



# 已弃用的功能

## BlueXP classification

NetApp  
June 14, 2024

# 目录

已弃用的功能 .....	1
BlueXP分类已弃用的功能 .....	1
部署BlueXP分类弃用 .....	2
扫描数据弃用 .....	10
管理数据弃用 .....	31

# 已弃用的功能

## BlueXP分类已弃用的功能

BlueXP分类作为BlueXP中的核心功能提供、无需额外付费。通过将BlueXP分类作为所有客户均可使用的核心BlueXP功能、NetApp使您能够使用核心功能访问定制的数据管理。

从BlueXP核心版本1.31及更高版本开始、有些特性和功能已弃用、但在原有版本1.3及更低版本中仍受支持。

### 支持的数据源

数据源	原有版本1.3及更早版本	BlueXP核心版本1.31及更高版本
Cloud Volumes ONTAP（部署在 AWS， Azure 或 GCP 中）	是的。	是的。
内部 ONTAP 集群	是的。	是的。
StorageGRID	是的。	否
Azure NetApp Files	是的。	是的。
适用于 ONTAP 的 Amazon FSX	是的。	是的。
Google Cloud NetApp卷	是的。	是的。
适用于 Google Cloud 的 Cloud Volumes Service	是的。	是的。
数据库	是的。	是的。
Amazon S3	是的。	否
Google Cloud 存储	是的。	否
OneDrive	是的。	否
SharePoint Online	是的。	否
SharePoint内部部署(SharePoint Server)	是的。	否
Google Drive	是的。	否

### 合规性功能

功能	原有版本1.3及更早版本	BlueXP核心版本1.31及更高版本
识别个人信息（PII）	是的。	是的。
识别敏感的个人数据	是的。	是的。
响应数据主体访问请求（DSAR）	是的。	是的。
创建已标识的自定义"个人数据"列表	是的。	否
当文件包含特定的身份信息时、通过电子邮件通知用户。(您可以使用定义此条件 <a href="#">"策略"</a> )	是的。	否

功能	原有版本1.3及更早版本	BlueXP核心版本1.3.1及更高版本
使用目录级筛选器	是的。	是的。
使用目录级的分析	是的。	否

## 用于管理数据的功能

功能	原有版本1.3及更早版本	BlueXP核心版本1.3.1及更高版本
移动、复制和删除源文件	是的。	否
使用状态标记对数据进行分类	是的。	否
使用AIP标签对数据进行分类	是的。	否
将文件分配给用户	是的。	否
按需重新扫描数据	是的。	否
创建自定义的定义器	是的。	否
从扫描中排除目录	是的。	是的。
搜索文件中的名称	是的。	是的。
从调查中将数据导出到NFS	是的。	否
从调查中将数据导出至CSV	是的。	是的。
支持多个扫描器	是的。	否
集成 Active Directory	是的。	是的。
使用权限分析和筛选器	是的。	是的。
使用文件卡	是的。	是的。
使用热图	是的。	是的。
使用信息板和文件卡上的操作	是的。	否
使用文件访问审核日志记录	是的。	否
从配置页面启用文件访问	是的。	否
使用某些预定义策略	是的。	否

## 部署BlueXP分类弃用

对于无法访问Internet的大型配置、在多个主机上安装BlueXP分类

完成几个步骤、在无法访问Internet的内部站点中的多个主机上安装BlueXP分类、也称为 `_private mode_`。此类安装非常适合您的安全站点。

对于需要扫描站点中数PB的数据而不能访问Internet的超大型配置、您可以包含多个主机以提供额外的处理能力。使用多个主机系统时，主系统称为 *Manager node*，提供额外处理能力的其他系统称为 扫描 程序 *nodes*。

在脱机环境中的多个内部主机上安装BlueXP分类软件时、请按照以下步骤进行操作。

\*注\*此信息仅与BlueXP分类的旧版版本1.3及更早版本相关。

#### 您需要的内容

- 验证管理器和扫描程序节点的所有Linux系统是否满足主机要求。
- 确认已安装两个必备软件包(Docker Engine或Podman以及Python 3)。
- 确保您在 Linux 系统上具有 root 权限。
- 验证脱机环境是否满足所需的权限和连接要求。
- 您必须具有计划使用的扫描程序节点主机的 IP 地址。
- 必须在所有主机上启用以下端口和协议：

Port	协议	Description
2377	TCP	集群管理通信
7946	TCP , UDP	节点间通信
4789	UDP	覆盖网络流量
50	电子服务	加密的 IPsec 覆盖网络 ( ESP ) 流量
111.	TCP , UDP	用于在主机之间共享文件的 NFS 服务器 (需要从每个扫描程序节点到管理器节点)
2049.	TCP , UDP	用于在主机之间共享文件的 NFS 服务器 (需要从每个扫描程序节点到管理器节点)

#### 步骤

1. 按照中的步骤 1 至 8 进行操作 **"单主机安装"** 在管理器节点上。
2. 如步骤 9 所示，在安装程序提示时，您可以在一系列提示中输入所需值，也可以将所需参数作为命令行参数提供给安装程序。

除了可用于单主机安装的变量之外，还会使用一个新选项 `* -n <node_IP>*` 来指定扫描程序节点的 IP 地址。多个节点 IP 以逗号分隔。

例如、此命令将添加3个扫描程序节点：

```
sudo ./install.sh -a <account_id> -c <client_id> -t <user_token> --host  
<ds_host> --manager-host <cm_host> -n <node_ip1>,<node_ip2>,<node_ip3> --no  
-proxy --darksite
```

3. 在管理器节点安装完成之前，将显示一个对话框，其中显示了扫描程序节点所需的安装命令。复制命令(例如：`sudo ./node_install.sh -m 10.11.12.13 -t ABCDEF-1-3u69m1-1s35212`)并将其保存在文本文件中。
4. 在 \* 每个 \* 扫描程序节点主机上：
  - a. 将Data sense安装程序文件(\* `cc_onprem_installer.tar.gz`\*)复制到主机。
  - b. 解压缩安装程序文件。

c. 粘贴并运行在步骤 3 中复制的命令。

在所有扫描程序节点上完成安装且这些节点已加入管理器节点后，管理器节点安装也会完成。

## 结果

BlueXP分类安装程序完成软件包安装并注册安装。安装可能需要 15 到 25 分钟。

## 下一步行动

在配置页面中，您可以选择本地 "内部 ONTAP 集群" 和本地 "数据库" 要扫描的。

## 将扫描程序节点添加到现有部署中

您可以将扫描程序节点添加到可访问Internet的Linux主机上的现有部署中。

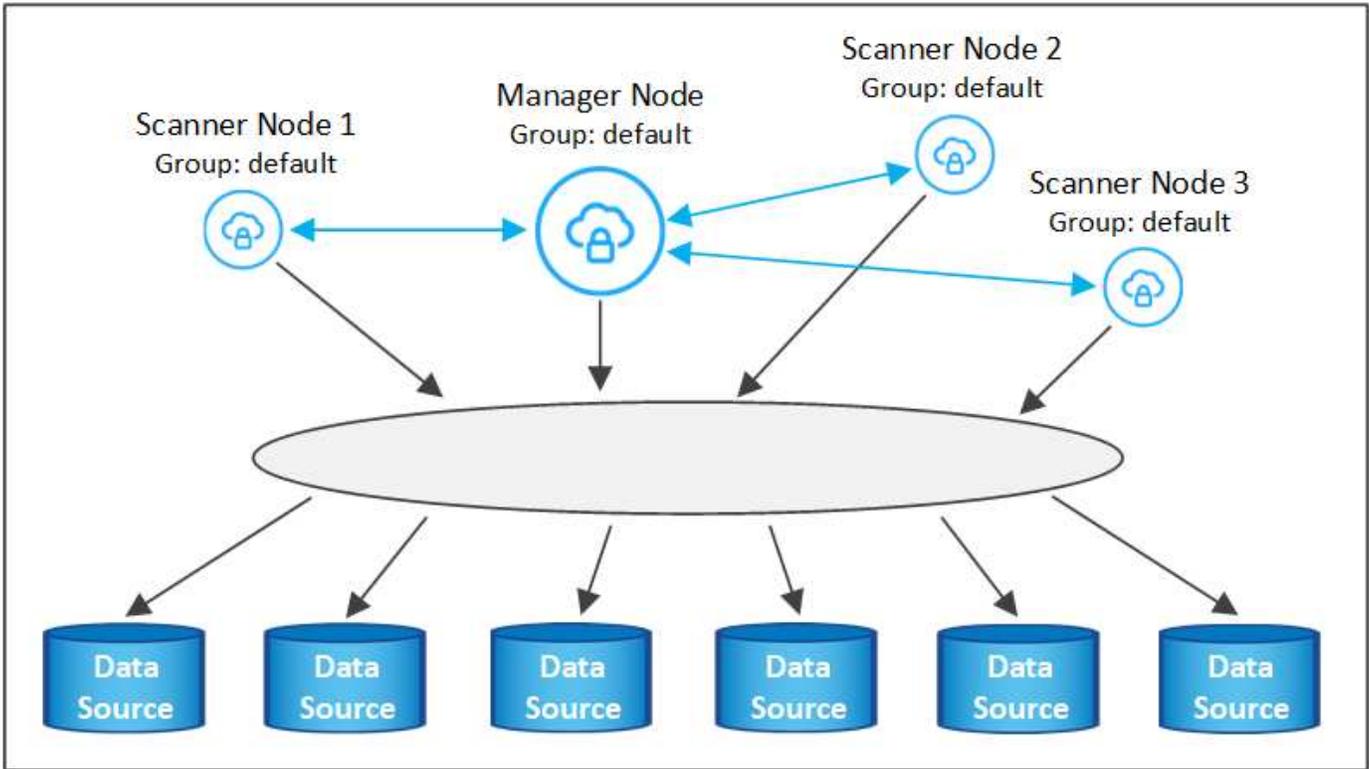
如果您发现扫描数据源需要更多扫描处理能力、则可以添加更多扫描程序节点。您可以在安装管理器节点后立即添加扫描程序节点、也可以稍后添加扫描程序节点。例如、如果您意识到一个数据源中的数据量在6个月后增加了一倍或增加了三倍、则可以添加一个新的扫描程序节点来协助进行数据扫描。

\*注\*此信息仅与BlueXP分类的旧版版本1.3及更早版本相关。

您可以通过两种方式添加其他扫描程序节点：

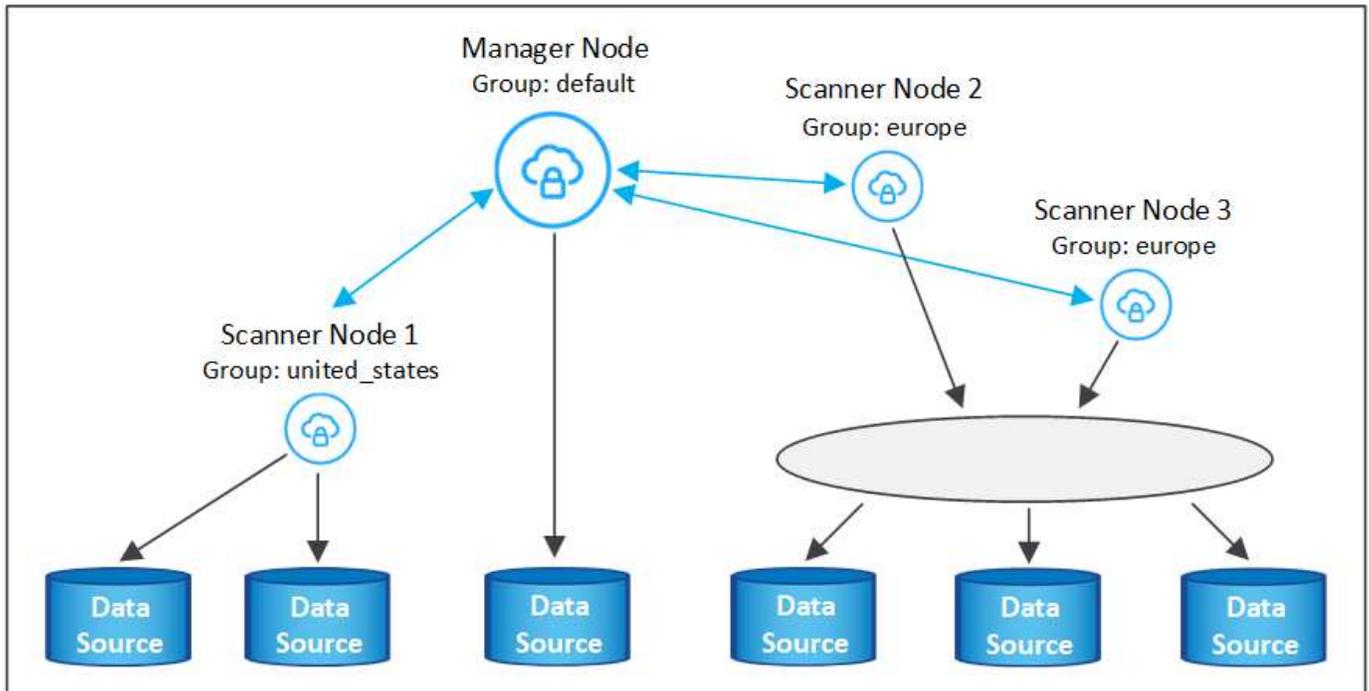
- 添加一个节点以协助扫描所有数据源
- 添加节点以协助扫描特定数据源或特定数据源组(通常基于位置)

默认情况下、您添加的任何新扫描程序节点都会添加到常规扫描资源池中。这称为"默认扫描程序组"。在下图中、"默认"组中有1个管理器节点和3个扫描程序节点、这些节点均扫描所有6个数据源中的数据。



如果您的某些数据源要由物理上更接近数据源的扫描程序节点扫描，则可以定义一个扫描程序节点或一组扫描程序节点、以扫描特定数据源或一组数据源。在下图中、有1个管理器节点和3个扫描程序节点。

- 管理器节点位于"默认"组中、它正在扫描1个数据源
- 扫描程序节点1位于"United States"组中、它正在扫描2个数据源
- 扫描程序节点2和3属于"欧洲"组、它们共享3个数据源的扫描任务



BlueXP分类扫描程序组可以定义为存储数据的单独地理区域。您可以在全球部署多个BlueXP分类扫描程序节

点、并为每个节点选择一个扫描程序组。这样、每个扫描程序节点都会扫描与其最接近的数据。扫描程序节点与数据的距离越近、越好、因为它可以在扫描数据时尽可能地减少网络延迟。

您可以选择要添加到BlueXP分类的扫描程序组、也可以选择其名称。BlueXP分类不会强制将映射到名为"Euro"的扫描程序组的节点部署在欧洲。

您将按照以下步骤安装其他BlueXP分类扫描程序节点：

1. 准备用作扫描程序节点的Linux主机系统
2. 将Data sense软件下载到这些Linux系统
3. 在管理器节点上运行命令以确定扫描程序节点
4. 按照以下步骤在扫描程序节点上部署软件(也可以为某些扫描程序节点定义"扫描程序组")
5. 如果定义了扫描程序组、请在管理器节点上：
  - a. 打开文件"工作环境\_to\_scanner\_group\_config.yml"、并定义每个扫描程序组要扫描的工作环境
  - b. 运行以下脚本、将此映射信息注册到所有扫描程序节点：

```
update_we_scanner_group_from_config_file.sh
```

