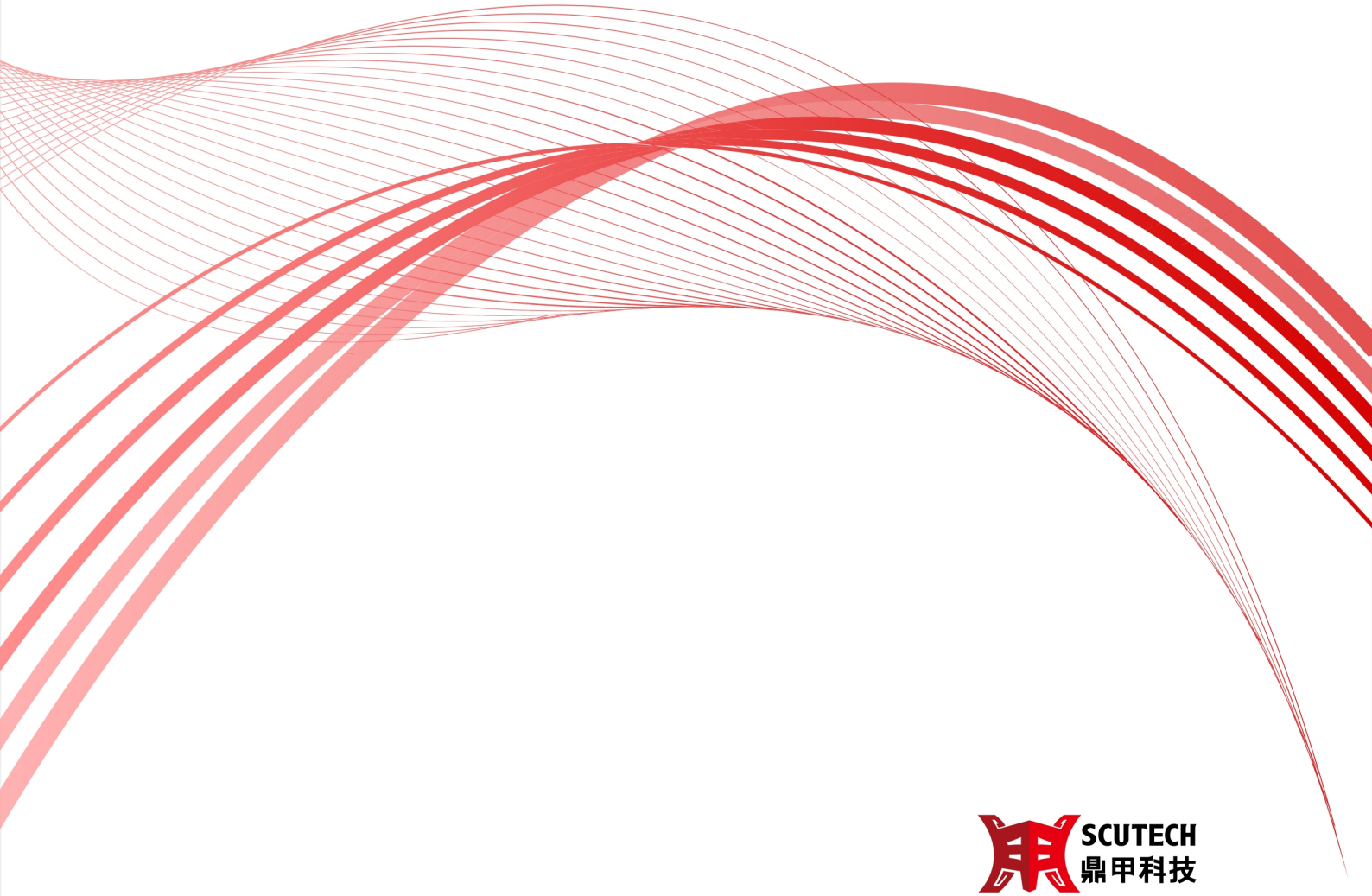
鼎甲迪备

Hadoop 备份恢复用户指南

Release V8.0-9

June, 2025



SCUTECH 鼎甲科技

目录

1 概述

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 Hadoop。

迪备支持 Hadoop 备份恢复主要特性包括：

- 备份内容

HDFS 文件系统单个或多个目录文件

- 备份类型

完全备份、增量备份、合成备份

- 备份目标

标准存储池、重删存储池、本地存储池、文件合成池、磁带库池、对象存储池、LAN-free 池

- 备份策略

迪备提供 7 种备份计划，立即、一次、手动、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、限制传输速度、限制备份速度、限制恢复速度、复制

- 恢复类型

时间点恢复、即时恢复、演练

- 恢复目标

原机恢复、异机恢复、跨系统恢复（Hadoop 和 Linux 系统互跨恢复）、异构恢复（Hadoop 恢复至操作系统的文件或对象存储）

- 恢复选项

备份主机、通道数、增量恢复、恢复路径（原路径、自定义路径）、同名文件处理、无效路径处理

2 计划和准备

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 Hadoop 备份及恢复，需要在能与 Hadoop 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 Hadoop 的环境已在鼎甲迪备的适配列表中。

迪备支持多种版本 Hadoop 备份恢复。支持的版本主要有:

- Hadoop 2.2.x/2.6.x/2.7.x/2.8.x/2.9.x/3.0.x/3.1.x/3.2.x
- CDH 6.0/6.1/6.2/6.3
- FusionInsight HD V100R002C50/V100R002C60/V100R002C70/V100R002C80
- 华为 MRS(Hadoop) 3.3.1

