NetApp Data Classification文档

NetApp Data Classification

NetApp
February 06, 2026

目录

NetApp Data Classification文档

发行说明

NetApp Data Classification的新功能

了解NetApp Data Classification的新功能。

2026年1月14日

1.50 版

本次数据分类版本更新包含错误修复和以下更新：

自定义分类改进

数据分类功能现在支持创建自定义数据类别。您可以上传文件来微调数据分类所使用的 AI 模型，以便将类别标记应用于数据。所有自定义分类的界面都得到了改进。

有关详细信息，请参阅 "创建自定义分类"。

自定义过期数据定义

数据分类功能现在允许您自定义过期数据的定义，以满足您组织的需求。此前，过时数据被定义为三年前最后一次修改的数据。现在，可以根据上次访问时间或上次修改时间来识别过时的数据；时间段可以从 6 个月前到 10 年前不等。

有关详细信息，请参阅 "自定义过期数据定义"。

性能提升

数据分类、数据映射报告和调查页面上的筛选器等所有页面的加载时间都已缩短。

调查报告预计完成时间

下载调查报告时，“数据分类”现在会显示下载完成的预计时间。

2025年12月8日

1.49 版

本次数据分类版本更新包含错误修复和以下更新：

在健康监测仪表板中监控指标和性能

数据分类现在提供了一个健康监控仪表板，可以实时监控您的资源，并提供有关内存使用情况、磁盘使用情况、磁盘利用率等方面的见解。借助健康监控仪表板提供的信息，您可以查看部署的基础架构，并获得优化存储和性能的见解。

有关详细信息，请参阅 "监控数据分类的运行状况"。

提升了装载性能

数据分类中所有页面的加载性能均已得到提升，从而创造了更高效的用户体验。

2025年11月10日

1.48 版

本次数据分类版本更新包括错误修复、安全改进和性能提升。

增强扫描进度清晰度

扫描配置现在包含对扫描完成情况的更深入洞察。以前，进度条只在扫描进行时显示。现在，扫描完成后进度条仍然可见，以确认扫描已成功完成。您还可以查看已映射和已扫描的文件数量。

有关扫描设置的更多信息，请参阅 "更改存储库的NetApp Data Classification扫描设置"。

2025年10月6日

1.47 版

BlueXP classification现为**NetApp Data Classification**

BlueXP classification已重命名为NetApp Data Classification。除了重命名之外，用户界面也得到了增强。

BlueXP现在是**NetApp Console**

BlueXP已重新命名并重新设计，以更好地反映其在管理数据基础设施中的作用。

NetApp Console提供企业级跨本地和云环境的存储和数据服务的集中管理，提供实时洞察、更快的工作流程和简化的管理。

有关更改的详细信息，请参阅 "NetApp Console发行说明"。

增强调查体验

使用新的可搜索过滤器、每个值的结果计数、总结关键发现的实时见解以及具有可自定义列和滑出详细信息窗格的刷新结果表，更快地查找和理解您的数据。

有关更多信息，请参阅"调查数据" 。

新的治理与合规仪表板

通过直观的小部件、更清晰的视觉效果和改进的加载性能更快地获得关键见解。有关详细信息，请参阅"审查有关您的数据的治理信息"和"查看有关您的数据的合规性信息"。

已保存查询的策略（预览）

数据分类现在使您能够通过条件操作实现治理自动化。您可以创建保留规则，设置自动删除和定期电子邮件通知，所有这些都可以通过更新的已保存查询页面进行管理。

有关更多信息，请参阅"创建策略" 。

操作（预览）

从调查页面直接控制 - 单独或批量删除、移动、复制或标记文件，以实现高效的数据管理和补救。

有关更多信息，请参阅"调查数据" 。

支持**Google Cloud NetApp Volumes**

数据分类现在支持在Google Cloud NetApp Volumes上进行扫描。从NetApp Console轻松添加Google Cloud

NetApp Volumes，实现无缝数据扫描和分类。有关详细信息，请参阅 "扫描Google Cloud NetApp Volumes"。

2025年8月11日

1.46 版

此数据分类版本包括错误修复和以下更新：

审计页面中增强的扫描事件洞察

审计页面现在支持对BlueXP classification的扫描事件的增强洞察。审计页面现在显示系统扫描的开始时间、系统状态以及任何问题。共享和系统的状态仅适用于映射扫描。

有关审计页面的更多信息，请参阅"监控NetApp Console操作"。

支持 RHEL 9.6

此版本增加了对 Red Hat Enterprise Linux v9.6 的支持，用于手动本地安装BlueXP classification，包括暗站部署。

以下操作系统需要使用 Podman 容器引擎，并且需要BlueXP classification版本 1.30 或更高版本：Red Hat Enterprise Linux 版本 8.8、8.10、9.0、9.1、9.2、9.3、9.4 和 9.5。

2025年7月14日

1.45 版

此BlueXP classification版本包括优化资源利用率的代码更改以及：

改进了添加文件共享进行扫描的工作流程

将文件共享添加到文件共享组的工作流程已经简化。该流程现在还根据身份验证类型（Kerberos 或 NTLM）区分 CIFS 协议支持。

有关更多信息，请参阅"扫描文件共享" 。

增强文件所有者信息

您现在可以在“调查”选项卡中查看有关捕获的文件的文件所有者的更多信息。在“调查”选项卡中查看文件的元数据时，找到文件所有者，然后选择“查看详细信息”以查看用户名、电子邮件和 SAM 帐户名称。您还可以查看该用户拥有的其他物品。此功能仅适用于具有 Active Directory 的工作环境。

有关更多信息，请参阅"调查组织中存储的数据" 。

2025年6月10日

1.44 版

此次BlueXP classification发布包括：

改进了治理仪表板的更新时间

治理仪表板各个组件的更新时间已得到改善。下表显示了每个组件的更新频率。

组件	更新时间
数据时代	24 小时
类别	24 小时
数据概览	5分钟
重复文件	2 小时
文件类型	24 小时
非业务数据	2 小时
开放权限	24 小时
保存的搜索	2 小时
敏感数据和广泛权限	24 小时
数据大小	24 小时
陈旧数据	2 小时
按敏感度级别划分的顶级数据存储库	2 小时

您可以查看上次更新的时间，并按敏感度级别手动更新重复文件、非业务数据、已保存的搜索、陈旧数据和顶级数据存储库组件。有关治理仪表板的更多信息，请参阅"查看有关组织中存储的数据的治理详细信息"。

性能和安全性改进

已做出改进以提高BlueXP分类的性能、内存消耗和安全性。

错误修复

Redis 已升级，以提高BlueXP classification的可靠性。BlueXP classification现在使用 Elasticsearch 来提高扫描期间文件计数报告的准确性。

2025年5月12日

1.43 版

本次BlueXP分类版本包含：

优先进行分类扫描

数据分类除了支持仅映射扫描之外，还支持对映射和分类扫描进行优先排序的功能，使您可以选择首先完成哪些扫描。在扫描开始期间和开始之前，支持对地图和分类扫描进行优先排序。如果您选择在扫描过程中确定扫描的优先级，则映射扫描和分类扫描都会被优先处理。

有关更多信息，请参阅"优先扫描" 。

支持加拿大个人身份信息 **(PII)** 数据类别

数据分类扫描识别加拿大 PII 数据类别。这些类别包括加拿大所有省份和地区的银行信息、护照号码、社会保险号码、驾驶执照号码和健康卡号码。

有关更多信息，请参阅"个人数据类别" 。

自定义分类（预览）

数据分类支持地图和分类扫描的自定义分类。通过自定义分类，您可以定制数据分类扫描，以使用正则表达式捕获特定于您的组织的数据。此功能目前处于预览状态。

有关更多信息，请参阅"添加自定义分类" 。

已保存的搜索标签

政策 选项卡已重命名"已保存的搜索"。功能没有改变。

将扫描事件发送到审核页面

数据分类支持发送分类事件（扫描启动时和扫描结束时）到"NetApp Console 审计页面"。

安全更新

- Keras 包已更新，缓解了漏洞（BDSA-2025-0107 和 BDSA-2025-1984）。
- Docker 容器配置已更新。容器不再有权访问主机的网络接口来制作原始网络数据包。通过减少不必要的访问，此更新可减轻潜在的安全风险。

性能增强

已经实施了代码增强，以减少 RAM 使用率并提高数据分类的整体性能。

错误修复

导致StorageGRID扫描失败、调查页面过滤选项无法加载以及无法下载大容量评估的数据发现评估的错误已得到修复。

2025年4月14日

1.42 版

此次BlueXP classification发布包括：

工作环境批量扫描

BlueXP classification支持工作环境的批量操作。您可以选择启用映射扫描、启用映射和分类扫描、禁用扫描或在工作环境中跨卷创建自定义配置。如果您对单个卷进行选择，它将覆盖批量选择。要执行批量操作，请导航到配置页面并进行选择。

本地下载调查报告

BlueXP classification支持将数据调查报告下载到本地以便在浏览器中查看。如果选择本地选项，数据调查仅以 CSV 格式提供，并且仅显示前 10,000 行数据。

有关更多信息，请参阅"使用BlueXP classification调查组织中存储的数据" 。

2025年3月10日

1.41 版

此BlueXP classification版本包括一般改进和错误修复。它还包括：

扫描状态

BlueXP classification跟踪卷上的初始映射和分类扫描的实时进度。单独的进度条跟踪映射和分类扫描，显示扫描文件总数的百分比。您还可以将鼠标悬停在进度条上来查看已扫描的文件数和文件总数。跟踪扫描状态可以更

深入地了解扫描进度，使您能够更好地规划扫描并了解资源分配。

要查看扫描状态，请导航到BlueXP classification中的配置，然后选择工作环境配置。每卷的进度均按行显示。

2025年2月19日

1.40 版

此BlueXP classification版本包括以下更新。

支持 RHEL 9.5

此版本除了支持以前支持的版本外，还提供对 Red Hat Enterprise Linux v9.5 的支持。这适用于BlueXP classification的任何手动本地安装，包括暗站部署。

以下操作系统需要使用 Podman 容器引擎，并且需要BlueXP classification版本 1.30 或更高版本：Red Hat Enterprise Linux 版本 8.8、8.10、9.0、9.1、9.2、9.3、9.4 和 9.5。

优先进行仅映射扫描

当进行仅映射扫描时，您可以优先考虑最重要的扫描。当您拥有多个工作环境并希望确保首先完成高优先级扫描时，此功能会有所帮助。

默认情况下，扫描按照启动的顺序排队。通过设置扫描优先级，您可以将扫描移至队列的最前面。可以对多个扫描进行优先排序。优先级按先进先出的顺序指定，这意味着您优先考虑的第一个扫描将移至队列的最前面；您优先考虑的第二个扫描将成为队列中的第二个扫描，依此类推。

优先权是一次性授予的。映射数据的自动重新扫描按照默认顺序进行。

优先级仅限于"仅映射扫描"；它不适用于地图和分类扫描。

有关更多信息，请参阅"优先扫描" 。

重试所有扫描

BlueXP classification支持批量重试所有失败扫描的功能。

您可以使用全部重试功能以批量操作的方式重新尝试扫描。如果分类扫描由于网络中断等临时问题而失败，您可以使用一个按钮同时重试所有扫描，而不必单独重试。可以根据需要重试扫描多次。

要重试所有扫描：

1. 从BlueXP classification菜单中，选择 配置。
2. 要重试所有失败的扫描，请选择*重试所有扫描*。

提高分类模型的准确性

机器学习模型的准确率"预定义类别"提高了11%。

2025年1月22日

1.39 版

此BlueXP classification版本更新了数据调查报告的导出流程。此导出更新对于对您的数据执行额外分析、对数

据创建额外可视化或与他人共享数据调查结果很有用。

以前，数据调查报告导出限制为 10,000 行。在此版本中，限制已被取消，以便您可以导出所有数据。此更改使您能够从数据调查报告中导出更多数据，从而为您的数据分析提供更大的灵活性。

您可以选择工作环境、卷、目标文件夹以及 JSON 或 CSV 格式。导出的文件名包含时间戳，以帮助您识别数据的导出时间。

支持的工作环境包括：

- Cloud Volumes ONTAP
- 适用于ONTAP的 FSx
- ONTAP
- 共享组

从数据调查报告中导出数据有以下限制：

- 每种类型（文件、目录和表）最多可下载 5 亿条记录
- 预计导出一百万条记录大约需要 35 分钟。

有关数据调查和报告的详细信息，请参阅 "调查组织中存储的数据"。

2024年12月16日

1.38 版

此BlueXP classification版本包括一般改进和错误修复。

2024年11月4日

1.37 版

此BlueXP classification版本包括以下更新。

支持 **RHEL 8.10**

此版本除了支持以前支持的版本外，还提供对 Red Hat Enterprise Linux v8.10 的支持。这适用于BlueXP classification的任何手动本地安装，包括暗站部署。

以下操作系统需要使用 Podman 容器引擎，并且需要BlueXP classification版本 1.30 或更高版本：Red Hat Enterprise Linux 版本 8.8、8.10、9.0、9.1、9.2、9.3 和 9.4。

详细了解 "BlueXP classification"。

支持 **NFS v4.1**

此版本除了支持以前支持的版本外，还提供对 NFS v4.1 的支持。

详细了解 "BlueXP classification"。

2024年10月10日

1.36 版

支持 **RHEL 9.4**

此版本除了支持以前支持的版本外，还提供对 Red Hat Enterprise Linux v9.4 的支持。这适用于BlueXP classification的任何手动本地安装，包括暗站部署。

以下操作系统需要使用 Podman 容器引擎，并且需要BlueXP classification版本 1.30 或更高版本：Red Hat Enterprise Linux 版本 8.8、9.0、9.1、9.2、9.3 和 9.4。

详细了解 "BlueXP classification部署概述"。

改进的扫描性能

此版本提供了改进的扫描性能。

2024年9月2日

1.35 版

扫描**StorageGRID**数据

BlueXP classification支持扫描StorageGRID中的数据。

有关详细信息，请参阅"扫描StorageGRID数据"。

2024年8月5日

1.34 版

此BlueXP classification版本包括以下更新。

从 **CentOS** 更改为 **Ubuntu**

BlueXP classification已将其针对 Microsoft Azure 和 Google Cloud Platform (GCP) 的 Linux 操作系统从 CentOS 7.9 更新为 Ubuntu 22.04。

有关部署详细信息，请参阅 "在具有互联网访问权限的Linux主机上安装并准备Linux主机系统"。

2024年7月1日

1.33 版

支持 **Ubuntu**

此版本支持 Ubuntu 24.04 Linux 平台。

地图扫描收集元数据

在映射扫描期间从文件中提取以下元数据，并将其显示在治理、合规性和调查仪表板上：

- 工作环境

- 工作环境类型
- 存储库
- 文件类型
- 已用容量
- 文件数
- 文件大小
- 文件创建
- 文件上次访问
- 文件上次修改时间
- 文件发现时间
- 权限提取

仪表板中的附加数据

此版本更新了映射扫描期间治理、合规性和调查仪表板中显示的数据。

有关详细信息，请参阅"映射和分类扫描之间有什么区别" 。

2024年6月5日

1.32 版

配置页面中的新映射状态列

此版本现在在配置页面中显示一个新的映射状态列。新列可帮助您识别映射是否正在运行、排队、暂停或更多。

有关状态的解释，请参阅 "更改扫描设置"。

2024年5月15日

1.31 版

分类是**BlueXP**中的一项核心服务

BlueXP classification现在作为BlueXP中的一项核心功能提供，每个连接器最多可免费扫描 500 TiB 的数据。无需分类许可或付费订阅。由于我们将BlueXP classification功能的重点放在新版本扫描NetApp存储系统上，因此某些旧功能将仅对之前已支付许可证费用的客户可用。当付费合同到期时，这些旧功能的使用将失效。

数据分类不会对其可以扫描的数据量施加限制。每个控制台代理支持扫描和显示 500 TiB 的数据。要扫描超过 500 TiB 的数据，"安装另一个控制台代理"然后"部署另一个数据分类实例"。 + 控制台 UI 显示来自单个连接器的数据。有关查看来自多个控制台代理的数据的提示，请参阅"使用多个控制台代理"。

2024年4月1日

1.30 版

增加了对 **RHEL v8.8** 和 **v9.3 BlueXP classification**的支持

此版本除了之前支持的 9.x 之外，还支持 Red Hat Enterprise Linux v8.8 和 v9.3，它需要 Podman，而不是 Docker 引擎。这适用于BlueXP classification的任何手动本地安装。

以下操作系统需要使用 Podman 容器引擎，并且需要BlueXP classification版本 1.30 或更高版本：Red Hat Enterprise Linux 版本 8.8、9.0、9.1、9.2 和 9.3。

详细了解 "BlueXP classification部署概述"。

如果您在本地的 RHEL 8 或 9 主机上安装连接器，则支持BlueXP classification。如果 RHEL 8 或 9 主机位于 AWS、Azure 或 Google Cloud 中，则不受支持。

删除了激活审计日志收集的选项

激活审计日志收集的选项已被禁用。

扫描速度提高

辅助扫描节点的扫描性能得到了改善。如果您需要额外的扫描处理能力，您可以添加更多扫描仪节点。有关详细信息，请参阅 "在可以访问互联网的主机上安装BlueXP classification"。

自动升级

如果您在具有互联网访问权限的系统上部署了BlueXP classification，则系统会自动升级。以前，升级发生在自上次用户活动以来经过特定时间之后。在此版本中，如果当地时间在凌晨 1:00 至凌晨 5:00 之间， BlueXP classification将自动升级。如果当地时间不在这些时间范围内，则升级将在用户上次活动后经过特定时间后进行。有关详细信息，请参阅 "在可以访问互联网的 Linux 主机上安装"。

如果您在没有互联网访问的情况下部署了BlueXP classification，则需要手动升级。有关详细信息，请参阅 "在没有互联网访问的 Linux 主机上安装BlueXP classification"。

2024年3月4日

1.29 版

现在您可以排除驻留在特定数据源目录中的扫描数据

如果您希望BlueXP classification排除驻留在特定数据源目录中的扫描数据，则可以将这些目录名称添加到BlueXP classification的配置文件中。此功能使您能够避免扫描不必要的目录，或避免返回错误的个人数据结果。

"了解更多" 。

超大型实例支持现已合格

如果您需要BlueXP classification来扫描超过 2.5 亿个文件，您可以在云部署或本地安装中使用超大实例。这种系统最多可以扫描 5 亿个文件。

"了解更多" 。

2024年1月10日

1.27 版

调查页面结果显示总大小以及项目总数

调查页面中的过滤结果除了显示文件总数外，还显示项目的总大小。这在移动文件、删除文件等操作时很有帮助。

将其他组 ID 配置为“向组织开放”

现在，如果组最初没有设置该权限，您可以直接从BlueXP classification将 NFS 中的组 ID 配置为“向组织开放”。任何附加了这些组 ID 的文件和文件夹都将在调查详情页面中显示为“向组织开放”。了解如何"添加其他组 ID 作为“对组织开放”"。

2023年12月14日

版本 1.26.6

此版本包含一些小的改进。

该版本还删除了以下选项：

- 激活审计日志收集的选项已被禁用。
- 在目录调查期间，无法使用目录计算个人身份信息 (PII) 数据数量的选项。请参阅"调查组织中存储的数据"。
- 使用 Azure 信息保护 (AIP) 标签集成数据的选项已被禁用。

2023年11月6日

版本 1.26.3

此版本已修复以下问题

- 修复了仪表板中显示系统扫描的文件数量不一致的问题。
- 通过处理和报告名称和元数据中带有特殊字符的文件和目录来改进扫描行为。

2023年10月4日

1.26 版

支持在 RHEL 版本 9 上本地安装BlueXP classification

Red Hat Enterprise Linux 8 和 9 版本不支持 Docker 引擎；而BlueXP classification安装需要该引擎。我们现在支持在 RHEL 9.0、9.1 和 9.2 上使用 Podman 版本 4 或更高版本作为容器基础设施进行BlueXP classification安装。如果您的环境需要使用最新版本的 RHEL，现在您可以在使用 Podman 时安装BlueXP classification（版本 1.26 或更高版本）。

目前，在使用 RHEL 9.x 时，我们不支持暗站安装或分布式扫描环境（使用主节点和远程扫描器节点）。

2023年9月5日

1.25 版

中小型部署暂时不可用

当您在 AWS 中部署BlueXP classification实例时，此时无法选择 部署 > 配置 并选择小型或中型实例。您仍然可以通过选择*部署>部署*来使用大实例大小部署实例。

在调查结果页面中为最多 **100,000** 个项目添加标签

过去，您一次只能在调查结果页面中将标签应用于单个页面（20 个项目）。现在您可以在调查结果页面中选择*所有*项目并将标签应用于所有项目 - 一次最多 100,000 个项目。

识别最小文件大小为 **1 MB** 的重复文件

BlueXP classification仅用于在文件大小为 50 MB 或更大时识别重复文件。现在可以识别以 1 MB 开头的重复文件。您可以使用调查页面过滤器“文件大小”和“重复项”来查看您的环境中哪些特定大小的文件是重复的。

2023年7月17日

1.24 版

BlueXP classification识别出两种新的德国个人数据

BlueXP classification可以识别和分类包含以下类型数据的文件：

- 德国身份证 (Personalausweisnummer)
- 德国社会安全号码 (Sozialversicherungsnummer)

"查看BlueXP classification可以在您的数据中识别的所有类型的个人数据" 。

BlueXP classification在限制模式和私人模式下完全受支持

BlueXP classification现在完全支持没有互联网访问（私人模式）和有限的出站互联网访问（受限模式）的站点。"了解有关连接器的BlueXP部署模式的更多信息" 。

升级**BlueXP classification**的私人模式安装时可以跳过版本

现在，即使 BlueXP 分类不是连续的，您也可以升级到较新版本的BlueXP classification。这意味着不再需要当前一次升级BlueXP classification的一个版本的限制。此功能从 1.24 版本开始适用。

BlueXP classificationAPI 现已可用

BlueXP classificationAPI 使您能够执行操作、创建查询以及导出有关您正在扫描的数据的信息。交互式文档可通过 Swagger 获取。该文档分为多个类别，包括调查、合规、治理和配置。每个类别都是对BlueXP classificationUI 中的选项卡的引用。

"了解有关BlueXP classificationAPI 的更多信息" 。

2023年6月6日

1.23 版

搜索数据主体名称时现在支持日语

现在，在响应数据主体访问请求 (DSAR) 时搜索主体名称时可以输入日语名称。您可以生成"数据主体访问请求报告"以及由此产生的信息。您还可以在"数据调查页面中的“数据主体”过滤器"识别包含主题名称的文件。

Ubuntu 现在是受支持的 **Linux** 发行版，您可以在其上安装**BlueXP classification**

Ubuntu 22.04 已被认定为BlueXP classification的支持操作系统。您可以在网络中的 Ubuntu Linux 主机上安装BlueXP classification，或者使用安装程序 1.23 版本在云中的 Linux 主机上安装。 "查看如何在安装了 Ubuntu 的主机上安装BlueXP classification" 。

新的**BlueXP classification**安装不再支持 **Red Hat Enterprise Linux 8.6** 和 **8.7**

这些版本不支持新的部署，因为 Red Hat 不再支持 Docker，而 Docker 是先决条件。如果您有在 RHEL 8.6 或 8.7 上运行的现有BlueXP classification机器， NetApp将继续支持您的配置。

BlueXP classification可以配置为 **FPolicy** 收集器，以从**ONTAP**系统接收 **FPolicy** 事件

您可以启用文件访问审计日志功能，在BlueXP classification系统上收集在工作环境中的卷上检测到的文件访问事件。 BlueXP classification可以捕获以下类型的 FPolicy 事件以及对您的文件执行操作的用户：创建、读取、写入、删除、重命名、更改所有者/权限以及更改 SACL/DACL。

暗网现已支持 **Data Sense BYOL** 许可证

现在，您可以将 Data Sense BYOL 许可证上传到暗站中的BlueXP digital wallet中，以便在许可证不足时收到通知。

2023年4月3日

1.22 版

新数据发现评估报告

数据发现评估报告对扫描环境进行了高级分析，以突出显示系统的发现并显示关注区域和潜在的补救步骤。本报告的目标是提高人们对数据治理问题、数据安全漏洞以及数据集的数据合规性差距的认识。 "了解如何生成和使用数据发现评估报告" 。

能够在云中的较小实例上部署**BlueXP classification**

在 AWS 环境中从BlueXP连接器部署BlueXP classification时，您现在可以从两个比默认实例更小的实例类型中进行选择。如果您正在扫描小型环境，这可以帮助您节省云成本。但是，使用较小的实例时存在一些限制。 "查看可用的实例类型和限制" 。

现在可以使用独立脚本在**BlueXP classification**安装之前验证您的 **Linux** 系统

如果您想独立于运行BlueXP classification安装来验证您的 Linux 系统是否满足所有先决条件，您可以下载一个单独的脚本，该脚本仅测试先决条件。 "了解如何检查您的 Linux 主机是否已准备好安装BlueXP classification" 。

2023年3月7日

1.21 版

从**BlueXP classificationUI** 添加您自己的自定义类别的新功能

BlueXP classification现在允许您添加自己的自定义类别，以便BlueXP classification能够识别适合这些类别的文件。 BlueXP classification有很多 "预定义类别"，因此此功能使您能够添加自定义类别，以识别在数据中找到组织独有的信息的位置。

现在您可以从**BlueXP classificationUI** 添加自定义关键字

BlueXP classification已经能够添加自定义关键字， BlueXP classification将在未来的扫描中识别这些关键字。但是，您需要登录BlueXP classificationLinux 主机并使用命令行界面添加关键字。在此版本中，添加自定义关键字的功能位于BlueXP classificationUI 中，这使得添加和编辑这些关键字变得非常容易。

当“上次访问时间”发生变化时， **BlueXP classification**不会扫描文件

默认情况下，如果BlueXP classification没有足够的“写入”权限，系统将不会扫描卷中的文件，因为BlueXP classification无法将“上次访问时间”恢复为原始时间戳。但是，如果您不介意将上次访问时间重置为文件中的原始时间，则可以在配置页面中覆盖此行为，以便BlueXP classification可以扫描卷，而不管权限如何。

与此功能结合，添加了名为“扫描分析事件”的新过滤器，以便您可以查看未分类的文件，因为BlueXP classification无法恢复上次访问时间，或者即使BlueXP classification无法恢复上次访问时间也已分类的文件。

"详细了解“上次访问时间戳”以及BlueXP classification所需的权限" 。

BlueXP classification识别三种新的个人数据类型

BlueXP classification可以识别和分类包含以下类型数据的文件：

- 博茨瓦纳身份证（奥芒）号码
- 博茨瓦纳护照号码
- 新加坡国民登记身份证（NRIC）

"查看BlueXP classification可以在您的数据中识别的所有类型的个人数据" 。

更新了目录的功能

- 数据调查报告的“精简版 CSV 报告”选项现在包含来自目录的信息。
- “上次访问”时间过滤器现在显示文件和目录的上次访问时间。

安装增强功能

- 对于没有互联网访问的网站（暗站）， BlueXP classification安装程序现在会执行预检查，以确保您的系统和网络要求满足成功安装的要求。
- 安装审计日志文件现在已保存；它们被写入 `/ops/netapp/install_logs`。

2023年2月5日

1.20 版

能够向任何电子邮件地址发送基于策略的通知电子邮件

在BlueXP classification的早期版本中，当某些关键策略返回结果时，您可以向您帐户中的BlueXP用户发送电子邮件警报。此功能使您能够在不在线时收到通知以保护您的数据。现在，您还可以从策略向不在您的BlueXP帐户中的任何其他用户（最多 20 个电子邮件地址）发送电子邮件警报。

"详细了解如何根据策略结果发送电子邮件提醒" 。

现在您可以从**BlueXP classificationUI** 添加个人模式

BlueXP classification已经能够添加自定义“个人数据”， BlueXP classification将在未来的扫描中识别这些数据。但是，您需要登录BlueXP classificationLinux 主机并使用命令行添加自定义模式。在此版本中，使用正则表达式添加个人模式的功能位于BlueXP classificationUI 中，从而可以非常轻松地添加和编辑这些自定义模式。

使用**BlueXP classification**可以移动 **1500** 万个文件

过去，您可以通过BlueXP classification将最多 100,000 个源文件移动到任何 NFS 共享。现在您一次最多可以移动 1500 万个文件。

能够查看有权访问 **SharePoint Online** 文件的用户数量

过滤器“具有访问权限的用户数量”现在支持存储在 SharePoint Online 存储库中的文件。过去仅支持 CIFS 共享上的文件。请注意，此时不基于活动目录的 SharePoint 组将不会计入此过滤器。

操作状态面板中添加了新的“部分成功”状态

新的“部分成功”状态表示BlueXP classification操作已完成，一些项目失败，一些项目成功，例如，当您移动或删除 100 个文件时。此外，“完成”状态已重命名为“成功”。过去，“完成”状态可能会列出成功和失败的操作。现在“成功”状态意味着所有项目上的所有操作都成功。 "了解如何查看操作状态面板" 。

2023年1月9日

1.19 版

能够查看包含敏感数据和过于宽松的文件图表

治理仪表板添加了一个新的“敏感数据和广泛权限”区域，该区域提供了包含敏感数据（包括敏感数据和敏感个人数据）且过于宽松的文件的热图。这可以帮助您了解敏感数据可能存在的风险。 "了解更多" 。

数据调查页面新增三个过滤器

新的过滤器可用于优化数据调查页面中显示的结果：

- “具有访问权限的用户数”过滤器显示哪些文件和文件夹对一定数量的用户开放。您可以选择一个数字范围来优化结果 - 例如，查看 51-100 个用户可以访问哪些文件。
- 现在，“创建时间”、“发现时间”、“上次修改时间”和“上次访问时间”过滤器允许您创建自定义日期范围，而不仅仅是选择预定义的日期范围。例如，您可以查找“创建时间”超过 6 个月的文件，或“上次修改时间”在“最近 10 天”内的文件。
- 现在，“文件路径”过滤器使您能够指定要从过滤查询结果中排除的路径。如果您输入包含和排除某些数据的路径， BlueXP classification会首先在包含的路径中找到所有文件，然后从排除的路径中删除文件，然后显示结果。

"查看可用于调查数据的所有过滤器的列表" 。

BlueXP classification可以识别日本个人编号

BlueXP classification可以识别和分类包含日本个人编号（也称为 My Number）的文件。这包括个人和企业我的号码。 "查看BlueXP classification可以在您的数据中识别的所有类型的个人数据" 。

NetApp Data Classification的已知限制

已知限制标识了此版本中不受支持或无法正确互操作的功能。仔细审查这些限制。

NetApp Data Classification禁用选项

2023 年 12 月（版本 1.26.6）版本删除了以下选项：

- 激活审计日志收集的选项已被禁用。
- 在目录调查期间，无法使用目录计算个人身份信息 (PII) 数据数量的选项。
- 使用 Azure 信息保护 (AIP) 标签集成数据的选项已被禁用。

数据分类扫描

数据分类扫描存在以下限制。

数据分类仅扫描卷下的一个共享

如果单个卷下有多个文件共享，数据分类将扫描具有最高层次的共享。例如，如果您有如下共享：

- /一个
- /A/B
- /C
- /D/E

在此配置中，仅扫描 /A 中的数据。 /C 和 /D 中的数据未被扫描。

临时解决策

有一种解决方法可以确保您扫描卷中所有共享的数据。按照下面的步骤进行操作：

1. 在系统中，添加要扫描的卷。
2. 数据分类完成扫描卷后，转到“数据调查”页面并创建过滤器以查看正在扫描的共享：

   通过“系统名称”和“目录类型 = 共享”过滤数据，以查看正在扫描哪个共享。

3. 获取卷中存在的共享的完整列表，以便您可以看到哪些共享未被扫描。
4. "将剩余的共享添加到共享组" 。

   单独添加所有共享，例如：

   ```
   /C
   /D
   ```

5. 对系统中具有多个共享的每个卷执行这些步骤。

上次访问的时间戳

当数据分类对目录进行扫描时，扫描会影响目录的上次访问字段。当您查看上次访问字段时，该元数据反映扫描的日期和时间或用户上次访问目录的时间。

开始使用

了解NetApp Data Classification

NetApp Data Classification是NetApp Console的一项数据治理服务，它可以扫描您的企业内部和云数据源以映射和分类数据并识别私人信息。这可以帮助降低您的安全和合规风险，降低存储成本，并协助您的数据迁移项目。

从 1.31 版开始，数据分类作为NetApp Console中的一项核心功能提供。无需额外付费。无需分类许可或订阅。 + 如果您一直在使用旧版本 1.30 或更早版本，则该版本在您的订阅到期之前可用。

NetApp Console

可以通过NetApp Console访问数据分类。

NetApp Console提供企业级跨本地和云环境的NetApp存储和数据服务的集中管理。需要控制台才能访问和使用NetApp数据服务。作为管理界面，它使您能够从一个界面管理许多存储资源。控制台管理员可以控制企业内所有系统的存储和服务的访问。

您不需要许可证或订阅即可开始使用NetApp Console，并且只有当您需要在云中部署控制台代理以确保与存储系统或NetApp数据服务的连接时才需要付费。但是，一些可从控制台访问的NetApp数据服务是需要许可或基于订阅的。

详细了解"NetApp Console"。

功能

数据分类使用人工智能 (AI)、自然语言处理 (NLP) 和机器学习 (ML) 来理解其扫描的内容，以便提取实体并对内容进行相应的分类。这使得数据分类能够提供以下功能领域。

"了解数据分类的用例" 。

保持合规

数据分类提供了多种工具，可以帮助您实现合规性。您可以使用数据分类来：

- 识别个人身份信息 (PII)。
- 根据 GDPR、CCPA、PCI 和 HIPAA 隐私法规的要求识别广泛的敏感个人信息。
- 根据姓名或电子邮件地址响应数据主体访问请求 (DSAR)。

加强安全

数据分类可以识别可能被犯罪分子访问的数据。您可以使用数据分类来：

- 识别向整个组织或公众公开的所有具有开放权限的文件和目录（共享和文件夹）。
- 识别位于初始专用位置之外的敏感数据。
- 遵守数据保留政策。

- 使用*Policies*自动检测新的安全问题，以便安全人员可以立即采取行动。

优化存储使用情况

数据分类提供可帮助您降低存储总拥有成本 (TCO) 的工具。您可以使用数据分类来：

- 通过识别重复或与业务无关的数据来提高存储效率。
- 通过识别可以分层到较便宜的对象存储的非活动数据来节省存储成本。 "了解有关Cloud Volumes ONTAP系统分层的更多信息" 。 "了解有关本地ONTAP系统分层的更多信息" 。

支持的系统和数据源

数据分类可以扫描和分析来自以下类型的系统和数据源的结构化和非结构化数据：

系统

- Amazon FSx for NetApp ONTAP管理
- Azure NetApp Files
- Cloud Volumes ONTAP （部署在 AWS、Azure 或 GCP 中）
- 本地ONTAP集群
- StorageGRID
- Google Cloud NetApp Volumes

数据来源

- NetApp文件共享
- 数据库：
  - 亚马逊关系数据库服务 (Amazon RDS)
  - MongoDB
  - MySQL
  - Oracle
  - PostgreSQL
  - SAP HANA
  - SQL 服务器 (MSSQL)

数据分类支持 NFS 版本 3.x、4.0 和 4.1，以及 CIFS 版本 1.x、2.0、2.1 和 3.0。

成本

数据分类可以免费使用。无需分类许可或付费订阅。

基础设施成本

- 在云中安装数据分类需要部署云实例，这会导致部署云的云提供商收取费用。看为每个云提供商部署的实例类型 。如果您在本地系统上安装数据分类，则无需付费。

- 数据分类要求您部署控制台代理。在许多情况下，由于您在控制台中使用其他存储和服务，因此您已经拥有控制台代理。控制台代理实例会导致其部署所在的云提供商收取费用。查看 "为每个云提供商部署的实例类型"。如果您在本地系统上安装控制台代理，则无需付费。

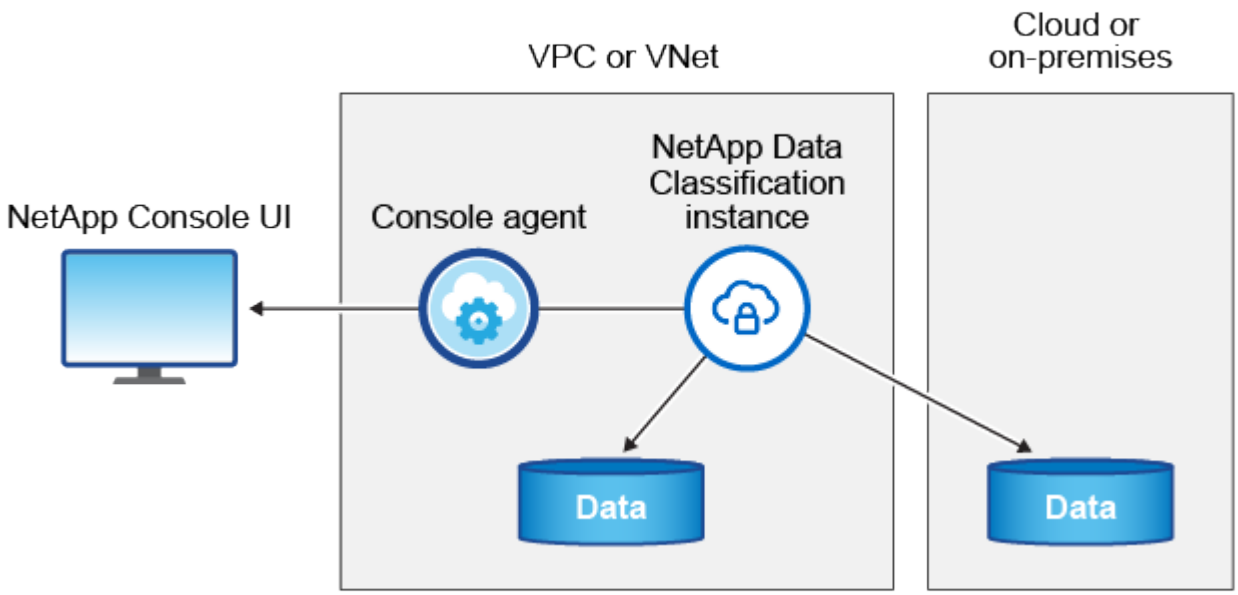
数据传输成本

数据传输成本取决于您的设置。如果数据分类实例和数据源位于同一可用区和区域，则没有数据传输成本。但是，如果数据源（例如Cloud Volumes ONTAP系统）位于不同的可用区或区域，那么您的云提供商将向您收取数据传输费用。请参阅以下链接以了解更多详细信息：

- "AWS：Amazon Elastic Compute Cloud (Amazon EC2) 定价"
- "Microsoft Azure：带宽定价详情"
- "Google Cloud：存储传输服务定价"

数据分类实例

当您在云中部署数据分类时，控制台会将实例部署在与控制台代理相同的子网中。 "了解有关控制台代理的更多信息。"



请注意有关默认实例的以下几点：

- 在 AWS 中，数据分类在 "m6i.4xlarge 实例"带有 500 GiB GP2 磁盘。操作系统映像是 Amazon Linux 2。在 AWS 中部署时，如果您要扫描少量数据，则可以选择较小的实例大小。
- 在 Azure 中，数据分类在"Standard_D16s_v3 VM"带有 500 GiB 磁盘。操作系统映像是 Ubuntu 22.04。
- 在 GCP 中，数据分类在"n2-standard-16 虚拟机"配备 500 GiB 标准持久磁盘。操作系统映像是 Ubuntu 22.04。
- 在默认实例不可用的区域中，数据分类在备用实例上运行。"查看替代实例类型" 。
- 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*
- 每个控制台代理仅部署一个数据分类实例。

您还可以在您的场所内的 Linux 主机上或您首选的云提供商的主机上部署数据分类。无论您选择哪种安装方法，软件的功能都完全相同。只要实例可以访问互联网，数据分类软件的升级就会自动进行。

实例应始终保持运行，因为数据分类会持续扫描数据。

在不同的实例类型上部署

查看实例类型的以下规范：

系统大小	规格	限制
特大号	32 个 CPU、128 GB RAM、1 TiB SSD	最多可扫描 5 亿个文件。
大（默认）	16 个 CPU、64 GB RAM、500 GiB SSD	最多可扫描 2.5 亿个文件。

在 Azure 或 GCP 中部署数据分类时，如果您想使用较小的实例类型，请发送电子邮件至 ng-contact-data-sense@netapp.com 寻求帮助。

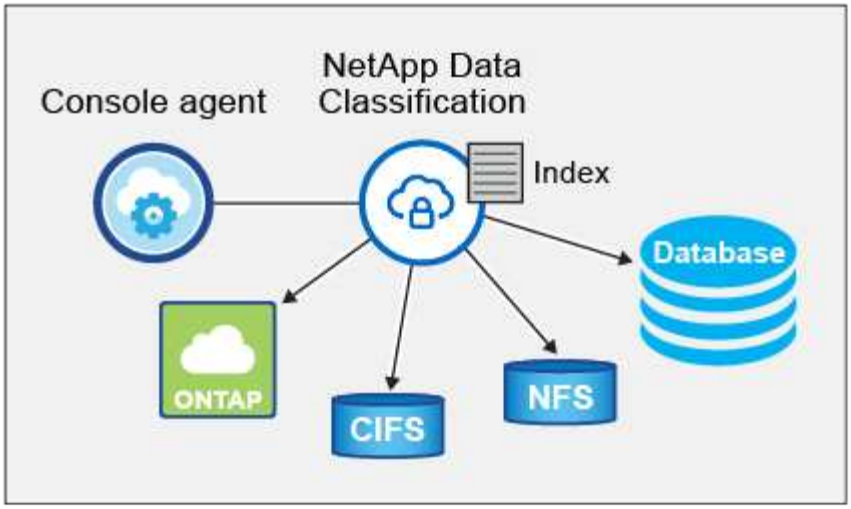
数据分类扫描的工作原理

从高层次来看，数据分类扫描的工作原理如下：

1. 您在控制台中部署数据分类实例。
2. 您可以在一个或多个数据源上启用高级映射（称为“仅映射”扫描）或深层扫描（称为“映射和分类”扫描）。
3. 数据分类使用人工智能学习过程扫描数据。
4. 您可以使用提供的仪表板和报告工具来帮助您实现合规性和治理工作。

启用数据分类并选择要扫描的存储库（这些是卷、数据库模式或其他用户数据）后，它会立即开始扫描数据以识别个人和敏感数据。在大多数情况下，您应该专注于扫描实时生产数据，而不是备份、镜像或 DR 站点。然后，数据分类映射您的组织数据，对每个文件进行分类，并识别和提取数据中的实体和预定义模式。扫描结果是个人信息、敏感个人信息、数据类别和文件类型的索引。

数据分类通过安装 NFS 和 CIFS 卷像任何其他客户端一样连接到数据。 NFS 卷自动以只读方式访问，而您需要提供 Active Directory 凭据来扫描 CIFS 卷。



初始扫描后，数据分类将以循环方式持续扫描您的数据以检测增量变化。这就是为什么保持实例运行很重要。

您可以在卷级别或数据库模式级别启用和禁用扫描。

数据分类不会对其可以扫描的数据量施加限制。每个控制台代理支持扫描和显示 500 TiB 的数据。要扫描超过 500 TiB 的数据，"安装另一个控制台代理"然后"部署另一个数据分类实例"。 + 控制台 UI 显示来自单个连接器的数据。有关查看来自多个控制台代理的数据的提示，请参阅"使用多个控制台代理"。

映射扫描和分类扫描之间有什么区别

您可以在数据分类中进行两种类型的扫描：

- 仅映射扫描仅提供数据的高级概览，并在选定的数据源上执行。仅映射扫描比映射和分类扫描花费的时间更少，因为它们不访问文件来查看其中的数据。您可能希望首先执行此操作来确定研究领域，然后对这些领域执行地图和分类扫描。
- 地图和分类扫描 为您的数据提供深层扫描。

有关映射扫描和分类扫描之间的差异的详细信息，请参阅"映射和分类扫描之间有什么区别？"。

数据分类所分类的信息

数据分类收集、索引并分配以下数据的类别：

- 关于文件的*标准元数据*：文件类型、大小、创建和修改日期等等。
- 个人数据：个人身份信息 (PII)，例如电子邮件地址、身份证号码或信用卡号，数据分类使用文件中的特定单词、字符串和模式进行识别。"了解有关个人数据的更多信息"。
- 敏感个人信息：《通用数据保护条例》（GDPR）和其他隐私法规定义的特殊类型的敏感个人信息（SPII），例如健康数据、种族血统或政治观点。"了解有关敏感个人数据的更多信息"。
- 类别：数据分类将扫描的数据分为不同类型的类别。类别是基于 AI 对每个文件的内容和元数据的分析的主题。"了解有关类别的更多信息"。
- 名称实体识别：数据分类使用人工智能从文档中提取人们的自然姓名。"了解如何响应数据主体访问请求"。

网络概述

数据分类可以在您选择的任何地方部署单个服务器或集群：在云端或本地。服务器通过标准协议连接到数据源，并在 Elasticsearch 集群中对结果进行索引，该集群也部署在同一服务器上。这使得能够支持多云、跨云、私有云和本地环境。

控制台使用安全组部署数据分类实例，该安全组启用来自控制台代理的入站 HTTP 连接。

当您在 SaaS 模式下使用控制台时，与控制台的连接通过 HTTPS 提供，并且您的浏览器和数据分类实例之间发送的私人数据使用 TLS 1.2 进行端到端加密保护，这意味着NetApp和第三方无法读取它。

出站规则完全开放。需要互联网访问来安装和升级数据分类软件以及发送使用情况指标。

如果您有严格的网络要求，"了解数据分类联系的端点" 。

访问NetApp Data Classification

您可以通过NetApp Console访问NetApp Data Classification。

要登录控制台，您可以使用您的NetApp支持站点凭据，也可以使用您的电子邮件和密码注册NetApp Console登录。"了解有关登录控制台的更多信息" 。

特定任务需要特定的控制台用户角色。"了解所有服务的控制台访问角色" 。

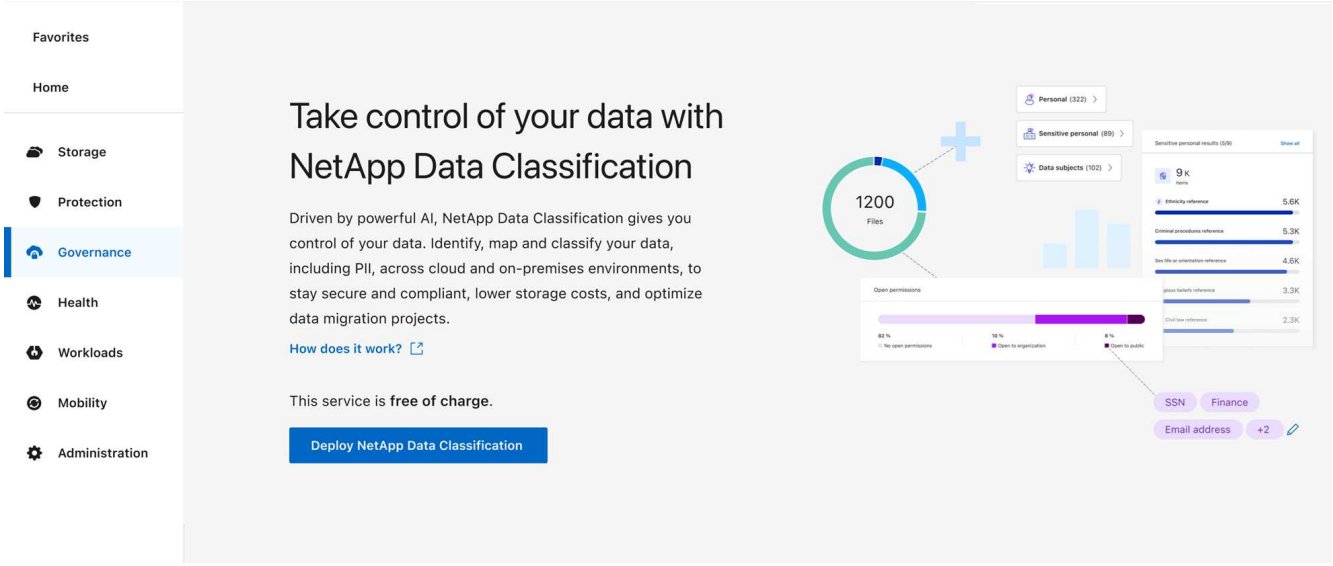
开始之前

- "您应该添加一个控制台代理。"
- "了解哪种数据分类部署样式适合您的工作负载。"

步骤

1. 在 Web 浏览器中，导航至"控制台"。
2. 登录控制台。
3. 从NetApp Console主页中，选择“治理”>“数据分类”。
4. 如果这是您第一次访问数据分类，则会出现登录页面。

   选择*在本地或云中部署分类*以开始部署您的分类实例。有关详细信息，请参阅"您应该使用哪种数据分类部署？"



否则，将出现数据分类仪表板。

部署数据分类

您应该使用哪种**NetApp Data Classification**部署?