您需要的内容

- 验证扫描程序节点的所有Linux系统是否满足主机要求。
- 验证系统是否已安装两个必备软件包(Docker Engine或Podman以及Python 3)。
- 确保您在 Linux 系统上具有 root 权限。
- 验证您的环境是否满足所需的权限和连接。
- 您必须具有要添加的扫描程序节点主机的IP地址。
- 您必须知道BlueXP分类管理器节点主机系统的IP地址
- 您必须具有连接器系统的IP地址或主机名、NetApp帐户ID、连接器客户端ID和用户访问令牌。如果您计划使用扫描程序组、则需要知道帐户中每个数据源的工作环境ID。要获取此信息、请参见下面的\*前提条件步骤\_\*
- 必须在所有主机上启用以下端口和协议：

Port	协议	Description
2377	TCP	集群管理通信
7946	TCP , UDP	节点间通信
4789	UDP	覆盖网络流量
50	电子服务	加密的 IPsec 覆盖网络 ( ESP ) 流量
111.	TCP , UDP	用于在主机之间共享文件的 NFS 服务器 (需要从每个扫描程序节点到管理器节点)
2049.	TCP , UDP	用于在主机之间共享文件的 NFS 服务器 (需要从每个扫描程序节点到管理器节点)

- 如果您使用的是 ... firewalld 在BlueXP分类计算机上、我们建议您在安装BlueXP分类之前启用它。运行以下命令进行配置 firewalld 以便与BlueXP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 --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
firewall-cmd --reload

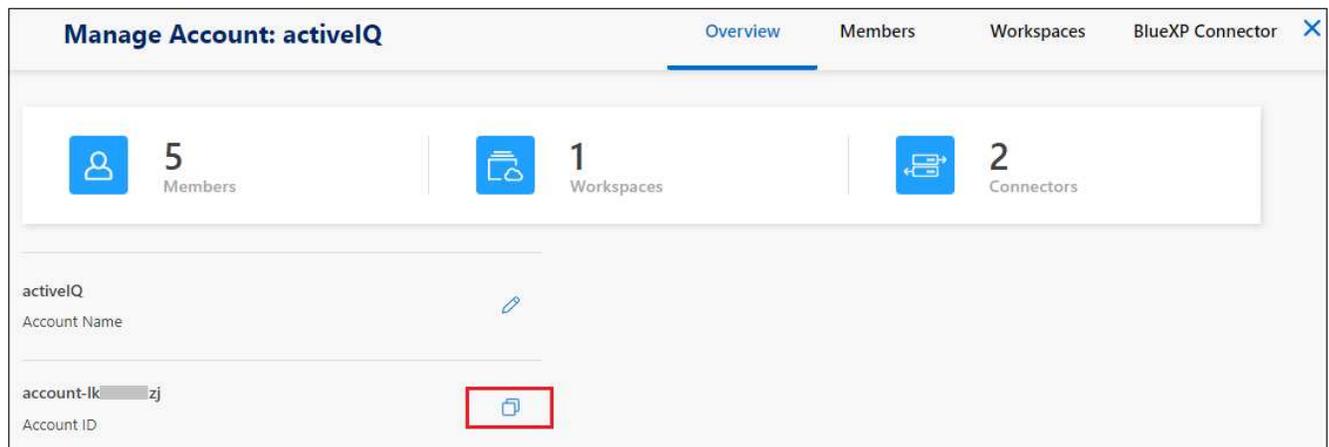
```

请注意、每当启用或更新时、都必须重新启动Docker或Podman firewalld 设置。

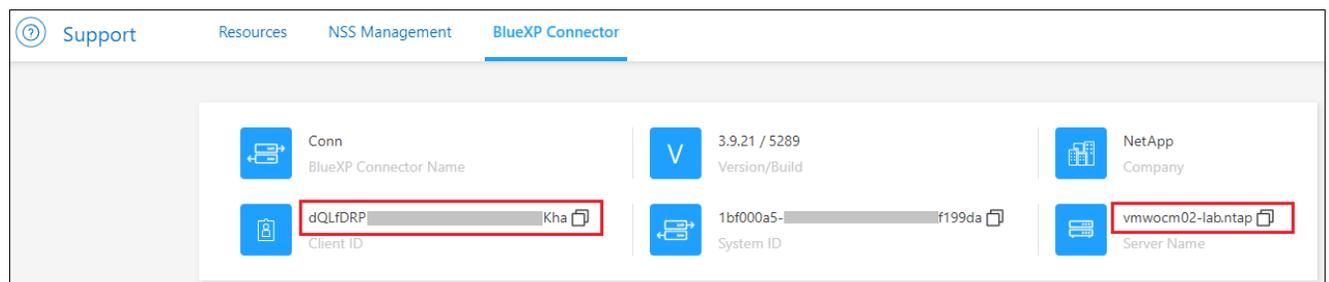
### 前提条件步骤

按照以下步骤获取添加扫描程序节点所需的NetApp帐户ID、连接器客户端ID、连接器服务器名称和用户访问令牌。

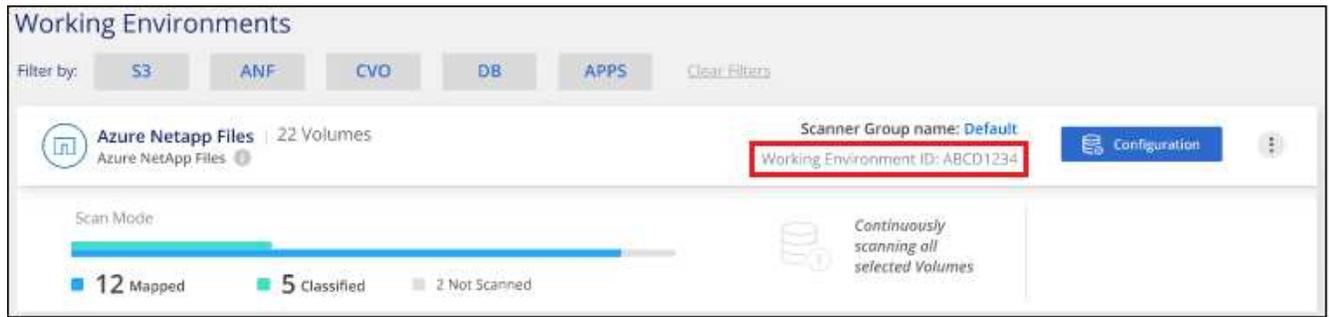
1. 从BlueXP菜单栏中、单击\*帐户>管理帐户\*。



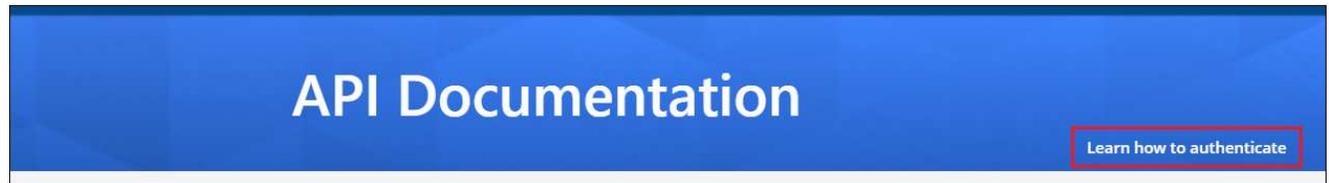
2. 复制\_Account ID\_。
3. 在BlueXP菜单栏中、单击\*帮助>支持> BlueXP连接器\*。



4. 复制connector *Client ID*\_和\_*Server Name*。
5. 如果您计划使用扫描程序组、请从"BlueXP分类配置"选项卡中、复制您计划添加到扫描程序组的每个工作环境的工作环境ID。



6. 转至 "API文档开发人员中心" 然后单击\*了解如何进行身份验证\*。



7. 按照身份验证说明、在"username"和"password"参数中使用帐户管理员的用户名和密码。

8. 然后、从响应中复制\_access token\_。

#### 步骤

1. 在BlueXP分类管理器节点上、运行脚本"add\_saner\_node.sh"。例如、此命令将添加2个扫描程序节点：

```
sudo ./add_scanner_node.sh -a <account_id> -c <client_id> -m <cm_host> -h
<ds_manager_ip> -n <node_private_ip_1,node_private_ip_2> -t <user_token>
```

#### 变量值：

- *account\_id* = NetApp 帐户 ID
  - *\_client\_id* = 连接器客户端ID (将后缀"clients"添加到在前提条件步骤中复制的客户端ID)
  - *cm\_host* = 连接器系统的IP地址或主机名
  - *ds\_manager\_IP* = BlueXP分类管理器节点系统的专用IP地址
  - *NODE\_PRIVIGE\_IP* = BlueXP分类扫描程序节点系统的IP地址(多个扫描程序节点IP以逗号分隔)
  - *user\_token* = JWT用户访问令牌
2. 在add\_scanner\_node脚本完成之前、会显示一个对话框、其中显示了扫描程序节点所需的安装命令。复制命令(例如：`sudo ./node_install.sh -m 10.11.12.13 -t ABCDEF1s35212 -u red95467j`)并将其保存在文本文件中。
3. 在 \* 每个 \* 扫描程序节点主机上：
- a. 将数据感知安装程序文件(\* `datasENSE-installer-cp.tar.gz*<version>`)复制到主机(使用`scp`或其他方法)。
  - b. 解压缩安装程序文件。
  - c. 粘贴并执行步骤2中复制的命令。
  - d. 如果要将扫描程序节点添加到"扫描程序组"中、请将参数\*`-r <scanner_group_name>`\*添加到命令中。否则、扫描程序节点将添加到"默认"组。

在所有扫描程序节点上完成安装且这些节点已加入管理器节点后、"add\_scanner\_node.sh"脚本也会完成。安装可能需要10到20分钟。

4. 如果将任何扫描程序节点添加到扫描程序组中、请返回到管理器节点并执行以下2项任务：

- a. 打开文件"/opt/NetApp/config/custom\_configuration/working\_扫描程序\_group\_config.yml"并输入要扫描特定工作环境的扫描程序组的映射。您需要为每个数据源提供\_Working Environment ID\_。例如、以下条目会将2个工作环境添加到"欧洲"扫描程序组、并将2个添加到"美国"扫描程序组：

```
scanner_groups:  
  europe:  
    working_environments:  
      - "working_environment_id1"  
      - "working_environment_id2"  
  united_states:  
    working_environments:  
      - "working_environment_id3"  
      - "working_environment_id4"
```

未添加到列表中的任何工作环境均由"default"组进行扫描-您必须在"default"组中至少有一个管理器或扫描程序节点。

- b. 运行以下脚本、将此映射信息注册到所有扫描程序节点：

```
/opt/netapp/Datasense/tools/update_we_scanner_group_from_config_file.sh
```

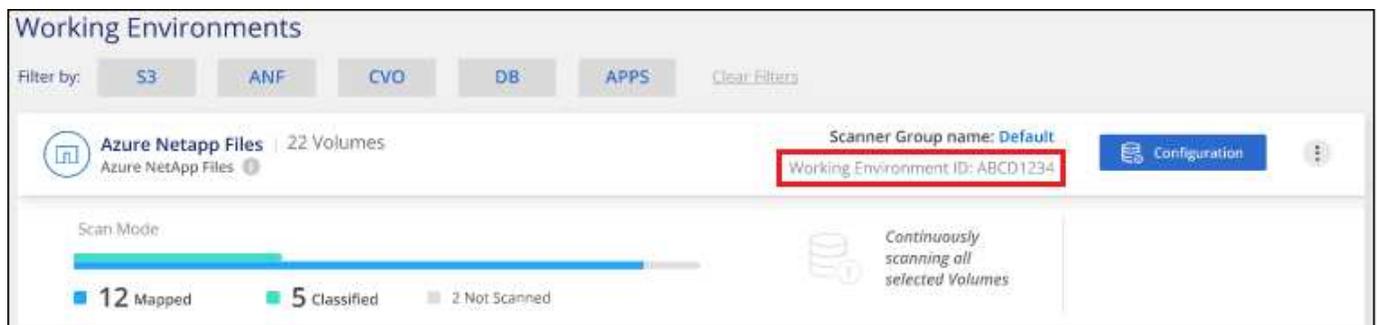
## 结果

BlueXP分类可通过管理器和扫描程序节点进行设置、以扫描所有数据源。

## 下一步行动

如果尚未扫描、您可以从配置页面中选择要扫描的数据源。如果创建了扫描程序组、则每个数据源都会由相应组中的扫描程序节点进行扫描。

您可以在配置页面中查看每个工作环境的扫描程序组名称。



The screenshot displays the 'Working Environments' configuration interface. At the top, there are filter buttons for 'S3', 'ANF', 'CVO', 'DB', and 'APPS', along with a 'Clear Filters' option. The main content area shows 'Azure Netapp Files' with '22 Volumes'. A 'Scanner Group name: Default' is selected, and the 'Working Environment ID: ABCD1234' is highlighted with a red box. A 'Configuration' button is visible. Below, a 'Scan Mode' progress bar shows 12 Mapped, 5 Classified, and 2 Not Scanned volumes. A note indicates 'Continuously scanning all selected Volumes'.

您还可以在配置页面底部查看所有扫描程序组的列表以及组中每个扫描程序节点的IP地址和状态。

Scanner Groups

Search

Scanner Group: Default Scanner nodes

2 Scanner nodes

Scanner node host name	IP	Last active time	Status	Error
ip-172-████████.us-west-2.compute	172-████████	23/09/2022 14:32	Active	
ip-172-████████.us-west-2.compute	172-████████	23/09/2022 14:32	Active	

Scanner Group: United\_States Scanner nodes

2 Scanner nodes

Scanner node host name	IP	Last active time	Status	Error
ip-172-████████.us-west-2.compute	172-████████	23/09/2022 14:32	Active	
ip-172-████████.us-west-2.compute	172-████████	23/09/2022 14:32	Active	

Scanner Group: Europe Scanner nodes

## 扫描数据弃用

### 扫描Amazon S3存储分段

BlueXP分类可以扫描Amazon S3存储分段、以确定驻留在S3对象存储中的个人数据和敏感数据。BlueXP分类可以扫描帐户中的任何存储分段、而不管该存储分段是否是NetApp解决方案创建的。

\*注\*此信息仅与BlueXP分类的旧版版本1.3及更早版本相关。

### 快速入门

按照以下步骤快速入门，或者向下滚动到其余部分以了解完整详细信息。

1

在云环境中设置 **S3** 要求

确保您的云环境可以满足BlueXP分类的要求、包括准备IAM角色以及设置从BlueXP分类到S3的连接。 [请参见完整列表。](#)

2

部署**BlueXP**分类实例

"部署BlueXP分类" 如果尚未部署实例。

**3**

在S3工作环境中激活BlueXP分类

选择 Amazon S3 工作环境，单击 \* 启用 \*，然后选择包含所需权限的 IAM 角色。

**4**

选择要扫描的分段

选择要扫描的存储分段、BlueXP分类将开始扫描它们。

查看 **S3** 前提条件

以下要求特定于扫描 S3 存储分段。

为BlueXP分类实例设置IAM角色

BlueXP分类需要有权限连接到您帐户中的S3存储分段并对其进行扫描。设置一个包含以下权限的 IAM 角色。在Amazon S3工作环境中启用BlueXP分类时、BlueXP会提示您选择IAM角色。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:Get*",
        "s3:List*",
        "s3:PutObject"
      ],
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": [
        "iam:GetPolicyVersion",
        "iam:GetPolicy",
        "iam:ListAttachedRolePolicies"
      ],
      "Resource": [
        "arn:aws:iam::*:policy/*",
        "arn:aws:iam::*:role/*"
      ]
    }
  ]
}
```