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。

- 解压 Hadoop 运行环境离线包:

```
$ sudo tar -axf hadoop-2.10.0.tar.xz -C <dir>
```

解压出目录为：hadoop-2.10.0

- 安装 OpenJDK，以 Ubuntu20.04 为例:

```
$ 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-2.10.0
$ Configure Hive? [y/N] n
$ Restarting dbackup3-agent (via systemctl):                  [  OK  ]
$ [ ok ] Restarting dbackup3-agent (via systemctl): dbackup3-agent.service.
```

3.2.2 FusionInsight Manager 集群环境

MRS 集群客户端是用于 FusionInsight Manager（下文简称 MRS）整个环境备份所需环境变量及相关依赖。在 MRS 8.3.1 版本以下通过手动方式安装，在 MRS 8.3.1 版本及以上版本时，支持远程自动部署。

3.2.2.1 MRS 版本低于 8.3.1

MRS 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 MRS 环境的兼容性问题，一个 MRS 客户端只能对应一个 MRS 集群。如果现场同时存在多个 MRS 集群，则应根据具体需求部署多个 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 MRS 客户端安装目录下的 hosts 的内容到 `/etc/hosts`。
- 将 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 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 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 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
```

3.2.2.2 MRS 版本等于或高于 8.3.1

MRS 版本大于或等于 8.3.1 时，MRS 已支持远程自动部署客户端，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
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

- 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，在主机的终端粘贴安装命令，按回车进行代理端安装。如:

```
curl "http://IP:80/d2/update/script?modules=hadoop&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。

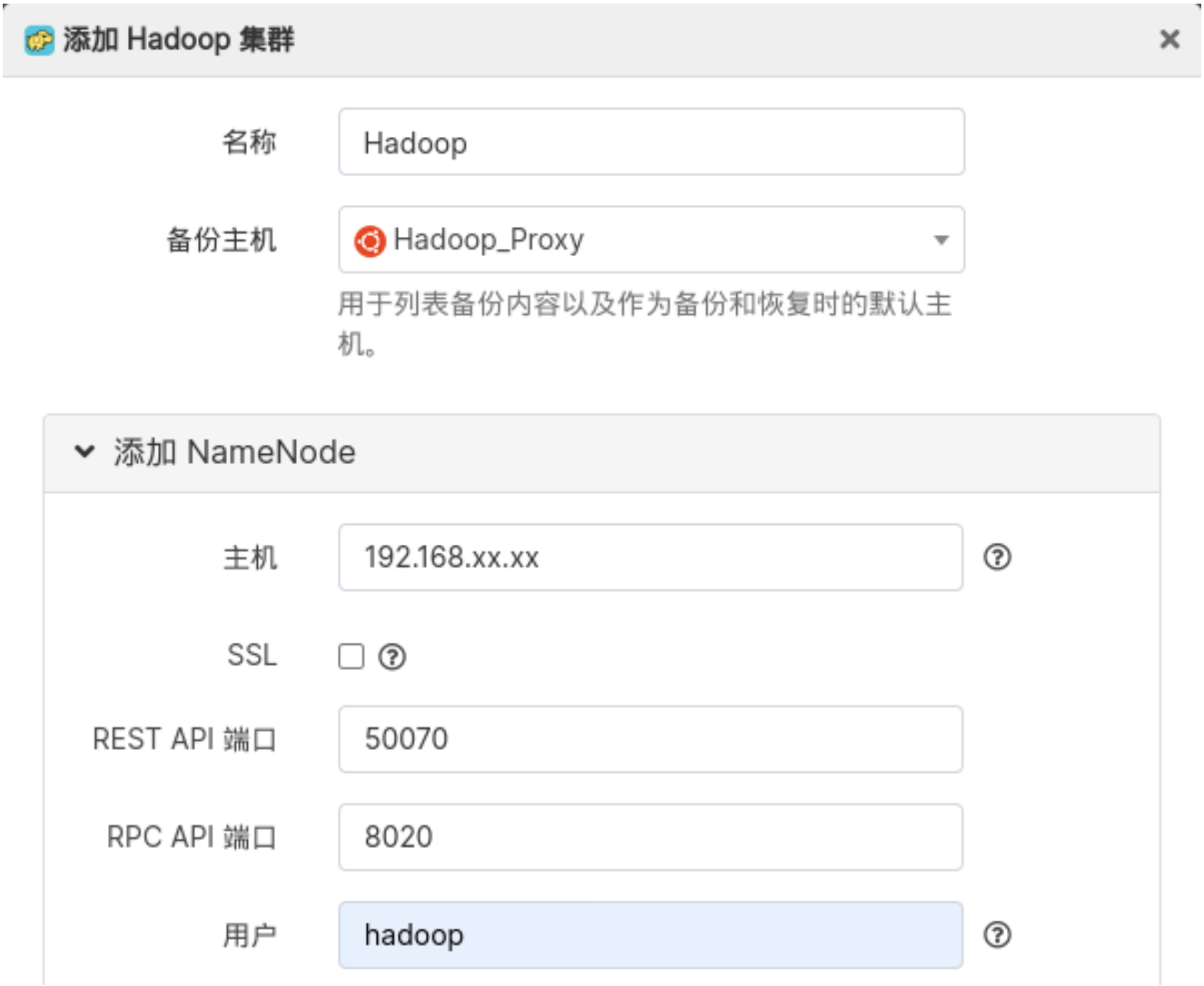
备注：

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。具体请参考《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）。

