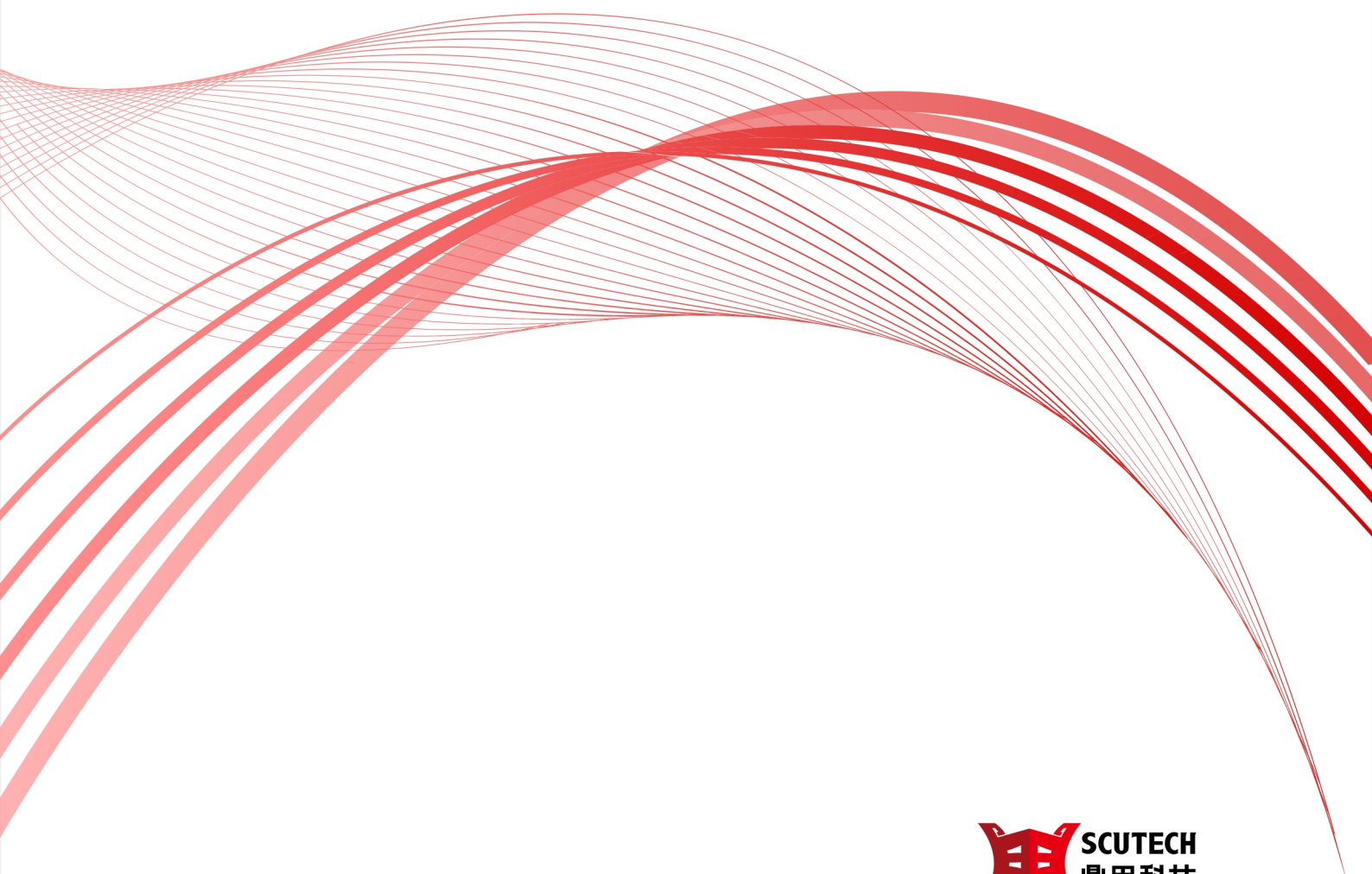


鼎甲迪备

# Hive 备份恢复用户指南

**Release V8.0-9**

November, 2024



# 目录

<b>1</b>	<b>概述</b>	<b>1</b>
<b>2</b>	<b>计划和准备</b>	<b>2</b>
<b>3</b>	<b>代理端安装和配置</b>	<b>3</b>
3.1	验证兼容性 . . . . .	3
3.2	安装迪备代理端 . . . . .	3
3.2.1	Apache Hadoop . . . . .	3
3.2.2	FusionInsight MRS 集群环境 . . . . .	4
<b>4</b>	<b>激活代理端许可证和分配授权</b>	<b>7</b>
<b>5</b>	<b>添加和激活 Hive</b>	<b>8</b>
5.1	添加 Hadoop 集群 . . . . .	8
5.2	添加 Hive . . . . .	11
5.3	激活 HDFS 和 Hive . . . . .	13
<b>6</b>	<b>备份</b>	<b>14</b>
6.1	备份类型 . . . . .	14
6.2	备份策略 . . . . .	14
6.3	开始之前 . . . . .	14
6.4	创建备份作业 . . . . .	15
6.5	备份选项 . . . . .	16
<b>7</b>	<b>恢复</b>	<b>18</b>
7.1	开始之前 . . . . .	18
7.2	创建数据库时间点恢复作业 . . . . .	18
7.3	恢复选项 . . . . .	20
<b>8</b>	<b>术语表</b>	<b>21</b>

该文档主要描述如何安装配置迪备代理端以及如何正确使用迪备备份和恢复 Hive。

迪备支持 Hive 备份恢复主要特性包括：

- 备份内容

数据库、用户

- 备份类型

完全备份、增量备份

- 备份目标

标准存储池、重删存储池、本地存储池、磁带库池、对象存储池、LAN-Free 池等

- 备份策略

迪备提供 6 种备份计划，立即、一次、每小时、每天、每周、每月

- 数据处理

数据压缩、数据加密、多通道、断点续传、限速、复制

- 恢复类型

数据库时间点恢复

- 恢复目标

原机恢复、异机恢复

- 恢复选项

备份主机、通道数、同名数据库处理方式、临时目录、后置重建表分区

在安装迪备代理端之前，需确保满足以下要求：

1. 确保所有备份组件都已安装和部署，包括备份服务器、存储服务器。
2. 迪备控制台创建一个至少具备操作员和管理员角色的用户，使用此用户登录迪备控制台并对资源进行备份恢复。

**备注：**管理员角色用于代理端安装和配置、激活许可证和授权用户。操作员角色用于创建备份和恢复作业、副本管理。

要实现 Hive 备份及恢复，需要在能与 Hadoop 和 Hive 通信的主机安装迪备代理端。

### 3.1 验证兼容性

在安装代理端之前，需先确保 Hive 的环境已在鼎甲迪备的适配列表中。

迪备支持多种版本 Hive 备份恢复。支持的版本主要有：