## 提供从BlueXP分类到Amazon S3的连接

BlueXP分类需要连接到Amazon S3。提供此连接的最佳方式是通过 VPC 端点连接到 S3 服务。有关说明，请参见 ["AWS 文档：创建网关端点"](#)。

创建VPC端点时、请务必选择与BlueXP分类实例对应的区域、VPC和路由表。您还必须修改安全组才能添加出站 HTTPS 规则、该规则允许通信到 S3 端点。否则、BlueXP分类无法连接到S3服务。

如果遇到任何问题，请参见 ["AWS支持知识中心：为什么我无法使用网关VPC端点连接到S3存储分段？"](#)

另一种方法是使用 NAT 网关提供连接。



您无法使用代理通过 Internet 访问 S3 。

## 部署BlueXP分类实例

["在BlueXP中部署BlueXP分类"](#) 如果尚未部署实例。

您需要使用部署在AWS中的Connector部署此实例、以便BlueXP自动发现此AWS帐户中的S3存储分段并将其显示在Amazon S3工作环境中。

\*注意：\*扫描S3存储分段时、当前不支持在内部位置部署BlueXP分类。

只要该实例具有Internet连接、BlueXP分类软件的升级就会自动进行。

## 在S3工作环境中激活BlueXP分类

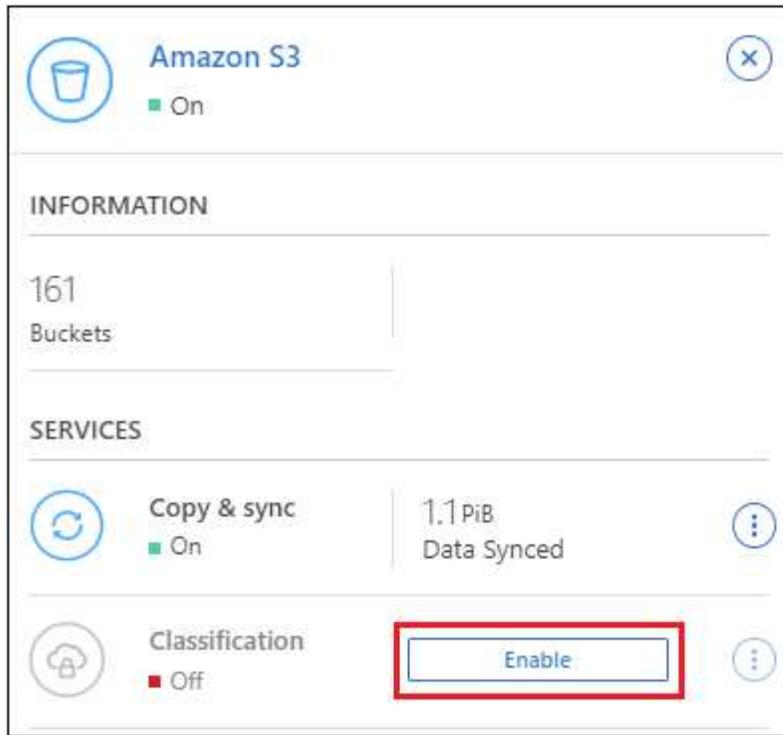
验证前提条件后、在Amazon S3上启用BlueXP分类。

### 步骤

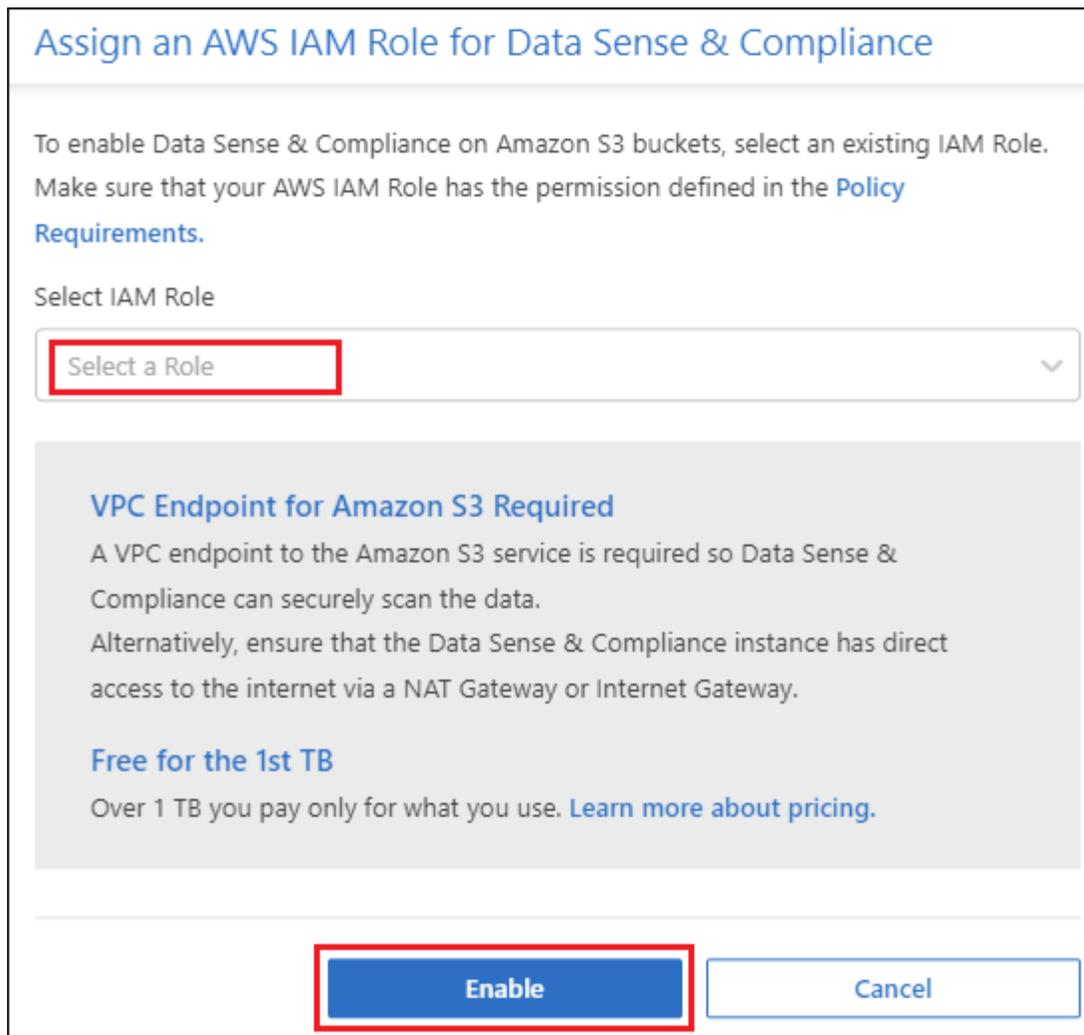
1. 从BlueXP左侧导航菜单中、单击\*存储>画布\*。
2. 选择 Amazon S3 工作环境。



3. 在右侧的服务窗格中、单击\*分类\*旁边的\*启用\*。



4. 出现提示时、将IAM角色分配给具有BlueXP分类实例 [所需权限](#)。



5. 单击 \* 启用 \* 。



您也可以通过单击在配置页面中为工作环境启用合规性扫描  按钮并选择\*激活BlueXP分类\*。

结果

BlueXP将IAM角色分配给实例。

在 **S3** 存储分段上启用和禁用合规性扫描

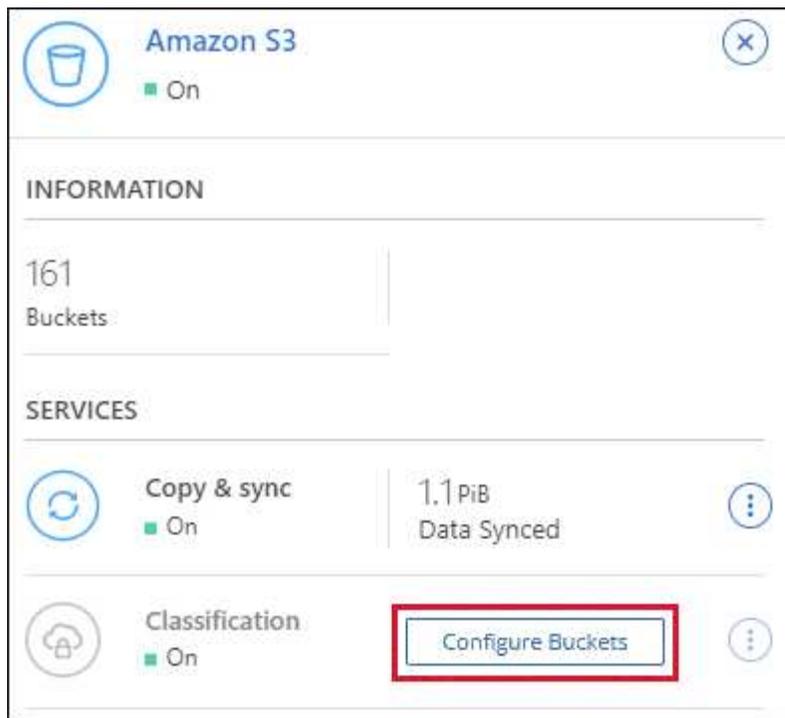
在BlueXP在Amazon S3上启用BlueXP分类后、下一步是配置要扫描的分段。

当BlueXP在包含要扫描的S3存储分段的AWS帐户中运行时、它会发现这些存储分段并将其显示在Amazon S3工作环境中。

BlueXP分类也可以 [扫描位于不同 AWS 帐户中的 S3 存储分段](#)。

步骤

1. 选择 Amazon S3 工作环境。
2. 在右侧的服务窗格中、单击\*配置分段\*。



3. 在存储分段上启用仅映射扫描或映射和分类扫描。

Amazon S3 Configuration			
15/28 Buckets in Scan Scope.			
Scan	Bucket Name	Status	Required Action
Off   Map   <b>Map &amp; Classify</b>	BucketName1	● Not Scanning	Add Credentials
Off   <b>Map</b>   Map & Classify	BucketName2	● Continuously Scanning	
<b>Off</b>   Map   Map & Classify	BucketName3	● Not Scanning	

收件人：	执行以下操作：
在存储分段上启用仅映射扫描	单击 * 映射 *
对存储分段启用完全扫描	单击 * 映射和分类 *
禁用对存储分段的扫描	单击 * 关闭 *

## 结果

BlueXP分类开始扫描您启用的S3分段。如果存在任何错误，它们将显示在状态列中，并显示修复此错误所需的操作。

## 从其他 **AWS** 帐户扫描存储分段

您可以通过从其他AWS帐户分配角色来扫描该帐户下的S3分段、以访问现有BlueXP分类实例。

## 步骤

1. 转到要扫描 S3 存储分段的目标 AWS 帐户，然后选择 \* 其他 AWS 帐户 \* 来创建 IAM 角色。

## Create role



### Select type of trusted entity

 <b>AWS service</b> EC2, Lambda and others	 <b>Another AWS account</b> Belonging to you or 3rd party	 <b>Web identity</b> Cognito or any OpenID provider	 <b>SAML 2.0 federation</b> Your corporate directory
--	---	---	--

Allows entities in other accounts to perform actions in this account. [Learn more](#)

### Specify accounts that can use this role

Account ID\*

- Options  Require external ID (Best practice when a third party will assume this role)  
 Require MFA 