   （1）创建 Hadoop 资源

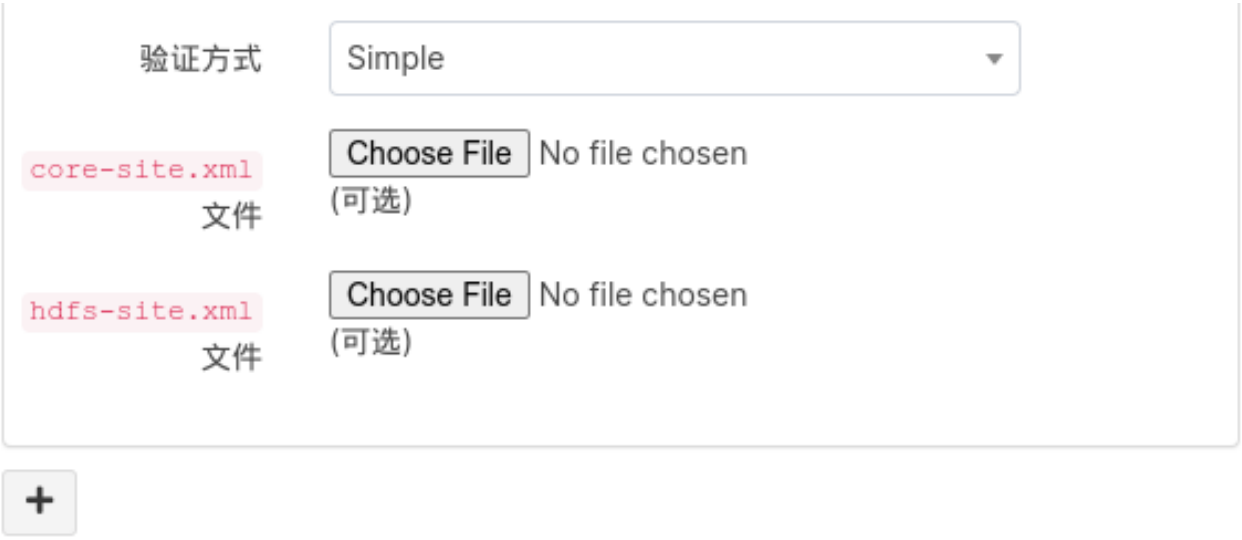


- 名称：自定义资源名称。
- 主机：输入 NameNode 节点所在的主机 IP 或主机名。如果配置 Kerberos 认证时 Principal 使用主机名进行创建，那么该字段需要填写主机名，并且代理端所在的机器的 hosts 文件也要添加该主机的 IP 及其对应的主机名解析。
- 安全连接：使用 SSL 安全连接。该选项需要 Hadoop 集群配置和启用 HTTPS 服务才能使用，否则不需要勾选该选项。
- RPC API 端口：默认为 8020，如果集群配置为其他端口那么该选项需要根据实际端口进行修改。可在 Hadoop 服务 `core-site.xml` 文件里面查看参数 `fs.defaultFS` 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。

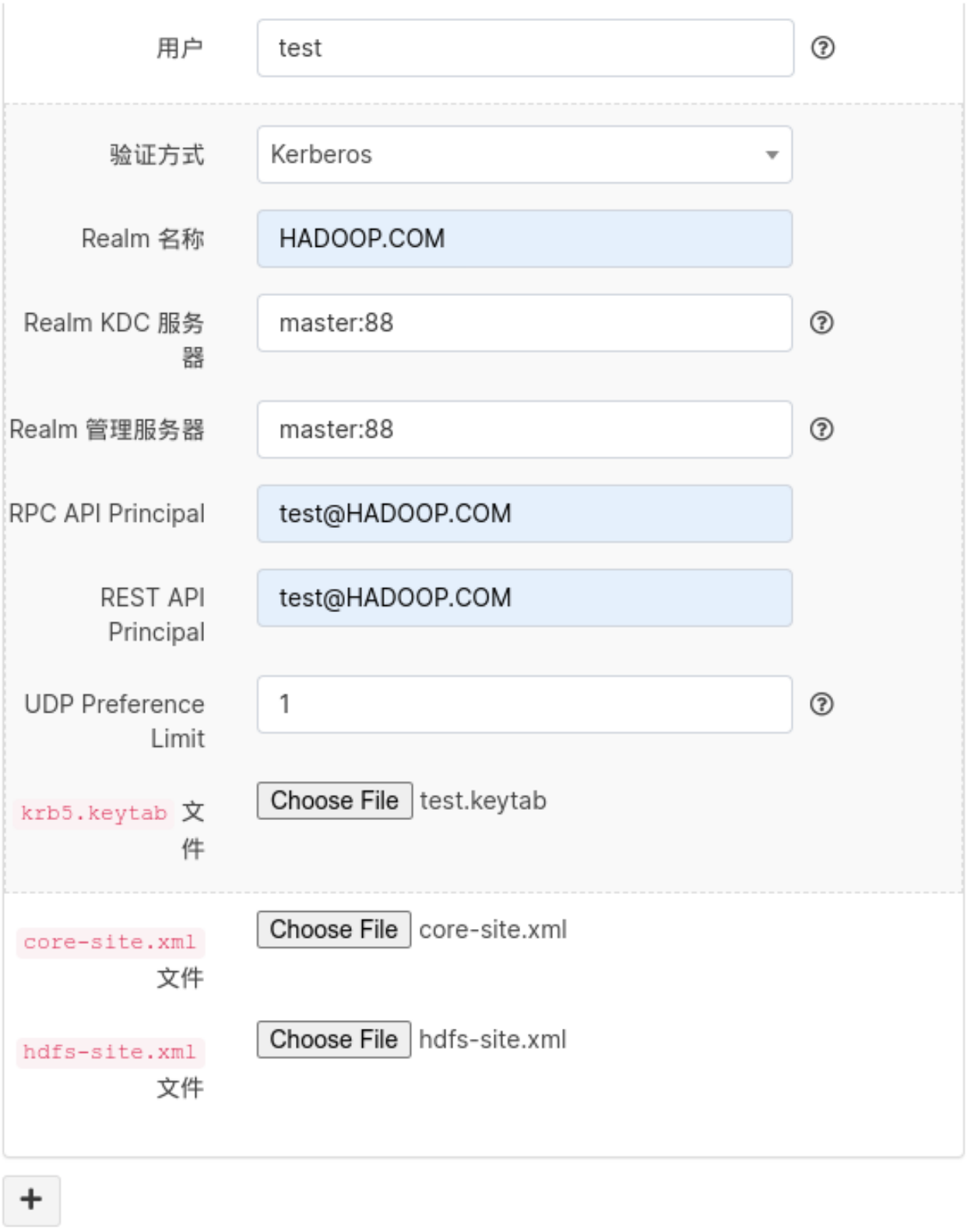
- 用户：HDFS 文件系统所属用户，有 Kerberos 认证，需要填写认证到 keytab 文件里面的 Principal 的用户。如：test@HADOOP.COM，就填写 test。

