激活数据源扫描
NetApp Data Classification

NetApp
February 11, 2026

目录

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
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。

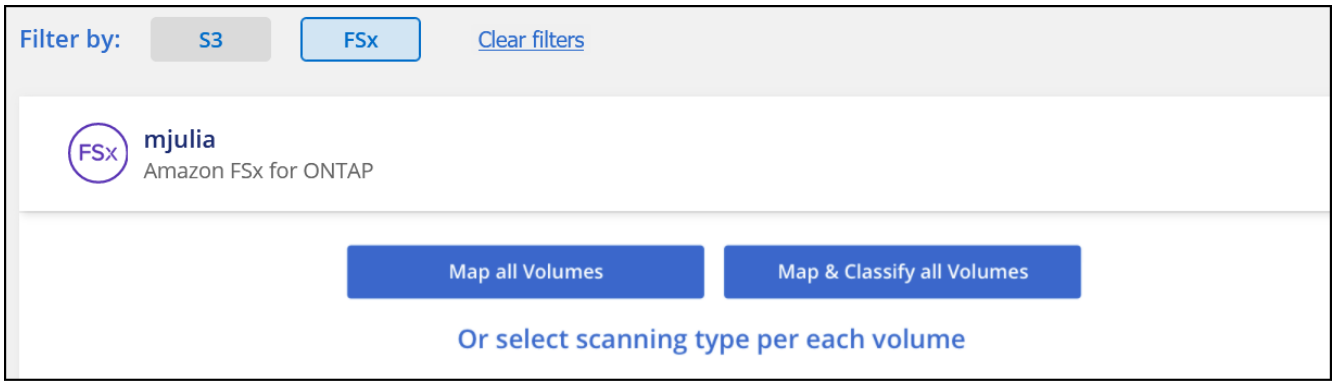
*注意：*扫描 FSx 卷时，当前不支持在本地位置部署数据分类。

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

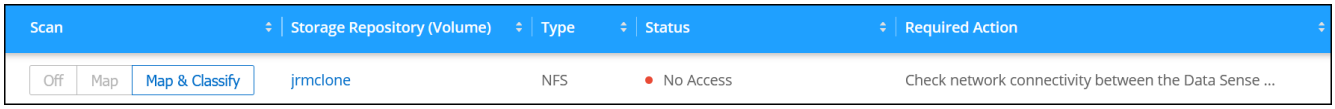
通过检查网络、安全组和导出策略，确保数据分类可以访问卷。