请务必执行以下操作：

- 输入BlueXP分类实例所在帐户的ID。
- 将 \* 最大 CLI/API 会话持续时间 \* 从 1 小时更改为 12 小时，然后保存此更改。
- 附加BlueXP分类IAM策略。确保它具有所需的权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:Get*",
        "s3:List*",
        "s3:PutObject"
      ],
      "Resource": "*"
    }
  ]
}
```

2. 转到BlueXP分类实例所在的源AWS帐户、然后选择附加到该实例的IAM角色。
  - a. 将 \* 最大 CLI/API 会话持续时间 \* 从 1 小时更改为 12 小时，然后保存此更改。
  - b. 单击 \* 附加策略 \* ，然后单击 \* 创建策略 \* 。
  - c. 创建一个包含 "STS : AssumeRole" 操作的策略，并指定您在目标帐户中创建的角色 ARN 。

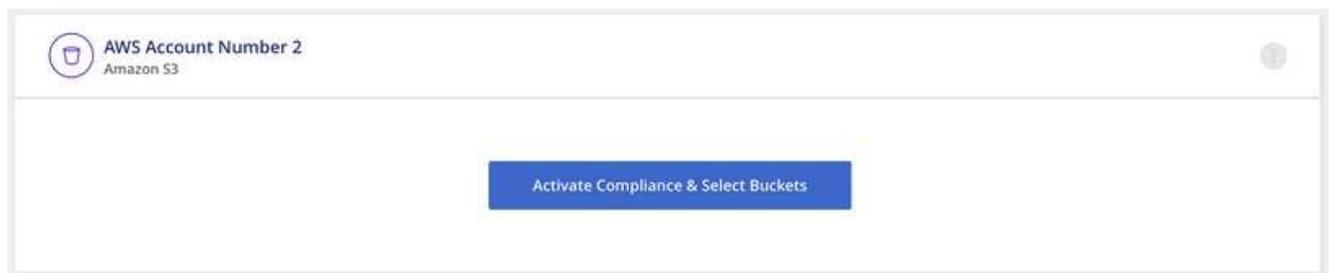
```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "sts:AssumeRole",
      "Resource": "arn:aws:iam::<ADDITIONAL-ACCOUNT-
ID>:role/<ADDITIONAL_ROLE_NAME>"
    },
    {
      "Effect": "Allow",
      "Action": [
        "iam:GetPolicyVersion",
        "iam:GetPolicy",
        "iam:ListAttachedRolePolicies"
      ],
      "Resource": [
        "arn:aws:iam::*:policy/*",
        "arn:aws:iam::*:role/*"
      ]
    }
  ]
}

```

BlueXP分类实例配置文件帐户现在可以访问其他AWS帐户。

3. 转到 \* Amazon S3 Configuration\* 页面，此时将显示新的 AWS 帐户。请注意、BlueXP分类可能需要几分钟时间来同步新帐户的工作环境并显示此信息。



4. 单击\*激活BlueXP分类并选择存储分段\*，然后选择要扫描的存储分段。

结果

BlueXP分类开始扫描您启用的新S3分段。

## 扫描OneDrive帐户

完成几个步骤、开始扫描具有BlueXP分类的用户OneDrive文件夹中的文件。

\*注\*此信息仅与BlueXP分类的旧版版本1.3及更早版本相关。

## 快速入门

按照以下步骤快速入门，或者向下滚动到其余部分以了解完整详细信息。

1

查看 **OneDrive** 前提条件

确保您拥有登录到 OneDrive 帐户所需的管理员凭据。

2

部署**BlueXP**分类实例

"部署BlueXP分类" 如果尚未部署实例。

3

添加 **OneDrive** 帐户

使用管理员用户凭据登录到要访问的 OneDrive 帐户，以便将其添加为新的工作环境。

4

添加用户并选择扫描类型

从 OneDrive 帐户中添加要扫描的用户列表，然后选择扫描类型。一次最多可以添加 100 个用户。

## 查看 **OneDrive** 要求

在启用BlueXP分类之前、请查看以下前提条件、以确保您的配置受支持。

- 您必须具有OneDrive for Business帐户的管理员登录凭据、该帐户可提供对用户文件的读取访问权限。
- 您需要列出要扫描其OneDrive文件夹的所有用户的电子邮件地址、并以行分隔。

## 部署**BlueXP**分类实例

如果尚未部署实例、请部署BlueXP分类。

BlueXP分类可以是 "[部署在云中](#)" 或 "[位于可访问 Internet 的内部位置](#)"。

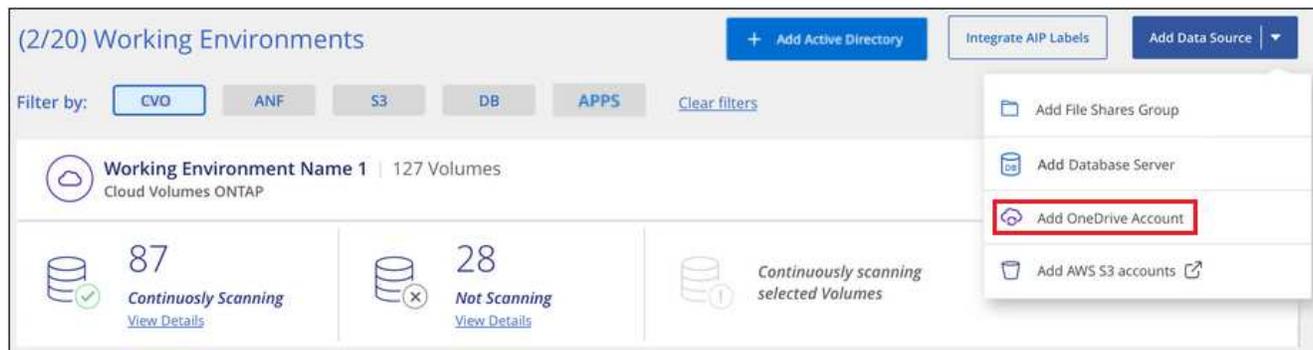
只要该实例具有Internet连接、BlueXP分类软件的升级就会自动进行。

## 正在添加 **OneDrive** 帐户

添加用户文件所在的 OneDrive 帐户。

## 步骤

1. 在工作环境配置页面中，单击 \* 添加数据源 \* > \* 添加 OneDrive 帐户 \* 。



2. 在添加 OneDrive 帐户对话框中，单击 \* 登录到 OneDrive\* 。
3. 在显示的Microsoft页面中、选择OneDrive帐户并输入所需的管理人员用户和密码、然后单击\*接受\*以允许BlueXP分类从此帐户读取数据。

OneDrive 帐户将添加到工作环境列表中。

将 **OneDrive** 用户添加到合规性扫描

您可以添加单个OneDrive用户或所有OneDrive用户、以便他们的文件将按BlueXP分类进行扫描。

步骤

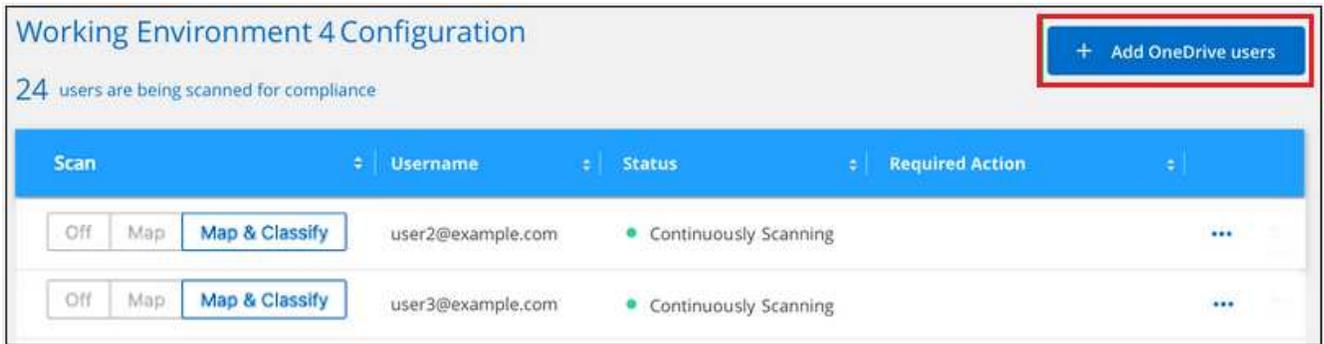
1. 在 *Configuration* 页面中，单击 OneDrive 帐户的 \* 配置 \* 按钮。



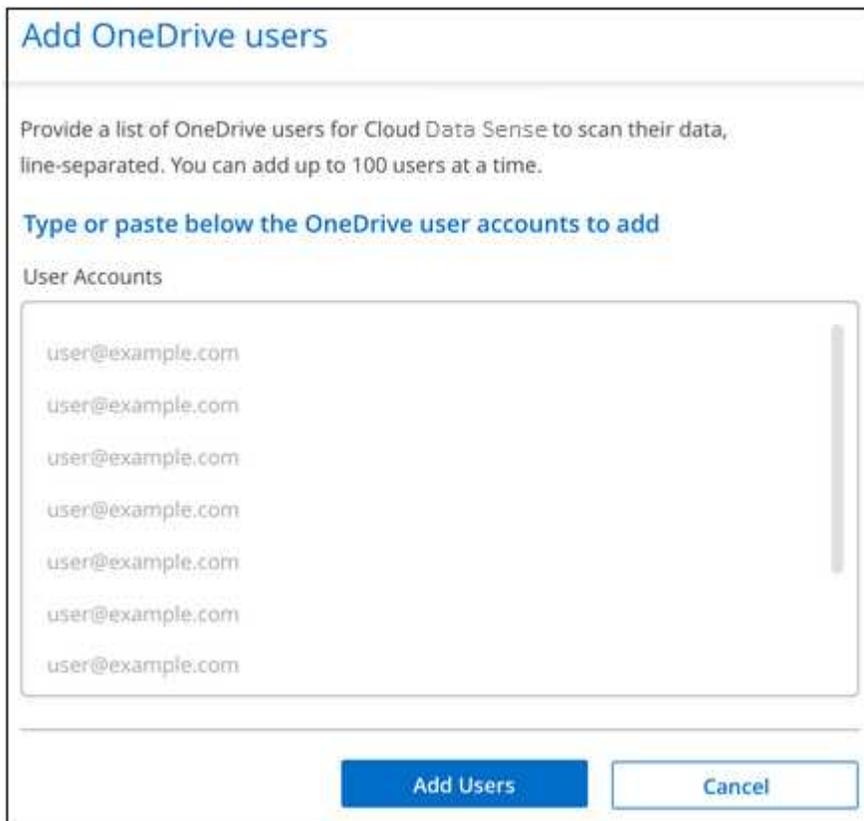
2. 如果这是首次为此 OneDrive 帐户添加用户，请单击 \* 添加您的首个 OneDrive 用户 \* 。



如果要从 OneDrive 帐户添加其他用户，请单击 \* 添加 OneDrive 用户 \* 。



3. 为要扫描其文件的用户添加电子邮件地址 - 每行一个电子邮件地址（每个会话最多 100 个） - 然后单击 \* 添加用户 \*。



确认对话框将显示已添加的用户数。

如果此对话框列出了任何无法添加的用户，请捕获此信息，以便解析问题描述。在某些情况下，您可以使用更正后的电子邮件地址重新添加用户。

4. 对用户文件启用仅映射扫描或映射和分类扫描。

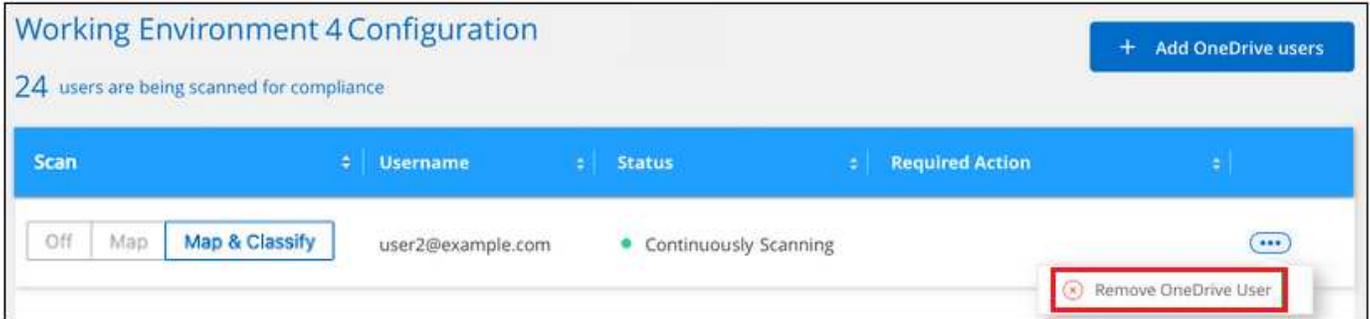
收件人：	执行以下操作：
对用户文件启用仅映射扫描	单击 * 映射 *
对用户文件启用完全扫描	单击 * 映射和分类 *
禁用对用户文件的扫描	单击 * 关闭 *

结果

BlueXP分类开始扫描您添加的用户的文件、结果将显示在信息板和其他位置。

### 从合规性扫描中删除 **OneDrive** 用户

如果用户离开公司或其电子邮件地址发生变化，您可以随时删除单个 OneDrive 用户的文件扫描功能。只需从配置页面中单击 \* 删除 OneDrive 用户 \* 即可。



### 扫描SharePoint帐户

完成几个步骤、开始扫描具有BlueXP分类的SharePoint Online和SharePoint内部部署帐户中的文件。

\*注\*此信息仅与BlueXP分类的旧版版本1.3及更早版本相关。

### 快速入门

按照以下步骤快速入门，或者向下滚动到其余部分以了解完整详细信息。

1

查看**SharePoint**前提条件

请确保您拥有可登录到SharePoint帐户的合格凭据、并且拥有要扫描的SharePoint站点的URL。

2

部署**BlueXP**分类实例

"部署BlueXP分类" 如果尚未部署实例。

3

登录到**SharePoint**帐户

使用符合条件的用户凭据登录到要访问的SharePoint帐户、以便将其添加为新的数据源/工作环境。

4

添加要扫描的**SharePoint**站点URL

在 SharePoint 帐户中添加要扫描的 SharePoint 站点 URL 列表，然后选择扫描类型。您一次最多可以添加100个URL、每个帐户总共最多可以添加1、000个站点。

## 查看SharePoint要求

查看以下前提条件、确保您已准备好在SharePoint帐户上激活BlueXP分类。

- 您必须具有可对所有SharePoint站点进行读取访问的SharePoint帐户的管理员用户登录凭据。
  - 对于SharePoint Online、您可以使用非管理员帐户、但该用户必须具有访问要扫描的所有SharePoint站点的权限。
- 对于内部部署的SharePoint、您还需要SharePoint Server的URL。
- 对于要扫描的所有数据，您需要一个以行分隔的 SharePoint 站点 URL 列表。

## 部署BlueXP分类实例

如果尚未部署实例、请部署BlueXP分类。

