在源和目标之间同步数据
Cloud Sync

NetApp
April 21, 2022

目录

在源和目标之间同步数据

创建同步关系

创建同步关系时、 Cloud Sync 服务会将文件从源复制到目标。初始副本完成后、服务将每 24 小时同步所有更改的数据。

在创建某些类型的同步关系之前，您需要先在 Cloud Manager 中创建一个工作环境。

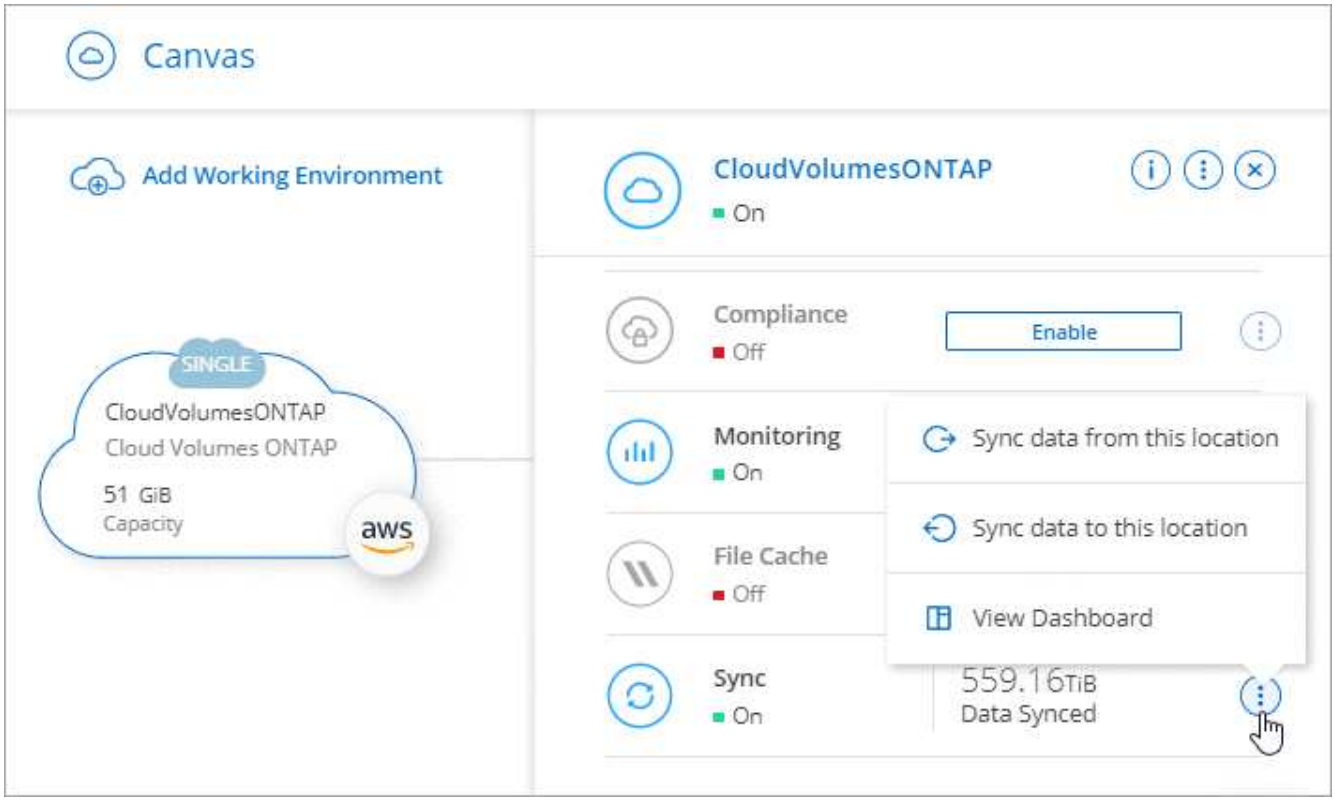
为特定类型的工作环境创建同步关系

如果要为以下任一项创建同步关系，则首先需要创建或发现工作环境：

- 适用于 ONTAP 的 Amazon FSX
- Azure NetApp Files
- Cloud Volumes ONTAP
- 内部 ONTAP 集群

步骤

1. 创建或发现工作环境。
    - "创建适用于 ONTAP 的 Amazon FSX 工作环境"
    - "设置和发现 Azure NetApp Files"
    - "在 AWS 中启动 Cloud Volumes ONTAP"
    - "在 Azure 中启动 Cloud Volumes ONTAP"
    - "在 Google Cloud 中启动 Cloud Volumes ONTAP"
    - "添加现有 Cloud Volumes ONTAP 系统"
    - "发现 ONTAP 集群"
2. 单击 * 画布 * 。
3. 选择与上述任何类型匹配的工作环境。
4. 选择 Sync 旁边的操作菜单。



5. 选择 * 从此位置同步数据 * 或 * 将数据同步到此位置 * ，然后按照提示设置同步关系。

创建其他类型的同步关系

使用以下步骤将数据同步到或不同步适用于 ONTAP ， Azure NetApp Files ， Cloud Volumes ONTAP 或内部 ONTAP 集群的 Amazon FSX 以外的受支持存储类型。以下步骤提供了一个示例，说明如何设置从 NFS 服务器到 S3 存储分段的同步关系。