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。

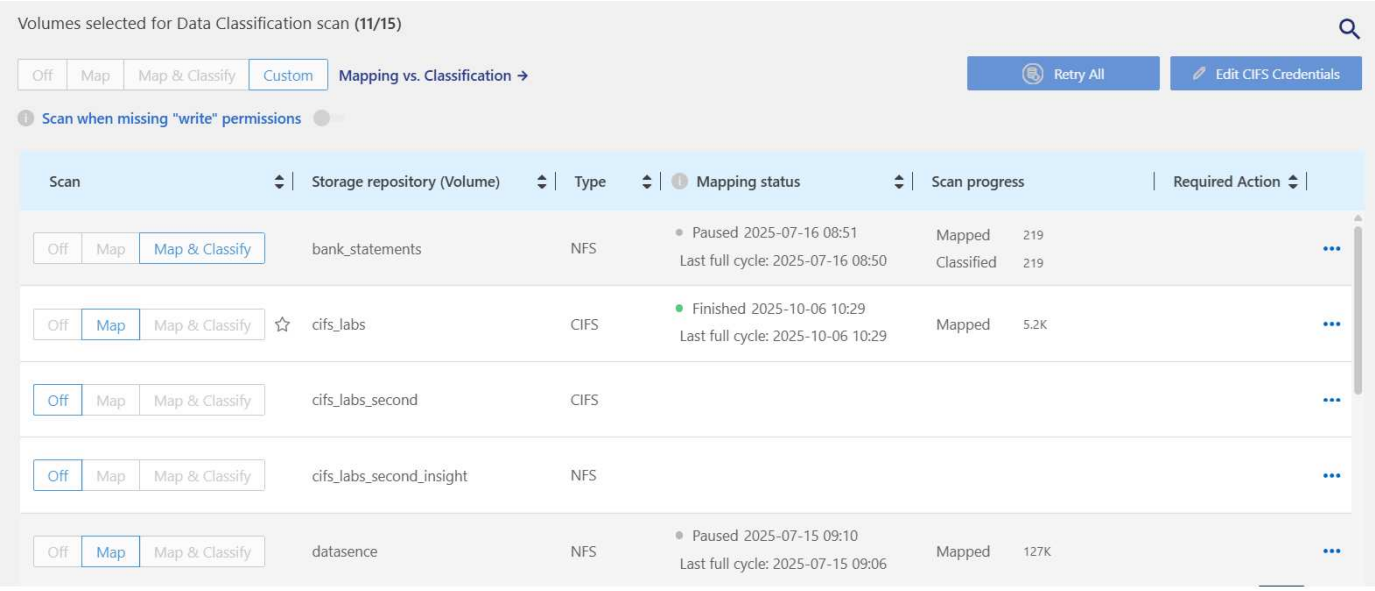
启用和禁用卷上的扫描

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。

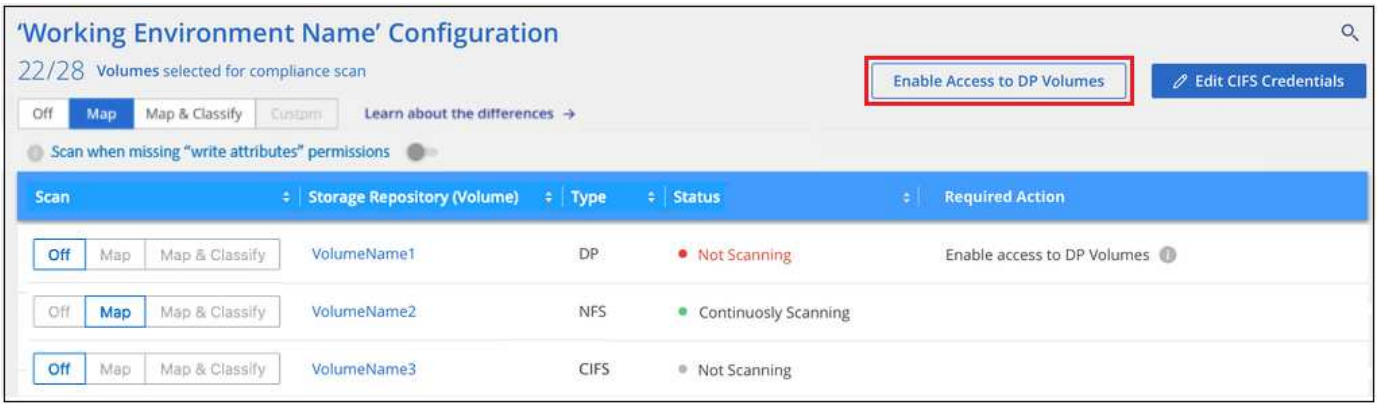
结果

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。

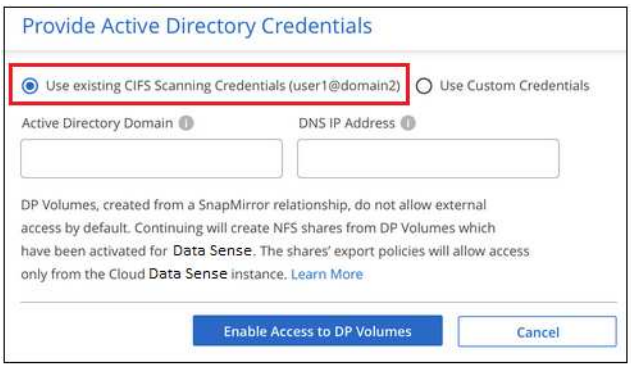
最初，卷列表将这些卷标识为_类型_ **DP**，其_状态_ 未扫描*和_所需操作_ *启用对 **DP** 卷的访问。

