 3 Add the backup host in the following format:

Backup_Host=<hostname>

Alternatively, you can also add a backup host using the following command:

```
bpplinclude PolicyName -add "Backup_Host=hostname"
```

- 4 As a best practice, add the entries of all the Hprimary servers and Region servers to the `/etc/hosts` file on all the backup hosts. You must add the host name in FQDN format.

OR

Add the appropriate DNS entries in the `/etc/resolve.conf` file.

Remove a backup host

To remove a backup host

- 1 Open the policy that you want to edit.
- 2 In the **Backup selections** tab, select the backup select that contains the backup host that you want to remove.
- 3 Select the backup host and click **Delete**.

Alternatively, you can also remove a backup host using the following command:

```
bpplinclude PolicyName -delete "Backup_Host=hostname"
```

Adding a NetBackup client to the allowed list

To use the NetBackup client as a backup host, you must add it to the allowed list. Perform this procedure on the NetBackup primary server.

Adding a host to an allowed list is a security practice used for restricting systems from running software or applications unless these have been approved for safe execution.

To add a NetBackup client to the allowed list

- ◆ Run the following command on the NetBackup primary server:

- For UNIX

The directory path to the command:

```
/usr/opensv/var/global/bin/admincmd/bpsetconfig
```

```
bpsetconfig -h primaryserver
bpsetconfig> APP_PROXY_SERVER = clientname.domain.org
bpsetconfig>
UNIX systems: <ctl-D>
```

- For Windows

The directory path to the command:

```
<install_path>\NetBackup\bin\admincmd\bpsetconfig
bpsetconfig -h primaryserver
bpsetconfig> APP_PROXY_SERVER = clientname1.domain.org
bpsetconfig> APP_PROXY_SERVER = clientname2.domain.org
bpsetconfig>
Windows systems: <ctl-Z>
```

This command sets the `APP_PROXY_SERVER = clientname` entry in the backup configuration (`bp.conf`) file or the Windows registry.

For more information about the `APP_PROXY_SERVER = clientname`, refer to the *Configuration options for NetBackup clients* section in the [NetBackup Administrator's Guide, Volume I](#).

Configure a NetBackup Appliance as a backup host

Review the following articles if you want to use NetBackup Appliance as a backup host:

- [Using NetBackup Appliance as the backup host of HBase with Kerberos authentication.](#)
- [Using NetBackup Appliance as the backup host with highly-available HBase cluster.](#)

Adding HBase credentials in NetBackup

To establish a seamless communication between HBase clusters and NetBackup for successful backup and restore operations, you must add and update HBase credentials to the NetBackup primary server.

Use the `tpconfig` command to add credentials in NetBackup primary server.

For HBase you need to provide the RSA fingerprint when you add the credentials.

For more information about the `tpconfig` command, see the [NetBackup Commands Reference Guide](#).

Consider the following when you add HBase credentials:

- For a highly-available HBase cluster, ensure that the user for the primary and the failover HMaster is the same.
- Use the credentials of the application server that you want to configure in the BigData policy.
- For the HBase cluster that uses Kerberos, specify the actual Kerberos username as `application_server_user_id` value.
- The RSA key must be in the SHA-256 format.
- Ensure that RSA is supported on the backup host and to obtain the RSA key, run the following command:

```
ssh_host_rsa_key.pub | awk '{print $2}' | base64 -d | sha256sum  
| awk '{print $1}'
```

This utility is available at `/etc/ssh`.

To add credentials in NetBackup

- 1 Run `tpconfig` command from the following directory paths:
 On UNIX systems, `/usr/opensv/volmgr/bin/`
 On Windows systems, `install_path\Volmgr\bin\`
- 2 Run the `tpconfig --help` command. A list of options which are required to add, update, and delete HBase credentials is displayed.
- 3 Run the `tpconfig -add -application_server application_server_name -application_server_user_id user_ID -application_type hbase -password password` command by providing appropriate values for each parameter to add HBase credentials.