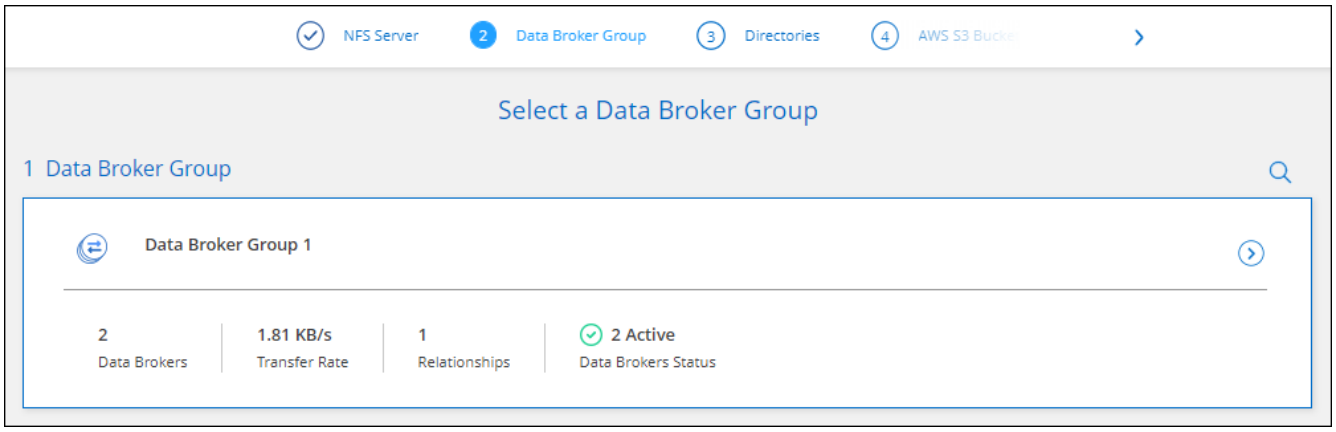
1. 在 Cloud Manager 中，单击 * 同步 * 。
2. 在 * 定义同步关系 * 页面上，选择源和目标。

   以下步骤提供了如何创建从 NFS 服务器到 S3 存储区的同步关系的示例。

3. 在 * NFS Server* 页面上，输入要同步到 AWS 的 NFS 服务器的 IP 地址或完全限定域名。
4. 在 * 数据代理组 * 页面上，按照提示在 AWS ， Azure 或 Google Cloud Platform 中创建数据代理虚拟机，或者在现有 Linux 主机上安装数据代理软件。

   有关详细信息，请参阅以下页面：

   - "在 AWS 中创建数据代理"
   - "在 Azure 中创建数据代理"
   - "在 Google Cloud 中创建数据代理"
   - "在 Linux 主机上安装数据代理"
5. 安装数据代理后，单击 * 继续 * 。

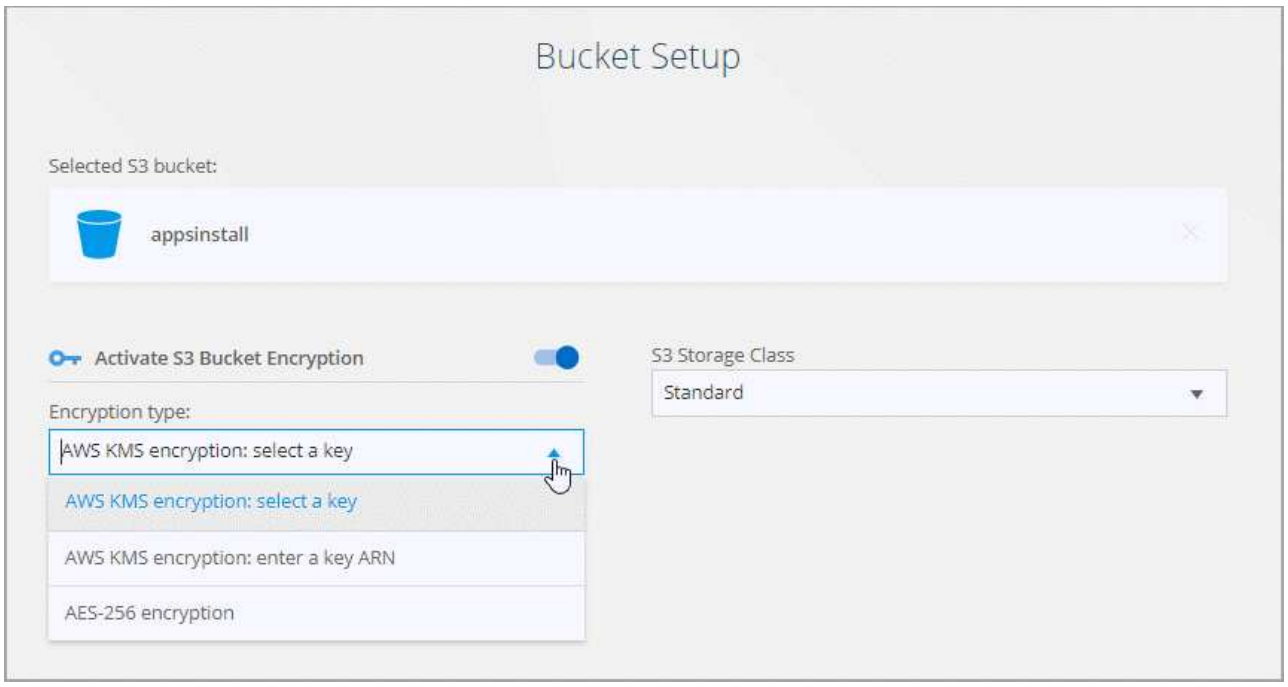


6. 在 * 目录 * 页面上，选择顶级目录或子目录。

   如果 Cloud Sync 无法检索导出，请单击 * 手动添加导出 * 并输入 NFS 导出的名称。

   如果要同步 NFS 服务器上的多个目录、则必须在完成后创建其他同步关系。

7. 在 * AWS S3 Bucket* 页面上，选择一个存储分段：
   - 向下钻取以选择存储桶中的现有文件夹或选择在存储桶内创建的新文件夹。
   - 单击 * 添加到列表 * 以选择与您的 AWS 帐户无关的 S3 存储分段。 "必须将特定权限应用于 S3 存储区。"。
8. 在 * 分段设置 * 页面上，设置分段：
   - 选择是否启用 S3 存储分段加密，然后选择 AWS KMS 密钥，输入 KMS 密钥的 ARN 或选择 AES-256 加密。
   - 选择 S3 存储类。 "查看支持的存储类。"。