- Hive 3.1.2
- FusionInsight MRS Hive 3.1.0

### 3.2 安装迪备代理端

迪备代理端安装在 Linux，支持在线安装和本地安装代理端，推荐在线安装方式。

1. 在线安装：迪备支持用 curl 或 wget 命令在 Linux 主机安装代理。
2. 本地安装：参考《代理端安装用户指南》的“本地安装”章节。

#### 3.2.1 Apache Hadoop

安装 Hadoop 备份代理端前需要在代理端主机安装 Hadoop 运行环境。

**备注：**如果同时存在 2.x 和 3.x 版本的 Hadoop 集群，而且需要支持 Hive 3.x 备份与恢复，则应根据具体需求部署两个代理端。一个代理端部署 Hadoop 2.10.0 组件版本，该代理端支持 Hadoop 2.x 和 3.x，不支持 Hive 的备份与恢复功能；另一个代理端部署 Hadoop 3.2.1 组件版本，该代理端支持 Hadoop 3.x 和 Hive 3.x 的备份与恢复功能。

- 解压 Hadoop 和 Hive 运行环境离线包：

```
$ sudo tar -axf hadoop-3.2.1.tar.xz -C <dir>
$ sudo tar -axf apache-hive-3.1.2-bin.tar.gz -C <dir>
```

解压出目录为：hadoop-3.2.1、apache-hive-3.1.2-bin

- 安装 OpenJDK：

```
$ sudo tar -axf Ubuntu20.04-OpenJDK11-AMD64.tar.gz
```

解压出目录为：openjdk11 只需要安装 openjdk-11-jre-headless 环境即可

```
$ sudo dpkg -i openjdk11/*.deb
```

**备注：**jre 的目录以及版本根据实际的安装情况，默认在 /usr/lib/jvm/ 目录下

- 主机配置 Hadoop 客户端时，需执行 config 配置环境变量，以压缩包解压后目录 '/opt/hadoopclient' 为例，

```
$ /etc/init.d/dbackup3-agent config hadoop
$ Configure Huawei MRS? [y/N] n
$ Please input JRE home []: /usr/lib/jvm/java-8-openjdk-amd64
$ Please input Hadoop home []:/opt/hadoopclient/hadoop-3.2.1
$ Configure Hive? [y/N] y
$ Please input Hive home []:/opt/hadoopclient/apache-hive-3.1.2-bin
$ Restarting dbackup3-agent (via systemctl): [ OK ]
$ [ ok ] Restarting dbackup3-agent (via systemctl): dbackup3-agent.service.
```