For example, if you want to add the credentials for the HBase server that has *application_server_name* as *HBase1*, then run the following command using the appropriate *<user_ID>* and *<password>* details.

```
tpconfig -add -application_server HBase1 -application_type hbase  
-application_server_user_id HBase -password password
```

- 4 You are prompted to enter the password.

```
5  tpconfig -add -application_server testserver.veritas.com
    -application_server_user_id hadoop-application_type hbase
    -requiredport 60010 -password hadoop -host_user_id hadoop
    -host_password hadoop-host_RSA_key
    b2352722053ac9f40bc1XXXXXXXXXXXXXXXXXXXXXXXXXXXX419fa241ba9431fd6b9
```

Note: Ensure that the HBase user has Admin permissions for SSH and for HBase folders.

6 Run the `tpconfig -dappservers` command to verify if the NetBackup primary server has the HBase credentials added.

Configuring the HBase plug-in using the HBase configuration file

The backup hosts use the `hbase.conf` file to save the configuration settings of the HBase plug-in. You need to create a separate file for each backup host and copy it to the `/usr/opensv/var/global/`. You need to manually create the `hbase.conf` file in JSON format. This file is not available by default with the installer.

Note: You must not provide a blank value for any of the parameters, or the backup job fails.

With this release, the following plug-in settings can be configured:

- See [“Configuring NetBackup for a highly-available HBase cluster”](#) on page 23.
- See [“Configuring communication between NetBackup and HBase clusters that have SSL enabled \(HTTPS\)”](#) on page 25.

Following is an example of the `hbase.conf` file.

Note: For non-HA environment, the fail-over parameters are not required.

```
{
  "application_servers":
  {
    "hostname_of_the_primary_HMaster":
    {
      "failover_HMaster":
```

```
[
  {
    "hostname": "hostname_of_failover_HMaster"

  }
]

}
}

}
```

Configuring NetBackup for a highly-available HBase cluster

To protect a highly-available HBase cluster, when you configure NetBackup for HBase cluster:

- Specify one of the HMaster (primary) as the client in the BigData policy.
- Specify the same HMaster (primary and fail-over) as application server when you execute the `tpconfig` command.
- Create a `hbase.conf` file, update it with the details of the HMaster (primary and fail-over), and copy it to all the backup hosts. The `hbase.conf` file is in JSON format.
- Hostname and port of the HMaster must be same as you have specified with the `http address` parameter in the `hbase-site.xml` of the HBase cluster.
- User name of the primary and fail-over HMaster must be same.
- Do not provide a blank value for any of the parameters, or the backup job fails.

To update the HBase.conf file for highly-available HBase cluster

- 1 Update the `hbase.conf` file with the following parameters:

```
{
  "application_servers":
  {
    "hostname_of_primary_HMaster1":
    {
      "failover_HMaster":
      [
        {
          "hostname": "hostname_of_failover_HMaster1"
        }
      ]
    }
  }
}
```

- 2 If you have multiple HBase clusters, use the same `hbase.conf` file to update the details. For example,

```
{
  "application_servers":
  {
    "hostname_of_primary_HMaster1":
    {
      "failover_HMaster":
      [
        {
          "hostname": "hostname_of_failover_HMaster1"

        }
      ],

    },

    "hostname_of_primary_HMaster2":
    {
      "failover_HMaster":
      [
        {
          "hostname": "hostname_of_failover_HMaster2",

        }
      ],

    }
  }
}
```

- 3 Copy this file to the following location on all the backup hosts:

```
/usr/opensv/var/global/
```

Configuring communication between NetBackup and HBase clusters that have SSL enabled (HTTPS)

To enable SSL communication between NetBackup and HBase, refer to the following topic for the *NetBackup for Hadoop Administrator's Guide*:

[Configuring communication between NetBackup and Hadoop clusters that are SSL-enabled \(HTTPS\)](#)

Configuration for a HBase cluster that uses Kerberos

For a HBase cluster that uses Kerberos, perform the following tasks on all the backup hosts:

- Ensure that the Kerberos package (krb5-workstation package) is present on all the backup hosts.
- Acquire the `keytab` file and copy it to a secure location on the backup host.
- Ensure that the `keytab` has the required principal.
- Manually update the `krb5.conf` file with the appropriate KDC server and realm details.