9. 【设置】在 * 设置 * 页面上，定义源文件和文件夹在目标位置的同步和维护方式：

计划

为将来的同步选择重复计划或关闭同步计划。您可以计划一个关系以每 1 分钟同步一次数据。

重试

定义在跳过文件之前、 Cloud Sync 应重试同步文件的次数。

比较依据

选择 Cloud Sync 在确定文件或目录是否已更改并应重新同步时是否应比较某些属性。

即使取消选中这些属性， Cloud Sync 仍会通过检查路径，文件大小和文件名来将源与目标进行比较。如果有任何更改，则会同步这些文件和目录。

您可以选择启用或禁用 Cloud Sync 以比较以下属性：

- * mtime* ：文件的上次修改时间。此属性对目录无效。
- * uid* ， * gid* 和 * 模式 * ： Linux 的权限标志。

复制对象

启用此选项可复制对象存储元数据和标记。如果用户更改了源上的元数据，则 Cloud Sync 会在下次同步时复制此对象，但如果用户更改了源上的标记（而不是数据本身），则 Cloud Sync 不会在下次同步时复制对象。

创建关系后，您无法编辑此选项。

包含 S3 兼容端点（ S3 ， StorageGRID 或 IBM 云对象存储）的同步关系支持复制标记。

以下任一端点之间的 " 云到云 " 关系支持复制元数据：

- AWS S3
- Azure Blob
- Google Cloud 存储
- IBM 云对象存储
- StorageGRID

最近修改的文件

选择排除在计划同步之前最近修改的文件。

删除源上的文件

选择在 Cloud Sync 将文件复制到目标位置后从源位置删除文件。此选项包括数据丢失的风险，因为源文件会在复制后被删除。

如果启用此选项，则还需要更改数据代理上 local.json 文件中的参数。打开文件并按如下所示进行更新：

```
{
"workers":{
"transferrer":{
"delete-on-source": true
}
}
}
```

删除目标上的文件

如果文件已从源文件中删除，请选择从目标位置删除这些文件。默认情况下，从不从目标位置删除文件。

文件类型

定义要包括在每个同步中的文件类型：文件、目录和符号链接。

排除文件扩展名

通过键入文件扩展名并按 * 输入 * 来指定要从同步中排除的文件扩展名。例如，键入 *log* 或 *.log* 排除 * 。log 文件。多个扩展不需要分隔符。以下视频提供了简短演示：

► https://docs.netapp.com/zh-cn/cloud-manager-sync//media/video_file_extensions.mp4 *(video)*

文件大小

选择同步所有文件、无论文件大小如何、还是仅同步特定大小范围内的文件。

修改日期

选择所有文件，无论其上次修改日期、在特定日期之后修改的文件、特定日期之前或时间范围之间的文件。

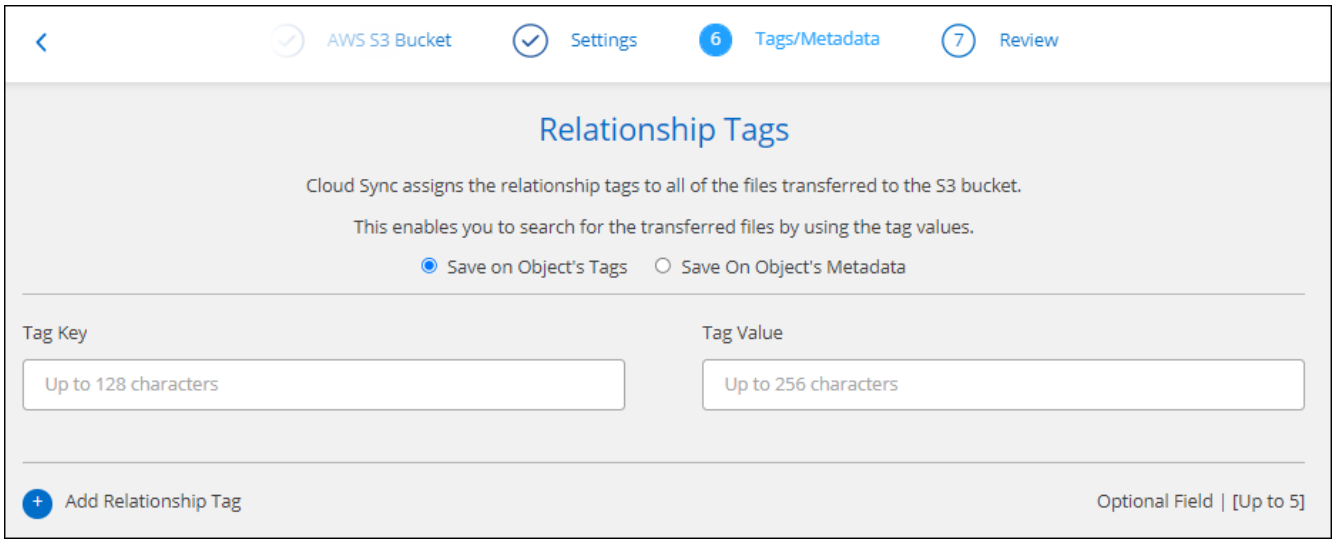
创建日期

如果 SMB 服务器是源服务器，则可以通过此设置在特定日期之后，特定日期之前或特定时间范围之间同步创建的文件。

ACL —访问控制列表

通过在创建关系时或创建关系后启用设置，从 SMB 服务器复制 ACL 。

10. 在 * 标记 / 元数据 * 页面上，选择是将密钥值对另存为传输到 S3 存储分段的所有文件的标记，还是为所有文件分配元数据密钥值对。



将数据同步到 StorageGRID 和 IBM 云对象存储时，也可以使用此功能。对于 Azure 和 Google Cloud Storage ，只有元数据选项可用。