您可以通过不同的方式部署NetApp Data Classification。了解哪种方法可以满足您的需求。

数据分类可以通过以下方式部署：

- "使用控制台在云中部署" 。控制台将数据分类实例部署在与控制台代理相同的云提供商网络中。
- "在可以访问互联网的 Linux 主机上安装" 。在您的网络中的 Linux 主机上或云中的 Linux 主机上安装数据分类，前提是该主机可以访问互联网。如果您希望使用同样位于本地的数据分类实例来扫描本地ONTAP系统，则这种类型的安装可能是一个不错的选择，尽管这不是必需的。
- "在没有互联网访问的本地站点的 Linux 主机上安装"，也称为_私人模式_。这种安装类型使用安装脚本，与控制台 SaaS 层没有连接。

> BlueXP私有模式（传统BlueXP接口）通常用于没有互联网连接的本地环境和安全云区域，其中包括 AWS Secret Cloud、AWS Top Secret Cloud 和 Azure IL6。NetApp继续通过传统的BlueXP界面支持这些环境。有关旧版BlueXP界面中的私有模式文档，请参阅"BlueXP私人模式的 PDF 文档"。

无论是在有互联网访问的 Linux 主机上安装，还是在没有互联网访问的 Linux 主机上进行本地安装，都使用安装脚本。脚本首先检查系统和环境是否满足先决条件。如果满足先决条件，则安装开始。如果您想独立于运行数据分类安装来验证先决条件，您可以下载一个单独的软件包，该软件包仅测试先决条件。

请参阅"检查您的 Linux 主机是否已准备好安装数据分类" 。

使用**NetApp Console**在云中部署**NetApp Data Classification**

您可以使用NetApp Console在云中部署NetApp Data Classification。控制台将数据分类实

例部署在与控制台代理相同的云提供商网络中。

请注意，您还可以"在可以访问互联网的 Linux 主机上安装数据分类"。如果您希望使用同样位于本地的数据分类实例来扫描本地ONTAP系统，则这种类型的安装可能是一个不错的选择 - 但这不是必需的。无论您选择哪种安装方法，软件的功能都完全相同。

快速启动

按照以下步骤快速开始，或者向下滚动到其余部分以获取完整详细信息。


创建控制台代理

如果您还没有控制台代理，请创建一个。看"在 AWS 中创建控制台代理"，"在 Azure 中创建控制台代理" ， 或者"在 GCP 中创建控制台代理"。

你也可以"在本地安装控制台代理"在您网络中的 Linux 主机上，或者在云端的 Linux 主机上。


先决条件

确保您的环境能够满足先决条件。这包括实例的出站互联网访问、Console 代理与 Data Classification 在端口 443 上的连接等。查看完整列表。


部署数据分类

启动安装向导以在云中部署数据分类实例。

创建控制台代理

如果您还没有控制台代理，请在您的云提供商中创建一个控制台代理。看 "在 AWS 中创建控制台代理"或者 "在 Azure 中创建控制台代理"， 或者 "在 GCP 中创建控制台代理"。大多数情况下，您可能需要在尝试激活数据分类之前设置好控制台代理，因为大多数 "控制台功能需要控制台代理"但有些情况下，你需要现在就设置一个。

在某些情况下，您必须使用部署在特定云提供商中的控制台代理：

- 在 AWS 中的Cloud Volumes ONTAP或Amazon FSx for ONTAP存储桶中扫描数据时，您可以使用 AWS 中的控制台代理。
- 在 Azure 中的Cloud Volumes ONTAP或Azure NetApp Files中扫描数据时，您可以使用 Azure 中的控制台代理。
  - 对于Azure NetApp Files，它必须部署在与您要扫描的卷相同的区域中。
- 在 GCP 中的Cloud Volumes ONTAP中扫描数据时，您可以使用 GCP 中的控制台代理。

使用任何这些云控制台代理时，都可以扫描本地ONTAP系统、 NetApp文件共享和数据库。

请注意，您也可以 "在本地安装控制台代理"在网络或云端的 Linux 主机上。一些计划在本地安装数据分类的用户可能还会选择在本地安装控制台代理。

可能有些情况下你需要使用 "多个控制台代理"。

数据分类不会对其可以扫描的数据量施加限制。每个控制台代理支持扫描和显示 500 TiB 的数据。要扫描超过 500 TiB 的数据，"安装另一个控制台代理"然后"部署另一个数据分类实例"。 + 控制台 UI 显示来自单个连接器的数据。有关查看来自多个控制台代理的数据的提示，请参阅"使用多个控制台代理"。

政府区域支持

当控制台代理部署在政府区域（AWS GovCloud、Azure Gov 或 Azure DoD）时，支持数据分类。以这种方式部署时，数据分类具有以下限制：

"了解如何在政府区域部署控制台代理"。

前提条件

在云中部署数据分类之前，请查看以下先决条件，以确保您具有受支持的配置。当您在云中部署数据分类时，它与控制台代理位于同一子网中。

启用数据分类的出站互联网访问

数据分类需要出站互联网访问。如果您的虚拟或物理网络使用代理服务器进行互联网访问，请确保数据分类实例具有出站互联网访问权限以联系以下端点。代理必须是非透明的。目前不支持透明代理。

根据您是在 AWS、Azure 还是 GCP 中部署数据分类，查看下面的相应表格。

AWS 所需的终端节点

端点	目的
\ https://api.console.netapp.com	与控制台服务（包括NetApp帐户）的通信。
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	与控制台网站通信，实现集中用户身份验证。
\ https://cloud-compliance-support-netapp.s3.us-west-2.amazonaws.com \ https://hub.docker.com \ https://auth.docker.io \ https://registry-1.docker.io \ https://index.docker.io/ \ https://dseasb33srnrn.cloudfront.net/ \ https://production.cloudflare.docker.com/	提供对软件映像、清单和模板的访问。
\ https://kinesis.us-east-1.amazonaws.com	使NetApp能够从审计记录中流式传输数据。
\ https://cognito-idp.us-east-1.amazonaws.com \ https://cognito-identity.us-east-1.amazonaws.com \ https://user-feedback-store-prod.s3.us-west-2.amazonaws.com \ https://customer-data-production.s3.us-west-2.amazonaws.com	启用数据分类来访问和下载清单和模板，以及发送日志和指标。

Azure 所需的终结点

端点	目的
\ https://api.console.netapp.com	与控制台服务（包括NetApp帐户）的通信。
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	与控制台网站通信，实现集中用户身份验证。
\ https://support.compliance.api.console.netapp.com/ \ https://hub.docker.com \ https://auth.docker.io \ https://registry-1.docker.io \ https://index.docker.io/ \ https://dseasb33srnrn.cloudfront.net/ \ https://production.cloudflare.docker.com/	提供对软件映像、清单、模板的访问以及发送日志和指标。
\ https://support.compliance.api.console.netapp.com/	使NetApp能够从审计记录中流式传输数据。

GCP 所需的端点

端点	目的
\ https://api.console.netapp.com	与控制台服务（包括NetApp帐户）的通信。
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	与控制台网站通信，实现集中用户身份验证。

端点	目的
\ https://support.compliance.api.console.netapp.com/ \ https://hub.docker.com \ https://auth.docker.io \ https://registry-1.docker.io \ https://index.docker.io/ \ https://dseasb33srnrn.cloudfront.net/ \ https://production.cloudflare.docker.com/	提供对软件映像、清单、模板的访问以及发送日志和指标。
\ https://support.compliance.api.console.netapp.com/	使NetApp能够从审计记录中流式传输数据。

确保数据分类具有所需的权限

确保数据分类具有部署资源和为数据分类实例创建安全组的权限。

- "Google Cloud 权限"
- "AWS 权限"
- "Azure 权限"

确保控制台代理可以访问数据分类

确保控制台代理和数据分类实例之间的连接。控制台代理的安全组必须允许通过端口 443 进出数据分类实例的入站和出站流量。此连接支持部署数据分类实例，并允许您查看“合规性和治理”选项卡中的信息。 AWS 和 Azure 的政府区域支持数据分类。

AWS 和 AWS GovCloud 部署需要额外的入站和出站安全组规则。看 "AWS 中的控制台代理规则"了解详情。

Azure 和 Azure 政府部署需要额外的入站和出站安全组规则。看 "Azure 中的控制台代理规则"了解详情。

确保数据分类能够持续运行

数据分类实例需要保持开启状态以持续扫描您的数据。

确保 **Web** 浏览器连接到数据分类

启用数据分类后，确保用户从与数据分类实例有连接的主机访问控制台界面。

数据分类实例使用私有 IP 地址来确保索引数据无法被互联网访问。因此，您用来访问控制台的 Web 浏览器必须连接到该私有 IP 地址。该连接可以来自与云提供商的直接连接（例如 VPN），也可以来自与数据分类实例位于同一网络内的主机。

检查您的 **vCPU** 限制

确保您的云提供商的 vCPU 限制允许部署具有必要数量的核心的实例。您需要验证控制台运行区域中相关实例系列的 vCPU 限制。"查看所需的实例类型" 。

有关 vCPU 限制的更多详细信息，请参阅以下链接：

- "AWS 文档：Amazon EC2 服务配额"
- "Azure 文档：虚拟机 vCPU 配额"
- "Google Cloud 文档：资源配额"

在云中部署数据分类

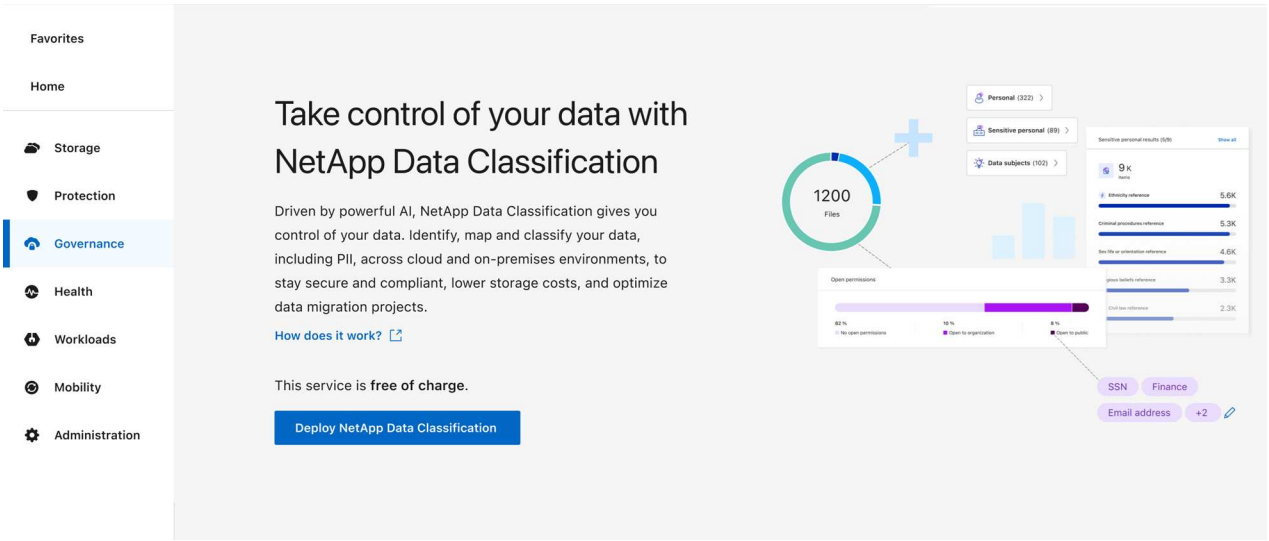
按照以下步骤在云中部署数据分类实例。控制台代理将在云中部署实例，然后在该实例上安装数据分类软件。

在默认实例类型不可用的区域中，数据分类在"备用实例类型"。

在 **AWS** 中部署

步骤

1. 从数据分类主页中，选择*在本地或云中部署分类*。

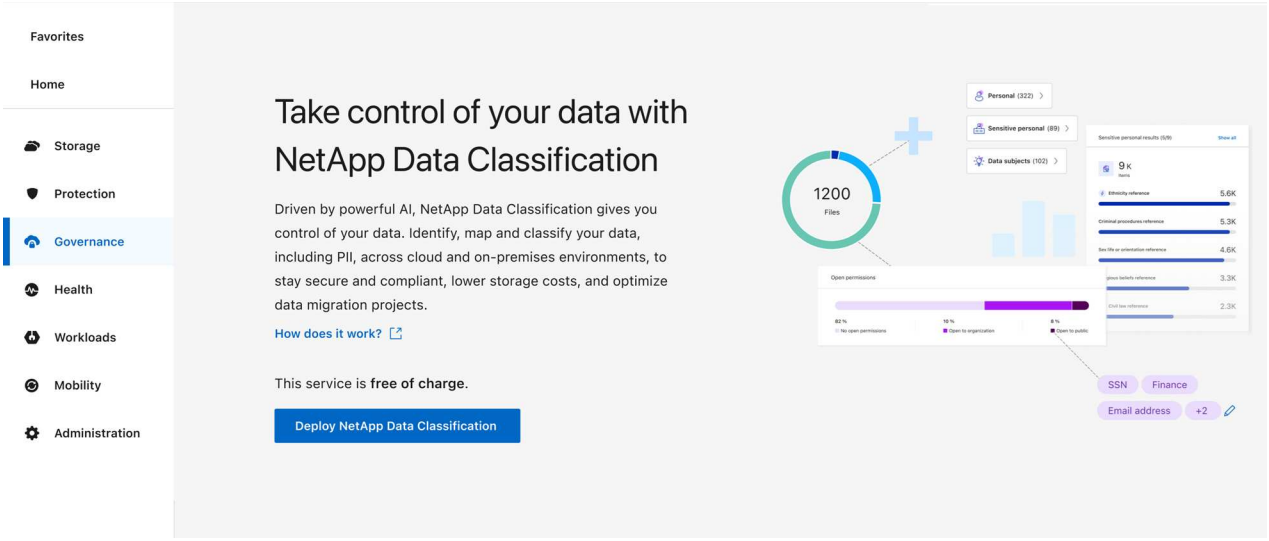


2. 从“安装”页面中，选择“部署”>“部署”以使用“大型”实例大小并启动云部署向导。
3. 向导在执行部署步骤时会显示进度。当需要输入或遇到问题时，系统会提示您。
4. 当实例部署完毕并安装数据分类后，选择“继续配置”进入“配置”页面。

在 **Azure** 中部署

步骤

1. 从数据分类主页中，选择*在本地或云中部署分类*。

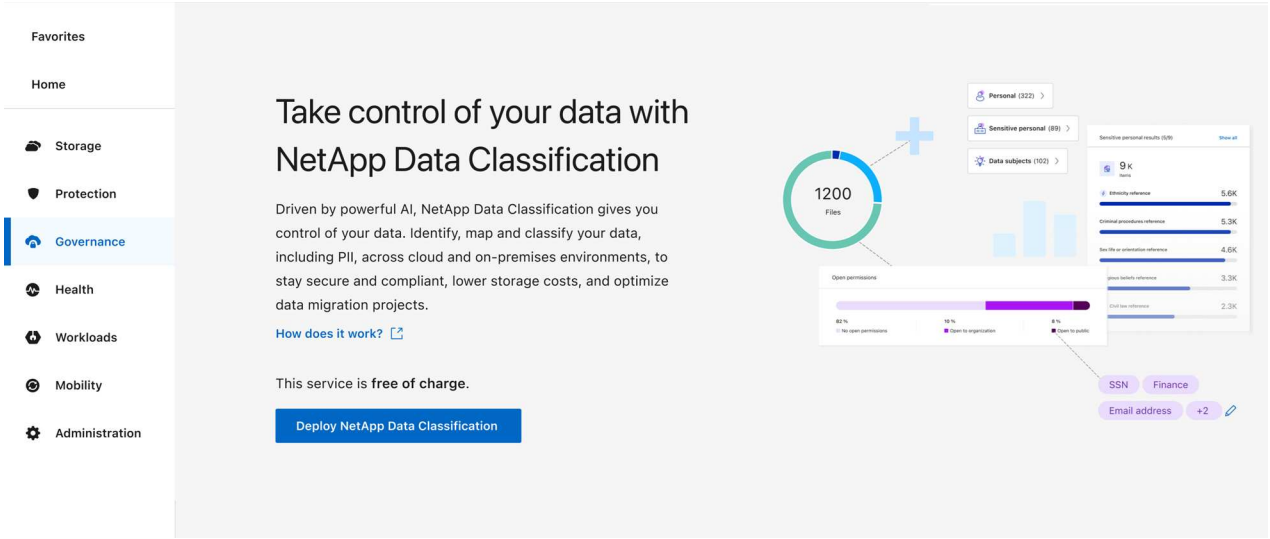


2. 选择*部署*以启动云部署向导。
3. 向导在执行部署步骤时会显示进度。如果遇到任何问题，它将停止并提示输入。
4. 当实例部署完毕并安装数据分类后，选择“继续配置”进入“配置”页面。

在 **Google Cloud** 中部署

步骤

1. 从数据分类主页中，选择*治理>分类*。
2. 选择*在本地或云中部署分类*。



3. 选择*部署*以启动云部署向导。
4. 向导在执行部署步骤时会显示进度。如果遇到任何问题，它将停止并提示输入。
5. 当实例部署完毕并安装数据分类后，选择“继续配置”进入“配置”页面。

结果

控制台在您的云提供商中部署数据分类实例。

只要实例具有互联网连接，控制台代理和数据分类软件的升级就会自动进行。

下一步

您可以从配置页面选择要扫描的数据源。

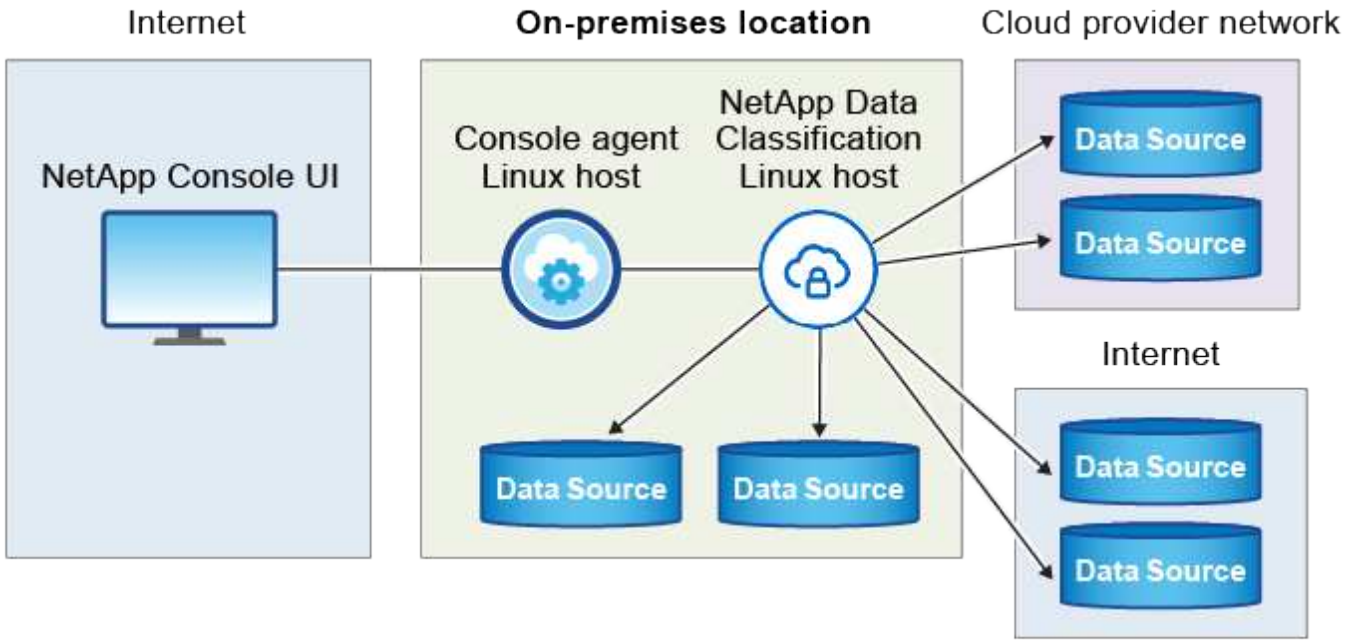
在可以访问互联网的主机上安装**NetApp Data Classification**

要在您的网络中的 Linux 主机或具有 Internet 访问权限的云中的 Linux 主机上部署NetApp Data Classification，您需要在您的网络或云中手动部署 Linux 主机。

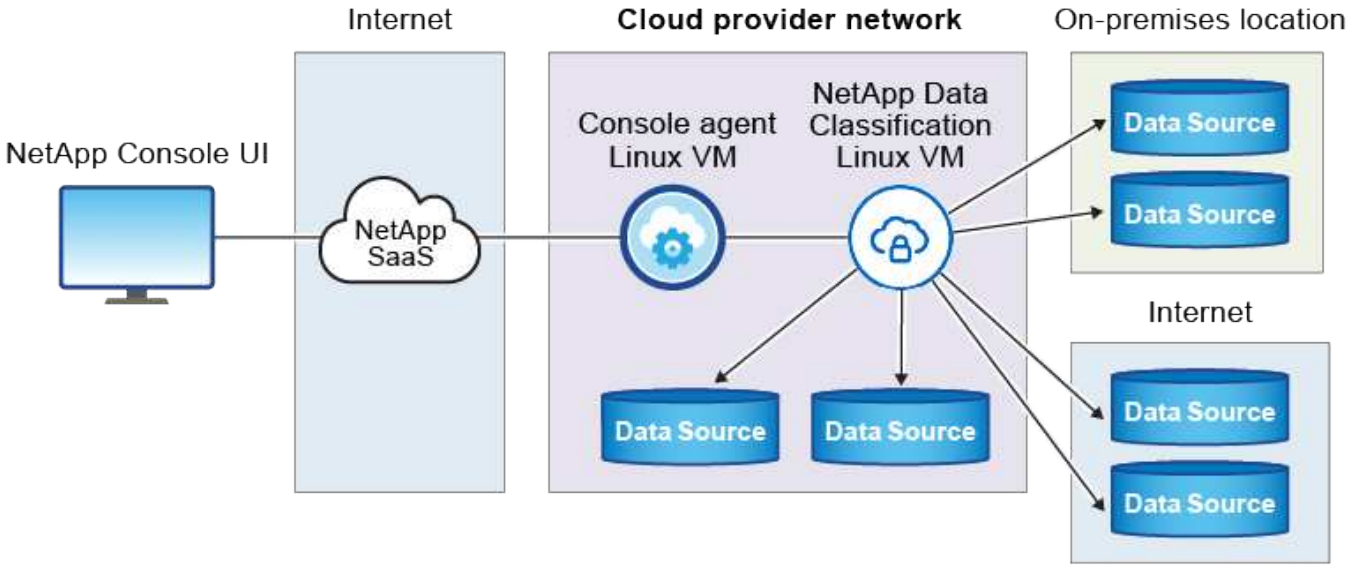
如果您希望使用同样位于本地的数据分类实例来扫描本地ONTAP系统，则本地安装是一个不错的选择。这不是必需的。无论选择哪种安装方法，软件的功能都是相同的。

数据分类安装脚本首先检查系统和环境是否满足所需的先决条件。如果所有先决条件都满足，则安装开始。如果您想独立于运行数据分类安装来验证先决条件，您可以下载一个单独的软件包，该软件包仅测试先决条件。"了解如何检查您的 Linux 主机是否已准备好安装数据分类" 。

您所在场所的 Linux 主机上的典型安装具有以下组件和连接。



云端 Linux 主机上的典型安装具有以下组件和连接。



快速启动

按照以下步骤快速开始，或者向下滚动到其余部分以获取完整详细信息。

1 创建控制台代理

如果您还没有控制台代理， "在本地部署控制台代理"在您的网络中的 Linux 主机上，或在云中的 Linux 主机上。

您还可以与您的云提供商一起创建控制台代理。看 "在 AWS 中创建控制台代理"， "在 Azure 中创建控制台代理" ， 或者 "在 GCP 中创建控制台代理"。

2 审查先决条件

确保您的环境能够满足先决条件。这包括实例的出站互联网访问、控制台代理和数据分类之间通过端口 443 的连接等等。查看完整列表 。

您还需要一个满足以下条件的 Linux 系统遵循要求。

3 下载并部署数据分类

从NetApp支持站点下载云数据分类软件，并将安装程序文件复制到您计划使用的 Linux 主机。然后启动安装向导并按照提示部署数据分类实例。

创建控制台代理

您需要先安装控制台代理，然后才能安装和使用数据分类。在大多数情况下，您可能会在尝试激活数据分类之前设置控制台代理，因为大多数 "控制台功能需要控制台代理"，但有些情况下您需要立即设置一个。

要在您的云提供商环境中创建一个，请参阅 "在 AWS 中创建控制台代理"， "在 Azure 中创建控制台代理" ， 或者 "在 GCP 中创建控制台代理"。

在某些情况下，您必须使用部署在特定云提供商中的控制台代理：

- 在 AWS 或Amazon FSx for ONTAP中的Cloud Volumes ONTAP中扫描数据时，您可以使用 AWS 中的控制台代理。
- 在 Azure 中的Cloud Volumes ONTAP或Azure NetApp Files中扫描数据时，您可以使用 Azure 中的控制台代理。

  对于Azure NetApp Files，它必须部署在与您要扫描的卷相同的区域中。

- 在 GCP 中的Cloud Volumes ONTAP中扫描数据时，您可以使用 GCP 中的控制台代理。

可以使用任何这些云控制台代理来扫描本地ONTAP系统、 NetApp文件共享和数据库帐户。

请注意，您还可以 "在本地部署控制台代理"在您的网络中的 Linux 主机上或云中的 Linux 主机上。一些计划在本地安装数据分类的用户可能还会选择在本地安装控制台代理。

安装数据分类时，您将需要控制台代理系统的 IP 地址或主机名。如果您在您的场所安装了控制台代理，您将获得此信息。如果控制台代理部署在云中，您可以从控制台中找到此信息：选择帮助图标，然后选择*支持*，然后选择控制台代理。

准备 **Linux** 主机系统

数据分类软件必须在满足特定操作系统要求、RAM 要求、软件要求等的主机上运行。 Linux 主机可以在您的网络中，也可以在云中。

确保您可以保持数据分类运行。数据分类机器需要保持开启状态以持续扫描您的数据。

- 数据分类必须运行在专用主机上。主机不能与其他应用程序或第三方软件（例如防病毒软件）共享。
- 选择与您计划使用数据分类扫描的数据集相符的大小。

系统大小	**CPU**	**RAM**（必须禁用交换内存）	磁盘
超大	32 个 CPU	128 GB 内存	• / 上 1 TiB SSD，或 /opt 上 100 GiB 可用 • /var/lib/docker 上可用 895 GiB • /tmp 上 5 GiB • 对于 **Podman**，**/var/tmp** 上有 **30 GB**
大的	16 个 CPU	64 GB 内存	• / 上 500 GiB SSD，或 /opt 上 100 GiB 可用 • /var/lib/docker 或 Podman /var/lib/containers 上可用 400 GiB • /tmp 上 5 GiB • 对于 **Podman**，**/var/tmp** 上有 **30 GB**

- 在云中为数据分类安装部署计算实例时，建议您使用满足上述“大型”系统要求的系统：
  - **Amazon Elastic Compute Cloud (Amazon EC2)** 实例类型：“m6i.4xlarge”。"查看其他 AWS 实例类型" 。
  - **Azure VM** 大小：“Standard_D16s_v3”。"查看其他 Azure 实例类型" 。
  - **GCP** 机器类型：“n2-standard-16”。"查看其他 GCP 实例类型" 。
- **UNIX** 文件夹权限：需要以下最低 UNIX 权限：

文件夹	最低权限
/tmp	`rwxrwxrwt`
/选择	`rwxr-xr-x`
/var/lib/docker	`rwx------`
/usr/lib/systemd/系统	`rwxr-xr-x`

- 操作系统：
  - 以下操作系统需要使用 Docker 容器引擎：
    - Red Hat Enterprise Linux 版本 7.8 和 7.9
    - Ubuntu 22.04（需要数据分类版本 1.23 或更高版本）
    - Ubuntu 24.04（需要数据分类版本 1.23 或更高版本）
  - 以下操作系统需要使用 Podman 容器引擎，并且需要数据分类版本 1.30 或更高版本：
    - Red Hat Enterprise Linux 版本 8.8、8.10、9.0、9.1、9.2、9.3、9.4、9.5 和 9.6。
  - 必须在主机系统上启用高级矢量扩展 (AVX2)。

- **Red Hat** 订阅管理：主机必须在 Red Hat 订阅管理中注册。如果未注册，系统将无法访问存储库来在安装期间更新所需的第三方软件。
- 附加软件：安装数据分类之前，必须在主机上安装以下软件：
  - 根据您使用的操作系统，您需要安装其中一个容器引擎：
    - Docker Engine 版本 19.3.1 或更高版本。 "查看安装说明" 。
    - Podman 版本 4 或更高版本。要安装 Podman，请输入(`sudo yum install podman netavark -y`）。
- Python 版本 3.6 或更高版本。 "查看安装说明" 。
  - **NTP** 注意事项： NetApp建议配置数据分类系统以使用网络时间协议 (NTP) 服务。数据分类系统和控制台代理系统之间的时间必须同步。
- **Firewalld** 注意事项：如果您计划使用 `firewalld`，我们建议您在安装数据分类之前启用它。运行以下命令进行配置 `firewalld`以便与数据分类兼容：

```
firewall-cmd --permanent --add-service=http
firewall-cmd --permanent --add-service=https
firewall-cmd --permanent --add-port=80/tcp
firewall-cmd --permanent --add-port=8080/tcp
firewall-cmd --permanent --add-port=443/tcp
firewall-cmd --reload
```

如果您计划使用其他数据分类主机作为扫描器节点，请在此时将这些规则添加到您的主系统：

```
firewall-cmd --permanent --add-port=2377/tcp
firewall-cmd --permanent --add-port=7946/udp
firewall-cmd --permanent --add-port=7946/tcp
firewall-cmd --permanent --add-port=4789/udp
```

请注意，每次启用或更新时都必须重新启动 Docker 或 Podman `firewalld`设置。

安装后，数据分类主机系统的 IP 地址无法更改。

启用数据分类的出站互联网访问

数据分类需要出站互联网访问。如果您的虚拟或物理网络使用代理服务器进行互联网访问，请确保数据分类实例具有出站互联网访问权限以联系以下端点。

端点	目的
\ https://api.console.netapp.com	与控制台的通信，其中包括NetApp帐户。
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	与控制台网站通信，实现集中用户身份验证。

端点	目的
\ https://support.compliance.api.bluexp.netapp.com/ \ https://hub.docker.com \ https://auth.docker.io \ https://registry-1.docker.io \ https://index.docker.io/ \ https://dseasb33srnrn.cloudfront.net/ \ https://production.cloudflare.docker.com/	提供对软件映像、清单、模板的访问以及发送日志和指标。
https://support.compliance.api.bluexp.netapp.com/	使NetApp能够从审计记录中流式传输数据。
https://github.com/docker https://download.docker.com	提供docker安装的必备包。
\ http://packages.ubuntu.com/ \ http://archive.ubuntu.com	提供 Ubuntu 安装的必备软件包。

验证所有必需的端口均已启用

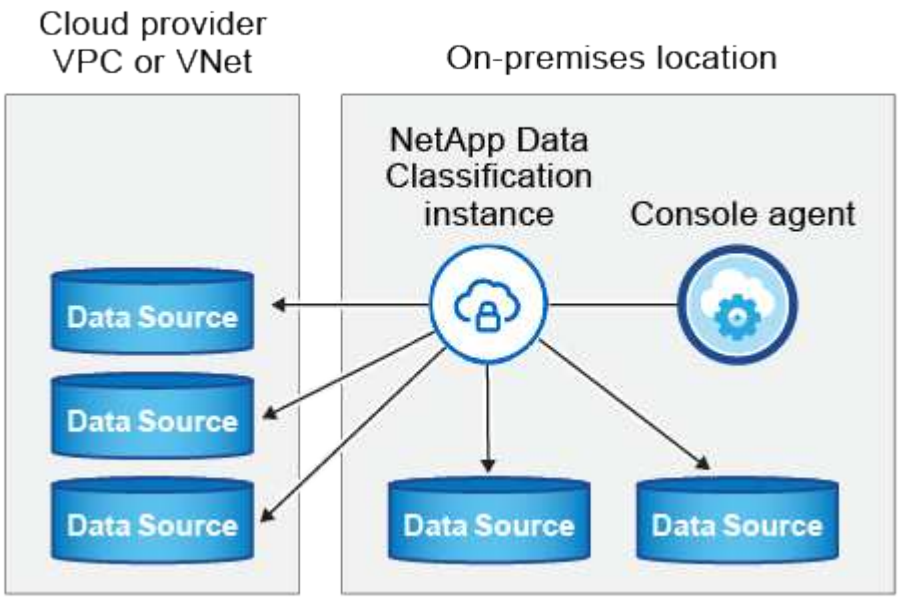
您必须确保所有必需的端口都已打开，以便控制台代理、数据分类、Active Directory 和数据源之间进行通信。

连接类型	端口	描述
控制台代理<>数据分类	8080 (TCP)、443 (TCP) 和 80。9000	控制台代理的防火墙或路由规则必须允许通过端口 443 进出数据分类实例的入站和出站流量。确保端口 8080 已打开，以便您可以在控制台中看到安装进度。如果 Linux 主机上使用防火墙，则 Ubuntu 服务器内的内部进程需要端口 9000。
控制台代理<> ONTAP集群 (NAS)	443（TCP）	控制台使用 HTTPS 发现ONTAP集群。如果您使用自定义防火墙策略，则它们必须满足以下要求： • 控制台代理主机必须允许通过端口 443 进行出站 HTTPS 访问。如果控制台代理位于云中，则预定义的防火墙或路由规则允许所有出站通信。 • ONTAP集群必须允许通过端口 443 进行入站 HTTPS 访问。默认的“mgmt”防火墙策略允许来自所有 IP 地址的入站 HTTPS 访问。如果您修改了此默认策略，或者创建了自己的防火墙策略，则必须将 HTTPS 协议与该策略关联并启用从控制台代理主机的访问。

连接类型	端口	描述
数据分类 <> ONTAP集群	• 对于 NFS - 111 (TCP\UDP) 和 2049 (TCP\UDP) • 对于 CIFS - 139 (TCP\UDP) 和 445 (TCP\UDP)	数据分类需要与每个Cloud Volumes ONTAP子网或本地ONTAP系统建立网络连接。 Cloud Volumes ONTAP的防火墙或路由规则必须允许来自数据分类实例的入站连接。 确保这些端口对数据分类实例开放： • 对于 NFS - 111 和 2049 • 对于 CIFS - 139 和 445 NFS 卷导出策略必须允许从数据分类实例进行访问。
数据分类<> Active Directory	389（TCP 和 UDP ）、636（TCP）、3268（TCP）和 3269（TCP）	您必须已经为公司用户设置了 Active Directory。此外，数据分类需要 Active Directory 凭据来扫描 CIFS 卷。 您必须具有 Active Directory 的信息： • DNS 服务器 IP 地址，或多个 IP 地址 • 服务器的用户名和密码 • 域名（Active Directory 名称） • 您是否使用安全 LDAP (LDAPS) • LDAP 服务器端口（LDAP 通常为 389，安全 LDAP 通常为 636）

在 **Linux** 主机上安装数据分类

对于典型配置，您将在单个主机系统上安装该软件。请参阅此处的步骤 。



看准备 Linux 主机系统和审查先决条件了解部署数据分类之前的完整要求列表。

只要实例具有互联网连接，数据分类软件的升级就会自动进行。

> 当软件安装在本地时，数据分类当前无法扫描 S3 存储桶、 Azure NetApp Files或 FSx for ONTAP 。在这些情况下，您需要在云中部署单独的控制台代理和数据分类实例，并且 "在连接器之间切换"适用于不同的数据源。

典型配置的单主机安装

在单个本地主机上安装数据分类软件时，请查看要求并遵循以下步骤。

"观看此视频"了解如何安装数据分类。

请注意，安装数据分类时会记录所有安装活动。如果您在安装过程中遇到任何问题，您可以查看安装审计日志的内容。它被写给 `/opt/netapp/install_logs/`。

开始之前

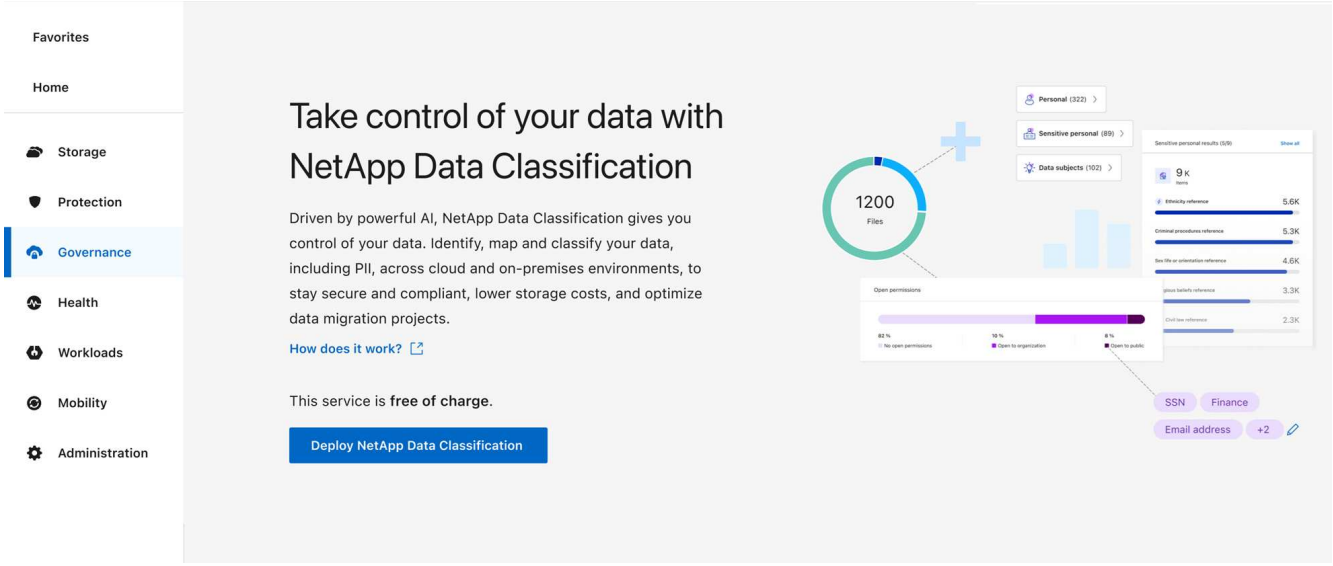
- 验证您的 Linux 系统是否满足主机要求。
- 验证系统是否安装了两个必备软件包（Docker Engine 或 Podman 和 Python 3）。
- 确保您在 Linux 系统上拥有 root 权限。
- 如果您使用代理访问互联网：
  - 您将需要代理服务器信息（IP 地址或主机名、连接端口、连接方案：https 或 http、用户名和密码）。
  - 如果代理正在执行 TLS 拦截，您需要知道数据分类 Linux 系统上存储 TLS CA 证书的路径。
  - 代理必须是非透明的。数据分类目前不支持透明代理。
  - 该用户必须是本地用户。不支持域用户。
- 验证您的离线环境是否满足要求权限和连接性。

步骤

1. 从下载数据分类软件 "NetApp 支持站点"。您应该选择的文件名为 **DATASENSE-INSTALLER-<version>.tar.gz**。
2. 将安装程序文件复制到您计划使用的 Linux 主机（使用 `scp`或其他方法）。
3. 在主机上解压安装程序文件，例如：

```
tar -xzf DATASENSE-INSTALLER-V1.25.0.tar.gz
```

4. 在控制台中，选择*治理>分类*。
5. 选择*在本地或云中部署分类*。



6. 根据您是在云中准备的实例上还是在本地准备的实例上安装数据分类，选择适当的*部署*选项来启动数据分类安装。
7. 将显示“在本地部署数据分类”对话框。复制提供的命令（例如： `sudo ./install.sh -a 12345 -c 27AG75 -t 2198qq` ）并将其粘贴到文本文件中，以便稍后使用。然后选择*关闭*以关闭对话框。
8. 在主机上，输入您复制的命令，然后按照一系列提示进行操作，或者您可以提供包含所有必需参数的完整命令作为命令行参数。

   请注意，安装程序会执行预检查以确保您的系统和网络要求满足，以便成功安装。 "观看此视频"了解预检信息和含义。

根据提示输入参数：	输入完整命令：
a. 粘贴从步骤 7 复制的命令： `sudo ./install.sh -a <account_id> -c <client_id> -t <user_token>` 如果您在云实例上安装（而不是在您的本地），请添加 `--manual-cloud-install <cloud_provider>`。 b. 输入数据分类主机的 IP 地址或主机名，以便控制台代理系统可以访问它。 c. 输入控制台代理主机的 IP 地址或主机名，以便数据分类系统可以访问它。 d. 根据提示输入代理详细信息。如果您的控制台代理已经使用代理，则无需在此处再次输入此信息，因为数据分类将自动使用控制台代理所使用的代理。	或者，您可以提前创建整个命令，提供必要的主机和代理参数： `sudo ./install.sh -a <account_id> -c <client_id> -t <user_token> --host <ds_host> --manager-host <cm_host> --manual-cloud-install <cloud_provider> --proxy-host <proxy_host> --proxy-port <proxy_port> --proxy-scheme <proxy_scheme> --proxy -user <proxy_user> --proxy-password <proxy_password> --cacert-folder-path <ca_cert_dir>`

变量值：

- *account_id* = NetApp帐户 ID

- *client_id* = 控制台代理客户端 ID（如果客户端 ID 中没有后缀“clients”，则添加后缀）
- *user_token* = JWT 用户访问令牌
- *ds_host* = 数据分类 Linux 系统的 IP 地址或主机名。
- *cm_host* = 控制台代理系统的 IP 地址或主机名。
- *cloud_provider* = 在云实例上安装时，根据云提供商输入“AWS”、“Azure”或“Gcp”。
- *proxy_host* = 如果主机位于代理服务器后面，则为代理服务器的 IP 或主机名。
- *proxy_port* = 连接到代理服务器的端口（默认为 80）。
- *proxy_scheme* = 连接方案：https 或 http（默认 http）。
- *proxy_user* = 如果需要基本身份验证，则经过身份验证的用户连接到代理服务器。用户必须是本地用户 - 不支持域用户。
- *proxy_password* = 您指定的用户名的密码。
- *ca_cert_dir* = 数据分类 Linux 系统上包含附加 TLS CA 证书包的路径。仅当代理执行 TLS 拦截时才需要。

结果

数据分类安装程序安装包、注册安装并安装数据分类。安装可能需要 10 到 20 分钟。

如果主机和控制台代理实例之间通过端口 8080 建立连接，您将在控制台的“数据分类”选项卡中看到安装进度。

下一步

您可以从配置页面选择要扫描的数据源。

在没有互联网访问的 **Linux** 主机上安装**NetApp Data Classification**

在没有互联网访问权限的本地站点的 Linux 主机上安装NetApp Data Classification称为_私有模式_。这种安装类型使用安装脚本，与NetApp Console SaaS 层没有连接。

BlueXP私有模式（传统BlueXP接口）通常用于没有互联网连接的本地环境和安全云区域，其中包括 AWS Secret Cloud、AWS Top Secret Cloud 和 Azure IL6。NetApp继续通过传统的BlueXP界面支持这些环境。有关旧版BlueXP界面中的私有模式文档，请参阅"BlueXP私人模式的 PDF 文档"。

检查您的 **Linux** 主机是否已准备好安装**NetApp Data Classification**

在 Linux 主机上手动安装NetApp Data Classification之前，可以选择在主机上运行脚本以验证安装数据分类的所有先决条件是否都已具备。您可以在网络中的 Linux 主机上或云中的 Linux 主机上运行此脚本。主机可以连接到互联网，或者主机可以驻留在没有互联网访问的站点（暗站）。

数据分类安装脚本包含一个测试脚本，以确保您的环境满足要求。您可以单独运行此脚本来验证 Linux 主机是否已准备就绪，然后再运行安装脚本。

入门指南

您将执行以下任务。

- 如果您尚未安装控制台代理，则可以选择安装一个。您可以在未安装控制台代理的情况下运行测试脚本，但脚本会检查控制台代理和数据分类主机之间的连接 - 因此建议您安装控制台代理。
- 准备主机并验证其是否满足所有要求。
- 启用数据分类主机的出站互联网访问。
- 验证所有系统上的所有必需端口是否都已启用。
- 下载并运行先决条件测试脚本。

创建控制台代理

您需要先安装控制台代理，然后才能安装和使用数据分类。但是，您可以在没有控制台代理的情况下运行先决条件脚本。

你可以 "在本地安装控制台代理" 在您网络中的 Linux 主机上，或者在云端的 Linux 主机上。如果控制台代理安装在本地，您也可以在本地安装数据分类。

要在您的云提供商环境中创建控制台代理，请参阅：

- "在 AWS 中创建控制台代理"
- "在 Azure 中创建控制台代理"
- "在 GCP 中创建控制台代理"

运行先决条件脚本时，需要控制台代理系统的 IP 地址或主机名。如果您在本地安装了控制台代理，则可以获取此信息。如果控制台代理部署在云端，您可以从控制台中找到此信息：选择帮助图标，然后选择“支持”；在“代理和审核”部分，选择“转到代理”。

验证主机要求

数据分类软件必须运行在满足特定操作系统要求、内存要求和软件要求的主机上。

- 数据分类必须运行在专用主机上。主机不能与其他应用程序或第三方软件（例如防病毒软件）共享。
- 选择与您计划使用数据分类扫描的数据集相符的大小。

系统大小	**CPU**	**RAM**（必须禁用交换内存）	磁盘
超大	32 个 CPU	128 GB 内存	• / 上 1 TiB SSD，或 /opt 上 100 GiB 可用 • /var/lib/docker 上可用 895 GiB • /tmp 上 5 GiB • 对于 **Podman**，**/var/tmp** 上有 **30 GB**

系统大小	**CPU**	**RAM**（必须禁用交换内存）	磁盘
大的	16 个 CPU	64 GB 内存	• / 上 500 GiB SSD，或 /opt 上 100 GiB 可用 • /var/lib/docker 或 Podman /var/lib/containers 上可用 400 GiB • /tmp 上 5 GiB • 对于 **Podman**，**/var/tmp** 上有 **30 GB**

- 在云中为数据分类安装部署计算实例时，建议您使用满足上述“大型”系统要求的系统：
  - **Amazon Elastic Compute Cloud (Amazon EC2)** 实例类型：“m6i.4xlarge”。"查看其他 AWS 实例类型" 。
  - **Azure VM** 大小：“Standard_D16s_v3”。"查看其他 Azure 实例类型" 。
  - **GCP** 机器类型：“n2-standard-16”。"查看其他 GCP 实例类型" 。
- **UNIX** 文件夹权限：需要以下最低 UNIX 权限：

文件夹	最低权限
/tmp	`rwxrwxrwt`
/选择	`rwxr-xr-x`
/var/lib/docker	`rwx------`
/usr/lib/systemd/系统	`rwxr-xr-x`

- 操作系统：
  - 以下操作系统需要使用 Docker 容器引擎：
    - Red Hat Enterprise Linux 版本 7.8 和 7.9
    - Ubuntu 22.04（需要数据分类版本 1.23 或更高版本）
    - Ubuntu 24.04（需要数据分类版本 1.23 或更高版本）
  - 以下操作系统需要使用 Podman 容器引擎，并且需要数据分类版本 1.30 或更高版本：
    - Red Hat Enterprise Linux 版本 8.8、8.10、9.0、9.1、9.2、9.3、9.4、9.5 和 9.6。
  - 必须在主机系统上启用高级矢量扩展 (AVX2)。
- **Red Hat** 订阅管理：主机必须在 Red Hat 订阅管理中注册。如果未注册，系统将无法访问存储库来在安装期间更新所需的第三方软件。
- 附加软件：安装数据分类之前，必须在主机上安装以下软件：
  - 根据您使用的操作系统，您需要安装其中一个容器引擎：
    - Docker Engine 版本 19.3.1 或更高版本。 "查看安装说明" 。
    - Podman 版本 4 或更高版本。要安装 Podman，请输入(`sudo yum install podman netavark -y`）。

- Python 版本 3.6 或更高版本。 "查看安装说明" 。
  - **NTP** 注意事项： NetApp建议配置数据分类系统以使用网络时间协议 (NTP) 服务。数据分类系统和控制台代理系统之间的时间必须同步。
- **Firewalld** 注意事项：如果您计划使用 `firewalld`，我们建议您在安装数据分类之前启用它。运行以下命令进行配置 `firewalld`以便与数据分类兼容：

```
firewall-cmd --permanent --add-service=http
firewall-cmd --permanent --add-service=https
firewall-cmd --permanent --add-port=80/tcp
firewall-cmd --permanent --add-port=8080/tcp
firewall-cmd --permanent --add-port=443/tcp
firewall-cmd --reload
```

如果您计划使用其他数据分类主机作为扫描器节点（在分布式模型中），请在此时将这些规则添加到您的主系统：

```
firewall-cmd --permanent --add-port=2377/tcp
firewall-cmd --permanent --add-port=7946/udp
firewall-cmd --permanent --add-port=7946/tcp
firewall-cmd --permanent --add-port=4789/udp
```