Note: Ensure that `default_cache_name` parameter is not set to the **KEYRING:persistent:%{uid}** value. You can comment the parameter to use the default or you can specify a file name such as, **FILE:/tmp/krb_file_name:%{uid}**.

- When you add HBase credentials in NetBackup, specify "**kerberos**" as `application_server_user_id` value. See ["Adding HBase credentials in NetBackup"](#) on page 20.
- To run backup and restore operations for a HBase cluster that uses Kerberos authentication, HBase needs a valid Kerberos ticket-granting ticket (TGT) to authenticate with the HBase cluster. See ["Prerequisites for running backup and restore operations for a HBase cluster with Kerberos authentication"](#) on page 30.

Create a BigData policy for HBase clusters

Use the following procedure to create a BigData policy.

To create a BigData policy for HBase clusters

- 1 Open the NetBackup web UI.
- 2 On the left, click **Protection > Policies**.
- 3 Click **Add**.
- 4 On the **Attributes** tab, select **BigData** as the policy type.

- 5 On the **Schedules** tab, click **Add** to create a schedule.

You can create a schedule for a **Full backup**, **Differential incremental backup**, or **Cumulative incremental backup**. After you set the schedule, HBase data is backed up automatically as per the set schedule without any further user intervention.

- 6 On the **Start window** tab, configure the window during which you want the backup to start.
- 7 On the **Clients** tab, enter the IP address or the host name of the NameNode.
- 8 On the **Backup selections** tab, enter the following parameters and their values as shown:

- *Application_Type=hbase*
The parameter values are case-sensitive.
- *Backup_Host=IP_address or hostname*
The backup host must be a Linux computer. The backup host can be a NetBackup client or a media server.
You can specify multiple backup hosts.
- Tables to back up
You can specify multiple tables.

Note: When you define a BigData policy with *Application_Type=hbase*, the table names cannot contain space or comma in their names.

- 9 Click **Create**.

For more information on using NetBackup for BigData applications, see the [NetBackup documentation](#) page.

Disaster recovery of a HBase cluster

For disaster recovery of the HBase cluster, perform the following tasks:

Table 3-2 Performing disaster recovery

Task	Description
After the HBase cluster and nodes are up, prepare the cluster for operations with NetBackup.	Perform the following tasks: Update firewall settings so that the backup hosts can communicate with the HBase cluster. Ensure that the webhbase service is enabled on the HBase cluster. See “Preparing the HBase cluster” on page 14.
To establish a seamless communication between HBase clusters and NetBackup for successful backup and restore operations, you must add and update HBase credentials to NetBackup master server.	Use <code>tpconfig</code> command to add HBase credentials in NetBackup master server. See “Adding HBase credentials in NetBackup” on page 20.
The backup hosts use the <code>HBase.conf</code> file to save the configuration settings of the HBase plug-in. You need to create a separate file for each backup host and copy it to <code>/usr/opensv/var/global/</code> . You need to create the <code>HBase.conf</code> file in JSON format.	See “Configuring NetBackup for a highly-available HBase cluster” on page 23.
Update the BigData policy with the original HMaster name.	See “Create a BigData policy for HBase clusters” on page 26.

Performing backups and restores of HBase

This chapter includes the following topics:

- [About backing up a HBase cluster](#)
- [About restoring an HBase cluster](#)
- [Restoring HBase data on an alternate HBase cluster](#)
- [Restoring truncated tables](#)
- [Best practices for restoring a HBase cluster](#)

About backing up a HBase cluster

Use NetBackup web UI to manage backup operations.

Table 4-1 Backing up HBase data

Task	Reference
Process understanding	See “Backing up HBase data” on page 8.
(Optional) Complete the pre-requisite for Kerberos	See “Prerequisites for running backup and restore operations for a HBase cluster with Kerberos authentication” on page 30.
Backing up a HBase cluster	See “Backing up a HBase cluster” on page 30.
Best practices	See “Best practices for backing up a HBase cluster” on page 31.

Table 4-1 Backing up HBase data (*continued*)

Task	Reference
Troubleshooting tips	For discovery and cleanup-related logs, review the following log file on the first backup host that triggered the discovery. <code>/usr/openv/var/global/logs/nbaapidiscv</code> For data transfer related logs, search for corresponding backup host (using the hostname) in the log files on the primary server. See “About NetBackup for HBase debug logging” on page 39.