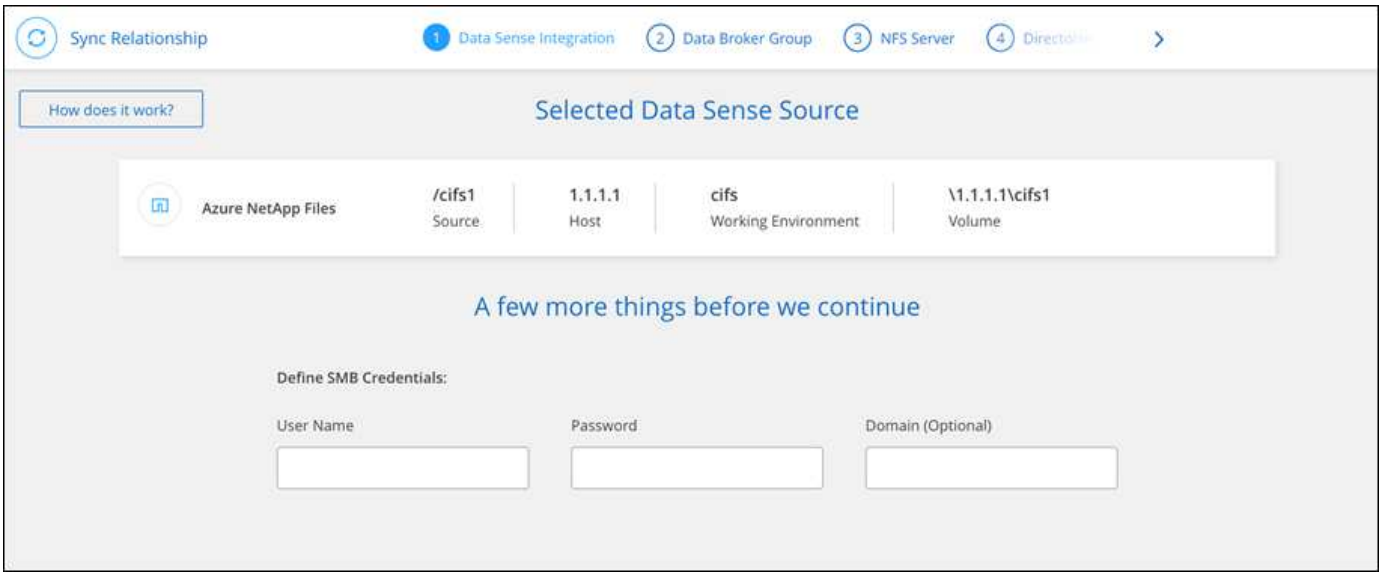
11. 查看同步关系的详细信息，然后单击 * 创建关系 * 。
    - 结果 *

Cloud Sync 开始在源和目标之间同步数据。

从 **Cloud Data sense** 创建同步关系

Cloud Sync 与云数据感知相集成。在 Data sense 中，您可以使用 Cloud Sync 选择要同步到目标位置的源文件。

从 Cloud Data sense 启动数据同步后，所有源信息都包含在一个步骤中，只需输入几个关键详细信息即可。然后，选择新同步关系的目标位置。



"了解如何从 Cloud Data sense 启动同步关系"。

从 SMB 共享复制 ACL

Cloud Sync 可以在源 SMB 共享和目标 SMB 共享之间复制访问控制列表（ACL），也可以从源 SMB 共享复制到对象存储（ONTAP S3 除外）。如果需要，您还可以选择使用 Robocopy 在 SMB 共享之间手动保留 ACL 。

