开始使用
NetApp Data Classification

NetApp
February 06, 2026

目录

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。

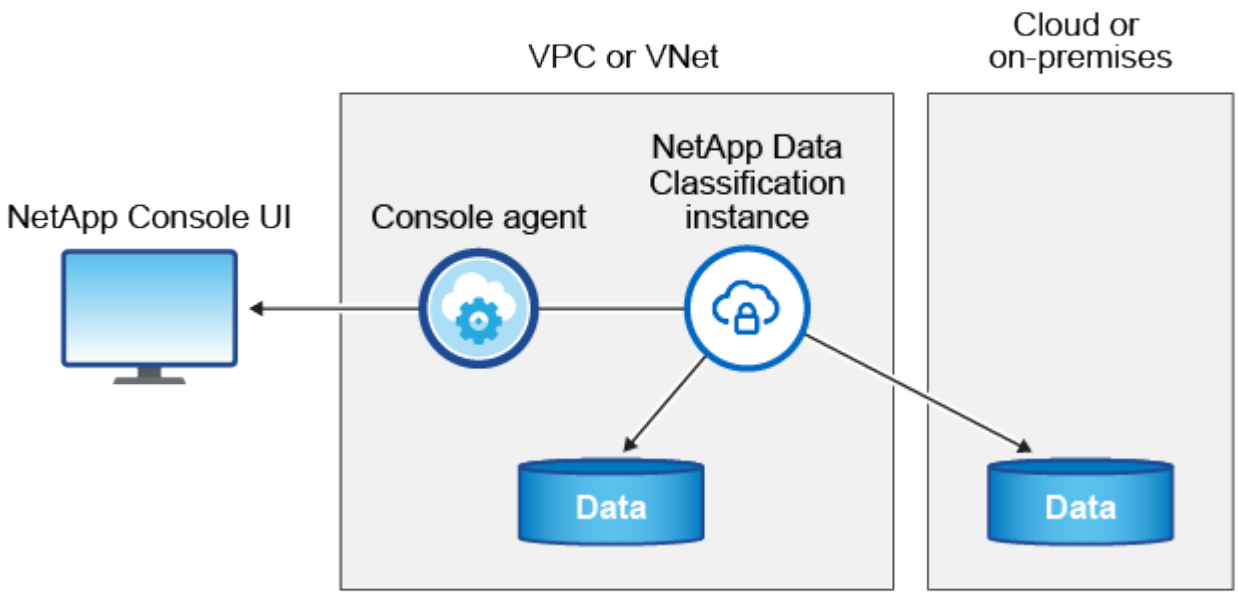
数据传输成本

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。

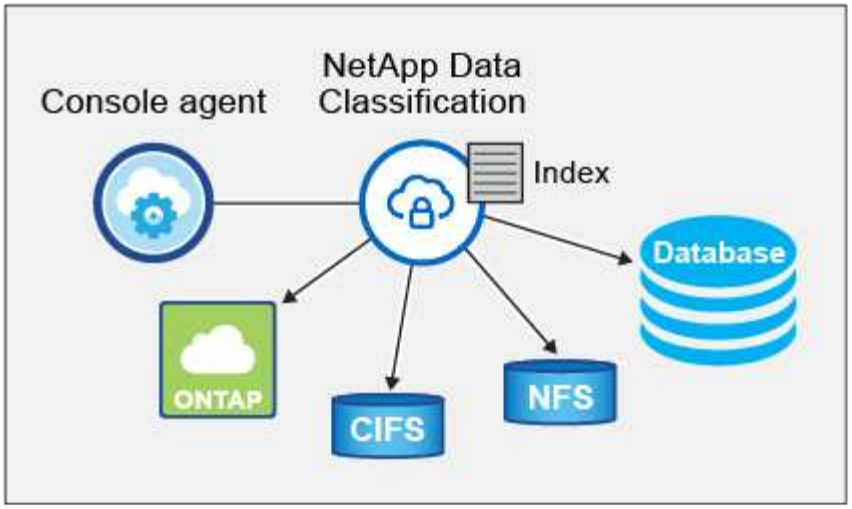
数据分类扫描的工作原理

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" 。

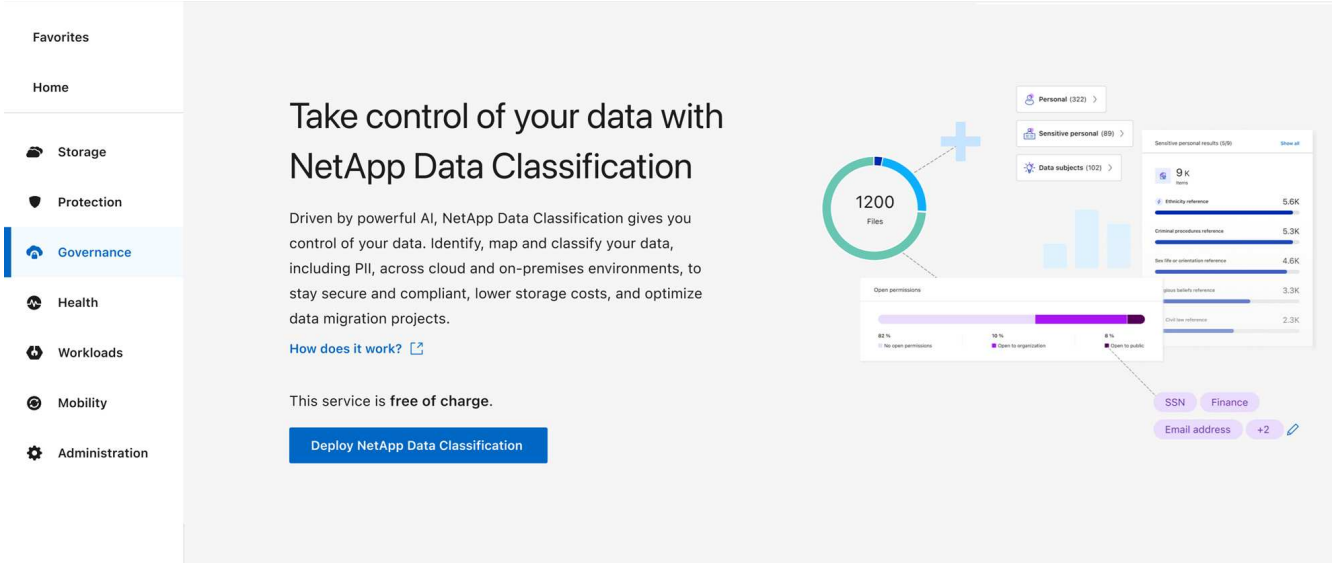
开始之前

- "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

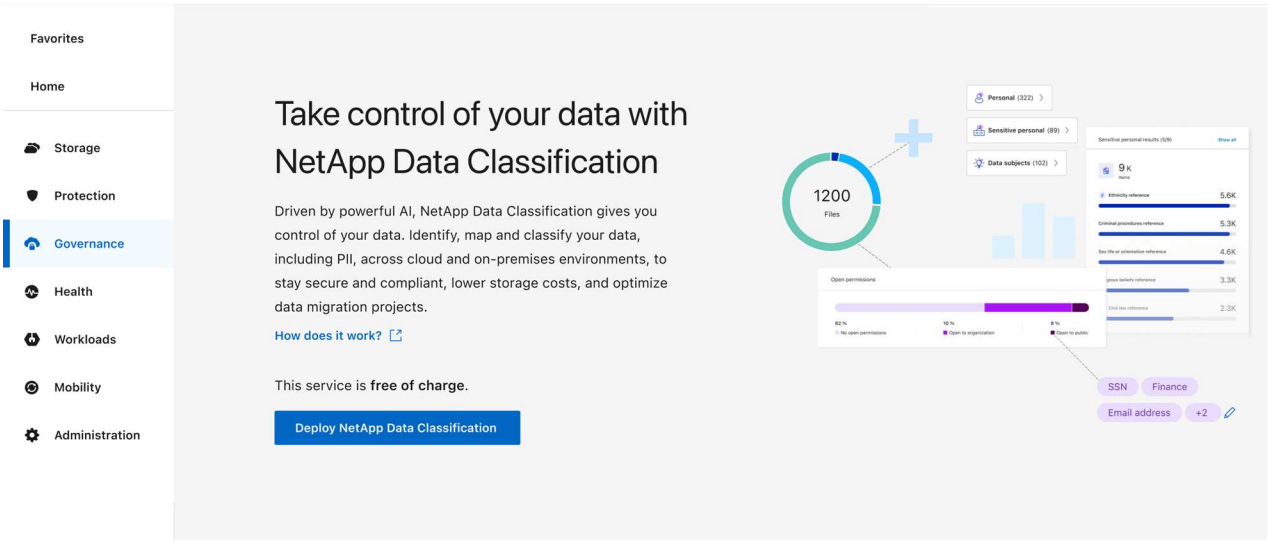
按照以下步骤在云中部署数据分类实例。控制台代理将在云中部署实例，然后在该实例上安装数据分类软件。