Prerequisites for running backup and restore operations for a HBase cluster with Kerberos authentication

To run backup and restore operations for a HBase cluster that uses Kerberos authentication, HBase needs a valid Kerberos ticket granting-ticket (TGT) to authenticate with the HBase cluster.

Note: During the backup and restore operations, the TGT must be valid. Thus, specify the TGT validity accordingly or renew it when required during the operation.

Run the following command to generate the TGT:

```
kinit -k -t /keytab_file_location/keytab_filename principal_name
```

For example,

```
kinit -k -t /usr/openv/var/global/nbusers/hbase_mykeytabfile.keytab  
hbase@MYCOMPANY.COM
```

Also review the configuration-related information. See [“Configuration for a HBase cluster that uses Kerberos”](#) on page 26.

Backing up a HBase cluster

You can schedule a backup job with a NetBackup policy. Or, you can run a backup job manually. See the [NetBackup Administrator's Guide, Volume I](#) for details on manual backups.

The backup process includes the following stages:

1. **Pre-processing:** In the pre-processing stage, the first backup host that you have configured with the BigData policy, starts the discovery. At this stage, a snapshot of the complete backup selection is generated. The snapshot details are visible on the Region server web interface.

2. Data transfer: During the data transfer process, one child job is created for each backup host.
3. Post-processing: As part of the post-processing, NetBackup cleans up the snapshots on the Region server.

An overview of the backup process is available.

See [“Backing up HBase data”](#) on page 8.

Considerations

- On the Hmaster, set the `PasswordAuthentication` field to **Yes** in the `/etc/ssh/sshd_config` file. After you update the file, restart `sshd`. Ensure that all the cluster servers support supports same hash key algorithm (RSA).
- Snapshots are not cleaned up if you cancel a job manually. After you cancel the job you must manually delete the snapshots from the HBase shell.
- If you take a backup of an empty table, you need to clean the snapshot manually from the HBase shell.
- See [“Best practices for backing up a HBase cluster”](#) on page 31.

Best practices for backing up a HBase cluster

Before backing up a HBase cluster, consider the following:

- Before you execute a backup job, ensure for a successful ping response from the backup hosts to hostname (FQDN) of all the nodes.
- Update the firewall settings so that the backup hosts can communicate with the HBase cluster.
- Ensure that the HBase table you want to protect is Snapshottable.
- HBase table folder should not be deleted from `hdfs` if snapshot was taken on that table. If deleted, snapshot loses the reference and won't be able to restore or recover data from that snapshot.
- Do not backup truncated or empty table. The backup job will fail.
- Namespace name and table name must not be the same. The backup job will fail.
- The tables specified for backup selection must not contain space or comma in their names.
table selection must be separated by colon. For example, `namespace:tablename`.
- The tables specified for backup selection must not be empty.

- Ensure that the local time on the HBase nodes and the backup host are synchronized with the NTP server.

About restoring an HBase cluster

Use the NetBackup web UI to manage restore operations.

Table 4-2 Restoring HBase data

Task	Reference
Process understanding	See “Restoring HBase data” on page 9.
Complete the pre-requisites for Kerberos	See “Prerequisites for running backup and restore operations for a HBase cluster with Kerberos authentication” on page 30.
Restoring HBase data on the same Hprimary or HBase cluster	See “Restoring HBase data on the same HBase cluster” on page 32.
Restoring HBase data to an alternate Hprimary or HBase cluster This task can be performed only using the <code>bprestore</code> command.	See “Restoring HBase data on an alternate HBase cluster” on page 34.
HBase has a limitation to restore truncated tables. As a workaround, you must restore to archive path.	See “Restoring truncated tables” on page 37.
Best practices	See “Best practices for restoring a HBase cluster” on page 38.
Troubleshooting tips	See “About NetBackup for HBase debug logging” on page 39.

Considerations

When you restoring disabled table, the table will be enabled after successful restore.