Cloud Sync 不支持将 ACL 从对象存储复制回 SMB 共享。

选项

- 设置 Cloud Sync 以自动复制 ACL
- 在 SMB 共享之间手动复制 ACL

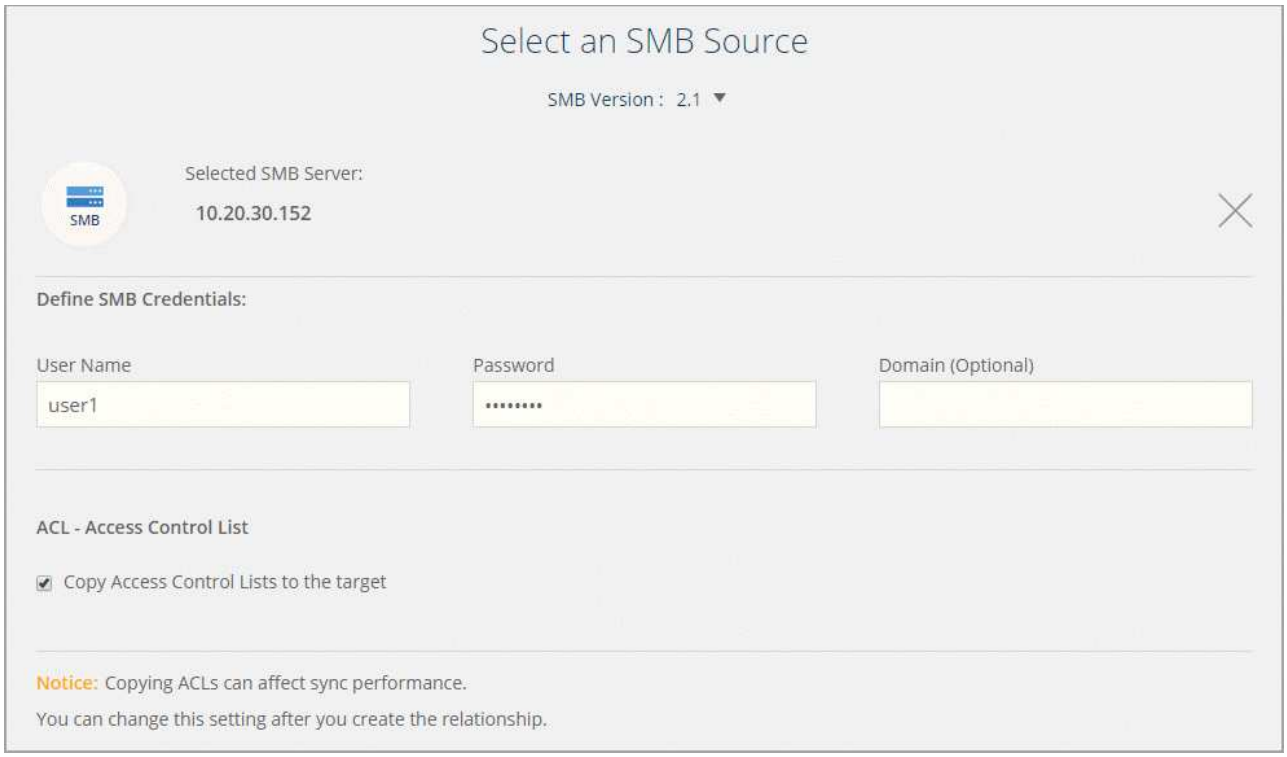
设置 Cloud Sync 以从 SMB 服务器复制 ACL

通过在创建关系时或创建关系后启用设置，从 SMB 服务器复制 ACL 。

此功能适用于 *any* 类型的数据代理： AWS ， Azure ， Google Cloud Platform 或内部数据代理。内部数据代理可以运行 "任何受支持的操作系统"。

新关系的步骤

1. 在 Cloud Sync 中，单击 * 创建新同步 * 。
2. 将 * SMB Server* 拖放到源，选择 SMB 服务器或对象存储作为目标，然后单击 * 继续 * 。
3. 在 * SMB Server* 页面上：
   a. 输入新的 SMB 服务器或选择现有服务器，然后单击 * 继续 * 。
   b. 输入 SMB 服务器的凭据。
   c. 选择 * 将访问控制列表复制到目标 * ，然后单击 * 继续 * 。



4. 按照其余提示创建同步关系。

   将 ACL 从 SMB 复制到对象存储时，您可以选择将 ACL 复制到对象的标记或对象的元数据上，具体取决于目标。对于 Azure 和 Google Cloud Storage ，只有元数据选项可用。

   以下屏幕截图显示了一个可以选择此操作的步骤示例。

现有关系的步骤

1. 将鼠标悬停在“同步关系”上，然后单击“操作”菜单。
2. 单击 * 设置 * 。
3. 选择 * 将访问控制列表复制到目标 * 。
4. 单击 * 保存设置 * 。

同步数据时， Cloud Sync 会保留源和目标 SMB 共享之间的 ACL ，或者从源 SMB 共享保留到对象存储。

在 SMB 共享之间手动复制 ACL

您可以使用 Windows robocopy 命令手动保留 SMB 共享之间的 ACL 。

步骤

1. 确定具有对 SMB 共享的完全访问权限的 Windows 主机。
2. 如果任一端点需要身份验证，请使用 * 网络使用 * 命令从 Windows 主机连接到这些端点。

   在使用 Robocopy 之前，必须执行此步骤。

3. 从 Cloud Sync 中，在源设备和目标 SMB 共享之间创建新关系或同步现有关系。
4. 数据同步完成后、从 Windows 主机运行以下命令以同步 ACL 和所有权：

   ```
   robocopy /E /COPY:SOU /secfix [source] [target] /w:0 /r:0 /XD ~snapshots
   /UNILOG:”[logfilepath]
   ```

   源 *sourc* 和目标 _ 均应使用 UNC 格式指定。例如： \\<server> \<share> \<path>

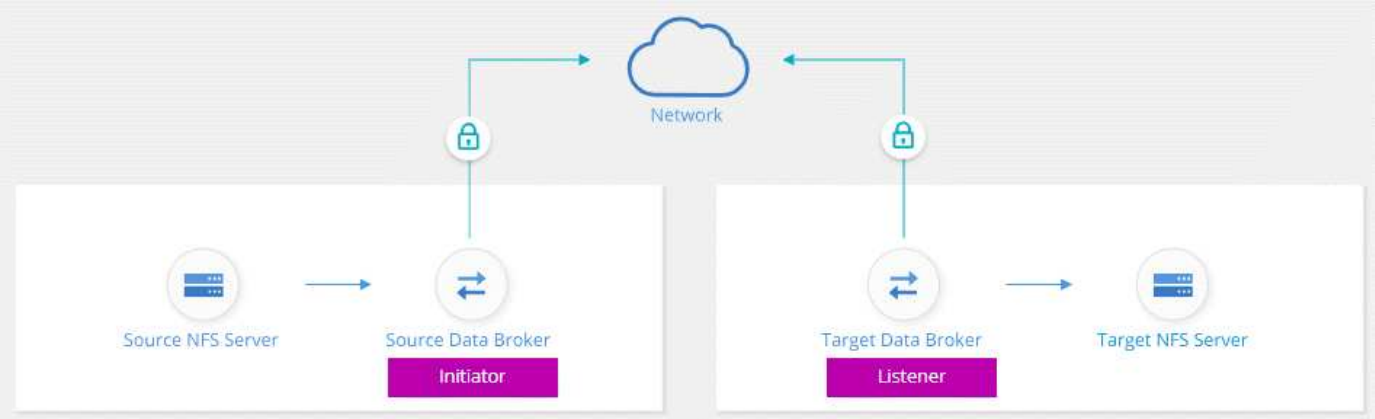
使用传输中数据加密同步 NFS 数据

如果您的企业拥有严格的安全策略，则可以使用数据传输加密来同步 NFS 数据。从 NFS 服务器到另一个 NFS 服务器以及从 Azure NetApp Files 到 Azure NetApp Files 均支持此功能。

例如，您可能希望在位于不同网络的两个 NFS 服务器之间同步数据。或者，您可能需要在子网或区域之间安全地传输 Azure NetApp Files 上的数据。

飞行中数据加密的工作原理

传输中数据加密在两个数据代理之间通过网络发送 NFS 数据时对其进行加密。下图显示了两个 NFS 服务器和两个数据代理之间的关系：



一个数据代理充当 *initiator* 。同步数据时，它会向另一个数据代理（即 *listener* ）发送连接请求。该数据代理会侦听端口 443 上的请求。如果需要，您可以使用其他端口，但请务必检查此端口是否未被其他服务使用。