请注意，每次启用或更新时都必须重新启动 Docker 或 Podman `firewalld`设置。

启用数据分类的出站互联网访问

数据分类需要出站互联网访问。如果您的虚拟或物理网络使用代理服务器进行互联网访问，请确保数据分类实例具有出站互联网访问权限以联系以下端点。

对于安装在没有互联网连接的站点的主机系统，不需要本节。

端点	目的
\ https://api.console.netapp.com	与控制台服务（包括NetApp帐户）的通信。
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	与控制台网站通信，实现集中用户身份验证。
\ https://support.compliance.api.console.netapp.com/ \ https://hub.docker.com \ https://auth.docker.io \ https://registry-1.docker.io \ https://index.docker.io/ \ https://dseasb33srnrn.cloudfront.net/ \ https://production.cloudflare.docker.com/	提供对软件映像、清单、模板的访问以及发送日志和指标。
\ https://support.compliance.api.console.netapp.com/	使NetApp能够从审计记录中流式传输数据。

端点	目的
https://github.com/docker https://download.docker.com	提供docker安装的必备包。
\ http://packages.ubuntu.com/ \ http://archive.ubuntu.com	提供 Ubuntu 安装的必备软件包。

验证所有必需的端口均已启用

您必须确保所有必需的端口都已打开，以便控制台代理、数据分类、Active Directory 和数据源之间进行通信。

连接类型	端口	描述
控制台代理<>数据分类	8080 (TCP)、443 (TCP)和 80。9000	控制台代理的防火墙或路由规则必须允许通过端口 443 进出数据分类实例的入站和出站流量。确保端口 8080 已打开，以便您可以在控制台中看到安装进度。如果 Linux 主机上使用防火墙，则 Ubuntu 服务器内的内部进程需要端口 9000。
控制台代理<> ONTAP集群 (NAS)	443（TCP）	控制台使用 HTTPS 发现ONTAP集群。如果您使用自定义防火墙策略，控制台代理主机必须允许通过端口 443 进行出站 HTTPS 访问。如果控制台代理位于云中，则预定义的防火墙或路由规则允许所有出站通信。

运行数据分类先决条件脚本

按照以下步骤运行数据分类先决条件脚本。

"观看此视频"了解如何运行先决条件脚本并解释结果。

开始之前

- 验证您的 Linux 系统是否满足主机要求。
- 验证系统是否安装了两个必备软件包（Docker Engine 或 Podman 和 Python 3）。
- 确保您在 Linux 系统上拥有 root 权限。

步骤

1. 从下载数据分类先决条件脚本 "NetApp 支持站点"。您应该选择的文件名为 **standalone-pre-requisite-tester-<version>**。
2. 将文件复制到您计划使用的 Linux 主机（使用 `scp`或其他方法）。
3. 分配运行脚本的权限。

```
chmod +x standalone-pre-requisite-tester-v1.25.0
```

4. 使用以下命令运行脚本。

```
./standalone-pre-requisite-tester-v1.25.0 <--darksite>
```

仅当您在没有互联网访问的主机上运行脚本时才添加选项“--darksite”。当主机未连接到互联网时，某些先决条件测试将被跳过。

5. 该脚本会提示您输入数据分类主机的 IP 地址。
   - 输入 IP 地址或主机名。
6. 该脚本会提示您是否安装了控制台代理。
   - 如果您没有安装控制台代理，请输入 **N**。
   - 如果您确实安装了控制台代理，请输入 **Y**。然后输入控制台代理的 IP 地址或主机名，以便测试脚本可以测试此连接。
7. 该脚本在系统上运行各种测试，并在运行过程中显示结果。完成后，它会将会话日志写入名为 `prerequisites-test-<timestamp>.log`在目录中 `/opt/netapp/install_logs`。

结果

如果所有先决条件测试均成功运行，您可以在准备就绪后在主机上安装数据分类。

如果发现任何问题，则将其归类为“建议”或“需要”修复。推荐的问题通常是会使数据分类扫描和分类任务运行速度变慢的项目。这些项目不需要更正 - 但您可能需要解决它们。

如果您有任何“必需”问题，您应该修复这些问题并再次运行先决条件测试脚本。

激活数据源扫描

使用NetApp Data Classification扫描数据源

NetApp Data Classification会扫描您选择的存储库（卷、数据库模式或其他用户数据）中的数据，以识别个人和敏感数据。然后，数据分类会映射您的组织数据、对每个文件进行分类并识别数据中的预定义模式。扫描结果是个人信息、敏感个人信息、数据类别和文件类型的索引。

初始扫描后，数据分类将以循环方式持续扫描您的数据以检测增量变化。这就是为什么保持实例运行很重要。

您可以在卷级别或数据库模式级别启用和禁用扫描。

映射扫描和分类扫描之间有什么区别

您可以在数据分类中进行两种类型的扫描：

- 仅映射扫描仅提供数据的高级概览，并在选定的数据源上执行。仅映射扫描比映射和分类扫描花费的时间更少，因为它们不访问文件来查看其中的数据。您可能希望首先执行此操作来确定研究领域，然后对这些领域执行地图和分类扫描。
- 地图和分类扫描 为您的数据提供深层扫描。

下表显示了一些差异：

功能	映射和分类扫描	仅映射扫描
扫描速度	慢	快

功能	映射和分类扫描	仅映射扫描
定价	可用	可用
容量	限制为 500 TiB*	限制为 500 TiB*
文件类型和已用容量列表	是	是
文件数量和已用容量	是	是
文件的年龄和大小	是	是
能够运行"数据映射报告"	是	是
数据调查页面查看文件详细信息	是	否
在文件中搜索名称	是	否
创造"已保存的查询"提供自定义搜索结果	是	否
能够运行其他报告	是	否
能够查看文件中的元数据**	否	是

* 数据分类不会对其可以扫描的数据量施加限制。每个控制台代理支持扫描和显示 500 TiB 的数据。要扫描超过 500 TiB 的数据，"安装另一个控制台代理"然后"部署另一个数据分类实例"。 + 控制台 UI 显示来自单个连接器的数据。有关查看来自多个控制台代理的数据的提示，请参阅"使用多个控制台代理"。

** 在映射扫描期间从文件中提取以下元数据：

- 系统
- 系统类型
- 存储库
- 文件类型
- 已用容量
- 文件数
- 文件大小
- 文件创建
- 文件上次访问
- 文件上次修改时间
- 文件发现时间
- 权限提取

治理仪表板差异：

功能	地图和分类	映射
过时的数据	是	是
非业务数据	是	是
重复文件	是	是
预定义保存的查询	是	否
默认保存的查询	是	是
DDA 报告	是	是
地图报告	是	是
灵敏度等级检测	是	否
具有广泛权限的敏感数据	是	否
开放权限	是	是
数据时代	是	是
数据大小	是	是
类别	是	否
文件类型	是	是

合规性仪表板差异：

功能	地图和分类	映射
个人信息	是	否
敏感个人信息	是	否
隐私风险评估报告	是	否
HIPAA 报告	是	否
PCI DSS 报告	是	否

调查过滤器差异：

功能	地图和分类	映射
已保存的查询	是	是
系统类型	是	是
系统	是	是
存储库	是	是
文件类型	是	是
文件大小	是	是
创建时间	是	是
发现时间	是	是
上次修改时间	是	是
上次访问	是	是
开放权限	是	是
文件目录路径	是	是
类别	是	否
敏感度等级	是	否
标识符数量	是	否
个人数据	是	否
敏感个人数据	是	否
数据主体	是	否
重复项	是	是
分类状态	是	状态始终为“见解有限”
扫描分析事件	是	是
文件哈希	是	是
有访问权限的用户数	是	是
用户/组权限	是	是
文件所有者	是	是
目录类型	是	是

使用**NetApp Data Classification**扫描**Amazon FSx for ONTAP**卷

完成几个步骤即可使用NetApp Data Classification扫描Amazon FSx for ONTAP卷。

开始之前

- 您需要 AWS 中一个活动的控制台代理来部署和管理数据分类。
- 创建系统时选择的安全组必须允许来自数据分类实例的流量。您可以使用连接到 FSx for ONTAP文件系统的 ENI 找到关联的安全组，并使用 AWS 管理控制台对其进行编辑。

  "Linux 实例的 AWS 安全组"

  "Windows 实例的 AWS 安全组"

  "AWS 弹性网络接口 (ENI)"

- 确保以下端口对数据分类实例开放：
  - 对于 NFS – 端口 111 和 2049。
  - 对于 CIFS – 端口 139 和 445。

部署数据分类实例

"部署数据分类"如果尚未部署实例。

您应该在与 AWS 控制台代理和要扫描的 FSx 卷相同的 AWS 网络中部署数据分类。

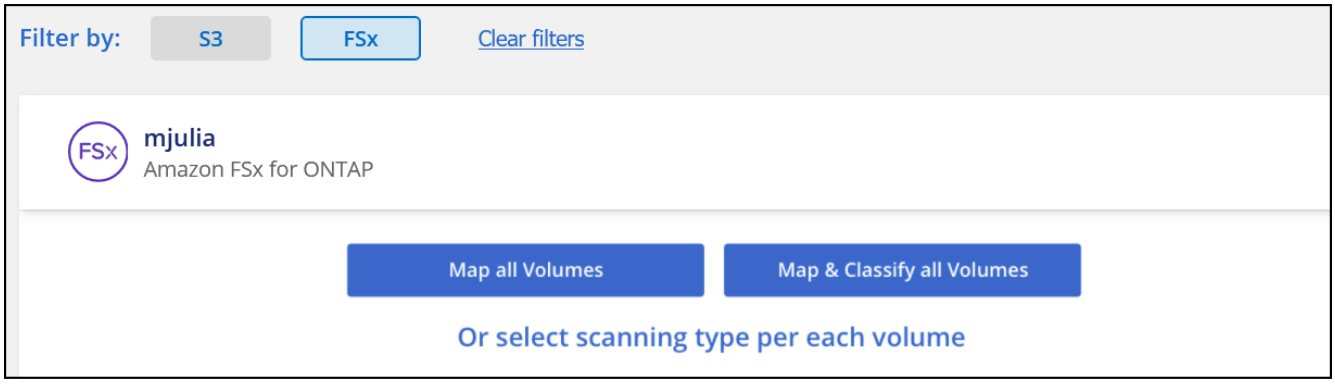
*注意：*扫描 FSx 卷时，当前不支持在本地位置部署数据分类。

只要实例具有互联网连接，数据分类软件的升级就会自动进行。

在您的系统中启用数据分类

您可以为 FSx for ONTAP卷启用数据分类。

1. 从NetApp Console，治理 > 分类。
2. 从数据分类菜单中，选择*配置*。



3. 选择如何扫描每个系统中的卷。"了解映射和分类扫描"：
   - 要映射所有卷，请选择*映射所有卷*。
   - 要映射和分类所有卷，请选择*映射和分类所有卷*。
   - 要自定义每个卷的扫描，请选择*或选择每个卷的扫描类型*，然后选择要映射和/或分类的卷。

4. 在确认对话框中，选择“批准”以使数据分类开始扫描您的卷。

结果

数据分类开始扫描您在系统中选择的卷。一旦数据分类完成初始扫描，结果将在合规性仪表板中提供。所需时间取决于数据量——可能是几分钟或几小时。您可以通过导航到配置菜单然后选择系统配置来跟踪初始扫描的进度。在进度条中跟踪每次扫描的进度；您可以将鼠标悬停在进度条上，以查看相对于卷中总文件数的扫描文件数。

- 默认情况下，如果数据分类在 CIFS 中没有写入属性权限，或者在 NFS 中没有写入权限，系统将不会扫描卷中的文件，因为数据分类无法将“上次访问时间”恢复为原始时间戳。如果您不介意上次访问时间是否重置，请选择*或为每个卷选择扫描类型*。结果页面有一个您可以启用的设置，以便数据分类可以扫描卷，而不管权限如何。
- 数据分类仅扫描卷下的一个文件共享。如果您的卷中有多个共享，则需要将这些其他共享作为共享组单独扫描。"查看有关此数据分类限制的更多详细信息" 。

验证数据分类是否有权访问卷

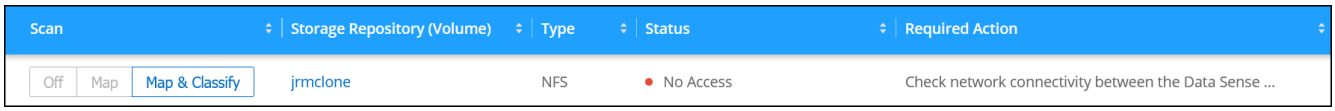
通过检查网络、安全组和导出策略，确保数据分类可以访问卷。

您需要向数据分类提供 CIFS 凭据，以便它可以访问 CIFS 卷。

步骤

1. 从数据分类菜单中，选择*配置*。
2. 在配置页面上，选择*查看详细信息*以查看状态并纠正任何错误。

   例如，下图显示由于数据分类实例和卷之间的网络连接问题，数据分类无法扫描卷。



3. 确保数据分类实例与包含 FSx for ONTAP卷的每个网络之间存在网络连接。

   对于 FSx for ONTAP，数据分类只能扫描与控制台位于同一区域的卷。

4. 确保 NFS 卷导出策略包含数据分类实例的 IP 地址，以便它可以访问每个卷上的数据。
5. 如果您使用 CIFS，请向数据分类提供 Active Directory 凭据，以便它可以扫描 CIFS 卷。

   a. 从数据分类菜单中，选择*配置*。

   b. 对于每个系统，选择*编辑 CIFS 凭据*并输入数据分类访问系统上的 CIFS 卷所需的用户名和密码。

      凭据可以是只读的，但提供管理员凭据可确保数据分类可以读取任何需要提升权限的数据。凭证存储在数据分类实例上。

      如果您想确保文件的“上次访问时间”不会因数据分类扫描而改变，建议用户在 CIFS 中具有写入属性权限或在 NFS 中具有写入权限。如果可能，请将 Active Directory 用户配置为组织中具有所有文件权限的父组的一部分。

   输入凭据后，您应该会看到一条消息，表明所有 CIFS 卷均已成功验证。

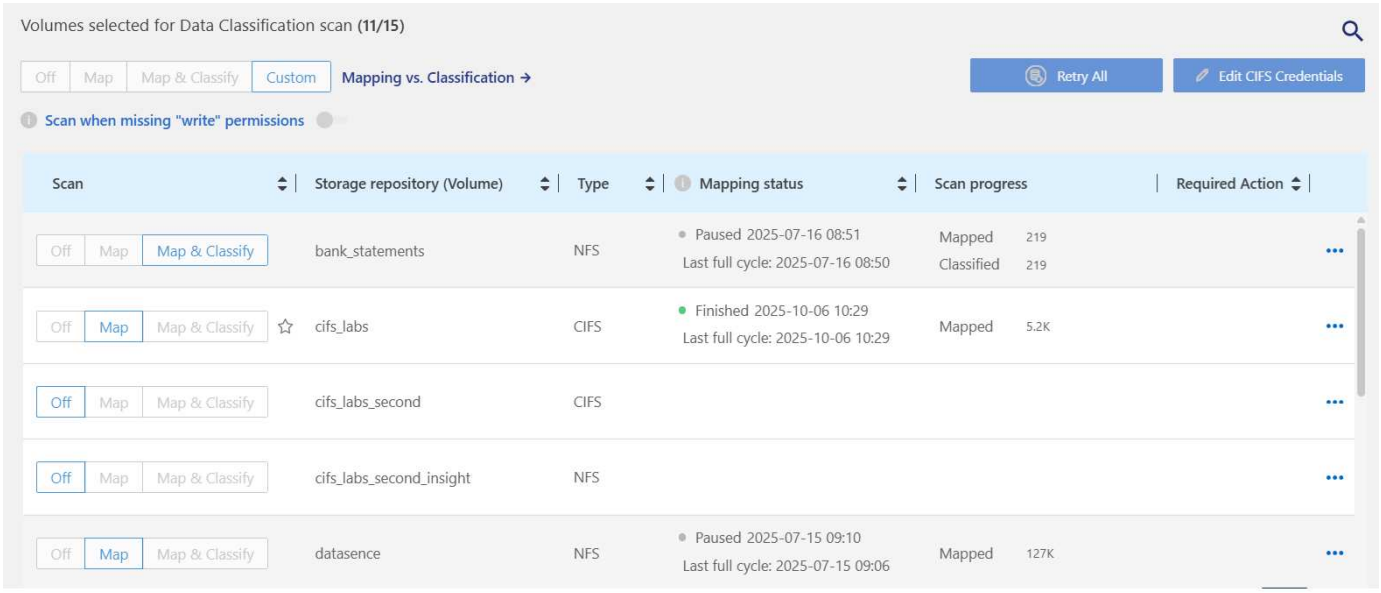
启用和禁用卷上的扫描

您可以随时从配置页面启动或停止任何系统上的扫描。您还可以将扫描从仅映射扫描切换到映射和分类扫描，反之亦然。建议您扫描系统中的所有卷。

仅当您在标题区域中选择了 **Map** 或 **Map & Classify** 设置时，才会自动扫描添加到系统的新卷。当在标题区域设置为*自定义*或*关闭*时，您需要在系统中添加的每个新卷上激活映射和/或完整扫描。

页面顶部的“缺少“写入”权限时扫描”开关默认处于禁用状态。这意味着，如果数据分类在 CIFS 中没有写属性权限或在 NFS 中没有写权限，系统将不会扫描文件，因为数据分类无法将“上次访问时间”恢复为原始时间戳。如果您不介意是否重置上次访问时间，请打开开关，无论权限如何，都会扫描所有文件。"了解更多"。

仅当您在标题区域中设置了 **Map** 或 **Map & Classify** 设置时，才会自动扫描添加到系统的新卷。当所有卷的设置都是“自定义”或“关闭”时，您需要为添加的每个新卷手动激活扫描。



步骤

1. 从数据分类菜单中，选择*配置*。
2. 选择一个系统，然后选择*配置*。
3. 要启用或禁用所有卷的扫描，请在所有卷上方的标题中选择映射、映射和分类或关闭。

   要启用或禁用对单个卷的扫描，请在列表中找到该卷，然后选择卷名称旁边的映射、映射和分类或关闭。

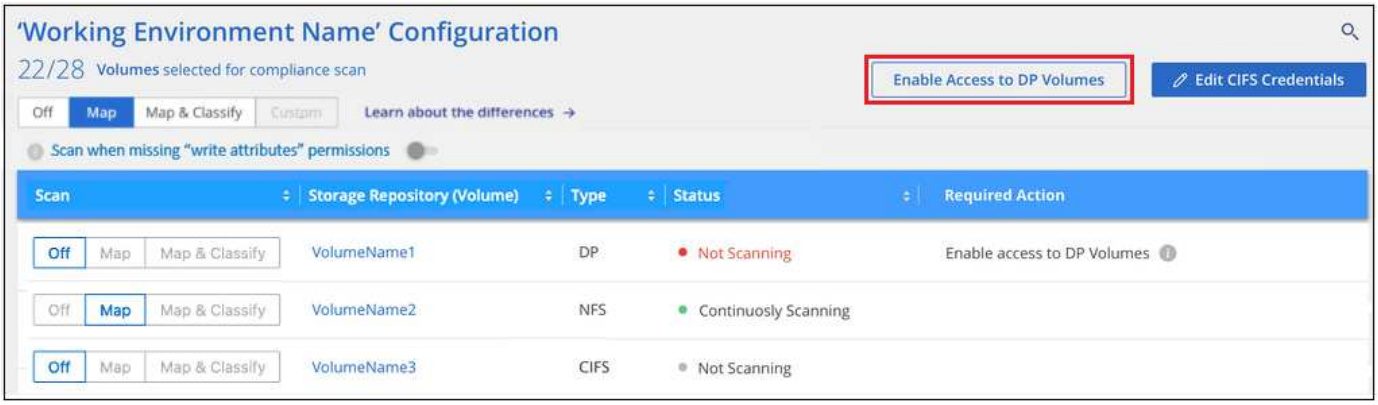
结果

当您启用扫描时，数据分类将开始扫描您在系统中选择的卷。一旦数据分类开始扫描，结果就会开始出现在合规性仪表板中。扫描完成时间取决于数据量，从几分钟到几小时不等。

扫描数据保护卷

默认情况下，不会扫描数据保护 (DP) 卷，因为它们未暴露在外部，并且数据分类无法访问它们。这些是来自FSx for ONTAP文件系统的SnapMirror操作的目标卷。

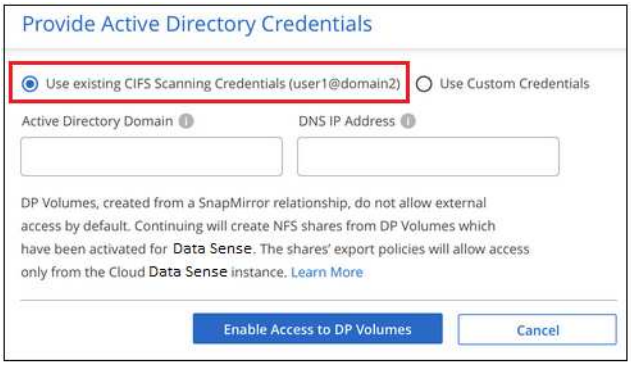
最初，卷列表将这些卷标识为_类型_ **DP**，其_状态_ 未扫描*和_所需操作_ *启用对 **DP** 卷的访问。

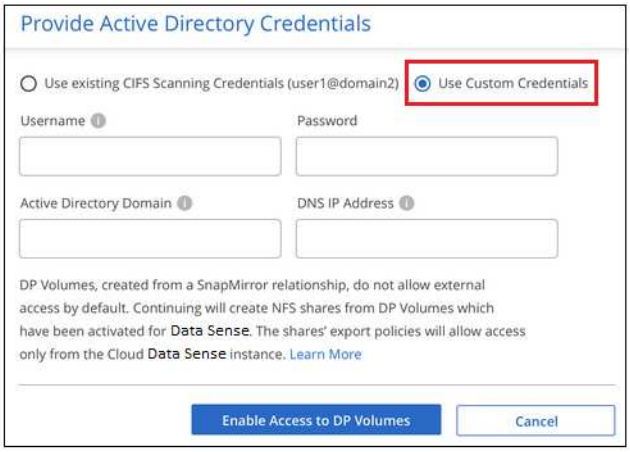


步骤

如果要扫描这些数据保护卷：

1. 从数据分类菜单中，选择*配置*。
2. 选择页面顶部的"启用对 DP 卷的访问"*。
3. 查看确认消息并再次选择*启用对 DP 卷的访问*。
   - 最初在源 FSx for ONTAP文件系统中创建为 NFS 卷的卷已启用。
   - 最初在源 FSx for ONTAP文件系统中创建为 CIFS 卷的卷要求您输入 CIFS 凭据来扫描这些 DP 卷。如果您已经输入了 Active Directory 凭据，以便数据分类可以扫描 CIFS 卷，您可以使用这些凭据，或者您可以指定一组不同的管理员凭据。





4. 激活您想要扫描的每个 DP 卷。

结果

一旦启用，数据分类将从每个激活扫描的 DP 卷创建一个 NFS 共享。共享导出策略仅允许从数据分类实例进行访问。

如果您在最初启用对 DP 卷的访问时没有 CIFS 数据保护卷，后来又添加了一些，则按钮 启用对 **CIFS DP** 的访问 将出现在配置页面的顶部。选择此按钮并添加 CIFS 凭据以启用对这些 CIFS DP 卷的访问。

Active Directory 凭据仅在第一个 CIFS DP 卷的存储 VM 中注册，因此该 SVM 上的所有 DP 卷都将被扫描。驻留在其他 SVM 上的任何卷都不会注册 Active Directory 凭据，因此不会扫描这些 DP 卷。

使用NetApp Data Classification扫描Azure NetApp Files卷

完成几个步骤即可开始使用适用于Azure NetApp Files的NetApp Data Classification。

发现要扫描的Azure NetApp Files系统Discover the Azure NetApp Files system that you want to scan

如果要扫描的Azure NetApp Files系统尚未作为系统出现在NetApp Console中，"将其添加到系统页面" 。

部署数据分类实例

"部署数据分类"如果尚未部署实例。

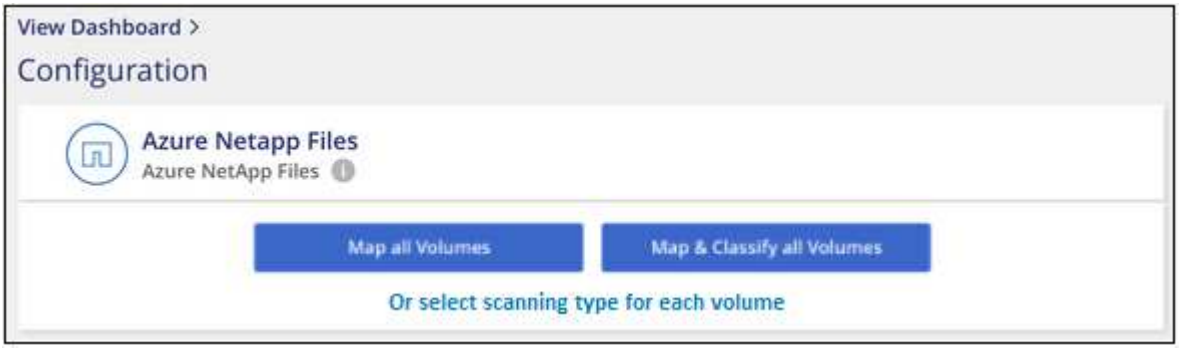
扫描Azure NetApp Files卷时，数据分类必须部署在云中，并且必须部署在与要扫描的卷相同的区域中。

*注意：*扫描Azure NetApp Files卷时，当前不支持在本地位置部署数据分类。

在您的系统中启用数据分类

您可以在Azure NetApp Files卷上启用数据分类。

1. 从数据分类菜单中，选择*配置*。



2. 选择如何扫描每个系统中的卷。"了解映射和分类扫描"：
   - 要映射所有卷，请选择*映射所有卷*。
   - 要映射和分类所有卷，请选择*映射和分类所有卷*。
   - 要自定义每个卷的扫描，请选择*或选择每个卷的扫描类型*，然后选择要映射或映射和分类的卷。

     看启用或禁用卷扫描了解详情。

3. 在确认对话框中，选择*批准*。

结果

数据分类开始扫描您在系统中选择的卷。一旦数据分类完成初始扫描，结果就会显示在合规性仪表板中。所需时间取决于数据量——可能是几分钟或几小时。您可以通过导航到配置菜单然后选择系统配置来跟踪初始扫描的进

度。数据分类显示每次扫描的进度条。您可以将鼠标悬停在进度条上，以查看相对于卷中文件总数的已扫描文件数。

- 默认情况下，如果数据分类在 CIFS 中没有写入属性权限，或者在 NFS 中没有写入权限，系统将不会扫描卷中的文件，因为数据分类无法将“上次访问时间”恢复为原始时间戳。如果您不介意上次访问时间是否重置，请选择*或为每个卷选择扫描类型*。结果页面有一个您可以启用的设置，以便数据分类可以扫描卷，而不管权限如何。
- 数据分类仅扫描卷下的一个文件共享。如果您的卷中有多个共享，则需要将这些其他共享作为共享组单独扫描。"了解此数据分类限制" 。

验证数据分类是否有权访问卷

通过检查网络、安全组和导出策略，确保数据分类可以访问卷。您需要为数据分类提供 CIFS 凭据，以便它可以访问 CIFS 卷。

对于Azure NetApp Files，数据分类只能扫描与控制台位于同一区域的卷。

清单

- 确保数据分类实例与包含Azure NetApp Files卷的每个网络之间存在网络连接。
- 确保以下端口对数据分类实例开放：
  - 对于 NFS – 端口 111 和 2049。
  - 对于 CIFS – 端口 139 和 445。
- 确保 NFS 卷导出策略包含数据分类实例的 IP 地址，以便它可以访问每个卷上的数据。

步骤

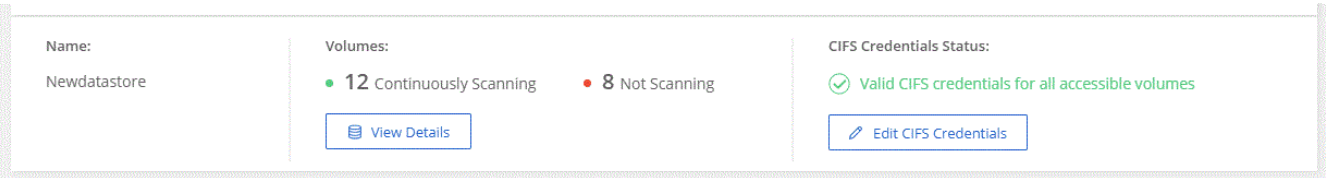
1. 从数据分类菜单中，选择*配置*。

   a. 如果您使用 CIFS（SMB），请确保 Active Directory 凭据正确。对于每个系统，选择*编辑 CIFS 凭据*，然后输入数据分类访问系统上的 CIFS 卷所需的用户名和密码。

   凭据可以是只读的；提供管理员凭据可确保数据分类可以读取任何需要提升权限的数据。凭证存储在数据分类实例上。

   如果您想确保文件的“上次访问时间”不会因数据分类扫描而改变，建议用户在 CIFS 中具有写入属性权限或在 NFS 中具有写入权限。如果可能，请将 Active Directory 用户配置为组织中具有所有文件权限的父组的一部分。

   输入凭据后，您应该会看到一条消息，表明所有 CIFS 卷均已成功验证。



2. 在配置页面上，选择*查看详细信息*以查看每个 CIFS 和 NFS 卷的状态。如有必要，请纠正任何错误，例如网络连接问题。

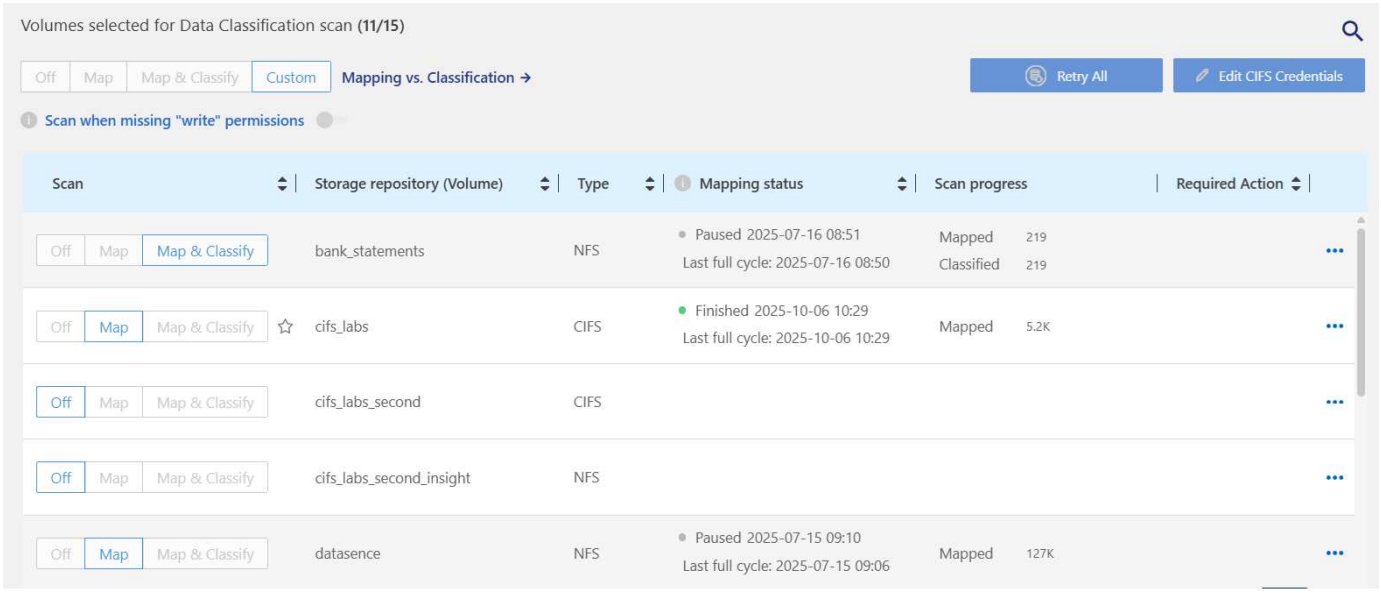
启用或禁用卷扫描

您可以随时从配置页面启动或停止任何系统上的扫描。您还可以将扫描从仅映射扫描切换到映射和分类扫描，反之亦然。建议您扫描系统中的所有卷。

仅当您在标题区域中选择了 **Map** 或 **Map & Classify** 设置时，才会自动扫描添加到系统的新卷。当在标题区域设置为*自定义*或*关闭*时，您需要在系统中添加的每个新卷上激活映射和/或完整扫描。

页面顶部的“缺少“写入”权限时扫描”开关默认处于禁用状态。这意味着，如果数据分类在 CIFS 中没有写属性权限或在 NFS 中没有写权限，系统将不会扫描文件，因为数据分类无法将“上次访问时间”恢复为原始时间戳。如果您不介意是否重置上次访问时间，请打开开关，无论权限如何，都会扫描所有文件。"了解更多"。

仅当您在标题区域中设置了 **Map** 或 **Map & Classify** 设置时，才会自动扫描添加到系统的新卷。当所有卷的设置都是“自定义”或“关闭”时，您需要为添加的每个新卷手动激活扫描。



步骤

1. 从数据分类菜单中，选择*配置*。
2. 选择一个系统，然后选择*配置*。
3. 要启用或禁用所有卷的扫描，请在所有卷上方的标题中选择映射、映射和分类或关闭。

   要启用或禁用对单个卷的扫描，请在列表中找到该卷，然后选择卷名称旁边的映射、映射和分类或关闭。

结果

当您启用扫描时，数据分类将开始扫描您在系统中选择的卷。一旦数据分类开始扫描，结果就会开始出现在合规性仪表板中。扫描完成时间取决于数据量，从几分钟到几小时不等。

使用NetApp Data Classification扫描Cloud Volumes ONTAP和本地ONTAP卷

完成几个步骤即可开始使用NetApp Data Classification扫描您的Cloud Volumes ONTAP和本地ONTAP卷。

前提条件

在启用数据分类之前，请确保您具有受支持的配置。

- 如果您正在扫描可通过互联网访问的Cloud Volumes ONTAP和本地ONTAP系统，您可以"在云中部署数据分类"或者"在可以访问互联网的本地位置"。
- 如果您要扫描安装在没有互联网访问的暗站中的本地ONTAP系统，则需要"在没有互联网访问的同一本地位置部署数据分类"。这要求将控制台代理部署在同一本地位置。

验证数据分类是否有权访问卷

通过检查网络、安全组和导出策略，确保数据分类可以访问卷。您需要向数据分类提供 CIFS 凭据，以便它可以访问 CIFS 卷。

清单

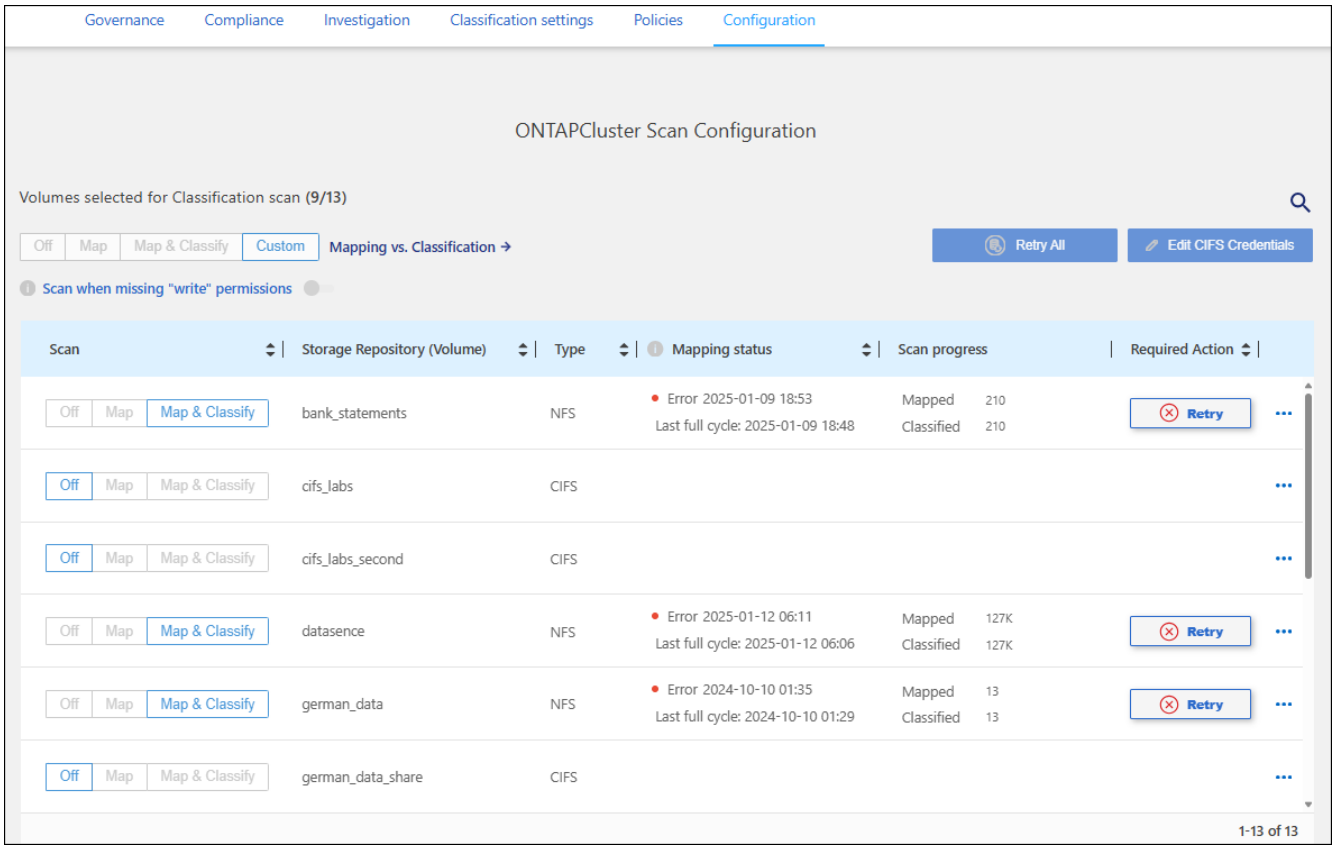
- 确保数据分类实例与包含Cloud Volumes ONTAP或本地ONTAP集群的卷的每个网络之间存在网络连接。
- 确保Cloud Volumes ONTAP的安全组允许来自数据分类实例的入站流量。

  您可以为来自数据分类实例的 IP 地址的流量打开安全组，也可以为来自虚拟网络内部的所有流量打开安全组。

- 确保 NFS 卷导出策略包含数据分类实例的 IP 地址，以便它可以访问每个卷上的数据。

步骤

1. 从数据分类菜单中，选择*配置*。



2. 如果您使用 CIFS，请向数据分类提供 Active Directory 凭据，以便它可以扫描 CIFS 卷。对于每个系统，选择*编辑 CIFS 凭据*并输入数据分类访问系统上的 CIFS 卷所需的用户名和密码。

   凭据可以是只读的，但提供管理员凭据可确保数据分类可以读取任何需要提升权限的数据。凭证存储在数据分类实例上。

   如果您想确保文件的“上次访问时间”不会因数据分类扫描而改变，建议用户在 CIFS 中具有写入属性权限或在 NFS 中具有写入权限。如果可能，请将 Active Directory 用户配置为组织中具有所有文件权限的父组的一部分。

   如果您正确输入了凭据，则会出现一条消息确认所有 CIFS 卷均已成功验证。

3. 在配置页面上，选择*配置*以查看每个 CIFS 和 NFS 卷的状态并纠正任何错误。

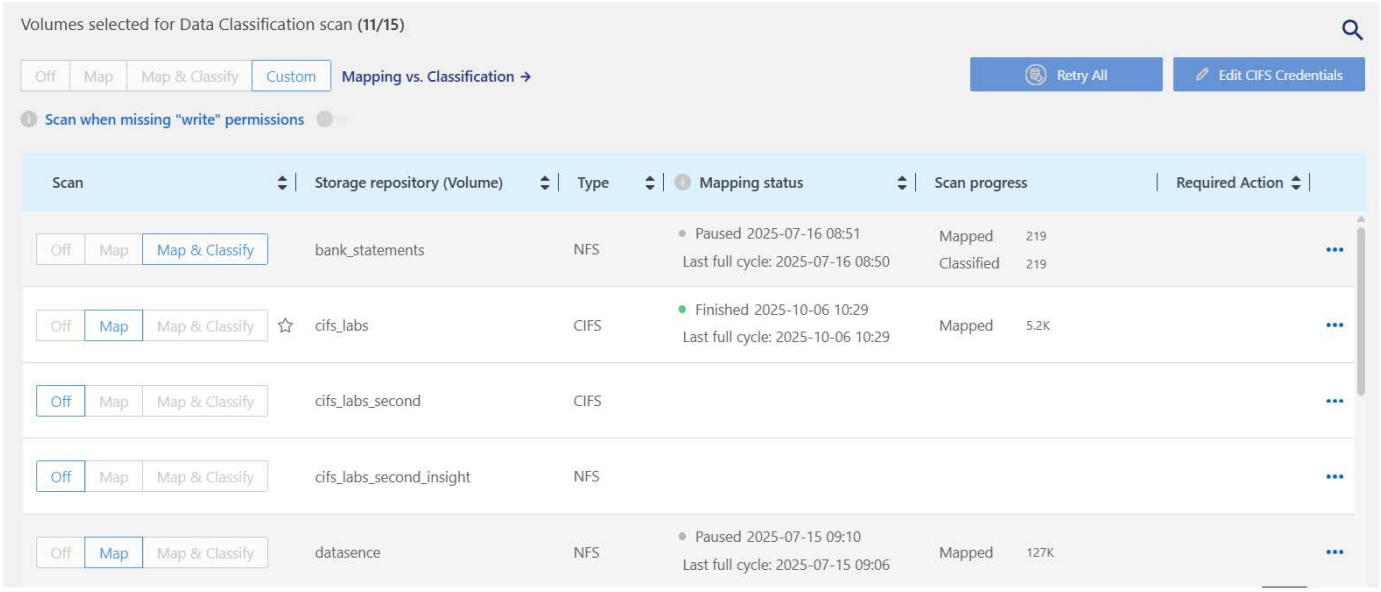
启用或禁用卷扫描

您可以随时从配置页面启动或停止任何系统上的扫描。您还可以将扫描从仅映射扫描切换到映射和分类扫描，反之亦然。建议您扫描系统中的所有卷。

仅当您在标题区域中选择了 **Map** 或 **Map & Classify** 设置时，才会自动扫描添加到系统的新卷。当在标题区域设置为*自定义*或*关闭*时，您需要在系统中添加的每个新卷上激活映射和/或完整扫描。

页面顶部的“缺少“写入”权限时扫描”开关默认处于禁用状态。这意味着，如果数据分类在 CIFS 中没有写属性权限或在 NFS 中没有写权限，系统将不会扫描文件，因为数据分类无法将“上次访问时间”恢复为原始时间戳。如果您不介意是否重置上次访问时间，请打开开关，无论权限如何，都会扫描所有文件。"了解更多"。

仅当您在标题区域中设置了 **Map** 或 **Map & Classify** 设置时，才会自动扫描添加到系统的新卷。当所有卷的设置都是“自定义”或“关闭”时，您需要为添加的每个新卷手动激活扫描。



步骤

1. 从数据分类菜单中，选择*配置*。
2. 选择一个系统，然后选择*配置*。

3. 要启用或禁用所有卷的扫描，请在所有卷上方的标题中选择映射、映射和分类或关闭。

   要启用或禁用对单个卷的扫描，请在列表中找到该卷，然后选择卷名称旁边的映射、映射和分类或关闭。

结果

当您启用扫描时，数据分类将开始扫描您在系统中选择的卷。一旦数据分类开始扫描，结果就会开始出现在合规性仪表板中。扫描完成时间取决于数据量，从几分钟到几小时不等。

数据分类仅扫描卷下的一个文件共享。如果您的卷中有多个共享，则需要将这些其他共享作为共享组单独扫描。"查看有关此数据分类限制的更多详细信息" 。

使用NetApp Data Classification

完成几个步骤即可开始使用NetApp Data Classification扫描数据库模式。

审查先决条件

在启用数据分类之前，请查看以下先决条件，以确保您具有受支持的配置。

支持的数据库

数据分类可以扫描以下数据库中的模式：

- 亚马逊关系数据库服务 (Amazon RDS)
- MongoDB
- MySQL
- Oracle
- PostgreSQL
- SAP HANA
- SQL 服务器 (MSSQL)

数据库中必须启用统计信息收集功能。

数据库要求

任何与数据分类实例连接的数据库都可以被扫描，无论它托管在何处。您只需要以下信息即可连接到数据库：

- IP 地址或主机名
- 端口
- 服务名称（仅用于访问 Oracle 数据库）
- 允许读取架构的凭证

  选择用户名和密码时，务必选择对要扫描的所有模式和表具有完全读取权限的用户名和密码。我们建议您为数据分类系统创建一个具有所有必需权限的专用用户。

对于 MongoDB，需要只读管理员角色。

部署数据分类实例

如果尚未部署实例，则部署数据分类。

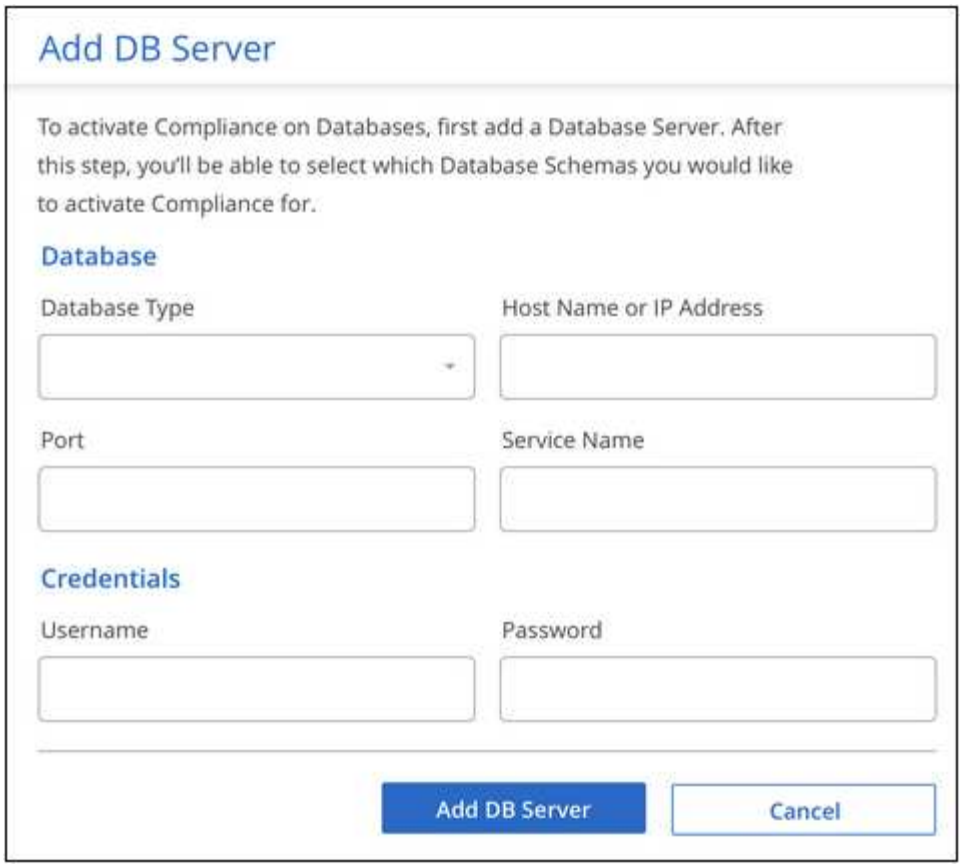
如果您正在扫描可通过互联网访问的数据库模式，您可以"在云中部署数据分类"或者"在可以访问互联网的本地位置部署数据分类"。

如果您正在扫描安装在没有互联网访问的暗网中的数据库模式，则需要"在没有互联网访问的同一本地位置部署数据分类"。这还要求将控制台代理部署在同一本地位置。

添加数据库服务器

添加架构所在的数据库服务器。

1. 从数据分类菜单中，选择*配置*。
2. 在配置页面中，选择*添加系统* > 添加数据库服务器。
3. 输入所需信息以识别数据库服务器。
    a. 选择数据库类型。
    b. 输入端口和主机名或 IP 地址以连接到数据库。
    c. 对于 Oracle 数据库，输入服务名称。
    d. 输入凭据以便数据分类可以访问服务器。
    e. 选择*添加数据库服务器*。