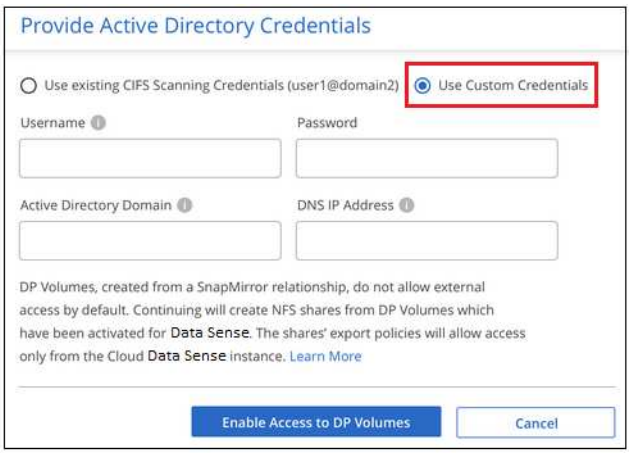


步骤

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

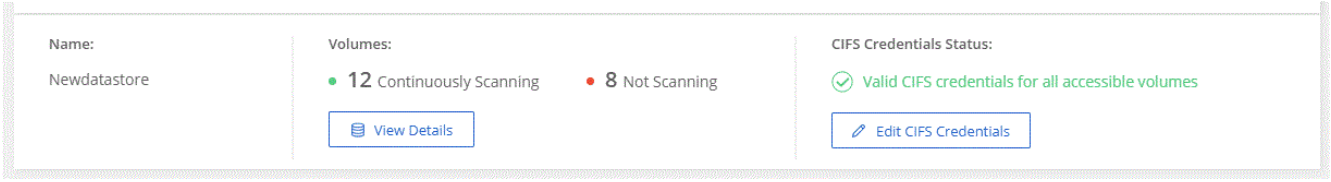
1. 从数据分类菜单中，选择*配置*。

   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。