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*。

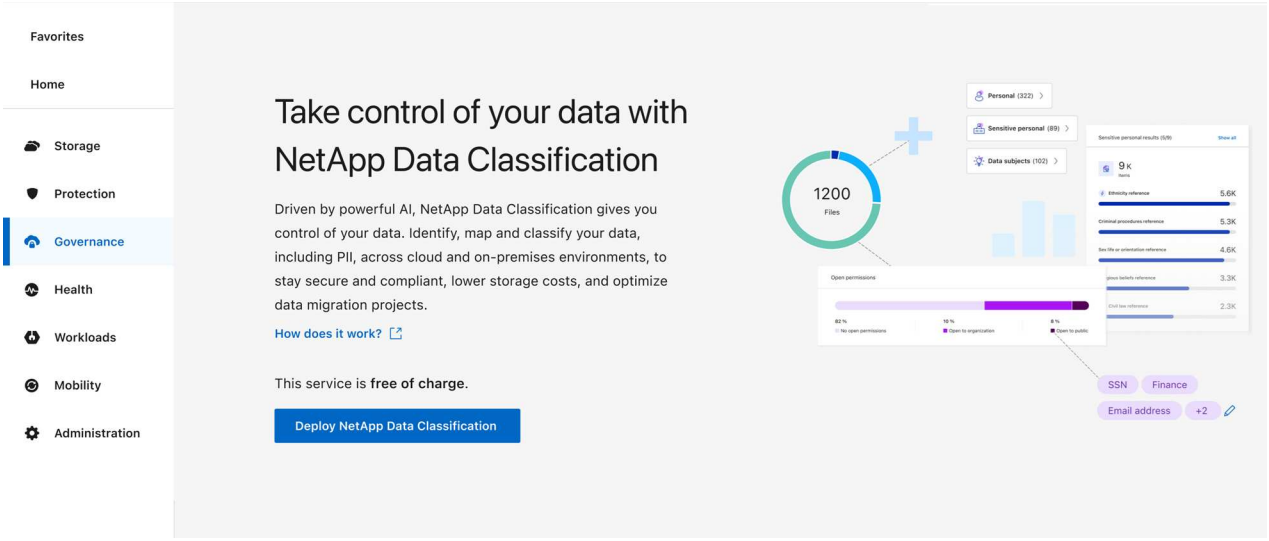


2. 从“安装”页面中，选择“部署”>“部署”以使用“大型”实例大小并启动云部署向导。
3. 向导在执行部署步骤时会显示进度。当需要输入或遇到问题时，系统会提示您。
4. 当实例部署完毕并安装数据分类后，选择“继续配置”进入“配置”页面。

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*。

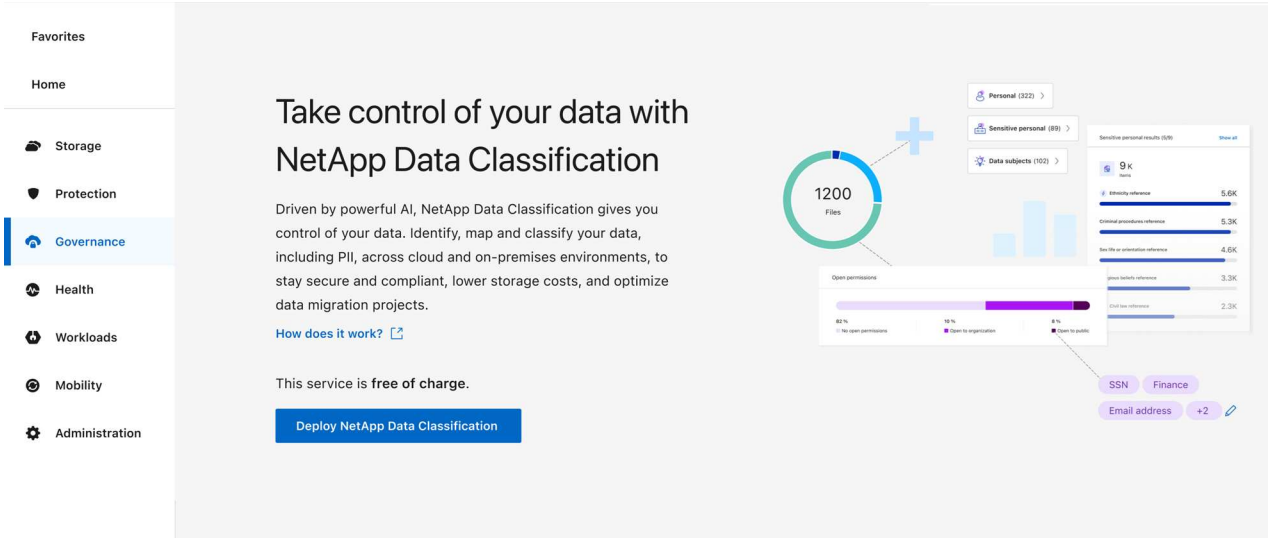


2. 选择*部署*以启动云部署向导。
3. 向导在执行部署步骤时会显示进度。如果遇到任何问题，它将停止并提示输入。
4. 当实例部署完毕并安装数据分类后，选择“继续配置”进入“配置”页面。

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。

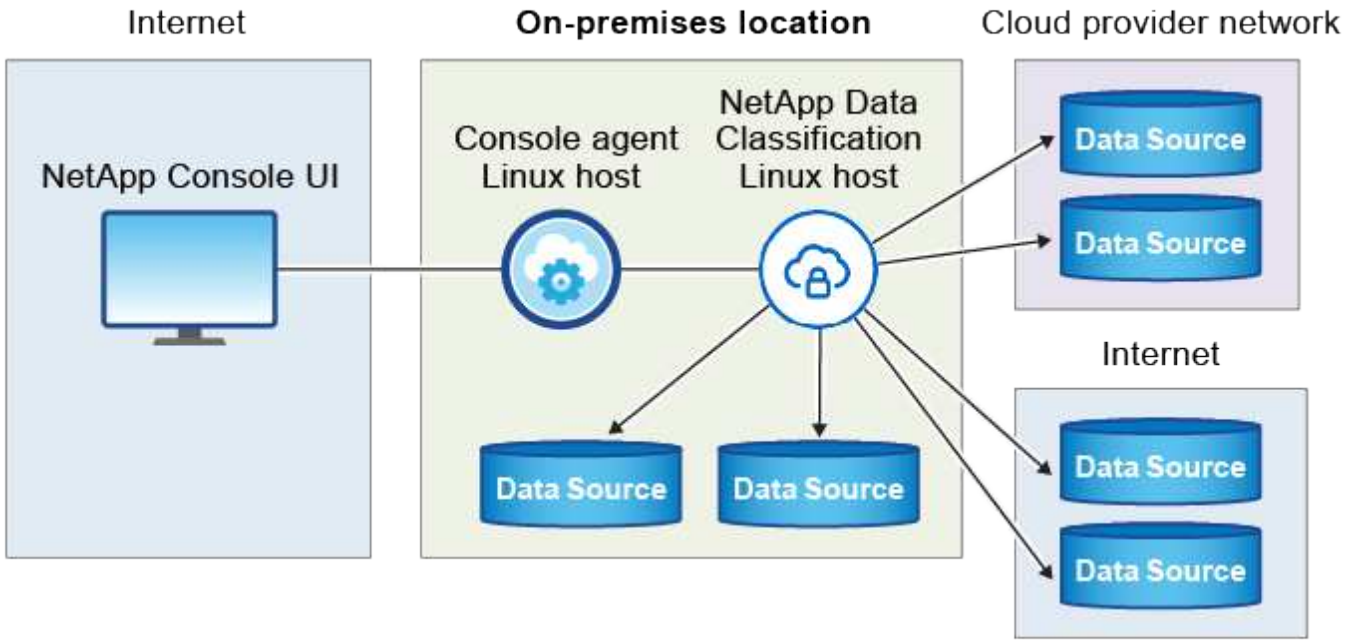
在可以访问互联网的主机上安装**NetApp Data Classification**

