

# Veritas NetBackup for Hadoop 管理指南

**UNIX、Windows 和 Linux**

**版本 8.1.2**

**VERITAS™**

# Veritas Hadoop 管理指南

上次更新时间： 2018-10-19

文档版本： NetBackup 8.1.2

## 法律声明

Copyright © 2018 Veritas Technologies LLC. © Veritas Technologies LLC 版权所有。All rights reserved. 保留所有权利。

Veritas、Veritas 徽标和 NetBackup 是 Veritas Technologies LLC 或其附属机构在美国和其他国家/地区的商标或注册商标。其他名称可能为其各自所有者的商标，特此声明。

本产品可能包括 Veritas 必须向第三方支付许可费的第三方软件（以下称“第三程序”）。部分第三程序会根据开源或免费软件许可证提供。软件随附的授权许可协议不会改变这些开源或免费软件许可证赋予您的任何权利或义务。请参考此 Veritas 产品随附的或以下链接提供的第三方法律声明文档：

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

本文中介绍的产品根据限制其使用、复制、分发和反编译/逆向工程的许可证进行分发。未经 Veritas Technologies LLC 及其许可方（如果存在）事先书面授权，不得以任何方式任何形式复制本文档的任何部分。

本文档按“现状”提供，对于所有明示或暗示的条款、陈述和保证，包括任何适销性、针对特定用途的适用性或无侵害知识产权的暗示保证，均不提供任何担保，除非此类免责声明的范围在法律上视为无效。Veritas Technologies LLC 不对任何与性能或使用本文档相关的伴随或后果性损害负责。本文档所含信息如有更改，恕不另行通知。

无论由 Veritas 作为内部服务还是托管服务提供，根据 FAR 12.212 中的定义，授权许可的软件和文档被视为“商业计算机软件”，受 FAR Section 52.227-19 “Commercial Computer Software - Restricted Rights”（商业计算机软件受限权利）和 DFARS 227.7202 等

“Commercial Computer Software and Commercial Computer Software Documentation”（商业计算机软件和商业计算机软件文档）中的适用规定，以及所有后续法规中规定的权利的制约。美国政府仅可根据本协议的条款对授权许可的软件和文档进行使用、修改、发布复制、执行、显示或披露。

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

<http://www.veritas.com>

## 技术支持

技术支持负责维护全球的支持中心。所有支持服务将会根据您的支持协议以及当时最新的企业技术支持政策进行交付。有关支持产品和服务以及如何联系技术支持的信息，请访问我们的网站：

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

您可以在下列 URL 上管理 Veritas 帐户信息：

<https://my.veritas.com>

如有关于现有支持协议有任何问题，请按如下所示给您所在区域的支持协议管理团队发送电子邮件：

全球（日本除外）

[CustomerCare@veritas.com](mailto:CustomerCare@veritas.com)

日本

[CustomerCare\\_Japan@veritas.com](mailto:CustomerCare_Japan@veritas.com)

## 文档

请确保您的文档是最新版本。每个文档在第 2 页显示上次更新日期。可以在 Veritas 网站上获取最新文档：

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

## 文档反馈

您的反馈对我们非常重要。请提出您对本文档的改进建议，或者就本文档中的错误或疏漏进行报告。请注明所报告文本的文档标题、文档版本和章节标题。请将您的反馈发送至：

[NB.docs@veritas.com](mailto:NB.docs@veritas.com)

您也可以在以下 Veritas 社区站点中查看相关文档信息或进行提问：

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

## Veritas Services and Operations Readiness Tools (SORT)

Veritas Services and Operations Readiness Tools (SORT) 是一个网站，提供的信息和工具有助于自动处理及简化某些耗时的管理任务。根据具体产品，SORT 会帮助您准备安装和升级、识别您数据中心的风险并提高操作效率。要了解 SORT 为您的产品提供了哪些服务和工具，请参见数据表：

[https://sort.veritas.com/data/support/SORT\\_Data\\_Sheet.pdf](https://sort.veritas.com/data/support/SORT_Data_Sheet.pdf)

# 目录

<b>第 1 章</b>	<b>简介</b> .....	6
	使用 NetBackup 保护 Hadoop 数据 .....	6
	备份 Hadoop 数据 .....	8
	还原 Hadoop 数据 .....	9
	NetBackup for Hadoop 术语 .....	10
	限制 .....	11
<b>第 2 章</b>	<b>部署 Hadoop 插件</b> .....	12
	关于部署 Hadoop 插件 .....	12
	Hadoop 插件的先决条件 .....	12
	操作系统和平台兼容性 .....	13
	NetBackup 服务器和客户端要求 .....	13
	使用 NetBackup 的 Hadoop 插件时需要的许可证 .....	13
	准备 Hadoop 群集 .....	13
	部署 Hadoop 插件的最佳做法 .....	14
	验证 Hadoop 插件的部署 .....	14
<b>第 3 章</b>	<b>配置 NetBackup for Hadoop</b> .....	15
	关于配置 NetBackup for Hadoop .....	15
	管理备份主机 .....	16
	在 NetBackup 主服务器上添加 NetBackup 客户端到白名单 .....	18
	将 NetBackup 设备配置为备份主机 .....	19
	在 NetBackup 中添加 Hadoop 凭据 .....	19
	使用 Hadoop 配置文件配置 Hadoop 插件 .....	20
	为高可用 Hadoop 群集配置 NetBackup .....	21
	配置 Hadoop 群集的自定义端口 .....	23
	为备份主机配置线程数 .....	24
	配置使用 Kerberos 的 Hadoop 群集 .....	24
	为 Hadoop 插件配置 NetBackup 策略 .....	25
	创建 BigData 备份策略 .....	25
	Hadoop 群集的灾难恢复 .....	30

<b>第 4 章</b>	<b>执行 Hadoop 的备份和还原</b> .....	31
	关于备份 Hadoop 群集 .....	31
	对使用 Kerberos 身份验证的 Hadoop 群集运行备份和还原操作的 先决条件 .....	32
	备份 Hadoop 群集 .....	32
	备份 Hadoop 群集的最佳做法 .....	33
	关于还原 Hadoop 群集 .....	33
	在同一 Hadoop 群集上还原 Hadoop 数据 .....	34
	在备用 Hadoop 群集上还原 Hadoop 数据 .....	37
	还原 Hadoop 群集的最佳做法 .....	40
<b>第 5 章</b>	<b>故障排除</b> .....	42
	关于对 NetBackup for Hadoop 问题进行故障排除 .....	42
	关于 NetBackup for Hadoop 调试日志记录 .....	42
	对 Hadoop 数据的备份问题进行故障排除 .....	43
	备份操作失败，并发生错误 112 .....	44
	备份操作失败，并发生错误 6609 .....	44
	备份操作失败，并发生错误 6618 .....	44
	备份操作失败，并发生错误 6647 .....	45
	对于 Hadoop，不会备份或还原扩展属性 (xattrs) 和访问控制列表 (ACL) .....	45
	备份操作失败，并发生错误 6654 .....	46
	备份操作失败，并发生 bpbrm 错误 8857 .....	46
	备份操作失败，并发生错误 6617 .....	46
	备份操作失败，并发生错误 6616 .....	46
	对 Hadoop 数据的还原问题进行故障排除 .....	47
	还原失败，并显示错误代码 2850 .....	47
	针对 Hadoop 的 NetBackup 还原作业部分完成 .....	47
	对于 Hadoop，不会备份或还原扩展属性 (xattrs) 和访问控制列表 (ACL) .....	47
	如果备份主机上缺少 Hadoop 插件文件，还原操作将失败 .....	48
	还原失败，并发生 bpbrm 错误 54932 .....	48
	还原操作失败，并发生 bpbrm 错误 21296 .....	48
<b>索引</b>	.....	49

# 简介

本章节包括下列主题：

- [使用 NetBackup 保护 Hadoop 数据](#)
- [备份 Hadoop 数据](#)
- [还原 Hadoop 数据](#)
- [NetBackup for Hadoop 术语](#)
- [限制](#)