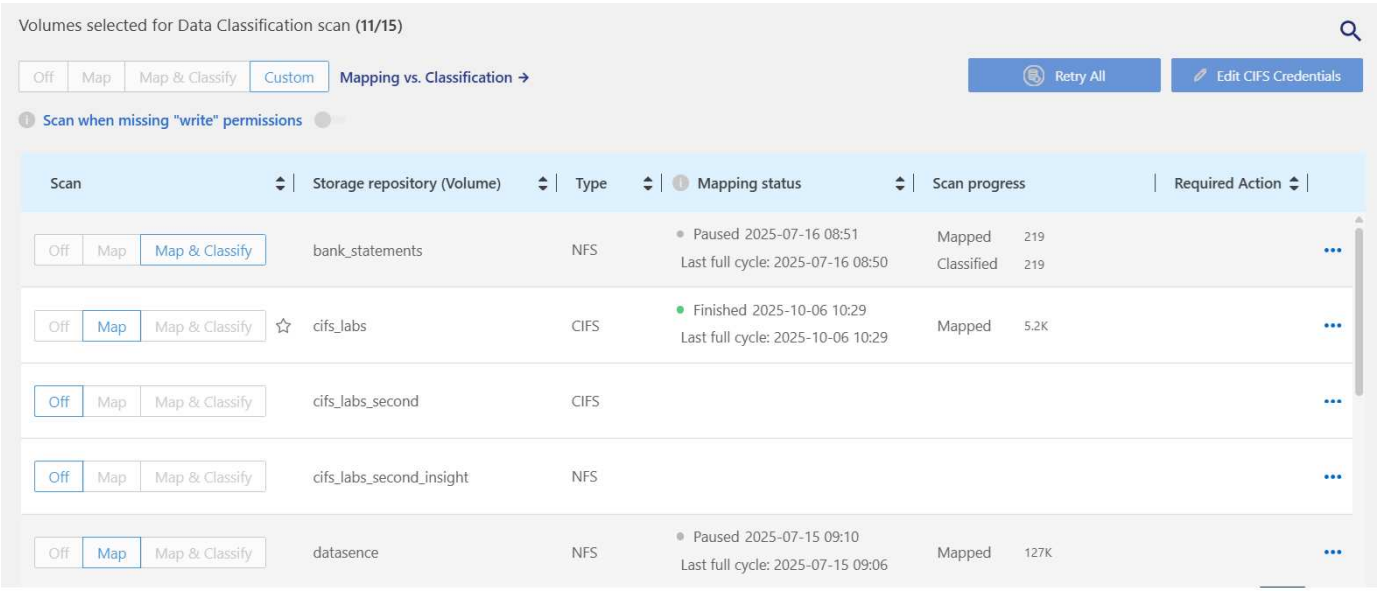
启用或禁用卷扫描

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

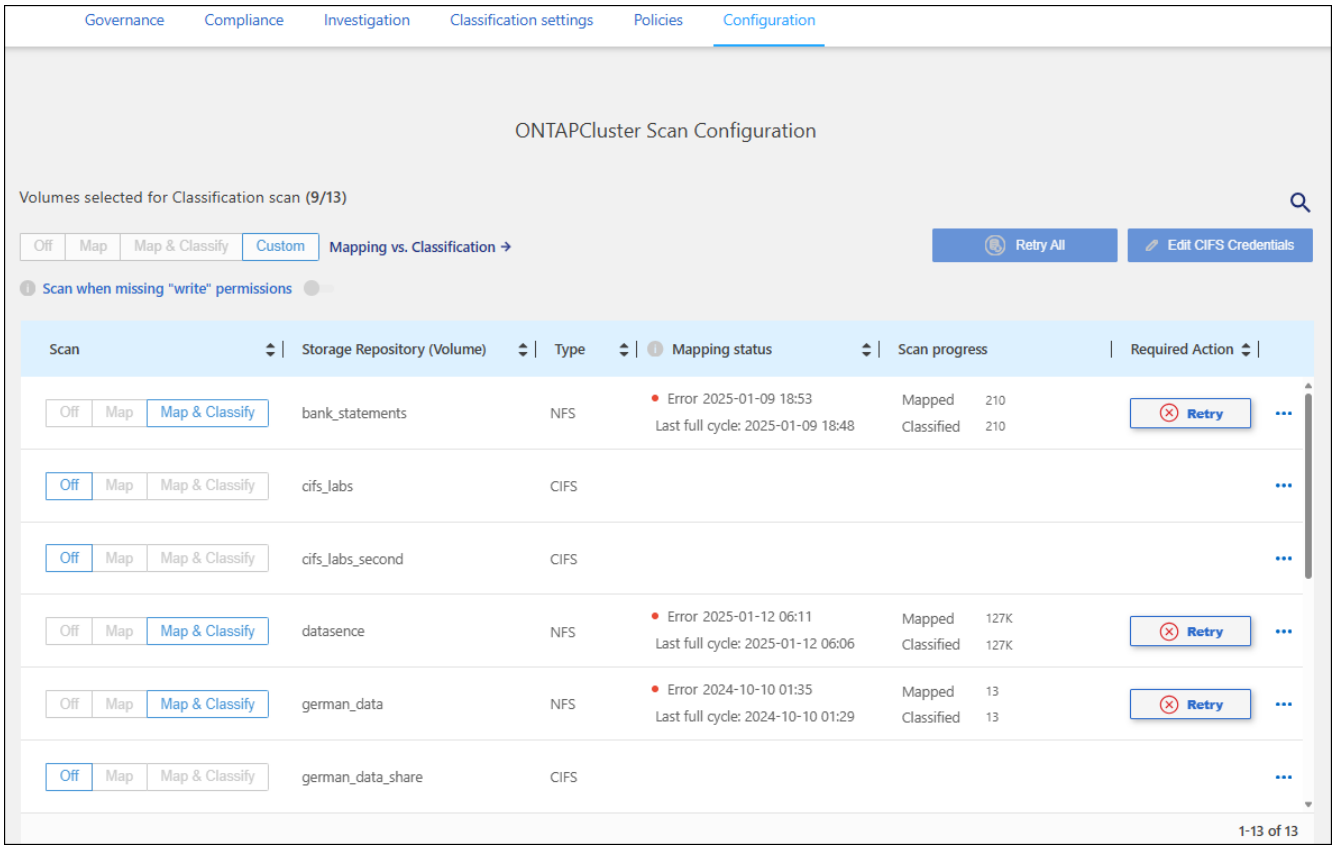
- 确保数据分类实例与包含Cloud Volumes ONTAP或本地ONTAP集群的卷的每个网络之间存在网络连接。
- 确保Cloud Volumes ONTAP的安全组允许来自数据分类实例的入站流量。