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。

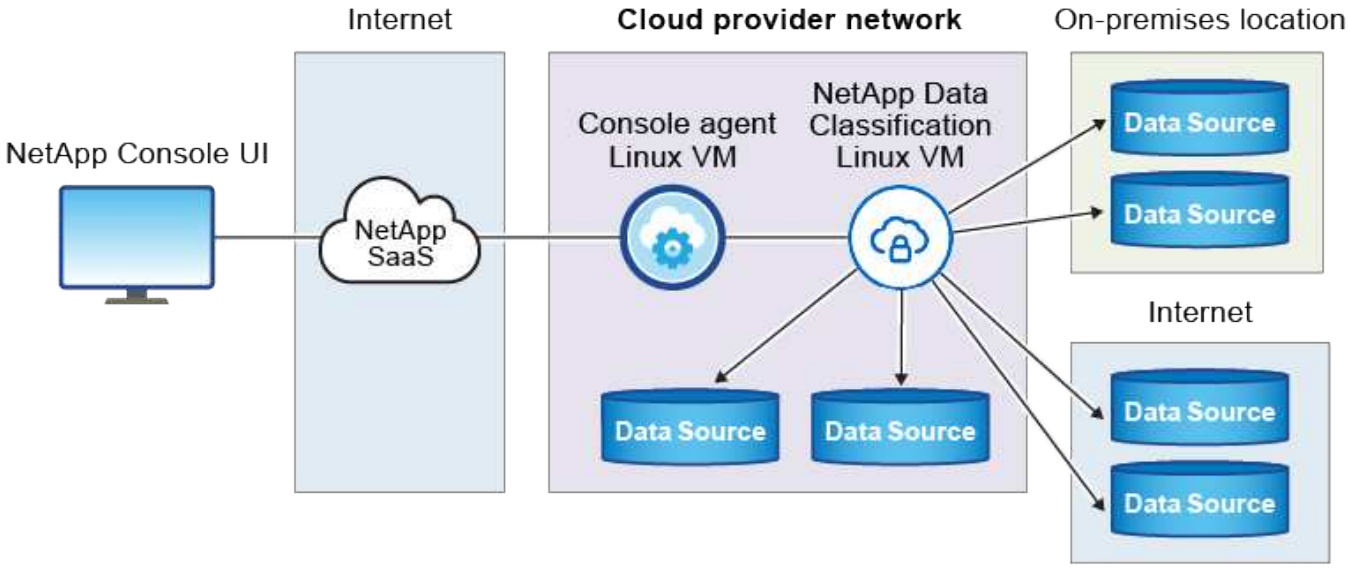
如果您希望使用同样位于本地的数据分类实例来扫描本地ONTAP系统，则本地安装是一个不错的选择。这不是必需的。无论选择哪种安装方法，软件的功能都是相同的。

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

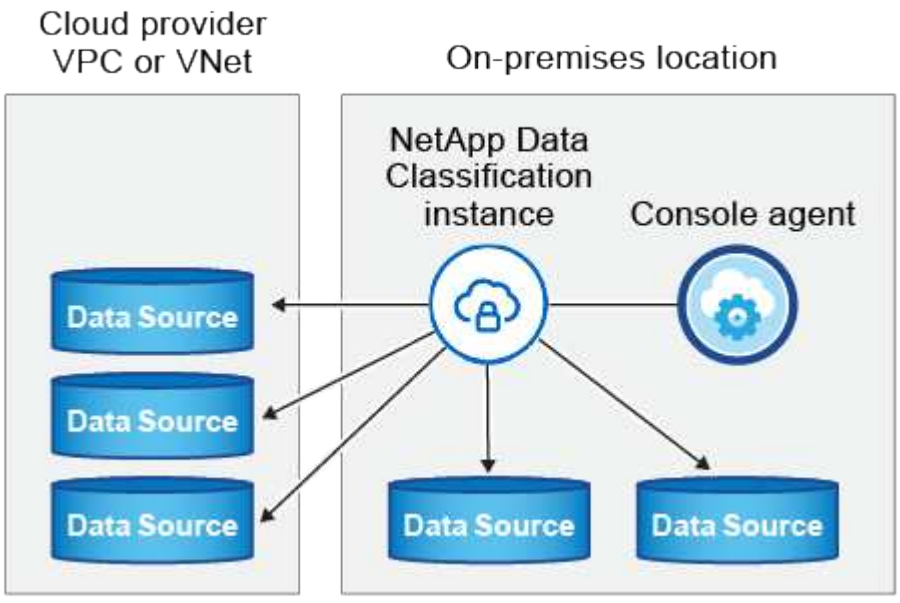
您必须确保所有必需的端口都已打开，以便控制台代理、数据分类、Active Directory 和数据源之间进行通信。

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

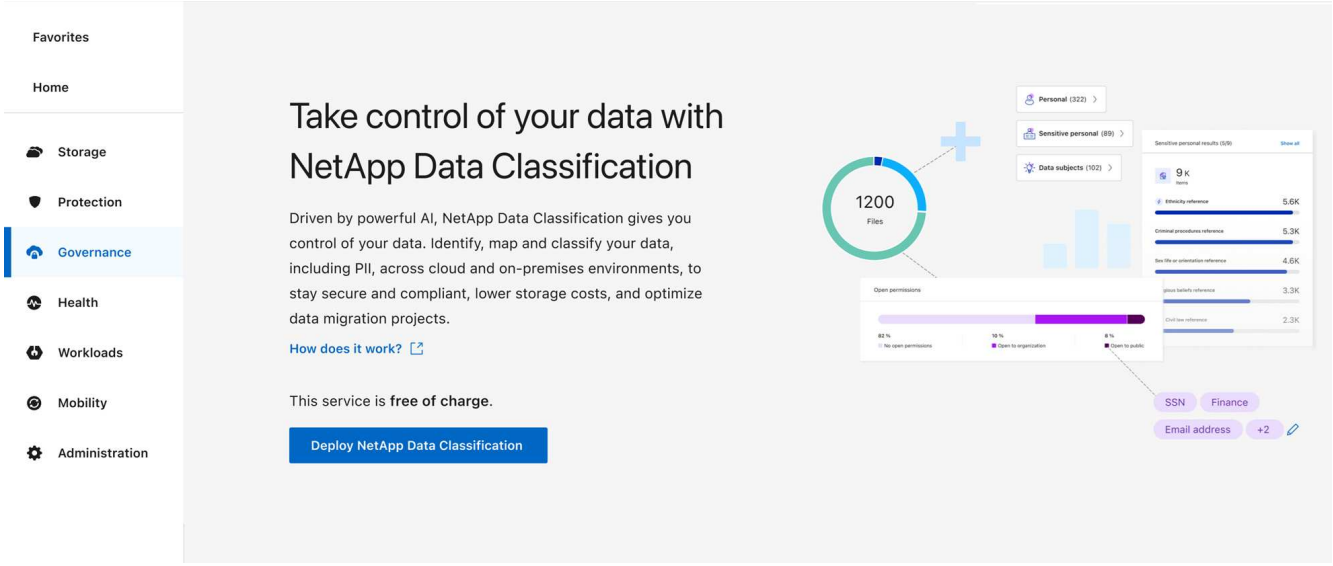
- 验证您的 Linux 系统是否满足主机要求。
- 验证系统是否安装了两个必备软件包（Docker Engine 或 Podman 和 Python 3）。
- 确保您在 Linux 系统上拥有 root 权限。
- 如果您使用代理访问互联网：
  - 您将需要代理服务器信息（IP 地址或主机名、连接端口、连接方案：https 或 http、用户名和密码）。
  - 如果代理正在执行 TLS 拦截，您需要知道数据分类 Linux 系统上存储 TLS CA 证书的路径。
  - 代理必须是非透明的。数据分类目前不支持透明代理。
  - 该用户必须是本地用户。不支持域用户。
- 验证您的离线环境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