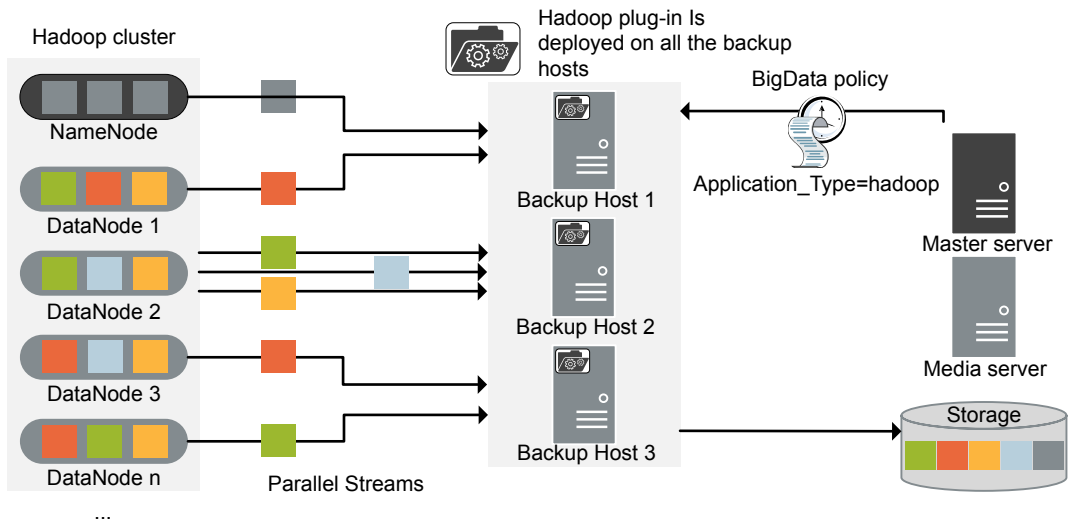
## 使用 NetBackup 保护 Hadoop 数据

现在，通过 NetBackup 并行流框架 (PSF)，可使用 NetBackup 保护 Hadoop 数据。

下图概述了 NetBackup 保护 Hadoop 数据的过程。

另请查看术语定义。请参见第 10 页的“[NetBackup for Hadoop 术语](#)”。

图 1-1 架构概述



如图所示：

- 数据通过并行流进行备份，其中 **DataNode** 将数据块同时流式传输到多台备份主机。由于采用多台备份主机和并行流，因此加快了作业处理速度。
- 通过使用 **NetBackup Plug-in for Hadoop**，支持 Hadoop 群集与 **NetBackup** 进行通信。  
该插件作为 **NetBackup** 安装的一部分进行安装。
- 对于 **NetBackup** 通信，需要配置 **BigData** 策略并添加相关的备份主机。
- 可以将 **NetBackup** 介质服务器、客户端或主服务器配置为备份主机。此外，根据 **DataNode** 的数量，可以添加或删除备份主机。可以通过添加更多的备份主机轻松扩展您的环境。
- **NetBackup** 并行流框架支持进行无代理备份，其中，备份和还原操作在备份主机上运行。群集节点上不存在任何代理信息。此外，**NetBackup** 也不受 Hadoop 群集升级或维护的影响。

更多信息：

- 请参见第 8 页的“[备份 Hadoop 数据](#)”。
- 请参见第 9 页的“[还原 Hadoop 数据](#)”。
- 请参见第 11 页的“[限制](#)”。

- 有关 NetBackup 并行流框架 (PSF) 的信息, 请参考《NetBackup 管理指南, 第 I 卷》。

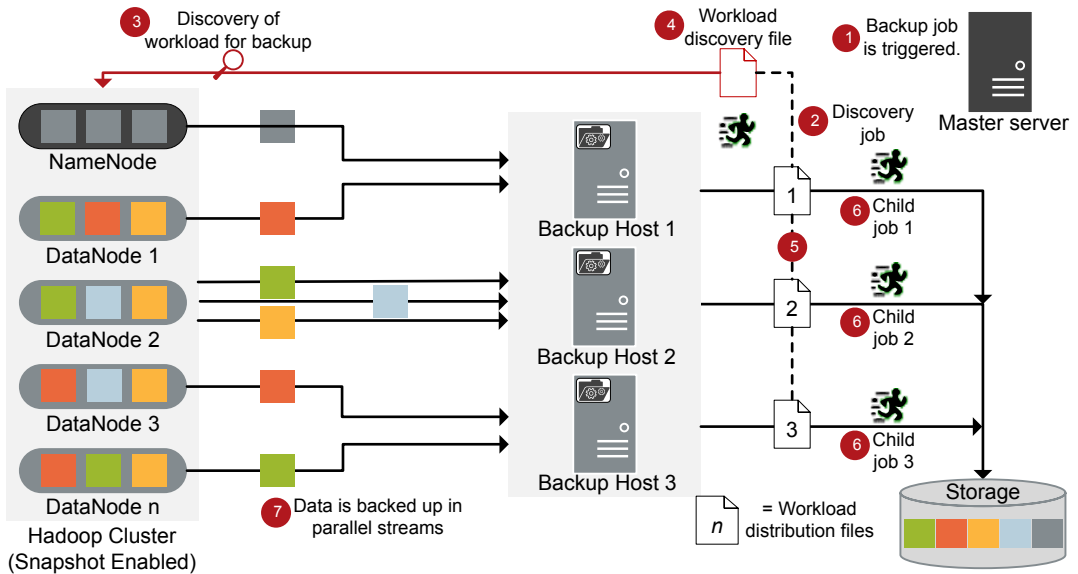
## 备份 Hadoop 数据

Hadoop 数据通过并行流进行备份, 其中 Hadoop DataNode 将数据块同时流式传输到多台备份主机。

**注意:** 在 Hadoop 备份选择中指定的所有目录必须在备份前启用快照。

下图提供了备份流程的概述:

图 1-2 备份流程



如下图所示:

1. 从主服务器触发了预定的备份作业。
2. Hadoop 数据的备份作业是一项复合作业。触发备份作业后, 首先运行发现作业。
3. 在发现期间, 第一台备份主机连接到 NameNode 并执行发现, 以获取需要备份的数据的详细信息。
4. 在备份主机上创建工作量发现文件。该工作量发现文件包含需要从不同 DataNode 备份的数据的详细信息。

5. 备份主机使用该工作量发现文件，并确定工作量在备份主机间的分发方式。为每台备份主机创建工作量分发文件。
6. 针对每台备份主机执行单个子作业。数据按照工作量分发文件中所述进行备份。
7. 数据块从不同 **DataNode** 同时流式传输到多台备份主机。

所有子作业都完成后，该复合备份作业才算完成。子作业完成后，**NetBackup** 会清除 **NameNode** 中的所有快照。在该清理活动完成后，该复合备份作业才算完成。

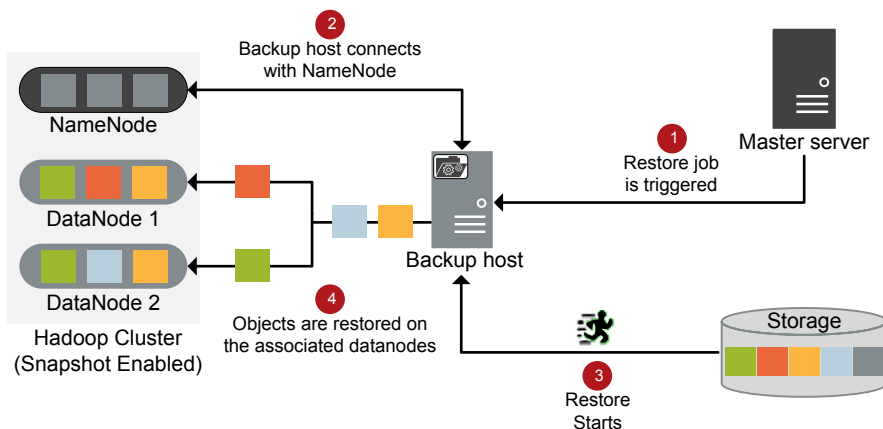
请参见第 31 页的“关于备份 **Hadoop** 群集”。

## 还原 Hadoop 数据

在还原过程中，只使用一台备份主机。

下图提供了还原流程的概述。

图 1-3 还原流程



如图所示：

1. 从主服务器触发了还原作业。
2. 备份主机连接到 **NameNode**。备份主机也是目标客户端。
3. 存储介质中的实际数据还原开始。

4. 数据块在 **DataNode** 上还原。

请参见第 33 页的“关于还原 **Hadoop** 群集”。

## NetBackup for Hadoop 术语

下表定义了在使用 **NetBackup** 保护 **Hadoop** 群集时会遇到的术语。

表 1-1 NetBackup 术语

术语	定义
复合作业	<b>Hadoop</b> 数据的备份作业是一项复合作业。 <ul style="list-style-type: none"><li>■ 备份作业会运行发现作业来获取要备份的数据的信息。</li><li>■ 将为执行实际数据传输的每台备份主机创建子作业。</li><li>■ 备份完成后，该作业会清理 <b>NameNode</b> 上的快照，然后标记为完成。</li></ul>
发现作业	执行备份作业时，会首先创建发现作业。发现作业可与 <b>NameNode</b> 通信，并收集需要备份的数据块及关联的 <b>DataNode</b> 的信息。发现结束后，该作业会填充 <b>NetBackup</b> 随后用于在备份主机之间分发工作量的工作量发现文件。
子作业	在备份过程中，会为每台备份主机创建单独的子作业，用于将数据传输到存储介质。一个子作业可从多个 <b>DataNode</b> 传输数据块。
工作量发现文件	在发现期间，当备份主机与 <b>NameNode</b> 通信时，将创建工作量发现文件。该文件中包含要备份的数据块及关联的 <b>DataNode</b> 的信息。
工作量分发文件	发现完成后， <b>NetBackup</b> 会为每台备份主机创建工作量分发文件。这些文件包含由相应的备份主机传输的数据的信息。
并行流	使用 <b>NetBackup</b> 并行流框架，可以同时通过多台备份主机备份多个 <b>DataNode</b> 的数据块。
备份主机	备份主机充当代理客户端。所有备份和还原操作均通过备份主机执行。可将介质服务器、客户端或主服务器配置为备份主机。备份主机在还原过程中还可以用作目标客户端。
BigData 策略	引入 <b>BigData</b> 策略是为了： <ul style="list-style-type: none"><li>■ 指定应用程序类型。</li><li>■ 允许备份分布式多节点环境。</li><li>■ 关联备份主机。</li><li>■ 执行工作量分发。</li></ul>
应用程序服务器	<b>Namenode</b> 在 <b>NetBackup</b> 中称为应用程序服务器。

