



管理数据分类

NetApp Data Classification

NetApp
January 14, 2026

目录

管理数据分类	1
从NetApp Data Classification扫描中排除特定目录	1
支持的数据源	1
定义要排除在扫描之外的目录	1
示例	2
转义文件夹名称中的特殊字符	3
查看当前排除列表	3
在NetApp Data Classification其他组 ID 定义为对组织开放	4
为群组 ID 添加“向组织开放”权限	4
查看当前组ID列表	4
在NetApp Data Classification中自定义过期数据定义	5
从NetApp Data Classification中删除数据源	5
停用系统扫描	6
从数据分类中删除数据库	6
从数据分类中删除一组文件共享	6
卸载NetApp Data Classification	6
从云提供商处卸载数据分类	6
从本地部署中卸载数据分类	7

管理数据分类

从NetApp Data Classification扫描中排除特定目录

如果您希望NetApp Data Classification从扫描中排除特定目录，则可以将这些目录名称添加到配置文件中。应用此更改后，数据分类引擎将从扫描中排除这些目录。



默认情况下，数据分类扫描排除与卷中的源相同的卷快照数据。

支持的数据源

以下数据源中的 NFS 和 CIFS 共享支持从数据分类扫描中排除特定目录：

- 本地ONTAP
- Cloud Volumes ONTAP
- Amazon FSx for NetApp ONTAP
- Azure NetApp Files
- 常规文件共享

定义要排除在扫描之外的目录

在将目录排除在分类扫描之外之前，您需要登录数据分类系统，以便可以编辑配置文件并运行脚本。了解如何[“登录数据分类系统”](#)取决于您是否在 Linux 机器上手动安装了该软件，或者是否在云中部署了该实例。

注意事项

- 每个数据分类系统最多可以排除 50 个目录路径。
- 排除目录路径可能会影响扫描时间。

步骤

1. 在数据分类系统上，转到“/opt/netapp/config/custom_configuration”，然后打开文件 `data_provider.yaml`。
2. 在“exclude:”行下的“data_providers”部分中，输入要排除的目录路径。例如：

```
exclude:  
- "folder1"  
- "folder2"
```

请勿修改此文件中的任何其他内容。

3. 保存对文件的更改。
4. 转到“/opt/netapp/Datasense/tools/customer_configuration/data_providers”并运行以下脚本：

```
update_data_providers_from_config_file.sh
```

- + 此命令将要排除在扫描范围之外的目录提交给分类引擎。

结果

对您的数据进行的所有后续扫描都将排除对这些指定目录的扫描。

您可以使用相同的步骤从排除列表中添加、编辑或删除项目。运行脚本提交更改后，修改后的排除列表将会更新。

示例

配置1：

名称中包含“folder1”的每个文件夹都将被排除在所有数据源之外。

```
data_providers:  
  exclude:  
    - "folder1"
```

将被排除的路径的预期结果：

- /CVO1/文件夹1
- /CVO1/文件夹1名称
- /CVO1/文件夹10
- /CVO1/*文件夹1
- /CVO1/+文件夹1名称
- /CVO1/notfolder10
- /CVO22/文件夹1
- /CVO22/文件夹1名称
- /CVO22/文件夹10

不会被排除的路径示例：

- /CVO1/*文件夹
- /CVO1/文件夹名称
- /CVO22/*folder20

配置2：

仅在名称开头包含“*folder1”的每个文件夹都将被排除。

```
data_providers:  
  exclude:  
    - "\\\*folder1"
```

将被排除的路径的预期结果：

- /CVO/*文件夹1
- /CVO/*文件夹1名称
- /CVO/*folder10

不会被排除的路径示例：

- /CVO/文件夹1
- /CVO/文件夹1名称
- /CVO/not*folder10

配置3：

数据源“CVO22”中名称中包含“folder1”的每个文件夹都将被排除。

```
data_providers:  
  exclude:  
    - "CVO22/folder1"
```