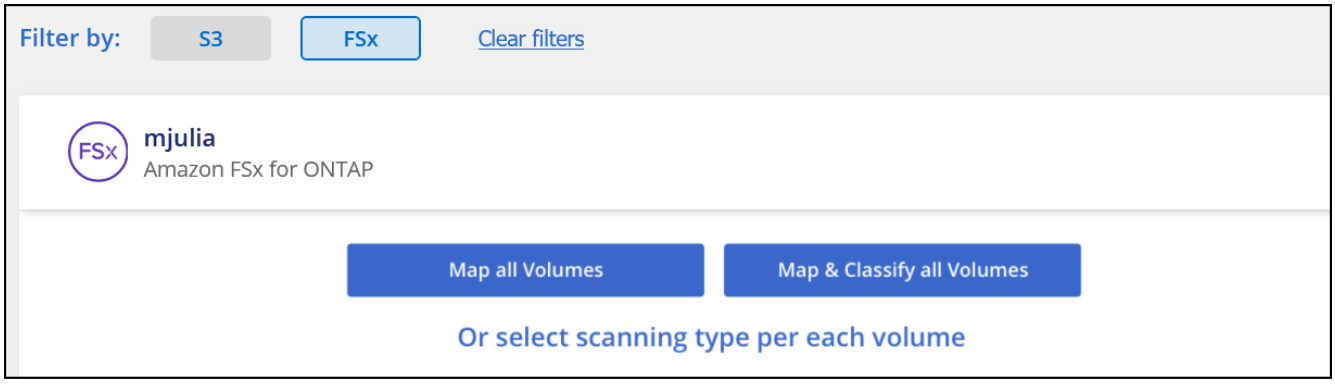
*注意：*扫描 FSx 卷时，当前不支持在本地位置部署数据分类。

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

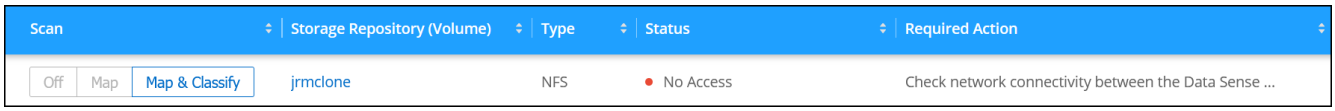
通过检查网络、安全组和导出策略，确保数据分类可以访问卷。

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。

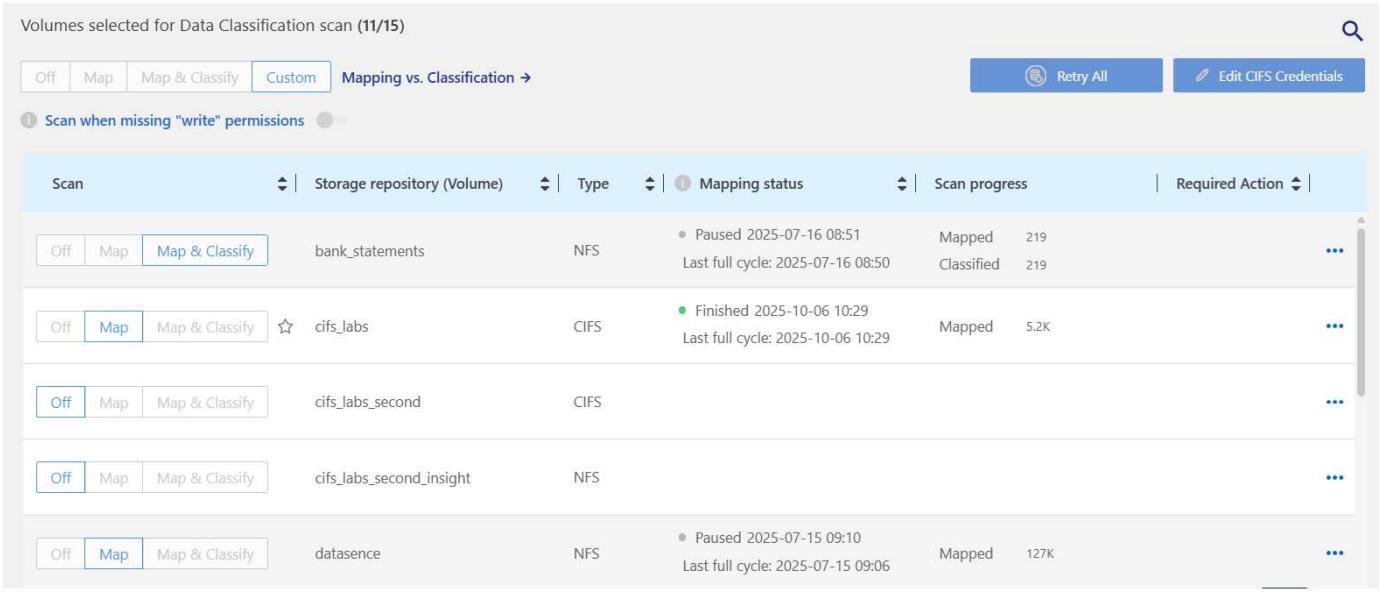
启用和禁用卷上的扫描

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。

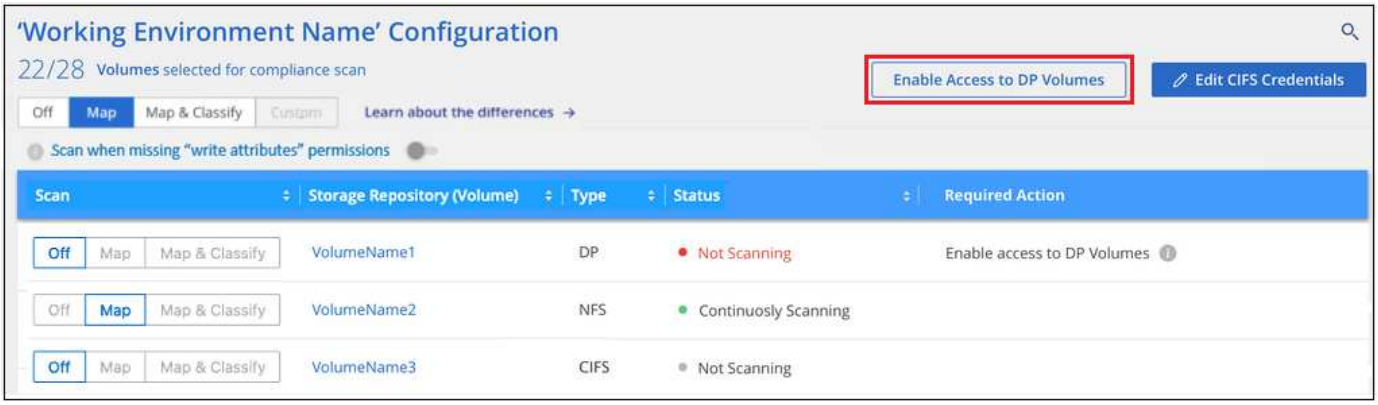
结果

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。

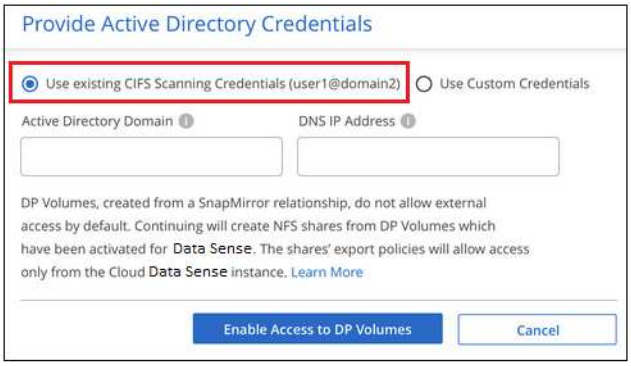
最初，卷列表将这些卷标识为_类型_ **DP**，其_状态_ 未扫描*和_所需操作_ *启用对 **DP** 卷的访问。