- 对于SharePoint Online、BlueXP分类可以是 "部署在云中"。
- 对于内部SharePoint、可以安装BlueXP分类 "位于可访问 Internet 的内部位置" 或 "位于无法访问Internet的内部位置"。

如果BlueXP分类安装在无法访问Internet的站点上、则BlueXP Connector也必须安装在无法访问Internet的同一站点上。 "了解更多信息。"。

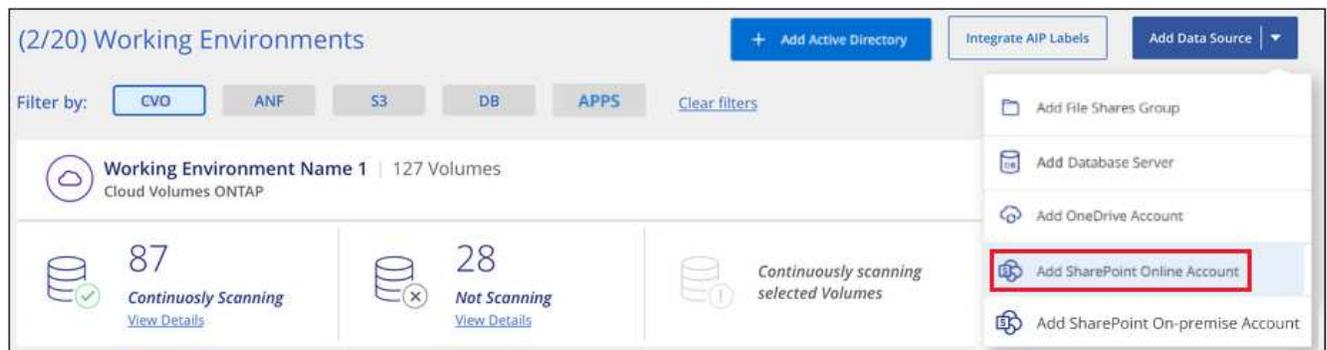
只要该实例具有Internet连接、BlueXP分类软件的升级就会自动进行。

## 添加SharePoint Online帐户

添加用户文件所在的SharePoint Online帐户。

### 步骤

1. 在工作环境配置页面中，单击 \* 添加数据源 \* > \* 添加 SharePoint Online 帐户 \* 。



2. 在添加 SharePoint Online 帐户对话框中，单击 \* 登录到 SharePoint\* 。
3. 在显示的Microsoft页中，选择SharePoint帐户并输入用户和密码(管理员用户或其他有权访问SharePoint站点的用户)，然后单击\*Accept (接受)\*以允许BlueXP分类从该帐户读取数据。

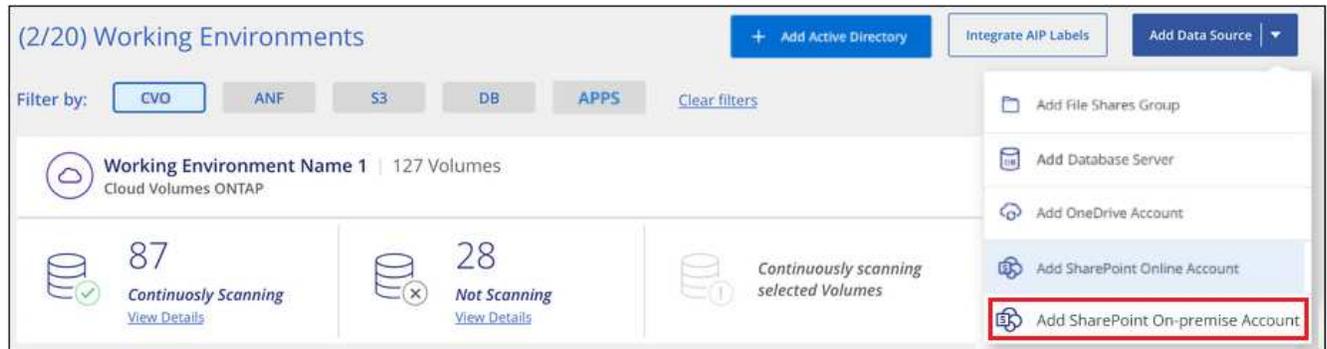
SharePoint Online帐户将添加到工作环境列表中。

## 添加SharePoint内部部署帐户

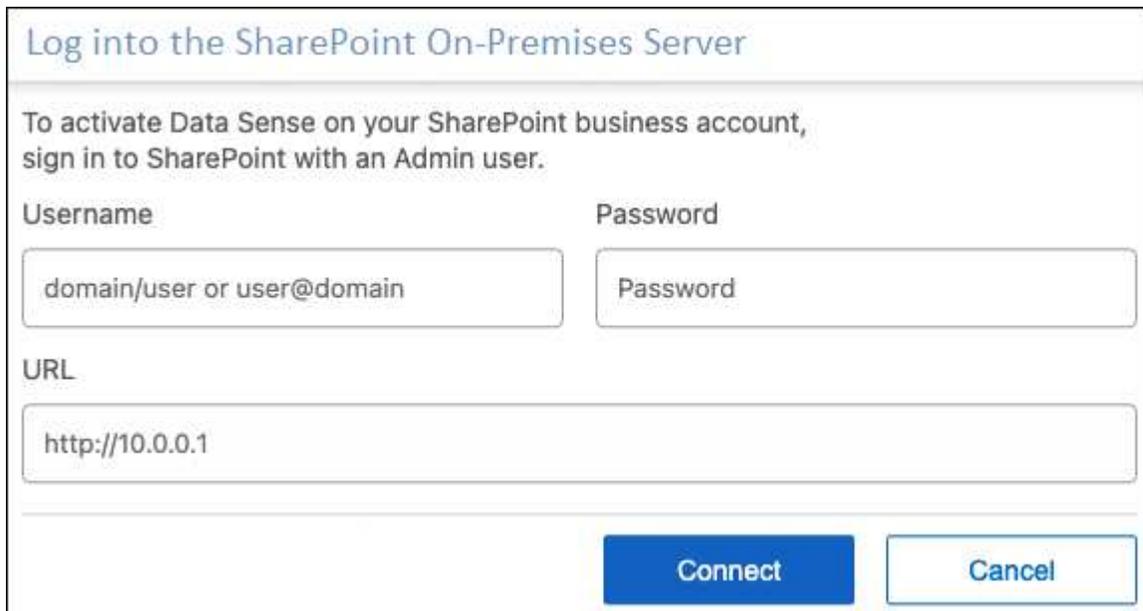
添加用户文件所在的SharePoint内部部署帐户。

### 步骤

1. 在工作环境配置页面中、单击\*添加数据源\*>\*添加SharePoint内部部署帐户\*。



2. 在登录到SharePoint内部部署服务器对话框中、输入以下信息：
  - 采用"域/用户"或"用户@域"格式的管理员用户、以及管理员密码
  - SharePoint Server的URL

The screenshot shows a dialog box titled 'Log into the SharePoint On-Premises Server'. The text inside says 'To activate Data Sense on your SharePoint business account, sign in to SharePoint with an Admin user.' There are three input fields: 'Username' with the placeholder 'domain/user or user@domain', 'Password' with the placeholder 'Password', and 'URL' with the placeholder 'http://10.0.0.1'. At the bottom, there are two buttons: 'Connect' (blue) and 'Cancel' (white with blue border).

3. 单击 \* 连接 \*。

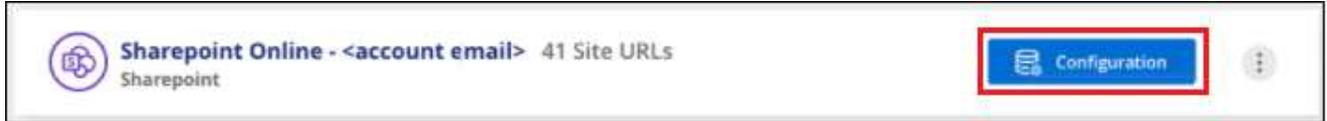
SharePoint内部部署帐户将添加到工作环境列表中。

## 将SharePoint站点添加到合规性扫描

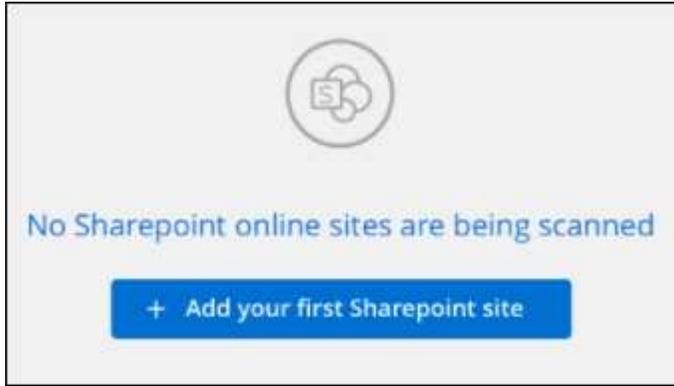
您可以在帐户中添加单个SharePoint网站或最多1,000个SharePoint网站、以便按BlueXP分类扫描关联的文件。无论您是添加SharePoint Online还是SharePoint内部部署网站、步骤都是相同的。

### 步骤

1. 在 *Configuration* 页面中，单击 SharePoint 帐户的 \* 配置 \* 按钮。



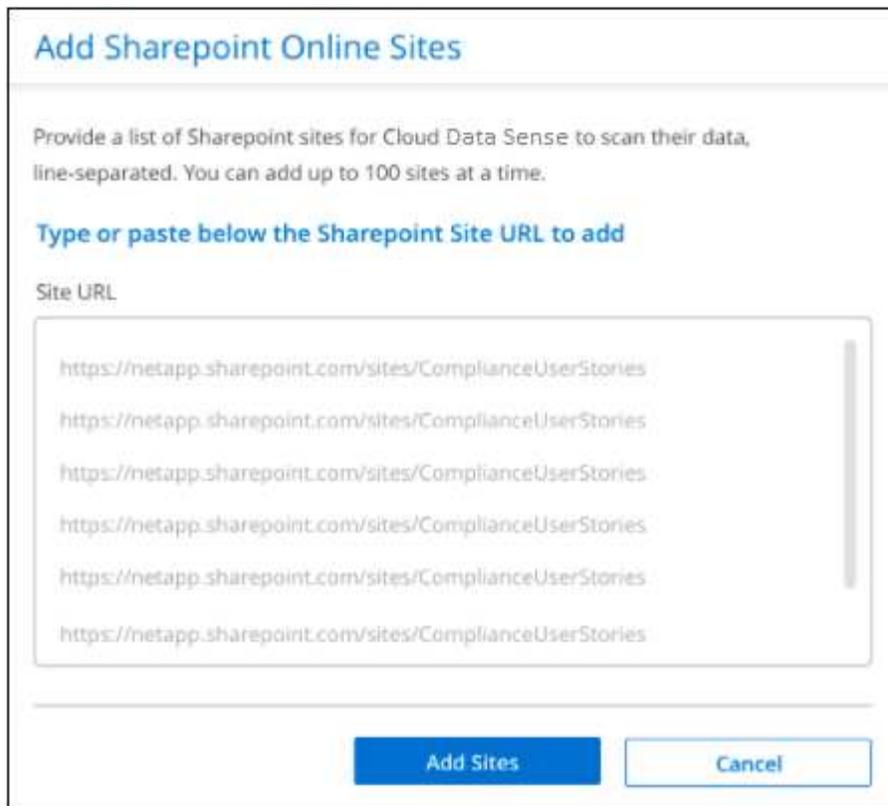
2. 如果这是首次为此 SharePoint 帐户添加站点，请单击 \* 添加您的第一个 SharePoint 站点 \*。



如果要从 SharePoint 帐户添加其他用户，请单击 \* 添加 SharePoint 站点 \*。



3. 为要扫描其文件的站点添加 URL - 每行一个 URL（每个会话最多 100 个 URL） - 然后单击 \* 添加站点 \*。



确认对话框将显示已添加的站点数量。

如果此对话框列出了任何无法添加的站点，请捕获此信息，以便您可以解析问题描述。在某些情况下，您可以使用更正后的 URL 重新添加此站点。

4. 如果您需要为此帐户添加100个以上的站点，只需再次单击\*添加SharePoint站点\*，直到您为此帐户添加了所有站点(每个帐户总共最多1000个站点)。
5. 对 SharePoint 站点中的文件启用仅映射扫描或映射和分类扫描。

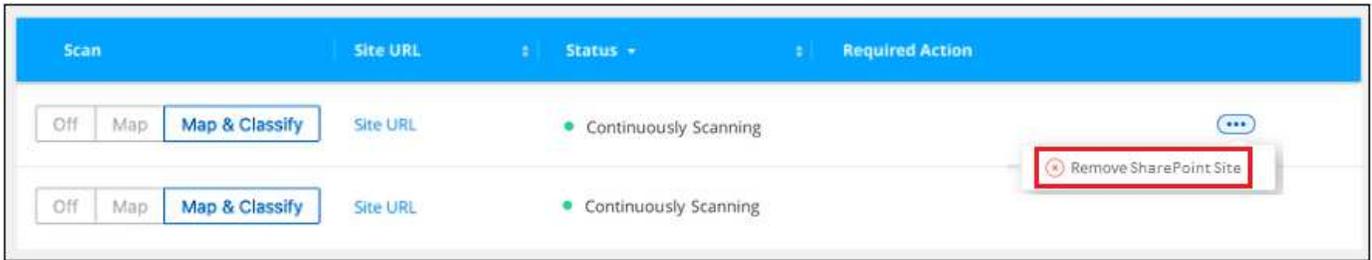
收件人：	执行以下操作：
对文件启用仅映射扫描	单击 * 映射 *
对文件启用完全扫描	单击 * 映射和分类 *
禁用文件扫描	单击 * 关闭 *

## 结果

BlueXP分类开始扫描您添加的SharePoint网站中的文件、结果将显示在信息板和其他位置。

## 从合规性扫描中删除SharePoint网站

如果您将来删除某个 SharePoint 站点，或者决定不扫描 SharePoint 站点中的文件，则可以随时删除各个 SharePoint 站点，使其无法扫描其文件。只需从配置页面中单击 \* 删除 SharePoint 站点 \* 即可。



请注意，您可以 ["从BlueXP分类中删除整个SharePoint帐户"](#) 如果您不想再扫描SharePoint帐户中的任何用户数据。

## 扫描Google Drive帐户

完成几个步骤、开始扫描具有BlueXP分类的Google Drive帐户中的用户文件。

\*注\*此信息仅与BlueXP分类的旧版版本1.3及更早版本相关。

### 快速入门

按照以下步骤快速入门，或者向下滚动到其余部分以了解完整详细信息。

1

查看**Google Drive**前提条件

确保您拥有登录到Google Drive帐户所需的**管理员凭据**。

2

部署**BlueXP**分类

["部署BlueXP分类"](#) 如果尚未部署实例。

3

登录到**Google Drive**帐户

使用**管理员用户凭据**登录到要访问的Google Drive帐户、以便将其添加为新的数据源。

4

选择扫描用户文件的类型

选择要对用户文件执行的扫描类型；映射或映射和分类。

### 查看**Google Drive**要求

查看以下前提条件、确保您已准备好在Google Drive帐户上启用BlueXP分类。

- 您必须拥有Google Drive帐户的**管理员登录凭据**、该帐户可提供对用户文件的读取访问权限

### 当前限制

Google Drive文件当前不支持以下BlueXP分类功能：

- 在"数据调查"页面中查看文件时、按钮栏中的操作未处于活动状态。您不能复制、移动、删除等任何文件。
- 无法在Google Drive中的文件中标识权限、因此调查页面中不会显示任何权限信息。

## 部署BlueXP分类

如果尚未部署实例、请部署BlueXP分类。

BlueXP分类可以是 "部署在云中" 或 "位于可访问 Internet 的内部位置"。

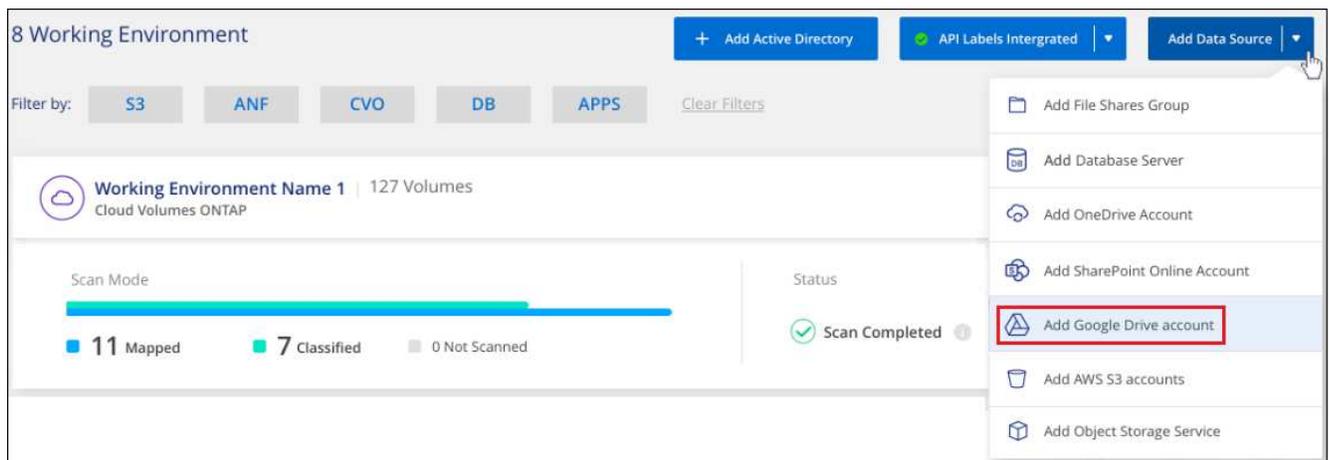
只要该实例具有Internet连接、BlueXP分类软件的升级就会自动进行。

## 添加Google Drive帐户

添加用户文件所在的Google Drive帐户。如果要扫描多个用户的文件、则需要对每个用户运行此步骤。

### 步骤

1. 在工作环境配置页面中、单击\*添加数据源\*>\*添加Google Drive帐户\*。



2. 在添加Google Drive帐户对话框中、单击\*登录到Google Drive\*。
3. 在显示的Google页面中、选择Google Drive帐户并输入所需的管理员用户和密码、然后单击\*接受\*以允许BlueXP分类从此帐户读取数据。

Google Drive帐户将添加到工作环境列表中。

### 选择扫描用户数据的类型

选择BlueXP分类将对用户数据执行的扫描类型。