例如，如果将数据从内部 NFS 服务器同步到基于云的 NFS 服务器，则可以选择哪个数据代理侦听连接请求并将其发送。

以下是动态加密的工作原理：

1. 创建同步关系后、启动程序将启动与其他数据代理的加密连接。
2. 源数据代理使用 TLS 1.3 对源中的数据进行加密。
3. 然后通过网络将数据发送到目标数据代理。
4. 目标数据代理会先对数据进行解密，然后再将其发送到目标。
5. 初始副本完成后、服务将每 24 小时同步所有更改的数据。如果有要同步的数据，则进程将从启动程序打开与其他数据代理的加密连接开始。

   如果您希望更频繁地同步数据， "您可以在创建关系后更改此计划"。

支持的 NFS 版本

- 对于 NFS 服务器， NFS 版本 3 ， 4.0 ， 4.1 和 4.2 支持传输中数据加密。
- 对于 Azure NetApp Files ， NFS 版本 3 和 4.1 支持传输中数据加密。

代理服务器限制

如果创建加密同步关系，加密数据将通过 HTTPS 发送，不能通过代理服务器路由。

您将需要什么才能开始

请确保具有以下内容：

- 两个 NFS 服务器相结合 "源和目标要求" 或 Azure NetApp Files 。
- 服务器的 IP 地址或完全限定域名。
- 两个数据代理的网络位置。

  您可以选择现有的数据代理、但它必须作为启动程序运行。侦听器数据代理必须是 *new* 数据代理。

  如果要使用现有数据代理组，该组必须只有一个数据代理。加密的同步关系不支持组中的多个数据代理。

  如果尚未部署数据代理、请查看数据代理要求。由于您具有严格的安全策略，请务必查看网络要求，其中包括端口 443 和的出站流量 "Internet 端点" 数据代理所联系的。

  - "查看 AWS 安装"
  - "查看 Azure 安装"
  - "查看 Google Cloud 安装"
  - "查看 Linux 主机安装"

使用传输中数据加密同步 **NFS** 数据

在两个 NFS 服务器之间或 Azure NetApp Files 之间创建新的同步关系，启用正在传输的加密选项，然后按照提示进行操作。

步骤

1. 单击 * 创建新同步 * 。
2. 将 * NFS Server* 拖放到源位置和目标位置，或者将 * Azure NetApp Files * 拖放到源位置和目标位置，然后选择 * 是 * 以启用传输中数据加密。
3. 按照提示创建关系：

   a. * NFS Server*/* Azure NetApp Files* ： 选择 NFS 版本，然后指定新的 NFS 源或选择现有服务器。

   b. * 定义数据代理功能 * ： 定义哪个数据代理侦听端口上的连接请求以及哪个数据代理启动连接。根据您的网络要求做出选择。

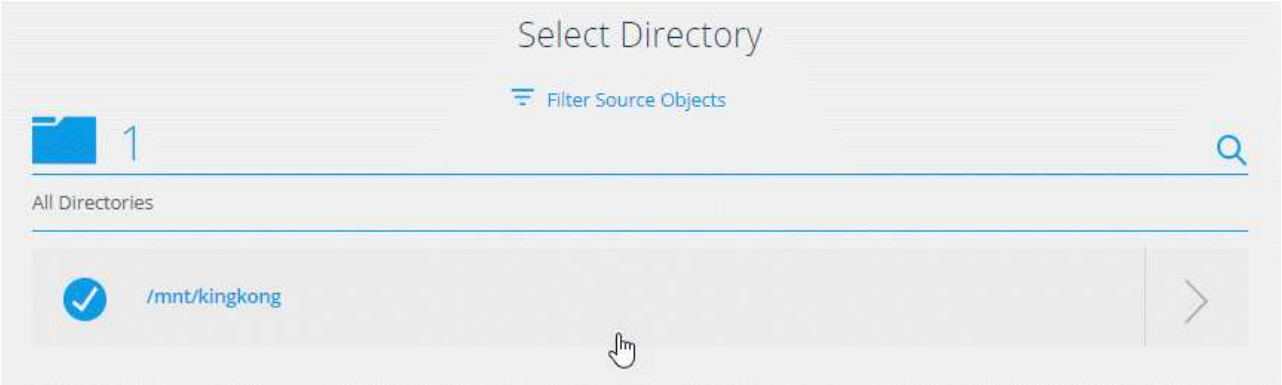
   c. * 数据代理 * ： 按照提示添加新的源数据代理或选择现有数据代理。

      请注意以下事项：

      - 如果要使用现有数据代理组，该组必须只有一个数据代理。加密的同步关系不支持组中的多个数据代理。
      - 如果源数据代理充当侦听程序、则它必须是新的数据代理。
      - 如果需要新的数据代理、 Cloud Sync 会提示您安装说明。您可以在云中部署数据代理或为自己的 Linux 主机下载安装脚本。