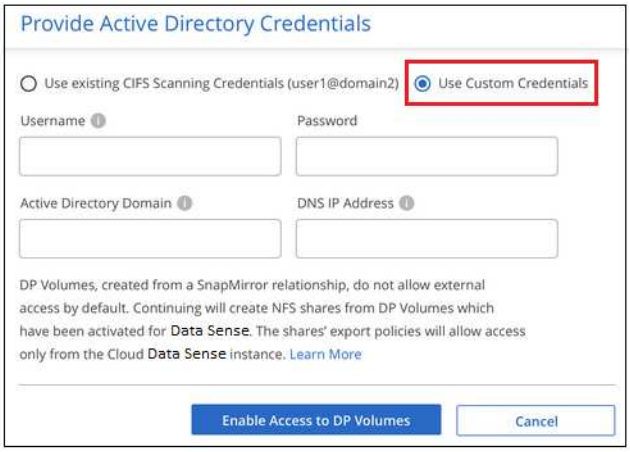


步骤

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。

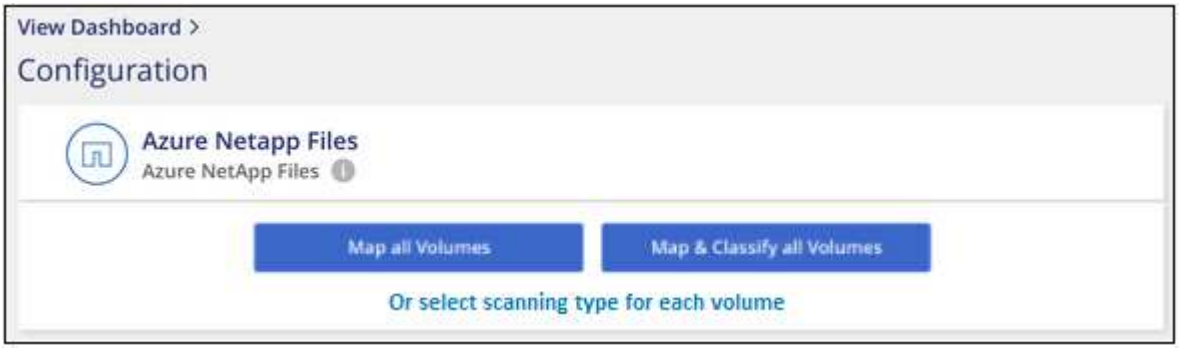
扫描Azure NetApp Files卷时，数据分类必须部署在云中，并且必须部署在与要扫描的卷相同的区域中。

*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

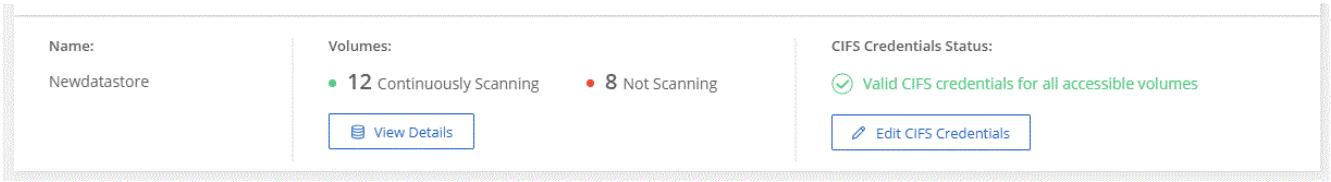
1. 从数据分类菜单中，选择*配置*。

   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。

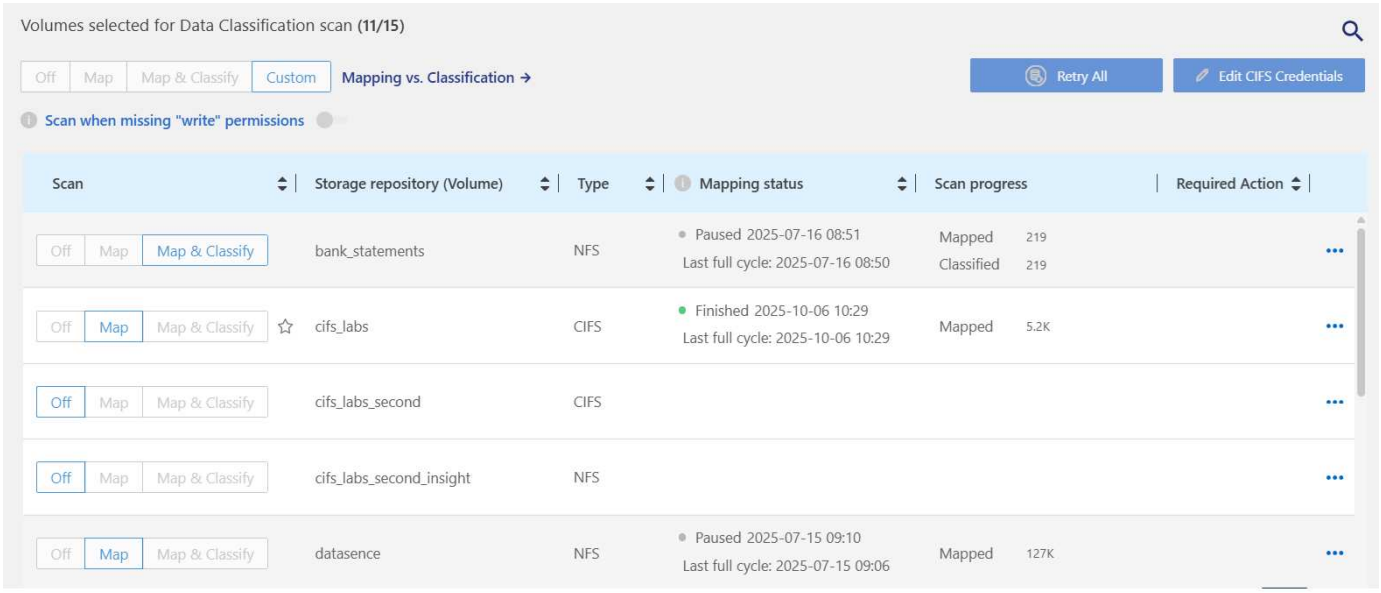
启用或禁用卷扫描

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

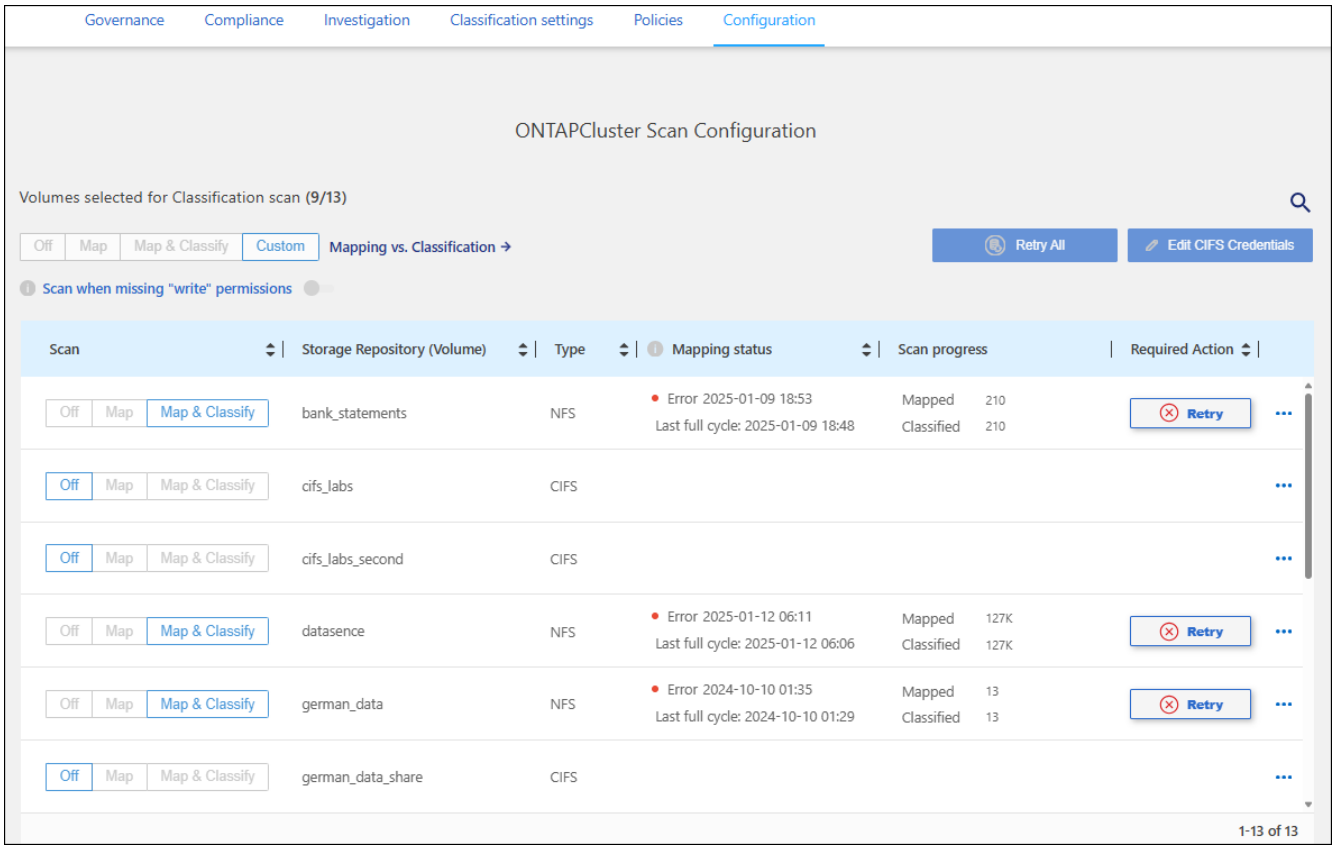
- 确保数据分类实例与包含Cloud Volumes ONTAP或本地ONTAP集群的卷的每个网络之间存在网络连接。
- 确保Cloud Volumes ONTAP的安全组允许来自数据分类实例的入站流量。