Restoring HBase data on the same HBase cluster

To restore HBase data on the same HBase cluster, consider the following:

- Use the NetBackup web UI to initiate HBase data restore operations. This interface lets you select the NetBackup server from which the objects are restored and the client whose backup images you want to browse. Based upon these

selections, you can browse the backup image history, select individual items and initiate a restore.

- The restore browser is used to display HBase directory objects. A hierarchical display is provided where objects can be selected for restore. The objects (HBase directory or files) that make up a HBase cluster are displayed by expanding an individual directory.
- An administrator can browse for and restore HBase directories and individual items. Objects that users can restore include HBase files and folders.

To restore HBase data on the same HBase cluster

- 1 Open the NetBackup web UI.
- 2 On the left, click **Recovery**.
- 3 Under **Regular recovery**, click **Start recovery**.
- 4 On the **Policies** tab, click **Add**.
- 5 On the **Basic properties** tab, enter the following:
 - Select the **Policy type** named **BigData > HBase**
 - Specify the HBase application server as the source for which you want to perform the restore operation.
From the **Source client** list, select the required Application server.
 - Specify the backup host as the destination client.
From the **Destination client** list, select the required backup host. Restore is faster if the backup host is the media server that had backed up the node.
 - Click **Next**.
- 6 On the **Recovery details** tab, do the following:
 - Select the appropriate date range to restore the complete data set. Or, click **Edit** to select **Use backup history** and then select the backup images that you want to restore.
 - Select the files and folders for restore.
 - Click **Next**.
- 7 On the **Recovery options** tab, do the following:
 - Select **Restore everything to original location** if you want to restore your files to the same location where you performed your backup.
 - Select **Restore everything to a different location** if you want to restore your files to a location which is not the same as your backup location.
Provide the path.

- Select **Restore individual directories and files to different locations** if you want to restore files and directories to separate locations.
 - Locate **Recovery options** and select the wanted options.
 - Click **Next**.
- 8 On the **Review** tab, verify the recovery options and click **Start recovery**.

Restoring HBase data on an alternate HBase cluster

NetBackup lets you restore HBase data to another HMaster or HBase cluster. This type of restore method is also referred to as redirected restores.

Consider the following when you perform an alternate restore

- To restoring HBase tables to different cluster, both cluster must have same HBase version deployed.
- Make sure that you have added the credentials for the alternate HMaster or HBase cluster in NetBackup primary server.

To perform redirected restore for HBase

- 1 Modify the values for *rename_file* and *listfile* as follows:

Parameter	Value
<i>rename_file</i>	Change /<<namespace:source_table_name> to /<<namespace:destination_table_name> ALT_APPLICATION_SERVER=<alternate name node>
<i>listfile</i>	List of all the HBase files to be restored

Note: /<namespace:source_table_name> and
/<namespace:destination_table_name> must be different.

- 2 Run the `bprestore -S primary_server -D backup_host -C client -R rename_file -t 44 -L progress_log -f listfile` command on the NetBackup primary server using the modified values for the mentioned parameters in step 1.

Where,

`-S primary_server`

Specifies the name of the NetBackup primary server.

`-D backup_host`

Specifies the name of the backup host.

`-C client`

Specifies a HMaster as a source to use for finding backups or archives from which to restore files. This name must be as it appears in the NetBackup catalog.

`-f listfile`

Specifies a file (listfile) that contains a list of files to be restored and can be used instead of the file names option. In listfile, list each file path must be on a separate line.

`-L progress_log`

Specifies the name of whitelisted file path in which to write progress information.

`-t 44`

Specifies BigData as the policy type.

`-R rename_file`

Specifies the name of a file with name changes for alternate-path restores.

Use the following form for entries in the rename file:

`change backup_tablename to restore_tablename`

`ALT_APPLICATION_SERVER=<Application Server Name>`

The file paths must start with / (slash).

Note: Ensure that you have whitelisted all the file paths such as `<rename_file_path>`, `<progress_log_path>` that are already not included as a part of NetBackup install path.

Restoring truncated tables

HBase has a limitation to restore truncated tables. As a workaround, follow the procedure.