   d. * 目录 * ： 选择要同步的目录，方法是选择所有目录，或者向下钻取并选择子目录。

      单击 * 筛选源对象 * 可修改用于定义源文件和文件夹在目标位置的同步和维护方式的设置。

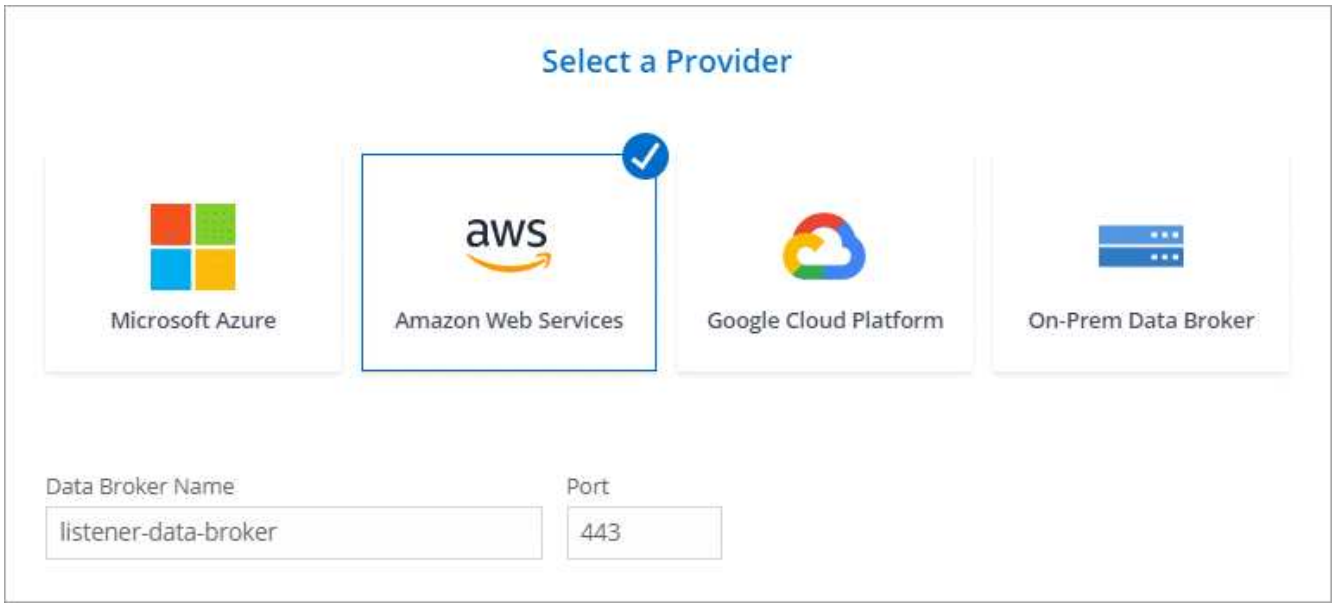
      

   e. * 目标 NFS 服务器 / 目标 NFS* ： 选择 Azure NetApp Files 版本，然后输入新的 目标或选择现有服务器。

   f. * 目标数据代理 * ： 按照提示添加新的源数据代理或选择现有数据代理。

      如果目标数据代理充当侦听程序、则它必须是新的数据代理。

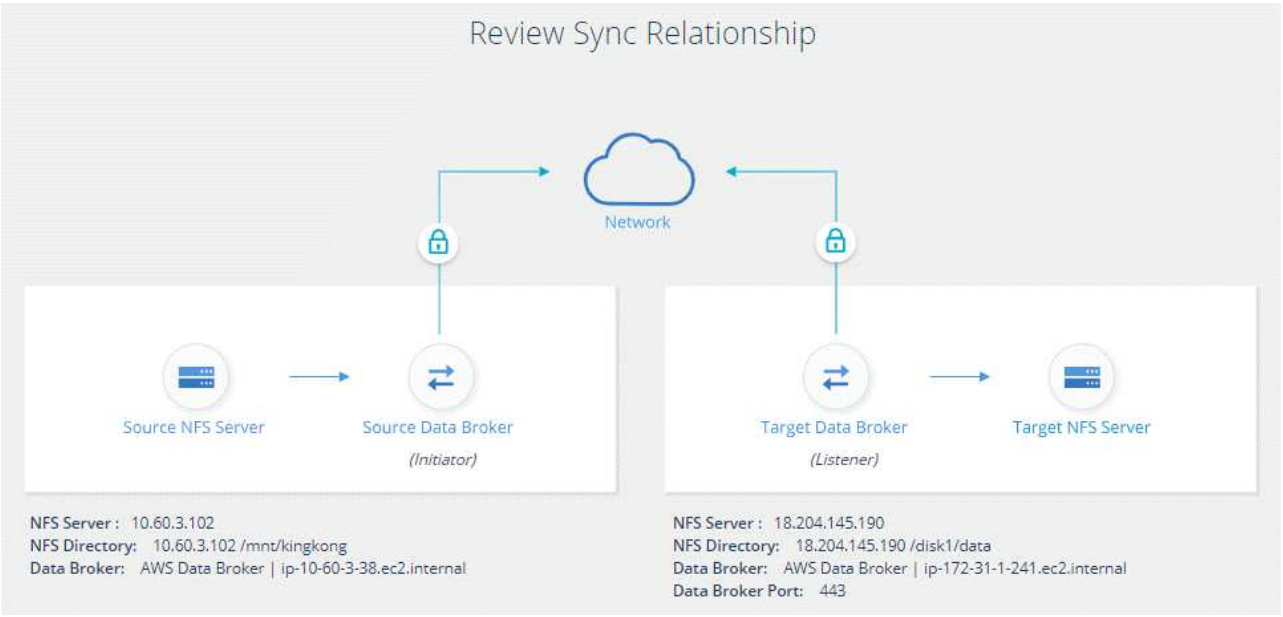
      以下是目标数据代理充当侦听器时的提示示例。请注意指定端口的选项。



a. * 目标目录 * ： 选择顶级目录，或者向下钻取以选择现有子目录或在导出中创建新文件夹。

b. * 设置 * ： 定义如何在目标位置同步和维护源文件和文件夹。

c. * 审阅 * ： 查看同步关系的详细信息，然后单击 * 创建关系 * 。



Cloud Sync 开始创建新的同步关系。完成后，单击 * 在信息板中查看 * 以查看有关新关系的详细信息。

设置数据代理组以使用外部 HashiCorp 存储

创建需要 Amazon S3 ， Azure 或 Google Cloud 凭据的同步关系时，您需要通过 Cloud Sync 用户界面或 API 指定这些凭据。另一种方法是设置数据代理组，以便直接从外部 HashiCorp 存储访问凭据（或 *sects* ）。