  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。

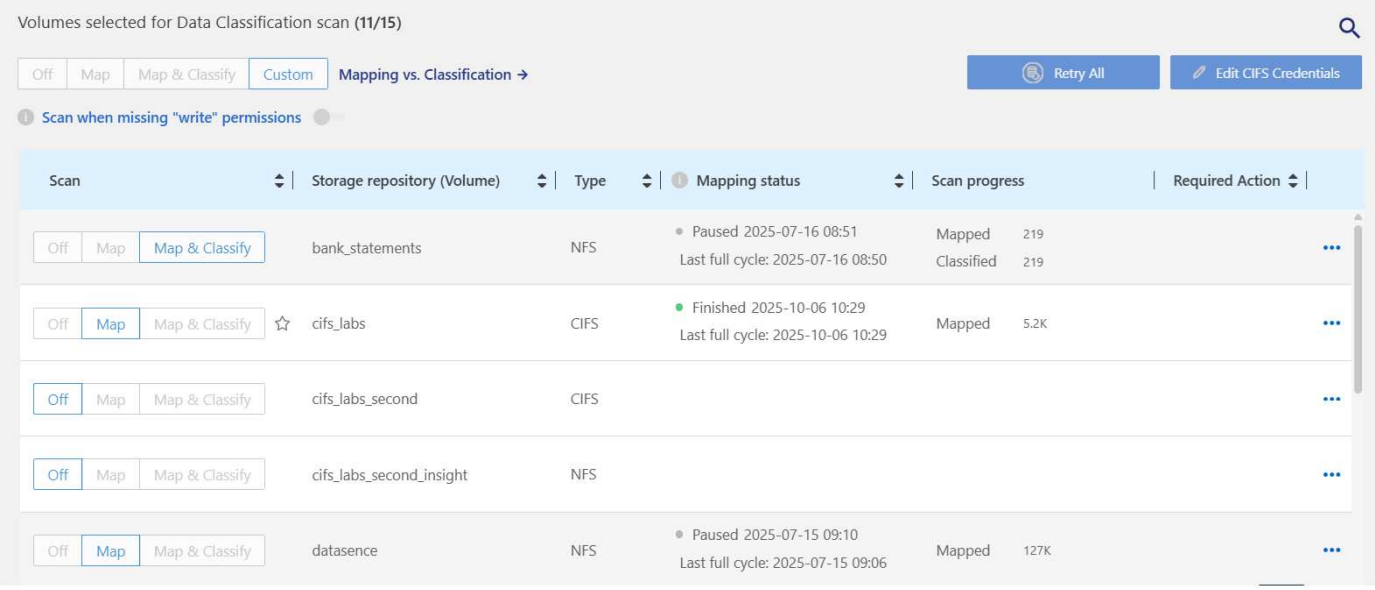
启用或禁用卷扫描

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。

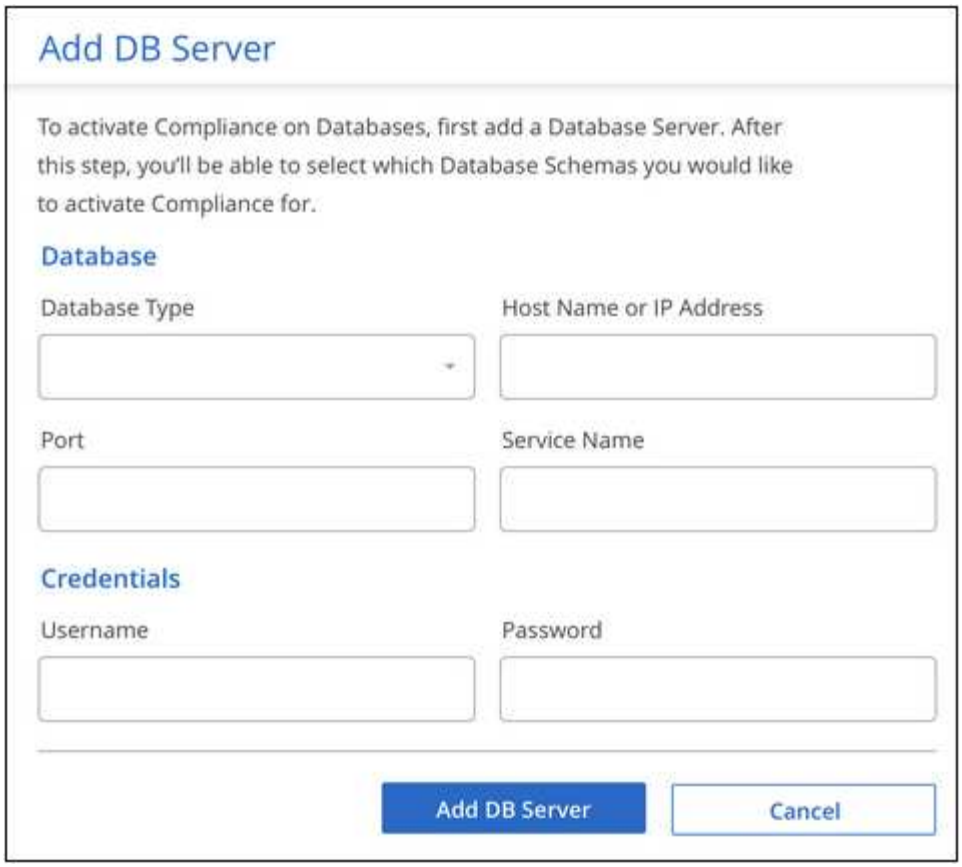
如果您正在扫描可通过互联网访问的数据库模式，您可以"在云中部署数据分类"或者"在可以访问互联网的本地位置部署数据分类"。

