Veritas NetBackup for HBase Administrator's Guide

UNIX, Windows, and Linux

8.1.1

VERITAS™

Veritas HBase Administartor's Guide

Last updated: 2018-04-24

Document version:NetBackup 8.1.1

Legal Notice

Veritas Technologies LLC
500 E Middlefield Road
Mountain View, CA 94043

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

This chapter includes the following topics:

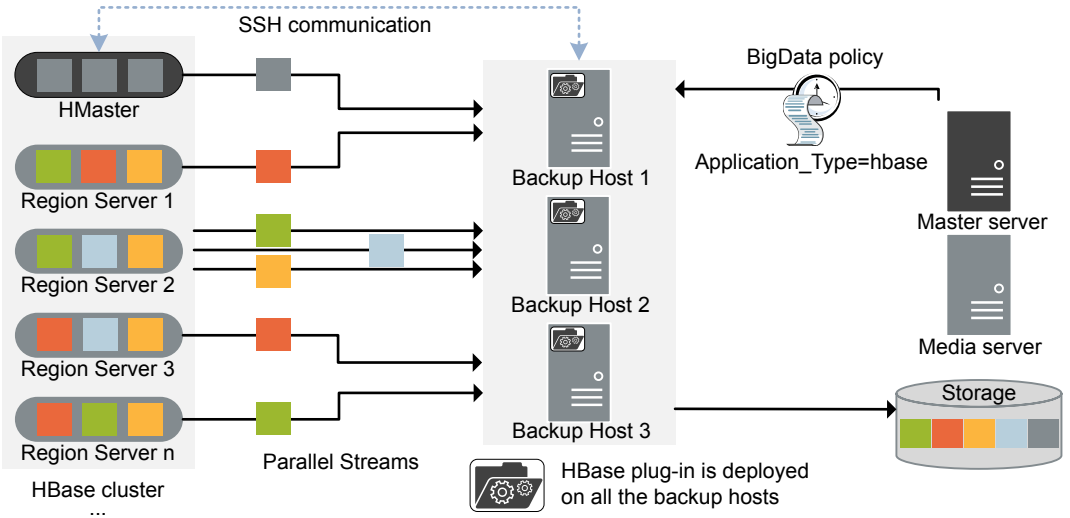
Protecting HBase data using NetBackup

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

The following diagram provides an overview of how HBase data is protected by NetBackup.

Also, review the definitions of terminologies.See “NetBackup for HBase terminologies” on page 11.

Figure 1-1 Architectural overview



As illustrated in the diagram:

- 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 available separately and must be installed on all the backup hosts.
- For NetBackup communication, you need to configure a Big Data policy and add the related backup hosts.
- You can configure a NetBackup media server, client, or master server as a backup host. Also, 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 wherein the backup and restore operations run on the backup hosts. There is no agent footprint on the cluster nodes. Also, NetBackup is not affected by the HBase cluster upgrades or maintenance.

For more information:

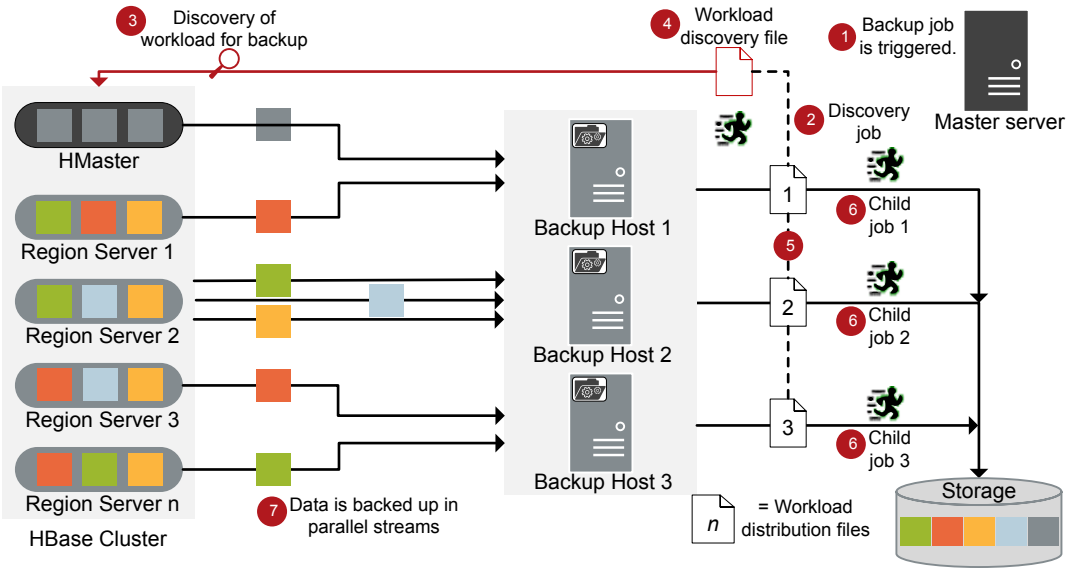
- See “Backing up HBase data” on page 8.
- See “Restoring HBase data” on page 9.
- See “Limitations” on page 12.
- For information about the NetBackup Parallel Streaming Framework (PSF) refer to the *NetBackup Administrator's Guide, Volume I*.