### 步骤

1. 在\_Configuration\_页面中、单击Google Drive帐户的\*配置\*按钮。



2. 对Google Drive帐户中的文件启用仅映射扫描或映射和分类扫描。



收件人：	执行以下操作：
对文件启用仅映射扫描	单击 * 映射 *
对文件启用完全扫描	单击 * 映射和分类 *
禁用文件扫描	单击 * 关闭 *

## 结果

BlueXP分类开始扫描您添加的Google Drive帐户中的文件、结果将显示在信息板和其他位置。

## 从合规性扫描中删除Google Drive帐户

由于只有一个用户的Google Drive文件属于一个Google Drive帐户、因此、如果要停止扫描用户的Google Drive帐户中的文件、则应执行此操作 "[从BlueXP分类中删除Google Drive帐户](#)"。

## 扫描使用S3协议的对象存储

完成几个步骤、即可开始使用BlueXP分类直接扫描对象存储中的数据。BlueXP分类可以扫描使用Simple Storage Service (S3)协议的任何对象存储服务中的数据。其中包括NetApp StorageGRID、IBM云对象存储、Linode、B2云存储、Amazon S3等。

\*注\*此信息仅与BlueXP分类的旧版版本1.3及更早版本相关。

## 快速入门

按照以下步骤快速入门，或者向下滚动到其余部分以了解完整详细信息。

1

查看对象存储前提条件

您需要具有端点 URL 才能连接到对象存储服务。

您需要具有对象存储提供程序的访问密钥和机密密钥、以便BlueXP分类可以访问分段。

2

部署BlueXP分类实例

["部署BlueXP分类"](#) 如果尚未部署实例。

3

### 添加对象存储服务

将对象存储服务添加到BlueXP分类。

4

### 选择要扫描的分段

选择要扫描的存储分段、BlueXP分类将开始扫描它们。

查看对象存储要求

在启用BlueXP分类之前、请查看以下前提条件、以确保您的配置受支持。

- 您需要具有端点 URL 才能连接到对象存储服务。
- 您需要具有对象存储提供程序的访问密钥和机密密钥、以便BlueXP分类可以访问分段。

### 部署BlueXP分类实例

如果尚未部署实例、请部署BlueXP分类。

如果要从可通过 Internet 访问的 S3 对象存储扫描数据，则可以 ["在云中部署BlueXP分类"](#) 或 ["在可访问Internet的内部位置部署BlueXP分类"](#)。

如果要从安装在无法访问 Internet 的非公开站点中的 S3 对象存储扫描数据，则需要执行以下操作 ["在无法访问Internet的同一内部位置部署BlueXP分类"](#)。这还要求在同一内部位置部署BlueXP Connector。

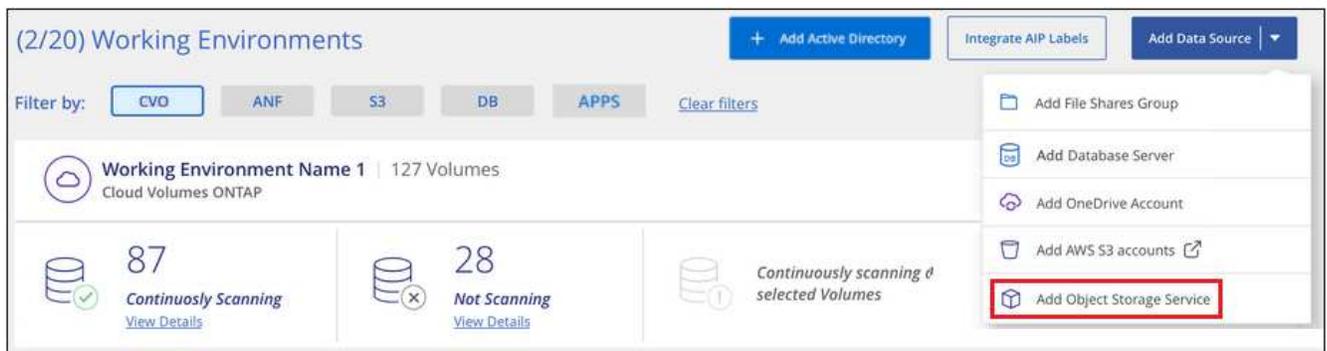
只要该实例具有Internet连接、BlueXP分类软件的升级就会自动进行。

### 将对象存储服务添加到BlueXP分类

添加对象存储服务。

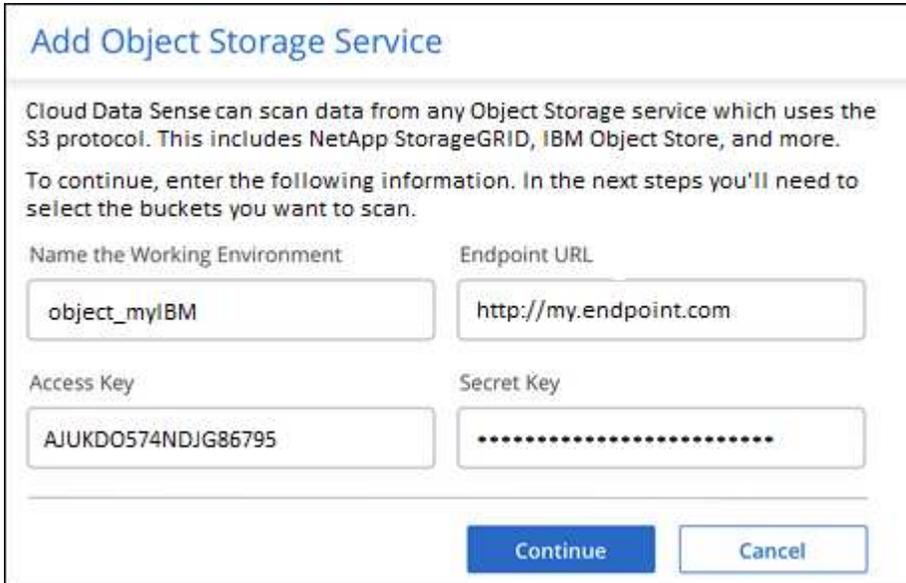
#### 步骤

1. 在工作环境配置页面中，单击 \* 添加数据源 \* > \* 添加对象存储服务 \*。



2. 在添加对象存储服务对话框中，输入对象存储服务的详细信息，然后单击 \* 继续 \*。
  - a. 输入要用于工作环境的名称。此名称应反映要连接到的对象存储服务的名称。
  - b. 输入端点 URL 以访问对象存储服务。

- c. 输入访问密钥和机密密钥、以便BlueXP分类可以访问对象存储中的分段。



结果

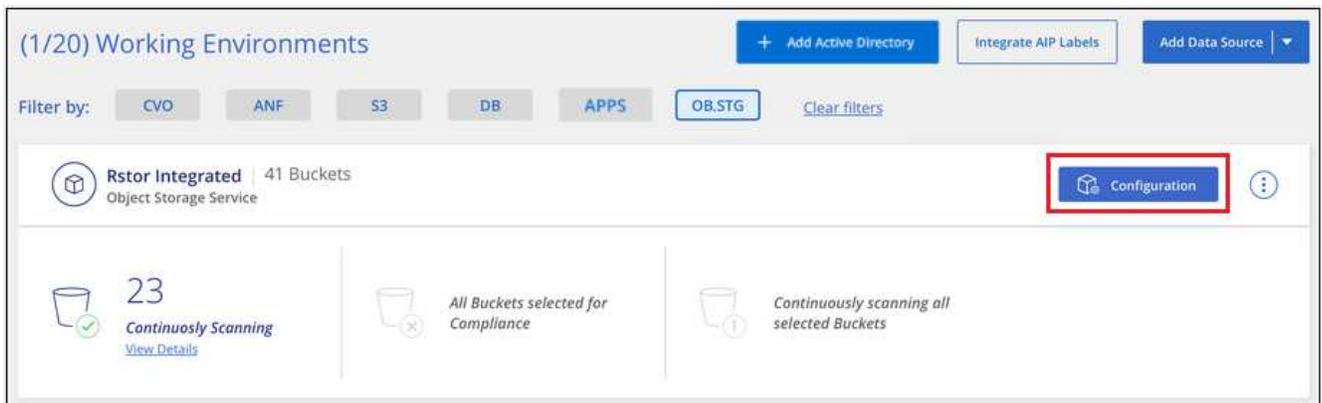
新的对象存储服务将添加到工作环境列表中。

启用和禁用对象存储分段上的合规性扫描

在对象存储服务上启用BlueXP分类后、下一步是配置要扫描的分段。BlueXP分类可发现这些分段、并在您创建的工作环境中显示它们。

步骤

1. 在配置页面中，单击对象存储服务工作环境中的 \* 配置 \* 。



2. 在存储分段上启用仅映射扫描或映射和分类扫描。

Rstor Integrated Configuration			
3/55 Buckets selected for Compliance scan			
Scan	Storage Repository (Bucket) ↓↑	Status ↓↑	Required Action ↓↑
<input type="checkbox"/> Off	Map	Map & Classify	logs-759995470648-us-east-1 ● Not Scanning
<input type="checkbox"/> Off	Map	Map & Classify	logs-759995470648-us-west-2 ● Not Scanning
<input type="checkbox"/> Off	<input checked="" type="checkbox"/> Map	Map & Classify	carstock ● Continuously Scanning

收件人：	执行以下操作：
在存储分段上启用仅映射扫描	单击 * 映射 *
对存储分段启用完全扫描	单击 * 映射和分类 *
禁用对存储分段的扫描	单击 * 关闭 *

## 结果

BlueXP分类开始扫描您启用的分段。如果存在任何错误，它们将显示在状态列中，并显示修复此错误所需的操作。

## 管理数据弃用

使用监管信息板查看有关数据的监管详细信息

控制与组织存储资源上的数据相关的成本。BlueXP分类可确定系统中陈旧数据、非业务数据、重复文件和非常大的文件的数量、以便您可以决定是删除某些文件还是将其分层到成本较低的对象存储。

此外、如果您计划将数据从内部位置迁移到云、则可以在移动数据之前查看数据的大小以及任何数据是否包含敏感信息。

\*注\*此信息仅与BlueXP分类的旧版版本1.3及更早版本相关。

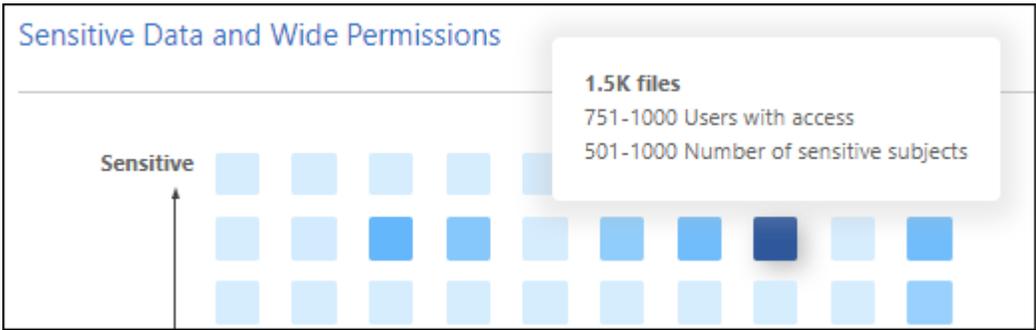
在监管信息板上按敏感度和广泛权限列出的数据

"监管"信息板上的"敏感数据和广泛权限"区域提供了包含敏感数据(包括敏感和敏感个人数据)且过于宽松的文件的 heatmap。这可以帮助您了解敏感数据可能存在哪些风险。



此适用场景BlueXP分类版本为1.3及更早版本。

根据有权访问X轴上文件的用户数量(从低到高)以及Y轴上文件中的敏感标识符数量(从低到高)、对文件进行评级。这些块表示与X轴和Y轴中的项匹配的文件数。颜色较浅的块很好；能够访问文件的用户较少、每个文件的敏感标识符较少。较暗的块是您可能需要调查的项目。例如、以下屏幕显示了深蓝色块的工具提示文本。它表明、您有1、500个文件、其中751-100个用户可以访问这些文件、并且每个文件有501到100个敏感标识符。



您可以单击感兴趣的块以在“调查”页面中查看受影响文件的筛选结果、以便进一步调查。

如果尚未将身份服务与BlueXP分类集成、则此面板中不会显示任何数据。 ["了解如何将Active Directory服务与BlueXP分类集成"](#)。



此面板支持CIFS共享、OneDrive和SharePoint数据源中的文件。目前不支持数据库、Google Drive、Amazon S3和通用对象存储。

信息板上显示**AIP**标签的分类区域

信息板上的 `_Classification_` 区域提供了扫描数据中最能识别的Azure信息保护(AIP)标签的列表。

如果您已订阅 Azure 信息保护（AIP），则可以通过将标签应用于内容来对文档和文件进行分类和保护。通过查看分配给文件的最常用 AIP 标签，您可以查看文件中最常使用的标签。

请参见 ["AIP 标签"](#) 有关详细信息 ...

## 组织您的私有数据

BlueXP分类提供了多种管理和组织私有数据的方式。这样可以更轻松地查看对您最重要的数据。

\*注\*此信息仅与BlueXP分类的旧版版本1.3及更早版本相关。2023年12月(v1.26.6)版本删除了使用Azure信息保护(AIP)标签集成数据的选项。

- 如果您已订阅 **"Azure 信息保护（AIP）"** 要对文件进行分类和保护、您可以使用BlueXP分类来管理这些AIP标签。
- 您可以将标记添加到要标记用于组织或某种类型的跟进的文件中。
- 您可以将一个BlueXP用户分配给一个特定文件或多个文件、以便此用户可以负责管理此文件。