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。

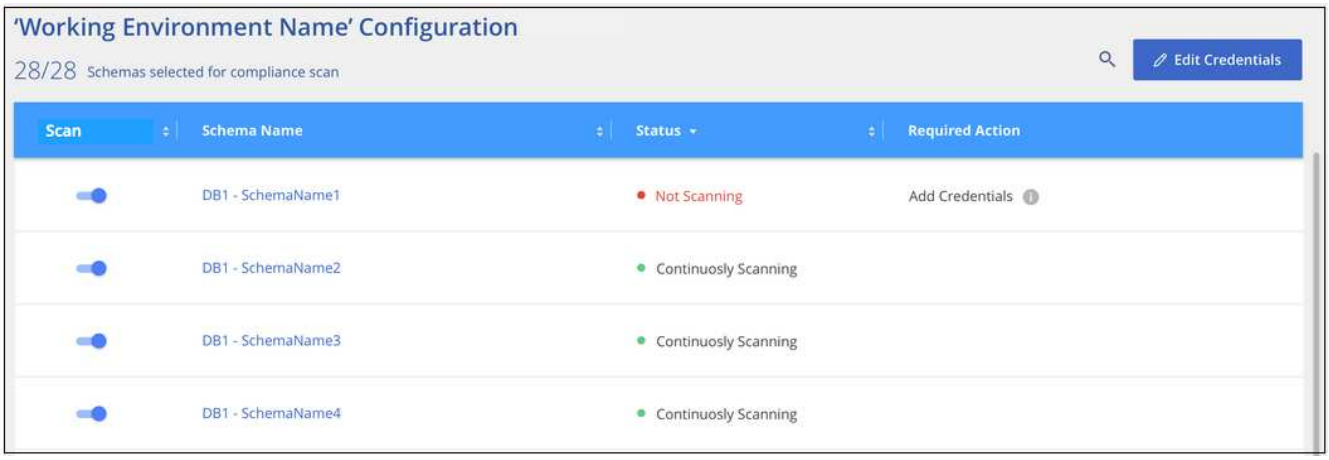
启用和禁用数据库模式扫描

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

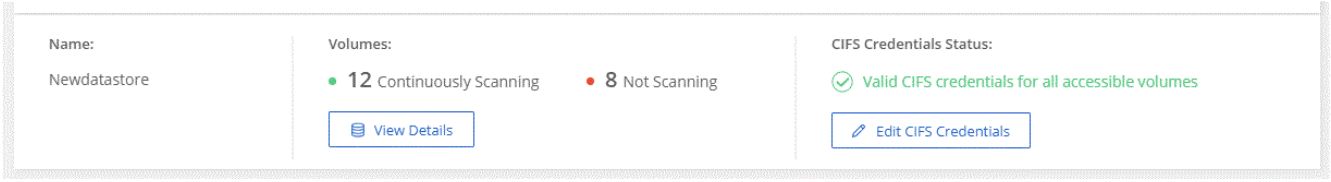
1. 从数据分类菜单中，选择*配置*。

   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。

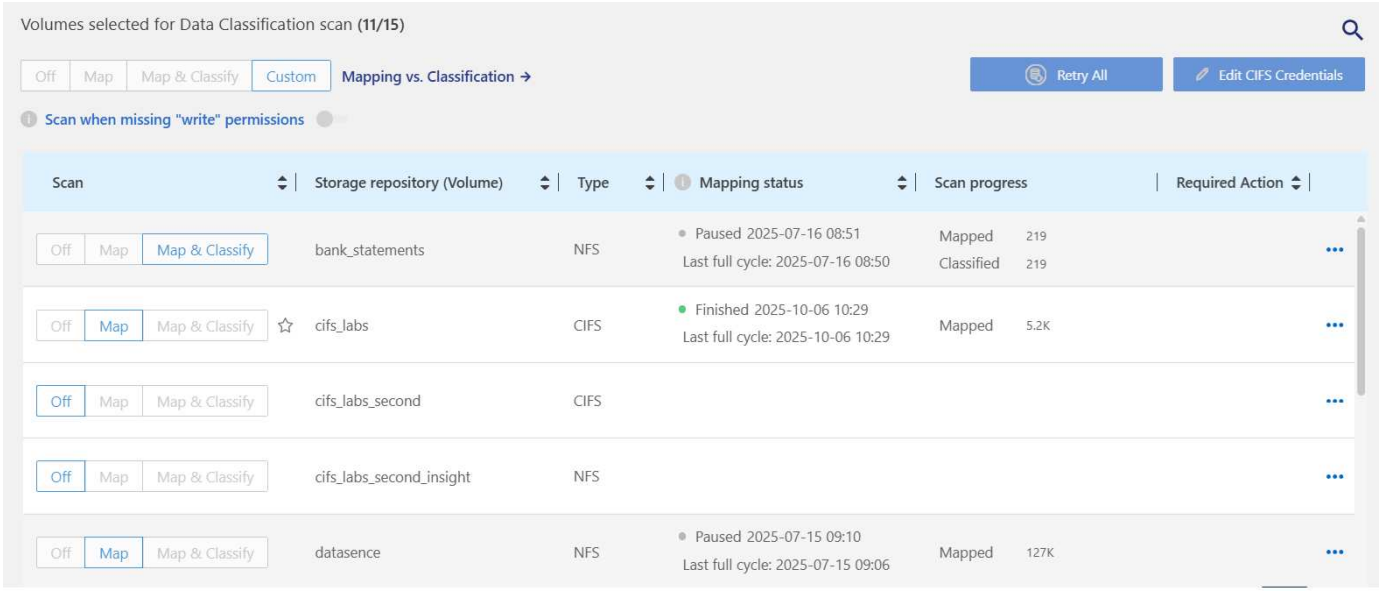
启用和禁用卷上的扫描

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

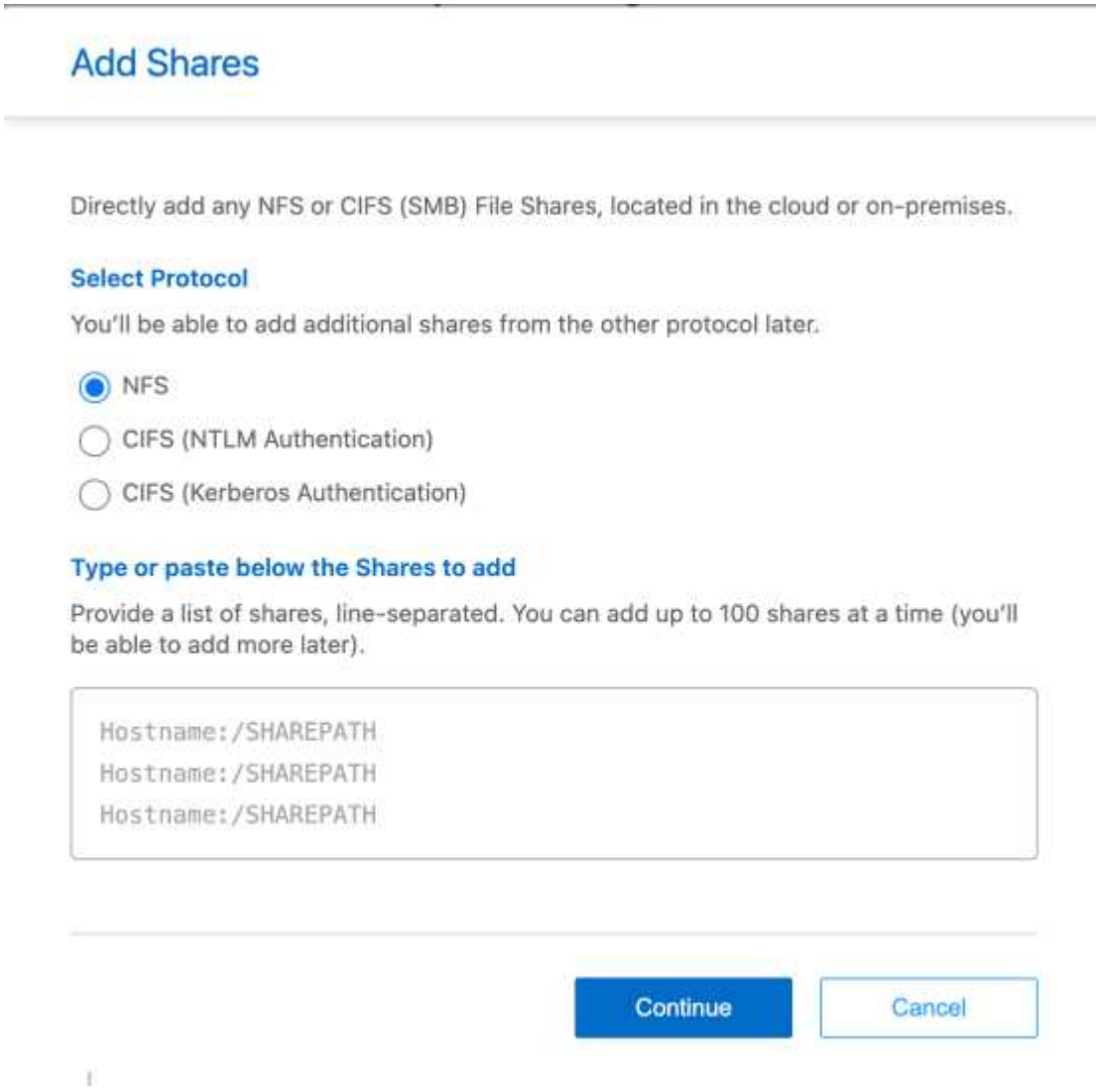
将文件共享添加到组时，必须使用格式 `<host_name>:/<share_path>`。