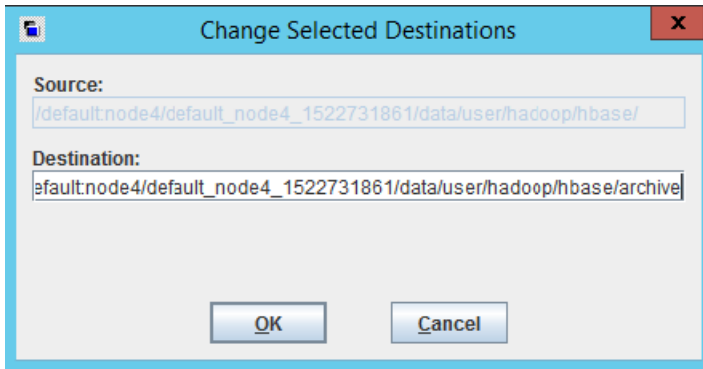
To restore truncated tables

- 1 Open the **Backup, Archive, and Restore** interface.
- 2 Select the appropriate date range to restore the complete data set.
- 3 In the **Browse** directory, specify the root directory ("/") as the path to browse.
- 4 From the File menu (Windows) or Actions menu (UNIX), choose **Specify NetBackup Machines and Policy Type**.
- 5 On the **Specify NetBackup Machines and Policy Type** wizard, enter the source and destination details for restore.
 - Specify the HBase HMaster as the source for which you want to perform the restore operation.
From the **Source client for restores** list, select the required HMaster.
 - Specify the backup host as the destination client.
From the **Destination client for restores** list, select the required backup host.
 - On the **Specify NetBackup Machines and Policy Type** wizard, enter the policy type details for restore.
From the **Policy type for restores** list, choose **BigData** as the policy type for restore.
Click **Ok**.
- 6 Go to the **Backup History** and select the backup images that you want to restore.
- 7 In the **Directory Structure** pane, expand the **Directory**.
All the subsequent files and folders under the directory are displayed in the **Contents of Selected Directory** pane.
- 8 In the **Contents of Selected Directory** pane, select the check box for the HBase files that you want to restore.
- 9 Click **Restore**.
- 10 In the **Restore Marked Files** dialog box, select **Restore individual directories and files to different locations**.
- 11 Select the source HBase directory.

12 Click **Change Selected Destination(s)...**

The **Change Selected Destinations** dialog box is displayed.

13 In the **Destinations** field, add archive at the end of the destination directory.



14 Click **OK**.

15 Click **Start Restore**.

16 Verify the restored files.

Best practices for restoring a HBase cluster

When restoring a HBase cluster, consider the following:

- Before you execute a restore job, ensure that there is sufficient space on the cluster to complete the restore job.
- Update firewall settings so that the backup hosts can communicate with the HBase cluster.
- When restoring large tables make sure timeout values are set to larger values accordingly on the backup hosts.

Troubleshooting

This chapter includes the following topics:

- [About NetBackup for HBase debug logging](#)
- [Backup fails with error 6609](#)
- [Backup fails with error 6601](#)
- [Backup fails with error 6623](#)
- [Restore fails with error 2850](#)
- [Backup fails with error 20](#)
- [Backup fails with error 112](#)
- [Backup operation fails with error 6654](#)
- [NetBackup configuration and certificate files do not persist after the container-based NetBackup appliance restarts](#)
- [Configuration file is not recovered after a disaster recovery](#)

About NetBackup for HBase debug logging

NetBackup maintains process-specific logs for the various processes that are involved in the backup and restore operations. Examining these logs can help you to find the root cause of an issue.

These log folders must already exist in order for logging to occur. If these folders do not exist, you must create them.

The log folders reside on the following directories.

- On Windows: `install_path\NetBackup\logs`

- On UNIX or Linux: `/usr/openv/var/global/logs`

Table 5-1 NetBackup logs related to HBase

Log Folder	Messages related to	Logs reside on
<code>install_path/NetBackup/logs/bpVMutil</code>	Policy configuration	Primary server
<code>install_path/NetBackup/logs/nbaapidiscv</code>	BigData framework, discovery, and HBase configuration file logs	Backup host
<code>install_path/NetBackup/logs/bpbm</code>	Policy validation, backup, and restore operations.	Media server
<code>install_path/NetBackup/logs/bpbkar</code>	Backup	Backup host
<code>install_path/NetBackup/logs/tar</code>	Restore and HBase configuration file.	Backup host

For more details, refer to the [NetBackup Logging Reference Guide](#).