  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。

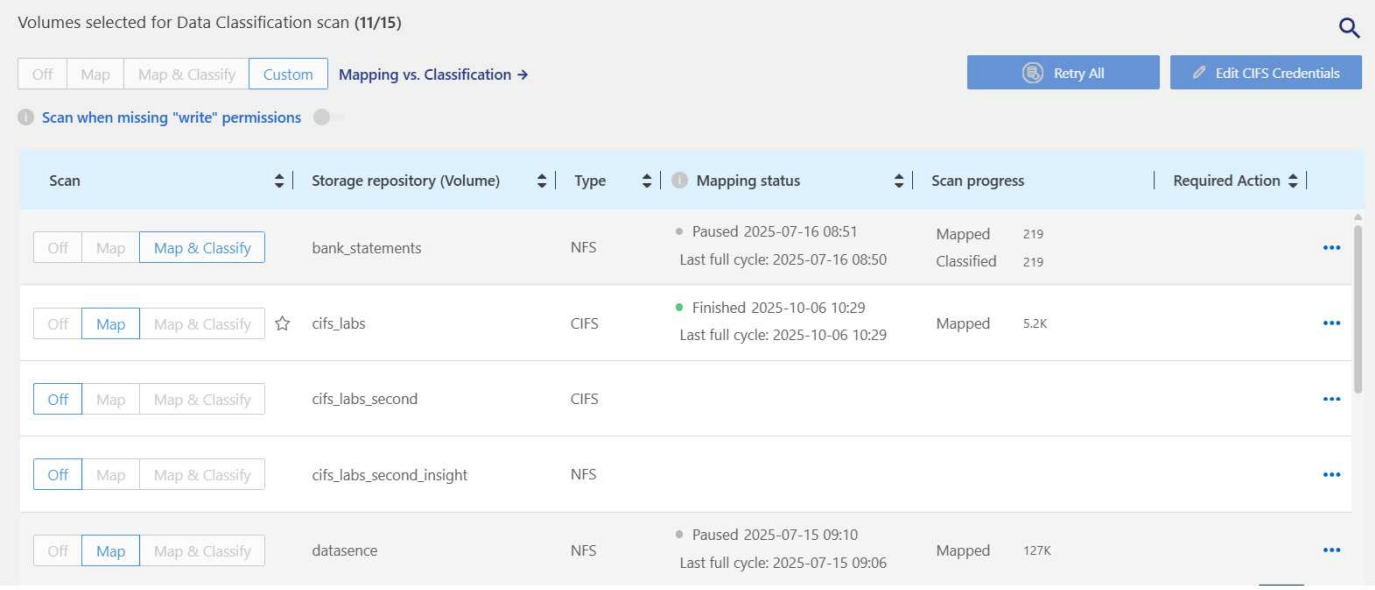
启用或禁用卷扫描

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。

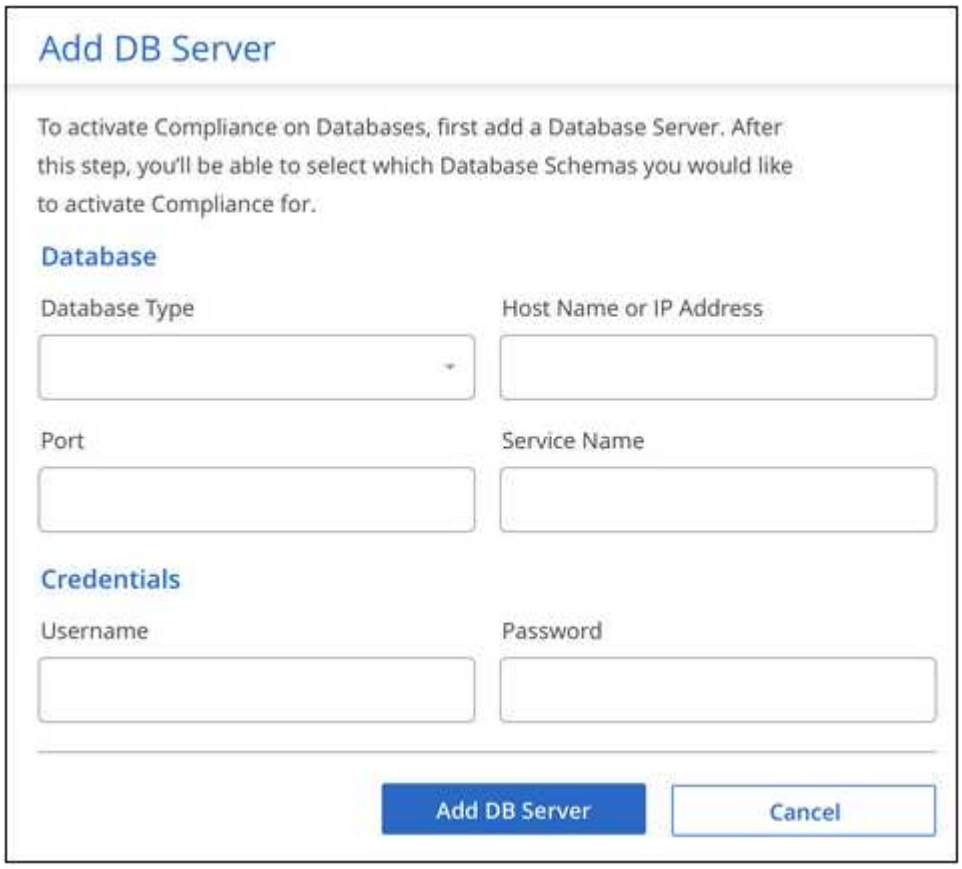
添加数据库服务器