- 使用 **"策略"** 功能，您可以创建自己的自定义搜索查询，以便通过单击一个按钮轻松查看结果。
- 当某些关键策略返回结果时、您可以向BlueXP用户或任何其他电子邮件地址发送电子邮件警报。



只有在选择对数据源执行完整分类扫描后，才可以使用本节所述的功能。已执行仅映射扫描的数据源不会显示文件级详细信息。

是否应使用标记或标签？

下面是BlueXP分类标记和Azure信息保护标记的比较。

Tags	标签
文件标记是BlueXP分类的一个集成部分。	要求您已订阅 Azure 信息保护（AIP）。
此标记仅保存在BlueXP分类数据库中、不会写入文件。它不会更改文件或文件访问或修改时间。	此标签是文件的一部分，当此标签发生更改时，此文件将发生更改。此更改还会更改文件的访问和修改时间。
一个文件可以包含多个标记。	一个文件可以有一个标签。
该标记可用于内部BlueXP分类操作、例如复制、移动、删除、运行策略、等等	可以读取文件的其他系统可以看到标签，可用于实现额外的自动化。
仅使用一个 API 调用来查看文件是否具有标记。	

### 使用 AIP 标签对数据进行分类

如果您订阅了BlueXP分类正在扫描的文件、则可以管理这些文件中的AIP标签 "[Azure 信息保护（AIP）](#)"。AIP允许您通过将标签应用于内容来对文档和文件进行分类和保护。通过BlueXP分类、您可以查看已分配给文件的标签、向文件添加标签以及在标签已存在时更改标签。

BlueXP分类支持以下文件类型中的AIP标签：.DOC、.docx、.PDF、.PPTX、.xls、.xlsx。



- 您当前无法更改大于 30 MB 的文件中的标签。对于OneDrive、SharePoint和Google Drive帐户、最大文件大小为4 MB。
- 如果文件的标签在AIP中已不存在、BlueXP分类会将其视为不带标签的文件。
- 如果您已在政府地区或无法访问Internet的内部位置(也称为非公开站点)部署BlueXP分类、则AIP标签功能不可用。

### 在工作空间中集成AIP标签

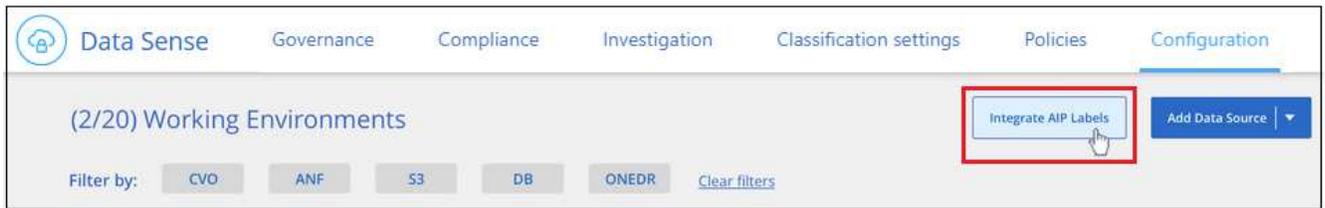
在管理AIP标签之前、您需要登录到现有Azure帐户、将AIP标签功能集成到BlueXP分类中。启用后、您可以管理所有文件中的 AIP 标签 "[数据源](#)" 在您的BlueXP工作空间中。

### 要求

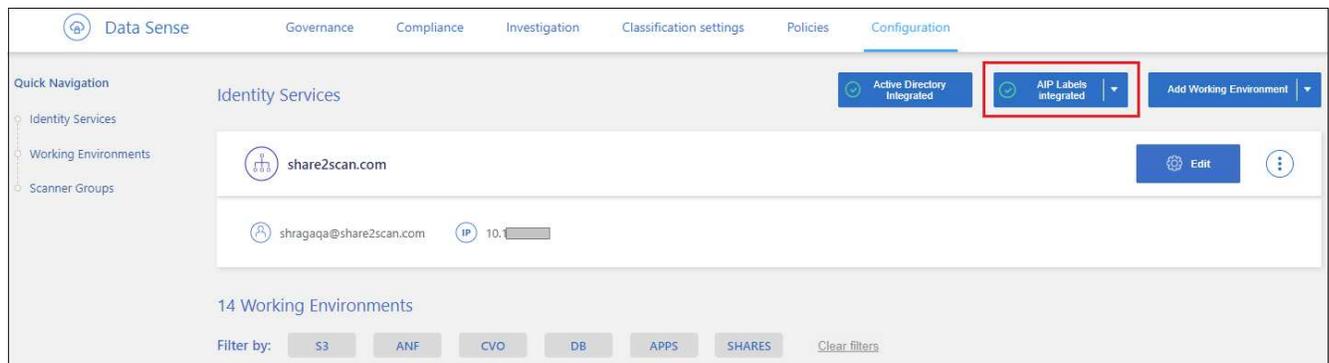
- 您必须拥有帐户和 Azure 信息保护许可证。
- 您必须具有 Azure 帐户的登录凭据。
- 如果您计划更改 Amazon S3 存储分段中文件的标签，请确保 IAM 角色中包含权限 `s3 : PutObject`。请参见 "[设置 IAM 角色](#)"。

### 步骤

1. 从"BlueXP分类配置"页面中，单击\*集成AIP标签\*。



2. 在集成 AIP 标签对话框中，单击 \* 登录到 Azure\* 。
3. 在显示的 Microsoft 页面中，选择帐户并输入所需的凭据。
4. 返回到"BlueXP分类"选项卡、您将看到消息"\_AIP标签已成功与帐户<account\_name>\_集成"。
5. 单击 \* 关闭 \* ，您将在页面顶部看到文本 \_AIP Labels Integrated \_ 。



## 结果

您可以从调查页面的结果窗格中查看和分配 AIP 标签。您还可以使用策略为文件分配 AIP 标签。

### 查看文件中的AIP标签

您可以查看分配给文件的当前 AIP 标签。

在数据调查结果窗格中，单击  文件以展开文件元数据详细信息。



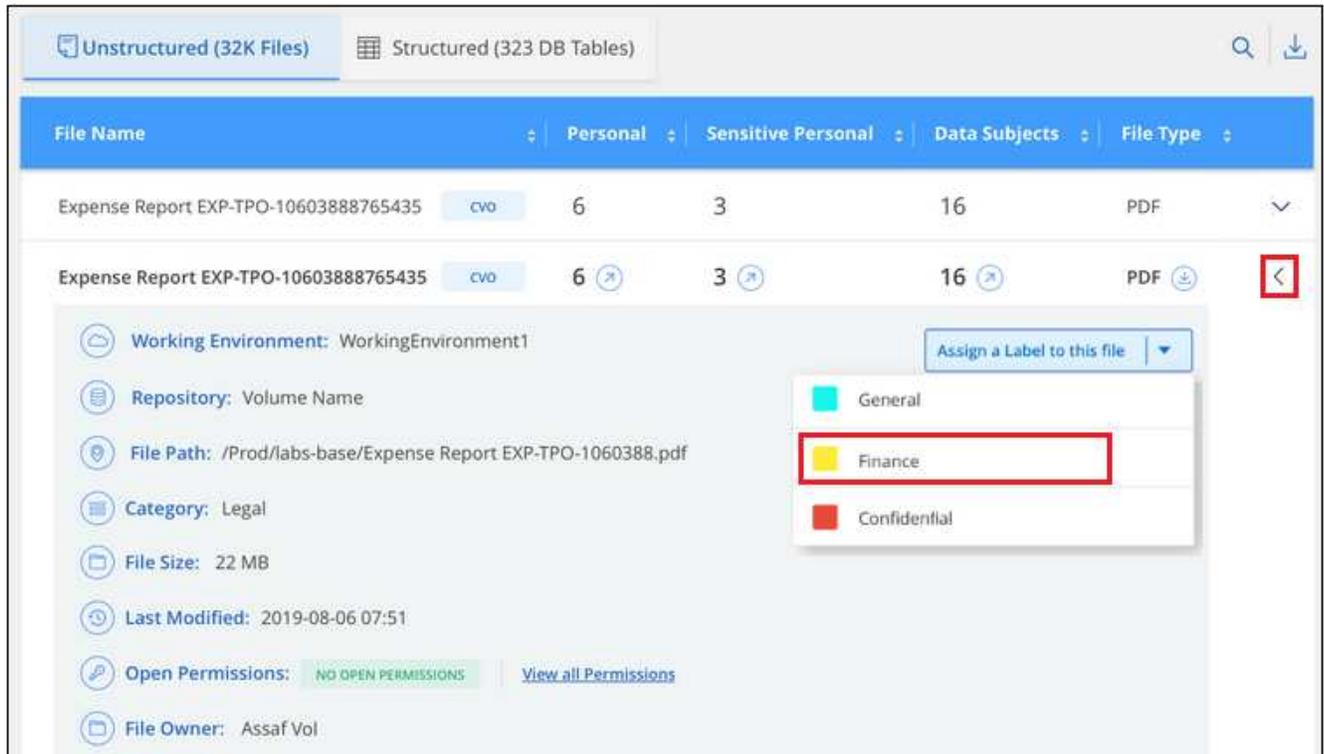
### 手动分配AIP标签

您可以使用BlueXP分类在文件中添加、更改和删除AIP标签。

按照以下步骤为单个文件分配 AIP 标签。

步骤

1. 在数据调查结果窗格中，单击  文件以展开文件元数据详细信息。



2. 单击 \* 为此文件分配标签 \* ，然后选择此标签。

此标签将显示在文件元数据中。

按照以下步骤为多个文件分配AIP标签。请注意、一次最多可以为20个文件分配AIP标签(UI中的一页)。

步骤

1. 在数据调查结果窗格中，选择要标记的一个或多个文件。



- 要选择单个文件，请选中每个文件 ( Volume\_1) 。
- 要选择当前页面上的所有文件，请选中标题行 ( File Name) 。

2. 从按钮栏中，单击 \* 标签 \* 并选择 AIP 标签：



AIP 标签将添加到所有选定文件的元数据中。

#### 删除AIP集成

如果您不再需要管理文件中的AIP标签、可以从BlueXP分类界面中删除AIP帐户。

请注意、不会对使用BlueXP分类添加的标签进行任何更改。文件中存在的标签将保持当前存在的状态。

#### 步骤

1. 从 *Configuration* 页面中，单击 \* 集成 AIP 标签 > 删除集成 \* 。



2. 从确认对话框中单击 \* 删除集成 \* 。

#### 应用标记以管理扫描的文件

您可以向要标记为某种类型的跟进的文件添加标记。例如，您可能已发现一些重复文件，并且希望删除其中一个文件，但需要检查应删除哪个文件。您可以向文件添加一个标记 "Check to delete"，以使您知道此文件需要进行一些研究并在未来执行某种类型的操作。

通过BlueXP分类、您可以查看分配给文件的标记、在文件中添加或删除标记以及更改名称或删除现有标记。

请注意，标记添加到文件中的方式与 AIP 标签是文件元数据的一部分不同。BlueXP用户只能使用BlueXP分类查看该标记、以便您查看是否需要删除文件或检查某个类型的跟进。

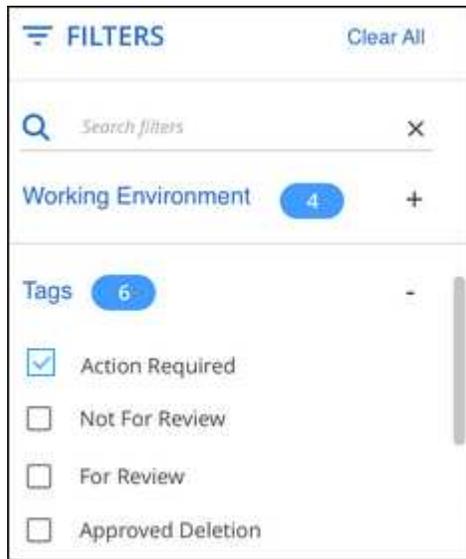


按BlueXP分类分配给文件的标记与可添加到资源(例如卷或虚拟机实例)的标记无关。BlueXP分类标记在文件级别应用。

#### 查看应用了某些标记的文件

您可以查看已分配特定标记的所有文件。

1. 单击BlueXP分类中的\*调查\*选项卡。
2. 在数据调查页面中，单击筛选器窗格中的 \* 标记 \*，然后选择所需的标记。



"调查结果"窗格将显示已分配这些标记的所有文件。

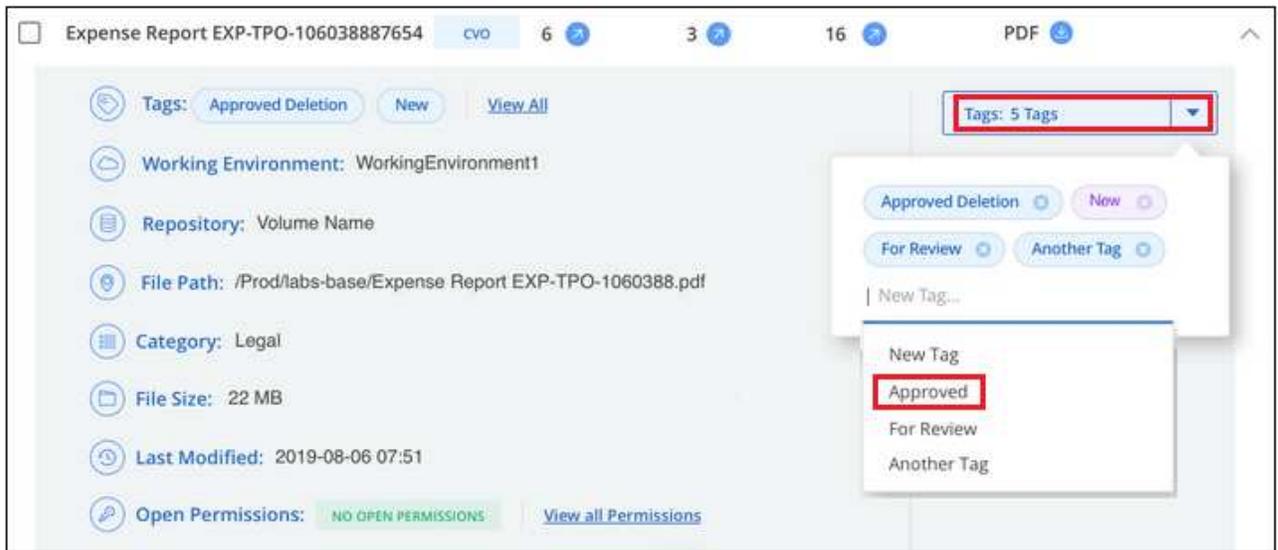
为文件分配标记

您可以向单个文件或一组文件添加标记。

向单个文件添加标记：

步骤

1. 在数据调查结果窗格中，单击  文件以展开文件元数据详细信息。
2. 单击 \* 标记 \* 字段，此时将显示当前已分配的标记。
3. 添加一个或多个标记：
  - 要分配现有标记，请单击 \* 新标记 ... \* 字段，然后开始键入标记的名称。出现要查找的标记时，请选择该标记并按 \* 输入 \*。
  - 要创建新标记并将其分配给文件，请单击 \* 新标记 ... \* 字段，输入新标记的名称，然后按 \* 输入 \*。



此标记将显示在文件元数据中。

向多个文件添加标记：

步骤

1. 在数据调查结果窗格中，选择要标记的一个或多个文件。



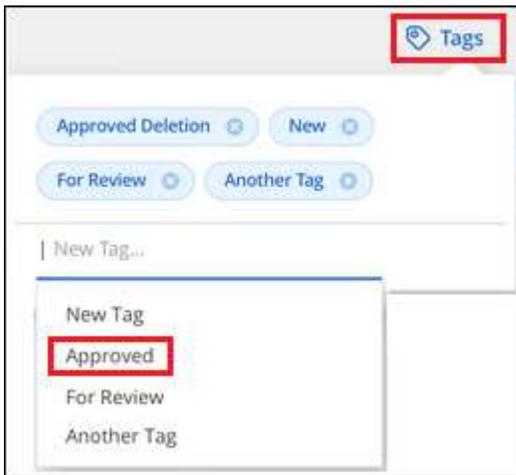
- 要选择单个文件，请选中每个文件（ Volume\_1）。
- 要选择当前页面上的所有文件，请选中标题行（ File Name）。
- 要选择所有页面上的所有文件，请选中标题行（ File Name），然后显示在弹出消息中 **All 20 Items on this page selected** [Select all Items in list \(63K Items\)](#) 下，单击 \* 选择列表中的所有项（xxx 项）\*。

一次最多可以对100、000个文件应用标记。

2. 在按钮栏中，单击 \* 标记 \*，此时将显示当前已分配的标记。

3. 添加一个或多个标记：

- 要分配现有标记，请单击 \* 新标记 ... \* 字段，然后开始键入标记的名称。出现要查找的标记时，请选择该标记并按 \* 输入 \*。
- 要创建新标记并将其分配给文件，请单击 \* 新标记 ... \* 字段，输入新标记的名称，然后按 \* 输入 \*。



4. 批准在确认对话框中添加标记，标记将添加到所有选定文件的元数据中。

从文件中删除标记

如果不再需要使用某个标记，可以将其删除。

只需单击现有标记的 \* x \* 即可。



如果选择了多个文件，则标记将从所有文件中删除。

分配用户以管理某些文件