### 3.2.2 FusionInsight MRS 集群环境

FusionInsight Manager 版本低于 8.3.1 时，只能手动在 Hadoop 备份代理端的主机安装 MRS 集群客户端。

**备注：**由于 FusionInsight MRS 环境的兼容性问题，一个 FusionInsight MRS 客户端只能对应一个 FusionInsight MRS 集群。如果现场同时存在多个 FusionInsight MRS 集群，则应根据具体需求部署多个 FusionInsight MRS 客户端和代理端，以确保各集群能够正常注册。

- 准备：安装 NTP 软件用于同步 MRS 集群系统时间。

```
$ yum install ntp ntpdate
```

- 复制 FusionInsight MRS 客户端安装目录下的 hosts 的内容到 /etc/hosts。
- 将 FusionInsight MRS 集群的 /etc/ntp.conf 文件内容拷贝增加到客户端的 /etc/ntp.conf 文件内。
- 将 FusionInsight MRS 集群的 /etc/ntp/ntpkeys 拷贝到 /etc/ntp/ 目录下。(如果 MRS 集群的 NTP 服务有 ntpkeys 文件执行该步骤，否则可跳过不做)
- 重启 NTP 服务使上述修改生效，并重启 Chrony 服务。

```
$ systemctl restart ntpd
$ systemctl start chronyd
```

- 解压 FusionInsight MRS 客户端运行环境离线包：

```
$ tar -axf FusionInsight_Cluster_1_Services_Client.tar -C <dir>
$ tar -axf FusionInsight_Cluster_1_Services_ClientConfig.tar -C <dir>
```

解压出目录为：FusionInsight\_Cluster\_1\_Services\_ClientConfig。

- 进入 FusionInsight\_Cluster\_1\_Services\_ClientConfig 安装 FusionInsight MRS 客户端，安装路径为 '/opt/hadoopclient'。

```
$ cd <dir>/FusionInsight_Cluster_1_Services_ClientConfig
$ mkdir -p /opt/hadoopclient
$ ./install.sh /opt/hadoopclient
```

FusionInsight Manager 版本大于或等于 8.3.1 时，FusionInsight Manager 已支持远程自动部署 FusionInsight MRS 客户端，具体信息可参考如下内容；

关于

FusionInsight Manager

版本：8.3.1

电子信箱：[support@huawei.com](mailto:support@huawei.com)

网址：<http://www.huawei.com>

版权所有 © 华为技术有限公司2024 保留所有权限

下载集群客户端

下载MRS\_dingjia\_WX... 的客户端，集群的客户端包括了所有服务

选择客户端类型：

完整客户端

仅配置文件

选择平台类型：

☒ x86\_64

☐ aarch64

选择下载位置：

☐ 服务器端

☐ 浏览器

☒ 远端节点

保存到如下路径：

/tmp/FusionInsight-Client-Remote/

?

主机IP：

10

.

200

.

29

.

223

主机端口：

3322

登录用户名称：

root

登录方式：

☒ 密码

☐ SSH私钥

☐ 免密

登录用户密码：

dingjia

👁

自动部署到节点：

☐ 否

☒ 是

客户端部署路径：

/opt/hadoopclient

?

确定

取消

3.2. 安装迪备代理端

5

- 主机配置 Hadoop 客户端时，需执行 config 配置环境变量，以压缩包解压后目录 ‘/opt/hadoopclient’ 为例，

```
$ /etc/init.d/dbackup3-agent config hadoop
$ Configure Huawei MRS? [y/N] y
$ Please input Huawei MRS home []: /opt/hadoopclient/
$ Huawei MRS JRE_HOME: /opt/hadoopclient/JDK/jdk1.8.0_402
$ Huawei MRS HADOOP_HOME: /opt/hadoopclient/HDFS/hadoop
$ Huawei MRS HIVE_HOME: /opt/hadoopclient/Hive/Beeline
$ Restarting dbackup3-agent (via systemctl): [ OK ]
$ [ ok ] Restarting dbackup3-agent (via systemctl): dbackup3-agent.service.
```