术语	定义
主 NameNode	在高可用性方案中，需要使用 <b>BigData</b> 策略和 <code>tpconfig</code> 命令指定一个 NameNode。此 NameNode 称为主 NameNode。
故障转移 NameNode	在高可用性方案中，主 NameNode 以外的其他 NameNode 在 <code>hadoop.conf</code> 文件中更新，称为故障转移 NameNode。

表 1-2 Hadoop 术语

术语	定义
NameNode	NameNode 在还原期间还用作源客户端。
DataNode	DataNode 负责在 Hadoop 中存储实际数据。
启用了快照的目录 (snapshottable)	为目录启用快照后，可在任何目录上创建快照。 <ul style="list-style-type: none"><li>■ 每个启用了快照的目录均可容纳 65,536 个同步快照。对于启用了快照的目录，不存在数量限制。</li><li>■ 管理员可将任何目录设置为启用快照。</li><li>■ 如果启用了快照的目录中有快照，则在删除所有快照前，无法删除或重命名该目录。</li><li>■ 如果目录的某个上级或下级是已启用快照的目录，则该目录无法启用快照。</li></ul>

## 限制

部署 Hadoop 插件之前，请查看以下限制：

- Hadoop 群集和备份主机仅支持 RHEL 和 SUSE 平台。
- Hadoop 群集不支持委派令牌身份验证方法。
- Hadoop 插件在备份期间不会捕获对象的扩展属性 (xattrs) 或访问控制列表 (ACL)，因此不会在还原的文件或文件夹上设置这些内容。
- 对于高可用 Hadoop 群集，如果执行备份或还原操作期间出现故障转移，则作业将失败。
- 如果针对备份操作执行发现作业时手动取消备份作业，则不会从 Hadoop Web 图形用户界面 (GUI) 中删除快照条目。

# 部署 Hadoop 插件

本章节包括下列主题：

- [关于部署 Hadoop 插件](#)
- [Hadoop 插件的先决条件](#)
- [准备 Hadoop 群集](#)
- [部署 Hadoop 插件的最佳做法](#)
- [验证 Hadoop 插件的部署](#)

## 关于部署 Hadoop 插件

Hadoop 插件随 NetBackup 一起安装。请查看以下主题以完成部署。

表 2-1 部署 Hadoop 插件

任务	参考信息
先决条件和要求	请参见第 12 页的“ <a href="#">Hadoop 插件的先决条件</a> ”。
准备 Hadoop 群集	请参见第 13 页的“ <a href="#">准备 Hadoop 群集</a> ”。
最佳做法	请参见第 14 页的“ <a href="#">部署 Hadoop 插件的最佳做法</a> ”。
验证部署	请参见第 14 页的“ <a href="#">验证 Hadoop 插件的部署</a> ”。
配置	请参见第 15 页的“ <a href="#">关于配置 NetBackup for Hadoop</a> ”。

## Hadoop 插件的先决条件

在使用 Hadoop 插件之前，请确保满足以下先决条件：

- 请参见第 13 页的“操作系统和平台兼容性”。
- 请参见第 13 页的“使用 NetBackup 的 Hadoop 插件时需要的许可证”。

## 操作系统和平台兼容性

在此版本中，Hadoop 群集和 NetBackup 备份主机支持 RHEL 和 SUSE 平台。  
有关更多信息，请参见 [NetBackup 主要兼容性列表](#)。

## NetBackup 服务器和客户端要求

对于 NetBackup 服务器，请验证是否满足以下要求：

## 使用 NetBackup 的 Hadoop 插件时需要的许可证

要使用 NetBackup 的 Hadoop 插件执行备份和还原操作，需要应用程序和数据库包许可证。

我们提供了有关如何添加许可证的更多信息。

请参见 [NetBackup 管理指南，第 I 卷](#)

## 准备 Hadoop 群集

要为 NetBackup 准备 Hadoop 群集，请执行以下任务：

- 确保 Hadoop 目录已启用快照。  
要使目录支持快照，请在 NameNode 上运行以下命令：  

```
hdfs dfsadmin -allowSnapshot directory_name
```

---

**注意：**如果目录的某个上级或下级是已启用快照的目录，则该目录无法启用快照。

---

有关更多信息，请参考 Hadoop 文档。

- 更新防火墙设置（默认为端口 50070），以便备份主机可与 Hadoop 群集通信。
- 在所有备份主机的 `/etc/hosts` 文件中添加所有 NameNode 和 DataNode 的条目。必须添加 FQDN 格式的主机名。  
或者  
在 `/etc/resolve.conf` 文件中添加相应的 DNS 条目。
- 确保在 Hadoop 群集上启用了 webhdfs 服务。

## 部署 Hadoop 插件的最佳做法

部署 Hadoop 插件和配置 NetBackup for Hadoop 时，请考虑以下事项：

- 对备份主机、介质服务器和主服务器的主机名使用一致约定。例如，如果使用的主机名为 **hadoop.veritas.com**（FQDN 格式），请在所有位置均使用此格式。
- 在所有备份主机的 `/etc/hosts` 文件中添加所有 **NameNode** 和 **DataNode** 的条目。必须添加 FQDN 格式的主机名。  
或者  
在 `/etc/resolve.conf` 文件中添加相应的 DNS 条目。
- 始终指定 FQDN 格式的 **NameNode** 和 **DataNode**。
- 从备份主机对所有节点（使用 FQDN）执行 ping 操作。
- **NameNode** 的主机名和端口必须与您在 **Hadoop** 群集的 `core-site.xml` 中使用 `http` 地址参数指定的主机名和端口相同。

## 验证 Hadoop 插件的部署

安装 NetBackup 后，将在 `/usr/opensv/lib` 目录中部署下列文件：

- `libaapihdfs.so`
- `libaapiipgnhadoop.so`

文件 `/usr/opensv/tmp/install_trace` 中包含安装跟踪信息。在确认部署成功后，可以删除此文件。

# 配置 NetBackup for Hadoop

本章节包括下列主题：

- [关于配置 NetBackup for Hadoop](#)
- [管理备份主机](#)
- [在 NetBackup 中添加 Hadoop 凭据](#)
- [使用 Hadoop 配置文件配置 Hadoop 插件](#)
- [配置使用 Kerberos 的 Hadoop 群集](#)
- [为 Hadoop 插件配置 NetBackup 策略](#)
- [Hadoop 群集的灾难恢复](#)

## 关于配置 NetBackup for Hadoop

表 3-1 配置 NetBackup for Hadoop

任务	参考信息
添加备份主机	请参见第 16 页的“ <a href="#">管理备份主机</a> ”。 如果要将在 NetBackup 客户端作为备份主机，需要将主服务器上的 NetBackup 客户端添加到白名单。 请参见第 18 页的“ <a href="#">在 NetBackup 主服务器上添加 NetBackup 客户端到白名单</a> ”。
在 NetBackup 中添加 Hadoop 凭据	请参见第 19 页的“ <a href="#">在 NetBackup 中添加 Hadoop 凭据</a> ”。

任务	参考信息
使用 Hadoop 配置文件配置 Hadoop 插件	请参见第 20 页的“使用 Hadoop 配置文件配置 Hadoop 插件”。 请参见第 21 页的“为高可用 Hadoop 群集配置 NetBackup”。 请参见第 24 页的“为备份主机配置线程数”。
为使用 Kerberos 的 Hadoop 群集配置备份主机	请参见第 24 页的“配置使用 Kerberos 的 Hadoop 群集”。
为 Hadoop 插件配置 NetBackup 策略	请参见第 25 页的“为 Hadoop 插件配置 NetBackup 策略”。

## 管理备份主机

备份主机充当代理客户端，用于承载 Hadoop 群集的所有备份和还原操作。对于 NetBackup 的 Hadoop 插件，无需在 Hadoop 群集上安装任何单独代理，备份主机即可执行所有备份和还原操作。

备份主机必须是 Linux 计算机。NetBackup 8.1 版本仅支持 RHEL 和 SUSE 平台作为备份主机。

备份主机可以是 NetBackup 客户端、介质服务器或主服务器。Veritas 建议将介质服务器用作备份主机。

添加备份主机之前，请考虑以下事项：

- 对于备份操作，可以添加一台或多台备份主机。
- 对于还原操作，只能添加一台备份主机。
- 主服务器、介质服务器或客户端可以执行备份主机的角色。
- 在所有备份主机上安装 NetBackup 的 Hadoop 插件。