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。

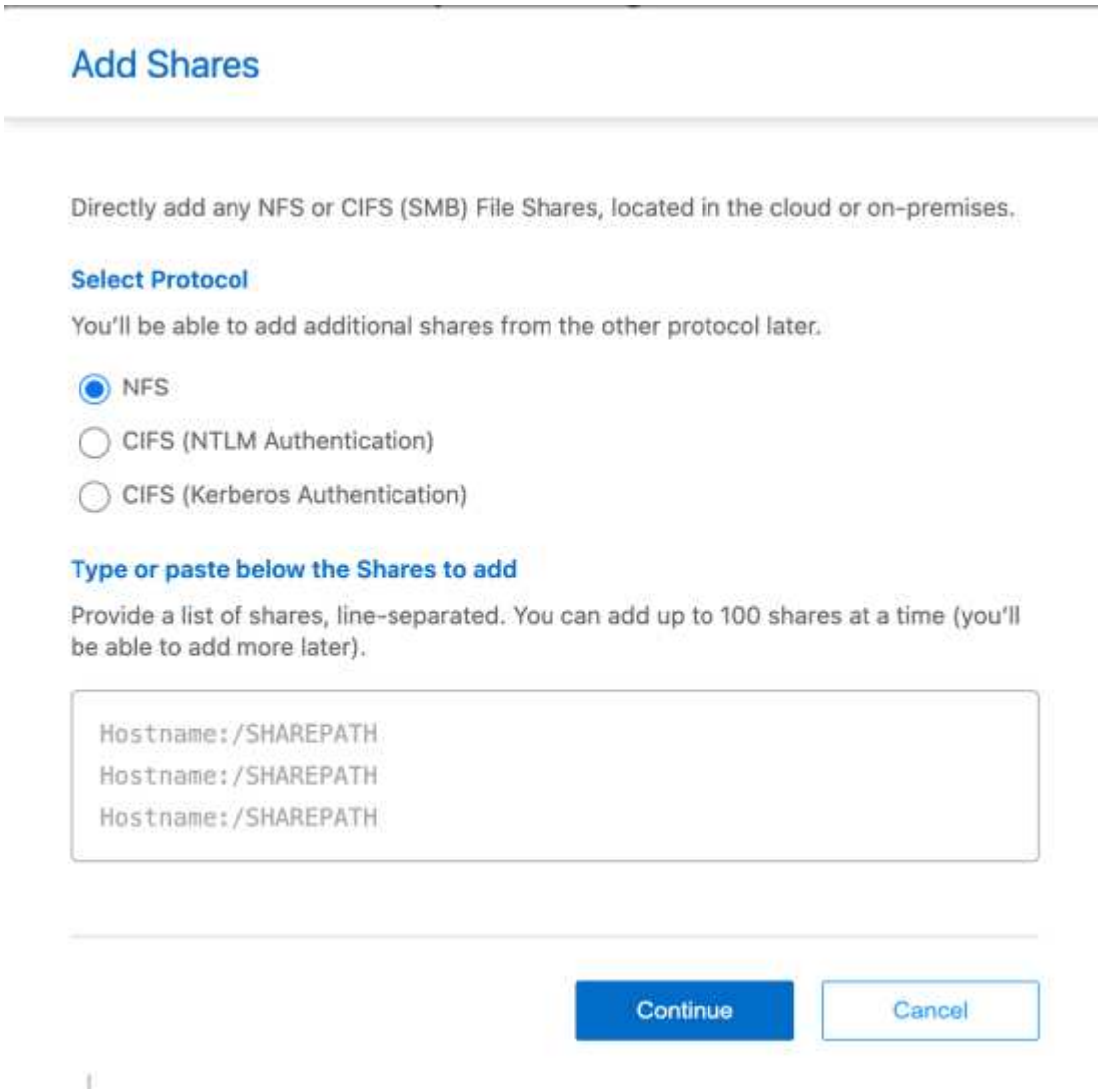
编辑 **CIFS** 协议配置

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

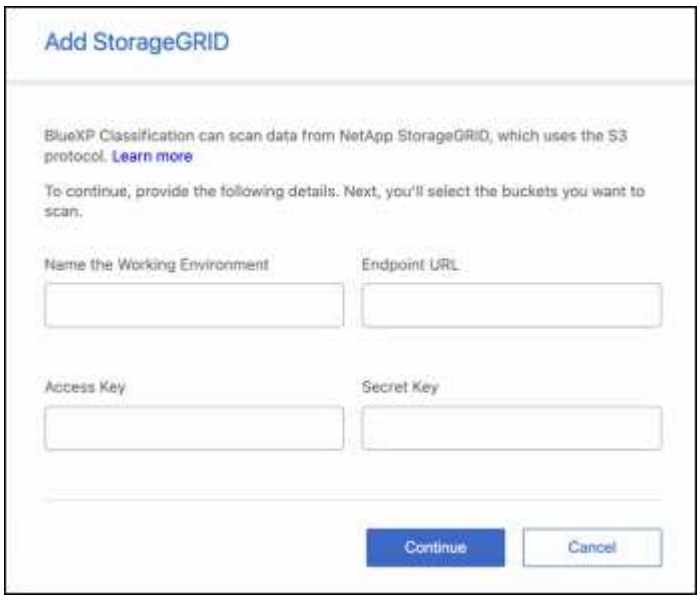
添加StorageGRID服务。

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：

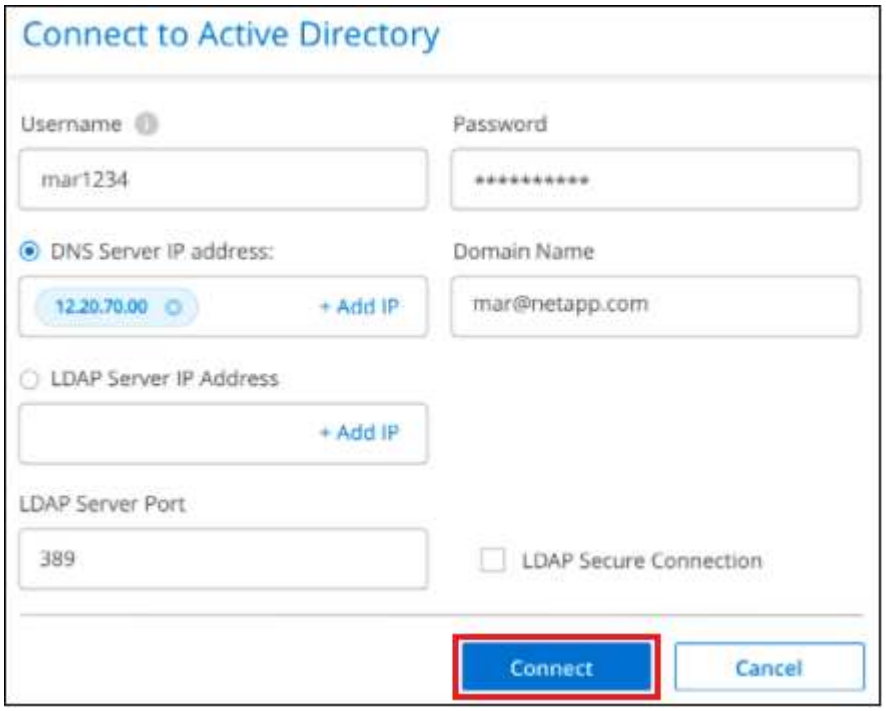
协议	端口	目标	目的
TCP 和 UDP	389	Active Directory	LDAP
TCP	636	Active Directory	基于 SSL 的 LDAP
TCP	3268	Active Directory	全局目录
TCP	3269	Active Directory	通过 SSL 的全局目录

步骤

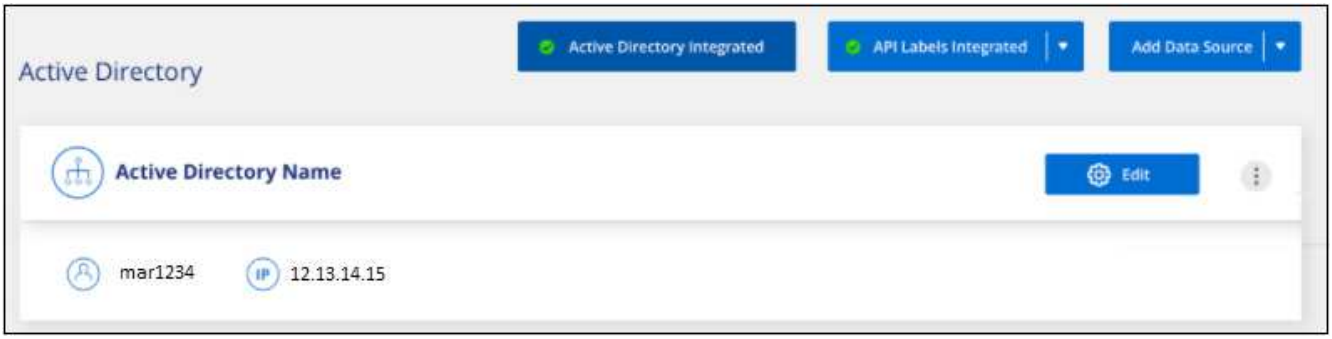
1. 在数据分类配置页面中，单击*添加 Active Directory*。

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*。