在线安装代理的步骤如下：

1. 登录迪备控制台。
2. 在菜单栏中，点击【资源】，进入【资源】页面。
3. 在工具栏中，点击【安装代理端】按钮，进入【安装代理端】页面。
4. 【选择系统】选择“Linux”，【选择模块】选择“Hadoop”。

**备注：**（1）如果您想在 Linux 主机安装完代理端后自动删除下载的安装包，需勾选【删除安装包】。（2）如果勾选【忽略 SSL 错误】选项，程序将会忽略证书等错误。若没勾选，程序将会维持当前逻辑。出现错误时提示用户输入 Y/N 以选择是否继续执行。

5. 选择使用 curl 或 wget，点击【复制】按钮，复制安装命令。
6. 使用 root 登录 Linux 主机，在主机终端粘贴安装命令，按回车进行代理端安装。如：

```
root@ubuntu:~# curl -o- "http://IP:80/d2/update/script?modules=h3c-cas&ignore_ssl_
↪error=&access_key=7dc57757b7e675f2ec5495180f90ac70&rm=&tool=curl" | sh
```

7. 等待安装完成。



## 4 激活代理端许可证和分配授权

代理端安装成功后，返回迪备控制台【资源】页面，列表中会出现安装代理端的主机。在备份恢复之前，需要在迪备控制台注册主机、激活 Hadoop 备份许可证，并授权用户。

操作步骤如下：

1. 在菜单栏中，点击【资源】，进入资源页面。
2. 在主机列表中，找到 Hadoop\_Proxy 所在的主机，点击主机的【注册】按钮，会弹出【激活】窗口，点击【提交】。
3. 在【激活】页面点击【提交】后，会弹出【配置】窗口，设置主机名称、选取数据网络，选择首选网络出口，授权用户组，点击【提交】。
  - 名称：可自定义设置主机名称。
  - 数据网络：可选取已在“存储 - 网络”处添加的网络。
  - 首选网络出口：设置该主机的首选备份数据的网络流量出口 IP 地址，支持 IPv4/IPv6。
  - 用户组：授权该资源给用户组。

### 备注：

1. 若提示“许可证不足”，需联系迪备管理员增加许可证。
2. 若代理端数量较多，建议对所有代理端先完成安装，再使用【批量注册】、【批量激活】和【批量授权】，以减少操作次数。具体查看《管理员用户指南》的“批量注册/激活/授权”章节。

5.1 添加 Hadoop 集群

- 1. 在菜单栏中，点击【资源】，进入资源页面，点击左上角的“+”添加 Hadoop 集群资源。
- 2. 添加 Hadoop 集群时支持使用 Simple 和 Kerberos 两种验证方式。如果需要添加的 Hadoop 集群如果配置了 Kerberos 认证服务，那么添加集群时需要选择 Kerberos 验证方式添加，如果没有配置 Kerberos 认证服务的直接使用默认的 Simple 验证方式即可。注册页面左下角的“+”是用来扩展填写多个 NameNode 节点（HA）。
- 3. 勾选 Hive 选项，切换到 Hive 信息页面，验证方式与 Hadoop 集群相同，连接方式有主机和 Zookeeper 两种连接方式。

(1) 创建 Hadoop 资源

添加 Hadoop 集群

名称

Hadoop

备份主机

Hadoop\_Proxy

用于列表备份内容以及作为备份和恢复时的默认主机。

添加 NameNode

主机

192.168.xx.xx

SSL

☐

REST API 端口

50070

RPC API 端口

8020

用户

hadoop

- 名称：自定义资源名称。
- 主机：输入 NameNode 节点所在的主机 IP 或主机名。如果配置 Kerberos 认证时 Principal 使用主机名进行创建，那么该字段需要填写主机名，并且代理端所在的机器的 hosts 文件也要添加该主机的 IP 及其对应的主机名解析。
- 安全连接：使用 SSL 安全连接。该选项需要 Hadoop 集群配置和启用 HTTPS 服务才能使用，否则不需要勾选该选项。

- RPC API 端口：默认为 8020，如果集群配置为其他端口那么该选项需要根据实际端口进行修改。可在 Hadoop 服务 `core-site.xml` 文件里面查看参数 `fs.defaultFS` 配置的端口，没有配置即为默认端口。
- REST API 端口：HTTP 默认为 50070，HTTPS 默认为 50470，如果集群配置为其他端口那么该选项需要根据实际端口进行修改。可在 Hadoop 服务 `hdfs-site.xml` 文件里面查看参数 `dfs.namenode.http(s)-address` 配置的端口，没有配置即为默认端口。
- 用户：HDFS 文件系统所属用户，有 Kerberos 认证，需要填写认证到 Keytab 文件里面的 Principal 的用户。如：test@HADOOP.COM，就填写 test。