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 master 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.

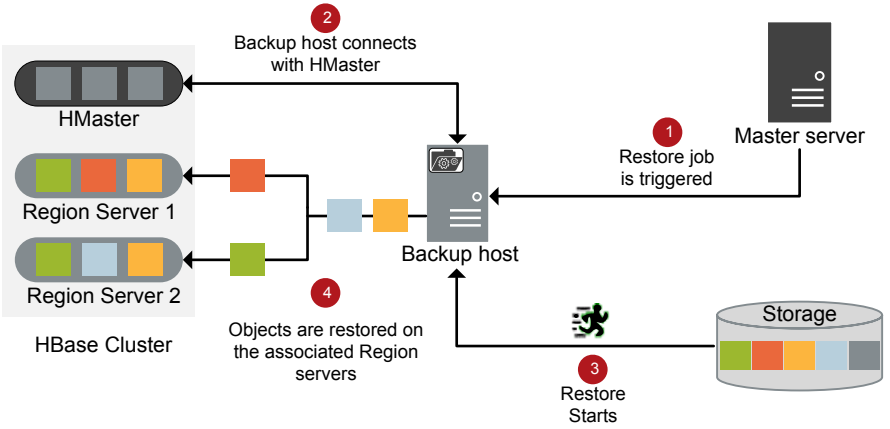
See "About backing up a HBase cluster" on page 37.

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 master 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.

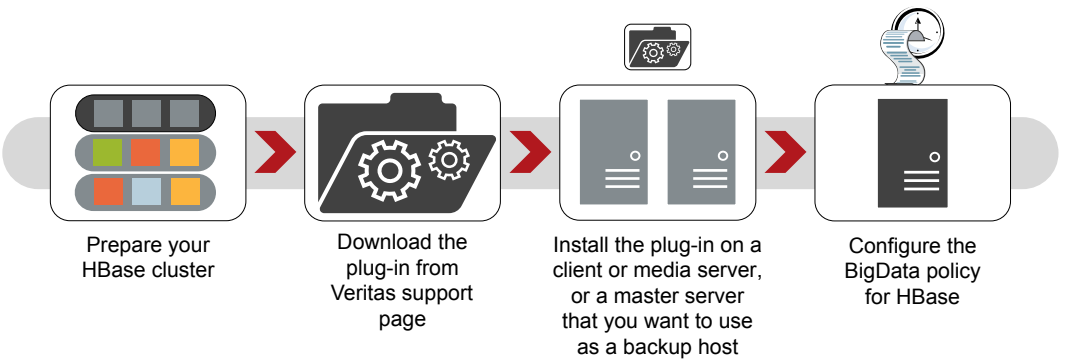
See "About restoring a HBase cluster" on page 40.

Deploying the HBase plug-in

The availability of the plug-in is not tied to any NetBackup release.

The following diagram provides an overview of how you can deploy the plug-in.

Figure 1-4 Deployment flow



As illustrated in the diagram, to deploy the plug-in:

1. Prepare the HBase cluster for NetBackup.
2. Download the plug-in package from *Veritas Entitlement Management System*.
3. Install the plug-in on all the backup hosts.
4. Configure the BigData policy. This policy is introduced to:
   - Specify the application type.
   - Allow backing up of distributed multi-node environments.