（2）Simple 验证方式



- 验证方式：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



- 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，使用非默认端口时需填写端口信息。
- Realm 管理服务器：Realm 管理服务器 IP 或主机名。默认端口为：88，使用非默认端口时需填写端口信息。
- RPC API Principal：输入 RPC API Principal 名称。可在 keytab 文件中查询，如：`klist -k -t test.keytab`。
- REST API Principal：输入 REST API Principal 名称。可在 keytab 文件中查询，如：`klist -k -t test.keytab`。

- UDP Preference Limit：指定 UDP 传输包的最大值。默认值为 1，当数据包大于该值时使用 TCP 进行传输，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 验证方式必须上传。

5.2 激活 Hadoop

1. 添加 Hadoop 集群成功后，会弹出【Hadoop 许可证】激活窗口，点击【**激活**】按钮。
2. 激活后，点击 Hadoop 资源【**授权**】按钮，进行授权。
3. 在【**授权**】窗口，可对 Hadoop 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 集群修改。

6.1 备份类型

迪备为 Hadoop 备份提供完全备份、增量备份两种常规的备份类型。除此之外，还提供 Hadoop 的高级备份类型：合成备份。

- 完全备份

备份 Hadoop 目录或文件。对某一个时间点的所有目录文件进行的一个完全拷贝。

- 增量备份

增量备份基于完全备份创建。备份上一次备份后（包含全量备份、增量备份），所有发生变化的文件。

- 合成备份

首次合成备份作业是全备份，后续每次为增量备份。达到合成条件时，最新全备份与后续增量备份合成在一起，生成一个新的全备份副本。合成备份主要用于提高恢复的性能。您可以通过“即时恢复”直接将副本挂载到目标机，无需增加物理拷贝并占用额外的存储空间。

6.2 备份策略

迪备提供 7 种备份计划，立即、一次、手动、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。
- 手动：作业创建后可手动启动作业执行。
- 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。

若要使用高级的合成备份，推荐用户使用以下备份策略：

合成备份：每天执行一次**合成备份**，保证每天有个可恢复的时间点。

6.3 开始之前

在备份恢复 Hadoop 之前，需保证已完成如下操作：

1. 检查存储池

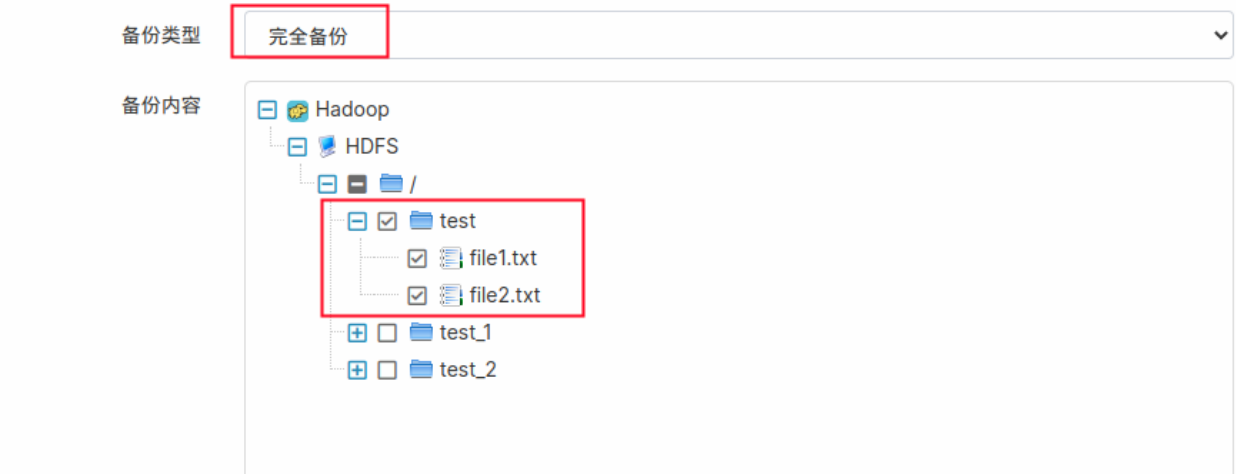
（1）在迪备菜单栏中，点击**【存储池】**，进入**【存储池】**页面。

（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》的创建存储池章节，创建存储池并授权给当前控制台用户。