该数据库已添加到系统列表中。

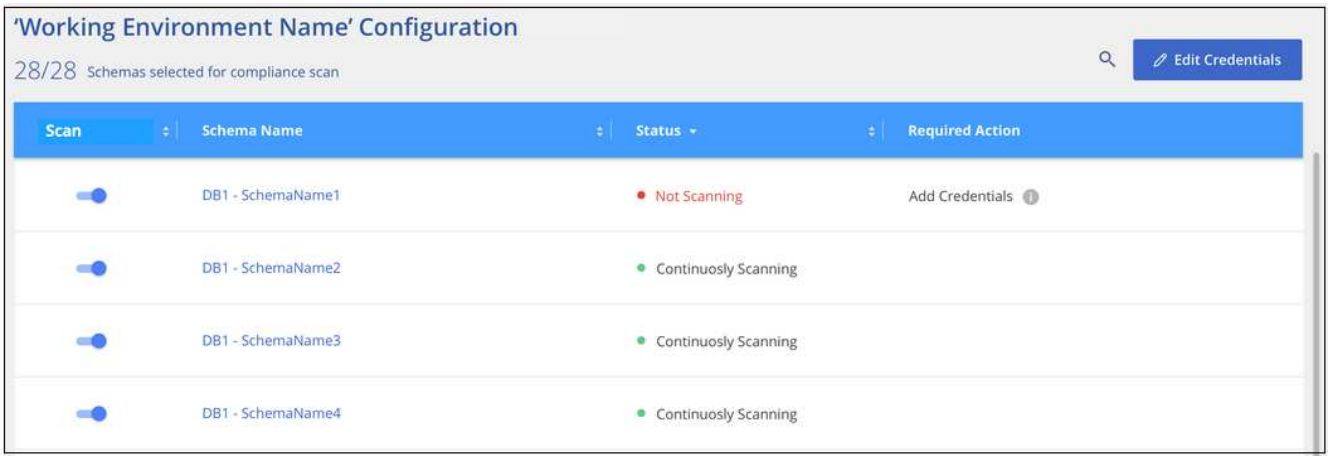
启用和禁用数据库模式扫描

您可以随时停止或开始对您的模式进行全面扫描。

没有选项可以选择仅映射数据库模式的扫描。

1. 在配置页面中，选择要配置的数据库的*配置*按钮。

2. 通过向右移动滑块来选择要扫描的模式。



结果

数据分类开始扫描您启用的数据库模式。您可以通过导航到配置菜单然后选择系统配置来跟踪初始扫描的进度。每次扫描的进度都显示为进度条。您还可以将鼠标悬停在进度条上，查看相对于卷中文件总数的已扫描文件数。如果有任何错误，它们将出现在“状态”列中，并显示修复错误所需的操作。

数据分类每天扫描您的数据库一次；数据库不像其他数据源那样被连续扫描。

使用NetApp Data Classification扫描Google Cloud NetApp Volumes

NetApp Data Classification支持Google Cloud NetApp Volumes作为一个系统。了解如何扫描您的Google Cloud NetApp Volumes系统。

发现您要扫描的Google Cloud NetApp Volumes系统

如果您要扫描的Google Cloud NetApp Volumes系统尚未作为系统出现在NetApp Console中，"将其添加到系统页面" 。

部署数据分类实例

"部署数据分类"如果尚未部署实例。

扫描Google Cloud NetApp Volumes时，数据分类必须部署在云中，并且必须部署在与您要扫描的卷相同的区域。

*注意：*扫描Google Cloud NetApp Volumes时，目前不支持在本地位置部署数据分类。

在您的系统中启用数据分类

您可以在Google Cloud NetApp Volumes系统上启用数据分类。

1. 从数据分类菜单中，选择*配置*。
2. 选择如何扫描每个系统中的卷。"了解映射和分类扫描"：
   - 要映射所有卷，请选择*映射所有卷*。
   - 要映射和分类所有卷，请选择*映射和分类所有卷*。
   - 要自定义每个卷的扫描，请选择*或选择每个卷的扫描类型*，然后选择要映射和/或分类的卷。

看启用和禁用卷上的扫描了解详情。

3. 在确认对话框中，选择*批准*。

结果

数据分类开始扫描您在系统中选择的卷。一旦数据分类完成初始扫描，结果就会显示在合规性仪表板中。所需时间取决于数据量：几分钟到几个小时。您可以在配置菜单的系统配置部分跟踪初始扫描的进度。数据分类显示每次扫描的进度条。您还可以将鼠标悬停在进度条上，查看相对于卷中总文件数的已扫描文件数。

- 默认情况下，如果数据分类在 CIFS 中没有写入属性权限，或者在 NFS 中没有写入权限，系统将不会扫描卷中的文件，因为数据分类无法将“上次访问时间”恢复为原始时间戳。如果您不介意上次访问时间是否重置，请选择*或为每个卷选择扫描类型*。结果页面有一个您可以启用的设置，以便数据分类可以扫描卷，而不管权限如何。
- 数据分类仅扫描卷下的一个文件共享。如果您的卷中有多个共享，则需要将这些其他共享作为共享组单独扫描。"了解此数据分类限制"。

验证数据分类是否有权访问卷

通过检查您的网络、安全组和导出策略，确保数据分类可以访问卷。对于 CIFS 卷，您需要为数据分类提供 CIFS 凭据。

对于Google Cloud NetApp Volumes，数据分类只能扫描与控制台位于同一区域的卷。

清单

- 确保数据分类实例与包含Google Cloud NetApp Volumes的每个网络之间存在网络连接。
- 确保以下端口对数据分类实例开放：
  - 对于 NFS – 端口 111 和 2049。
  - 对于 CIFS – 端口 139 和 445。
- 确保 NFS 卷导出策略包含数据分类实例的 IP 地址，以便它可以访问每个卷上的数据。

步骤

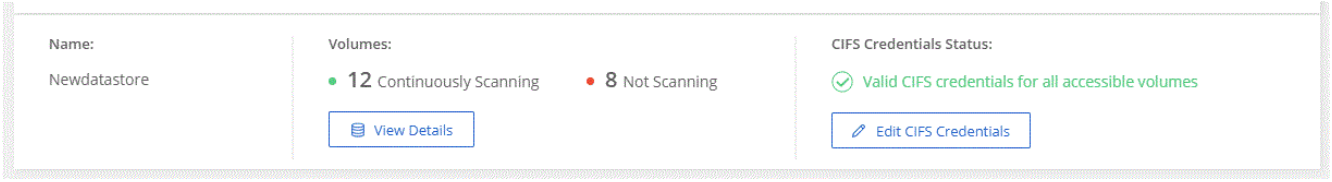
1. 从数据分类菜单中，选择*配置*。

   a. 如果您使用 CIFS（SMB），请确保 Active Directory 凭据正确。对于每个系统，选择*编辑 CIFS 凭据*，然后输入数据分类访问系统上的 CIFS 卷所需的用户名和密码。

      凭据可以是只读的，但提供管理员凭据可确保数据分类可以读取任何需要提升权限的数据。凭证存储在数据分类实例上。

      如果您想确保文件的“上次访问时间”不会因数据分类扫描而改变，建议用户在 CIFS 中具有写入属性权限或在 NFS 中具有写入权限。如果可能，请将 Active Directory 用户配置为组织中具有所有文件权限的父组的一部分。

      输入凭据后，您应该会看到一条消息，表明所有 CIFS 卷均已成功验证。



2. 在配置页面上，选择*查看详细信息*以查看每个 CIFS 和 NFS 卷的状态并纠正任何错误。

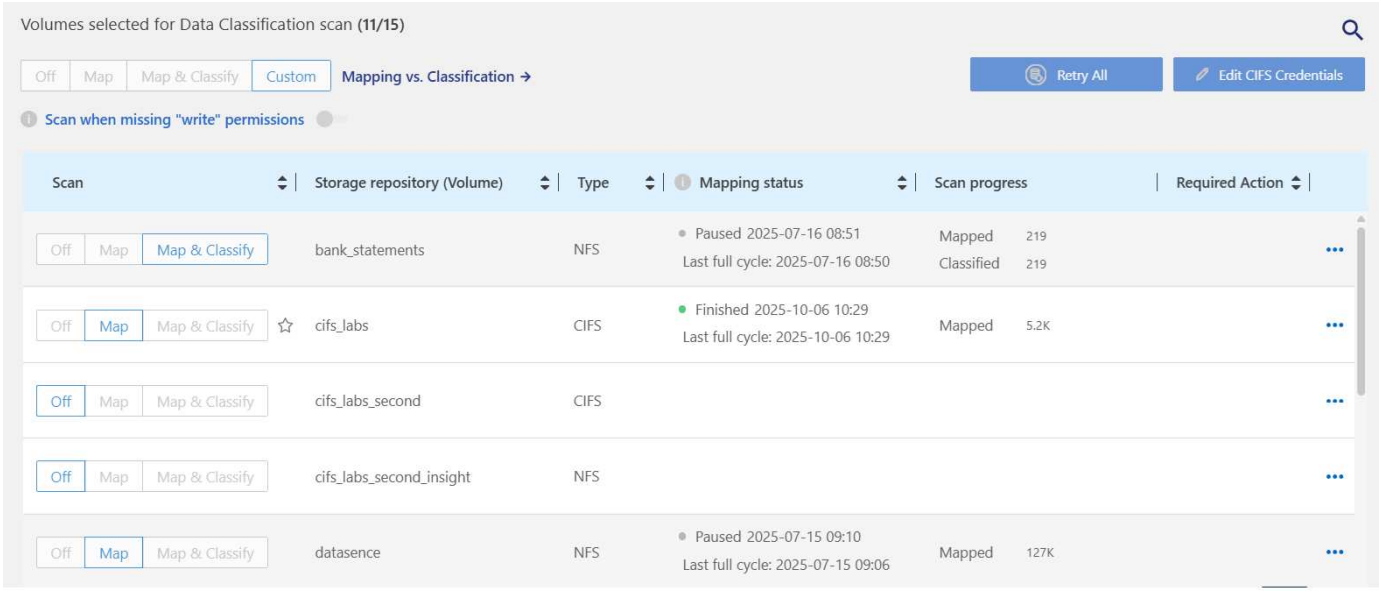
启用和禁用卷上的扫描

您可以随时从配置页面启动或停止任何系统上的扫描。您还可以将扫描从仅映射扫描切换到映射和分类扫描，反之亦然。建议您扫描系统中的所有卷。

仅当您在标题区域中选择了 **Map** 或 **Map & Classify** 设置时，才会自动扫描添加到系统的新卷。当在标题区域设置为*自定义*或*关闭*时，您需要在系统中添加的每个新卷上激活映射和/或完整扫描。

页面顶部的“缺少“写入”权限时扫描”开关默认处于禁用状态。这意味着，如果数据分类在 CIFS 中没有写属性权限或在 NFS 中没有写权限，系统将不会扫描文件，因为数据分类无法将“上次访问时间”恢复为原始时间戳。如果您不介意是否重置上次访问时间，请打开开关，无论权限如何，都会扫描所有文件。"了解更多"。

仅当您在标题区域中设置了 **Map** 或 **Map & Classify** 设置时，才会自动扫描添加到系统的新卷。当所有卷的设置都是“自定义”或“关闭”时，您需要为添加的每个新卷手动激活扫描。



步骤

1. 从数据分类菜单中，选择*配置*。
2. 选择一个系统，然后选择*配置*。
3. 要启用或禁用所有卷的扫描，请在所有卷上方的标题中选择映射、映射和分类或关闭。

   要启用或禁用对单个卷的扫描，请在列表中找到该卷，然后选择卷名称旁边的映射、映射和分类或关闭。

结果

当您启用扫描时，数据分类将开始扫描您在系统中选择的卷。一旦数据分类开始扫描，结果就会开始出现在合规性仪表板中。扫描完成时间取决于数据量，从几分钟到几小时不等。

使用NetApp Data Classification扫描文件共享

要扫描文件共享，您必须首先在NetApp Data Classification中创建一个文件共享组。文件共享组适用于本地或云中托管的 NFS 或 CIFS (SMB) 共享。

数据分类核心版本不支持扫描非NetApp文件共享的数据。

前提条件

在启用数据分类之前，请查看以下先决条件，以确保您具有受支持的配置。

- 共享可以托管在任何地方，包括云端或本地。可以将旧版NetApp 7-模式存储系统中的 CIFS 共享扫描为文件共享。
  - 数据分类无法从 7 模式系统中提取权限或“上次访问时间”。
  - 由于某些 Linux 版本和 7-模式系统上的 CIFS 共享之间存在已知问题，因此您必须将共享配置为仅使用启用了 NTLM 身份验证的 SMBv1。
- 数据分类实例和共享之间需要有网络连接。
- 您可以将 DFS（分布式文件系统）共享添加为常规 CIFS 共享。由于数据分类不知道共享是基于组合为单个 CIFS 共享的多个服务器/卷构建的，因此当消息实际上仅适用于位于不同服务器/卷上的一个文件夹/共享时，您可能会收到有关共享的权限或连接错误。
- 对于 CIFS (SMB) 共享，请确保您拥有可提供共享读取访问权限的 Active Directory 凭据。如果数据分类需要扫描任何需要提升权限的数据，则最好使用管理员凭据。

  如果您想确保文件的“上次访问时间”不会因数据分类扫描而改变，建议用户在 CIFS 中具有写入属性权限或在 NFS 中具有写入权限。如果可能，请将 Active Directory 用户配置为组织中具有所有文件权限的父组的一部分。

- 组中的所有 CIFS 文件共享必须使用相同的 Active Directory 凭据。
- 您可以混合使用 NFS 和 CIFS（使用 Kerberos 或 NTLM）共享。您必须单独将共享添加到组中。也就是说，您必须完成该过程两次 - 每个协议一次。
  - 您不能创建混合 CIFS 身份验证类型（Kerberos 和 NTLM）的文件共享组。
- 如果您使用带有 Kerberos 身份验证的 CIFS，请确保数据分类可以访问所提供的 IP 地址。如果 IP 地址无法访问，则无法添加文件共享。

创建文件共享组

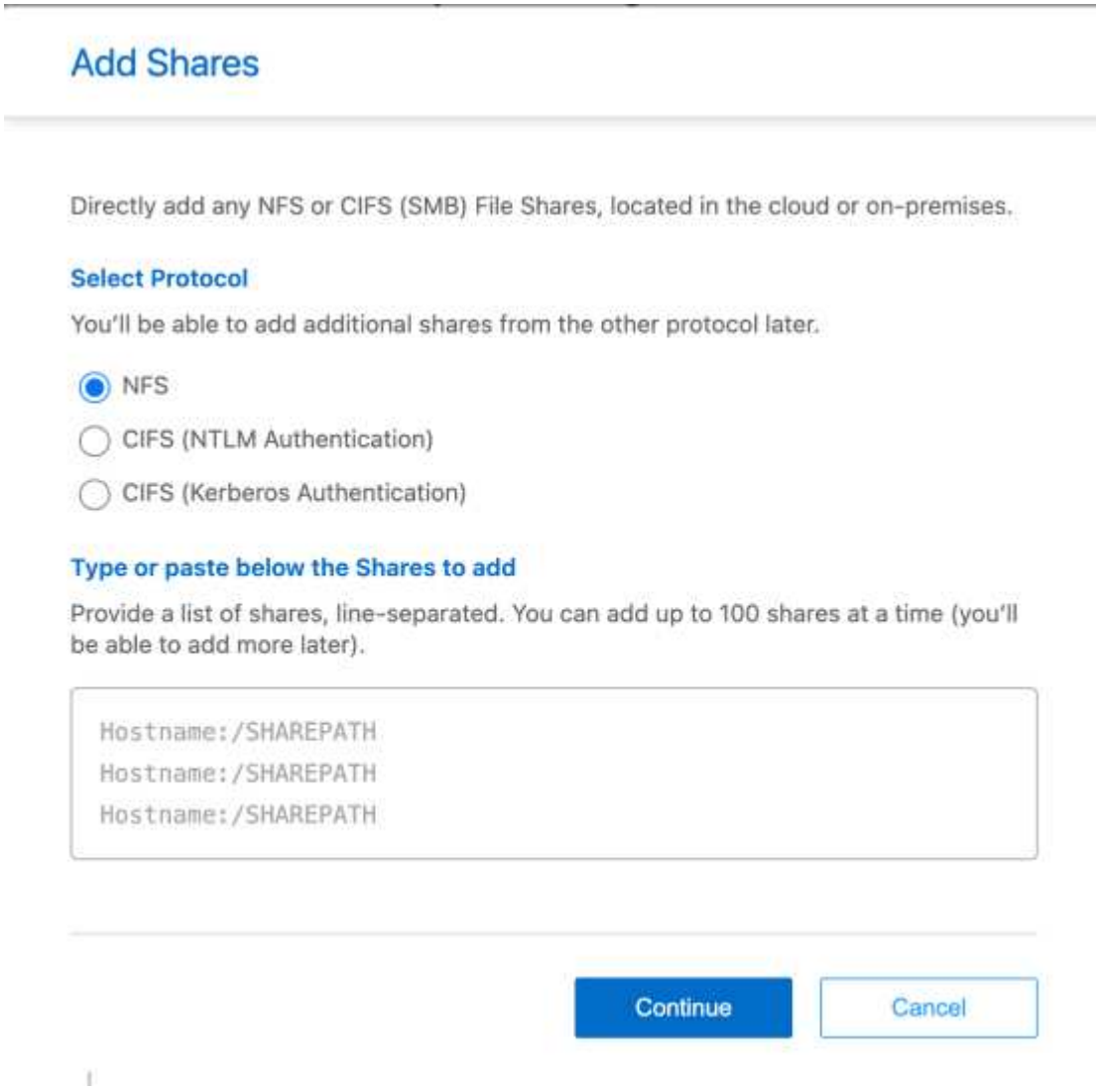
将文件共享添加到组时，必须使用格式 `<host_name>:/<share_path>`。

您可以单独添加文件共享，也可以输入要扫描的文件共享的行分隔列表。您一次最多可以添加 100 股。

步骤

1. 从数据分类菜单中，选择*配置*。
2. 从配置页面中，选择*添加系统*>*添加文件共享组*。

3. 在添加文件共享组对话框中，输入共享组的名称，然后选择*继续*。
4. 选择要添加的文件共享的协议。



   a. 如果您要添加具有 NTLM 身份验证的 CIFS 共享，请输入 Active Directory 凭据以访问 CIFS 卷。尽管支持只读凭据，但建议您使用管理员凭据提供完全访问权限。选择保存。
5. 添加要扫描的文件共享（每行一个文件共享）。然后选择继续。
6. 确认对话框显示已添加的共享数量。

   如果对话框列出了任何无法添加的共享，请捕获此信息以便解决问题。如果问题与命名约定有关，您可以使用更正的名称重新添加共享。

7. 配置卷上的扫描：
   - 要对文件共享启用仅映射扫描，请选择*映射*。
   - 要对文件共享启用完整扫描，请选择*Map & Classify*。
   - 要禁用文件共享上的扫描，请选择“关闭”。

页面顶部的“缺少“写入属性”权限时扫描”开关默认处于禁用状态。这意味着，如果数据分类在 CIFS 中没有写属性权限或在 NFS 中没有写权限，系统将不会扫描文件，因为数据分类无法将“上次访问时间”恢复为原始时间戳。+ 如果将*缺少“写入属性”权限时扫描*切换为*开*，则扫描将重置上次访问时间并扫描所有文件，而不管权限如何。+ 要了解有关上次访问时间戳的更多信息，请参阅"从数据分类中的数据源收集的元数据"。

结果

数据分类开始扫描您添加的文件共享中的文件。你可以跟踪扫描进度并在仪表板中查看扫描结果。

如果对于使用 Kerberos 身份验证的 CIFS 配置的扫描未成功完成，请检查配置选项卡中是否存在错误。

编辑文件共享组

创建文件共享组后，您可以编辑 CIFS 协议或添加和删除文件共享。

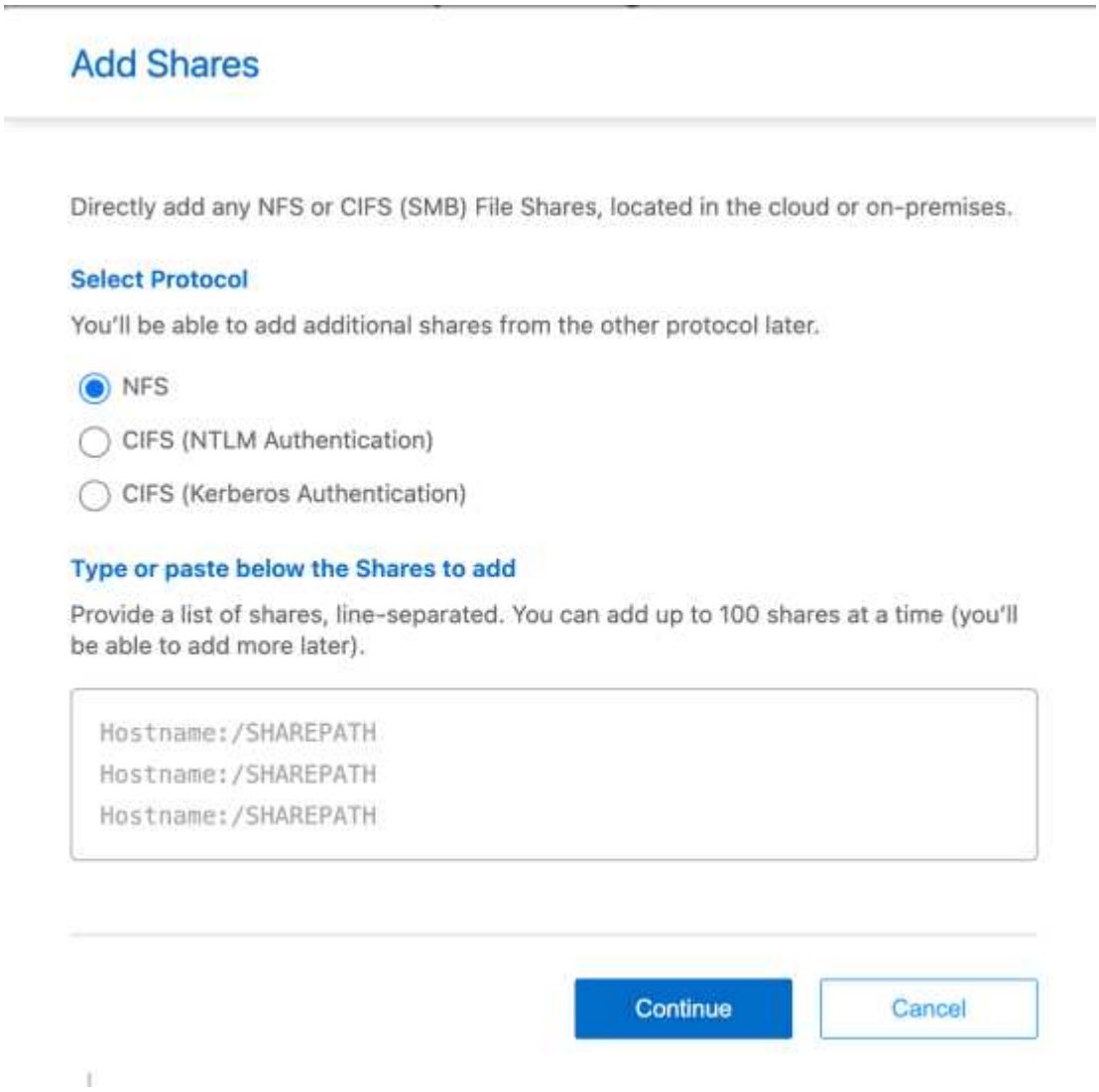
编辑 **CIFS** 协议配置

1. 从数据分类菜单中，选择*配置*。
2. 从配置页面中，选择要修改的文件共享组。
3. 选择编辑 **CIFS** 凭证。

4. 选择身份验证方法：**NTLM** 或 **Kerberos**。
5. 输入 Active Directory 用户名和密码。
6. 选择保存以完成该过程。

将文件共享添加到扫描

1. 从数据分类菜单中，选择*配置*。
2. 从配置页面中，选择要修改的文件共享组。
3. 选择 **+** 添加共享。
4. 选择要添加的文件共享的协议。



如果您要将文件共享添加到已配置的协议，则无需进行任何更改。

如果您要使用第二种协议添加文件共享，请确保您已正确配置身份验证，详情请见"前提条件"。

5. 使用以下格式添加要扫描的文件共享（每行一个文件共享） `<host_name>:/<share_path>` 。
6. 选择继续以完成添加文件共享。

从扫描中删除文件共享

1. 从数据分类菜单中，选择*配置*。
2. 选择要从中删除文件共享的系统。
3. 选择*配置*。
4. 在配置页面中，选择操作 ... 对于要删除的文件共享。
5. 从操作菜单中，选择*删除共享*。

跟踪扫描进度

您可以跟踪初始扫描的进度。

1. 选择配置菜单。
2. 选择系统配置。
3. 对于存储库，检查扫描进度列以查看其状态。

使用NetApp Data Classification扫描StorageGRID数据

完成几个步骤即可直接使用NetApp Data Classification开始扫描StorageGRID内的数据。

查看StorageGRID要求

在启用数据分类之前，请查看以下先决条件，以确保您具有受支持的配置。

- 您需要有端点 URL 才能连接对象存储服务。
- 您需要拥有来自StorageGRID的访问密钥和密钥，以便数据分类可以访问存储桶。

部署数据分类实例

如果尚未部署实例，则部署数据分类。

如果您正在扫描可通过互联网访问的StorageGRID数据，您可以"在云中部署数据分类"或者"在可以访问互联网的本地位置部署数据分类"。

如果您要扫描安装在没有互联网访问的暗站中的StorageGRID数据，则需要"在没有互联网访问的同一本地位置部署数据分类"。这还要求将控制台代理部署在同一本地位置。

将StorageGRID服务添加到数据分类

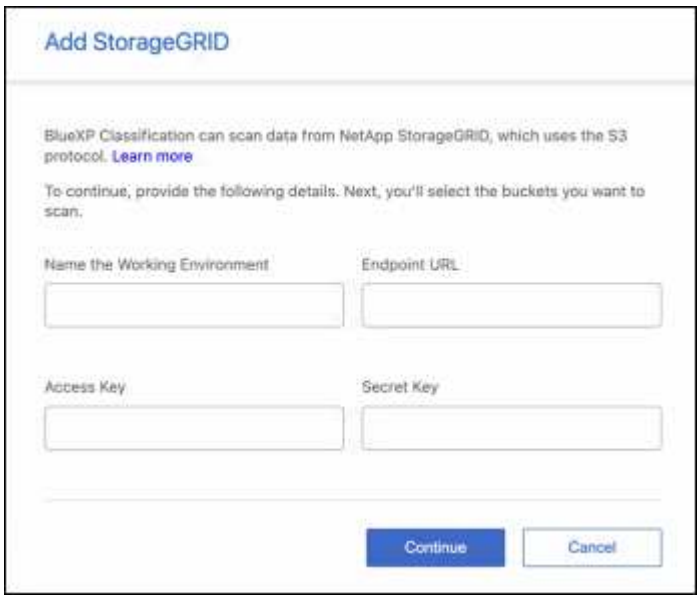
添加StorageGRID服务。

步骤

1. 从数据分类菜单中，选择*配置*选项。
2. 在配置页面中，选择“添加系统”>“添加StorageGRID”。
3. 在添加StorageGRID服务对话框中，输入StorageGRID服务的详细信息并选择*继续*。
    a. 输入您想要使用的系统名称。此名称应反映您要连接的StorageGRID服务的名称。

b. 输入 Endpoint URL 以访问对象存储服务。

c. 输入访问密钥和密钥，以便数据分类可以访问StorageGRID中的存储桶。



结果

StorageGRID已添加到系统列表中。

启用和禁用**StorageGRID**桶上的扫描

在StorageGRID上启用数据分类后，下一步是配置要扫描的存储桶。数据分类发现这些存储桶并将它们显示在您创建的系统中。

步骤

1. 在配置页面中，找到StorageGRID系统。
2. 在StorageGRID系统图块上，选择 配置。
3. 完成以下步骤之一来启用或禁用扫描：
   - 要对存储桶启用仅映射扫描，请选择*Map*。
   - 要对存储桶启用完整扫描，请选择*Map & Classify*。
   - 要禁用对存储桶的扫描，请选择“关闭”。

结果

数据分类开始扫描您启用的存储桶。您可以通过导航到配置菜单然后选择系统配置来跟踪初始扫描的进度。每次扫描的进度都显示为进度条。您还可以将鼠标悬停在进度条上，查看相对于卷中总文件数的已扫描文件数。如果有任何错误，它们将出现在“状态”列中，并显示修复错误所需的操作。

将您的 **Active Directory** 与**NetApp Data Classification**集成

您可以将全局 Active Directory 与NetApp Data Classification集成，以增强数据分类报告有关文件所有者以及哪些用户和组有权访问您的文件的结果。

当您设置某些数据源（如下所列）时，您需要输入 Active Directory 凭据以便数据分类扫描 CIFS 卷。此集成为数据分类提供了驻留在这些数据源中的数据的文件所有者和权限详细信息。为这些数据源输入的 Active Directory 可能与您在此处输入的全局 Active Directory 凭据不同。数据分类将在所有集成的活动目录中查找用户和权限详细信息。

此集成在数据分类的以下位置提供了附加信息：

- 您可以使用“文件所有者”"筛选"并在调查窗格中的文件元数据中查看结果。它不是包含 SID（安全标识符）的文件所有者，而是填充实际的用户名。

  您还可以查看有关文件所有者的更多详细信息：帐户名称、电子邮件地址和 SAM 帐户名称，或查看该用户拥有的项目。

- 你可以看到"完整文件权限"当您单击“查看所有权限”按钮时，每个文件和目录都会显示所有权限。
- 在"治理仪表板"，打开权限面板将显示有关您的数据的更详细的详细信息。

本地用户 SID 和来自未知域的 SID 不会转换为实际用户名。

支持的数据源

Active Directory 与数据分类的集成可以识别来自以下数据源的数据：

- 本地ONTAP系统
- Cloud Volumes ONTAP
- Azure NetApp Files
- 适用于ONTAP的 FSx

连接到您的 **Active Directory** 服务器

部署数据分类并激活数据源扫描后，您可以将数据分类与 Active Directory 集成。可以使用 DNS 服务器 IP 地址或 LDAP 服务器 IP 地址访问 Active Directory。

Active Directory 凭据可以是只读的，但提供管理员凭据可确保数据分类可以读取任何需要提升权限的数据。凭证存储在数据分类实例上。

对于 CIFS 卷/文件共享，如果您想确保文件的“上次访问时间”不会因数据分类扫描而改变，则用户应该具有写入属性权限。如果可能的话，我们建议将 Active Directory 配置的用户作为组织中对所有文件具有权限的父组的一部分。

要求

- 您必须已经为公司用户设置了 Active Directory。
- 您必须具有 Active Directory 的信息：
  - DNS 服务器 IP 地址，或多个 IP 地址

    或

  LDAP 服务器 IP 地址，或多个 IP 地址

- 访问服务器的用户名和密码
- 域名（Active Directory 名称）
- 您是否使用安全 LDAP (LDAPS)
- LDAP 服务器端口（LDAP 通常为 389，安全 LDAP 通常为 636）

- 必须打开以下端口以供数据分类实例进行出站通信：

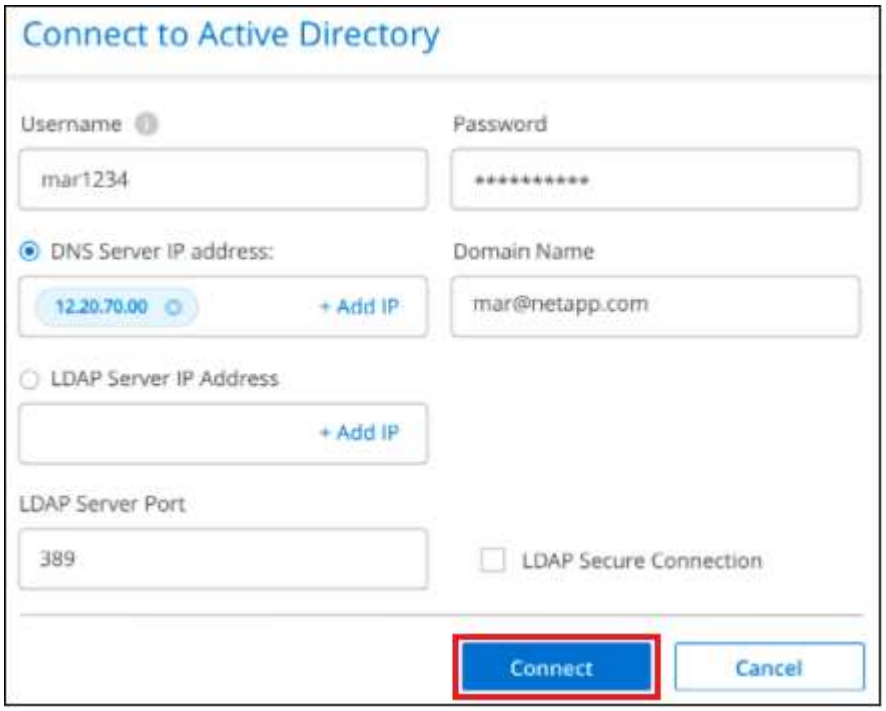
协议	端口	目标	目的
TCP 和 UDP	389	Active Directory	LDAP
TCP	636	Active Directory	基于 SSL 的 LDAP
TCP	3268	Active Directory	全局目录
TCP	3269	Active Directory	通过 SSL 的全局目录

步骤

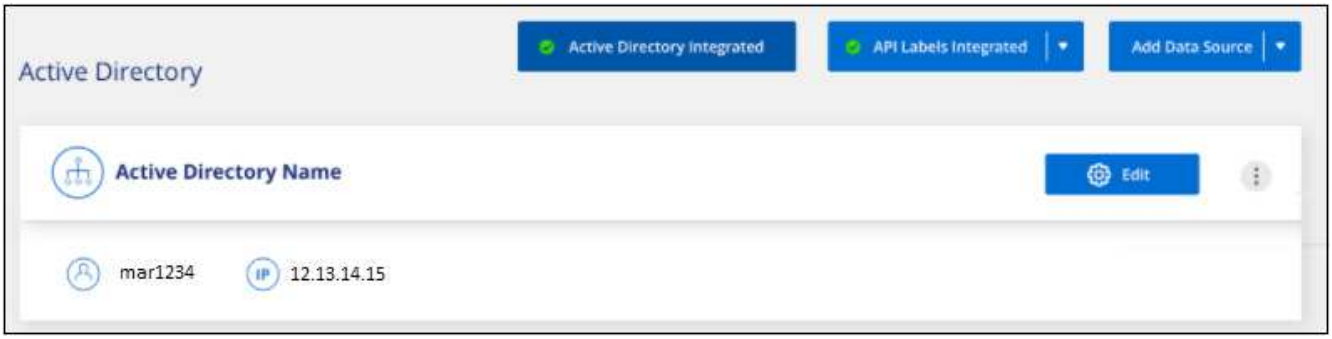
1. 在数据分类配置页面中，单击*添加 Active Directory*。

2. 在“连接到 Active Directory”对话框中，输入 Active Directory 详细信息，然后单击“连接”。

   如果需要，您可以通过选择“添加 IP”来添加多个 IP 地址。



   数据分类集成到 Active Directory，并在配置页面中添加一个新部分。



管理您的 **Active Directory** 集成

如果您需要修改 Active Directory 集成中的任何值，请单击“编辑”按钮并进行更改。

您还可以选择 按钮，然后*删除 Active Directory*。

使用数据分类

使用NetApp Data Classification查看组织中存储的数据的治理详细信息

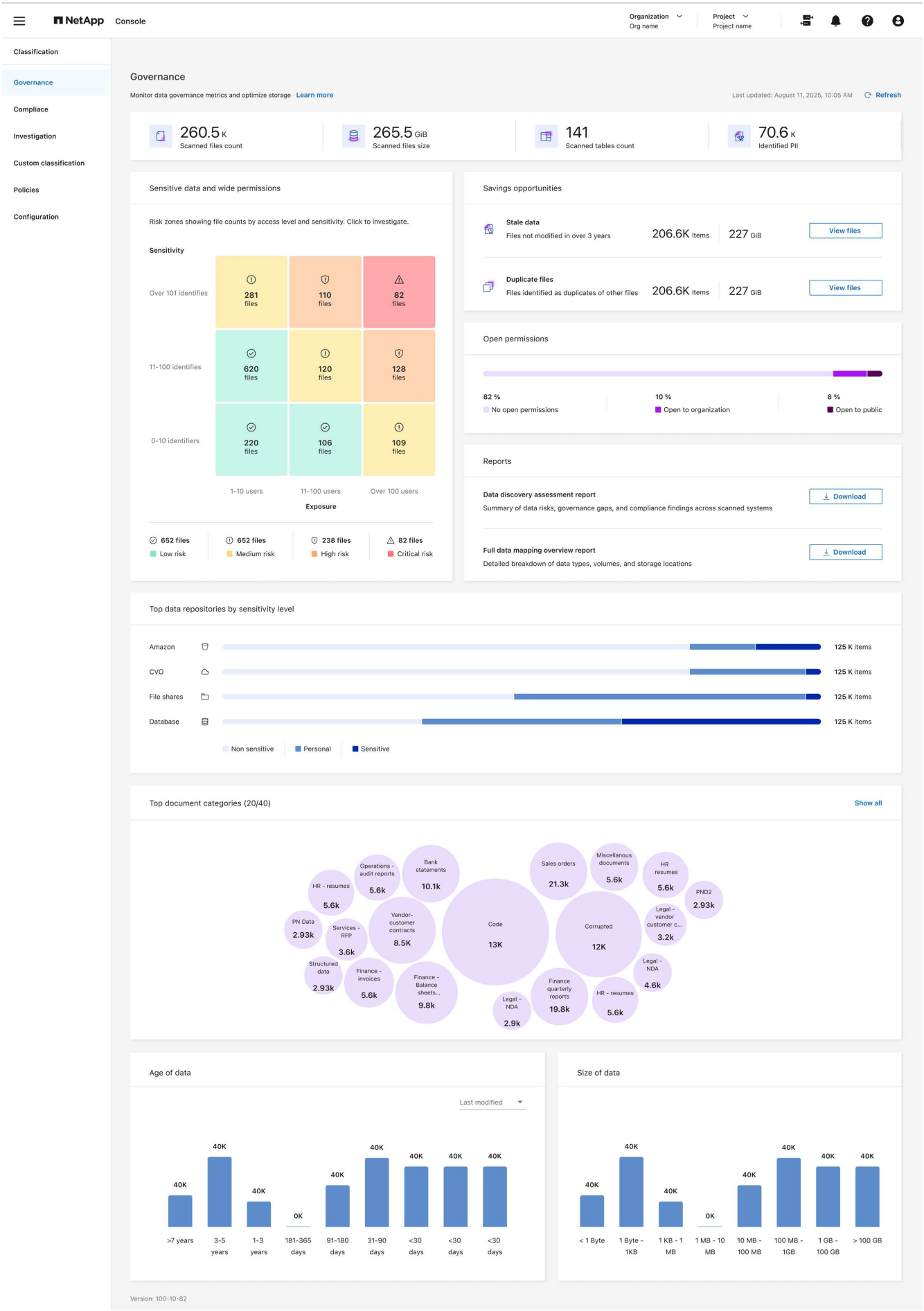
控制与组织存储资源上的数据相关的成本。 NetApp Data Classification可识别系统中陈旧数据、重复文件和超大文件的数量，以便您可以决定是否要删除某些文件或将某些文件分层到成本较低的对象存储中。

您应该从这里开始您的研究。从治理仪表板中，您可以选择一个区域进行进一步调查。

此外，如果您计划将数据从本地位置迁移到云端，则可以在移动数据之前查看数据的大小以及其中是否有任何数据包含敏感信息。

查看治理仪表板

治理仪表板提供信息，以便您可以提高效率并控制与存储在存储资源上的数据相关的成本。


NetApp Console
Organization
Org name
Project
Project name
Classification
Governance
Compliace
Investigation
Custom classification
Policies
Configuration
Governance
Monitor data governance metrics and optimize storage Learn more
Last updated: August 11, 2025, 10:05 AM
Refresh
260.5 K
Scanned files count
265.5 GiB
Scanned files size
141
Scanned tables count
70.6 K
Identified PII
Sensitive data and wide permissions
Risk zones showing file counts by access level and sensitivity. Click to investigate.
Sensitivity
Over 101 identifies
11-100 identifies
0-10 identifiers
281 files
110 files
82 files
620 files
120 files
128 files
220 files
106 files
109 files
1-10 users
11-100 users
Over 100 users
Exposure
652 files
Low risk
652 files
Medium risk
238 files
High risk
82 files
Critical risk
Savings opportunities
Stale data
Files not modified in over 3 years
206.6K Items
227 GiB
View files
Duplicate files
Files identified as duplicates of other files
206.6K Items
227 GiB
View files
Open permissions
82 %
No open permissions
10 %
Open to organization
8 %
Open to public
Reports
Data discovery assessment report
Summary of data risks, governance gaps, and compliance findings across scanned systems
Download
Full data mapping overview report
Detailed breakdown of data types, volumes, and storage locations
Download
Top data repositories by sensitivity level
Amazon
CVO
File shares
Database
125 K items
Non sensitive
Personal
Sensitive
Top document categories (20/40)
Show all
Operations - audit reports 5.6k
Bank statements 10.1k
Sales orders 21.3k
Miscellanous documents 5.6k
HR resumes 5.6k
PND2 2.93k
HR - resumes 5.6k
PN Data 2.93k
Services - RFP 3.6k
Vendor-customer contracts 8.5K
Code 13K
Corrupted 12K
Legal - vendor customer c... 3.2k
Structured data 2.93k
Finance - invoices 5.6k
Finance - Balance sheets... 9.8k
Legal - NDA 2.9k
Finance quarterly reports 19.8k
HR - resumes 5.6k
Legal - NDA 4.6k
Age of data
Last modified
40K
>7 years
40K
3-5 years
40K
1-3 years
0K
181-365 days
40K
91-180 days
40K
31-90 days
40K
<30 days
40K
<30 days
40K
<30 days
Size of data
40K
< 1 Byte
40K
1 Byte - 1KB
40K
1 KB - 1 MB
0K
1 MB - 10 MB
40K
10 MB - 100 MB
40K
100 MB - 1GB
40K
1 GB - 100 GB
40K
> 100 GB
Version: 100-10-82

步骤

1. 从NetApp Console菜单中，选择 治理 > 分类。
2. 选择*治理*。

   出现治理仪表板。

审查节省机会

_节省机会_组件显示您可以删除或分层到较便宜的对象存储的数据。 《节省机会》中的数据每 2 小时更新一次。您也可以手动更新数据。

步骤

1. 从数据分类菜单中，选择*治理*。
2. 在治理仪表板的每个节省机会图块中，选择*优化存储*以在调查页面中查看过滤的结果。要发现您应该删除或分层到较便宜的存储的任何数据，请调查_节省机会_。
   - 过期数据 - 默认情况下，如果数据上次修改时间超过 3 年，则该数据被视为过期数据。您可以[自定义过期数据的定义](task-stale-data.html)。
   - 重复文件 - 您正在扫描的数据源中其他位置重复的文件。"查看显示的重复文件类型" 。

如果您的任何数据源实现了数据分层，则可以在“陈旧数据”类别中识别已经驻留在对象存储中的旧数据。

创建数据发现评估报告

数据发现评估报告对扫描环境进行了高级分析，以显示关注区域和潜在的补救步骤。结果基于数据的映射和分类。本报告的目标是提高您对数据集三个重要方面的认识：

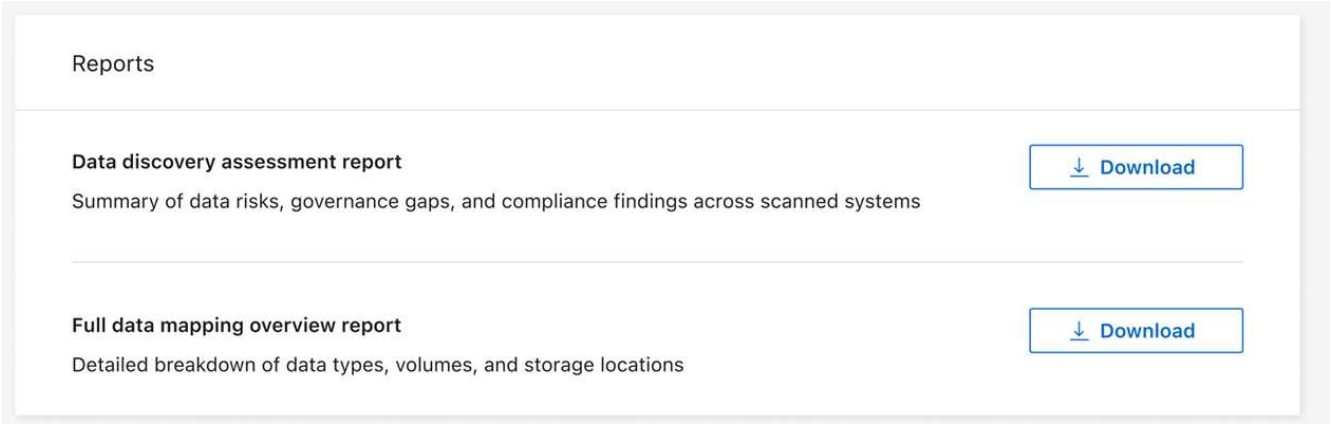
功能	描述
数据治理问题	您拥有的所有数据以及可以减少数据量以节省成本的区域的详细图片。
数据安全风险	由于访问权限广泛，您的数据可能受到内部或外部攻击的区域。
数据合规性差距	您的个人或敏感个人信息位于何处，以满足安全和 DSAR（数据主体访问请求）。

通过该报告，您可以采取以下行动：

- 通过更改保留策略或移动或删除某些数据（陈旧或重复的数据）来降低存储成本。
- 通过修改全局组管理策略来保护具有广泛权限的数据。
- 通过将 PII 移动到更安全的数据存储来保护包含个人或敏感个人信息的数据。

步骤

1. 从数据分类中，选择*治理*。
2. 在报告图块中，选择“数据发现评估报告”。



结果

数据分类会生成一份您可以查看和共享的 PDF 报告。

创建数据映射概览报告

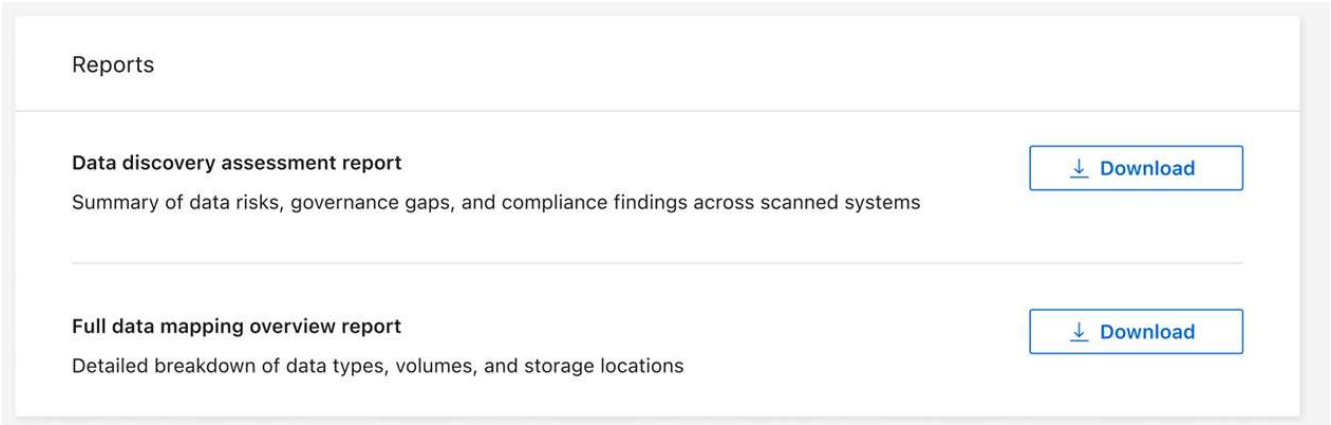
数据映射概览报告提供了存储在公司数据源中的数据的概览，以帮助您做出迁移、备份、安全和合规流程的决策。该报告总结了所有系统和数据源。它还为每个系统提供了分析。

该报告包含以下信息：

类别	描述
使用容量	对于所有系统：列出每个系统的文件数量和已用容量。对于单个系统：列出使用最多容量的文件。
数据时代	提供三个图表和图形，分别表示文件的创建时间、上次修改时间或上次访问时间。根据特定日期范围列出文件数量及其已用容量。
数据大小	列出系统中存在于特定大小范围内的文件数。

步骤

1. 从数据分类中，选择*治理*。
2. 在报告图块中，选择*完整数据映射概览报告*。



结果

数据分类会生成一份 PDF 报告，您可以根据需要查看并发送给其他组。

如果报告大于 1 MB，则 PDF 文件将保留在数据分类实例上，您将看到有关确切位置的弹出消息。当数据分类安装在您本地的 Linux 机器上或在云中部署的 Linux 机器上时，您可以直接导航到 PDF 文件。当数据分类部署在云端时，您需要使用 SSH 授权数据分类实例下载 PDF 文件。