- Associate backup hosts.
- Perform workload distribution.
- Specify HBase table names to be protected

See "About installing and deploying the HBase plug-in" on page 15.

See "About configuring NetBackup for HBase" on page 20.

See "Limitations" on page 12.

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 master 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 `tpconfig` 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 `hbase.conf` 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 highly available HBase cluster, if fail-over happens during a backup or restore operation, the job fails.
- If you cancel a backup and restore job manually while the discovery job for a backup operation is in progress, the snapshot entry does not get removed from the HBase web graphical user interface (GUI).

- When you provide credentials colon (:) is not supported.
- HBase plug-in supports HTTP (or non-SSL) configured HBase backup only. Backup job for HBase configured with HTTPS (or SSL) fails.
- Backup of read-only tables is not supported.

Chapter 2

Installing and deploying HBase plug-in for NetBackup

This chapter includes the following topics:

About installing and deploying the HBase plug-in

Table 2-1 Installing and deploying the HBase plug-in

Task	Reference
Pre-requisites and requirements	See "Pre-requisites for installing the HBase plug-in" on page 15.
Preparing the HBase cluster	See "Preparing the HBase cluster" on page 16.
Best practices	See "Best practices for deploying the HBase plug-in" on page 18.
Downloading the HBase plug-in	See "Downloading the plug-in" on page 17.
Performing the installation	See "Installing the HBase plug-in" on page 17.
Verifying the installation	See "Verifying the installation of the HBase plug-in " on page 19.
Configuring	See "About configuring NetBackup for HBase" on page 20.

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 15.
- See "License for HBase plug-in for NetBackup" on page 15.

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.

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 HMaster 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.
- Ensure that HBase service is enabled on the HBase cluster.
- HMaster user should be able to do SSH
- Set the following environment variables for Hmaster in the `.bashrc` file
  - `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/*`

Downloading the plug-in

You can download the following HBase plug-in package for NetBackup from the Veritas Support site.

`NetBackup_PSF_HBase_Plugin_ReleaseVersion-BuildID.tar.gz`

To download the HBase plug-in:

1. Go to https://www.veritas.com/support site.
2. Click **Licensing**. You are directed to the **Veritas Account Manager** page to access your Veritas account.
3. Enter your user credentials to access your Veritas account. You are directed to the *Veritas Entitlement Management System* site.
4. On the **Entitlements** menu, use your **Entitlement IDHBase** to locate and download the following file for HBase plug-in for NetBackup.

   `NetBackup_PSF_HBase_Plugin_ReleaseVersion-BuildID.tar.gz`

   Alternatively, on the **Downloads** menu, locate

   `NetBackup_PSF_HBase_Plugin_ReleaseVersion-BuildID.tar.gz`

   The list of software or plug-in package that is available to download may vary across user accounts based on the entitlements within each account.
5. In the **Actions** column against the software or plug-in package you want to download, click **Download**.
6. Save the downloaded file in a local directory on the intended backup host.

Installing the HBase plug-in

Log on to the Veritas Support site to download the HBase plug-in and store it in a local directory.

Use your **Entitlement ID** to locate the PSF plug-ins and download the `NetBackup_PSF_HBase_Plugin_ReleaseVersion-BuildID.tar.gz` file.

See “Downloading the plug-in” on page 17.

You must install HBase plug-in on all the media servers, clients, or master server you want to use as backup host.

To install HBase plug-in on a backup host

1. Login to the backup host with root privileges.
2. Extract the plug-in. Run the following command:

```
tar -xvf package_name
```

Following files are extracted:

- `plug-in_pkg.linuxR_x86.tar`
- `plug-in_pkg.linuxS_x86.tar`
- `install`
- `LICENSE`

3. Run the installer. Run the following command:

```
./install
```

4. Accept the End User License Agreement (EULA).

Enter **y** for the prompt.