备注： 合成备份的环境要求比较复杂，必须满足以下条件：
（1）您需要申请迪备的“Hadoop 合成备份”、“Hadoop 副本管理”的高级许可。
（2）您需要在具备管理员权限的用户“存储池”处创建“文件合成池”并确保当前用户可以使用文件合成池。

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**】，勾选您希望备份的文件，点击【**下一步**】。



（1）【**备份类型**】选择完全备份、增量备份或合成备份。

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，无需再次选择目录和文件。

（2）点击 + 可以展开文件夹，勾选备份的文件或文件夹。
（3）如果要过滤【**备份内容**】中选定的文件和文件夹，可点击【**备份内容**】下方的【**过滤器**】，会弹出【**过滤器**】窗口。

- 默认情况下禁用**排除**选项。如果要从【**备份内容**】中排除某些目录或文件，可以勾选**排除**复选框并输入目录路径名和文件名。
- 如果要从**排除**的目录和文件中保留某些目录和文件，您可以选择**包含**复选框并输入目录路径名和文件名。

备注： 例如，有目录 `/test` 和 `/data`，在 `/test` 中有数百个文件。有部分文件是 `.txt`，有部分文件是 `.dat`。需备份整个 `/data` 和 `/test` 下的所有 `.txt` 文件。

1. 首先在【**备份内容**】中选择 `/test` 和 `/data`。然后打开【**过滤器**】窗口。
2. 在**排除**中输入 `/test`。
3. 选择**包含**复选框并输入 `*.txt`。
4. 备份结果将是包含 `/data` 中所有数据和仅包含 `/test` 下的 `.txt` 文件。

对于通配符 * 过滤举例说明，假设存在以下结构目录文件上传到 HDFS：

```
root@ubuntu:/# tree /backup/
/backup/
└── test
        ├── group_1
        │   └── sub_group
        │       ├── file1.dat
```

(续下页)

(接上页)

```
        │       └── file1.txt
        ├── group_2
        │   └── sub_group
        │       ├── file2.dat
        │       └── file2.txt
        └── no_group
```

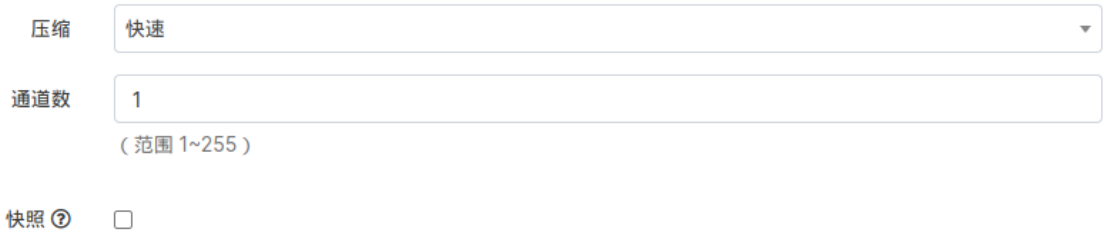
表 1：过滤器例子

排除	包含	结果
/backup/*	/backup/test/group_*/*	备份 group_1 和 group_2 目录及其所有子目录文件。
/backup/*	*.txt	备份以.txt 结尾的文件且包含 no_group 目录。
*.txt	不勾选包含	备份除.txt 结尾的文件以外的所有目录文件。

4. 在【**备份目标**】页面，选择一个备份主机和存储池，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。

5. 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“手动”，作业创建后可手动启动作业执行。
   - 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。
6. 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**】。



7. 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**】。
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。

6.5 备份选项

迪备为 Hadoop 提供以下备份选项：

- 常规选项

表 2：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。	仅完全备份和合成备份支持。
快照	开启该选项可进行 Hadoop 快照备份，默认不开启。	仅完全备份、增量备份和合成备份支持。创建增量备份时无需手动启用【快照】，因为增量备份会自动继承基准备份的快照设置。
不执行备份	根据设置的 HDFS 剩余存储空间阈值，只要任一阈值条件不满足时将不执行备份。该选项要求开启【快照】选项。	仅完全备份、增量备份和合成备份支持。
重删模式	可选择代理端重删或服务端重删。选择代理端重删时，备份数据在代理端进行重删，仅传输唯一数据块至存储服务器；选择服务端重删时，备份数据先传输至存储服务器，再进行重删。为避免在处理重复数据块时（例如代理端压缩或加密）消耗代理端的计算资源，建议仅在首次备份或增量备份等重复数据较少的场景下使用服务端重删。	备份目标中选择存储池为重删池时出现该选项。

- 高级选项：

表 3：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。	
限制传输速度	可分时段限制数据传输速度。单位为 KiB/s、MiB/s 或 GiB/s。	
限制备份速度	可分时段限制磁盘读速度。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。	

续下页

表 3 – 接上页

功能	描述	限制性说明
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。	

7 恢复

针对不同需求，迪备提供多种 Hadoop 的恢复方式，包括：

- 时间点恢复

当 Hadoop 的文件夹或文件丢失时，可以通过时间点恢复功能将 Hadoop 恢复到指定的时间点状态。Hadoop 时间点恢复支持本机和异机恢复，可以原路径恢复或自定义路径恢复。