查看按数据敏感度列出的顶级数据存储库

数据映射概览报告中的“按敏感度级别排列的顶级数据存储库”区域列出了包含最敏感项目的前四个数据存储库（系统和数据源）。每个系统的条形图分为：

- 非敏感数据
- 个人数据
- 敏感个人数据

该数据每两小时刷新一次，可以手动刷新。

步骤

1. 要查看每个类别中的项目总数，请将光标放在栏的每个部分上。
2. 要过滤调查页面中显示的结果，请选择栏中的每个区域并进一步调查。

审查敏感数据和广泛的权限

治理仪表板的“敏感数据和广泛权限”区域显示包含敏感数据和具有广泛权限的文件的数量。该表显示以下类型的权限：

- 从横轴上最严格的权限到最宽松的限制。
- 纵轴上从最不敏感的数据到最敏感的数据。

步骤

1. 要查看每个类别中的文件总数，请将光标放在每个框上。
2. 要过滤调查页面中显示的结果，请选择一个框并进一步调查。

查看按开放权限类型列出的数据

数据映射概览报告的“打开权限”区域显示正在扫描的所有文件中每种权限的百分比。该图表显示以下类型的权限：

- 无开放权限
- 向组织开放
- 向公众开放
- 未知访问

步骤

1. 要查看每个类别中的文件总数，请将光标放在每个框上。
2. 要过滤调查页面中显示的结果，请选择一个框并进一步调查。

审查数据的年龄和大小

您可以调查数据映射概览报告的“*Age*”和“*Size*”图表中的项目，看看是否有任何数据应该删除或分层到较便宜的对象存储。

步骤

1. 在数据年龄图表中，要查看有关数据年龄的详细信息，请将光标放在图表中的某个点上。
2. 要按年龄或尺寸范围进行过滤，请选择该年龄或尺寸。
    - 数据年龄图 - 根据数据创建时间、上次访问时间或上次修改时间对数据进行分类。
    - 数据大小图 - 根据大小对数据进行分类。

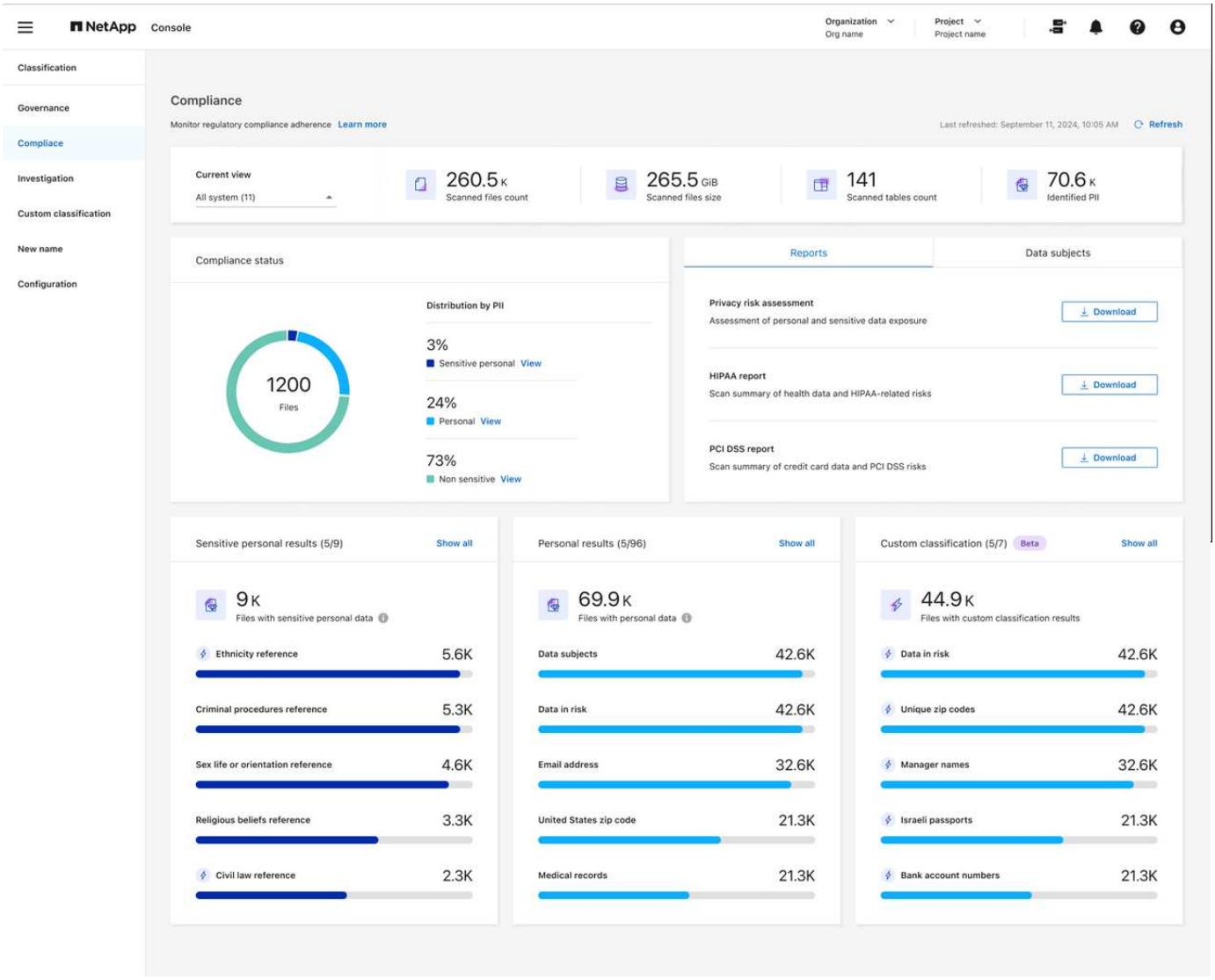
如果您的任何数据源实现了数据分层，则已驻留在对象存储中的旧数据可能会在“数据年龄”图中被识别。

使用NetApp Data Classification查看组织中存储的私人数据的合规性详细信息

通过查看组织中的个人数据 (PII) 和敏感个人数据 (SPII) 的详细信息来控制您的私人数据。您还可以通过查看NetApp Data Classification在您的数据中找到的类别和文件类型来获得可见性。

仅当您执行完整分类扫描时，才可获得文件级合规性详细信息。仅映射扫描不会产生文件级详细信息。

默认情况下，数据分类仪表板显示所有系统和数据库的合规性数据。要仅查看部分系统的数据，请选择它们。



您可以从数据调查页面过滤结果，并将结果报告下载为 CSV 文件。看"在数据调查页面中过滤数据"了解详情。

查看包含个人数据的文件

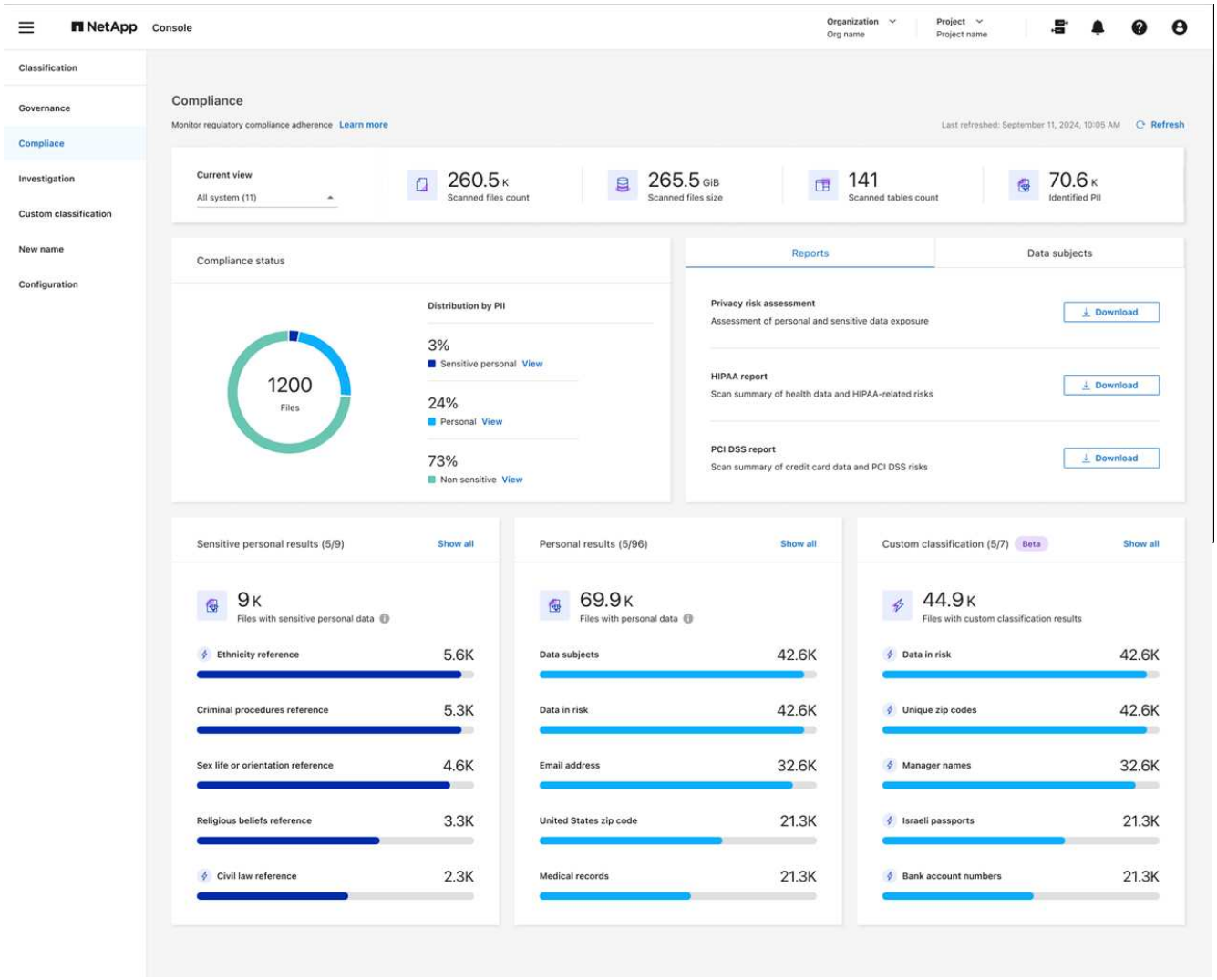
数据分类自动识别数据中的特定单词、字符串和模式（正则表达式）。"例如，信用卡号、社会保险号、银行账号、密码等等。"数据分类可在单个文件、目录（共享和文件夹）内的文件以及数据库表中识别此类信息。

您还可以创建自定义搜索词来识别特定于您组织的个人数据。有关更多信息，请参阅"创建自定义分类" 。

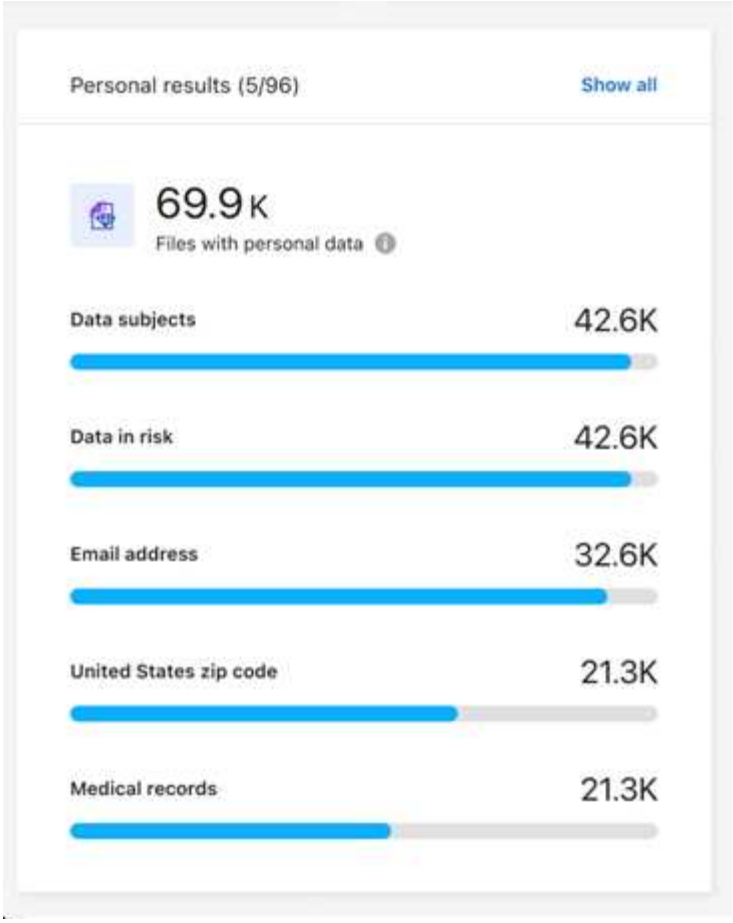
对于某些类型的个人数据，数据分类使用_邻近验证_来验证其发现。通过查找与找到的个人数据接近的一个或多个预定义关键字来进行验证。例如，如果数据分类看到旁边有一个近似词（例如，*SSN* 或 *social security*），它就会将美国社会保障号码 (SSN) 识别为 SSN。"个人资料表"显示数据分类何时使用邻近验证。

步骤

1. 从数据分类菜单中，选择“合规性”选项卡。
2. 要调查所有个人数据的详细信息，请选择个人数据百分比旁边的图标。

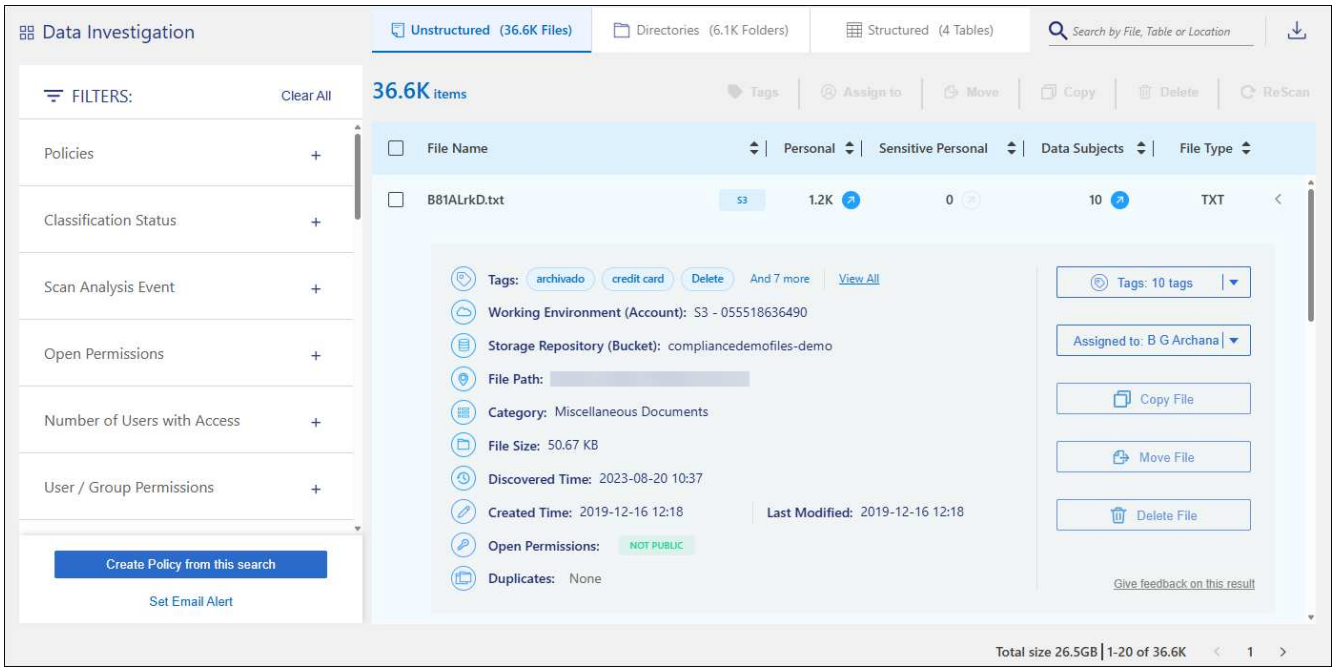


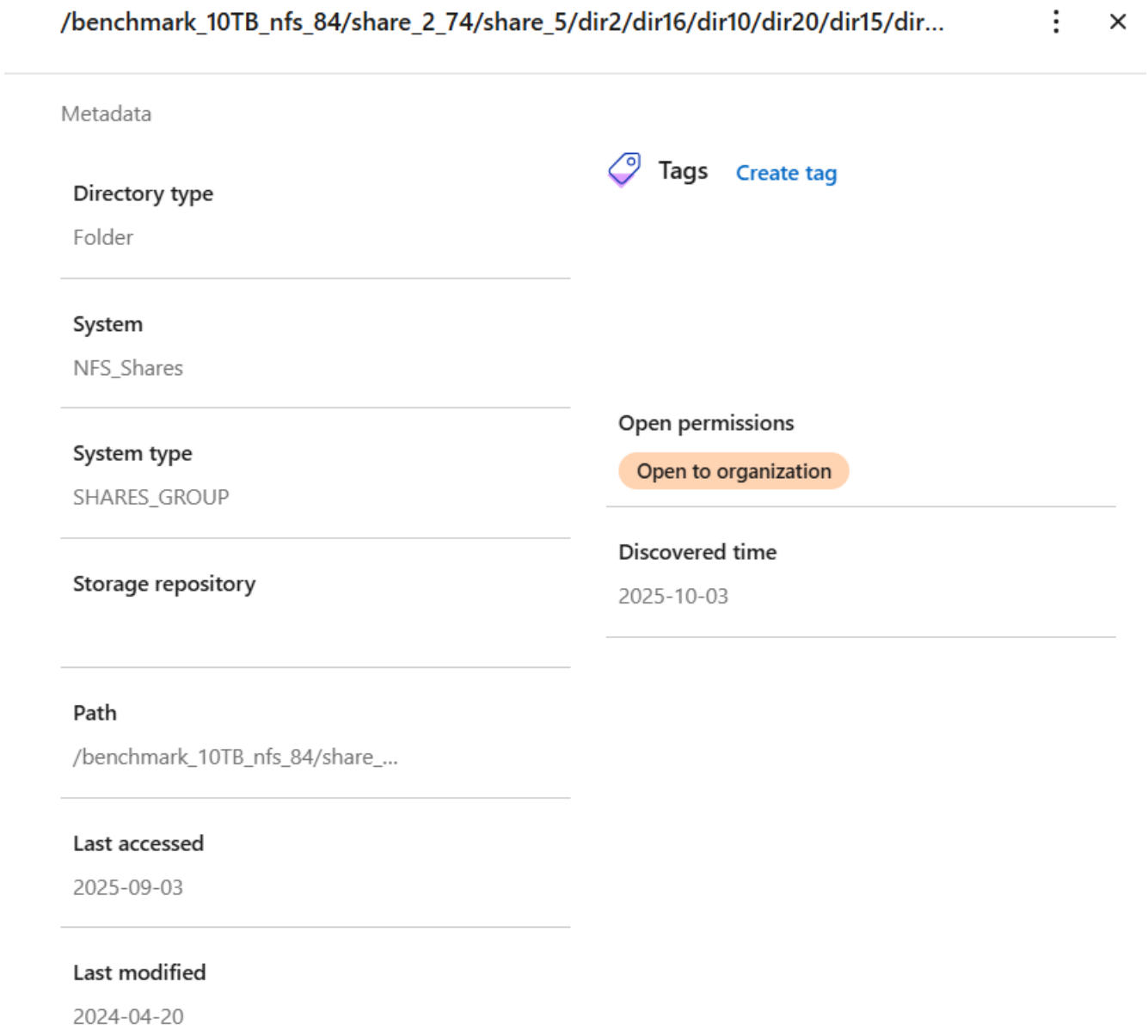
3. 要调查特定类型的个人数据的详细信息，请选择*查看全部*，然后选择特定类型的个人数据（例如电子邮件地址）的*调查结果*箭头图标。



4. 通过搜索、排序、扩展特定文件的详细信息、选择“调查结果”箭头查看屏蔽信息或下载文件列表来调查数据。

   下图显示在目录（共享和文件夹）中找到的个人数据。在“结构化”选项卡中，您可以查看数据库中的个人数据。在“非结构化”选项卡中，您可以查看文件级数据。





查看包含敏感个人数据的文件

数据分类会自动识别隐私法规所定义的特殊类型的敏感个人信息，例如 "GDPR 第 9 条和第 10 条"。例如，有关一个人的健康、种族或性取向的信息。"查看完整列表" 。数据分类可在单个文件、目录（共享和文件夹）内的文件以及数据库表中识别此类信息。

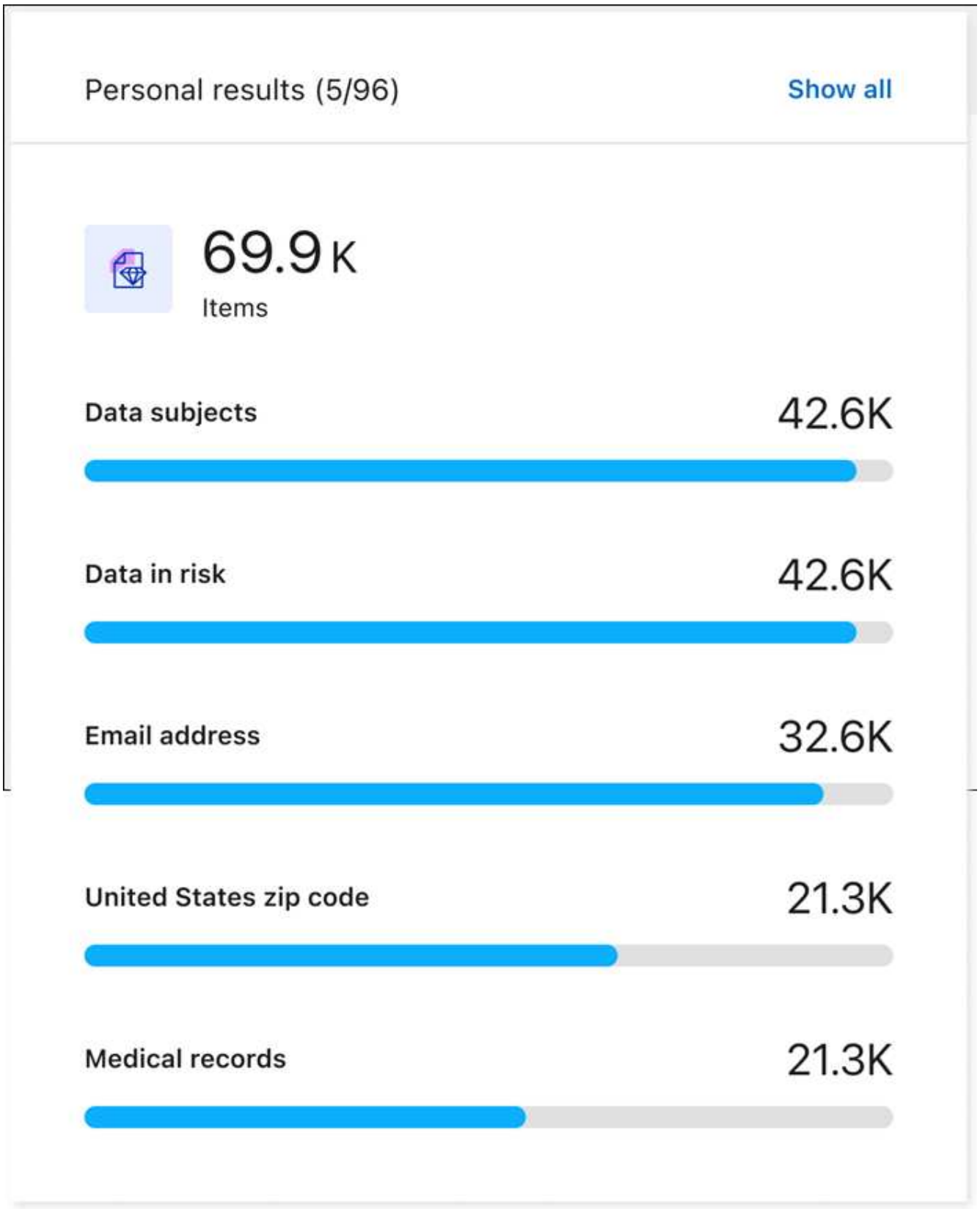
数据分类使用人工智能、自然语言处理 (NLP)、机器学习 (ML) 和认知计算 (CC) 来理解其扫描的内容的含义，以便提取实体并对其进行相应的分类。

例如，GDPR 数据的一个敏感类别是种族来源。由于其 NLP 能力，数据分类可以区分“乔治是墨西哥人”（表示 GDPR 第 9 条规定的敏感数据）和“乔治正在吃墨西哥食物”之间的区别。

扫描敏感个人数据时仅支持英语。稍后将添加对更多语言的支持。

步骤

1. 从数据分类菜单中，选择*合规性*。
2. 要调查所有敏感个人数据的详细信息，请找到敏感个人信息结果卡，然后选择显示全部。



。

3. 要调查特定类型的敏感个人数据的详细信息，请选择“查看全部”，然后选择特定类型的敏感个人数据的“调查结果”箭头图标。
4. 通过搜索、排序、扩展特定文件的详细信息、单击“调查结果”查看屏蔽信息或下载文件列表来调查数据。

NetApp Data Classification中的私有数据类别

NetApp Data Classification可以在您的卷和数据库中识别多种类型的私有数据。

数据分类识别两种类型的个人数据：

- 个人身份信息（**PII**）
- 敏感个人信息（**SPII**）

如果您需要数据分类来识别其他私人数据类型，例如额外的国民身份证号码或医疗保健标识符，请联系您的客户经理。

个人数据的类型

文件中的个人数据或_个人身份信息_（PII）可以是一般个人数据或国家标识符。下表第三列标识数据分类是否使用"接近度验证"验证其对标识符的发现。

表中标明了可以识别这些项目的语言。

类型	标识符	接近度验证?	英语	德语	西班牙语	法语	日语
常规	信用卡号码	是	✓	✓	✓		✓
	数据主体	否	✓	✓	✓		
	电子邮件地址	否	✓	✓	✓		✓
	IBAN 号码（国际银行账户号码）	否	✓	✓	✓		✓
	IP 地址	否	✓	✓	✓		✓
	密码	是	✓	✓	✓		✓

类型	标识符	接近度验证?	英语	德语	西班牙语	法语	日语
国家标识符							

类型	标识符	接近度验证?	英语	德语	西班牙语	法语	日语

类型	标识符	接近度验证?	英语	德语	西班牙语	法语	日语
	希腊身份证	是	✓	✓	✓		
	匈牙利税务识别号	是	✓	✓	✓		
	爱尔兰身份证 (PPS)	是	✓	✓	✓		
	以色列身份证	是	✓	✓	✓		
	意大利税务识别号	是	✓	✓	✓		
	日本个人身份证号码（个人和公司）	是	✓	✓	✓		✓
	拉脱维亚身份证	是	✓	✓	✓		
	立陶宛身份证	是	✓	✓	✓		
	卢森堡身份证	是	✓	✓	✓		
	马耳他身份证	是	✓	✓	✓		
	国家医疗服务体系 (NHS) 号码	是	✓	✓	✓		
	新西兰银行账户	是	✓	✓	✓		
	新西兰驾驶执照	是	✓	✓	✓		
	新西兰税务局 (IRD) 号码（税号）	是	✓	✓	✓		
	新西兰 NHI（国民健康指数）号码	是	✓	✓	✓		
	新西兰护照号码	是	✓	✓	✓		
	波兰身份证 (PESEL)	是	✓	✓	✓		
	葡萄牙税务识别号（NIF）	是	✓	✓	✓		
	罗马尼亚身份证 (CNP)	是	✓	✓	✓		
	新加坡国民登记身份证（NRIC）	是	✓	✓	✓		
	斯洛文尼亚身份证 (EMSO)	是	✓	✓	✓		
	南非身份证	是	✓	✓	✓		
	西班牙税务识别号	是	✓	✓	✓		
	瑞典身份证	是	✓	✓	✓		
	英国身份证（NINO）	是	✓	✓	✓		
	美国加州驾驶执照	是	✓	✓	✓		
	美国印第安纳州驾驶执照	是	✓	✓	✓		
	美国纽约州驾驶执照	是	✓	✓	✓		
	美国德克萨斯州驾驶执照	是	✓	✓	✓		
	美国社会安全号码（SSN）	是	✓	✓	✓		

敏感个人数据的类型

数据分类可以在文件中找到以下敏感个人信息（SPII）。

以下 SPII 目前仅能以英文识别：

- 刑事诉讼参考：有关自然人的刑事定罪和犯罪的数据。

- 种族参考：有关自然人的种族或民族血统的数据。
- 健康参考：有关自然人健康的数据。
- **ICD-9-CM** 医疗代码：医疗保健行业使用的代码。
- **ICD-10-CM** 医疗代码：医疗保健行业使用的代码。
- 哲学信仰参考：有关自然人的哲学信仰的数据。
- 政治观点参考：有关自然人政治观点的数据。
- 宗教信仰参考：有关自然人的宗教信仰的数据。
- 性生活或性取向参考：有关自然人的性生活或性取向的数据。

类别类型

数据分类将您的数据分类如下。

大多数类别都可以用英语、德语和西班牙语识别。

类别	类型	英语	德语	西班牙语
金融	资产负债表	✓	✓	✓
	采购订单	✓	✓	✓
	发票	✓	✓	✓
	季度报告	✓	✓	✓
人力资源	背景调查	✓		✓
	薪酬计划	✓	✓	✓
	员工合同	✓		✓
	员工评价	✓		✓
	运行状况	✓		✓
	简历	✓	✓	✓
合法的	保密协议	✓	✓	✓
	供应商-客户合同	✓	✓	✓
营销	活动	✓	✓	✓
	会议	✓	✓	✓
操作	审计报告	✓	✓	✓
销售额	销售订单	✓	✓	
服务	射频干扰	✓		✓
	征求建议书	✓		✓
	母猪	✓	✓	✓
	培训	✓	✓	✓
支持	投诉和票务	✓	✓	✓

以下元数据也使用相同的受支持语言进行分类和识别：

- 应用程序数据
- 存档文件
- 声音的
- 数据分类业务应用数据中的面包屑
- CAD 文件
- 代码
- 腐败
- 数据库和索引文件
- 设计文件
- 电子邮件应用程序数据
- 加密（具有高熵值的文件）
- 可执行文件
- 财务应用数据
- 健康应用数据
- 图片
- 日志
- 杂项文件
- 杂项演示
- 杂项电子表格
- 杂项“未知”
- 受密码保护的文件
- 结构化数据
- 视频
- 零字节文件

文件类型

数据分类扫描所有文件的类别和元数据洞察，并在仪表板的文件类型部分显示所有文件类型。当数据分类检测个人身份信息 (PII) 或执行 DSAR 搜索时，仅支持以下文件格式：

```
.CSV, .DCM, .DOC, .DOCX, .JSON, .PDF, .PPTX, .RTF, .TXT, .XLS, .XLSX, Docs,
Sheets, and Slides
```

所发现信息的准确性

NetApp无法保证数据分类识别的个人数据和敏感个人数据 100% 的准确性。您应该始终通过查看数据来验证信息。

根据我们的测试，下表显示了数据分类发现的信息的准确性。我们根据_精度_和_召回率_来细分它：

精确

数据分类发现的内容被正确识别的概率。例如，个人数据的准确率为 90%，意味着在被识别为包含个人信息的 10 个文件中，有 9 个实际上包含个人信息。 10 个文件中会有 1 个是误报。

记起

数据分类找到其应有内容的概率。例如，个人数据的召回率为 70%，意味着数据分类可以识别出组织中 10 个文件中实际包含个人信息的 7 个。数据分类会遗漏 30% 的数据，并且不会出现在仪表板中。

我们正在不断提高结果的准确性。这些改进将在未来的数据分类版本中自动提供。

类型	精确	记起
个人数据 - 一般	90%-95%	60%-80%
个人数据 - 国家标识符	30%-60%	40%-60%
敏感个人数据	80%-95%	20%-30%
类别	90%-97%	60%-80%

在NetApp Data Classification中创建自定义分类

NetApp Data Classification允许您创建自定义类别或个人标识符，以识别特定于您组织监管和合规要求的数据。

数据分类支持两种类型的自定义分类器：类别和个人标识符。自定义类别是根据您上传的一组文件创建的，数据分类功能会根据这些文件创建一个 AI 模型，以识别您组织中的类似数据（例如，一家健康研究公司可能会创建一个临床分析类别）。使用关键字列表或正则表达式 (regex) 创建自定义个人标识符，以识别贵组织特有的、可能构成合规风险的信息。

所有自定义分类都可以在自定义分类控制面板中找到。

创建自定义个人标识符

数据分类功能允许您使用上下文关键字或正则表达式创建自定义个人标识符，以识别贵组织特有的数据。

关键词要求

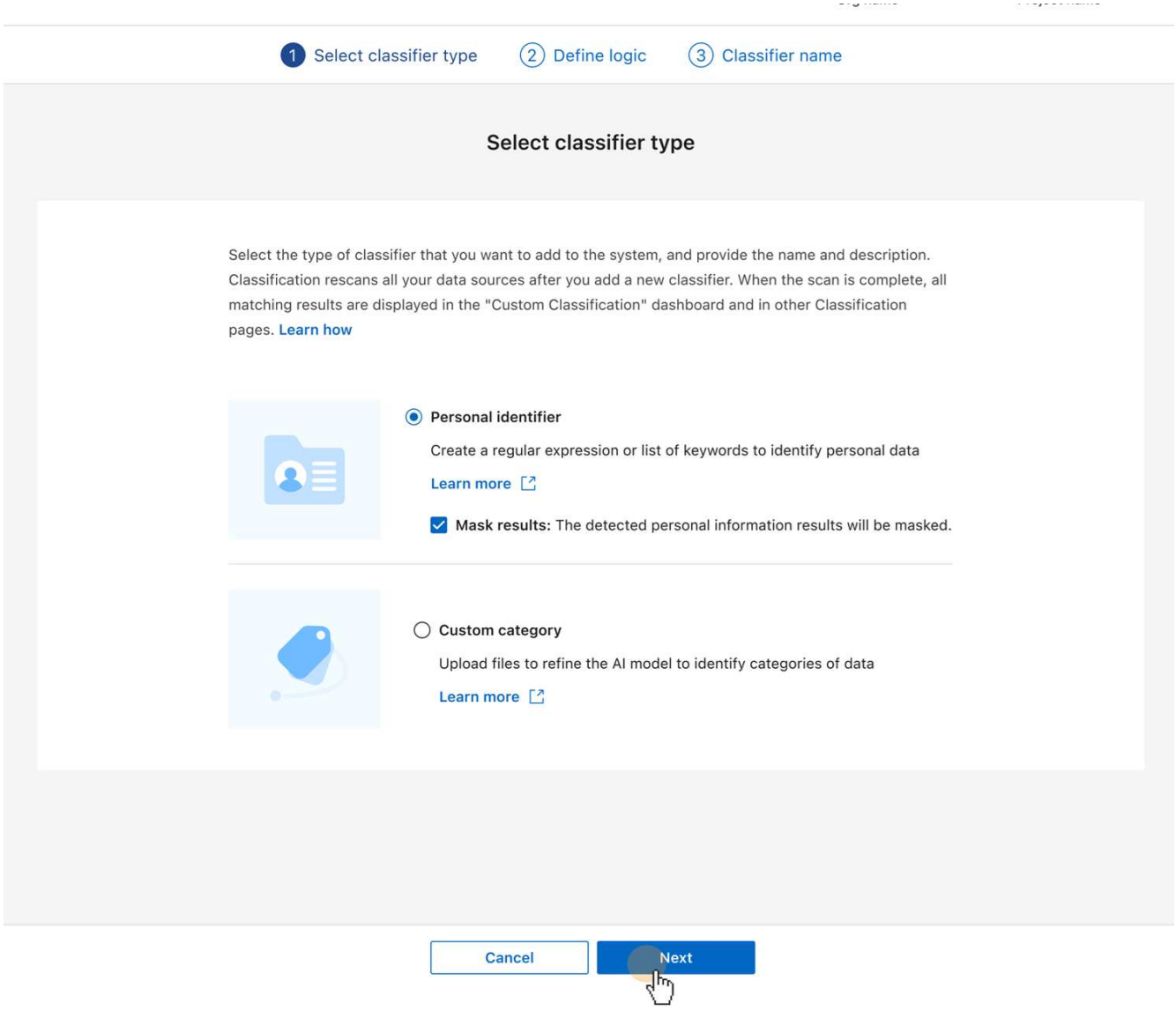
如果您使用关键词列表创建个人标识符，则该列表必须满足以下要求：

- 关键词输入不区分大小写。
- 关键词必须至少包含三个字符。长度少于三个字符的单词将被忽略。
- 重复的词语只会添加一次。
- 关键词总数不能超过 50 万个字符。列表中必须至少包含一个关键词。

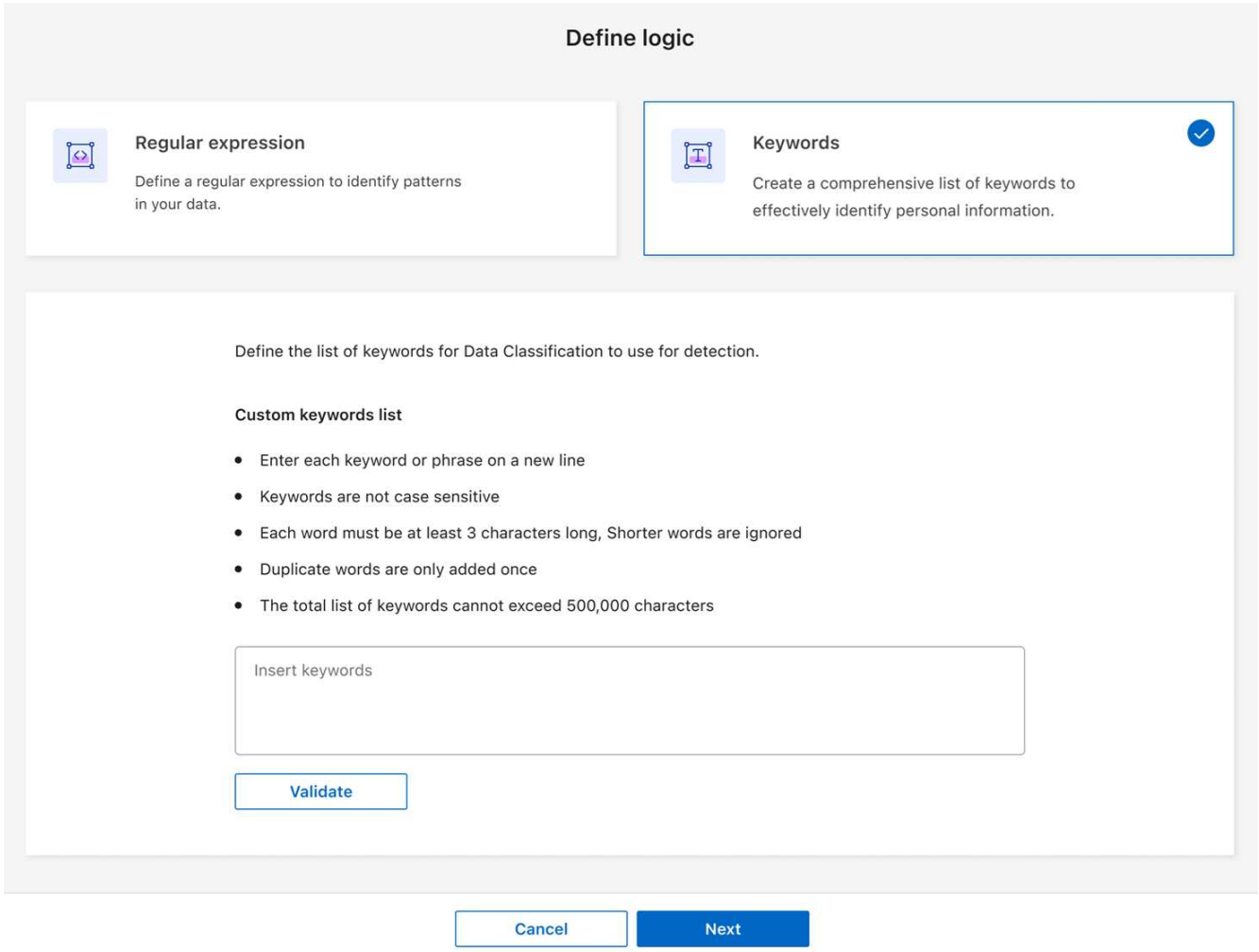
步骤

1. 选择自定义分类选项卡。
2. 选择**+** 新建分类器以创建自定义分类器。

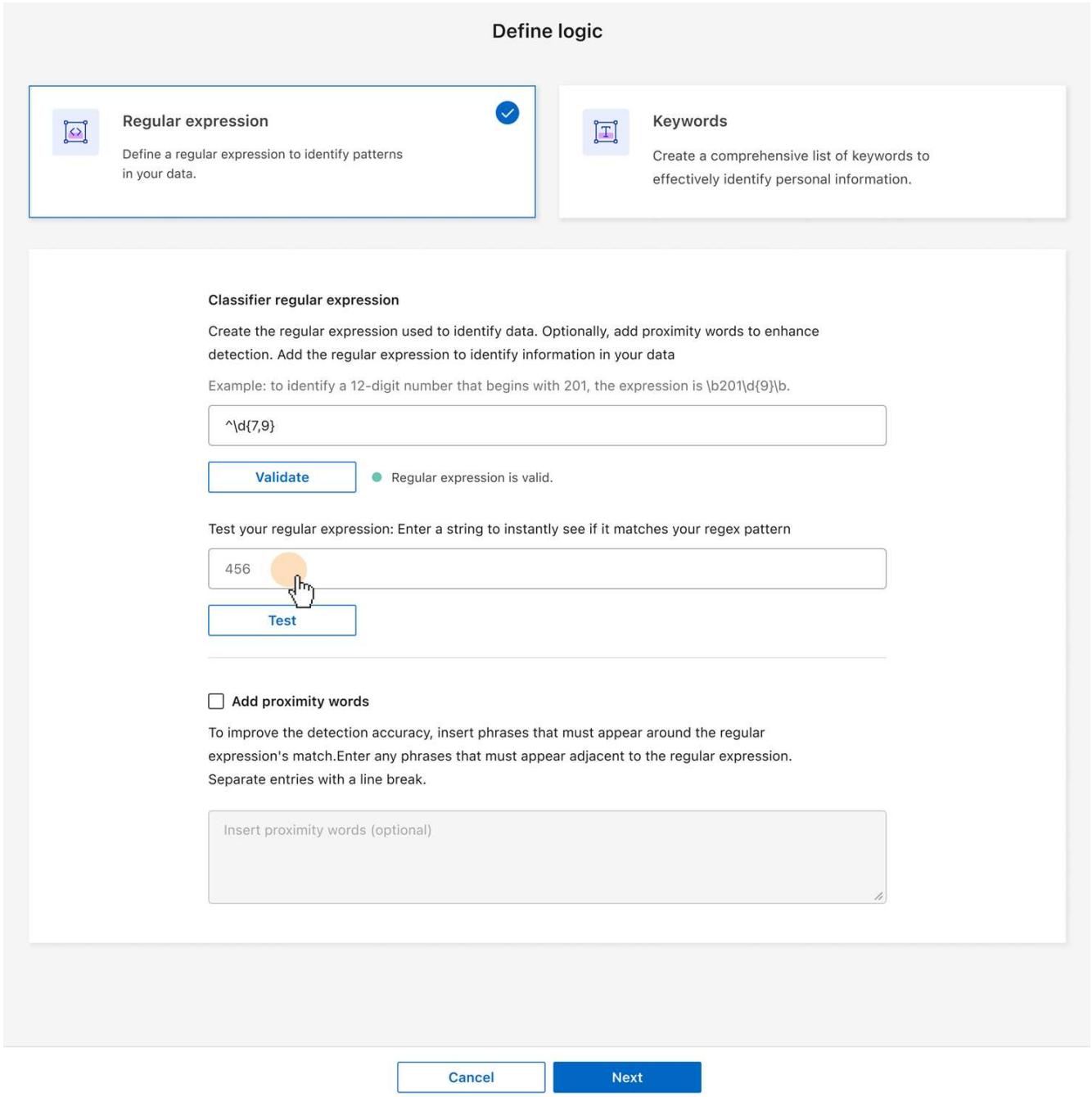
3. 请选择*个人标识符*。（可选）选择“屏蔽结果”以屏蔽检测到的个人数据。
4. 选择下一步。



5. 要添加带关键词的分类器，请选择关键词。请输入关键词列表，每个关键词占一行。请确保关键词符合要求。



要将分类器添加为正则表达式，请选择正则表达式，然后添加一个模式来检测数据的特定信息。选择验证以确认您输入的语法正确。



a. （可选）输入一个应该与正则表达式模式匹配的示例字符串，然后选择测试进行检查。

b. （可选）添加邻近词。如果添加邻近词，数据分类仅在邻近词与匹配字符串相邻时才标记正则表达式模式。

6. 选择下一步。
7. 输入分类器名称和描述，以便在仪表板中标识自定义类别。
8. 选择保存以创建自定义个人标识符。

创建自定义个人标识符后，其结果将在下次计划扫描中捕获。为了更快地获取结果，请执行按需扫描。要查看结果，请参阅 生成合规性报告。

创建自定义类别

通过自定义类别，您可以对特定于您组织的数据进行分类。自定义类别是根据您上传的文本文件创建的，数据分类功能会根据这些文件创建一个人工智能模型，以识别其他文件中的类似信息。

训练数据要求

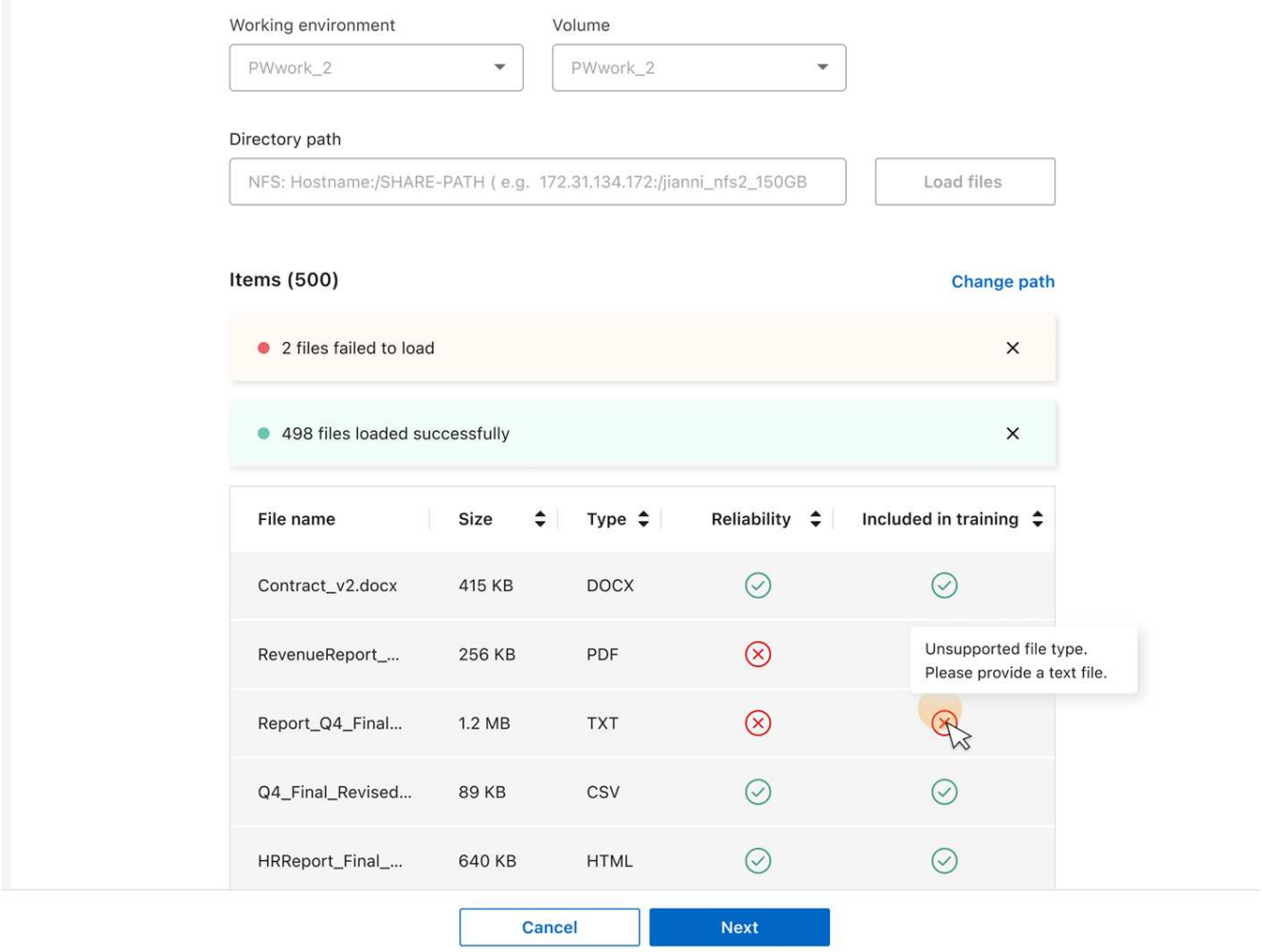
- 训练数据集必须至少包含 25 个文件。最大文件数为 1,000。
- 所有文件必须直接位于您提供的文件路径中。
- 所有文件必须大于 100 字节。
- 数据分类训练数据必须是以下文件类型之一：CSV、DOCX、DOC、GZ、JSON、PDF、PPTX、TXT、RTT、XLS 或 XLSX。您可以上传所有支持的文件类型的组合。