5. Following plugin-in binaries are installed:

- `lusr/openv/lib/libcrypto.so.nb.0.0`
- `usr/openv/lib/libaapipgnhbase.so`
- `usr/openv/lib/lgpl/libssh_threads.so.4.4.2`
- `usr/openv/lib/lgpl/libssh.so.4`
- `usr/openv/lib/lgpl/libssh_threads.so.4`
- `usr/openv/lib/lgpl/README`
- `usr/openv/lib/lgpl/libssh_threads.so`
- `usr/openv/lib/lgpl/libssh.so.4.4.2`
- `usr/openv/lib/lgpl/libssh.so`

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 master 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 server 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 server in FQDN format.
- Ping all the nodes (use FQDN) from the backup hosts.

Post installation procedures

Complete the post-installation procedures:

See "Verifying the installation of the HBase plug-in " on page 19.

See "Configuration for a HBase cluster that uses Kerberos" on page 29.

See "Configuring NetBackup for a highly-available HBase cluster" on page 27.

Verifying the installation of the HBase plug-in

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

- `lusr/openv/lib/libcrypto.so.nb.0.0`
- `usr/openv/lib/libaapipgnhbase.so`
- `usr/openv/lib/lgpl/libssh_threads.so.4.4.2`
- `usr/openv/lib/lgpl/libssh.so.4`
- `usr/openv/lib/lgpl/libssh_threads.so.4`
- `usr/openv/lib/lgpl/README`
- `usr/openv/lib/lgpl/libssh_threads.so`
- `usr/openv/lib/lgpl/libssh.so.4.4.2`
- `usr/openv/lib/lgpl/libssh.so`

File `/usr/openv/tmp/install_trace` contains a trace of the installation. This file can be deleted after you are sure the installation was successful.

Chapter 3

Configuring NetBackup for HBase

This chapter includes the following topics:

About configuring NetBackup for HBase

Table 3-1 Configuring NetBackup for HBase

Task	Reference
Adding backup hosts	See "Managing backup hosts" on page 21. If you want to use NetBackup client as a backup host, you need to whitelist the NetBackup client on the master server. See "Whitelisting a NetBackup client on NetBackup master server" on page 23.

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

Task	Reference
Adding HBase credentials in NetBackup	See "Adding HBase credentials in NetBackup" on page 24.
Configuring the HBase plug-in using the HBase configuration file	See "Configuring the HBase plug-in using the HBase configuration file" on page 26. See "Configuring NetBackup for a highly-available HBase cluster" on page 27.
Configuring the backup hosts for HBase clusters that use Kerberos	See "Configuration for a HBase cluster that uses Kerberos" on page 29.
Configuring NetBackup policies for HBase plug-in	See "Configuring NetBackup policies for HBase plug-in" on page 30.

Managing backup hosts

A backup host acts as a proxy client which hosts all the backup and restore operations for HBase clusters. In case of HBase plug-in for NetBackup, backup host performs all the backup and restore operations without any separate agent installed 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 master server. Veritas recommends that you have 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 master, media, or client can perform the role of a backup host.
- HBase plug-in for NetBackup is installed on all the backup hosts.
- When using multiple backup host, make sure that all backup hosts are communicating with the media server.

You can add a backup host while configuring BigData policy using either the NetBackup Administration Console or Command Line Interface.

For more information on how to create a policy, see See "Creating a BigData backup policy" on page 30.

To add a backup host

1. In the **Backup Selections** tab, click **New** and add the backup host in the following format:

   Backup_Host=<IP_address or hostname>

   For more information on how to create a policy, See "Creating a BigData backup policy" on page 30.

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

   For Windows:

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

   For UNIX:

   ```
   bpplinclude PolicyName -add 'Backup_Host=IP_address or hostname'
   ```

   For more information, See "Using NetBackup Command Line Interface (CLI) to create a BigData policy for HBase clusters " on page 32.

2. As a best practice, add the entries of all the HMaster 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.

To remove a backup host

1. In the **Backup Selections** tab, select the backup host that you want to remove.
2. Right click the selected backup host and click **Delete**.

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

   For Windows:

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

   For UNIX:

   ```
   bpplinclude PolicyName -delete 'Backup_Host=IP_address or
   hostname'
   ```

Whitelisting a NetBackup client on NetBackup master server

To use the NetBackup client as a backup host, you must whitelist it. Perform the Whitelisting procedure on the NetBackup master server .