此功能通过 Cloud Sync API 提供支持，并且同步关系需要 Amazon S3 ， Azure 或 Google Cloud 凭据。

通过设置 URL 来准备存储以向数据代理组提供凭据。存储中的机密 URL 必须以 _Creds_ 结尾 。

通过修改组中每个数据代理的本地配置文件，使数据代理组做好准备，以便从外部存储提取凭据。

设置完所有内容后，您可以发送 API 调用来创建同步关系，以使用存储获取密码。

准备存储

您需要为 Cloud Sync 提供存储中机密的 URL 。通过设置这些 URL 来准备存储。您需要为计划创建的同步关系中的每个源和目标的凭据设置 URL 。

必须按如下所示设置 URL ：

` /<path>/<RequestId>/<Endpoint-protocol>Creds`

路径
: 密钥的前缀路径。这可以是您独有的任何值。

请求 **ID**
: 需要生成的请求 ID 。创建同步关系时，您需要在 API POST 请求中的一个标题中提供 ID 。

端点协议
: 定义的以下协议之一 "在关系后 v2 文档中"： S3 ， Azure 或 GCP （每个都必须大写）。

创建
: URL 必须以 _Creds_ 结尾 。

示例

以下示例显示了指向机密的 URL 。

源凭据的完整 **URL** 和路径示例
: http://example.vault.com:8200/my-path/all-secrets/hb312vdasr2/S3Creds

  如示例中所示，前缀路径为 //my-path/all-sects/ ，请求 ID 为 *hb312vdasr2* ，源端点为 S3 。

目标凭据的完整 **URL** 和路径示例
: http://example.vault.com:8200/my-path/all-secrets/n32hcbnejk2/AZURECreds

  前缀路径为 //my-path/all-sections/_ ，请求 ID 为 *n32hcbnejk2* ，目标端点为 Azure 。

正在准备数据代理组

通过修改组中每个数据代理的本地配置文件，使数据代理组做好准备，以便从外部存储提取凭据。

步骤

1. 通过 SSH 连接到组中的数据代理。
2. 编辑 /opt/netapp/databroker/config 中的 local.json 文件。
3. 将 enable 设置为 * true * ，并按如下所示在 *external*-积分 *.hashashicorp* 下设置配置参数字段：

enabled

- 有效值： true/false
- 类型：布尔值
- 默认值： false
- true ： 数据代理从您自己的外部 HashiCorp Vault 获取机密
- false ： 数据代理将凭据存储在其本地存储中

url

- 键入： string
- value ： 外部存储的 URL

path

- 键入： string
- value ： 使用凭据将路径前缀为密钥

拒绝 **-** 未授权

- 确定是否希望数据代理拒绝未经授权的外部存储
- 类型：布尔值
- 默认值： false

auth-method

- 数据代理从外部存储访问凭据时应使用的身份验证方法
- 键入： string
- 有效值： "AWS-iam"/"role-app"/"GCP-iam"

角色名称

- 键入： string
- 您的角色名称（如果您使用 AWS-iam 或 GCP-iam ）

Secretid 和 **rootid**

- type： string （如果使用 app-role ）

命名空间

- 键入： string
- 命名空间（如果需要，则为 X-Vault-Namespace 标头）

4. 对组中的任何其他数据代理重复上述步骤。

AWS 角色身份验证示例

```
{
          “external-integrations”: {
                  “hashicorp”: {
                          “enabled”: true,
                          “url”: “https://example.vault.com:8200”,
                          “path”: ““my-path/all-secrets”,
                          “reject-unauthorized”: false,
                          “auth-method”: “aws-role”,
                          “aws-role”: {
                                “role-name”: “my-role”
                          }
                  }
        }
}
```

GCP-iam 身份验证示例

```
{
"external-integrations": {
    "hashicorp": {
      "enabled": true,
      "url": http://ip-10-20-30-55.ec2.internal:8200,
      "path": "v1/secret",
      "namespace": "",
      "reject-unauthorized": true,
      "auth-method": "gcp-iam",
      "aws-iam": {
        "role-name": ""
      },
      "app-role": {
        "root_id": "",
        "secret_id": ""
      },
"gcp-iam": {
          "role-name": "my-iam-role"
      }
    }
  }
}
```

使用 **GCP-iam** 身份验证时设置权限

如果您使用的是 *gcp-iam* 身份验证方法，则数据代理必须具有以下 GCP 权限：

```
- iam.serviceAccounts.signJwt
```

"详细了解数据代理的 GCP 权限要求"。

使用存储中的密钥创建新的同步关系

设置完所有内容后，您可以发送 API 调用来创建同步关系，以使用存储获取密码。

使用 Cloud Sync REST API 发布关系。

```
Headers:
Authorization: Bearer <user-token>
Content-Type: application/json
x-account-id: <accountid>
x-netapp-external-request-id-src: request ID as part of path for source
credentials
x-netapp-external-request-id-trg: request ID as part of path for target
credentials
Body: post relationship v2 body
```

- 要获取用户令牌和 Cloud Central 帐户 ID， "请参见文档中的此页面"。
- 为您的后关系构建实体， "请参见 relationships-v2 API 调用"。

示例

POST 请求示例：

```
url: https://api.cloudsync.netapp.com/api/relationships-v2
headers:
"x-account-id": "CS-SasdW"
"x-netapp-external-request-id-src": "hb312vdasr2"
"Content-Type": "application/json"
"Authorization": "Bearer eyJhbGciOiJSUzI1NiIsInR5cCI6IkpXVCIsImtpZCI6Ik…"
Body:
{
"dataBrokerId": "5e6e111d578dtyuu1555sa60",
"source": {
        "protocol": "s3",
        "s3": {
                "provider": "sgws",
                "host": "1.1.1.1",
                "port": "443",
                "bucket": "my-source"
     },
"target": {
        "protocol": "s3",
        "s3": {
                "bucket": "my-target-bucket"
        }
    }
}
```