步骤

1. 在NetApp Data Classification中，选择“自定义分类”。
2. 选择 **+** 新建分类器。
3. 选择“自定义类别”作为分类器类型，然后下一步。
4. 使用一系列基于文本的文件来定义自定义类别的逻辑。请提供*工作地址*的IP地址，然后从下拉菜单中选择*音量*。

   输入包含训练数据的目录的目录路径。

5. 选择“加载文件”进行数据分类，以执行文件检查。您可以查看文件摘要，其中列出了文件名、大小、类型和备注（如果该文件被认为适合用于培训）。



a. 要更改文件路径或重新上传文件，请选择更改路径，然后输入数据并再次加载文件。

6. 当您对上传的文件满意后，请选择下一步。
7. 输入分类器名称和描述，以便在仪表板中标识自定义类别。
8. 选择保存以创建自定义类别。

结果

创建自定义类别后，其结果将在下次计划扫描中捕获。为了更快地获取结果，请手动启动扫描。

编辑自定义分类器

创建个人标识符后，您可以修改其逻辑。您无法更改个人标识符的类型或逻辑类型；例如，您无法将自定义类别更改为自定义个人标识符。您也不能将基于关键字的自定义标识符更改为基于正则表达式的自定义标识符。

步骤

1. 在NetApp Data Classification中，选择“自定义分类”。
2. 确定要删除的分类器，然后选择操作菜单 ... 在它那一行的末尾。
3. 选择编辑逻辑。
4. 如果要修改关键词，请添加、删除或编辑相应的关键词。如果要修改正则表达式，请输入新的正则表达式并进行验证。（可选）添加邻近关键词。

5. 选择“保存”以应用更改。

删除自定义分类器

1. 在NetApp Data Classification中，选择“自定义分类”。
2. 确定要删除的分类器，然后选择操作菜单 ... 在它那一行的末尾。
3. 选择删除分类器。

下一步

- 生成合规性报告

使用NetApp Data Classification调查组织中存储的数据

数据调查仪表板显示文件和目录级别的数据洞察，使您能够对结果进行排序和过滤。数据调查页面提供有关文件和目录元数据和权限的见解以及识别重复文件。通过文件、目录和数据库级别的洞察，您可以采取措施来提高组织的合规性并节省存储空间。数据调查页面还支持移动、复制和删除文件。

要从调查页面获得见解，您必须对数据源执行完整的分类扫描。仅进行过映射扫描的数据源不会显示文件级别的详细信息。

数据调查结构

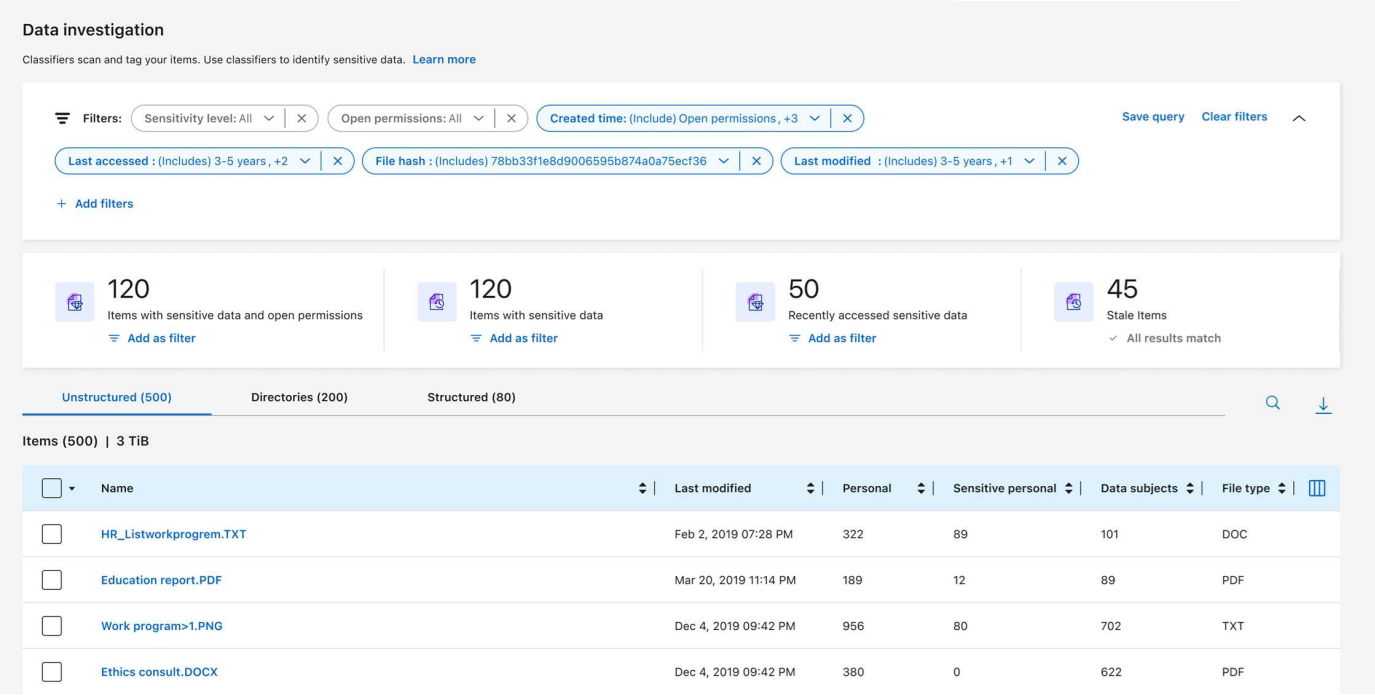
数据调查页面将数据分类到三个选项卡中：

- 非结构化数据：文件数据
- 目录：文件夹和文件共享
- 结构化：数据库

数据过滤器

数据调查页面提供了许多过滤器来对您的数据进行分类，以便您可以找到所需的数据。您可以同时使用多个过滤器。

要添加过滤器，请选择添加过滤器按钮。



过滤敏感度和内容

使用以下过滤器查看您的数据中包含多少敏感信息。

筛选器	详细信息
类别	选择"类别类型"。
敏感度等级	选择敏感度级别：个人、敏感个人或非敏感。
标识符数量	选择每个文件检测到的敏感标识符的范围。包括个人数据和敏感个人数据。在目录中过滤时，数据分类会将每个文件夹（和子文件夹）中所有文件的匹配结果汇总。注意：2023 年 12 月（版本 1.26.6）版本删除了按目录计算个人身份信息 (PII) 数据数量的选项。
个人资料	选择"个人数据的类型"。
敏感个人信息	选择"敏感个人数据的类型"。
数据主体	输入数据主体的全名或已知标识符。"在此处了解有关数据主体的更多信息" 。

过滤用户所有者和用户权限

使用以下过滤器查看文件所有者和访问数据的权限。

筛选器	详细信息
开放权限	选择数据和文件夹/共享内的权限类型。
用户/组权限	选择一个或多个用户名和/或组名，或者输入部分名称。
文件所有者	输入文件所有者名称。
有访问权限的用户数	选择一个或多个类别范围以显示哪些文件和文件夹对一定数量的用户开放。

按时间顺序过滤

使用以下过滤器根据时间标准查看数据。

筛选器	详细信息
创建时间	选择文件创建的时间范围。您还可以指定自定义时间范围来进一步优化搜索结果。
发现时间	选择数据分类发现文件的时间范围。您还可以指定自定义时间范围来进一步优化搜索结果。
上次修改	选择文件最后修改的时间范围。您还可以指定自定义时间范围来进一步优化搜索结果。
上次访问	选择文件或目录*上次被访问的时间范围。您还可以指定自定义时间范围来进一步优化搜索结果。对于数据分类扫描的文件类型，这是数据分类最后一次扫描该文件的时间。

{星号} 目录的上次访问时间仅适用于 NFS 或 CIFS 共享。

过滤元数据

使用以下过滤器根据位置、大小和目录或文件类型查看数据。

筛选器	详细信息
文件路径	输入最多 20 条要在查询中包含或排除的部分或完整路径。如果同时输入包含路径和排除路径，数据分类会首先在包含路径中找到所有文件，然后从排除路径中删除文件，然后显示结果。请注意，在此过滤器中使用“*”没有任何效果，并且您无法从扫描中排除特定文件夹 - 配置共享下的所有目录和文件都将被扫描。
目录类型	选择目录类型；“共享”或“文件夹”。
文件类型	选择"文件类型"。
文件大小	选择文件大小范围。
文件哈希	输入文件的哈希值即可查找特定文件，即使名称不同。

过滤器存储类型

使用以下过滤器按存储类型查看数据。

筛选器	详细信息
系统类型	选择系统类型。
系统环境名称	选择特定系统。
存储库	选择存储库，例如卷或模式。

过滤查询

使用以下过滤器按已保存的查询查看数据。

筛选器	详细信息
已保存的查询	选择一个或多个已保存的查询。前往"已保存的查询选项卡"查看现有已保存查询的列表并创建新查询。
标签	选择"一个或多个标签"分配给您的文件。

过滤分析状态

使用以下过滤器按数据分类扫描状态查看数据。

筛选器	详细信息
分析状态	选择一个选项来显示“等待首次扫描”、“已完成扫描”、“等待重新扫描”或“扫描失败”的文件列表。
扫描分析事件	选择是否要查看由于数据分类无法恢复上次访问时间而未分类的文件，或者即使数据分类无法恢复上次访问时间但已分类的文件。

"查看有关“上次访问时间”时间戳的详细信息"有关使用扫描分析事件进行过滤时调查页面中出现的项目的更多信息。

按重复项过滤数据

使用以下过滤器查看存储中重复的文件。

筛选器	详细信息
重复项	选择文件是否在存储库中重复。

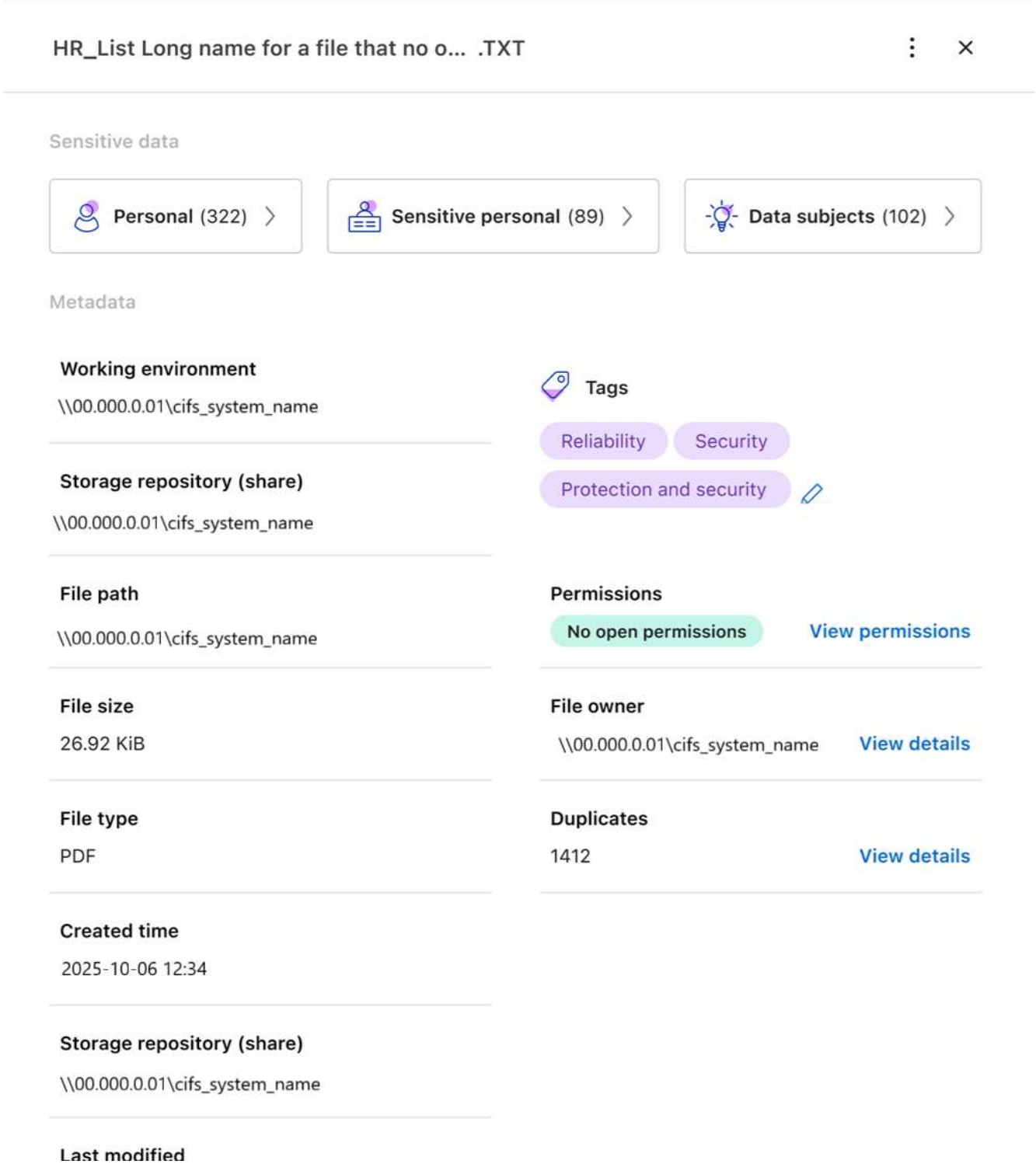
查看文件元数据

除了显示文件所在的系统和卷之外，元数据还显示更多信息，包括文件权限、文件所有者以及该文件是否有重复。如果您打算"创建已保存的查询"因为您可以看到可用于过滤数据的所有信息。

信息的可用性取决于数据来源。例如，数据库文件的卷名和权限不共享。

步骤

1. 从数据分类菜单中，选择*调查*。
2. 在右侧的数据调查列表中，选择向下插入符号 在任意单个文件的右侧查看文件元数据。



3. 或者，您可以使用*创建标签*按钮为文件创建或添加标签。从下拉菜单中选择一个现有标签或使用 + 添加 按钮添加一个新标签。标签可用于过滤数据。

查看文件和目录的用户权限

要查看有权访问文件或目录的所有用户或组的列表以及他们拥有的权限类型，请选择“查看所有权限”。此选项仅适用于 CIFS 共享中的数据。

如果您使用安全标识符 (SID) 而不是用户名和组名，则应该将 Active Directory 集成到数据分类中。有关更多信

息，请参阅"将 Active Directory 添加到数据分类" 。

步骤

1. 从数据分类菜单中，选择*调查*。
2. 在右侧的数据调查列表中，选择向下插入符号 在任意单个文件的右侧查看文件元数据。
3. 要查看有权访问文件或目录的所有用户或组的列表以及他们拥有的权限类型，请在“打开权限”字段中选择“查看所有权限”。

   数据分类在列表中显示最多 100 个用户。

4. 选择向下插入符号 任何群组的按钮即可查看属于该群组的用户列表。

   您可以展开该组的某个级别来查看属于该组的用户。

5. 选择用户或组的名称以刷新调查页面，以便您可以看到该用户或组有权访问的所有文件和目录。

检查存储系统中的重复文件

您可以检查存储系统中是否存储了重复的文件。如果您想确定可以节省存储空间的区域，这将非常有用。确保具有特定权限或敏感信息的某些文件不会在存储系统中不必要地重复也是很好的。

数据分类会比较所有文件（数据库除外）是否存在重复项，如果存在重复项，则进行以下操作：

- 1 MB 或更大
- 或包含个人信息或敏感个人信息

数据分类使用散列技术来确定重复文件。如果一个文件的哈希码与另一个文件相同，即使文件名不同，这两个文件也是完全相同的副本。

步骤

1. 从数据分类菜单中，选择*调查*。
2. 在“过滤器”窗格中，选择“文件大小”以及“重复项”（“有重复项”）以查看您的环境中哪些特定大小范围的文件是重复的。
3. 或者，下载重复文件的列表并将其发送给存储管理员，以便他们可以决定可以删除哪些文件（如果有）。
4. 您可以选择删除、标记或移动重复的文件。选择您想要执行操作的文件，然后选择适当的操作。

查看特定文件是否重复

您可以查看单个文件是否有重复。

步骤

1. 从数据分类菜单中，选择*调查*。
2. 在数据调查列表中，选择 在任意单个文件的右侧查看文件元数据。

   如果文件存在重复，则此信息将显示在“*Duplicates*”字段旁边。

3. 要查看重复文件的列表及其位置，请选择“查看详细信息”*。

4. 在下一页中选择“查看重复项”以查看调查页面中的文件。
5. 您可以选择删除、标记或移动重复的文件。选择您想要执行操作的文件，然后选择适当的操作。

您可以使用此页面提供的“文件哈希”值并将其直接输入到调查页面中，以便随时搜索特定的重复文件 - 或者您可以在已保存的查询中使用它。

下载您的报告

您可以以 CSV 或 JSON 格式下载过滤结果。

如果数据分类正在扫描文件（非结构化数据）、目录（文件夹和文件共享）和数据库（结构化数据），则最多可以下载三个报告文件。

文件被分割成具有固定行数或记录数的文件：

- JSON：每份报告 100,000 条记录，生成大约需要 5 分钟
- CSV：每份报告 200,000 条记录，生成大约需要 4 分钟

您可以下载 CSV 文件的版本以在此浏览器中查看。此版本限制为 10,000 条记录。

可下载报告包含的内容

*非结构化文件数据报告*包含有关您的文件的以下信息：

- 文件名
- 位置类型
- 系统名称
- 存储库（例如，卷、存储桶、共享）
- 存储库类型
- 文件路径
- 文件类型
- 文件大小（MB）
- 创建时间
- 上次修改时间
- 上次访问
- 文件所有者
  - 配置 Active Directory 时，文件所有者数据包括帐户名称、SAM 帐户名称和电子邮件地址。
- 类别
- 个人信息
- 敏感个人信息
- 开放权限

- 扫描分析错误
- 删除检测日期

  删除检测日期标识文件被删除或移动的日期。这使您能够识别敏感文件何时被移动。已删除的文件不会计入仪表板或调查页面上显示的文件数量。这些文件仅出现在 CSV 报告中。

*非结构化目录数据报告*包括有关您的文件夹和文件共享的以下信息：

- 系统类型
- 系统名称
- 目录名称
- 存储库（例如文件夹或文件共享）
- 目录所有者
- 创建时间
- 发现时间
- 上次修改时间
- 上次访问
- 开放权限
- 目录类型

*结构化数据报告*包含有关数据库表的以下信息：

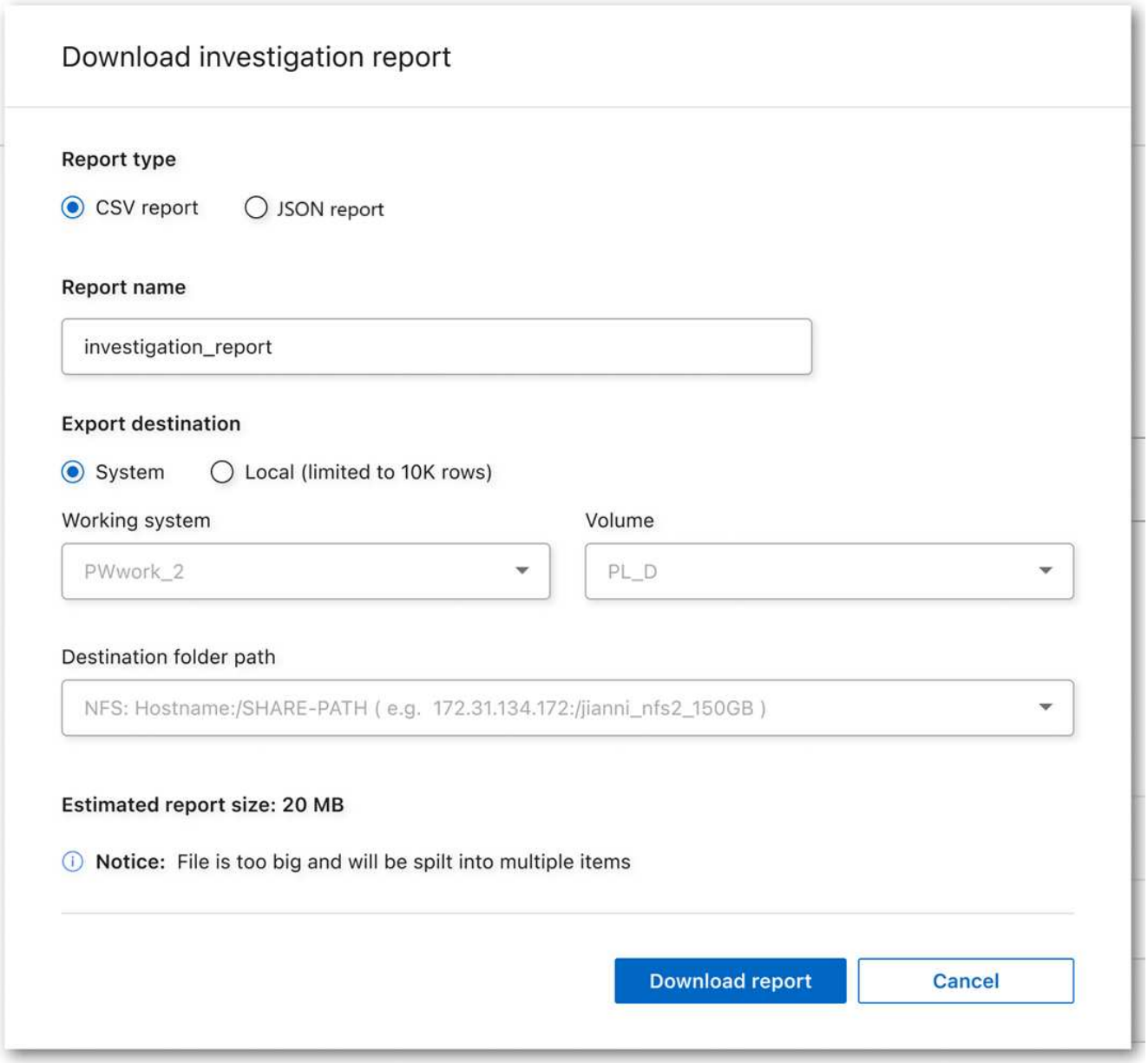
- 数据库表名称
- 位置类型
- 系统名称
- 存储库（例如，架构）
- 列数
- 行数
- 个人信息
- 敏感个人信息

生成报告的步骤

1. 从数据调查页面中，选择 页面右上方的按钮。
2. 选择报告类型：CSV 或 JSON。
3. 输入报告名称。
4. 要下载完整的报告，请选择系统，然后从相应的下拉菜单中选择系统和卷。提供目标文件夹路径。

   要在浏览器中下载报告，请选择本地。请注意，此选项将报告限制为前 10,000 行，并且仅限于 **CSV** 格式。如果您选择本地，则无需填写任何其他字段。

5. 选择下载报告。



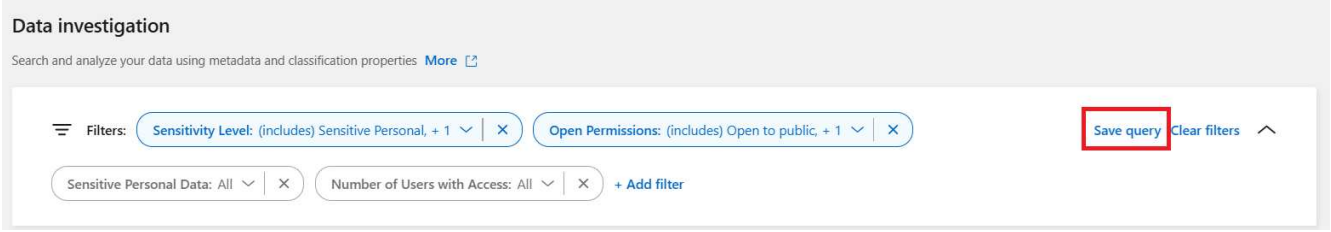
结果

对话框中将显示一条消息，提示正在下载报告。

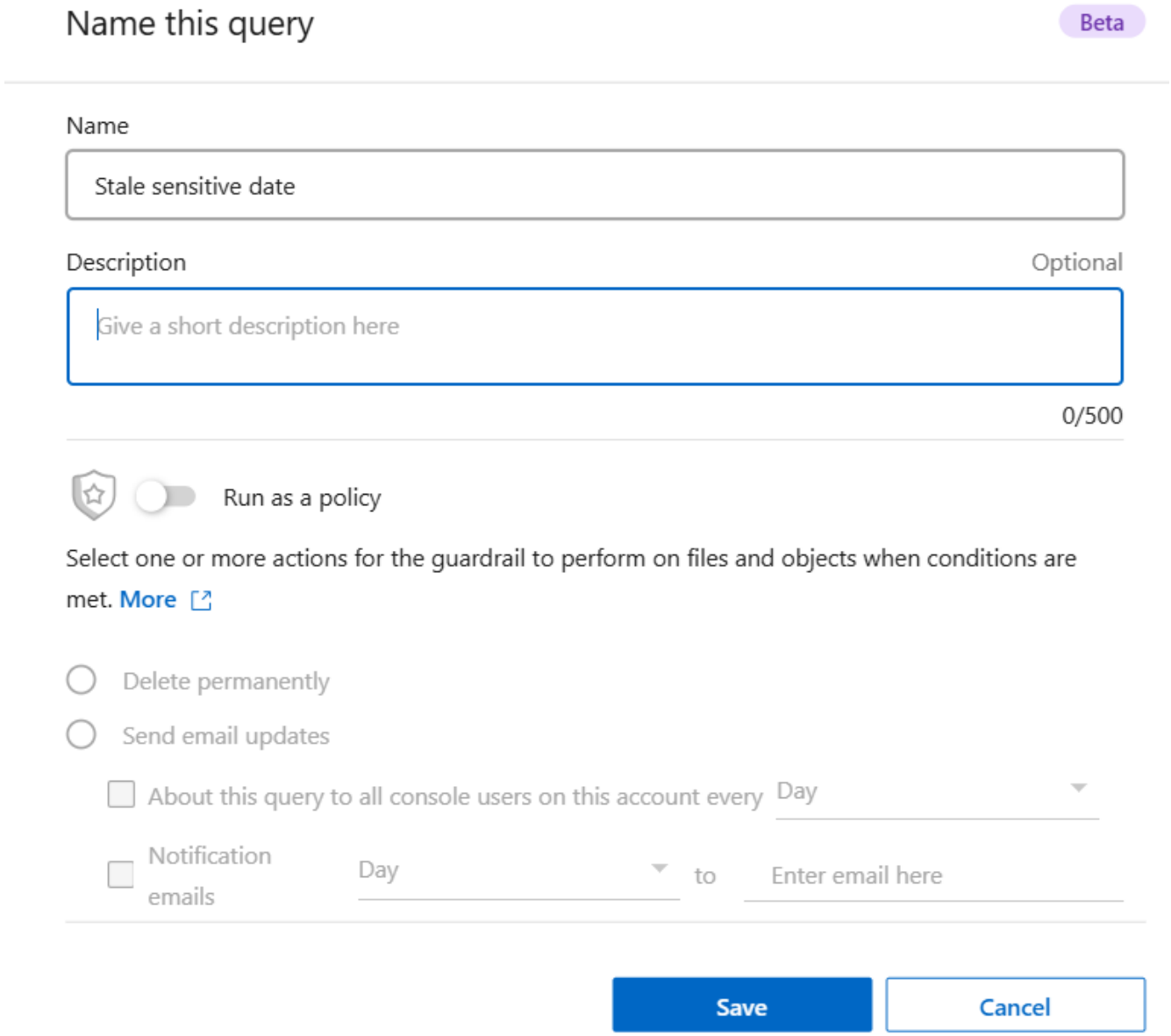
根据选定的过滤器创建已保存的查询

步骤

1. 在调查选项卡中，通过选择要使用的过滤器来定义搜索。看"在调查页面中过滤数据"了解详情。
2. 一旦您根据自己的喜好设置了所有过滤器特性，请选择*保存查询*。



3. 为保存的查询命名并添加描述。该名称必须是唯一的。
4. 您可以选择将查询保存为策略：
   a. 要将查询保存为策略，请切换*作为策略运行*开关。
   b. 选择*永久删除*或*发送电子邮件更新*。如果您选择电子邮件更新，您可以每天、每周或每月通过电子邮件将查询结果发送给所有控制台用户。或者，您可以以相同的频率将通知发送到特定的电子邮件地址。
5. 选择*保存*。



创建搜索或策略后，您可以在已保存的查询选项卡中查看它。

结果可能需要最多 15 分钟才会显示在“已保存的查询”页面上。

使用NetApp Data Classification管理已保存的查询

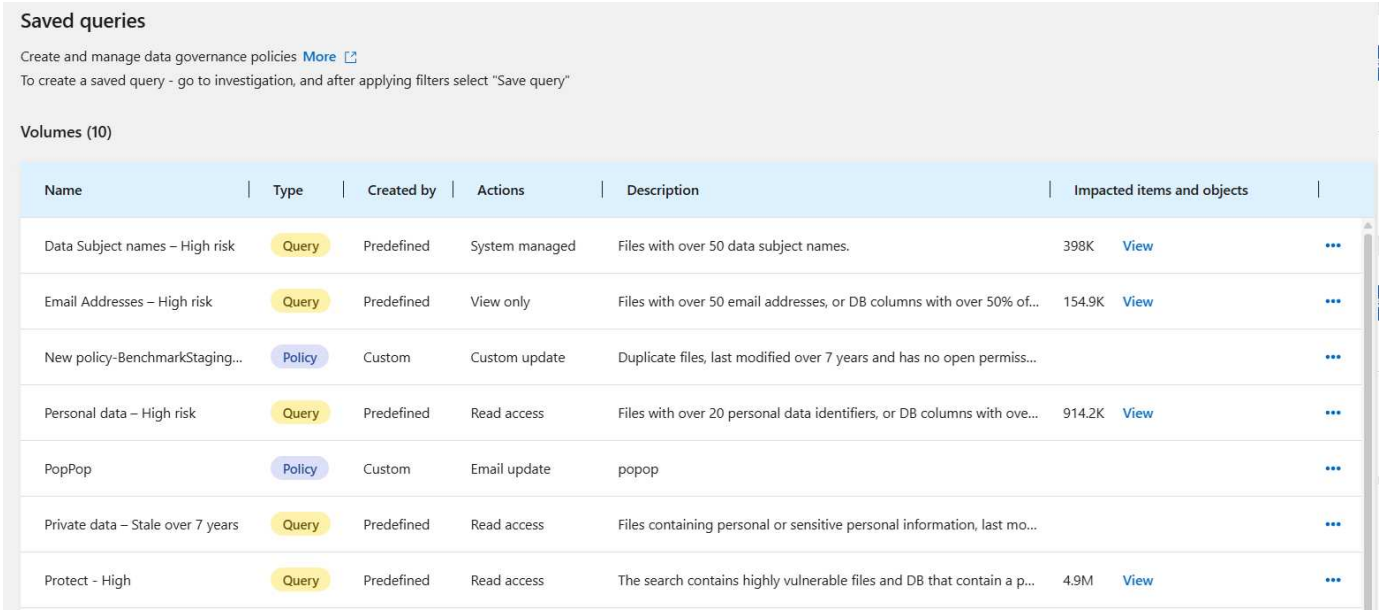
NetaApp 数据分类支持保存您的搜索查询。使用已保存的查询，您可以创建自定义过滤器来对数据调查页面的常见查询进行排序。数据分类还包括基于常见请求的预定义保存的查询。

合规性仪表板中的“已保存的查询”选项卡列出了此数据分类实例上可用的所有预定义和自定义已保存查询。

已保存的查询也可以保存为策略。查询过滤数据，而策略允许您对数据采取行动。通过策略：您可以删除发现的

数据或发送有关发现的数据的电子邮件更新。

已保存的查询也会出现在调查页面的过滤器列表中。



在调查页面中查看已保存的查询结果

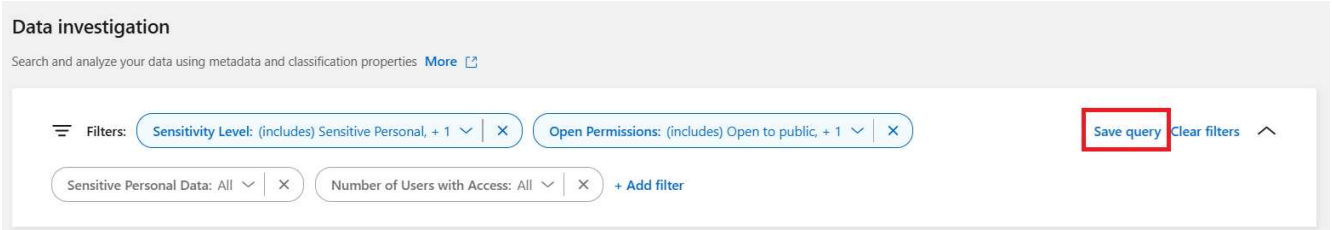
要在调查页面中显示已保存查询的结果，请选择 按钮进行特定搜索，然后选择*调查结果*。

创建已保存的查询和策略

您可以创建自己的自定义已保存查询，以提供特定于您组织的查询结果。返回符合搜索条件的所有文件和目录（共享和文件夹）的结果。

步骤

1. 在调查选项卡中，通过选择要使用的过滤器来定义搜索。看"在调查页面中过滤数据"了解详情。
2. 一旦您根据自己的喜好设置了所有过滤器特性，请选择*保存查询*。

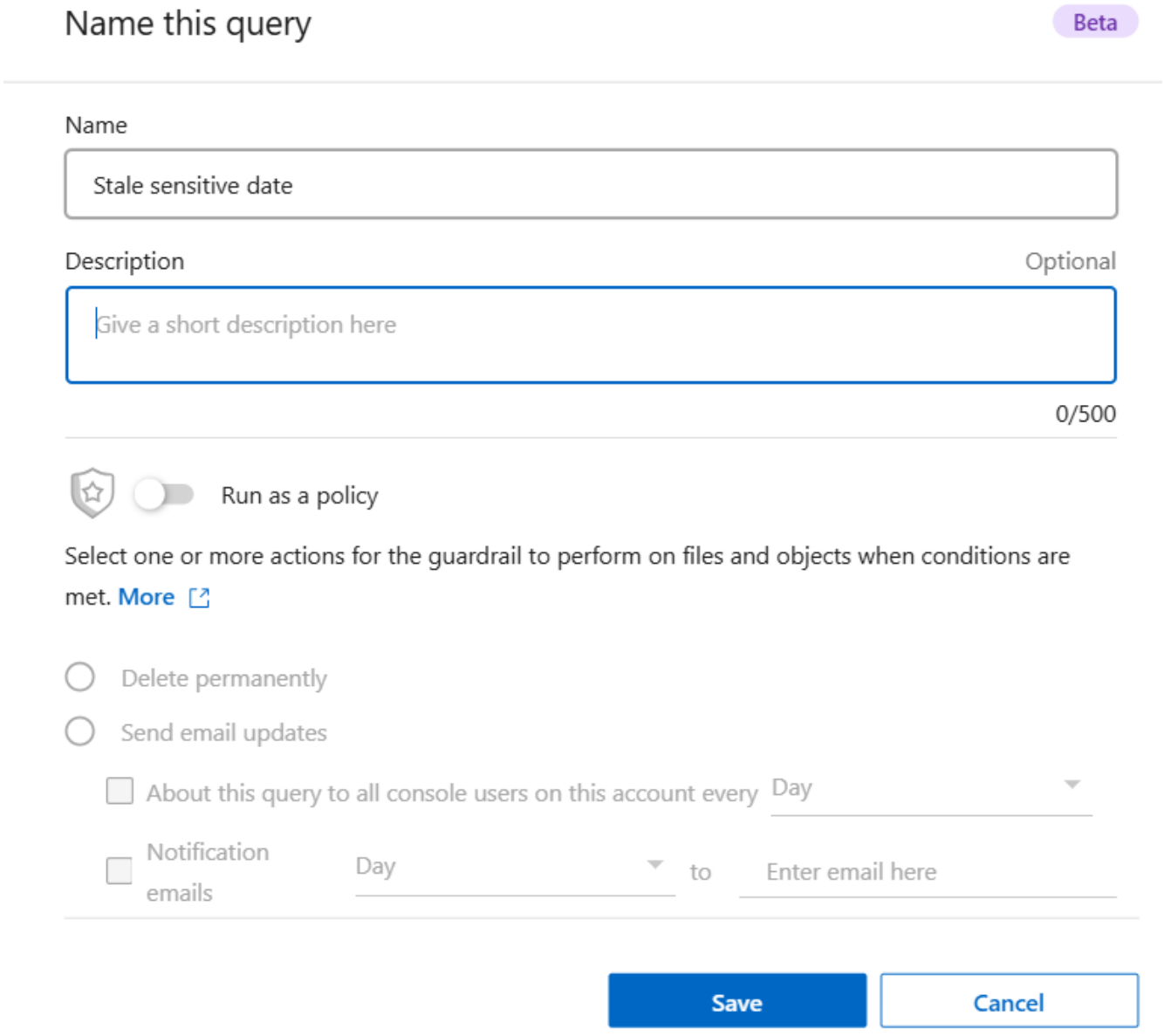


3. 为保存的查询命名并添加描述。该名称必须是唯一的。
4. 您可以选择将查询保存为策略：

a. 要将查询保存为策略，请切换*作为策略运行*开关。

b. 选择*永久删除*或*发送电子邮件更新*。如果您选择电子邮件更新，您可以每天、每周或每月通过电子邮件将查询结果发送给所有控制台用户。或者，您可以以相同的频率将通知发送到特定的电子邮件地址。

5. 选择*保存*。



创建搜索或策略后，您可以在已保存的查询选项卡中查看它。

编辑已保存的查询或策略

您可以修改已保存查询的名称和描述。您还可以将查询转换为策略，反之亦然。

您不能修改默认保存的查询。您不能修改已保存查询的过滤器。您可以交替查看已保存查询的调查结果，更改或修改过滤器，然后将其保存为新查询或策略。

步骤

1. 在“已保存的查询”页面中，选择要更改的搜索的“编辑搜索”。

2. 对名称和描述字段进行更改。仅更改名称和描述字段。

   您可以选择将查询转换为策略，或将策略转换为已保存的查询。根据需要切换*作为策略运行*开关。..如果您要将查询转换为策略，请选择*永久删除*或*发送电子邮件更新*。如果您选择电子邮件更新，您可以每天、每周或每月通过电子邮件将查询结果发送给所有控制台用户。或者，您可以以相同的频率将通知发送到特定的电子邮件地址。

3. 选择“保存”以完成更改。

删除已保存的查询

如果您不再需要任何自定义保存的查询或策略，可以将其删除。您不能删除默认保存的查询。

要删除已保存的查询，请选择 按钮进行特定搜索，选择*删除查询*，然后在确认对话框中再次选择*删除查询*。

默认查询

数据分类提供以下系统定义的搜索查询：

- 数据主体姓名 - 高风险

  包含超过 50 个数据主体名称的文件

- 电子邮件地址 - 高风险

  包含超过 50 个电子邮件地址的文件或数据库列中超过 50% 的行包含电子邮件地址

- 个人数据 - 高风险

  包含超过 20 个个人数据标识符的文件或数据库列中超过 50% 的行包含个人数据标识符

- 私人数据 - 已过期 **7** 年以上

  包含个人或敏感个人信息的文件，上次修改时间超过 7 年

- 保护 - 高

  包含密码、信用卡信息、IBAN 号码或社会安全号码的文件或数据库列

- 保护 - 低

  超过 3 年未访问的文件

- 保护 - 中等

  包含具有个人数据标识符（包括身份证号码、税务识别号、驾驶执照号码、药品 ID 或护照号码）的文件或数据库列的文件

- 敏感个人数据 - 高风险

  包含超过 20 个敏感个人数据标识符的文件或数据库列中超过 50% 的行包含敏感个人数据

更改存储库的NetApp Data Classification扫描设置

您可以管理在每个系统和数据源中如何扫描数据。您可以在“存储库”基础上进行更改；这意味着您可以根据正在扫描的数据源类型对每个卷、模式、用户等进行更改。

您可以更改的一些内容包括是否扫描存储库，以及NetApp Data Classification是否正在执行"映射扫描或映射和分类扫描"。您还可以暂停和恢复扫描，例如，如果您需要在一段时间内停止扫描某个卷。

查看存储库的扫描状态

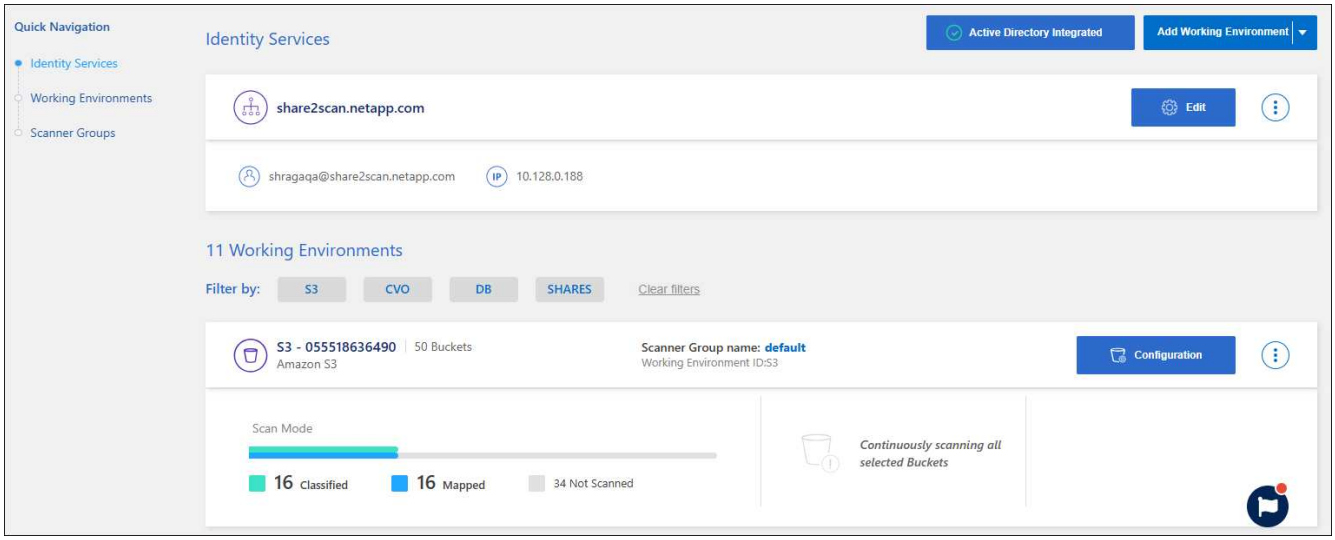
您可以查看NetApp Data Classification正在为每个系统和数据源扫描的各个存储库（卷、存储桶等）。您还可以看到有多少已被“映射”，有多少已被“分类”。分类需要更长的时间，因为所有数据都进行了完整的 AI 识别。

您可以在配置页面查看各个工作环境的扫描状态：

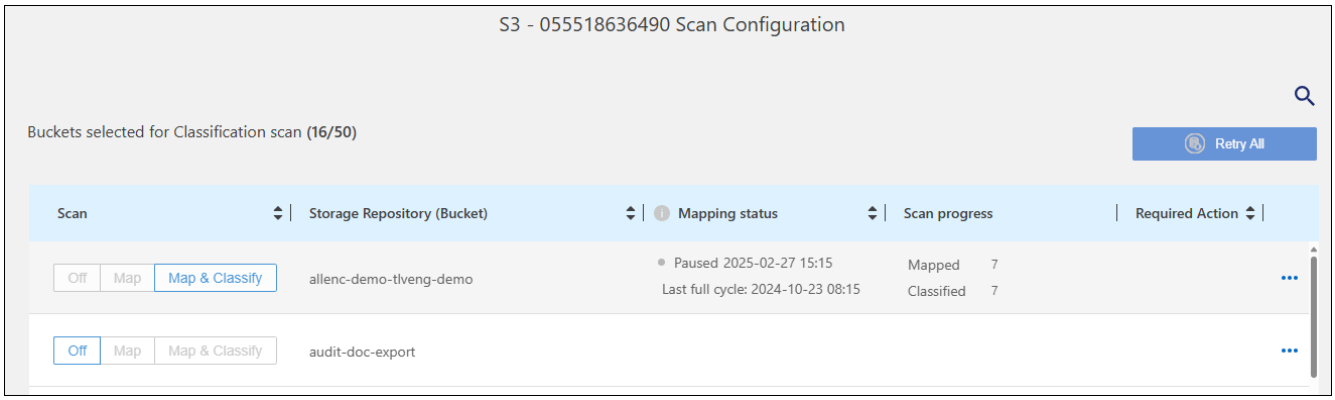
- 初始化（浅蓝色点）：地图或分类配置已激活。此状态会短暂显示，然后过渡到“待处理队列”状态。
- 待处理队列（橙色圆点）：扫描任务正在等待列入扫描队列。
- 已排队（橙色圆点）：任务已成功添加到扫描队列。当队列中的卷轮到达时，系统将开始映射或分类该卷。
- 正在运行（绿点）：队列中的扫描任务正在选定的存储库上积极进行。
- 完成（绿点）：存储库扫描已完成。
- 已暂停（灰点）：您已暂停扫描。虽然系统中未显示音量变化，但扫描结果仍然可用。
- 错误（红点）：扫描无法完成，因为遇到了问题。如果您需要完成某项操作，错误将出现在“所需操作”列下的工具提示中。否则，系统将显示“错误”状态并尝试恢复。完成后，状态就会改变。
- 未扫描：选择了“关闭”卷配置，系统未扫描该卷。

步骤

1. 从数据分类菜单中，选择*配置*。



2. 从配置选项卡中，选择系统的*配置*按钮。
3. 在扫描配置页面中，查看所有存储库的扫描设置。



4. 扫描期间，将光标悬停在“映射状态”列中的进度条上，即可查看该存储库中待映射或分类的文件数量。

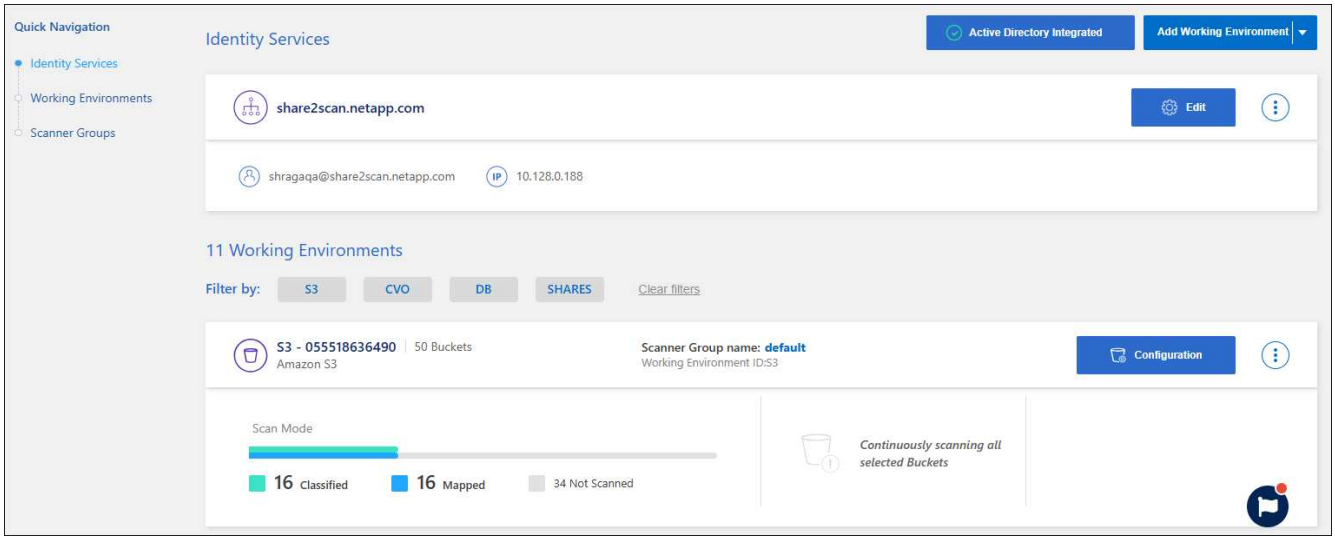
更改存储库的扫描类型

您可以随时从配置页面启动或停止系统中的仅映射扫描或映射和分类扫描。您还可以从仅映射扫描更改为映射和分类扫描，反之亦然。

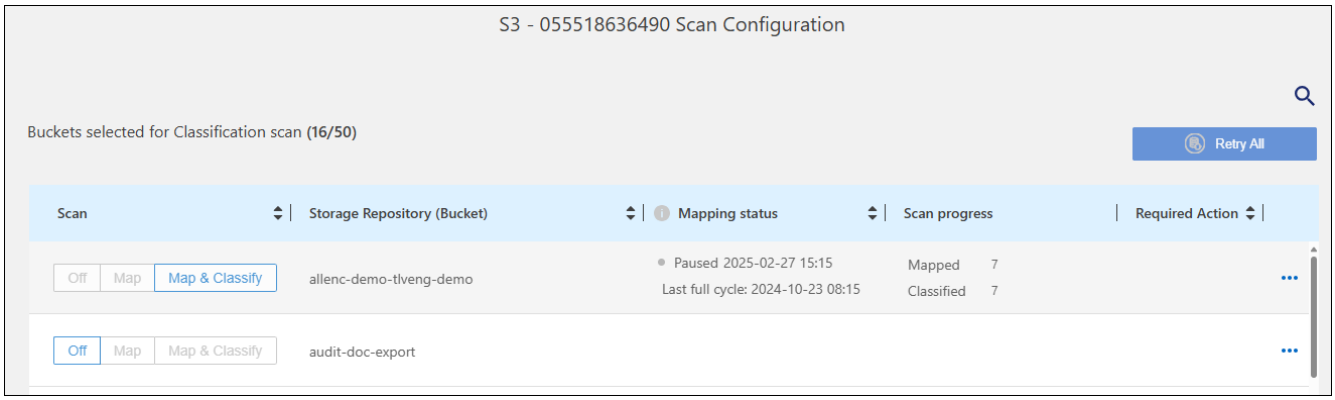
数据库不能设置为仅映射扫描。数据库扫描可以关闭或打开；其中“打开”相当于“映射和分类”。