Whitelisting is a security practice used for restricting systems from running software or applications unless these have been approved for safe execution.

To Whitelist a NetBackup client on NetBackup master server

- Run the following command on the NetBackup master server:
  - For UNIX

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

  - For Windows

    ```
    bpsetconfig -h masterserver
    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.

For more information about the *APP_PROXY_SERVER = clientname*, refer to the *Configuration options for NetBackup clients* section in *NetBackup Administrator's Guide, Volume I*

Veritas NetBackup Documentation

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 and hadoop credentials to the NetBackup master server.

Use the `tpconfig` command to add credentials in NetBackup master 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 fail-over HMaster is the same.
- Use the credentials of the application server that you will use when configuring the BigData policy.
- For a hadoop cluster that uses Kerberos, specify "**kerberos**" as `application_server_user_id` value.
- For a hbase cluster that uses Kerberos, specify the actual Kerberos user name as `application_server_user_id` value.
- RSA key must be in MD5 format.
- Ensure that RSA is supported on the backup host. Run the following command:
  `ssh-keygen -l -f ssh_host_rsa_key`

To obtain the RSA fingerprint for HBase

1. Run the following command:

   ```
   ssh-keyscan -t ssh-rsa application_server_name > tmp.keyhas
   2>/dev/null
   ```

2. Run the following command:

   ```
   ssh-keygen -l -f tmp.keyhas
   ```

3. Copy the RSA fingerprint. You need to provide this fingerprint when you add the HBase credentials.

To authorize HMaster servers without providing RSA key

1. Login to each of your backup hosts as root user.
2. Run the following command to make HMaster as known host.

   Note: Ensure that the `root/.ssh` directory is present.

   Remove any old keys for HMaster from `/root/.ssh/known_hosts` file before running the command.

   ```
   ssh-keyscan -t ssh-rsa <HMASTER_HOST_NAME> >> ~/.ssh/known_hosts
   ```

3. Run the `-tpconfig` command without the RSA fingerprint.

To add credentials in NetBackup

1. Run `tpconfig` command from the following directory paths:

   On UNIX systems, `/usr/openv/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` *`application_type`* `-requiredport` *`IP_port_number`* `[-password` *`password`* `[-key` *`encryption_key`*`]]` command by providing appropriate values for each parameter to add hadoop credentials.

   For example, if you want to add credentials for hadoop server which has *application_server_name* as *Hadoop1*, then run the following command using the appropriate <*user_ID*> and <*password*> details.

   ```
   tpconfig -add -application_server servername -application_type 1
   -application_server_user_id Hadoop -
   ```

4. You are prompted to enter the password.

5 Run the `tpconfig -add -application_server` *`application_server_name`* `-application_server_user_id` *`hmaster`*`:`*`username`*`:`*`RSA fingerprint`* `-application_type` *`application_type`* `-requiredport` *`IP_port_number`* `[-key` *`encryption_key`*`]` command by providing appropriate values for each parameter to add HBase credentials.

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

```
tpconfig -add -application_server servername -application_type 2
-application_server_user_id
hmaster:Hbase:ec:33:1c:28:9b:5f:3b:c8:59:61:a1:35:1c:4b:60:5a
```

6 You are prompted to enter the password. Enter the password as `Hmaster:`*`password`*.

7 Run the `tpconfig -dappservers` command to verify if the NetBackup master 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/openv/netbackup/`. 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 27.

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/openv/netbackup/`

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: Enure 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 24.
- 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 “Pre-requisite for running backup and restore operations for a HBase cluster with Kerberos authentication” on page 38.

Configuring NetBackup policies for HBase plug-in

Backup policies provide the instructions that NetBackup follows to back up clients. For configuring backup policies for HBase plug-in for NetBackup, use the **BigData** policy as the **Policy Type**.

You can create **BigData** policy using either the **NetBackup Administration Console** or the **Command Line Interface**.

For more information on how to create a BigData policy, See “Creating a BigData backup policy” on page 30.

Creating a BigData backup policy

Use the BigData policy to backup big data applications such as HBase clusters.

A BigData policy differs from other policies in the following respects:

- You must specify **BigData** as the policy type.
- The entries which are provided in the **Clients** tab and the **Backup Selections** differ based on the application that you want to back up.
- In the **Backup Selections** tab, you must specify certain parameters and their appropriate values.