将被排除的路径的预期结果：

- /CVO22/文件夹1
- /CVO22/文件夹1名称
- /CVO22/文件夹10

不会被排除的路径示例：

- /CVO1/文件夹1
- /CVO1/文件夹1名称
- /CVO1/文件夹10

转义文件夹名称中的特殊字符

如果您的文件夹名称包含以下特殊字符之一，并且您想要排除该文件夹中的数据进行扫描，则需要在文件夹名称前使用转义序列 \।

例如：
., +, *, ?, ^, \$, (,), [,], {, }, |

源中的路径： /project/*not_to_scan

排除文件中的语法： "*not_to_scan"

查看当前排除列表

内容可能 `data_provider.yaml` 配置文件与运行后实际提交的内容不同
`update_data_providers_from_config_file.sh` 脚本。要查看已从数据分类扫描中排除的当前目录列表，请从“/opt/netapp/Datasense/tools/customer_configuration/data_providers”运行以下命令：

```
get_data_providers_configuration.sh
```

在NetApp Data Classification其他组 ID 定义为对组织开放

当组 ID (GID) 附加到 NFS 文件共享中的文件或文件夹时，它们定义了文件或文件夹的权限；例如它们是否“对组织开放”。如果某些 GID 最初未设置“向组织开放”权限级别，您可以向 GID 添加该权限，以便任何附加了该 GID 的文件和文件夹都将被视为“向组织开放”。

在您进行此更改并且NetApp Data Classification重新扫描您的文件和文件夹后，任何附加了这些组 ID 的文件和文件夹都将在“调查详细信息”页面中显示此权限，并且它们还将出现在您显示文件权限的报告中。

要激活此功能，您需要登录数据分类系统，以便可以编辑配置文件并运行脚本。了解如何[“登录数据分类系统”](#)取决于您是否在 Linux 机器上手动安装了该软件，或者是否在云中部署了该实例。

为群组 ID 添加“向组织开放”权限

在开始此任务之前，您需要有组 ID 号 (GID)。

步骤

1. 在数据分类系统上，转到“/opt/netapp/config/custom_configuration”并打开文件 `data_provider.yaml`。
2. 在“organization_group_ids: []”行中添加组 ID。例如：

```
organization_group_ids: [1014, 1015, 21, 2021, 1013, 2020, 1018, 1019]
```

不要更改此文件中的任何其他内容。

3. 保存对文件的更改。
4. 转到“/opt/netapp/Datasense/tools/customer_configuration/data_providers”并运行以下脚本：

```
update_data_providers_from_config_file.sh
```

此命令将修改后的组 ID 权限提交给分类引擎。

结果

对您的数据进行的所有后续扫描都将识别出附加有这些组 ID 且标记为“向组织开放”的文件或文件夹。

您可以使用相同的步骤编辑组 ID 列表并删除过去添加的任何组 ID。运行脚本提交更改后，修改后的组 ID 列表将会更新。

查看当前组ID列表

内容可能 `data_provider.yaml` 配置文件与运行后实际提交的内容不同
`update_data_providers_from_config_file.sh` 脚本。要查看已添加到数据分类的当前组 ID 列表，请

从“/opt/netapp/Datasense/tools/customer_configuration/data_providers”运行以下命令：

```
get_data_providers_configuration.sh
```

在NetApp Data Classification中自定义过期数据定义

NetApp Data Classification可识别过时数据，帮助您发现节省成本的机会和治理风险。由于不同组织环境中对过时数据的定义可能有所不同，因此您可以自定义数据分类如何定义过时数据。

过期数据可以根据其上次访问时间或上次修改时间来定义。时间段选择范围从 6 个月前到 10 年前。

默认情况下，如果数据上次修改时间距今已有三年，则该数据被视为过期数据。

定义过期数据

1. 在勒索软件恢复能力中，选择配置。
2. 在“配置”页面中，滚动到“过期数据定义”标题。
3. 在“文件属性”下拉菜单中，选择是否要根据文件的“上次访问时间”或“上次修改时间”来定义过期数据。
4. 选择过期数据定义的时间段。