步骤

1. 从数据分类菜单中，选择*配置*。
2. 从配置选项卡中，选择系统的*配置*按钮。

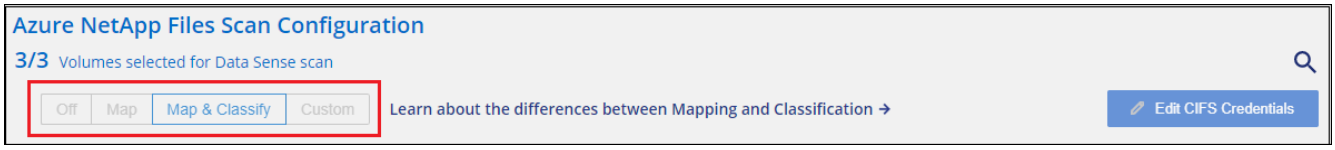


3. 在扫描配置页面中，更改任何存储库（本例中为存储桶）以执行*Map*或*Map & Classify*扫描。



某些类型的系统允许您使用页面顶部的按钮栏全局更改所有存储库的扫描类型。这对于Cloud Volumes ONTAP、本地ONTAP、 Azure NetApp Files和Amazon FSx for ONTAP系统有效。

下面的示例显示了Azure NetApp Files系统的按钮栏。



优先扫描

您可以优先考虑最重要的仅映射扫描或映射和分类扫描，以确保高优先级扫描首先完成。

默认情况下，扫描按照启动的顺序排队。通过设置扫描优先级，您可以将扫描移至队列的最前面。可以对多个扫描进行优先排序。优先级按先进先出的顺序指定，这意味着您优先考虑的第一个扫描将移至队列的最前面；您优先考虑的第二个扫描将成为队列中的第二个扫描，依此类推。

优先权是一次性授予的。映射数据的自动重新扫描按照默认顺序进行。

步骤

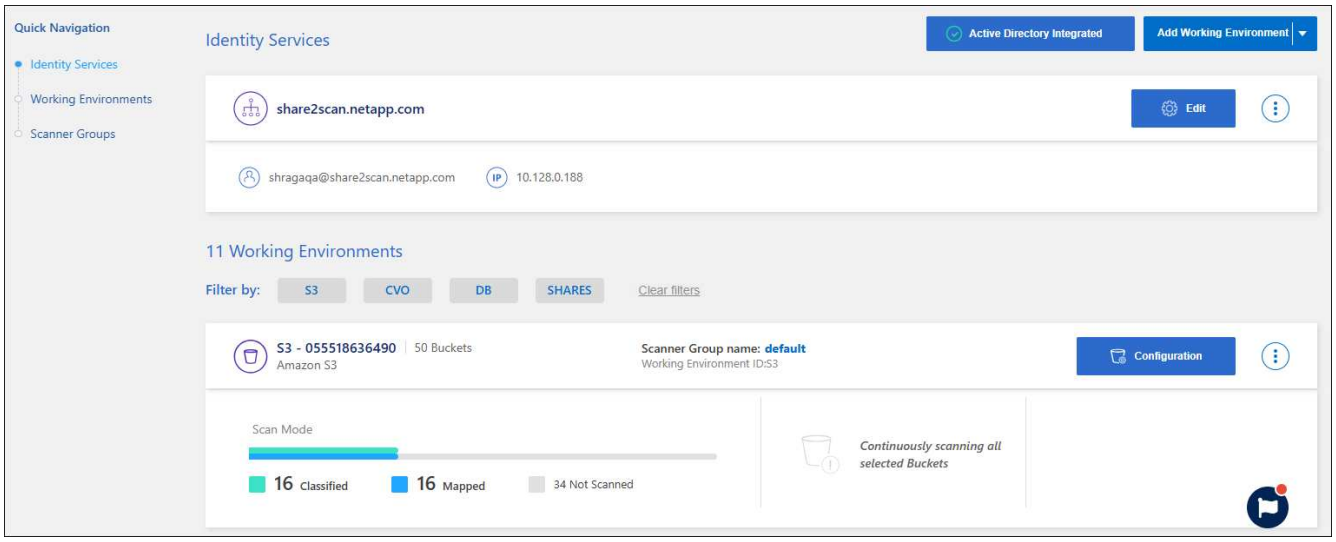
1. 从数据分类菜单中，选择*配置*。
2. 选择您想要优先考虑的资源。
3. 从行动 `…`选项，选择*优先扫描*。

停止扫描存储库

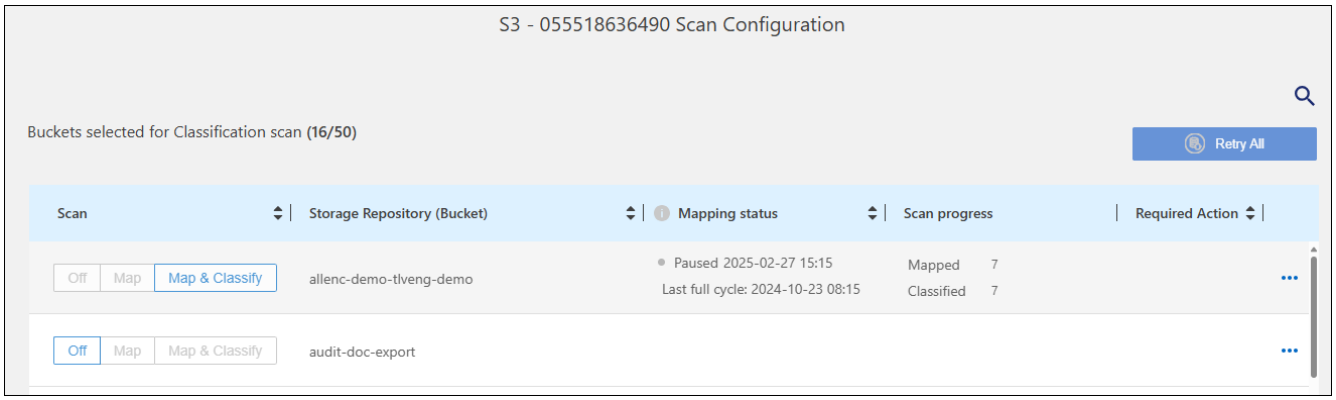
如果您不再需要监控存储库（例如卷）的合规性，则可以停止扫描它。您可以通过关闭扫描来实现此目的。当扫描关闭时，有关该卷的所有索引和信息都将从系统中删除，并且扫描数据的收费也将停止。

步骤

1. 从数据分类菜单中，选择*配置*。
2. 从配置选项卡中，选择系统的*配置*按钮。



3. 在扫描配置页面中选择“关闭”以停止扫描特定存储桶。



暂停并恢复存储库扫描

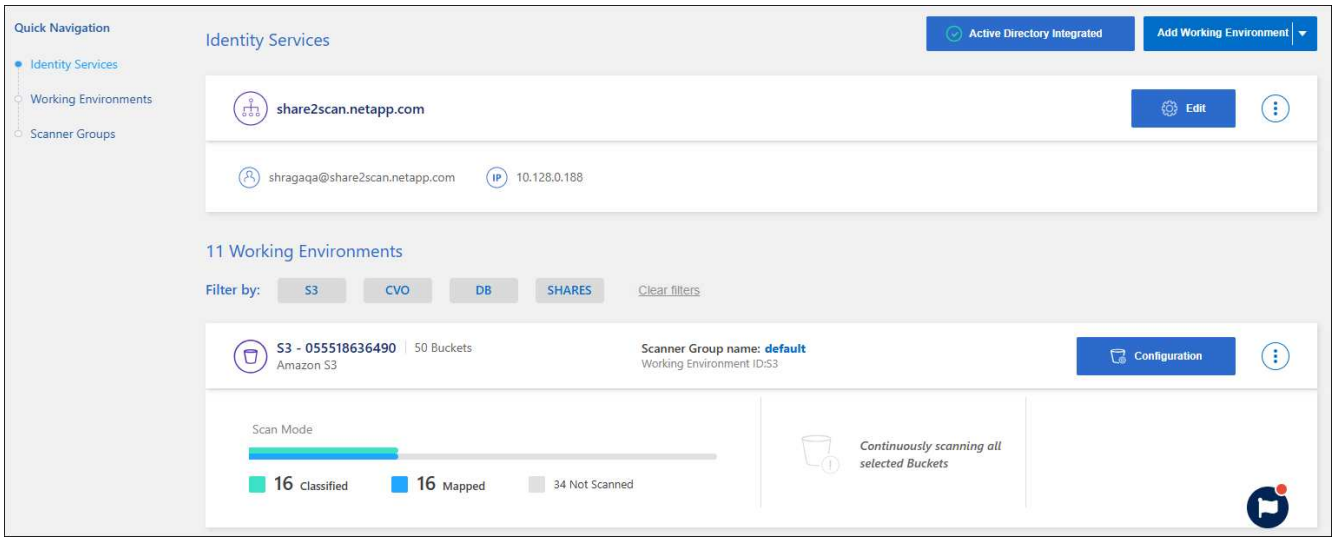
如果您想暂时停止扫描某些内容，您可以“暂停”存储库扫描。暂停扫描意味着数据分类将不再对存储库中的更改或添加执行任何未来的扫描。所有当前的扫描结果仍可在数据分类中查看。

即使暂停扫描，也不会免除计费费用，因为数据仍然保留在系统中。

您可以随时恢复扫描。

步骤

1. 从数据分类菜单中，选择*配置*。
2. 从配置选项卡中，选择系统的*配置*按钮。



3. 在扫描配置页面中，选择操作 ••• 图标。
4. 选择“暂停”暂停对卷的扫描，或选择“恢复”恢复对先前已暂停的卷的扫描。

查看NetApp Data Classification合规性报告

NetApp Data Classification提供报告，您可以使用这些报告更好地了解组织的数据隐私计划的状态。

默认情况下，数据分类仪表板显示所有系统、数据库和数据源的合规性和治理数据。如果您想要查看仅包含部分系统数据的报告，您可以进行筛选以仅查看这些系统的数据。

- 仅当您对数据源执行完整分类扫描时，才可获得合规性报告。已进行仅映射扫描的数据源只能生成数据映射报告。
- NetApp无法保证数据分类识别的个人数据和敏感个人数据 100% 的准确性。您应该始终通过查看数据来验证信息。

以下报告可用于数据分类：

- 数据发现评估报告：对扫描环境进行高级分析，以突出系统的发现并显示关注领域和潜在的补救步骤。此报告可在治理仪表板中找到。
- 完整数据映射概览报告：提供有关系统中文件的大小和数量的信息。这包括使用容量、数据年限、数据大小和文件类型。此报告可在治理仪表板中找到。
- 数据主体访问请求报告：使您能够提取包含有关数据主体的特定名称或个人标识符信息的所有文件的报告。此报告可在合规性仪表板中找到。
- **HIPAA** 报告：帮助您识别文件中健康信息的分布。此报告可在合规性仪表板中找到。

- **PCI DSS** 报告：帮助您识别文件中信用卡信息的分布。此报告可在合规性仪表板中找到。
- 隐私风险评估报告：提供来自您的数据的隐私见解和隐私风险评分。此报告可在合规性仪表板中找到。
- 特定信息类型的报告：可提供包含已识别文件（包含个人数据和敏感个人数据）详细信息的报告。您还可以查看按类别和文件类型细分的文件。

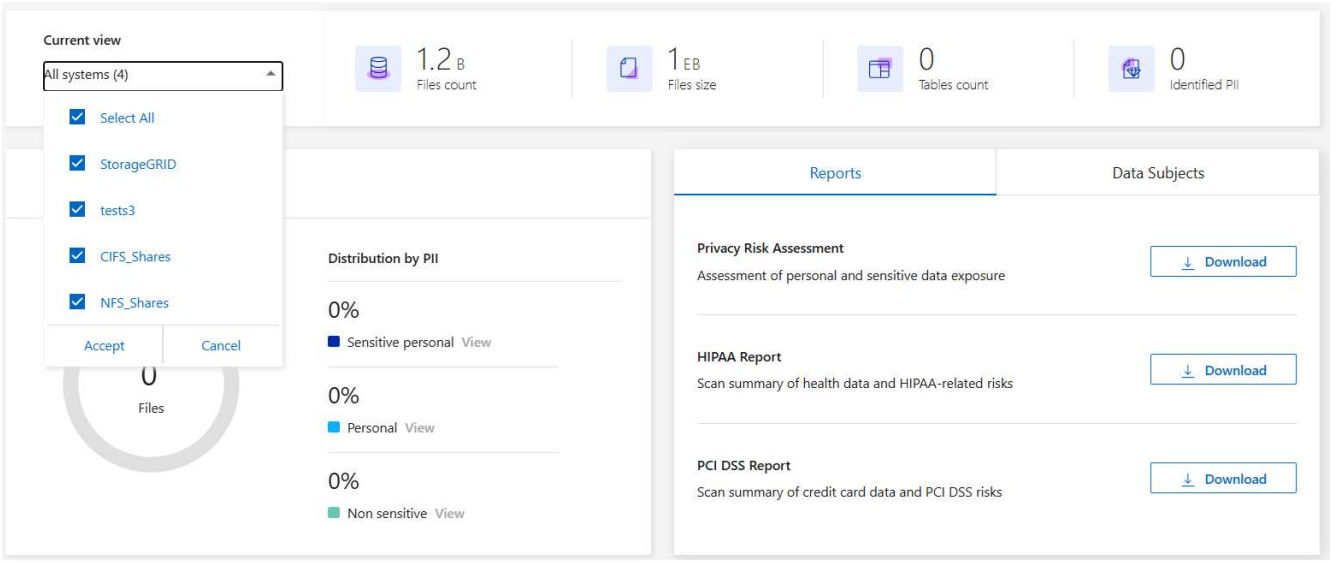
选择报告系统

您可以过滤数据分类合规性仪表板的内容，以查看所有系统和数据库的合规性数据，或仅查看特定系统的合规性数据。

当您过滤仪表板时，数据分类会将合规性数据和报告范围限定到您选择的系统。

步骤

1. 从数据分类菜单中，选择*合规性*。
2. 选择系统过滤器下拉菜单，然后选择系统。
3. 选择接受来确认您的选择。



数据主体访问请求报告

欧洲 GDPR 等隐私法规赋予数据主体（例如客户或员工）访问其个人数据的权利。当数据主体请求此信息时，这被称为 DSAR（数据主体访问请求）。各组织必须“毫不拖延”地回应这些请求，最迟不得超过收到请求后的一个月。

您可以通过搜索主题的全名或已知标识符（例如电子邮件地址）然后下载报告来回应 DSAR。该报告旨在帮助您的组织遵守 GDPR 或类似的数据隐私法。

数据分类如何帮助您响应 **DSAR**？

当您执行数据主体搜索时，数据分类会找到包含该人姓名或标识符的所有文件。数据分类检查最新的预索引数据的名称或标识符。它不会启动新的扫描。

搜索完成后，您可以下载数据主体访问请求报告的文件列表。该报告汇总了数据中的见解，并将其转化为法律术

语，以便您可以将其发送给相关人员。

目前数据库不支持数据主体搜索。

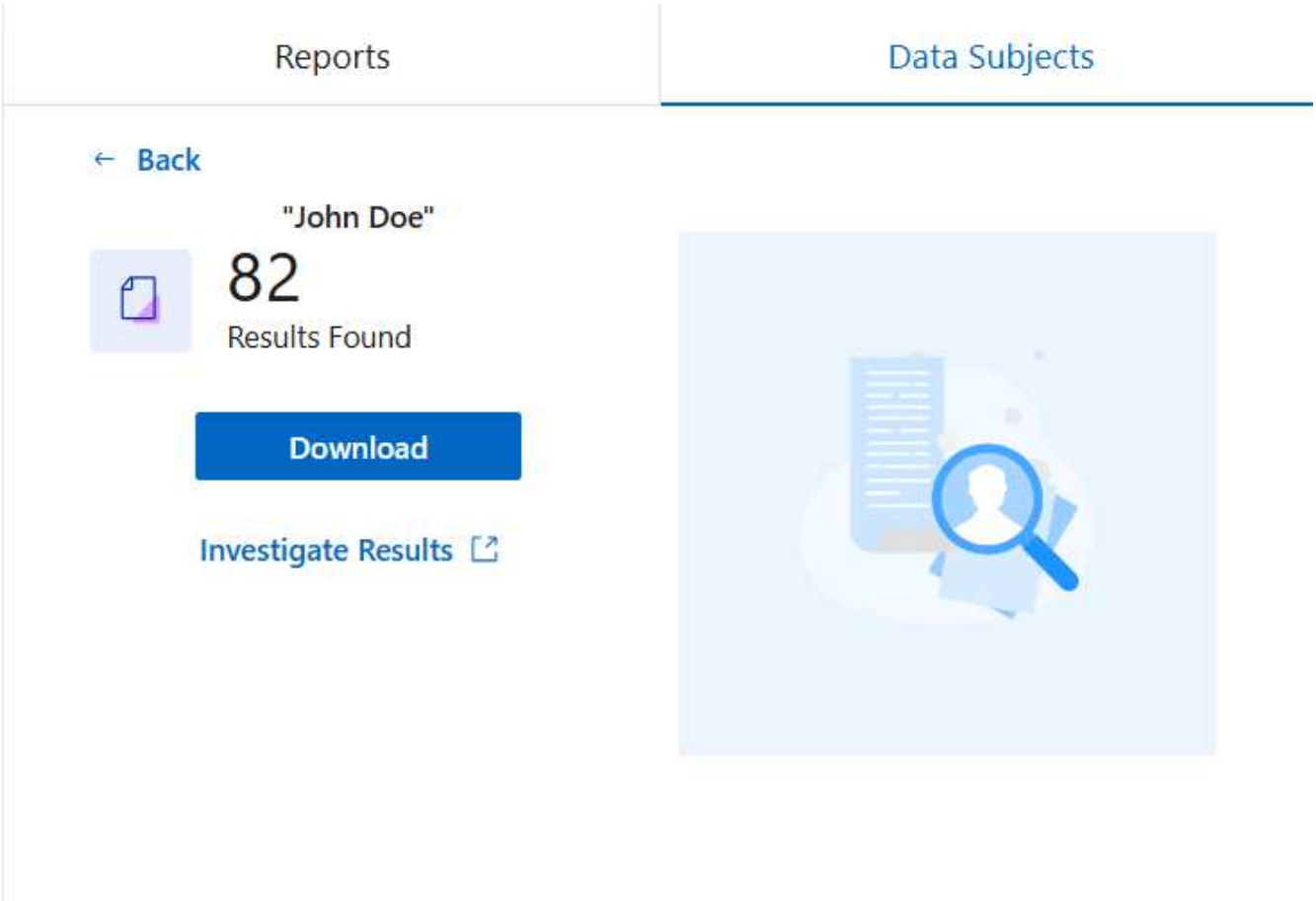
搜索数据主体并下载报告

搜索数据主体的全名或已知标识符，然后下载文件列表报告或 DSAR 报告。您可以通过以下方式搜索"任何个人信息类型"。

搜索数据主体的姓名时支持英语、德语、日语和西班牙语。稍后将添加对更多语言的支持。

步骤

1. 从数据分类菜单中，选择*合规性*。
2. 在合规性页面中，找到数据主体选项卡。
3. 在“数据主体”部分，输入名称或已知标识符，然后选择“搜索”。
4. 搜索完成后，选择下载以访问数据主体访问请求响应。选择调查结果以在数据调查页面中查看更多信息。



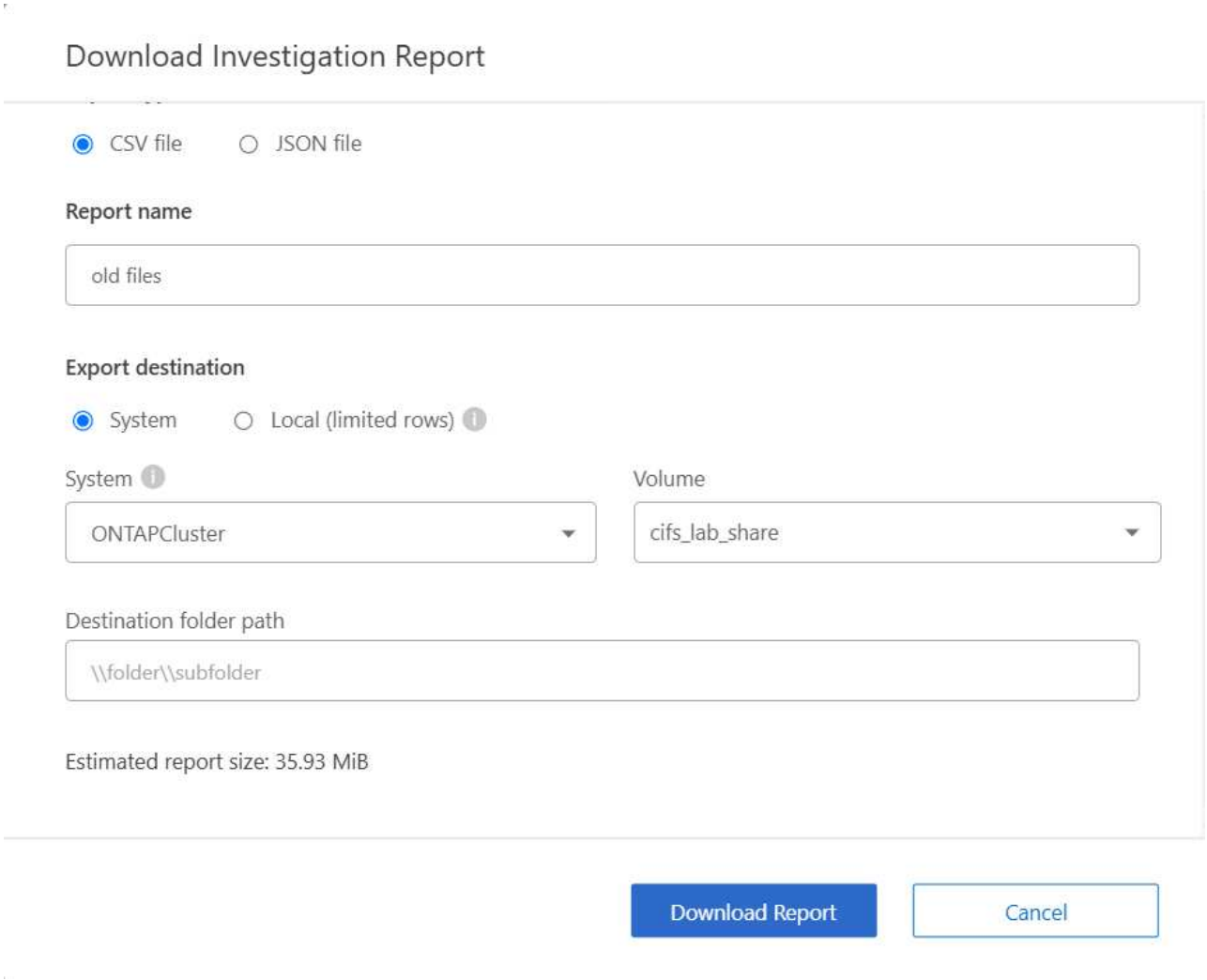
5. 查看数据分类中的结果或通过选择下载图标将其下载为报告。
   a. 选择下载图标后，配置您的下载设置：
      - 选择影片格式：CSV 或 JSON
      - 输入*报告名称*

- 选择导出目的地：*系统*或您的*本地*机器。

  如果您选择系统，则会下载所有数据。您还必须选择*系统*、卷*和*目标文件夹路径。

如果您选择*本地*，则会将报告限制为前 10,000 行非结构化数据；5,000 行非结构化数据和 1,000 行结构化数据。

a. 选择下载报告开始下载。



健康保险流通与责任法案（HIPAA）报告

健康保险流通与责任法案 (HIPAA) 报告可以帮助您识别包含健康信息的文件。它旨在帮助您的组织遵守 HIPAA 数据隐私法的要求。数据分类寻找的信息包括：

- 健康参考模式
- ICD-10-CM 医疗代码
- ICD-9-CM 医疗代码
- HR - 健康类别
- 健康应用数据类别

该报告包含以下信息：

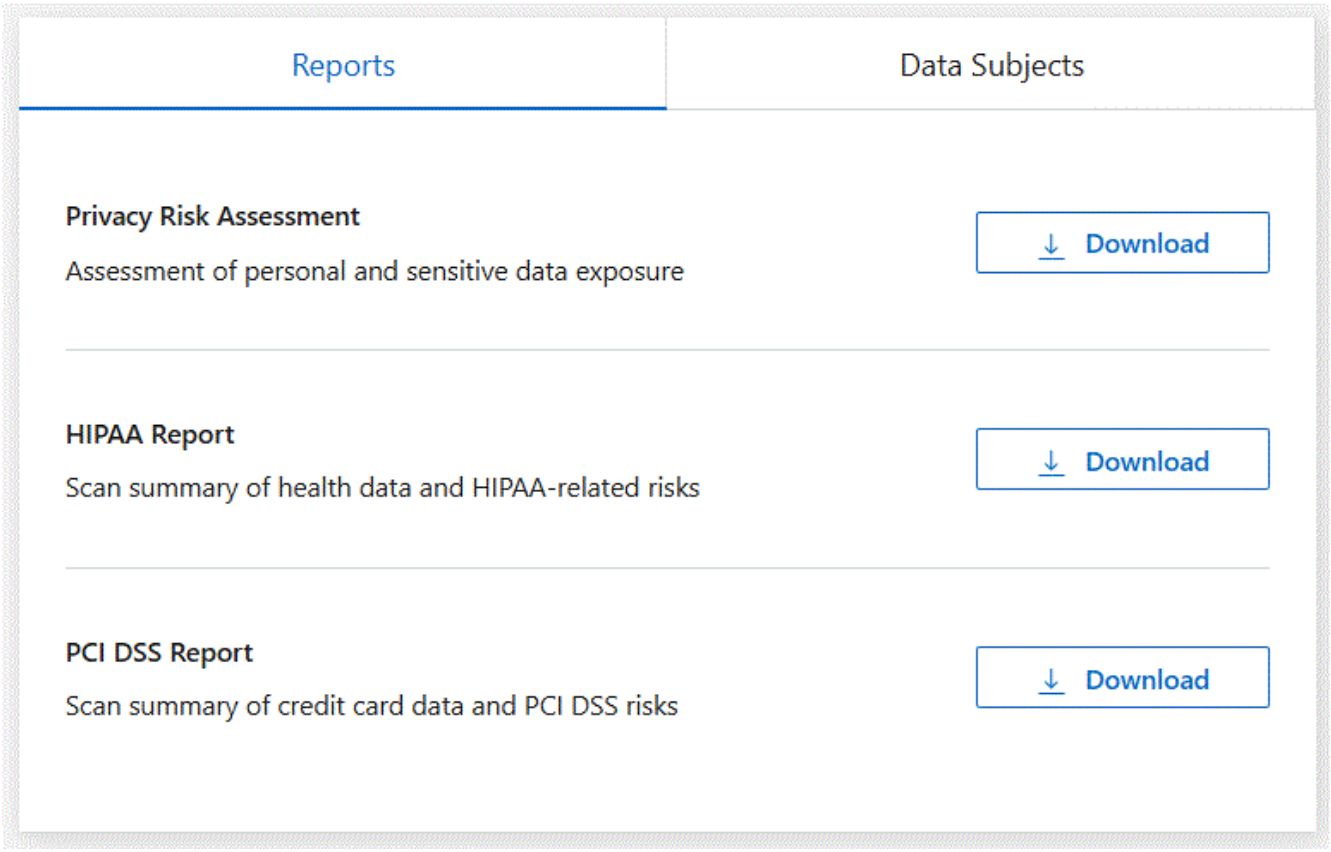
- 概述：有多少文件包含健康信息以及在哪些系统中。
- 加密：加密或未加密系统中包含健康信息的文件的百分比。此信息特定于Cloud Volumes ONTAP。
- 勒索软件防护：在启用或未启用勒索软件防护的系统上，包含健康信息的文件的百分比。此信息特定于Cloud Volumes ONTAP。
- 保留：文件最后修改的时间范围。这很有用，因为您不应该将健康信息保存超过处理所需的时间。
- 健康信息分发：发现健康信息的系统以及是否启用了加密和勒索软件保护。

生成 **HIPAA** 报告

转到“合规性”选项卡以生成报告。

步骤

1. 从数据分类菜单中，选择*合规性*。
2. 找到报告窗格。选择*HIPAA 报告*旁边的下载图标。



结果

数据分类生成 PDF 报告。

支付卡行业数据安全标准 (PCI DSS) 报告

支付卡行业数据安全标准 (PCI DSS) 报告可以帮助您识别信用卡信息在文件中的分布。

该报告包含以下信息：

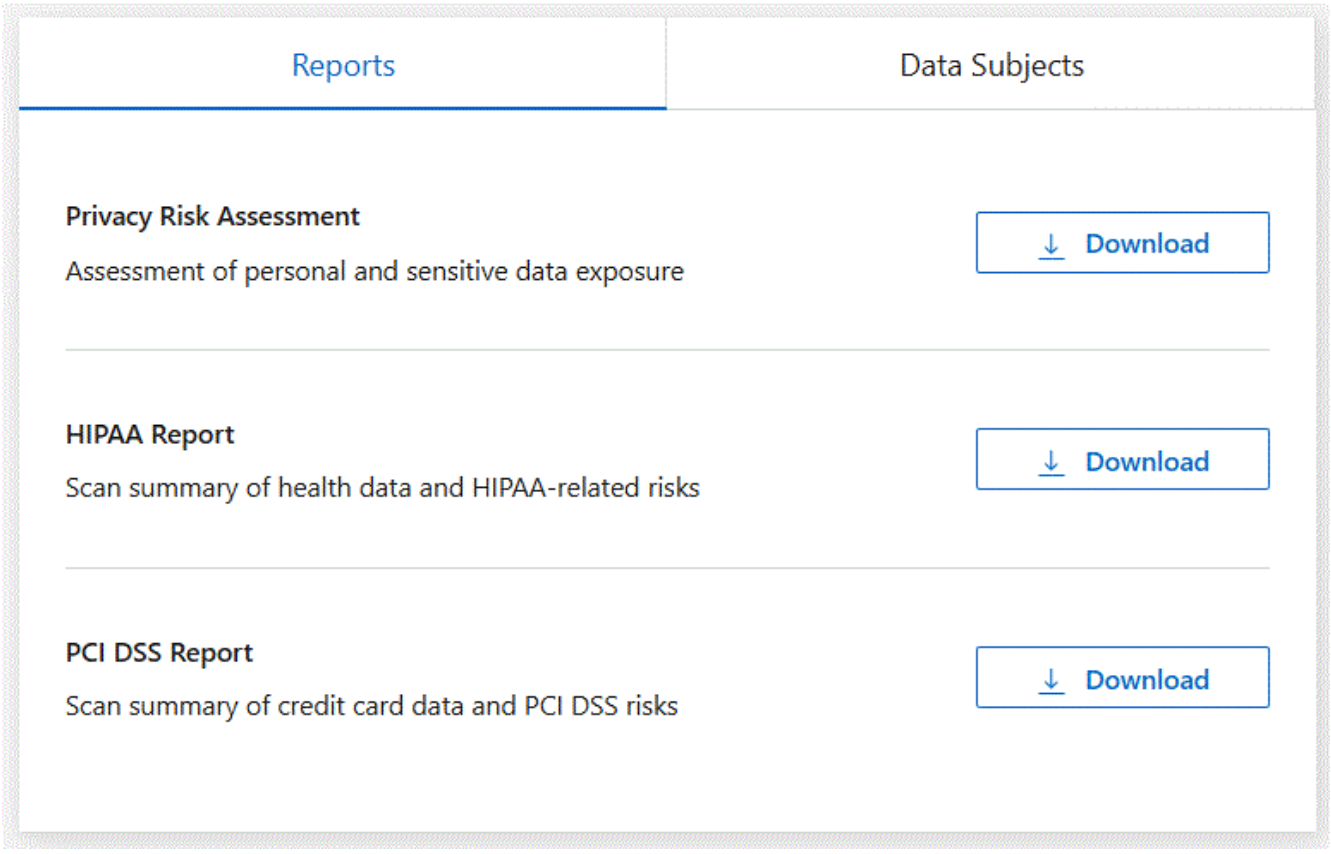
- 概述：有多少个文件包含信用卡信息以及在哪些系统中。
- 加密：加密或未加密系统中包含信用卡信息的文件的百分比。此信息特定于Cloud Volumes ONTAP。
- 勒索软件防护：在启用或未启用勒索软件防护的系统上，包含信用卡信息的文件的百分比。此信息特定于Cloud Volumes ONTAP。
- 保留：文件最后修改的时间范围。这很有用，因为您不应该将信用卡信息保存的时间超过处理所需的时间。
- 信用卡信息分发：发现信用卡信息的系统以及是否启用了加密和勒索软件保护。

生成 **PCI DSS** 报告

转到“合规性”选项卡以生成报告。

步骤

1. 从数据分类菜单中，选择*合规性*。
2. 找到报告窗格。选择*PCI DSS 报告*旁边的下载图标。



结果

数据分类会生成一份 PDF 报告，您可以根据需要查看并发送给其他组。

隐私风险评估报告

隐私风险评估报告概述了您组织的隐私风险状况，这是 GDPR 和 CCPA 等隐私法规所要求的。

该报告包含以下信息：

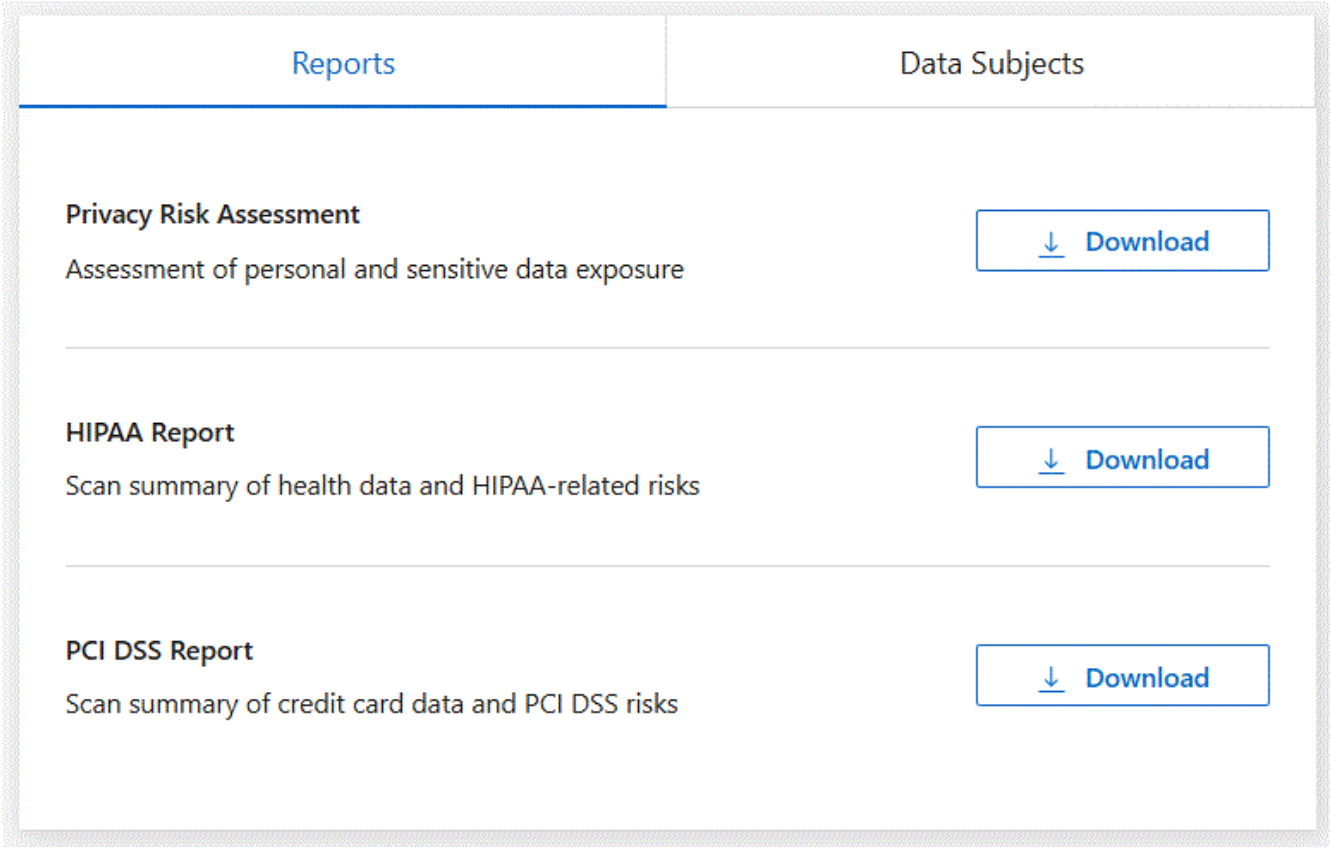
- 合规状态：严重性评分和数据分布，无论是非敏感数据、个人数据还是敏感个人数据。
- 评估概述：发现的个人数据类型以及数据类别的细分。
- 本次评估中的数据主体：按地点划分的已找到国家标识符的人数。

生成隐私风险评估报告

转到“合规性”选项卡以生成报告。

步骤

1. 从数据分类菜单中，选择*合规性*。
2. 找到报告窗格。选择*隐私风险评估报告*旁边的下载图标。



结果

数据分类会生成一份 PDF 报告，您可以根据需要查看并发送给其他组。

严重程度评分

数据分类根据三个变量计算隐私风险评估报告的严重性分数：

- 个人数据占所有数据的百分比。
- 敏感个人数据占所有数据的比例。
- 包含数据主体的文件百分比，由国家标识符（例如国民身份证、社会安全号码和税号）决定。

确定分数的逻辑如下：

严重程度评分	逻辑
0	所有三个变量都恰好为 0%
1	其中一个变量大于 0%
2	其中一个变量大于3%
3	其中两个变量大于 3%
4	其中三个变量大于 3%
5	其中一个变量大于6%
6	其中两个变量大于 6%
7	其中三个变量大于 6%
8	其中一个变量大于15%
9	其中两个变量大于 15%
10	其中三个变量大于 15%

监控NetApp Data Classification的运行状况

NetApp Data Classification健康监视器仪表板提供实时监控和性能洞察。健康监视器会捕获有关您的数据分类基础架构、系统运行状况、使用指标和利用率数据的信息，使您能够识别和解决问题。

健康监测洞察

健康监测仪表盘以四类信息呈现信息。

- 基础设施状况

  查看版本状态、系统稳定性、部署类型和机器规模等信息。

- 问题容器

  查看“问题容器”字段，以了解哪些容器已停止或频繁重启。利用这些信息调查具体的容器。

- 系统信息

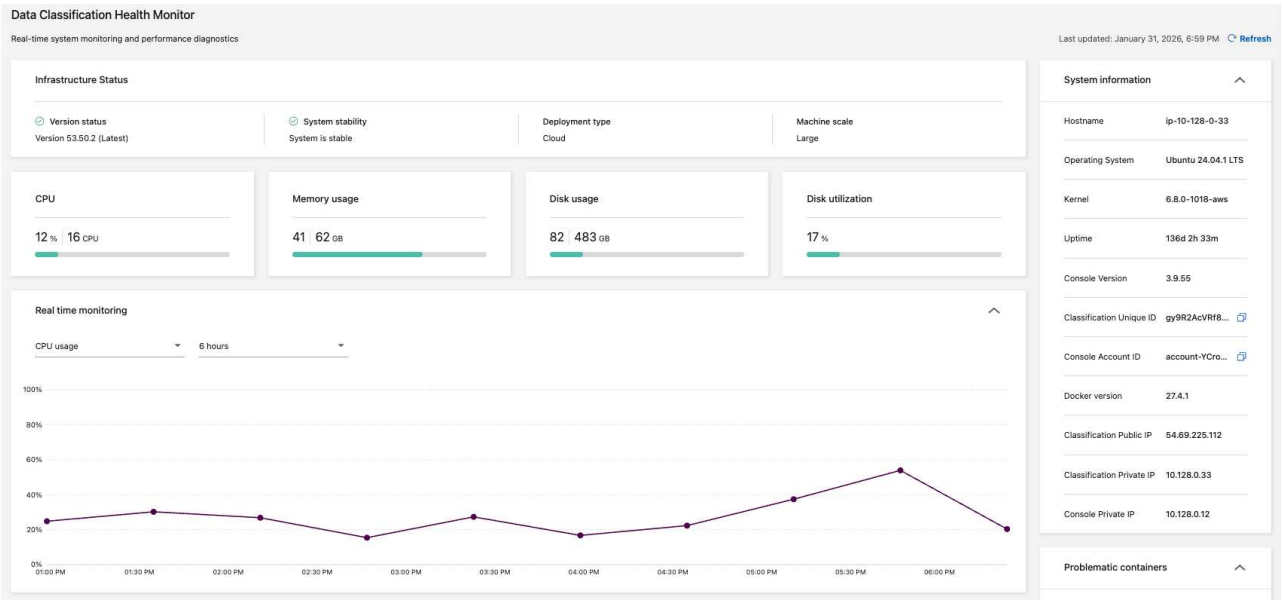
  系统信息面板捕获有关NetApp Console和数据分类的关键信息，例如公共和私有 IP 地址、主机名、操作系统、控制台版本和控制台 ID。

- 用途和使用方法

  查看 CPU 使用率、磁盘利用率、磁盘使用率和内存使用率。这些值以存储单位（GB）或总使用量的百分比显示。如果任何字段显示警告，请选择该警告以获取相关信息和补救建议。

访问健康监测仪表板

1. 在数据分类中，选择配置。
2. 在“配置”标题下，选择“数据分类运行状况监视器”。
3. 在健康监测仪表板中，您可以：
   - 审查使用情况和利用情况。如果任何使用情况或利用率指标显示警告，请选择该警告以获取解决问题的建议。
   - 切换图表以显示 CPU 使用率、磁盘利用率、磁盘使用率和内存使用率。您可以更改 x 轴，以按小时（6、12 或 24 小时）或天（2、7 或 14 天）显示内容。
   - 刷新仪表盘以查看最新数据指标。



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

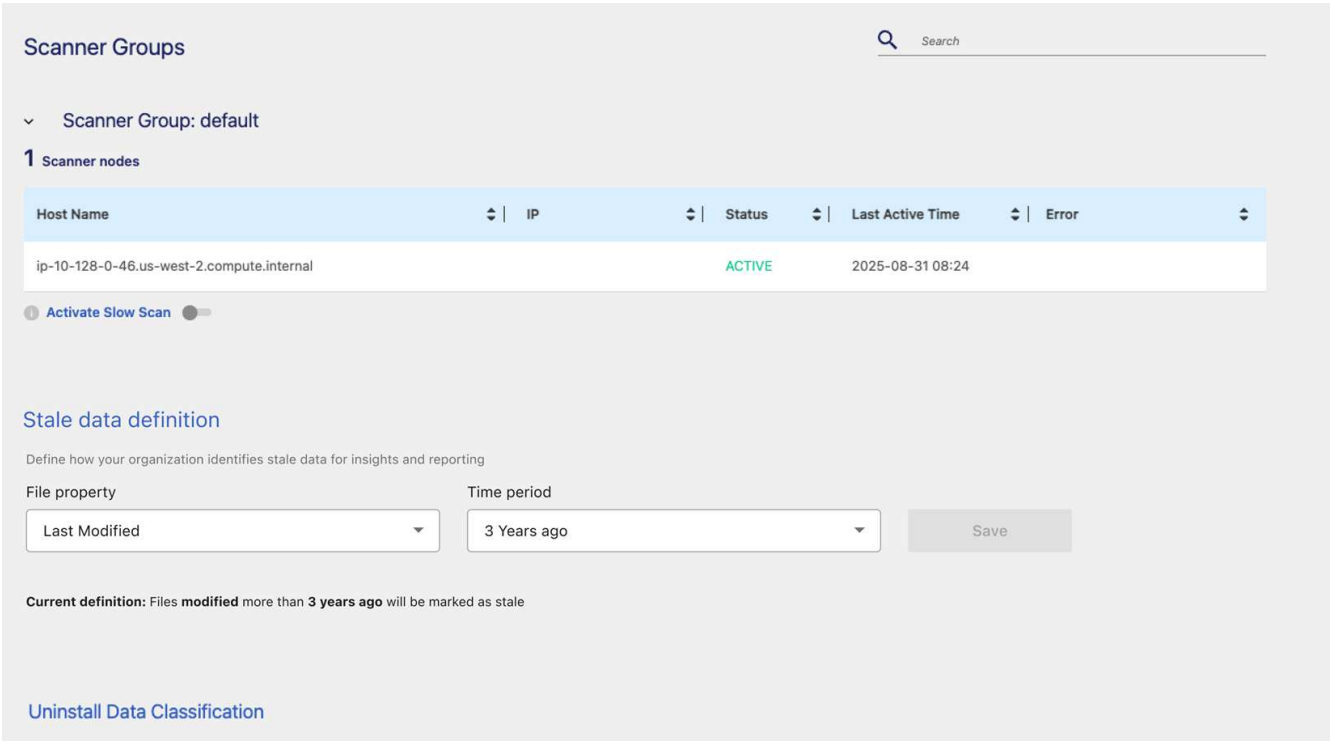
NetApp Data Classification可识别过时数据，帮助您发现节省成本的机会和治理风险。由于不同组织环境中对过时数据的定义可能有所不同，因此您可以自定义数据分类如何定义过时数据。