可以在使用 NetBackup 管理控制台或命令行界面配置 BigData 策略时添加备份主机。

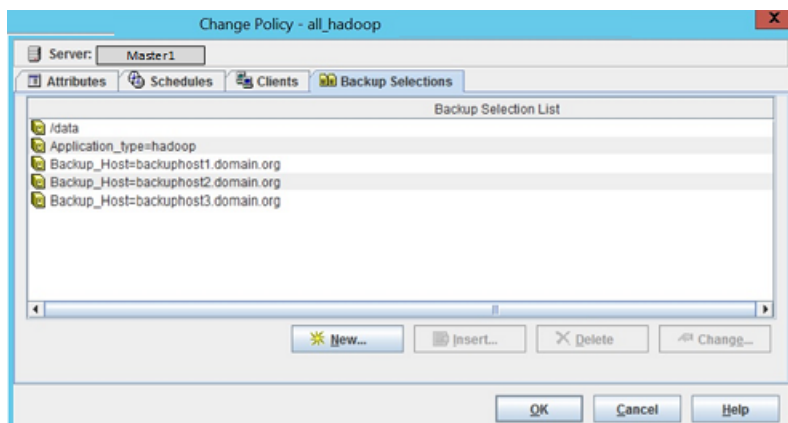
有关如何创建策略的更多信息，请参见第 25 页的“创建 BigData 备份策略”。

## 添加备份主机

- 1 在“备份选择”选项卡上，单击“新建”，然后按以下格式添加备份主机：

*Backup\_Host=<IP\_address or hostname>*

有关如何创建策略的更多信息，请参见第 25 页的“[创建 BigData 备份策略](#)”。



或者，也可以使用以下命令添加备份主机：

对于 Windows：

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

对于 UNIX：

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

有关更多信息，请参见第 27 页的“[使用 NetBackup 命令行界面 \(CLI\) 为 Hadoop 群集创建 BigData 策略](#)”。

- 2 最佳做法是，在所有备份主机的 `/etc/hosts` 文件中添加所有 NameNode 和 DataNode 的条目。必须以 FQDN 格式添加主机名。

或者

在 `/etc/resolve.conf` 文件中添加相应的 DNS 条目。

## 删除备份主机

- 1 在“备份选择”选项卡上，选择要删除的备份主机。
- 2 右键单击选定的备份主机，然后单击“删除”。

或者，也可以使用以下命令删除备份主机：

对于 Windows：

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

对于 UNIX：

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

## 在 NetBackup 主服务器上 将 NetBackup 客户端添加到白名单

要将 NetBackup 客户端用作备份主机，必须将其添加到白名单。在 NetBackup 主服务器上执行添加到白名单的过程。

白名单是一种用于限制系统运行软件或应用程序的安全做法，除非这些软件或应用程序已得到批准可安全执行。

### 在 NetBackup 主服务器上 将 NetBackup 客户端添加到白名单

- ◆ 在 NetBackup 主服务器上运行以下命令：

- 对于 UNIX

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

- 对于 Windows

```
bpsetconfig -h masterserver  
bpsetconfig> APP_PROXY_SERVER = clientname1.domain.org  
bpsetconfig> APP_PROXY_SERVER = clientname2.domain.org  
bpsetconfig>  
Windows systems: <ctl-Z>
```

此命令会在备份配置 (bp.conf) 文件中设置 `APP_PROXY_SERVER = clientname` 条目。

有关 `APP_PROXY_SERVER = clientname` 的更多信息，请参考《NetBackup 管理指南，第 I 卷》中的“NetBackup 客户端的配置选项”部分

[Veritas NetBackup 文档](#)

## 将 NetBackup 设备配置为备份主机

如果要将 NetBackup 设备用作备份主机，请查看以下文章：

- 将 NetBackup Appliance 用作使用 Kerberos 身份验证的 Hadoop 的备份主机  
有关详细信息，请与 Veritas 技术支持联系并让代表参考文章 100039992。
- 将 NetBackup Appliance 用作包含高可用 Hadoop 群集的备份主机  
有关详细信息，请与 Veritas 技术支持联系并让代表参考文章 100039990。

## 在 NetBackup 中添加 Hadoop 凭据

要在 Hadoop 群集与 NetBackup 之间建立无缝通信以成功执行备份和还原操作，必须在 NetBackup 主服务器中添加并更新 Hadoop 凭据。

使用 `tpconfig` 命令在 NetBackup 主服务器中添加 Hadoop 凭据。

有关使用 `tpconfig` 命令删除和更新凭据的参数的信息，请参见 [NetBackup 命令参考指南](#)。

添加 Hadoop 凭据时，请考虑以下事项：

- 对于高可用 Hadoop 群集，确保主 NameNode 和故障转移 NameNode 的用户相同。
- 使用在配置 BigData 策略时将使用的应用程序服务器的凭据。
- 对于使用 Kerberos 的 Hadoop 群集，将 **kerberos** 指定为 `application_server_user_id` 值。
- NameNode 的主机名和端口必须与您在 Hadoop 群集的 `core-site.xml` 中使用 `http` 地址参数指定的主机名和端口相同。
- 对于密码，提供任意随机值。例如，Hadoop。

### 在 NetBackup 中添加 Hadoop 凭据

1 从以下目录路径运行 `tpconfig` 命令：

在 UNIX 系统上： `/usr/opensv/volmgr/bin/`

在 Windows 系统上： `install_path\Volmgr\bin\`

2 运行 `tpconfig --help` 命令。将显示添加、更新和删除 Hadoop 凭据时所需的选项列表。

- 3 运行 `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]]` 命令，并为每个参数提供适当的值以添加 Hadoop 凭据。

例如，如果要为 Hadoop 服务器（该服务器的 `application_server_name` 为 `hadoop1`）添加凭据，则使用相应的 `<user_ID>` 和 `<password>` 详细信息运行以下命令。

```
tpconfig -add -application_server hadoop1 -application_type 1
-application_server_user_id Hadoop -password Hadoop
```

其中，为 `-application_type` 参数指定的数值 1 对应于 Hadoop。

- 4 运行 `tpconfig -dappservers` 命令，以验证 NetBackup 主服务器是否添加了 Hadoop 凭据。

## 使用 Hadoop 配置文件配置 Hadoop 插件

备份主机使用 `hadoop.conf` 文件保存 Hadoop 插件的配置设置。需要为每台备份主机创建单独的文件，并将其复制到 `/usr/opensv/netbackup/`。需要手动创建 JSON 格式的 `hadoop.conf` 文件。默认情况下，此文件不可用于安装程序。

---

**注意：**不得为任何参数提供空值，否则备份作业将失败。

---

在此版本中，可以配置以下插件设置：

- 请参见第 21 页的“为高可用 Hadoop 群集配置 NetBackup”。
- 请参见第 23 页的“配置 Hadoop 群集的自定义端口”。
- 请参见第 24 页的“为备份主机配置线程数”。

以下是 `hadoop.conf` 文件的示例。

---

**注意：**对于非 HA 环境，不需要故障转移参数。

---

```
{
  "application_servers":
  {
    "hostname_of_the_primary_namenode":
    {
      "failover_namenodes":
```

```
[
  {
    "hostname":"hostname_of_failover_namenode",
    "port":port_of_the_failover_namenode
  }
],
"port":port_of_the_primary_namenode
}
},
"number_of_threads":number_of_threads
}
```

## 为高可用 Hadoop 群集配置 NetBackup

要保护高可用 Hadoop 群集，在配置 NetBackup for Hadoop 群集时，请执行以下操作：

- 在 **BigData** 策略中指定其中一个 **NameNode**（主 **NameNode**）作为客户端。
- 执行 `tpconfig` 命令时，将同一 **NameNode**（主 **NameNode** 和故障转移 **NameNode**）指定为应用程序服务器。
- 创建 `hadoop.conf` 文件，使用 **NameNode**（主 **NameNode** 和故障转移 **NameNode**）的详细信息更新该文件，然后将其复制到所有备份主机。  
`hadoop.conf` 文件为 **JSON** 格式。
- **NameNode** 的主机名和端口必须与您在 Hadoop 群集的 `core-site.xml` 中使用 `http` 地址参数指定的主机名和端口相同。
- 主 **NameNode** 和故障转移 **NameNode** 的用户名必须相同。
- 不要为任何参数提供空值，否则备份作业将失败。

## 更新高可用 Hadoop 群集的 `hadoop.conf` 文件

1 使用以下参数更新 `hadoop.conf` 文件：

```
{
  "application_servers":
  {
    "hostname_of_primary_namenode1":
    {
      "failover_namenodes":
      [
        {
          "hostname": "hostname_of_failover_namenode1",
          "port": port_of_failover_namenode1
        }
      ],
      "port":port_of_primary_namenode1
    }
  }
}
```

- 2 如果有多个 Hadoop 群集，请使用相同的 `hadoop.conf` 文件更新详细信息。例如，

```
{
  "application_servers":
  {
    "hostname_of_primary_namenode1":
    {
      "failover_namenodes":
      [
        {
          "hostname": "hostname_of_failover_namenode1",
          "port": port_of_failover_namenode1
        }
      ],
      "port": port_of_primary_namenode1
    },
    "hostname_of_primary_namenode2":
    {
      "failover_namenodes":
      [
        {
          "hostname": "hostname_of_failover_namenode2",
          "port": port_of_failover_namenode2
        }
      ],
      "port": port_of_primary_namenode2
    }
  }
}
```