The screenshot shows the 'Scanner Groups' configuration page. Under 'Scanner Group: default', there is a table with one scanner node listed. The table columns are Host Name, IP, Status, Last Active Time, and Error. The node 'ip-10-128-0-46.us-west-2.compute.internal' is marked as ACTIVE and last active on 2025-08-31 08:24. Below the table is a button labeled 'Activate Slow Scan'. The 'Stale data definition' section follows, which includes a 'File property' dropdown set to 'Last Modified' and a 'Time period' dropdown set to '3 Years ago'. A note below states 'Current definition: Files modified more than 3 years ago will be marked as stale'. At the bottom of this section is a link 'Uninstall Data Classification'.

5. 选择保存。

从NetApp Data Classification中删除数据源

如果需要，您可以停止NetApp Data Classification扫描一个或多个系统、数据库或文件共

享组。

停用系统扫描

当您停用扫描时，数据分类不再扫描系统上的数据，并且它会从数据分类实例中删除索引的见解。系统本身的数据不会被删除。

1. 从“配置”页面中，选择  系统行中的按钮，然后*停用数据分类*。



您还可以在选择系统时从“服务”面板禁用系统扫描。

从数据分类中删除数据库

如果您不再需要扫描某个数据库，您可以从数据分类界面将其删除并停止所有扫描。

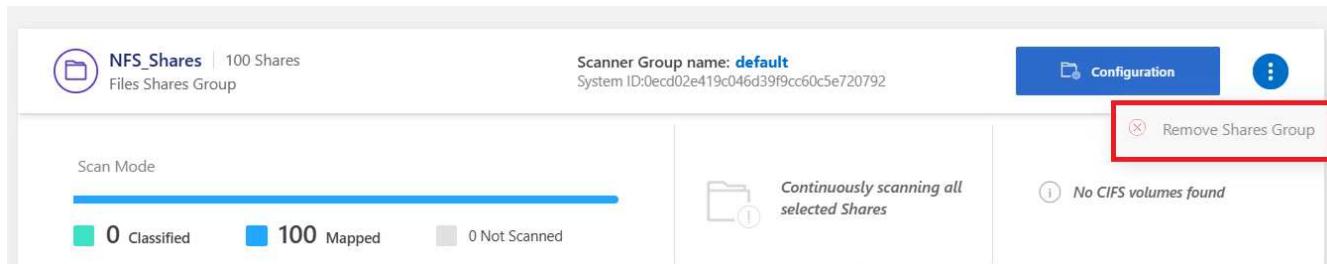
1. 从“配置”页面中，选择  数据库行中的按钮，然后*删除数据库服务器*。

从数据分类中删除一组文件共享

如果您不再想从文件共享组扫描用户文件，您可以从数据分类界面删除文件共享组并停止所有扫描。

步骤

1. 从“配置”页面中，选择  文件共享组行中的按钮，然后单击*删除文件共享组*。



2. 从确认对话框中选择*删除共享组*。

卸载NetApp Data Classification

您可以卸载NetApp Data Classification来解决问题或从主机中永久删除该软件。删除实例还会删除索引数据所在的关联磁盘，这意味着数据分类扫描的所有信息都将被永久删除。

您需要使用的步骤取决于您是在云中还是在本地主机上部署数据分类。

从云提供商处卸载数据分类

1. 从数据分类中，选择配置。
2. 在配置页面底部，选择卸载分类。

The screenshot shows the Data Classification configuration interface. On the left, a sidebar lists categories: Classification, Governance, Compliance, Investigation, Custom classification, New name, and Configuration. The Configuration tab is selected. The main area displays two scanner groups: 'Test' and 'testk'. Each group has 0 shares. Below the groups is a table titled 'Scanner Groups' showing one scanner node. The node details are: Host Name: ip-10-000-0-01.us-west.compute.internal, IP: 10.000.0.01, Status: ACTIVE, Last Active Time: 2025-08-31 08:24. At the bottom, there is a red button labeled 'Uninstall Classification'.