Creating BigData policy using the NetBackup Administration Console

If you prefer using the **NetBackup Administration Console** for creating BigData policy, you can use either of the following methods:

- Creating a BigData policy using the **Policy Configuration Wizard**
- Creating a BigData policy using the NetBackup **Policies** utility

The easiest method to set up a **BigData** policy is to use the **Policy Configuration Wizard**. This wizard guides you through the setup process by automatically choosing the best values for most configurations. Not all policy configuration options are presented through the wizard. For example, **calendar-based** scheduling and the **Data Classification** setting. After the policy is created, modify the policy in the Policies utility to configure the options that are not part of the wizard.

Using the Policy Configuration Wizard to create a BigData policy for HBase clusters

Use the following procedure to create a BigData policy with the Policy Configuration Wizard.

To create a BigData policy with the Policy Configuration Wizard

1. In the **NetBackup Administration Console**, in the left pane, click **NetBackup Management**.
2. In the right pane, click **Create a Policy** to begin the **Policy Configuration Wizard**.
3. Select the type of policy to create:
   - **BigData** policy : A policy to backup **Hbase** data
4. Select the storage unit type for BigData policy.
5. Click **Next** to start the wizard and follow the prompts.

   Click **Help** on any wizard panel for assistance while running the wizard.

Using the NetBackup Policies utility to create a BigData policy for HBase clusters

Use the following procedure to create a BigData policy with the NetBackup Policies utility.

To create a BigData policy with the NetBackup Policies utility

1. In the **NetBackup Administration Console**, in the left pane, expand **NetBackup Management > Policies**.
2. On the **Actions** menu, click **New > Policy**.

3 Type a unique name for the new policy in the **Add a New Policy** dialog box.

Click **OK**.

4 On the **Attributes** tab, select **BigData** as the policy type.

5 On the **Attributes** tab, select the storage unit for BigData policy type.

6 On the **Schedules** tab, click **New** to create a new schedule.

You can create a schedule for a **Full Backup**, **Differential Incremental Backup**, or **Cumulative Incremental Backup** for your BigData policy. Once you set the schedule, HBase data is backed up automatically as per the set schedule without any further user intervention.

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.
- File path or the directory to back up
  You can specify multiple file paths.

Note: The directory or folder specified for backup selection while defining BigData Policy with Application_Type=hbase must not contain space or comma in their names.

---

9 Click **OK** to save the changes.

For more information on using NetBackup for big data applications, refer to the Veritas NetBackup documentation page.

Using NetBackup Command Line Interface (CLI) to create a BigData policy for HBase clusters

You can also use the CLI method to create a BigData policy for HBase.

To create a BigData policy using NetBackup CLI method

1 Log on as an Administrator.

2 Navigate to `/usr/openv/netbackup/bin/admincmd`.

3. Create a new BigData policy using the default settings.

```
bppolicynew policyname
```

4. View the details about the new policy using the `-L` option.

```
bpplinfo policyname -L
```

5. Modify and update the policy type as **BigData**.

```
bpplinfo PolicyName -modify -v -M MasterServerName -pt BigData
```

6. Specify the *Application_Type* as HBase.

   For Windows:

```
bpplinclude PolicyName -add "Application_Type=hbase"
```

   For UNIX:

```
bpplinclude PolicyName -add 'Application_Type=hbase
```

   Note: The parameter values for *Application_Type=HBase* are case-sensitive.

7. Specify the backup host on which you want the backup operations to be performed for HBase.

   For Windows:

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

   For UNIX:

```
bpplinclude PolicyName -add 'Backup_Host=IP_address or hostname'
```

   Note: The backup host must be a Linux computer. The backup host can be a NetBackup client or a media server or a master server.

8 Specify the HBase directory or folder name that you want to backup.

For Windows:

```
bpplinclude PolicyName -add "/namespace:table_name"
```

For UNIX:

```
bpplinclude PolicyName -add '/namespace:table_name'
```

Note: Directory or folder used for backup selection while defining BigData Policy with Application_Type=hbase must not contain space or comma in their names.