- 3 将此文件复制到所有备份主机的以下位置：

```
/usr/opensv/netbackup/
```

## 配置 Hadoop 群集的自定义端口

可以使用 Hadoop 配置文件配置自定义端口。默认情况下，NetBackup 使用端口 50070。

## 配置 Hadoop 群集的自定义端口

1 使用以下参数更新 `hadoop.conf` 文件:

```
{
  "application_servers": {
    "hostname_of_namenode1":{

      "port":port_of_namenode1
    }
  }
}
```

2 将此文件复制到所有备份主机的以下位置:

```
/usr/opensv/netbackup/
```

## 为备份主机配置线程数

要增强备份性能，可以配置每台备份主机允许的线程（流）数。可以通过添加更多备份主机或增加每台备份主机的线程数来提高备份性能。

要确定线程数，请考虑以下事项:

- 默认值为 4。
- 可为每台备份主机设置最少 1 个线程，最多 32 个线程。
- 每台备份主机可以配置不同的线程数。
- 配置线程数时，请考虑可用的核心数和要使用的核心数。最佳做法是，应为每个核心配置 1 个线程。例如，如果有 8 个核心可用，并且您要使用 4 个核心，则应配置 4 个线程。

### 更新 `hadoop.conf` 文件以配置线程数

1 使用以下参数更新 `hadoop.conf` 文件:

```
{
  "number_of_threads": number_of_threads
}
```

2 将此文件复制到备份主机上的以下位置:

```
/usr/opensv/netbackup/
```

## 配置使用 Kerberos 的 Hadoop 群集

对于使用 Kerberos 的 Hadoop 群集，请在所有备份主机上执行以下任务:

- 确保所有备份主机上都存在 Kerberos 软件包。
  - 适用于 RHEL 的 krb5-workstation 软件包
  - 适用于 SUSE 的 krb5-client
- 获取 keytab 文件并将其复制到备份主机上的安全位置。
- 确保 keytab 具有所需的主体。
- 使用相应的 KDC 服务器和领域详细信息手动更新 krb5.conf 文件。

---

**注意：**确保 default\_ccache\_name 参数未设置为 **KEYRING:persistent:%{uid}** 值。可以注释该参数以使用默认值，也可以指定文件名，例如 **FILE:/tmp/krb\_file\_name:%{uid}**。

---

- 在 NetBackup 中添加 Hadoop 凭据时，请将 **kerberos** 指定为 application\_server\_user\_id 值。请参见第 19 页的“在 NetBackup 中添加 Hadoop 凭据”。
- 要对使用 Kerberos 身份验证的 Hadoop 群集运行备份和还原操作，Hadoop 需要有效的 Kerberos 票证授予票证 (TGT) 来对 Hadoop 群集进行身份验证。请参见第 32 页的“对使用 Kerberos 身份验证的 Hadoop 群集运行备份和还原操作的先决条件”。

## 为 Hadoop 插件配置 NetBackup 策略

备份策略提供 NetBackup 备份客户端时所遵循的说明。要为 NetBackup 的 Hadoop 插件配置备份策略，请使用 **BigData** 策略作为“策略类型”。

可以使用“**NetBackup 管理控制台**”或“**命令行界面**”创建 **BigData** 策略。

---

**注意：**NameNode 的主机名和端口必须与您在 Hadoop 群集的 core-site.xml 中使用 http 地址参数指定的主机名和端口相同。

---

有关如何创建 BigData 策略的更多信息，请参见第 25 页的“创建 BigData 备份策略”。

### 创建 BigData 备份策略

使用 BigData 策略备份大数据应用程序，例如 Hadoop 群集。

在以下方面，BigData 策略与其他策略存在差别：

- 必须将 **BigData** 指定为策略类型。

- “客户端”选项卡和“备份选择”中提供的条目因要备份的应用程序而异。
- 在“备份选择”选项卡中，必须指定某些参数及其适当值。

## 使用 NetBackup 管理控制台创建 BigData 策略

如果希望使用“NetBackup 管理控制台”创建 BigData 策略，可以使用以下方法之一：

- 使用“策略配置向导”创建 BigData 策略
- 使用 NetBackup “策略”实用程序创建 BigData 策略

设置 BigData 策略的最简单方法是使用“策略配置向导”。此向导通过自动选择多数配置的最佳值引导您完成设置过程。并非所有策略配置选项都会通过向导显示。例如，“基于日历的日程表”和“数据分类”设置。创建策略后，请在“策略”实用程序中修改策略，以配置向导中不存在的选项。

### 使用“策略配置向导”为 Hadoop 群集创建 BigData 策略

使用以下过程可通过“策略配置向导”创建 BigData 策略。

#### 利用“策略配置向导”创建 BigData 策略

- 1 在“NetBackup 管理控制台”的左窗格中，单击“NetBackup 管理”。
- 2 在右窗格中，单击“创建策略”以开始使用“策略配置向导”。
- 3 选择要创建的策略类型：
  - BigData 策略：备份 Hadoop 数据的策略
- 4 选择 BigData 策略的存储单元类型。
- 5 单击“下一步”启动向导，然后按照提示执行操作。

运行该向导时，单击任何向导面板上的“帮助”获取帮助。

### 使用 NetBackup 策略实用程序为 Hadoop 群集创建 BigData 策略

使用以下过程可通过 NetBackup 策略实用程序创建 BigData 策略。

#### 利用 NetBackup 策略实用程序创建 BigData 策略

- 1 在 NetBackup 管理控制台的左窗格中，展开“NetBackup 管理” > “策略”。
- 2 在“操作”菜单上，单击“新建” > “策略”。
- 3 在“添加新策略”对话框中，为新策略键入唯一的名称。  
单击“确定”。

- 4 在“属性”选项卡上，选择 **BigData** 作为策略类型。
- 5 在“属性”选项卡上，选择 **BigData** 策略类型的存储单元。
- 6 在“日程表”选项卡上，单击“新建”创建新的日程表。

可以为 **BigData** 策略创建“完全备份”、“差异增量式备份”或“累积增量式备份”的日程表。设置日程表后，会根据设置的日程表自动备份 **Hadoop** 数据，而无需任何进一步的用户干预。

- 7 在“客户端”选项卡上，输入 **NameNode** 的 IP 地址或主机名。
- 8 在“备份选择”选项卡上，输入以下参数及其值，如下所示：

- **Application\_Type=hadoop**  
这些参数值区分大小写。
- **Backup\_Host=IP\_address or hostname**  
备份主机必须是 **Linux** 计算机。备份主机可以是 **NetBackup** 客户端或介质服务器。  
您可以指定多台备份主机。
- 要备份的文件路径或目录  
您可以指定多个文件路径。

---

**注意：**使用 **Application\_Type=hadoop** 定义 **BigData** 策略时为备份选择指定的目录或文件夹的名称中不得包含空格或逗号。

---

- 9 单击“确定”保存更改。

有关将 **NetBackup** 用于大数据应用程序的更多信息，请参考 [Veritas NetBackup 文档](#) 页面。

## 使用 NetBackup 命令行界面 (CLI) 为 Hadoop 群集创建 BigData 策略

您还可以使用 CLI 方法为 **Hadoop** 创建 **BigData** 策略。

### 使用 NetBackup CLI 方法创建 BigData 策略

- 1 以管理员身份登录。
- 2 导航到 `/usr/opensv/netbackup/bin/admincmd`。
- 3 使用默认设置创建新的 **BigData** 策略。

```
bppolicynew policyname
```

- 4 使用 `-L` 选项查看有关新策略的详细信息。

```
bpplinfo policyname -L
```

- 5 修改策略类型并将其更新为 **BigData**。

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

- 6 将 *Application\_Type* 指定为 Hadoop。

对于 Windows:

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

对于 UNIX:

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

---

**注意:** *Application\_Type=hadoop* 的参数值区分大小写。

---

- 7 指定要在其上执行 Hadoop 的备份操作的备份主机。

对于 Windows:

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

对于 UNIX:

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

---

**注意:** 备份主机必须是 Linux 计算机。备份主机可以是 NetBackup 客户端、介质服务器或主服务器。

---

- 8 指定要备份的 Hadoop 目录或文件夹名称。

对于 Windows:

```
bpplinclude PolicyName -add "/hdfsfoldername"
```

对于 UNIX:

```
bpplinclude PolicyName -add '/hdfsfoldername'
```

---

**注意:** 使用 *Application\_Type=hadoop* 定义 BigData 策略时用于备份选择的目录或文件夹的名称不得包含空格或逗号。

---

- 9 修改并更新 BigData 策略的策略存储类型。

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

- 10 指定用于添加客户端详细信息的 **NameNode** 的 IP 地址或主机名。

对于 **Windows**:

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

对于 **UNIX**:

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

- 11 根据要求为所创建的 **BigData** 策略分配日程表。

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

其中, *sched\_type* 值可以按如下所示指定:

日程表类型	描述
FULL	完全备份
INCR	差异增量式备份
CINC	累积增量式备份
TLOG	事务日志
UBAK	用户备份
UARC	用户存档

*sched\_type* 的默认值为 **FULL**。

设置日程表后, 会根据设置的日程表自动备份 **Hadoop** 数据, 而无需任何进一步的用户干预。

- 12 或者, 也可以手动备份 **Hadoop** 数据。

要执行手动备份操作, 请执行所有步骤 (步骤 1 到步骤 11)。

- 13 对于手动备份操作, 请导航到 `/usr/opensv/netbackup/bin`

使用以下命令为现有 **BigData** 策略启动手动备份操作:

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

其中, `-p` 指的是策略, `-s` 指的是日程表, `-S` 指的是主服务器, `-t 44` 指的是 **BigData** 策略类型。

## Hadoop 群集的灾难恢复

要对 Hadoop 群集进行灾难恢复，请执行以下任务：

表 3-2 执行灾难恢复

任务	描述
启动 Hadoop 群集和节点后，为 NetBackup 操作准备群集。	执行以下任务：  更新防火墙设置，以便备份主机可与 Hadoop 群集通信。  确保在 Hadoop 群集上启用了 webhdfs 服务。 请参见第 13 页的“准备 Hadoop 群集”。
要在 Hadoop 群集与 NetBackup 之间建立无缝通信以成功执行备份和还原操作，必须在 NetBackup 主服务器中添加并更新 Hadoop 凭据。	使用 <code>tpconfig</code> 命令可在 NetBackup 主服务器中添加 Hadoop 凭据。  请参见第 19 页的“在 NetBackup 中添加 Hadoop 凭据”。
备份主机使用 <code>hadoop.conf</code> 文件保存 Hadoop 插件的配置设置。需要为每台备份主机创建单独的文件，并将其复制到 <code>/usr/opensv/netbackup/</code> 。需要创建 JSON 格式的 <code>hadoop.conf</code> 文件。	在此版本中，可以配置以下插件设置 <ul style="list-style-type: none"><li>■ 请参见第 21 页的“为高可用 Hadoop 群集配置 NetBackup”。</li><li>■ 请参见第 24 页的“为备份主机配置线程数”。</li></ul>
使用原始 NameNode 名称更新 BigData 策略。	请参见第 25 页的“为 Hadoop 插件配置 NetBackup 策略”。

# 执行 Hadoop 的备份和还原

本章节包括下列主题：

- [关于备份 Hadoop 群集](#)
- [关于还原 Hadoop 群集](#)

## 关于备份 Hadoop 群集

使用 NetBackup “备份、存档和还原” 控制台管理备份操作。

表 4-1 备份 Hadoop 数据

任务	参考信息
了解进程	请参见第 8 页的“ <a href="#">备份 Hadoop 数据</a> ”。
(可选) 完成 Kerberos 的先决条件	请参见第 32 页的“ <a href="#">对使用 Kerberos 身份验证的 Hadoop 群集运行备份和还原操作的先决条件</a> ”。
备份 Hadoop 群集	请参见第 32 页的“ <a href="#">备份 Hadoop 群集</a> ”。
最佳做法	请参见第 33 页的“ <a href="#">备份 Hadoop 群集的最佳做法</a> ”。

任务	参考信息
故障排除提示	<p>要发现和清理相关日志，请在触发生效的第一台备份主机上查看以下日志文件。</p> <pre>/usr/opensv/netbackup/logs/nbaapidiscv</pre> <p>对于数据传输相关的日志，请在主服务器上的日志文件中搜索相应的备份主机（使用主机名）。</p> <p>请参见第 43 页的“对 Hadoop 数据的备份问题进行故障排除”。</p>

## 对使用 Kerberos 身份验证的 Hadoop 群集运行备份和还原操作的先决条件

要对使用 Kerberos 身份验证的 Hadoop 群集运行备份和还原操作，Hadoop 需要有效的 Kerberos 票证授予票证 (TGT) 来对 Hadoop 群集进行身份验证。

---

**注意：**在备份和还原操作过程中，TGT 必须有效。因此，请相应地指定 TGT 有效期或在操作过程中根据需要续订 TGT。

---

运行以下命令以生成 TGT：

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

例如，

```
kinit -k -t /usr/opensv/netbackup/nbusers/hdfs_mykeytabfile.keytab
hdfs@MYCOMPANY.COM
```

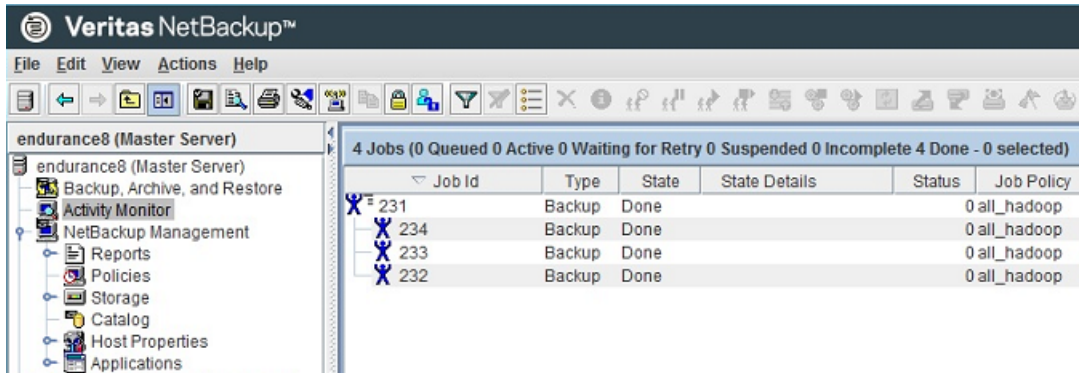
另请查看与配置相关的信息。请参见第 24 页的“配置使用 Kerberos 的 Hadoop 群集”。

## 备份 Hadoop 群集

可以调度备份作业或手动运行备份作业。请参见 [NetBackup 管理指南](#)，第 I 卷有关备份过程的概述，请参见第 8 页的“备份 Hadoop 数据”。

备份过程包括以下阶段：

1. **预处理：**在预处理阶段，已使用 **BigData** 策略配置的第一台备份主机将触发发现操作。在此阶段，将生成完整备份选择的快照。**NameNode Web** 界面中将显示快照详细信息。
2. **数据传输：**在数据传输过程中，会对每台备份主机创建一个子作业。



3. 后处理：在后处理过程中，NetBackup 将清理 NameNode 上的快照。

## 备份 Hadoop 群集的最佳做法

在备份 Hadoop 群集之前，请考虑以下事项：

- 要备份整个 Hadoop 文件系统，请提供“/”作为备份选择，并确保“/”已启用快照。
- 执行备份作业之前，请确保从备份主机针对所有节点的主机名 (FQDN) 执行的 ping 操作返回成功响应。
- 更新防火墙设置，以便备份主机可与 Hadoop 群集通信。

## 关于还原 Hadoop 群集

使用 NetBackup “备份、存档和还原” 控制台管理还原操作。

表 4-2 还原 Hadoop 数据

任务	参考信息
了解进程	请参见第 9 页的“还原 Hadoop 数据”。
完成 Kerberos 的先决条件	请参见第 32 页的“对使用 Kerberos 身份验证的 Hadoop 群集运行备份和还原操作的先决条件”。
在同一 NameNode 或 Hadoop 群集上还原 Hadoop 数据	<ul style="list-style-type: none"> <li>■ 请参见第 34 页的“使用“还原向导”在同一 Hadoop 群集上还原 Hadoop 数据”。</li> <li>■ 请参见第 35 页的“使用 bprestore 命令在同一 Hadoop 群集上还原 Hadoop 数据”。</li> </ul>

任务	参考信息
将 Hadoop 数据还原到备用 NameNode 或 Hadoop 群集  只能使用 bprestore 命令执行此任务。	请参见第 37 页的“在备用 Hadoop 群集上还原 Hadoop 数据”。
最佳做法	请参见第 40 页的“还原 Hadoop 群集的最佳做法”。
故障排除提示	请参见第 47 页的“对 Hadoop 数据的还原问题进行故障排除”。

## 在同一 Hadoop 群集上还原 Hadoop 数据

要在同一 Hadoop 群集上还原 Hadoop 数据，请考虑以下事项：

- 使用“备份、存档和还原”控制台启动 Hadoop 数据还原操作。使用此界面可选择从其还原对象的 NetBackup 服务器以及要浏览其备份映像的客户端。基于这些选择，您可以浏览备份映像历史记录、选择单个项目并启动还原。
- 还原浏览器用于显示 Hadoop 目录对象。还原浏览器提供分层显示，可供选择要还原的对象。通过展开单个目录可显示构成 Hadoop 群集的对象（Hadoop 目录或文件）。
- 管理员可以浏览和还原 Hadoop 目录和单个项目。用户可以还原的对象包括 Hadoop 文件和文件夹。

### 使用“还原向导”在同一 Hadoop 群集上还原 Hadoop 数据

本主题介绍如何使用“还原向导”在同一 Hadoop 群集上还原 Hadoop 数据。

#### 使用“还原向导”执行还原

- 1 打开“备份、存档和还原”界面。
- 2 选择适当的日期范围以还原完整的数据集。
- 3 在“浏览”目录中，指定根目录（“/”）作为浏览路径。
- 4 从“文件”菜单(Windows)或“操作”菜单(UNIX)中，选择“指定 NetBackup 计算机和策略类型”。
- 5 在“指定 NetBackup 计算机和策略类型”向导中，输入用于还原的源和目标详细信息。
  - 指定 Hadoop NameNode 作为要执行还原操作的源。