## (2) Simple 验证方式

验证方式

Simple

core-site.xml  
文件

Choose File No file chosen  
(可选)

hdfs-site.xml  
文件

Choose File No file chosen  
(可选)

+

- 验证方式：默认选择 Simple。
- core-site.xml 文件：上传集群的 `core-site.xml` 文件，使用 Simple 验证方式可以不用上传。
- hdfs-site.xml 文件：上传集群的 `hdfs-site.xml` 文件，使用 Simple 验证方式可以不用上传。

## (3) Kerberos 验证方式

主机	<input type="text" value="MRSDingjiaWX1175038-node-master3..."/>	②
	仅支持一个地址	
SSL	<input checked="" type="checkbox"/> 使用 SSL 连接 NameNode	
REST API 端口	<input type="text" value="25003"/>	
RPC API 端口	<input type="text" value="25006"/>	
用户	<input type="text" value="test"/>	②

---

验证方式	<div>Kerberos</div>	
Realm 名称	<input type="text" value="A147FF59_B917_4F70_885F_38C7DF8..."/>	
Realm KDC 服务器	<input type="text" value="10.200.29.251:21732,10.200.29.107:21..."/>	②
Realm 管理服务器	<input type="text" value="10.200.29.251:21730,10.200.29.107:21..."/>	②
RPC API Principal	<input type="text" value="test@A147FF59_B917_4F70_885F_38C..."/>	
REST API Principal	<input type="text" value="test@A147FF59_B917_4F70_885F_38C..."/>	
UDP Preference Limit	<input type="text" value="1465"/>	②
krb5.keytab 文件	<div>Choose File</div> user.keytab	

---

core-site.xml 文件	<div>Choose File</div> core-site.xml (可选)	
hdfs-site.xml 文件	<div>Choose File</div> hdfs-site.xml (可选)	

- 验证方式：选择 Kerberos。
- Realm 名称：配置 Kerberos 创建时的 Realm 名称。
- Realm KDC 服务器：Realm KDC 服务器 IP 或主机名。默认端口为：88，非默认端口，需要加上端口。
- Realm 管理服务器：Realm 管理服务器 IP 或主机名。默认端口为：88，非默认端口，需要加上端口。
- RPC API Principal：输入 RPC API Principal 名称。可在 Keytab 文件中查看到，如：`klist -k -t test.keytab`。
- REST API Principal：输入 REST API Principal 名称。可在 Keytab 文件中查看到，如：`klist -k -t test.keytab`。

- **UDP Preference Limit**: 该参数指定 UDP 传输包的最大值, 数据包大于该值使用 TCP 进行传输。默认值为 1, 默认使用 TCP 传输, 应根据 KDC 服务的 `/etc/krb5.conf` 中的参数进行调整。
- **krb5.keytab 文件**: 获取 Keytab 文件并将其复制到访问迪备控制台的主机上的安全位置。
- **core-site.xml 文件**: 上传集群的 `core-site.xml` 文件, 使用 Kerberos 验证方式必须要上传。
- **hdfs-site.xml 文件**: 上传集群的 `hdfs-site.xml` 文件, 使用 Kerberos 验证方式必须要上传。

## 5.2 添加 Hive

### (1) 主机连接方式

SSL

☐ 使用 SSL 连接 Hive

连接方式

☒ 主机 ☐ ZooKeeper

地址

192.168.xx.xxx

端口

10000

用户名

可选

密码

可选

验证方式

Simple

- **安全连接**: 使用 SSL 安全连接。该选项需要 Hive 配置和启用 HTTPS 服务才能使用, 否则不需要勾选该选项。
- **验证方式**: 选择主机。
- **名称**: 自定义资源名称。
- **主机**: Hiveserver 所在的主机 IP 或主机名。如果配置 Kerberos 认证时 Principal 使用主机名进行创建, 那么该字段需要填写主机名, 并且代理端所在的机器的 `hosts` 文件也要添加该主机的 IP 及其对应的主机名解析。
- **安全连接**: 使用 SSL 安全连接。该选项需要 Hive 配置和启用 HTTPS 服务才能使用, 否则不需要勾选该选项。
- **端口**: 默认为 10000, 如果集群配置为其他端口那么该选项需要根据实际端口进行修改。
- **用户名**: 默认为空, 用户如若有设置数据库用户则按实际填写。
- **密码**: 用户设置数据库用户时填写。

### (2) ZooKeeper 连接方式

SSL

☐ 使用 SSL 连接 Hive

连接方式

☐ 主机 ☒ ZooKeeper

服务器列表

10.200.29.228:24002,10.200.29.251:2...