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。

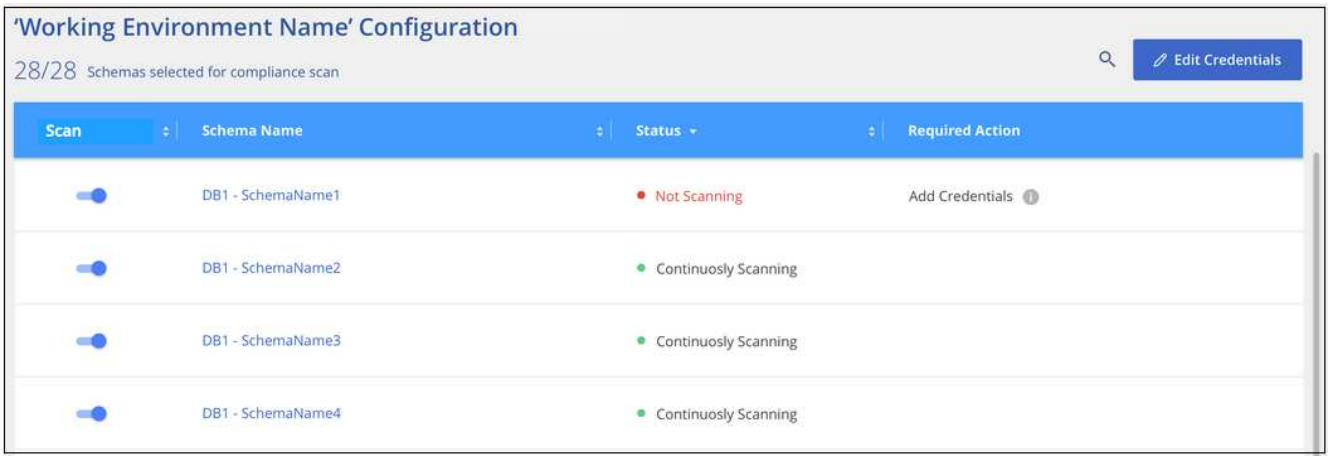
启用和禁用数据库模式扫描

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

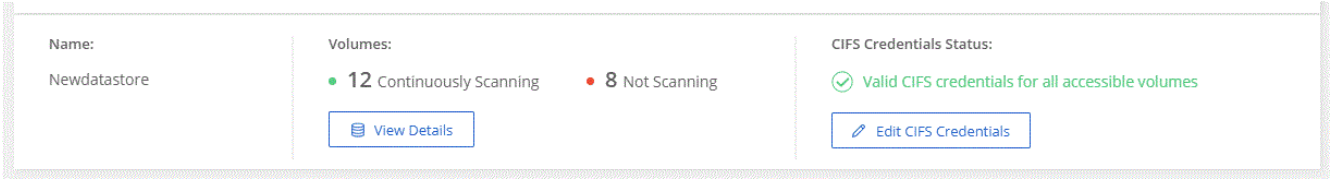
1. 从数据分类菜单中，选择*配置*。