---

9 Modify and update the policy storage type for BigData policy.

```
bpplinfo PolicyName -residence STUName -modify
```

10 Specify the IP address or the host name of the HMaster for adding the client details.

For Windows:

```
bpplclients PolicyName -M "MasterServerName" -add
"HBaseServerHMaster" "Linux" "RedHat"
```

For UNIX:

```
bpplclients PolicyName -M 'MasterServerName' -add
'HBaseServerHMaster' 'Linux' 'RedHat'
```

11 Assign a schedule for the created BigData policy as per your requirements.

```
bpplsched PolicyName -add Schedule_Name -cal 0 -rl 0 -st
sched_type -window 0 0
```

Here, *sched_type* value can be specified as follows:

Schedule Type	Description
FULL	Full backup
INCR	Differential Incremental backup
CINC	Cumulative Incremental backup
TLOG	Transaction Log
UBAK	User Backup
UARC	User Archive

The default value for *sched_type* is **FULL**.

Once you set the schedule, HBase data is backed up automatically as per the set schedule without any further user intervention.

12 Alternatively, you can also perform a manual backup for HBase data.

For performing a manual backup operation, execute all the steps from Step 1 to Step 11.

13 For a manual backup operation, navigate to `/usr/openv/netbackup/bin`

Initiate a manual backup operation for an existing BigData policy using the following command:

```
bpbackup -i -p PolicyName -s Schedule_Name -S MasterServerName
-t 44
```

Here, `-p` refers to policy, `-s` refers to schedule, `-S` refers to master server, and `-t 44` refers to BigData policy type.

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 webhbase service is enabled on the HBase cluster. See "Preparing the HBase cluster" on page 16.
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 `tpconfig` command to add HBase credentials in NetBackup master server. See "Adding HBase credentials in NetBackup" on page 24.
The backup hosts use the `HBase.conf` file to save the configuration settings of the HBase plug-in. You need to create separate file for each backup host and copy it to `/usr/openv/netbackup/`. You need to create the `HBase.conf` file in JSON format.	With this release, the following plug-in settings can be configured ■ See "Configuring NetBackup for a highly-available HBase cluster" on page 27.
Update the BigData policy with the original HMaster name.	See "Configuring NetBackup policies for HBase plug-in" on page 30.

Chapter 4

Performing backups and restores of HBase

This chapter includes the following topics:

About backing up a HBase cluster

Use the **NetBackup, Backup, Archive, and Restore** console 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 "Pre-requisite for running backup and restore operations for a HBase cluster with Kerberos authentication" on page 38.
Backing up a HBase cluster	See "Backing up a HBase cluster" on page 38.

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

Task	Reference
Best practices	See "Best practices for backing up a HBase cluster" on page 39.
Troubleshooting tips	For discovery and cleanup related logs, review the following log file on the first backup host that triggered the discovery. `/usr/openv/netbackup/logs/nbaapidiscv` For data transfer related logs, search for corresponding backup host (using the hostname) in the log files on the master server. See "About NetBackup for HBase debug logging" on page 49.