从“用于还原的源客户端”列表中，选择所需的 NameNode。

- 指定备份主机作为目标客户端。  
从“用于还原的目标客户端”列表中，选择所需的备份主机。
  - 在“指定 NetBackup 计算机和策略类型”向导中，输入用于还原的策略类型详细信息。  
从“用于还原的策略类型”列表中，选择 **BigData** 作为用于还原的策略类型。  
单击“确定”。
- 6 转至“备份历史记录”，选择要还原的备份映像。
  - 7 在“目录结构”窗格中，展开“目录”。  
该目录下的所有后续文件和文件夹都显示在“选定目录的内容”窗格中。
  - 8 在“选定目录的内容”窗格中，选择要还原的 Hadoop 文件对应的复选框。
  - 9 单击“还原”。
  - 10 在“还原标记的文件”对话框中，根据您的要求选择还原目标。
    - 如果要将文件还原到执行备份的相同位置，请选择“将所有内容还原到其原始位置”。
    - 如果要将文件还原到备份位置以外的其他位置，请选择“将所有内容还原到其他位置”。
  - 11 单击“开始还原”。
  - 12 验证还原的文件。

## 使用 `bprestore` 命令在同一 Hadoop 群集上还原 Hadoop 数据

可以通过 `bprestore` 命令还原已备份或存档的文件或文件列表。您也可以命名要还原的目录。如果包括一个目录名，`bprestore` 将还原此目录中的所有文件和子目录。可以使用以下方法排除在以前的还原操作中包括的文件或目录路径：在文件或目录路径的前面放置一个感叹号 (!)（不适用于 NDMP 还原）。例如，如果您希望在还原时排除目录的某一部分，这种排除功能将非常有用。

## 在与备份位置相同的位置还原 Hadoop 数据

- 1 分别基于 Windows 或 UNIX 系统以管理员身份或 root 用户身份登录。
- 2 通过提供相应的值，在 NetBackup 主服务器上运行以下命令：

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

其中，

```
-S master_server
```

指定 NetBackup 主服务器的名称。

```
-D backup host
```

指定备份主机的名称。

```
-C client
```

将 NameNode 指定为源，以用于查找从中还原文件的备份或存档。该名称必须与它在 NetBackup 目录库中显示的名称相同。

```
-f listfile
```

指定一个文件 (listfile)，其中包含要还原的文件列表，可用于替代文件名选项。在 listfile 中列出所有文件路径，每个路径必须单独占一行。

```
-L progress_log
```

指定要在其中写入进度信息的白名单文件路径的名称。

```
-t 44
```

指定 BigData 作为策略类型。

## 在备用位置还原 Hadoop 数据

- 1 以管理员身份登录。
- 2 通过提供相应的值，在 NetBackup 主服务器上运行以下命令：

```
bprestore -S master_server -D backup_host -C client -t 44 -L  
progress_log -R rename_file -f listfile
```

其中，

```
-S master_server
```

指定 NetBackup 主服务器的名称。

```
-D backup host
```

指定备份主机的名称。

```
-C client
```

将 NameNode 指定为源，以用于查找从中还原文件的备份或存档。该名称必须与它在 NetBackup 目录库中显示的名称相同。

```
-f listfile
```

指定一个文件 (listfile)，其中包含要还原的文件列表，可用于替代文件名选项。在 listfile 中列出所有文件路径，每个路径必须单独占一行。

```
-L progress_log
```

指定要在其中写入进度信息的白名单文件路径的名称。

```
-t 44
```

指定 BigData 作为策略类型。

```
-R rename_file
```

为备用路径还原操作指定更改后的文件名。

将 /<source\_folder\_path> 更改为 /<destination\_folder\_path>

## 在备用 Hadoop 群集上还原 Hadoop 数据

使用 NetBackup 可将 Hadoop 数据还原到另一个 NameNode 或 Hadoop 群集。此类型的还原方法也称为重定向还原。

---

**注意：**NetBackup 仅支持使用命令行界面 (CLI) 执行重定向还原。

---

---

**注意：**确保已在 NetBackup 主服务器中添加备用 NameNode 或 Hadoop 群集的凭据，并已在 NetBackup 主服务器上完成添加到白名单任务。有关如何在 NetBackup 中添加 Hadoop 凭据以及添加到白名单过程的更多信息，请参见第 19 页的“[在 NetBackup 中添加 Hadoop 凭据](#)”。请参见第 18 页的“[在 NetBackup 主服务器上将 NetBackup 客户端添加到白名单](#)”。

---

## 为 Hadoop 执行重定向还原

1 修改 *rename\_file* 和 *listfile* 的值，如下所示：

参数	值
<i>rename_file</i>	将 /<source_folder_path> 更改为 /<destination_folder_path> ALT_APPLICATION_SERVER=<alternate name node>
<i>listfile</i>	要还原的所有 Hadoop 文件的列表

- 2 在 NetBackup 主服务器上使用步骤 1 中所提及参数的修改值运行 `bprestore` `-S master_server -D backup_host -C client -R rename_file -t 44 -L progress_log -f listfile` 命令。

其中，

`-S master_server`

指定 NetBackup 主服务器的名称。

`-D backup_host`

指定备份主机的名称。

`-C client`

将 NameNode 指定为源，以用于查找从中还原文件的备份或存档。该名称必须与它在 NetBackup 目录库中显示的名称相同。

`-f listfile`

指定一个文件 (listfile)，其中包含要还原的文件列表，可用于替代文件名选项。在 listfile 中列出所有文件路径，每个路径必须单独占一行。

`-L progress_log`

指定要在其中写入进度信息的白名单文件路径的名称。

`-t 44`

指定 BigData 作为策略类型。

`-R rename_file`

为备用路径还原操作指定更改后的文件名。

重命名文件中的条目使用以下形式：

`change backup_filepath to restore_filepath`

`ALT_APPLICATION_SERVER=<Application Server Name>`

文件路径必须以 / (斜杠) 开头。

---

**注意：** 确保已将 NetBackup 安装路径中不包含的所有文件路径 (如 `<rename_file_path>`、`<progress_log_path>`) 添加到白名单中。

---

## 还原 Hadoop 群集的最佳做法

还原 Hadoop 群集时，请考虑以下事项：

- 执行还原作业之前，请确保群集上有足够的空间来完成还原作业。

- 更新防火墙设置，以便备份主机可与 Hadoop 群集通信。

# 故障排除

本章节包括下列主题：

- [关于对 NetBackup for Hadoop 问题进行故障排除](#)
- [关于 NetBackup for Hadoop 调试日志记录](#)
- [对 Hadoop 数据的备份问题进行故障排除](#)
- [对 Hadoop 数据的还原问题进行故障排除](#)

## 关于对 NetBackup for Hadoop 问题进行故障排除

表 5-1 对 NetBackup for Hadoop 问题进行故障排除

区域	参考
常规日志记录和调试	请参见第 42 页的“ <a href="#">关于 NetBackup for Hadoop 调试日志记录</a> ”。
备份问题	请参见第 43 页的“ <a href="#">对 Hadoop 数据的备份问题进行故障排除</a> ”。
还原问题	请参见第 47 页的“ <a href="#">对 Hadoop 数据的还原问题进行故障排除</a> ”。
要避免出现问题，另请查看最佳做法	请参见第 14 页的“ <a href="#">部署 Hadoop 插件的最佳做法</a> ”。 请参见第 33 页的“ <a href="#">备份 Hadoop 群集的最佳做法</a> ”。 请参见第 40 页的“ <a href="#">还原 Hadoop 群集的最佳做法</a> ”。

## 关于 NetBackup for Hadoop 调试日志记录

NetBackup 可为备份和还原操作中涉及各个进程维护特定于进程的日志。查看这些日志可以帮助您找到问题的根本原因。

要进行日志记录，这些日志文件夹“必须已经存在”。如果这些文件夹不存在，则必须进行创建。

日志文件夹位于以下目录

- 在 Windows 上：install\_path\NetBackup\logs
- 在 UNIX 或 Linux 上：/usr/opensv/netbackup/logs

表 5-2 与 Hadoop 相关的 NetBackup 日志

日志文件夹	消息相关项	日志所在位置
install_path/NetBackup/logs/bpVMutil	策略配置	主服务器
install_path/NetBackup/logs/nbaapidisv	BigData 框架、发现和 Hadoop 配置文件日志	备份主机
install_path/NetBackup/logs/bpbrm	策略验证、备份和还原操作	介质服务器
install_path/NetBackup/logs/bpbkar	备份	备份主机
install_path/NetBackup/logs/tar	还原和 Hadoop 配置文件	备份主机

有关更多详细信息，请参考 [NetBackup 日志记录参考指南](#)。

## 对 Hadoop 数据的备份问题进行故障排除

请查看以下主题：

- 请参见第 42 页的“关于 NetBackup for Hadoop 调试日志记录”。
- 请参见第 44 页的“备份操作失败，并发生错误 6609”。
- 请参见第 44 页的“备份操作失败，并发生错误 6618”。
- 请参见第 45 页的“备份操作失败，并发生错误 6647”。
- 请参见第 45 页的“对于 Hadoop，不会备份或还原扩展属性 (xattrs) 和访问控制列表 (ACL)”。
- 请参见第 46 页的“备份操作失败，并发生错误 6654”。
- 请参见第 46 页的“备份操作失败，并发生 bpbm 错误 8857”。
- 请参见第 46 页的“备份操作失败，并发生错误 6617”。
- 请参见第 46 页的“备份操作失败，并发生错误 6616”。