   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。

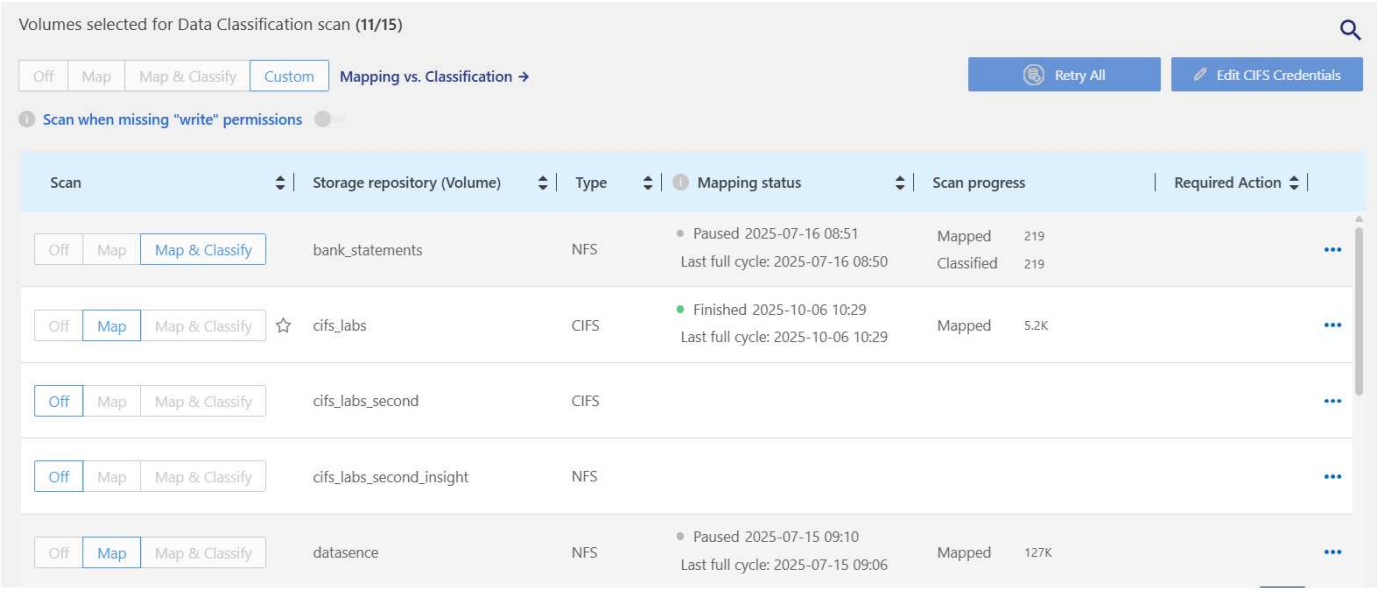
启用和禁用卷上的扫描

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

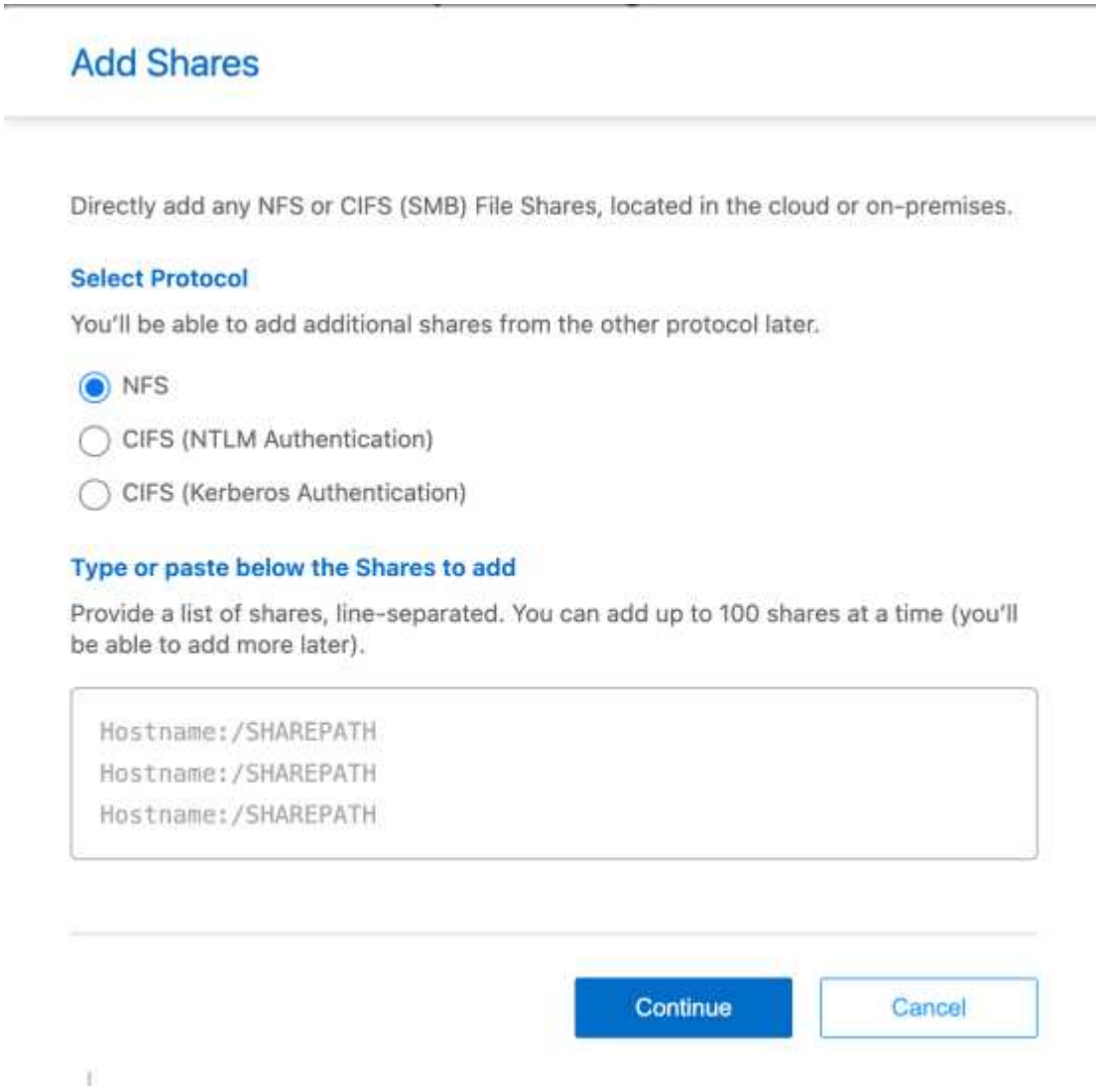
将文件共享添加到组时，必须使用格式 `<host_name>:/<share_path>`。