- 即时恢复

将存储服务器中的 Hadoop 备份集通过挂载方式实现快速恢复。Hadoop 即时恢复具有恢复速度快、资源消耗少、节省磁盘空间以及提高备份集的可用性等优点。

- 演练

结合每小时、每天、每周、每月恢复计划，支持将 Hadoop 的最新备份集周期地恢复到本机其他路径或异机实例。

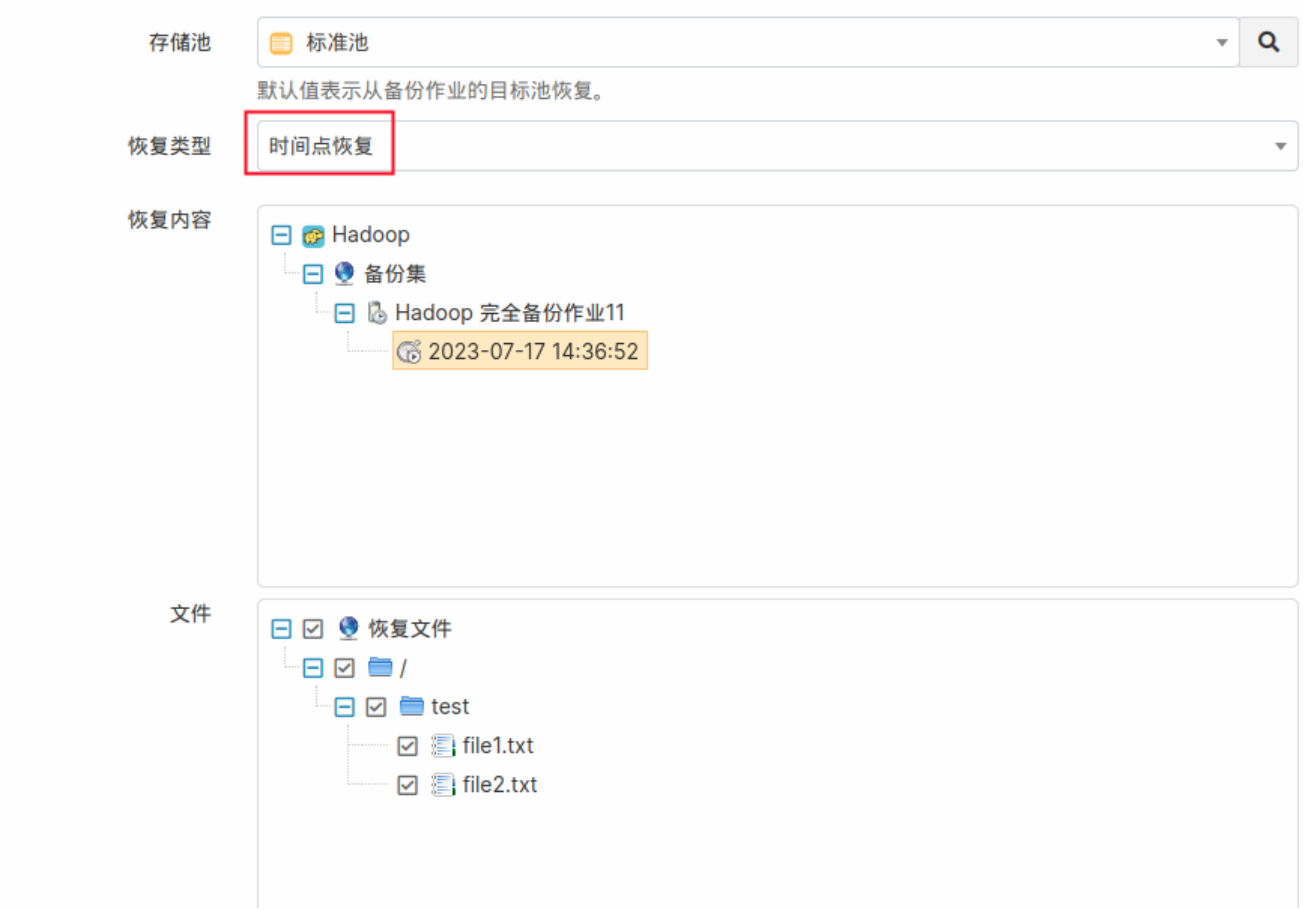
7.1 开始之前

如果要恢复 Hadoop 到其他主机，需先在该主机安装代理端或注册 Hadoop 资源，激活许可证，并将文件或 Hadoop 资源授权给当前迪备控制台用户。

7.2 创建时间点恢复作业

创建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 Hadoop 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：



（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。包括做池复制的源池和目的池。

（2）**【恢复类型】**选择**时间点恢复**。

（3）在**【恢复内容】**列表中，选择需要恢复的备份集时间点。

（4）选择**【文件】**。默认恢复备份集中的所有文件，也可以手动“取消/勾选”选择恢复部分文件。

备注：在恢复页面不支持列出本地存储池和 LAN-free 池备份集中的文件，可以通过选择作业时间点恢复对应的备份集。

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

   备注：如果选择 Hadoop 或 obs 主机和实例，在后续恢复选项页面中则会显示**【备份主机】**选项进行选择。

5. 在**【恢复计划】**页面，选择“立即”、“一次”或“手动”，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“手动”，作业创建后可手动启动作业执行。
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 创建即时恢复作业

备注：

1. Hadoop 即时恢复功能需要在存储服务器安装 dbackup3-nfsd 包。
2. Hadoop 即时恢复功能目前只支持选择标准存储池（未启用多存储或数据存储加密）和文件合成池里的备份集。

创建即时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 Hadoop 所在主机和实例，点击**【下一步】**。
3. 在**【备份集】**页面中，完成以下操作：