3. 在对话框中，输入“卸载”以继续断开数据分类实例与控制台代理的连接。选择卸载进行确认。
4. 在“卸载分类”对话框中，键入“卸载”以确认您要断开数据分类实例与控制台代理的连接，然后选择“卸载”。
5. 要完成卸载过程，请转到云提供商的控制台并删除数据分类实例。该实例名为 *CloudCompliance*，并带有与之连接的生成的哈希值（UUID）。例如：*CloudCompliance-16bb6564-38ad-4080-9a92-36f5fd2f71c7*

从本地部署中卸载数据分类

1. 从数据分类中，选择配置。
2. 在配置页面底部，选择卸载分类。

This screenshot is identical to the one above, showing the Data Classification configuration interface with two scanner groups ('Test' and 'testk') and a table of scanner nodes. The 'Uninstall Classification' button is also present at the bottom.

3. 在对话框中，输入“卸载”以继续断开数据分类实例与控制台代理的连接。选择卸载进行确认。

4. 要从主机卸载软件，请运行 `cleanup.sh` 数据分类主机上的脚本，例如：

```
cleanup.sh
```

该脚本位于 `/install/light_probe/onprem_installer/cleanup.sh` 目录。了解如何["登录数据分类主机"](#)。

版权信息

版权所有 © 2026 NetApp, Inc.。保留所有权利。中国印刷。未经版权所有者事先书面许可，本文档中受版权保护的任何部分不得以任何形式或通过任何手段（图片、电子或机械方式，包括影印、录音、录像或存储在电子检索系统中）进行复制。

从受版权保护的 NetApp 资料派生的软件受以下许可和免责声明的约束：

本软件由 NetApp 按“原样”提供，不含任何明示或暗示担保，包括但不限于适销性以及针对特定用途的适用性的隐含担保，特此声明不承担任何责任。在任何情况下，对于因使用本软件而以任何方式造成的任何直接性、间接性、偶然性、特殊性、惩罚性或后果性损失（包括但不限于购买替代商品或服务；使用、数据或利润方面的损失；或者业务中断），无论原因如何以及基于何种责任理论，无论出于合同、严格责任或侵权行为（包括疏忽或其他行为），NetApp 均不承担责任，即使已被告知存在上述损失的可能性。

NetApp 保留在不另行通知的情况下随时对本文档所述的任何产品进行更改的权利。除非 NetApp 以书面形式明确同意，否则 NetApp 不承担因使用本文档所述产品而产生的任何责任或义务。使用或购买本产品不表示获得 NetApp 的任何专利权、商标权或任何其他知识产权许可。

本手册中描述的产品可能受一项或多项美国专利、外国专利或正在申请的专利的保护。

有限权利说明：政府使用、复制或公开本文档受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中“技术数据权利 — 非商用”条款第 (b)(3) 条规定的限制条件的约束。

本文档中所含数据与商业产品和/或商业服务（定义见 FAR 2.101）相关，属于 NetApp, Inc. 的专有信息。根据本协议提供的所有 NetApp 技术数据和计算机软件具有商业性质，并完全由私人出资开发。美国政府对这些数据的使用权具有非排他性、全球性、受限且不可撤销的许可，该许可既不可转让，也不可再许可，但仅限在与交付数据所依据的美国政府合同有关且受合同支持的情况下使用。除本文档规定的情形外，未经 NetApp, Inc. 事先书面批准，不得使用、披露、复制、修改、操作或显示这些数据。美国政府对国防部的授权仅限于 DFARS 的第 252.227-7015(b)（2014 年 2 月）条款中明确的权利。

商标信息

NetApp、NetApp 标识和 <http://www.netapp.com/TM> 上所列的商标是 NetApp, Inc. 的商标。其他公司和产品名称可能是其各自所有者的商标。