ZooKeeper 服务器列表，用英文逗号分隔。例如：zk1.com:2181,zk2:2181,zk3:2181

命名空间

hiveserver2

Hive 在 ZooKeeper 中使用的命名空间

SSL

☐ 使用 SSL 连接 ZooKeeper

用户名

可选

密码

可选

验证方式

Kerberos

Realm 名称

A147FF59\_B917\_4F70\_885F\_38C7DF8...

Realm KDC 服务器

10.200.29.251:21732,10.200.29.107:21...?

Realm 管理服务器

10.200.29.251:21730,10.200.29.107:21...?

Principal

test@A147FF59\_B917\_4F70\_885F\_38C...

UDP Preference Limit

1465?

krb5.keytab 文件

Choose File

user.keytab

- 安全连接：使用 SSL 安全连接。该选项需要 Hive 配置和启用 HTTPS 服务才能使用，否则不需要勾选该选项。
- 验证方式：选择 ZooKeeper。
- 服务列表：ZooKeeper 服务器的列表，用英文逗号分隔。在 FusionInsight MRS 客户端主机上执行 echo \$CLIENT\_HIVE\_URI 获取相关信息。
- 命名空间：Hive 在 ZooKeeper 中使用的命名空间。在 FusionInsight MRS 客户端主机上执行 echo \$CLIENT\_HIVE\_URI 获取相关信息。
- 端口：默认为 10000，如果集群配置为其他端口那么该选项需要根据实际端口进行修改。
- 用户名：默认为空，用户如若有设置数据库用户则按实际填写。
- 密码：用户设置数据库用户时填写。

(3) Simple 验证方式和 Kerberos 验证方式与添加 Hadoop 集群一致。

## 5.3 激活 HDFS 和 Hive

1. 添加 Hadoop 集群和 Hive 成功后，会弹出【Hadoop 许可证】激活窗口，点击 HDFS、Hive 资源的【激活】按钮。
2. 激活后，点击 HDFS、Hive 资源【授权】按钮，进行授权。
3. 在【授权】窗口，可对 HDFS、Hive 资源进行授权用户操作。
  - 用户组：授权该资源给用户组。
  - 受保护：被标记为受保护的资源将无法用于恢复或数据复制的目标，除非管理员移除该标记。

### 备注：

1. 若提示“许可证不足”，需联系迪备管理员增加许可证。
2. 若已添加的集群参数发生变更，包括主机 IP、端口、验证方式等参数，用户可以通过点击【设置】对已添加的 Hadoop 集群和 Hive 进行修改。

### 6.1 备份类型

迪备为 Hive 备份提供完全备份、增量备份两种常规的备份类型。

- 完全备份

备份 Hive 实例下的所有数据。

- 增量备份

增量备份基于完全备份创建，备份相较于上次完全/增量备份以来改变的数据文件。若第一次增量备份前未做完全备份，则会做完全备份。

### 6.2 备份策略

迪备提供 6 种备份计划，立即、一次、每小时、每天、每周、每月。

- 立即：作业创建后就执行。
- 一次：作业在指定时间执行一次。
- 每小时：作业每天在设置的时间范围内以特定的小时/分钟间隔重复运行。
- 每天：作业以特定的天数间隔在特定时间重复运行。
- 每周：作业以特定的周数间隔在特定时间重复运行。
- 每月：作业在特定月份和时间重复运行。

针对用户的实际情况和需求，设置合理的备份策略。通常，推荐用户使用常规的备份策略：

1. 完全备份：每周在应用访问量较小的时间（例如周末）进行一次完全备份，以确保每周至少有一个可恢复的时间点。
2. 增量备份：每天在业务低峰期（例如凌晨 02:00）进行一次增量备份，可以更好地节省存储空间和备份时间，保证每天至少有一个可恢复的时间点。

### 6.3 开始之前

在备份恢复 Hive 之前，需保证已完成如下操作：

#### 1. 检查存储池

（1）在迪备菜单栏中，点击【存储池】，进入【存储池】页面。

（2）检查展示区是否存在存储池。如果没有，需参考《管理员用户指南》的“创建存储池”，创建存储池并授权给当前控制台用户。



## 6.4 创建备份作业

- 1. 在菜单栏中，点击【备份】，进入【备份】页面。
- 2. 在【主机和资源】页面，选择 Hadoop 主机和 Hive 实例，自动跳转【下一步】。
- 3. 在【备份内容】页面，选择一个【备份类型】，勾选您希望备份的数据库，点击【下一步】。