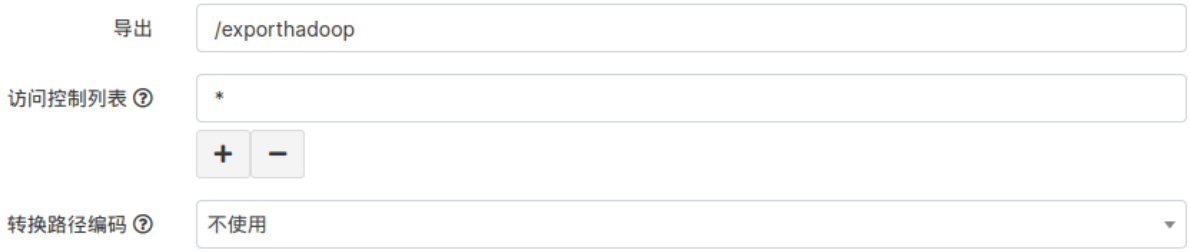
（1）**【存储池】**默认值表示从备份作业的目标池恢复，可选择任意已产生备份集的存储池。包括做池复制的源池和目的池。

（2）**【恢复类型】**选择**即时恢复**。

（3）在**【恢复内容】**列表中，选择需要恢复的备份集时间点。

（4）恢复信息设置完成，点击**【下一步】**。

4. 在**【导出】**页面中，完成以下操作。



（1）【**导出**】设置导出挂载点。路径使用字母、数字，只能以 / 开头，长度为 2-30。
（2）【**访问控制列表**】可以挂载访问该备份集的代理端列表，支持设置指定 IP 或网段，* 表示任何代理端都可以访问。
（3）【**转换路径编码**】文件即时恢复时默认不使用，不使用即为 UTF8 路径编码。可选择 GBK、GB18030 或 BIG5。
（4）【**高级选项**】默认不使用桥接。使用桥接网络导出备份集可以避免与系统的 NFS 服务产生冲突。

备注：

1. 高级选项中选择桥接需输入 IP 地址、子网掩码和默认网关，该 IP 地址必须是该网段未被使用的有效地址。
2. 桥接设置需要在存储服务器安装 bridge-utils，桥接网卡启动后，迪备才能识别到，编辑 `/etc/network/interfaces` 配置文件，添加以下内容：

```
auto br0
iface br0 inet static
address 192.168.88.10
netmask 255.255.255.0
gateway 192.168.88.1
bridge_ports bond0
bridge_stp off
bridge_fd 9
bridge_hello 2
bridge_maxage 12
```

5. 在【**完成**】页面，检查作业信息是否有误。点击【**提交**】。
6. 在【**提交**】后，进入【**副本管理**】页面，会弹出【**手动挂载流程**】帮助文档，备份时间点副本下会增加一条挂载副本，状态为已挂载。可参考查看副本章节。

7.4 创建演练作业

创建演练作业的步骤如下：

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 Hadoop 所在主机和实例，点击【**下一步**】。
3. 在【**备份集**】页面中，完成以下操作：

（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**演练**。

（3）在**【恢复内容】**列表中，选择需要恢复的备份集时间点。

（4）选择**【文件】**。默认恢复备份集中的所有文件，也可以手动“取消/勾选”选择恢复部分文件。演练作业将定期演练恢复所选文件或目录的最新备份集。

备注：在恢复页面不支持列出本地存储池和 LAN-free 池备份集中的文件，可以通过选择作业时间点演练恢复对应的备份集。

（5）恢复信息设置完成，点击【**下一步**】。

4. 在**【恢复目标】**页面，支持恢复到本机异目录或异机，自动跳转**【下一步】**。

   备注：如果选择 Hadoop 或 obs 主机和实例，在后续恢复选项页面中则会显示**【备份主机】**选项进行选择。

5. 在**【恢复计划】**页面，选择周期的演练计划。点击**【下一步】**。
   - 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。

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**】，并检查作业信息是否有误。点击【**提交**】。
8. 提交成功，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。

7.5 恢复选项

迪备为 Hadoop 提供以下恢复选项：

- 常规选项：

表 4：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 Hadoop 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。	
恢复路径	可设置恢复路径为原始路径或自定义路径，自定义路径可手动输入或者点击浏览在弹出框中直接选择目标文件夹。	
增量恢复	只有选择“增量备份集”时才有此项，默认不勾选。勾选后只恢复选中时间点的增量数据。	仅时间点恢复支持。
同名文件处理方式	当恢复路径中出现同名文件时，执行该选项的处理方式，默认覆盖	仅支持“覆盖”、“跳过”、“保留最新”三种策略。
恢复副本数	恢复时每个目标文件在集群中存储的副本数量。auto：默认值 auto 代表恢复个数使用 HDFS 配置文件中的参数值。自定义：提供自定义恢复副本数范围为 1~3，减少副本个数可提升恢复速度并减少空间占用，但也增加了数据丢失的风险，请根据实际需求选择。	当 DataNode 的节点数少于 3 时，恢复出来的文件副本数只能小于等于 DataNode 的节点数

- 高级选项：

表 5：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。	

续下页

表 5 – 接上页

功能	描述	限制性说明
限制传输速度	可分时段限制数据传输速度。单位为 KiB/s、MiB/s 或 GiB/s。	
限制恢复速度	可分时段限制磁盘写速度。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。	
无效路径处理	不检查与转换路径合法性。 忽略含有非法字符的路径。 移除非法字符。 转义非法字符。	