Pre-requisite 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/netbackup/nbusers/hbase_mykeytabfile.keytab
hbase@MYCOMPANY.COM
```

Also review the configuration-related information. See "Configuration for a HBase cluster that uses Kerberos" on page 29.

Backing up a HBase cluster

You can either schedule a backup job or run a backup job manually. See, NetBackup Administrator's Guide, Volume I

For overview of the backup process, See "Backing up HBase data" on page 8.

The backup process comprises of the following stages:

1. Pre-processing: In the pre-processing stage, the first backup host that you have configured with the BigData policy, triggers 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 Region server.

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 cancelling the job you must manually delete snapshots from the HBase shell.
- If you take 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 39.

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.

About restoring a HBase cluster

Use the **NetBackup, Backup, Archive, and Restore** console 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 "Pre-requisite for running backup and restore operations for a HBase cluster with Kerberos authentication" on page 38.
Restoring HBase data on the same HMaster or HBase cluster	■ See "Using the Restore Wizard to restore HBase data on the same cluster" on page 41. ■ See "Using the `bprestore` command to restore HBase data on the same HBase cluster" on page 42.
Restoring HBase data to an alternate HMaster or HBase cluster This task can be performed only using the `bprestore` command.	See "Restoring HBase data on an alternate HBase cluster" on page 43.
HBase has a limitation to restore truncated tables. As a workaround, you must restore to archive path.	See "Restoring truncated tables" on page 47.
Best practices	See "Best practices for restoring a HBase cluster" on page 48.
Troubleshooting tips	See "About NetBackup for HBase debug logging" on page 49.

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 following:

- Use the Backup, Archive, and Restore console 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.

Using the Restore Wizard to restore HBase data on the same cluster

This topic describes how to use the Restore Wizard to restore HBase data on the same HBase cluster.

To use the Restore Wizard to perform a restore

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 the destination for restore as per your requirement.
    - Select **Restore everything to its 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.
11. Click **Start Restore**.
12. Verify the restored files.

Using the `bprestore` command to restore HBase data on the same HBase cluster

The `bprestore` command lets you restore a backed up or archived file or list of files. You can also name directories to restore. If you include a directory name, `bprestore` restores all files and subdirectories of that directory. You can exclude a file or a directory path that was previously included in the restore by placing an exclamation mark (!) in front of the file or the directory path (does not apply to NDMP restores). For example, the exclude capability is useful if you want to exclude part of a directory from the restore.

To restore HBase data on the same location as your backup location

1. Log on as an Administrator or root user based on windows or UNIX system respectively.
2. Run the following command on the NetBackup master server by providing appropriate values:

   ```
   bprestore -S master_server -D backup_host-C client -t 44 -L
   progress log -f listfile
   ```

   Where,

   `-S master_server`

   Specifies the name of the NetBackup master 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.

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.

- NetBackup supports redirected restores only using the Command Line Interface (CLI).
- Make sure that you have added the credentials for the alternate HMaster or HBase cluster in NetBackup master server.

To perform redirected restore for HBase

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

Parameter	Value
rename_file	Change `/<namespace:source_table_name>` to `/<namespace:destination_table_name> ALT_APPLICATION_SERVER=<alternate name node>`
listfile	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 master_server -D backup_host-C client -R rename_file -t 44 -L progress log -f listfile` command on the NetBackup master server using the modified values for the mentioned parameters in step 1.

Where,

`-S master_server`

Specifies the name of the NetBackup master 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

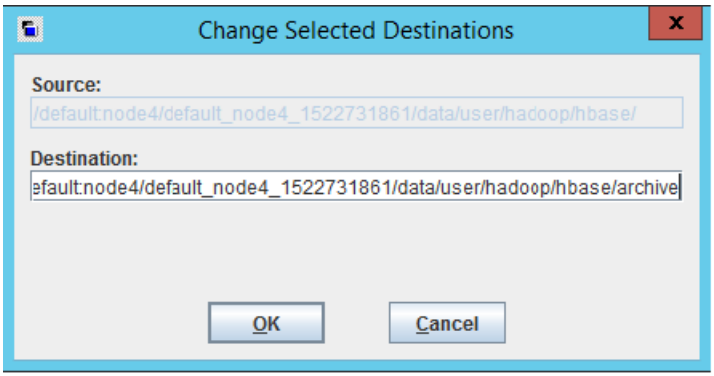
1. Open the **Backup, Archive, and Restore** interface.
2. Select the appropriate date range to restore the complete data set.
3. In the **Browse** directory, specify the root directory (“/”) as the path to browse.
4. From the File menu (Windows) or Actions menu (UNIX), choose **Specify NetBackup Machines 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.

Chapter 5

Troubleshooting

This chapter includes the following topics:

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/netbackup/logs`

Table 5-1 NetBackup logs related to HBase

Log Folder	Messages related to	Logs reside on
`install_path/NetBackup/logs/bpVMutil`	Policy configuration	Master server

Table 5-1 NetBackup logs related to HBase *(continued)*

Log Folder	Messages related to	Logs reside on
install_path/NetBackup/logs/nbaapidiscv	BigData framework, discovery, and HBase configuration file logs	Backup host
install_path/NetBackup/logs/bpbrm	Policy validation, backup, and restore operations	Media server
install_path/NetBackup/logs/bpbkar	Backup	Backup host
install_path/NetBackup/logs/tar	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
```

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 se
```

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.

Index