备份类型

完全备份

备份内容

Hive

数据库

☐ bigdata

☐ db1

☐ db10

☒ db100

☒ db100\_1722326909

☒ db10\_1721644727

☒ db10\_1721731972

☐ db10\_1722326853

☐ db11

☐ db11\_1721644779

☐ db11\_1722326963

☐ db12

☐ db12\_1722327018

☐ db13

☐ db13\_1722327076

☐ db14

☐ db15

☐ db16

☐ db17

☐ db18

☐ db19

☐ db1\_1721477646

☐ db1\_1721731918

☐ db2

(1) 【备份类型】选择完全备份、增量备份。

备注：对于增量备份，【备份内容】步骤只需要选择完全备份作为基准，无需再次选择数据库和表。

(2) 点击 + 可以展开数据库的表，勾选需要备份的表。

- 4. 在【备份目标】页面，选择一个备份主机和一个存储池，点击【下一步】。

**备注：**增量备份没有【备份目标】步骤，由于的它们的【备份目标】与【备份内容】步骤中选择的基准完全备份相同。

6. 在【备份计划】页面，选择一个计划类型，参考[备份策略](#)。点击【下一步】。
- 选择“立即”，作业创建后就执行。
  - 选择“一次”，设置作业的开始时间。
  - 选择“每小时”，设置开始时间和结束时间，用于指定作业一天内执行的时间范围。输入作业执行的时间间隔，单位可选择小时或分钟。
  - 选择“每天”，设置作业的开始时间。输入作业执行的时间间隔，单位为天。
  - 选择“每周”，设置作业的开始时间。输入作业执行的时间间隔，单位为周，并选择一周内具体执行的日期。
  - 选择“每月”，设置作业的开始时间。选择作业执行的月份。按每月的自然日，或每月的周选择具体日期。
7. 在【备份选项】页面，根据需要设置常规选项和高级选项，参考[备份选项](#)。点击【下一步】。

压缩

快速

通道数

1

范围 1~255

快照 ?

☐

数据库未配置复制策略时

☒ 发送警报信息并取消作业

☐ 添加复制策略

8. 在【完成】页面，设置【作业名】，并检查作业信息是否有误。点击【提交】。
9. 提交成功后，自动跳转到作业页面。您还可以对作业进行开始、修改、删除等管理操作。

## 6.5 备份选项

迪备为 Hive 提供以下备份选项：

- 常规选项

表 1：备份常规选项

功能	描述	限制性说明
压缩	默认启用快速压缩。 - 不压缩：备份过程中不压缩。 - 可调节：自定义压缩级别，需激活高级功能。 - 快速压缩：备份过程中压缩，使用快速压缩算法。	
通道数	开启该选项可提高备份效率。通道数默认为 1，选择范围为 1~255，单位为个。 一般建议跟 CPU 核心数一致，超过 CPU 核心数之后效率提高不明显。	该选项仅完全备份支持，增量备份和完全备份保持一致。

续下页

表 1 – 接上页

功能	描述	限制性说明
快照	开启该选项可进行 Hive 快照备份，默认不开启。	该选项仅完全备份支持，增量备份和完全备份保持一致。
数据库未配置复制策略时	开启该选项可选择发送警报信息并取消作业或者添加复制策略，Hive 3.x 以上不需配置选项即可进行备份	增量备份和完全备份保持一致。

- 高级选项：

表 2：备份高级选项

功能	描述	限制性说明
断线重连时间	支持 1~60，单位为分钟。在设置时间内网络发生异常复位后作业继续进行。	
断点续传缓冲区	设置断点续传缓冲区大小，默认为 10 MiB。加大缓冲区将消耗更多物理内存，但在高吞吐量场景下加大缓冲区可避免断点续传失效。	
速度限制	可分时段限制数据传输速度或磁盘读写速度。单位为 KiB/s、MiB/s 或 GiB/s。	
前置条件	作业开始前调用，当前置条件不成立时中止作业执行，作业变成空闲状态。	
前置/后置脚本	前置脚本在作业开始后资源进行备份前调用，后置脚本在资源进行备份后调用。	

针对不同需求，迪备提供多种 Hive 的恢复方式，包括：

- 时间点恢复

当 Hive 的数据库数据丢失时，可以通过时间点恢复功能将 Hive 恢复到指定的时间点状态。Hive 时间点恢复支持本机和异机恢复，可以覆盖恢复、重命名恢复或跳过重名数据库恢复。