8 副本管理

操作员用户可以通过副本管理界面对合成备份、即时恢复产生的数据副本进行管理，包括查看、创建、卸载、删除副本等操作。

8.1 查看副本

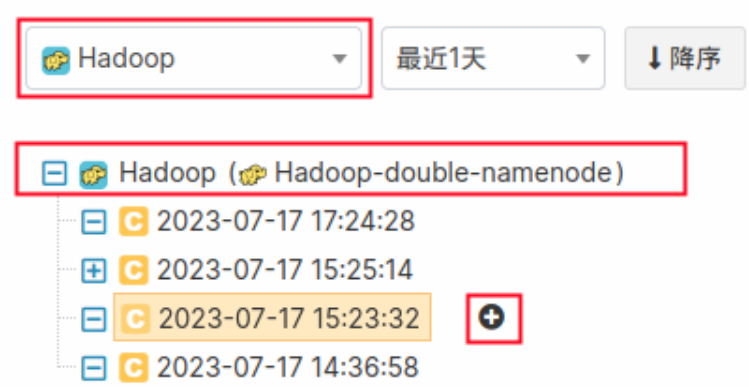
查看副本的步骤如下：

1. 在菜单栏中，点击【**副本管理**】，进入【**副本管理**】页面。
2. 在工具栏中，选择主机的 Hadoop 实例，设置副本生成的时间段。展示区会显示该实例在相应时间段内生成的副本。
3. 点击副本名称，页面右侧会显示该副本的详细信息。数据副本以创建时间命名，不同图标表示各种副本类型，包括：
   - 全备份副本：合成备份生产的数据副本。
   - 挂载副本：即时恢复生成的数据副本。

8.2 克隆副本

您可以直接通过【**克隆副本**】对 Hadoop 实例的合成副本创建即时恢复作业，生成一个新的挂载副本。步骤如下：

1. 在菜单栏中，点击【**副本管理**】，进入【**副本管理**】页面。
2. 在工具栏中，选择主机的 Hadoop 实例，在展示区会显示该实例在相应时间段内生成的副本。
3. 在展示区，点击 Hadoop 实例下面以创建时间命名的全副本。实例右侧会显示【**克隆副本**】按钮。



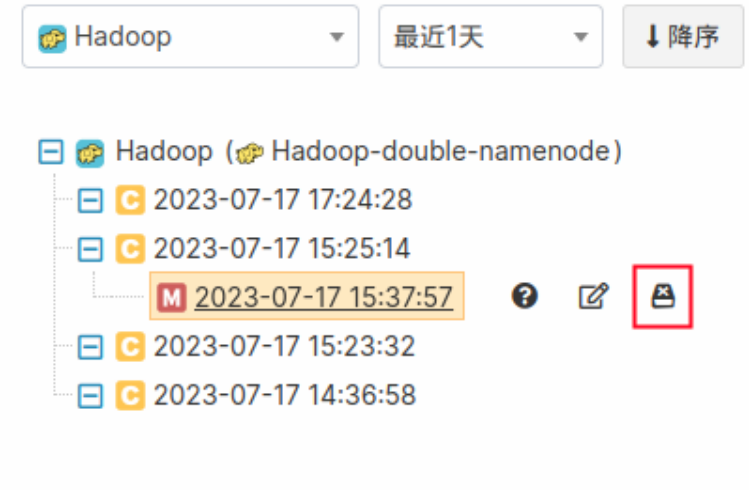
4. 点击【**克隆副本**】按钮，进入【**副本管理**】页面，参考创建即时恢复作业 操作步骤，依次设置该文件实例即时恢复作业的信息。
5. 即时恢复执行成功后，在【**副本管理**】页面可以查询全备份副本下增加一条挂载副本，状态为已挂载。

8.3 卸载副本

您可以使用【**卸载**】按钮对已挂载的副本进行解挂。这个操作会导致恢复目标机挂载的目录无法访问。

步骤如下：

1. 在菜单栏中，点击【**副本管理**】，进入【**副本管理**】页面。
2. 在工具栏中，选择主机的 Hadoop 实例，设置副本生成的时间段，在展示区会显示该实例在相应时间段内生成的副本。
3. 展开全备份副本，选择全备份副本下已挂载的副本。挂载副本右侧会显示【**卸载**】按钮。



4. 点击【**卸载**】按钮，弹出确认窗口。
5. 确认警告提示，输入验证码后，点击【**确定**】。
6. 卸载成功后，可以查询全备份副本下无此挂载副本记录。

表 6：限制性

功能	限制描述
Hadoop 时间点恢复	不支持 Hadoop HDFS 文件系统备份集与 Windows 平台文件互相恢复。
Hadoop 即时恢复	仅支持 Linux 存储服务器。 支持标准存储池（未启用多存储或数据存储加密）。 支持文件合成池。
Hadoop 演练	不支持 Hadoop HDFS 备份集演练恢复到 Windows 系统。 不支持 Hadoop HDFS 备份集演练恢复到对象存储。
Hadoop 备份恢复	Agent 需要能够通过网络访问 NameNode、DataNode。
恢复副本数	不适用华为 MRS 环境，因为该环境默认禁止自定义文件副本数，并且无法解除此限制。

10 术语表

表 7：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。
跨系统恢复	Hadoop 和 Linux 系统互相恢复。
异构恢复	Hadoop 可恢复至对象存储或操作系统的文件。

SCUTECH
鼎甲科技

全国销售热线：400-650-0081　　全国服务热线：400-003-3191

电话：+86 20 32053160　　网址：www.scutech.com

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