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。

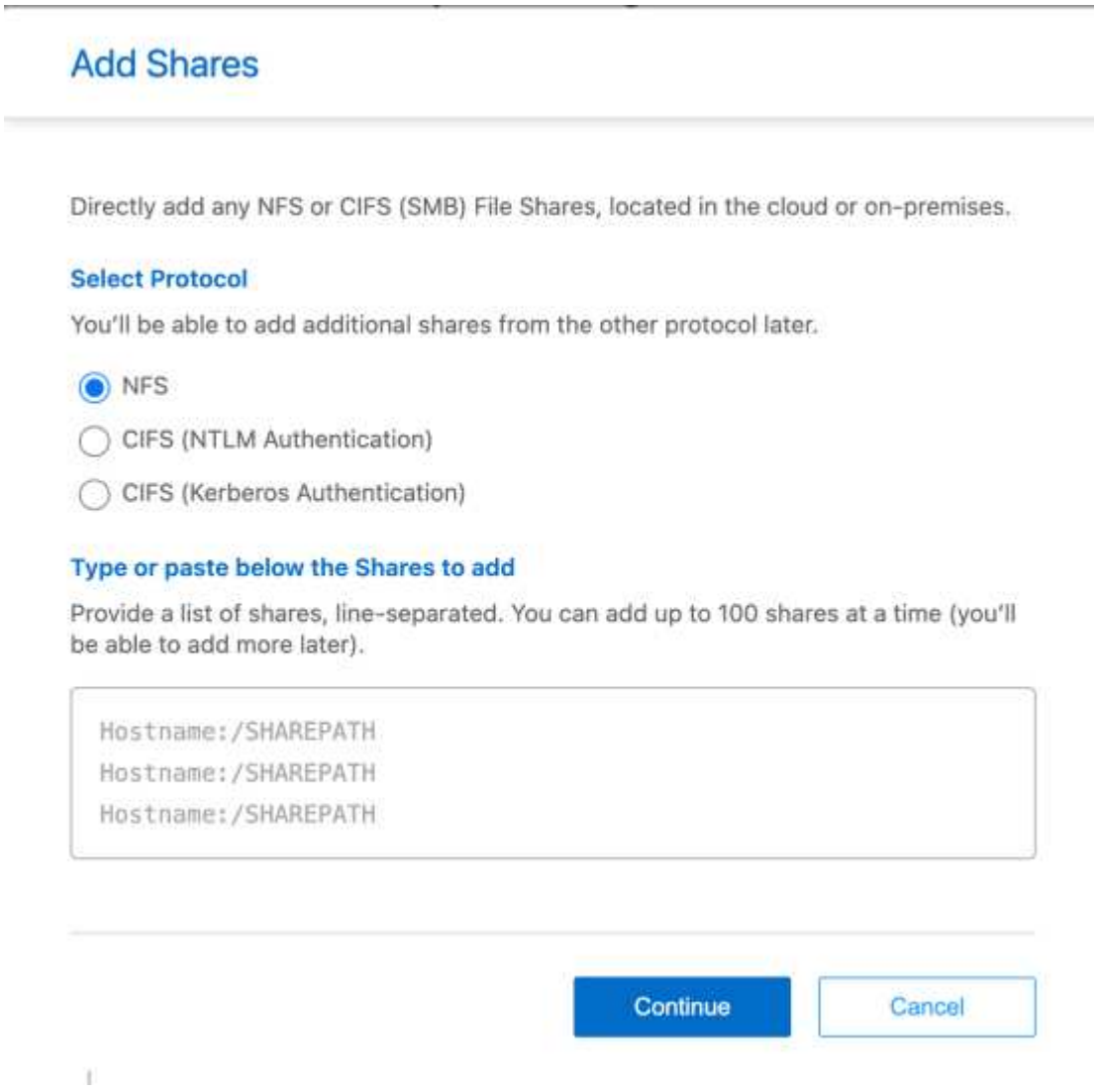
编辑 **CIFS** 协议配置

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。

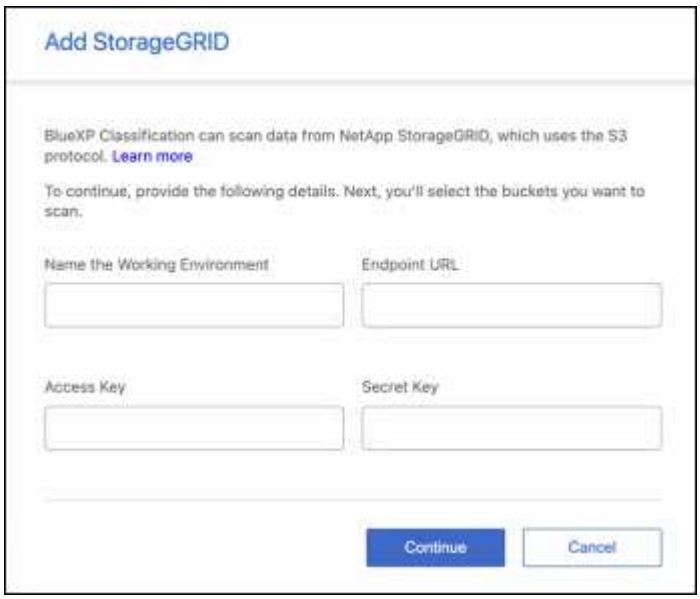
步骤

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。