### 7.1 开始之前

如果要恢复 Hive 到其他主机，需先注册其他主机的 Hive 资源，激活许可证，并将 Hive 资源授权给当前迪备控制台用户。

### 7.2 创建数据库时间点恢复作业

创建数据库时间点恢复作业的步骤如下：

1. 在菜单栏中，点击【恢复】，进入【恢复】页面。
2. 在【主机和资源】页面，选择 Hive 所在主机和实例，自动跳转【下一步】。
3. 在【备份集】页面中，完成以下操作：

存储池	[默认]
	默认值表示从备份作业的目标池恢复。
恢复类型	时间点恢复
还原类型	数据库
恢复内容	<div> <div>Hive</div> <div> <div>备份集</div> <div> <div>Hive 完全备份作业9</div> <div>Hive 完全备份作业8</div> <div>2024-07-20 19:29:35</div> <div>Hive 完全备份作业6</div> <div>Hive 完全备份作业5</div> <div>Hive 完全备份作业4</div> <div>Hive 完全备份作业3</div> <div>Hive 完全备份作业2</div> <div>Hive 完全备份作业1</div> </div> </div> </div>
数据库	<div> <div>数据库</div> <div> <div>bigdata</div> <div>db1</div> <div>db2</div> <div>mrsarc</div> </div> </div>

(1) **【存储池】**默认值表示从备份作业的目标池恢复，可选择任意已产生备份集的存储池。包括做池复制的目的池。

(2) **【恢复类型】**选择**时间点恢复**。

(3) **【还原类型】**选择**数据库**。

(4) 在**【恢复内容】**列表中，选择需要恢复的备份集时间点。

(5) 选择**【数据库】**。默认恢复备份集中的所有数据库，也可以手动“取消/勾选”选择恢复部分数据库。

4. 在**【恢复目标】**页面，支持恢复到本机异实例或异机实例。自动跳转**【下一步】**。

5. 在**【恢复计划】**页面，选择“立即”或“一次”，点击**【下一步】**。

- 选择“立即”，作业创建后就执行。
- 选择“一次”，设置作业的开始时间。

6. 在**【恢复选项】**页面，参考[恢复选项](#)，根据所需进行设置。点击**【下一步】**。

7. 在**【完成】**页面，设置作业名称，并确认恢复内容。点击**【提交】**，等待作业执行。

8. 提交成功后，自动跳转到作业页面。您还可以对作业进行开始、修改、删除等管理操作。

### 7.3 恢复选项

迪备为 Hive 提供以下恢复选项：

- 常规选项：

表 3：恢复常规选项

功能	描述	限制性说明
备份主机	可以修改备份主机。默认为 Hive 实例设置的主机。	
通道数	开启该选项可提高恢复效率。通道数默认为 1，选择范围的最大值不能超过备份集最大的通道数，单位为个。	
同名数据库处理方式	可选择覆盖恢复、跳过重名数据库的恢复和新数据库添加时间戳后缀	
临时目录	可自主选择临时目录，用于存放恢复过程中产生的临时文件。恢复完成后，将自动清理此作业产生的临时文件。	
后置重建表分区	默认开启，在恢复 Hive 元数据时先忽略分区结构，待元数据恢复完成后重建表分区。可避免 Hive 3.1.0 执行 REPL LOAD 时内存占用过高的问题。	

- 高级选项：

表 4：恢复高级选项

功能	描述	限制性说明
断线重连时间	支持 1~60，单位为分钟。在设置时间内网络发生异常复位后作业继续进行。	
断点续传缓冲区	默认为 10 MiB。设置断点续传缓冲区大小。加大缓冲区将消耗更多物理内存，但在高吞吐量场景下加大缓冲区可避免断点续传失效。	
速度限制	可分时段限制数据传输速度或磁盘读写速度。单位为 KiB/s、MiB/s 或 GiB/s。	
前置条件	作业开始前调用，当前置条件不成立时中止作业执行，作业变成空闲状态。	
前置/后置脚本	前置脚本在作业开始后资源进行恢复前调用，后置脚本在资源进行恢复后调用。	

表 5：术语表

术语	说明
快速压缩	备份过程中压缩，使用快速压缩算法。
异机恢复	两个 <b>Hive</b> 数据库互相恢复。



全国销售热线：400-650-0081

电话：+86 20 32053160

总部地址：广州市科学城科学大道243号总部经济区A5栋9楼

全国服务热线：400-003-3191

网址：[www.scutech.com](http://www.scutech.com)