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。



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。

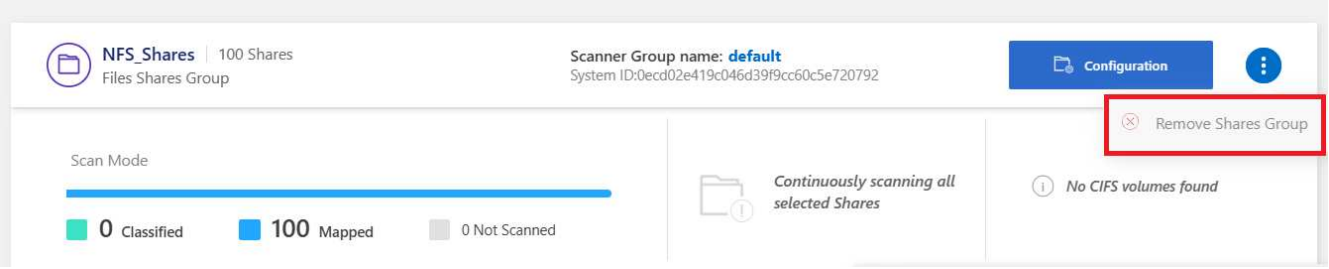
1. 从“配置”页面中，选择 数据库行中的按钮，然后*删除数据库服务器*。

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*。

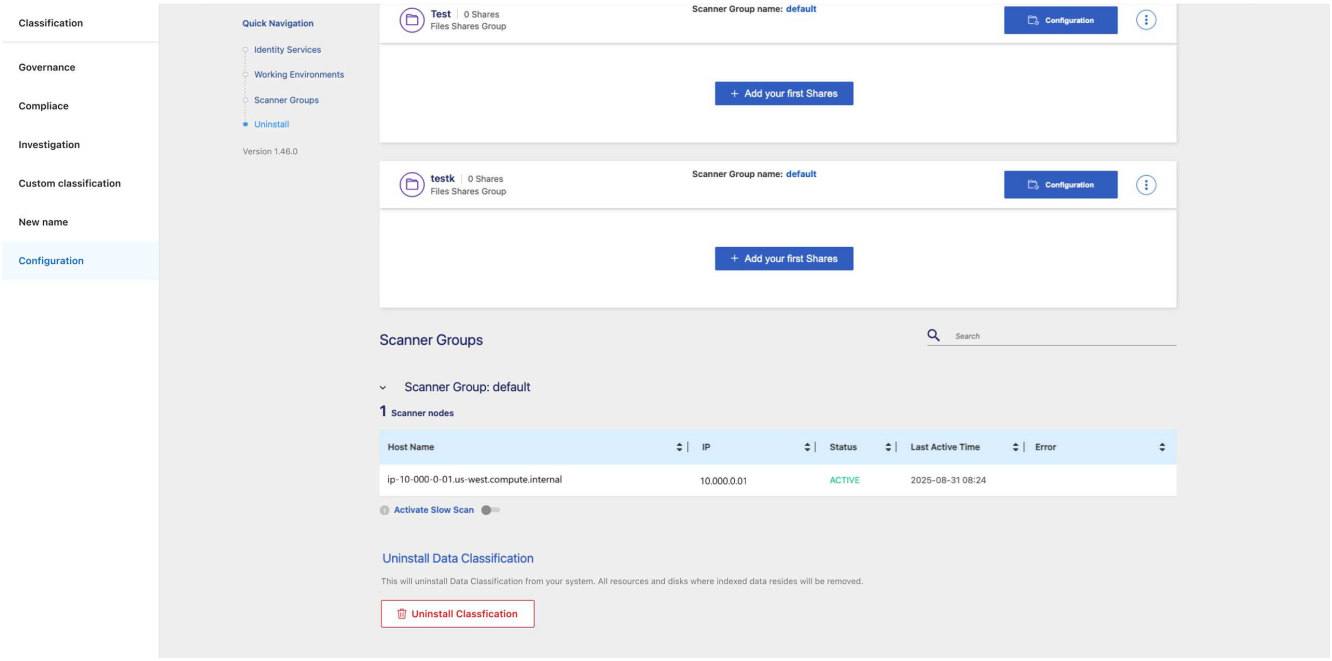
卸载NetApp Data Classification

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

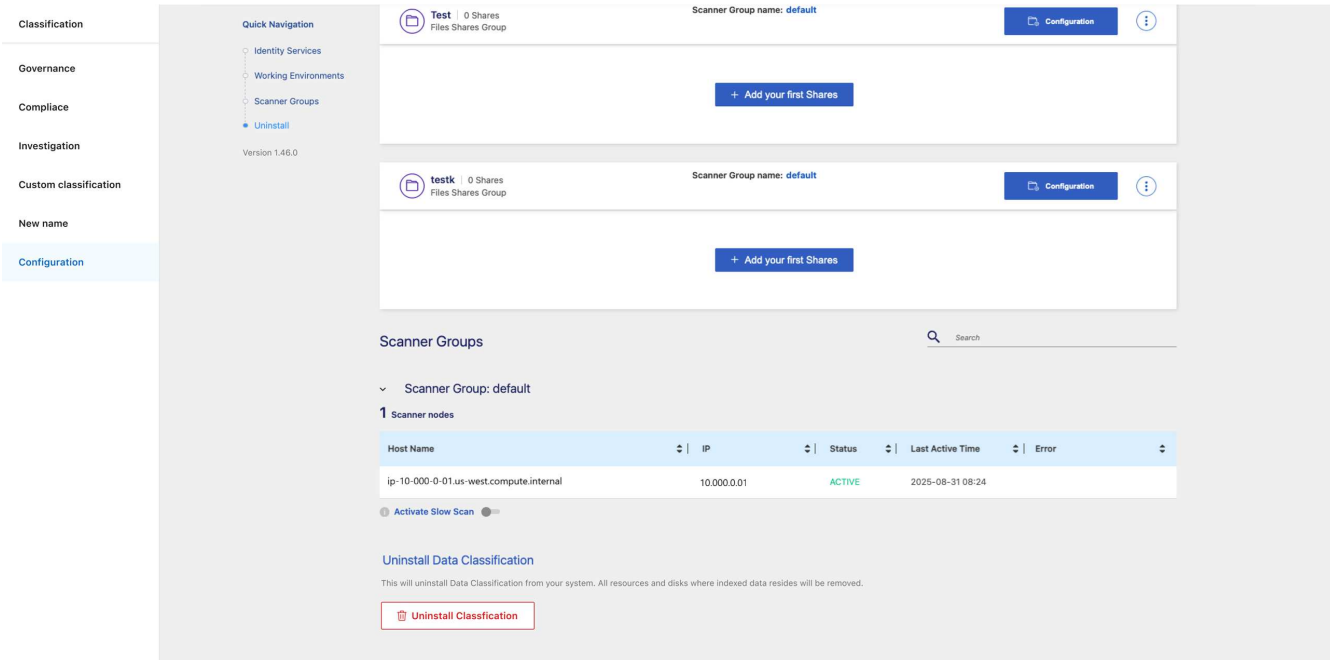
1. 从数据分类中，选择配置。
2. 在配置页面底部，选择卸载分类。



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。



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"。

参考

支持的NetApp Data Classification实例类型

NetApp Data Classification软件必须在满足特定操作系统要求、RAM 要求、软件要求等的主机上运行。在云中部署数据分类时，我们建议您使用具有“大型”特性的系统以实现全部功能。

您可以在具有较少 CPU 和较少 RAM 的系统上部署数据分类，但在使用这些功能较弱的系统时会受到一些限制。"了解这些限制" 。

在下表中，如果标记为“默认”的系统在您安装数据分类的区域中不可用，则将部署表中的下一个系统。

AWS 实例类型

系统大小	规格	实例类型
特大号	32 个 CPU、128 GB RAM、1 TiB gp3 SSD	"m6i.8xlarge"（默认）
大型	16 个 CPU、64 GB RAM、500 GiB SSD	"m6i.4xlarge"（默认）m6a.4xlarge m5a.4xlarge m5.4xlarge m4.4xlarge
中	8 个 CPU、32 GB RAM、200 GiB SSD	"m6i.2xlarge"（默认）m6a.2xlarge m5a.2xlarge m5.2xlarge m4.2xlarge
小型	8 个 CPU、16 GB RAM、100 GiB SSD	"c6a.2xlarge"（默认）c5a.2xlarge c5.2xlarge c4.2xlarge

Azure 实例类型

系统大小	规格	实例类型
特大号	32 个 CPU、128 GB RAM、OS 磁盘（2,048 GiB，最小吞吐量 250 MB/s）和数据磁盘（1 TiB SSD，最小吞吐量 750 MB/s）	"Standard_D32_v3"（默认）
大型	16 个 CPU、64 GB RAM、500 GiB SSD	"Standard_D16s_v3"（默认）

GCP 实例类型

系统大小	规格	实例类型
大型	16 个 CPU、64 GB RAM、500 GiB SSD	"n2-标准-16"（默认）n2d-standard-16 n1-standard-16

从NetApp Data Classification中的数据源收集的元数据

NetApp Data Classification在对来自数据源和系统的数据执行分类扫描时收集某些元数据。数据分类可以访问我们对数据进行分类所需的大部分元数据，但有些来源我们无法访问所需的数据。

	元数据	CIFS	NFS
时间戳	创建时间	可用	不可用（Linux 不支持）
	上次访问时间	可用	可用
	上次修改时间	可用	可用
权限	开放权限	如果“EVERYONE”组有权访问该文件，则视为“对组织开放”	如果“其他人”有权访问该文件，则视为“对组织开放”
	用户/组访问	用户和组信息取自 LDAP	不可用（NFS 用户通常在服务器本地进行管理，因此同一个人在每个服务器上可以有不同的 UID）

- 数据分类不会从数据库数据源中提取“上次访问时间”。
- 旧版本的 Windows 操作系统（例如 Windows 7 和 Windows 8）默认禁用“上次访问时间”属性的收集，因为它会影响系统性能。当未收集此属性时，基于“上次访问时间”的数据分类分析将受到影响。如果需要，您可以在这些较旧的 Windows 系统上启用上次访问时间的收集。

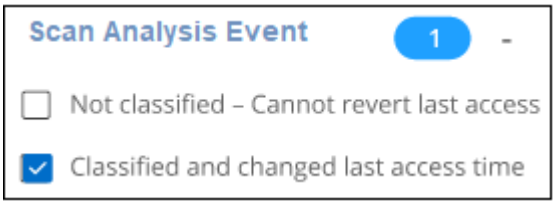
上次访问时间时间戳

当数据分类从文件共享中提取数据时，操作系统会将其视为访问数据，并相应地更改“上次访问时间”。扫描后，数据分类会尝试将上次访问时间恢复为原始时间戳。如果数据分类在 CIFS 中没有写入属性权限，或者在 NFS 中没有写入权限，则系统无法将上次访问时间恢复为原始时间戳。配置了SnapLock的ONTAP卷具有只读权限，并且无法将上次访问时间恢复为原始时间戳。

默认情况下，如果数据分类没有这些权限，系统将不会扫描卷中的这些文件，因为数据分类无法将“上次访问时间”恢复为原始时间戳。但是，如果您不介意文件中的最后访问时间是否重置为原始时间，则可以选择配置页面底部的*缺少“写入属性”权限时扫描*开关，以便数据分类无论权限如何都会扫描卷。

此功能适用于本地ONTAP系统、 Cloud Volumes ONTAP、 Azure NetApp Files、 Amazon FSx for NetApp ONTAP管理和第三方文件共享。

调查页面中有一个名为“扫描分析事件”的过滤器，它使您能够显示未分类的文件（因为数据分类无法恢复上次访问时间），或者即使数据分类无法恢复上次访问时间也已分类的文件。



过滤器选择包括：

- “未分类 - 无法恢复上次访问时间” - 这显示由于缺少写入权限而未分类的文件。
- “已分类并更新的上次访问时间” - 这显示已分类的文件，并且数据分类无法将上次访问时间重置回原始日期。此过滤器仅与您打开*缺少“写入属性”权限时扫描*的环境相关。

如果需要，您可以将这些结果导出到报告中，以便查看哪些文件由于权限原因而被扫描，哪些文件未被扫描。"了解有关数据调查报告的更多信息"。

登录NetApp Data Classification系统

您需要登录NetApp Data Classification系统，以便访问日志文件或编辑配置文件。

当数据分类安装在您本地的 Linux 机器上或在云中部署的 Linux 机器上时，您可以直接访问配置文件和脚本。

当数据分类部署在云中时，您需要通过 SSH 连接到数据分类实例。您可以通过输入用户和密码，或使用控制台代理安装期间提供的 SSH 密钥，通过 SSH 连接到系统。 SSH 命令是：

```
ssh -i <path_to_the_ssh_key> <machine_user>@<datasense_ip>
```

- `<path_to_the_ssh_key>`= ssh 身份验证密钥的位置
- `<machine_user>`:
  - 对于 AWS：使用 <ec2-user>
  - 对于 Azure：使用为控制台实例创建的用户
  - 对于 GCP：使用为控制台实例创建的用户
- `<datasense_ip>`= 虚拟机实例的 IP 地址

需要修改安全组入站规则才能访问云端的系统。有关详细信息，请参阅：

- "AWS 中的安全组规则"
- "Azure 中的安全组规则"
- "Google Cloud 中的防火墙规则"

NetApp Data ClassificationAPI

通过 Web UI 提供的NetApp Data Classification功能也可通过 REST API 提供。

数据分类中定义了四个类别，与 UI 中的选项卡相对应：

- 调查
- Compliance
- 治理
- 配置

Swagger 文档中的 API 允许您搜索、聚合数据、跟踪扫描以及执行复制、移动和删除等操作。

概述

该 API 使您能够执行以下功能：

- 导出信息
  - UI 中可用的所有内容都可以通过 API 导出（报告除外）
  - 数据以 JSON 格式导出（易于解析并推送到第三方应用程序，如 Splunk）
- 使用“AND”和“OR”语句创建查询，包括和排除信息等等。

  例如，您可以找到没有特定个人身份信息 (PII) 的文件（UI 中不可用的功能）。您还可以排除导出操作的特定字段。
- 执行操作
  - 更新 CIFS 凭证
  - 查看和取消操作
  - 重新扫描目录
  - 导出数据

API 是安全的，它使用与 UI 相同的身份验证方法。您可以在"REST API 文档"。

访问 **Swagger API** 参考

要进入 Swagger，您需要数据分类实例的 IP 地址。对于云部署，您将使用公共 IP 地址。然后你需要进入这个端点：

https://<classification_ip>/documentation

使用 **API** 的示例

以下示例显示了复制文件的 API 调用。

API 请求

您最初需要获取系统的所有相关字段和选项，以查看调查选项卡中的所有过滤器。

```
curl -X GET "http://{classification_ip}/api/{classification_version}
/search/options?data_mode=ALL_EXTRACTABLE" -H "accept: application/json"
-H "Authorization: Bearer eyJhbGciOiJSUzI1NiIsInR…… " -H "x-agent-id:
hOXsZNvnA5LsthwMILtjL9xZFYBQxAwMclients"
```

响应


```
{
  "options": [
    {
      "active_directory_affected": false,
      "data_mode": "ALL_SCANNED",
      "field": "string",
      "is_rulable": true,
      "name": "string",
      "operators": [
        "EQUALS"
      ],
      "optional_values": [
        {}
      ],
      "secondary": {},
      "server_data": false,
      "type": "TEXT"
    }
  ]
}
{
  "options": [
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "POLICIES",
      "name": "Policies",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "EXTRACTION_STATUS_RANGE",
```

```
    "name": "Scan Analysis Status",
    "operators": [
      "IN"
    ],
    "server_data": true,
    "type": "SELECT"
  },
  {
    "active_directory_affected": false,
    "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
    "field": "SCAN_ANALYSIS_ERROR",
    "name": "Scan Analysis Event",
    "operators": [
      "IN"
    ],
    "server_data": true,
    "type": "SELECT"
  },
  {
    "active_directory_affected": false,
    "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
    "field": "PUBLIC_ACCESS",
    "name": "Open Permissions",
    "operators": [
      "IN",
      "NOT_IN"
    ],
    "server_data": true,
    "type": "SELECT"
  },
  {
    "active_directory_affected": true,
    "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
    "field": "USERS_PERMISSIONS_COUNT_RANGE",
    "name": "Number of Users with Access",
    "operators": [
      "IN",
      "NOT_IN"
    ],
    "server_data": true,
    "type": "SELECT"
  },
  {
    "active_directory_affected": true,
    "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
    "field": "USER_GROUP_PERMISSIONS",
```

```
      "name": "User / Group Permissions",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "FILE_OWNER",
      "name": "File Owner",
      "operators": [
        "EQUALS",
        "CONTAINS"
      ],
      "server_data": true,
      "type": "TEXT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "ENVIRONMENT_TYPE",
      "name": "system-type",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "ENVIRONMENT",
      "name": "system",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_SCANNED",
```

```
    "field": "SCAN_TASK",
    "name": "Storage Repository",
    "operators": [
      "IN",
      "NOT_IN"
    ],
    "server_data": true,
    "type": "SELECT"
  },
  {
    "active_directory_affected": false,
    "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
    "field": "FILE_PATH",
    "name": "File / Directory Path",
    "operators": [
      "MULTI_CONTAINS",
      "MULTI_EXCLUDE"
    ],
    "server_data": true,
    "type": "MULTI_TEXT"
  },
  {
    "active_directory_affected": false,
    "data_mode": "ALL_DASHBOARD_EXTRACTABLE",
    "field": "CATEGORY",
    "name": "Category",
    "operators": [
      "IN",
      "NOT_IN"
    ],
    "server_data": true,
    "type": "SELECT"
  },
  {
    "active_directory_affected": false,
    "data_mode": "ALL_EXTRACTABLE",
    "field": "PATTERN_SENSITIVITY_LEVEL",
    "name": "Sensitivity Level",
    "operators": [
      "IN"
    ],
    "server_data": true,
    "type": "SELECT"
  },
  {
    "active_directory_affected": false,
```

```
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "NUMBER_OF_IDENTIFIERS",
      "name": "Number of identifiers",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "PATTERN_PERSONAL",
      "name": "Personal Data",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "PATTERN_SENSITIVE",
      "name": "Sensitive Personal Data",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "DATA_SUBJECT",
      "name": "Data Subject",
      "operators": [
        "EQUALS",
        "CONTAINS"
      ],
      "server_data": true,
      "type": "TEXT"
    },
```

```
    {
      "active_directory_affected": false,
      "data_mode": "DIRECTORIES",
      "field": "DIRECTORY_TYPE",
      "name": "Directory Type",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "FILE_TYPE",
      "name": "File Type",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "FILE_SIZE_RANGE",
      "name": "File Size",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "FILE_CREATION_RANGE_RETENTION",
      "name": "Created Time",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
```

```
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "DISCOVERED_TIME_RANGE",
      "name": "Discovered Time",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "FILE_LAST_MODIFICATION_RETENTION",
      "name": "Last Modified",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "FILE_LAST_ACCESS_RANGE_RETENTION",
      "name": "Last Accessed",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "FILES",
      "field": "IS_DUPLICATE",
      "name": "Duplicates",
      "operators": [
        "EQUALS",
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
```

```
    {
      "active_directory_affected": false,
      "data_mode": "FILES",
      "field": "FILE_HASH",
      "name": "File Hash",
      "operators": [
        "EQUALS",
        "IN"
      ],
      "server_data": true,
      "type": "TEXT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "USER_DEFINED_STATUS",
      "name": "Tags",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "ASSIGNED_TO",
      "name": "Assigned to",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    }
  ]
}
```


我们将在请求参数中使用该响应来过滤我们想要复制的文件。

您可以对多个项目应用一个操作。支持的操作类型包括：移动、删除和复制。

我们将创建复制动作：

API 请求

下一个 API 是操作 API，它允许您创建多个操作。

```
curl -X POST "http://
{classification_ip}/api//{classification_version}/actions" -H "accept:
application/json" -H "Authorization: Bearer eyJhbGciOiJSUzI1NiIsInR…… "
-H "x-agent-id: hOXsZNvnA5LsthwMILtjL9xZFYBQxAwMclients " -H "Content-
Type: application/json" -d "{ \"action_type\": \"COPY\", \"data_mode\":
\"FILES\", \"policy_id\": 0, \"request_params\": { destination_nfs_path:
"{ontap_ip}:/{share_name} " },
\"requested_query\":{"condition":"AND","rules":[{"field":"ENVIRONMENT_TYPE
","operator":"IN","value":["ONPREM"]},{"field":"CATEGORY","operator":"IN",
"value":["21"]}]}}"
```

响应

响应将返回操作对象，因此您可以使用获取和删除 API 来获取有关操作的状态，或取消它。

```
{
  "action_type": "COPY",
  "creation_time": "2023-08-08T12:37:21.705Z",
  "data_mode": "FILES",
  "end_time": "2023-08-08T12:37:21.705Z",
  "estimated_time_to_complete": 0,
  "id": 0,
  "policy_id": 0,
  "policy_name": "string",
  "priority": 0,
  "request_params": {},
  "requested_query": {},
  "result": {
    "error_message": "string",
    "failed": 0,
    "in_progress": 0,
    "succeeded": 0,
    "total": 0
  },
  "start_time": "2023-08-08T12:37:21.705Z",
  "status": "QUEUED",
  "title": "string",
  "user_id": "string"
}
```

知识和支持

注册NetApp Console支持

需要进行支持注册才能获得针对NetApp Console及其存储解决方案和数据服务的技术支持。还需要支持注册才能启用Cloud Volumes ONTAP系统的关键工作流程。

注册支持并不能使NetApp获得云提供商文件服务的支持。有关云提供商文件服务、其基础设施或使用该服务的任何解决方案的技术支持，请参阅该产品文档中的“获取帮助”。

- "适用于ONTAP 的Amazon FSx"
- "Azure NetApp Files"
- "Google Cloud NetApp Volumes"

支持注册概述

激活支持权利的注册方式有两种：

- 注册您的NetApp Console帐户序列号（您的 20 位 960xxxxxxxx 序列号位于控制台中的“支持资源”页面上）。

  这可作为控制台内任何服务的单一支持订阅 ID。每个控制台帐户都必须注册。

- 在您的云提供商市场中注册与订阅相关的Cloud Volumes ONTAP序列号（这些是 20 位 909201xxxxxxxx 序列号）。

  这些序列号通常称为_PAYGO 序列号_，由NetApp Console在Cloud Volumes ONTAP部署时生成。

注册两种类型的序列号可以实现开立支持票和自动生成案例等功能。通过将NetApp支持站点 (NSS) 帐户添加到控制台即可完成注册，如下所述。

注册NetApp Console以获取NetApp支持

要注册支持并激活支持权利，您的NetApp Console帐户中的一名用户必须将NetApp支持站点帐户与其控制台登录名关联。如何注册NetApp支持取决于您是否已经拥有NetApp支持站点 (NSS) 帐户。

拥有 **NSS** 帐户的现有客户

如果您是拥有 NSS 帐户的NetApp客户，则只需通过控制台注册即可获得支持。

步骤

1. 选择“管理”>“凭证”。
2. 选择*用户凭证*。
3. 选择*添加 NSS 凭据*并按照NetApp支持站点 (NSS) 身份验证提示进行操作。
4. 要确认注册过程是否成功，请选择“帮助”图标，然后选择“支持”。

   *资源*页面应显示您的控制台帐户已注册以获得支持。

请注意，如果其他控制台用户尚未将NetApp支持站点帐户与其登录名关联，他们将看不到相同的支持注册状态。但是，这并不意味着您的帐户没有注册支持。只要组织中的一名用户遵循了这些步骤，您的帐户就已注册。

现有客户但没有 **NSS** 账户

如果您是现有的NetApp客户，拥有现有许可证和序列号但没有 NSS 帐户，则需要创建一个 NSS 帐户并将其与您的控制台登录关联。

步骤

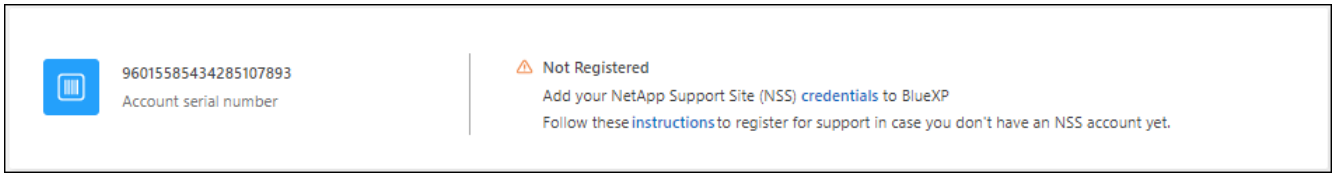
1. 通过完成以下操作创建NetApp支持站点帐户 "NetApp支持站点用户注册表"
   a. 请务必选择适当的用户级别，通常为* NetApp客户/最终用户*。
   b. 请务必复制上面用于序列号字段的控制台帐户序列号（960xxxx）。这将加快账户处理速度。
2. 完成以下步骤，将您的新 NSS 帐户与您的控制台登录名关联拥有 NSS 帐户的现有客户。

NetApp全新产品

如果您是NetApp新用户并且没有 NSS 帐户，请按照以下步骤操作。

步骤

1. 在控制台的右上角，选择“帮助”图标，然后选择“支持”。
2. 从支持注册页面找到您的帐户 ID 序列号。



3. 导航至 "NetApp 的支持注册网站"并选择*我不是注册的NetApp客户*。
4. 填写必填字段（带有红色星号的字段）。
5. 在*产品线*字段中，选择*云管理器*，然后选择适用的计费提供商。
6. 从上面的步骤 2 复制您的帐户序列号，完成安全检查，然后确认您已阅读 NetApp 的全球数据隐私政策。

   一封电子邮件会立即发送到提供的邮箱以完成此安全交易。如果几分钟内没有收到验证电子邮件，请务必检查您的垃圾邮件文件夹。

7. 从电子邮件中确认操作。

   确认向NetApp提交您的请求并建议您创建NetApp支持站点帐户。

8. 通过完成以下操作创建NetApp支持站点帐户 "NetApp支持站点用户注册表"
   a. 请务必选择适当的用户级别，通常为* NetApp客户/最终用户*。
   b. 请务必复制上面用于序列号字段的帐户序列号（960xxxx）。这将加快处理速度。

完成后

NetApp应该在此过程中与您联系。这是针对新用户的一次性入职培训。

拥有NetApp支持站点帐户后，请按照以下步骤将该帐户与您的控制台登录关联拥有 NSS 帐户的现有客户。

关联 NSS 凭据以获得Cloud Volumes ONTAP支持

需要将NetApp支持站点凭据与您的控制台帐户关联，才能为Cloud Volumes ONTAP启用以下关键工作流程：

- 注册即用即付Cloud Volumes ONTAP系统以获得支持

  需要提供您的 NSS 帐户才能激活对您的系统的支持并获得对NetApp技术支持资源的访问权限。

- 自带许可证 (BYOL) 时部署Cloud Volumes ONTAP

  需要提供您的 NSS 帐户，以便控制台可以上传您的许可证密钥并启用您购买的期限的订阅。这包括期限续订的自动更新。

- 将Cloud Volumes ONTAP软件升级到最新版本

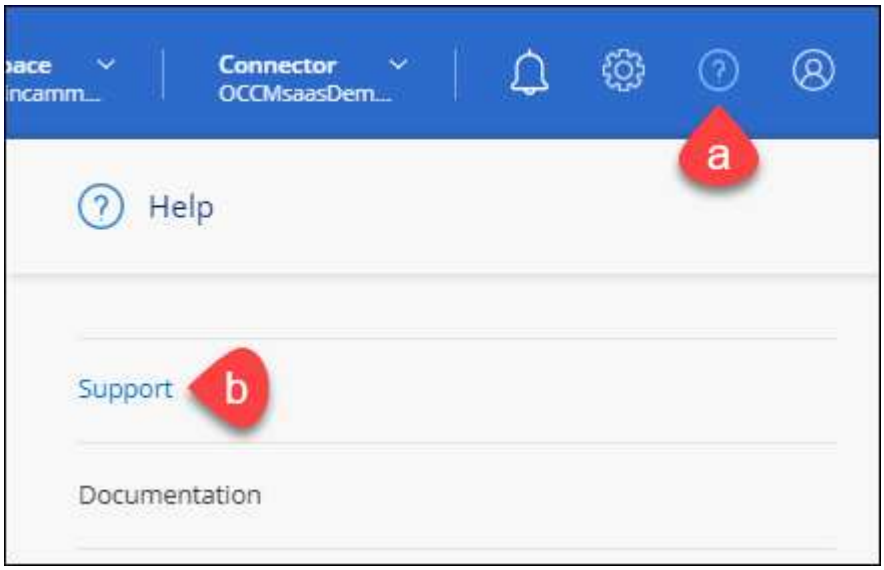
将 NSS 凭据与您的NetApp Console帐户关联与将 NSS 帐户与控制台用户登录关联不同。

这些 NSS 凭证与您的特定控制台帐户 ID 相关联。属于控制台组织的用户可以从*支持 > NSS 管理*访问这些凭据。

- 如果您有客户级帐户，则可以添加一个或多个 NSS 帐户。
- 如果您有合作伙伴或经销商帐户，则可以添加一个或多个 NSS 帐户，但不能与客户级帐户一起添加。

步骤

1. 在控制台的右上角，选择“帮助”图标，然后选择“支持”。



2. 选择*NSS 管理 > 添加 NSS 帐户*。
3. 当出现提示时，选择“继续”以重定向到 Microsoft 登录页面。

   NetApp使用 Microsoft Entra ID 作为特定于支持和许可的身份验证服务的身份提供者。

4. 在登录页面，提供您的NetApp支持站点注册的电子邮件地址和密码以执行身份验证过程。

   这些操作使控制台能够使用您的 NSS 帐户进行许可证下载、软件升级验证和未来支持注册等操作。

   请注意以下事项：

   - NSS 帐户必须是客户级帐户（不是访客或临时帐户）。您可以拥有多个客户级 NSS 帐户。
   - 如果该帐户是合作伙伴级别帐户，则只能有一个 NSS 帐户。如果您尝试添加客户级 NSS 帐户并且合作伙伴级帐户已存在，您将收到以下错误消息：

     “此帐户不允许使用 NSS 客户类型，因为已经存在不同类型的 NSS 用户。”

   如果您已有客户级 NSS 帐户并尝试添加合作伙伴级帐户，情况也是如此。

   - 成功登录后， NetApp将存储 NSS 用户名。

     这是系统生成的映射到您的电子邮件的 ID。在*NSS 管理*页面上，您可以显示来自 ••• 菜单。

   - 如果您需要刷新登录凭证令牌，还有一个*更新凭证*选项 ••• 菜单。

     使用此选项会提示您再次登录。请注意，这些帐户的令牌将在 90 天后过期。我们将发布通知来提醒您此事。

获取NetApp Data Classification帮助

NetApp以多种方式为NetApp Console及其云服务提供支持。全天候提供广泛的免费自助支持选项，例如知识库 (KB) 文章和社区论坛。您的支持注册包含通过网络工单获取的远程技术支持。

获取云提供商文件服务的支持

有关云提供商文件服务、其基础设施或使用该服务的任何解决方案的技术支持，请参阅该产品的文档。

- "适用于ONTAP 的Amazon FSx"
- "Azure NetApp Files"
- "Google Cloud NetApp Volumes"

要获得特定于NetApp及其存储解决方案和数据服务的技术支持，请使用下面描述的支持选项。

使用自助选项

这些选项每周 7 天、每天 24 小时免费提供：

- 文档

  您当前正在查看的NetApp Console文档。

- "知识库"

搜索NetApp知识库以查找有助于解决问题的文章。

- "社区"

  加入NetApp Console社区，关注正在进行的讨论或创建新的讨论。

向NetApp支持创建案例

除了上述自助支持选项之外，您还可以在激活支持后与NetApp支持专家合作解决任何问题。

开始之前

- 要使用“创建案例”功能，您必须首先将您的NetApp支持站点凭据与您的控制台登录关联。 "了解如何管理与控制台登录相关的凭据" 。
- 如果您要为具有序列号的ONTAP系统打开案例，那么您的 NSS 帐户必须与该系统的序列号相关联。

步骤

1. 在NetApp Console中，选择“帮助”>“支持”。
2. 在“资源”页面上，选择“技术支持”下的可用选项之一：

   a. 如果您想通过电话与某人交谈，请选择“致电我们”。您将被引导至 netapp.com 上的一个页面，其中列出了您可以拨打的电话号码。

   b. 选择“创建案例”向NetApp支持专家开具一张票：

      - 服务：选择与问题相关的服务。例如，* NetApp Console* 特定于控制台内的工作流或功能的技术支持问题。
      - 系统：如果适用于存储，请选择* Cloud Volumes ONTAP* 或 **On-Prem**，然后选择相关的工作环境。

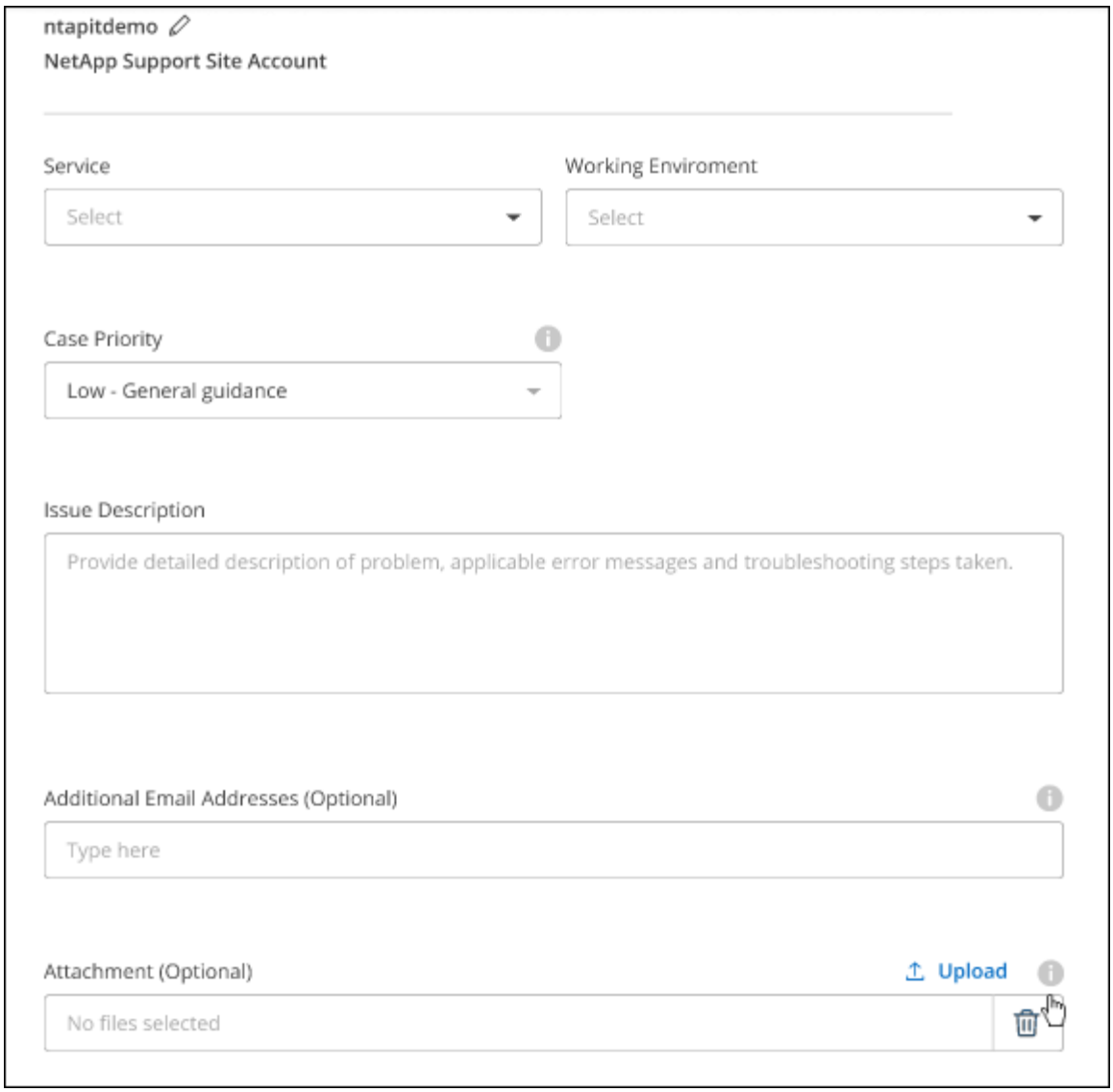
        系统列表位于控制台组织范围内，并且您在顶部横幅中选择了控制台代理。

      - 案例优先级：选择案例的优先级，可以是低、中、高或严重。

        要了解有关这些优先事项的更多详细信息，请将鼠标悬停在字段名称旁边的信息图标上。

      - 问题描述：提供问题的详细描述，包括任何适用的错误消息或您执行的故障排除步骤。
      - 其他电子邮件地址：如果您想让其他人知道此问题，请输入其他电子邮件地址。
      - 附件（可选）：一次最多上传五个附件。

        每个附件文件大小限制为 25 MB。支持以下文件扩展名：txt、log、pdf、jpg/jpeg、rtf、doc/docx、xls/xlsx 和 csv。



完成后

将会出现一个弹出窗口，其中显示您的支持案例编号。NetApp支持专家将审查您的案例并尽快回复您。

要查看支持案例的历史记录，您可以选择*设置>时间线*并查找名为“创建支持案例”的操作。最右边的按钮可让您展开操作以查看详细信息。

尝试创建案例时，您可能会遇到以下错误消息：

“您无权针对所选服务创建案例”

此错误可能意味着 NSS 帐户及其关联的记录公司与NetApp Console帐户序列号的记录公司不同（即。960xxxx）或工作环境序列号。您可以使用以下选项之一寻求帮助：

- 提交非技术案例 https://mysupport.netapp.com/site/help

管理您的支持案例

您可以直接从控制台查看和管理活动和已解决的支持案例。您可以管理与您的 NSS 帐户和公司相关的案例。

请注意以下事项：

- 页面顶部的案例管理仪表板提供两种视图：
  - 左侧视图显示了您提供的用户 NSS 帐户在过去 3 个月内打开的案件总数。
  - 右侧的视图根据您的用户 NSS 帐户显示了过去 3 个月内贵公司级别开设的案件总数。

  表中的结果反映了与您选择的视图相关的案例。
- 您可以添加或删除感兴趣的列，并且可以过滤优先级和状态等列的内容。其他列仅提供排序功能。

  请查看以下步骤以了解更多详细信息。
- 在每个案件级别，我们提供更新案件记录或关闭尚未关闭或待关闭状态的案件的功能。

步骤

1. 在NetApp Console中，选择“帮助”>“支持”。
2. 选择*案例管理*，如果出现提示，请将您的 NSS 帐户添加到控制台。

   案例管理*页面显示与您的控制台用户帐户关联的 **NSS** 帐户相关的未结案例。这与出现在 ***NSS** 管理 页面顶部的 NSS 帐户相同。
3. （可选）修改表中显示的信息：
   - 在“组织的案例”下，选择“查看”以查看与您的公司相关的所有案例。
   - 通过选择精确的日期范围或选择不同的时间范围来修改日期范围。
   - 过滤列的内容。
   - 通过选择然后选择您想要显示的列。
4. 通过选择管理现有案例并选择其中一个可用选项：
   - 查看案例：查看有关特定案例的完整详细信息。
   - 更新案例说明：提供有关您的问题的更多详细信息，或选择*上传文件*以附加最多五个文件。

     每个附件文件大小限制为 25 MB。支持以下文件扩展名：txt、log、pdf、jpg/jpeg、rtf、doc/docx、xls/xlsx 和 csv。
   - 结案：提供有关结案原因的详细信息，然后选择*结案*。

关于NetApp Data Classification的常见问题解答

如果您只是想快速找到问题的答案，此常见问题解答可以为您提供帮助。

NetApp Data Classification

以下问题提供了对数据分类的一般了解。

数据分类如何工作?

数据分类在NetApp Console系统和存储系统旁边部署了另一层 AI。然后，它会扫描卷、存储桶、数据库和其他存储帐户上的数据，并对找到的数据洞察进行索引。数据分类利用人工智能和自然语言处理，而不是通常围绕正则表达式和模式匹配构建的替代解决方案。

数据分类使用人工智能来提供数据的上下文理解，以便进行准确的检测和分类。它由人工智能驱动，因为它是针对现代数据类型和规模而设计的。它还了解数据背景，以便提供强大、准确的发现和分类。

"详细了解数据分类的工作原理" 。

数据分类是否有 **REST API**，它是否可以与第三方工具一起使用?

是的，数据分类有一个 REST API，用于支持控制台核心平台一部分的数据分类版本中的功能。看"API 文档" 。

数据分类是否可以通过云市场获得?

数据分类是NetApp Console核心功能的一部分，因此您不需要使用此服务的市场。

数据分类扫描和分析

以下问题与数据分类扫描性能和分析有关。

数据分类多久扫描一次我的数据?

虽然初始数据扫描可能需要一点时间，但后续扫描仅检查增量变化，从而减少系统扫描时间。数据分类以循环方式连续扫描您的数据，每次扫描六个存储库，以便所有更改的数据都能快速分类。

"了解扫描的工作原理" 。

数据分类每天仅扫描数据库一次；数据库不像其他数据源那样被连续扫描。

数据扫描对您的存储系统和数据的影响可以忽略不计。

扫描性能是否有所不同?

扫描性能可能因网络带宽和环境中的平均文件大小而异。它还取决于主机系统（在云端或本地）的大小特征。请参阅"数据分类实例"和"部署数据分类"了解更多信息。

在最初添加新的数据源时，您还可以选择仅执行“映射”（Mapping only）扫描，而不是完整的“分类”（Map &

Classify）扫描。由于它不需要访问文件来查看其中的数据，因此可以非常快速地在数据源上完成映射。"查看映射和分类扫描之间的区别" 。

我可以使用数据分类搜索我的数据吗?

数据分类提供了广泛的搜索功能，可以轻松地在所有连接的源中搜索特定文件或数据。数据分类使用户能够进行比元数据所反映的更深入的搜索。它是一种与语言无关的服务，还可以读取文件并分析多种敏感数据类型，例如名称和 ID。例如，用户可以在结构化和非结构化数据存储中进行搜索，以查找可能从数据库泄露到用户文件的数据，从而违反公司政策。可以保存搜索结果以供日后使用，并且可以创建策略以设定的频率搜索并对结果采取行动。

一旦找到感兴趣的文件，就可以列出其特征，包括标签、系统帐户、存储桶、文件路径、类别（来自分类）、文件大小、上次修改、权限状态、重复、敏感度级别、个人数据、文件内的敏感数据类型、所有者、文件类型、文件大小、创建时间、文件哈希、数据是否分配给寻求其关注的人等等。可以使用过滤器来筛选出不相关的特征。

如果存在正确的权限，数据分类还具有基于角色的访问控制（RBAC），允许移动或删除文件。如果没有正确的权限，则可以将任务分配给组织中具有正确权限的人员。

数据分类管理和隐私

以下问题提供了有关如何管理数据分类和隐私设置的信息。

如何启用或禁用数据分类?

首先，您需要在控制台或本地系统中部署数据分类实例。实例运行后，您可以从“配置”选项卡或通过选择特定系统在现有系统、数据库和其他数据源上启用该服务。"了解如何开始" 。

> 在数据源上激活数据分类将立即导致初始扫描。扫描结果很快就会显示。

您可以从数据分类配置页面禁用数据分类扫描单个系统、数据库或文件共享组。看"从数据分类中删除数据源" 。

要完全删除数据分类实例，请从云提供商的门户或本地位置手动删除数据分类实例。

该服务可以排除某些目录中的扫描数据吗?

是如果您希望数据分类排除驻留在特定数据源目录中的扫描数据，则可以将该列表提供给分类引擎。应用该更改后，数据分类将排除指定目录中的扫描数据。"了解更多" 。

是否扫描了位于ONTAP卷上的快照?

否。数据分类不会扫描快照，因为其内容与卷中的内容相同。

如果在ONTAP卷上启用了数据分层，会发生什么情况?

当数据分类使用仅映射扫描扫描具有分层到对象存储的冷数据的卷时，它会扫描所有数据 - 本地磁盘上的数据和分层到对象存储的冷数据。对于实施分层的非NetApp产品来说也是如此。

仅映射扫描不会使冷数据升温——它会保持冷状态并保留在对象存储中。另一方面，如果您执行地图和分类扫描，某些配置可能会使冷数据升温。

源系统和数据类型的类型

以下问题涉及可以扫描的存储类型以及扫描的数据类型。

在政府区域部署时有什么限制吗?

当控制台代理部署在政府区域（AWS GovCloud、Azure Gov 或 Azure DoD）时，支持数据分类 - 也称为“受限模式”。

如果我在没有互联网访问的站点安装数据分类，我可以扫描哪些数据源?

> BlueXP私有模式（传统BlueXP接口）通常用于没有互联网连接的本地环境和安全云区域，其中包括 AWS Secret Cloud、AWS Top Secret Cloud 和 Azure IL6。NetApp继续通过传统的BlueXP界面支持这些环境。有关旧版BlueXP界面中的私有模式文档，请参阅"BlueXP私人模式的 PDF 文档"。

数据分类只能扫描来自本地站点的数据源的数据。目前，数据分类可以以“私人模式”扫描以下本地数据源 - 也称为“暗”站点：

- 本地ONTAP系统
- 数据库模式
- 使用简单存储服务（S3）协议的对象存储

支持哪些文件类型?

数据分类扫描所有文件的类别和元数据洞察，并在仪表板的文件类型部分显示所有文件类型。

当数据分类检测到个人身份信息 (PII) 或执行 DSAR 搜索时，仅支持以下文件格式：

```
.CSV, .DCM, .DOC, .DOCX, .JSON, .PDF, .PPTX, .RTF, .TXT, .XLS, .XLSX, Docs,
Sheets, and Slides
```

数据分类捕获哪些类型的数据和元数据?

数据分类使您能够对数据源运行常规“映射”扫描或完整“分类”扫描。映射仅提供数据的高级概述，而分类提供数据的深层扫描。由于它不需要访问文件来查看其中的数据，因此可以非常快速地在数据源上完成映射。

- 数据映射扫描（仅映射扫描）：数据分类仅扫描元数据。这对于整体数据管理和治理、快速项目范围界定、大型地产和优先级排序很有用。数据映射基于元数据，被认为是一种*快速*扫描。

  快速扫描后，您可以生成数据映射报告。此报告概述了您公司数据源中存储的数据，以帮助您做出有关资源利用率、迁移、备份、安全性和合规性流程的决策。

- 数据分类深度扫描（地图和分类扫描）：数据分类使用标准协议和只读权限在整个环境中扫描数据。打开选定的文件并扫描其中的敏感业务相关数据、私人信息以及与勒索软件相关的问题。

  完整扫描后，您可以将许多附加数据分类功能应用于数据，例如在数据调查页面中查看和优化数据、在文件中搜索名称、复制、移动和删除源文件等。

数据分类捕获元数据，例如：文件名、权限、创建时间、上次访问和上次修改。这包括数据调查详情页面和数据调查报告中显示的所有元数据。

数据分类可以识别多种类型的私人数据，例如个人信息（PII）和敏感个人信息（SPII）。有关私人数据的详细信息，请参阅数据分类扫描的私人数据类别。

我可以将数据分类信息限制给特定用户吗?

是的，数据分类与NetApp Console完全集成。 NetApp Console用户只能查看根据其权限有资格查看的系统的信息。

此外，如果您希望允许某些用户仅查看数据分类扫描结果而无权管理数据分类设置，则可以为这些用户分配*分类查看器*角色（在标准模式下使用NetApp Console时）或*合规性查看器*角色（在受限模式下使用NetApp Console时）。"了解更多" 。

任何人都可以访问我的浏览器和数据分类之间发送的私人数据吗?

不可以。您的浏览器和数据分类实例之间发送的私人数据使用 TLS 1.2 进行端到端加密保护，这意味着NetApp和非NetApp方都无法读取它。除非您请求并批准访问，否则数据分类不会与NetApp共享任何数据或结果。

扫描的数据保留在您的环境中。

敏感数据如何处理?

NetApp无法访问敏感数据，也不会在 UI 中显示它。敏感数据被屏蔽，例如，显示信用卡信息的最后四位数字。

数据存储在哪里?

扫描结果存储在数据分类实例内的 Elasticsearch 中。

如何访问数据?

数据分类通过 API 调用访问存储在 Elasticsearch 中的数据，这些调用需要身份验证并使用 AES-128 加密。直接访问 Elasticsearch 需要 root 访问权限。

许可证和费用

以下问题涉及使用数据分类的许可和成本。

数据分类的费用是多少?

数据分类是NetApp Console的核心功能。沒有充電。

控制台代理部署

以下问题与控制台代理有关。

什么是控制台代理?

控制台代理是在您的云帐户或本地的计算实例上运行的软件，它使NetApp Console能够安全地管理云资源。您必须部署控制台代理才能使用数据分类。

控制台代理需要安装在哪里?

扫描数据时，需要在以下位置安装NetApp Console代理：

- 对于 AWS 中的Cloud Volumes ONTAP或Amazon FSx for ONTAP：控制台代理位于 AWS 中。
- 对于 Azure 或Azure NetApp Files中的Cloud Volumes ONTAP ： 控制台代理位于 Azure 中。
- 对于 GCP 中的Cloud Volumes ONTAP ： 控制台代理位于 GCP 中。
- 对于本地ONTAP系统：控制台代理位于本地。

如果您在这些位置有数据，您可能需要使用 "多个控制台代理"。

数据分类是否需要访问凭证?

数据分类本身不会检索存储凭证。相反，它们存储在控制台代理中。

数据分类使用数据平面凭证（例如 CIFS 凭证）在扫描之前挂载共享。

服务和控制台代理之间的通信是否使用 **HTTP**?

是的，数据分类使用 HTTP 与控制台代理进行通信。

数据分类部署

以下问题与单独的数据分类实例有关。

数据分类支持哪些部署模型?

NetApp Console允许用户在几乎任何地方扫描和报告系统，包括本地、云和混合环境。数据分类通常使用 SaaS 模型部署，其中服务通过控制台界面启用，不需要安装硬件或软件。即使在这种点击即运行的部署模式下，无论数据存储是在本地还是在公共云中，都可以进行数据管理。

数据分类需要什么类型的实例或虚拟机?

什么时候"部署在云端"：

- 在 AWS 中，数据分类在具有 500 GiB GP2 磁盘的 m6i.4xlarge 实例上运行。您可以在部署期间选择较小的实例类型。
- 在 Azure 中，数据分类在具有 500 GiB 磁盘的 Standard_D16s_v3 VM 上运行。
- 在 GCP 中，数据分类在具有 500 GiB 标准持久磁盘的 n2-standard-16 VM 上运行。

"详细了解数据分类的工作原理" 。

我可以在自己的主机上部署数据分类吗?

是您可以在网络或云中具有互联网访问权限的 Linux 主机上安装数据分类软件。一切运作相同，您可以继续通过控制台管理扫描配置和结果。看"在本地部署数据分类"了解系统要求和安装详情。

没有互联网接入的安全站点怎么样?

是的，也支持。你可以"在没有互联网访问权限的本地站点中部署数据分类"以获得完全安全的网站。

法律声明

法律声明提供对版权声明、商标、专利等的访问。

版权

"https://www.netapp.com/company/legal/copyright/"

商标

NETAPP、NETAPP 徽标和NetApp商标页面上列出的标志是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。

"https://www.netapp.com/company/legal/trademarks/"

专利

NetApp拥有的专利的最新列表可以在以下位置找到：

https://www.netapp.com/pdf.html?item=/media/11887-patentspage.pdf

隐私政策

"https://www.netapp.com/company/legal/privacy-policy/"

开源

通知文件提供有关NetApp软件中使用的第三方版权和许可的信息。

- "NetApp Console通知"
- "NetApp Data Classification声明"