## 备份操作失败，并发生错误 112

错误

```
Job activity monitor shows the Hadoop backup job as failed with error code 112.
```

原因

已在策略备份选择中添加跟踪 /。

故障排除

确保具有 Hadoop 应用程序类型的 Bigdata 策略没有选择用于备份的多个文件夹的跟踪 /。例如，备份选择应如下所示

```
Application_Type=hadoop  
Backup_Host=hostname.fqdn  
/folder1  
/folder2  
/folder3
```

## 备份操作失败，并发生错误 6609

在以下情况下会发生此错误：

1. 任何备份主机（一个或多个）中删除或缺少 Hadoop 插件文件。

解决方法：

下载并安装 Hadoop 插件。

2. Application\_Type 详细信息不正确。

解决方法：

指定 Application\_Type 时使用 `hadoop`（而不是 `Hadoop`）。

## 备份操作失败，并发生错误 6618

备份操作失败并发生错误 6618，其中显示以下错误：

```
NetBackup cannot find the file to complete the operation.(6618)
```

如果提供了无效目录作为备份选择，则会发生此错误。

解决方法：

请在 BigData 策略中提供有效的目录作为备份选择。

## 备份操作失败，并发生错误 6647

备份操作失败并发生错误 6647，其中显示以下错误：

```
Unable to create or access a directory or a path. (6647)
```

在下列其中一种情况下，会发生此错误：

- 目录未启用快照。
- 将策略配置为创建根文件夹快照作为备份选择，而其中一个子文件夹已启用快照。
- 将策略配置为创建子文件夹快照作为备份选择，而其中一个父文件夹已启用快照。
- 将策略配置为创建文件快照作为备份选择

解决方法：

Hadoop 中不允许使用启用了快照的嵌套目录。如果父目录已启用快照，则父目录下的任何其他子目录均无法启用快照。对于 Bigdata 策略类型中的备份选择，必须仅选择启用了快照的目录进行备份，不得选择任何其他子目录。

## 对于 Hadoop，不会备份或还原扩展属性 (xattrs) 和访问控制列表 (ACL)

扩展属性允许用户应用程序将附加元数据与 Hadoop 中的某个文件或目录相关联。默认情况下，Hadoop 分布式文件系统 (HDFS) 上已启用此功能。

除了标准权限之外，访问控制列表还提供了一种为特定的命名用户或命名组设置不同权限的方法。默认情况下，HDFS 上已禁用此功能。

Hadoop 插件在备份期间不会捕获对象的扩展属性或访问控制列表 (ACL)，因此不会在还原的文件或文件夹上设置这些内容。

解决方法：

如果在使用包含 `Application_Type = hadoop` 的 BigData 策略备份的任何文件或目录上设置了扩展属性，则必须对还原的数据显式设置扩展属性。

可以使用 Hadoop shell 命令（例如 `fs -getfattr` 和 `hadoop fs -setfattr`）设置扩展属性。

如果在使用包含 `Application_Type = hadoop` 的 BigData 策略备份的任何文件或目录上启用并设置了访问控制列表 (ACL)，则必须对还原的数据显式设置 ACL。

可以使用 Hadoop shell 命令（例如 `hadoop fs -getfacl` 和 `hadoop fs -setfacl`）设置 ACL。

## 备份操作失败，并发生错误 6654

在以下情况下会发生此错误：

1. 如果未在 NetBackup 主服务器中添加 Hadoop 凭据

解决方法：

确保在 NetBackup 主服务器中添加了 Hadoop 凭据。使用 `tpconfig` 命令。有关更多信息，请参见第 19 页的“在 NetBackup 中添加 Hadoop 凭据”。

2. 如果备份主机上未安装 Hadoop 插件文件。

解决方法：

确保所有备份主机上均已安装 Hadoop 插件文件，然后再开始备份操作。

## 备份操作失败，并发生 bpbrm 错误 8857

如果未将 NetBackup 主服务器上的 NetBackup 客户端添加到白名单，则会发生此错误。

解决方法：

如果要将在 NetBackup 客户端用作备份主机，则必须在 NetBackup 主服务器上执行添加到白名单过程。有关更多信息，请参见第 18 页的“在 NetBackup 主服务器上添加 NetBackup 客户端添加到白名单”。

## 备份操作失败，并发生错误 6617

备份操作失败并发生错误 6617，其中显示以下错误：

```
A system call failed.
```

对于启用了 Kerberos 的 Hadoop 群集，请验证备份主机是否具有有效的票证授予票证 (TGT)。

解决方法：

续订 TGT。

## 备份操作失败，并发生错误 6616

备份操作失败并发生错误 6616，其中记录了以下错误：

```
hadoopOpenConfig: Failed to Create Json Object From Config File.
```

解决方法：

验证 `hadoop.conf` 文件，以确保未对参数值使用空值或不正确的语法。

## 对 Hadoop 数据的还原问题进行故障排除

- 请参见第 47 页的“还原失败，并显示错误代码 2850”。
- 请参见第 47 页的“针对 Hadoop 的 NetBackup 还原作业部分完成”。
- 请参见第 45 页的“对于 Hadoop，不会备份或还原扩展属性 (xattrs) 和访问控制列表 (ACL)”。
- 请参见第 48 页的“如果备份主机上缺少 Hadoop 插件文件，还原操作将失败”。
- 请参见第 48 页的“还原失败，并发生 bpbrm 错误 54932”。
- 请参见第 48 页的“还原操作失败，并发生 bpbrm 错误 21296”。

### 还原失败，并显示错误代码 2850

在以下情况下会出现此错误：

- 错误：2850 "errno = 62 - Timer expired"  
解决方法：  
更新防火墙设置，以便备份主机可与 Hadoop 群集通信。
- 请求的文件未恢复。  
解决方法：  
对于启用了 Kerberos 的 Hadoop 群集，请验证备份主机是否具有有效的票证授予票证 (TGT)。续订 TGT。
- 应用程序服务器的值不正确且凭据无效。  
解决方法：  
确保还原期间输入的目标 Hadoop 群集的主机名正确无误。此主机名应与 `tpconfig` 命令中提供的主机名相同。

### 针对 Hadoop 的 NetBackup 还原作业部分完成

如果要还原的数据大小超过 Hadoop 群集上的可用空间，还原作业将部分完成。

解决方法：

清理 Hadoop 群集上的空间。

### 对于 Hadoop，不会备份或还原扩展属性 (xattrs) 和访问控制列表 (ACL)

有关此问题的更多信息，请参见第 45 页的“对于 Hadoop，不会备份或还原扩展属性 (xattrs) 和访问控制列表 (ACL)”。

## 如果备份主机上缺少 Hadoop 插件文件，还原操作将失败

在未安装 Hadoop 插件文件的备份主机上触发还原作业时，还原操作将失败并显示以下错误：

```
client restore EXIT STATUS 50: client process aborted
```

解决方法：下载并安装 Hadoop 插件。

## 还原失败，并发生 bpbrm 错误 54932

如果要还原的文件备份不成功，则会发生此错误。

解决方法：

在开始还原操作之前，请确保备份已成功完成。

或者，在“活动监视器”菜单上单击“作业状态”选项卡，以找到特定的作业 ID 并查看错误消息详细信息。

## 还原操作失败，并发生 bpbrm 错误 21296

如果在将 Hadoop 凭据添加到 NetBackup 主服务器时为 <application\_server\_name> 提供的值不正确，则会发生此错误。

解决方法：

验证为 <application\_server\_name> 提供的详细信息是否正确。

# 索引

## A

- 安装
  - 验证 14

## B

- BigData 策略
  - NetBackup 管理控制台 26
  - 策略配置向导 26
  - 策略实用程序 26
  - 命令行界面 27
- 备份 8, 32
  - Hadoop 31
- 并行流框架 6

## C

- 策略
  - 配置 25
- 创建
  - BigData 备份策略 25

## G

- 概述
  - 安装 6
  - 备份 6
  - 部署 6
  - 还原 6
  - 配置 6
- 故障排除
  - 备份 43
  - 还原 47

## H

- Hadoop 凭据
  - 添加 19
- 还原 9
  - bprestore 命令 35
  - Hadoop 33
  - Hadoop 群集 34
  - 备用 NameNode 37

## J

- 兼容性
  - 支持的操作系统 13

## K

- kerberos
  - 备份 32
  - 还原 32
- Kerberos
  - 安装后 24

## N

- NetBackup
  - 调试日志记录 42
  - 服务器和客户端要求 13
- NetBackup 设备
  - 备份主机 19

## S

- 删除
  - 备份主机 16
- 术语 10

## T

- 添加
  - 备份主机 16
- 添加到白名单
  - 备份主机 18

## X

- 限制 11
- 许可证
  - Hadoop 13

## Z

- 灾难恢复 30
- 准备
  - Hadoop 13