Backup fails with error 6609

Backup fails with the following error:

```
(6609) The NetBackup plug-in cannot complete the operation because
the object is invalid.
```

Workaround:

Download and install the HBase plug-in.

Backup fails with error 6601

Backup fails with the following error:

```
(6601) One or more of the input parameters or arguments are invalid.
```

Workaround:

Remove non-existing tables from the backup selection.

Backup fails with error 6623

Backup fails with the following error:

```
(6623) Failed to connect to the application server or the backup
host. The server is either shut down or not reachable.
```

Workaround:

HMaster or Data nodes are offline. Ensure that HMaster or Data nodes are online.

Restore fails with error 2850

Restore fails with the following error:

```
(2850) Restore error.
```

Workaround:

Ensure that the destination client is a backup host.

Backup fails with error 20

Backup fails with the following error:

```
(20) invalid command parameter.
```

Workaround:

Ensure that the backup host is online and connects to the HMaster.

Backup fails with error 112

Backup fails with the following error:

```
(112) no files specified in the file list.
```

Workaround

Either the HBase credentials are not added in NetBackup primary server, or the credentials added are invalid.

Ensure that the HBase credentials are added in NetBackup primary server.

You can use the NetBackup `tpconfig` command to add the credentials. See [“Adding HBase credentials in NetBackup”](#) on page 20.

Backup operation fails with error 6654

This error is encountered during the following scenarios:

- If HBase credentials are not added in NetBackup primary server.
Workaround:
Ensure that the HBase credentials are added in NetBackup primary server. Use the `tpconfig` command. For more information, See [“Adding HBase credentials in NetBackup”](#) on page 20.
- If HBase plug-in files are not installed on backup host.
Workaround:
Ensure that the HBase plug-in files are installed on all backup hosts before you begin backup operation.
- If a NetBackup client that is used as a backup host is not allowlisted.
Workaround:
Ensure that the NetBackup client that is used as a backup host is allowlisted before you begin backup operation.
See [“Adding a NetBackup client to the allowed list”](#) on page 19.

NetBackup configuration and certificate files do not persist after the container-based NetBackup appliance restarts

The NetBackup configuration files like `hadoop.conf` or `hbase.conf` or SSL certificate and CRL paths do not persist after the container-based NetBackup Appliance restarts for any reason. This issue is applicable where container-based NetBackup Appliance is used as a backup host to protect the Hadoop or HBase workload.

Reason:

In the NetBackup Appliance environments the files that are available in the docker host's persistent location are retained after restart operation. The `hadoop.conf` and `hbase.conf` files are custom configuration files and are not listed in the persistent location.

The configuration files are used for defining values like HA (high availability) nodes during a failover and number of threads for backup. If these files get deleted, backups use the default values for both HA and number of threads that are Primary Name Node and 4 respectively. Backup fails only if the primary node goes down in such a case as plug-in fails to find secondary server.

If the SSL certificates and CRL path files are stored at a location that is not persistent the appliance restart, the backups and restore operations fail.

Workaround:

If custom configuration files for Hadoop and HBase get deleted after a restart, you can manually create the files at the following location:

- **Hadoop:** `/usr/openv/var/global/hadoop.conf`
- **HBase:** `/usr/openv/var/global/hbase.conf`

You can store the CA certificate that has signed the Hadoop or HBase SSL certificate and CRL at the following location:

`/usr/openv/var/global/`

Configuration file is not recovered after a disaster recovery

When you use NetBackup primary server as a backup host for high availability with an HBase cluster or an HBase cluster that is SSL-enabled (HTTPS) and run a full catalog recovery, the `hbase.conf` configuration file is not recovered.

Create the configuration file manually. Use the following format for the configuration file:

```
{
  "application_servers":
  {
    "primary.host.com":
    {
      "use_ssl":true
      "failover_namenodes":
      [
        {
          "hostname":"secondary.host.com",
          "use_ssl":true
          "port":11111
        }
      ],
      "port":11111
    }
  },
  "number_of_threads":5
}
```