您可以将一个BlueXP用户分配给一个特定文件或多个文件、以便此用户可以负责对该文件执行任何后续操作。此功能通常与功能结合使用，用于向文件添加自定义状态标记。

例如，您的文件可能包含某些个人数据，这些数据允许过多的用户进行读写访问（打开权限）。因此，您可以将状态标记 " 更改权限 " 并将此文件分配给用户 "Joan Smith"，以便用户确定如何修复问题描述。修复问题描述后，他们可以将状态标记更改为 " 已完成 "。

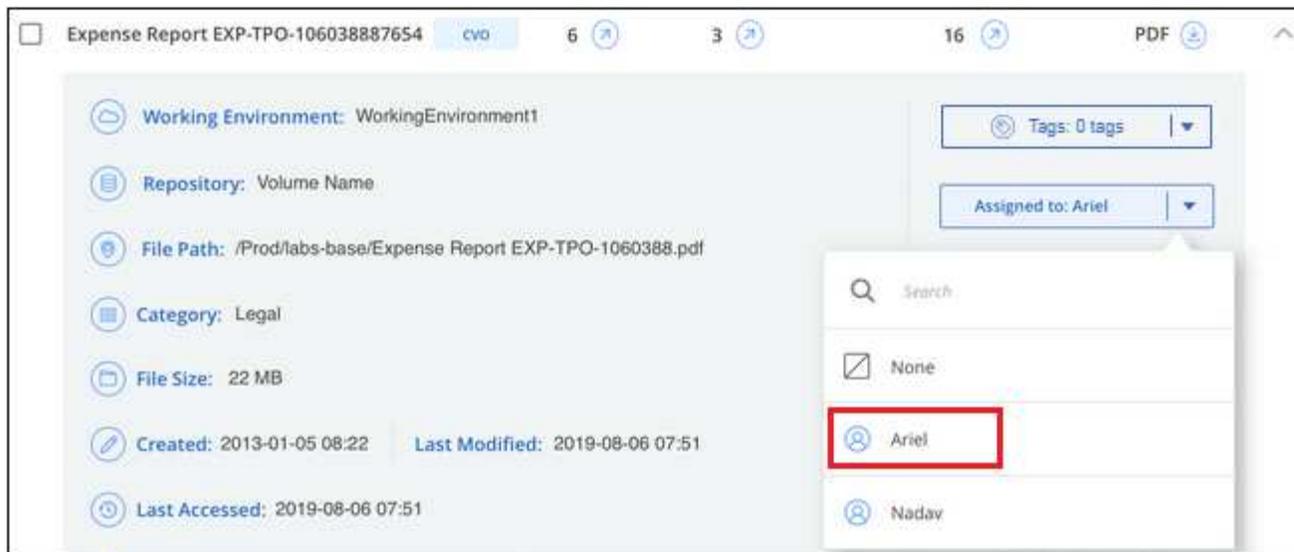
请注意、用户名不会作为文件元数据的一部分添加到文件中-使用BlueXP分类时、BlueXP用户只会看到该用户名。

通过 " 调查 " 页面中的新筛选器，您可以轻松查看 " 已分配给 " 字段中具有相同人员的所有文件。

按照以下步骤将用户分配给单个文件。

步骤

1. 在数据调查结果窗格中，单击  文件以展开文件元数据详细信息。
2. 单击 \* 已分配给 \* 字段并选择用户名。



用户名显示在文件元数据中。

按照以下步骤将用户分配给多个文件。请注意、一次最多可以为一个用户分配20个文件(UI中的一个页面)。

步骤

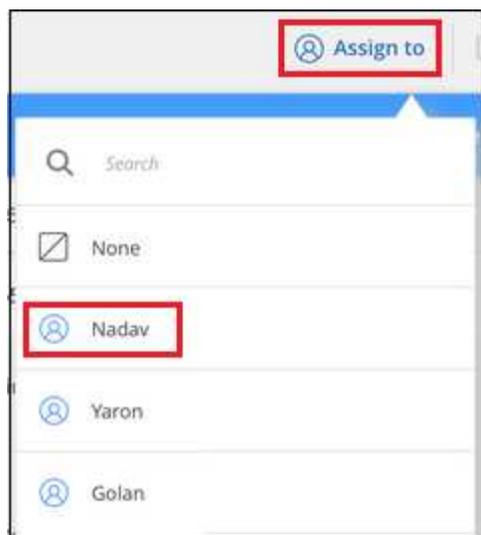
1. 在数据调查结果窗格中，选择要分配给用户的一个或多个文件。

The screenshot shows a file management interface with a table of files. The table has columns for File Name, Personal, Sensitive Personal, Data Subjects, and File Type. Two files are selected, indicated by red checkboxes. The interface also shows a toolbar with buttons for Tags, Assign to, Label, Copy, Move, and Delete.

<input type="checkbox"/>	File Name	Personal	Sensitive Personal	Data Subjects	File Type
<input checked="" type="checkbox"/>	Expense Report EXP-TPO-106038887654 cvo	6	3	16	PDF
<input checked="" type="checkbox"/>	Expense Report EXP-TPO-106038887654 cvo	6	3	6	PDF
<input type="checkbox"/>	Expense Report EXP-TPO-106038887654 cvo	6	3	6	PDF
<input type="checkbox"/>	Expense Report EXP-TPO-106038887654 cvo	6	3	6	PDF

- 要选择单个文件，请选中每个文件 ( Volume\_1)。
- 要选择当前页面上的所有文件，请选中标题行 ( File Name)。

2. 从按钮栏中，单击 \* 分配给 \* 并选择用户名：



用户将添加到所有选定文件的元数据中。

## 管理私有数据

BlueXP分类提供了多种管理私有数据的方式。某些功能使您可以更轻松地为迁移数据做准备，而其他功能则允许您对数据进行更改。

\*注\*此信息仅与BlueXP分类的旧版版本1.3及更早版本相关。

- 如果要为某些数据创建副本并将其移动到其他 NFS 位置，可以将文件复制到目标 NFS 共享。
- 您可以将 ONTAP 卷克隆到新卷，同时在新克隆的卷中仅包含源卷中的选定文件。如果您要迁移数据，并且希望从原始卷中排除某些文件，则此选项非常有用。
- 您可以将文件从源存储库复制并同步到特定目标位置的目录。如果您要将数据从一个源系统迁移到另一个源系统、而源文件上仍有一些最终活动、则此功能非常有用。
- 您可以将BlueXP分类正在扫描的源文件移动到任何NFS共享。
- 您可以删除那些看起来不安全或风险过高，无法留在存储系统中的文件，或者您已确定为重复的文件。



- 只有在选择对数据源执行完整分类扫描后，才可以使用本节所述的功能。已执行仅映射扫描的数据源不会显示文件级详细信息。
- 目前、Google Drive帐户中的数据无法使用上述任何功能。

## 复制源文件

您可以复制BlueXP分类正在扫描的任何源文件。根据您尝试完成的任务，复制操作有三种类型：

- \* 将文件 \* 从相同或不同的卷或数据源复制到目标 NFS 共享。

如果要为某些数据创建副本并将其移动到其他 NFS 位置，则此功能非常有用。

- \* 将 ONTAP 卷 \* 克隆到同一聚合中的新卷，但在新克隆的卷中仅包含源卷中的选定文件。

如果您要迁移数据，并且希望从原始卷中排除某些文件，则此选项非常有用。此操作将使用 "NetApp FlexClone" 用于快速复制卷，然后删除您 \* 未 \* 选择的文件的功能。

- 将文件 \* 从单个源存储库（ONTAP 卷，S3 存储分段，NFS 共享等）复制并同步到特定目标（目标）位置的目录。

这对于将数据从一个源系统迁移到另一个源系统的情况非常有用。初始副本之后、服务将根据您设置的计划同步所有更改的数据。此操作将使用 "NetApp BlueXP复制和同步" 用于将数据从源复制并同步到目标的功能。

#### 将源文件复制到NFS共享

您可以将BlueXP分类正在扫描的源文件复制到任何NFS共享。NFS共享不需要与BlueXP分类集成、您只需知道将以格式复制所有选定文件的NFS共享的名称即可 <host\_name>:/<share\_path>。



您不能复制数据库中的文件。

#### 要求

- 要复制文件，您必须具有帐户管理员或工作空间管理员角色。
- 复制文件要求目标NFS共享允许从BlueXP分类实例进行访问。
- 一次可以复制1到100、000个文件。

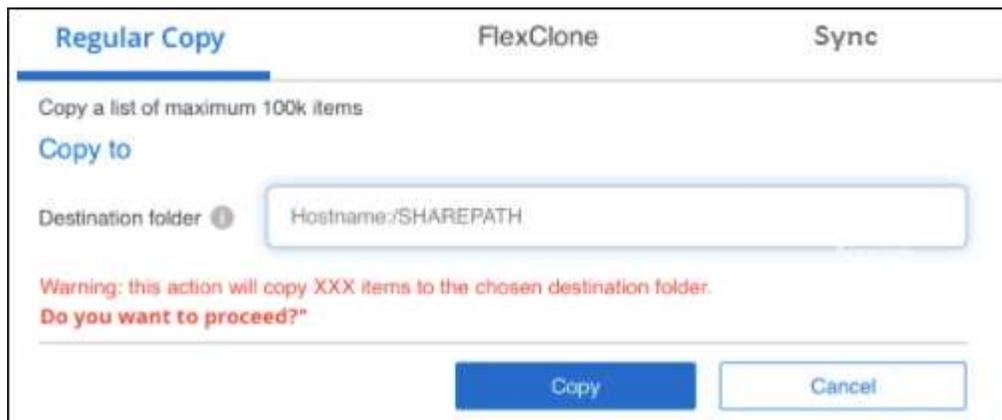
#### 步骤

1. 在数据调查结果窗格中，选择要复制的一个或多个文件，然后单击 \* 复制 \* 。



- 要选择单个文件，请选中每个文件（ Volume\_1）。
- 要选择当前页面上的所有文件，请选中标题行（ File Name）。
- 要选择所有页面上的所有文件，请选中标题行（ File Name），然后显示在弹出消息中 **All 20 Items on this page selected Select all Items in list (63K Items)** 下，单击 \* 选择列表中的所有项（xxx 项）\*。

2. 在 *Copy Files* 对话框中，选择 \* 常规复制 \* 选项卡。



3. 输入要复制所有选定文件的 NFS 共享的名称，格式为 ``<host_name> : /<share_path>``，然后单击 \* 复制 \*。

此时将显示一个对话框，其中包含复制操作的状态。

您可以在中查看复制操作的进度 "操作状态窗格"。

请注意，您也可以在查看文件的元数据详细信息时复制单个文件。只需单击 \* 复制文件 \*。



#### 将卷数据克隆到新卷

您可以使用 NetApp\_FlexCLONE 功能克隆 BlueXP 分类正在扫描的现有 ONTAP 卷。这样，您就可以快速复制卷，同时仅包含选定的文件。如果要迁移数据，并且要从原始卷中排除某些文件，或者要创建卷的副本以进行测试，则此功能非常有用。

新卷将在与源卷相同的聚合中创建。在启动此任务之前，请确保聚合中有足够的空间容纳此新卷。如有必要，请联系您的存储管理员。

- 注意： \* 无法克隆 FlexGroup 卷，因为 FlexClone 不支持这些卷。

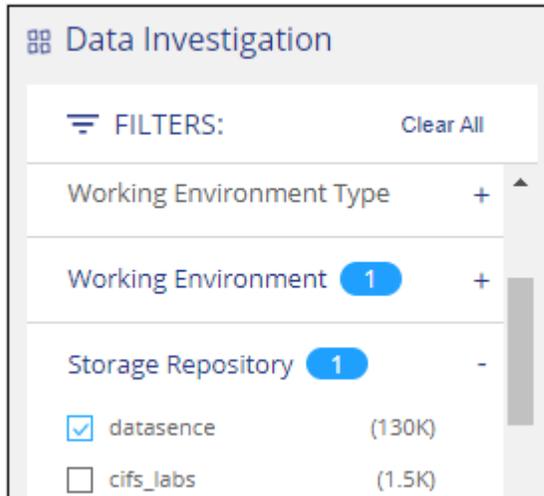
#### 要求

- 要复制文件，您必须具有帐户管理员或工作空间管理员角色。
- 您必须至少选择 20 个文件。
- 所有选定文件都必须来自同一个卷，并且该卷必须处于联机状态。

- 卷必须来自 Cloud Volumes ONTAP 或内部 ONTAP 系统。当前不支持任何其他数据源。
- 集群上必须安装 FlexClone 许可证。默认情况下，此许可证会安装在 Cloud Volumes ONTAP 系统上。

#### 步骤

1. 在 "数据调查" 窗格中，选择一个 \* 工作环境 \* 和一个 \* 存储库 \* 来创建筛选器，以确保所有文件都来自同一个 ONTAP 卷。



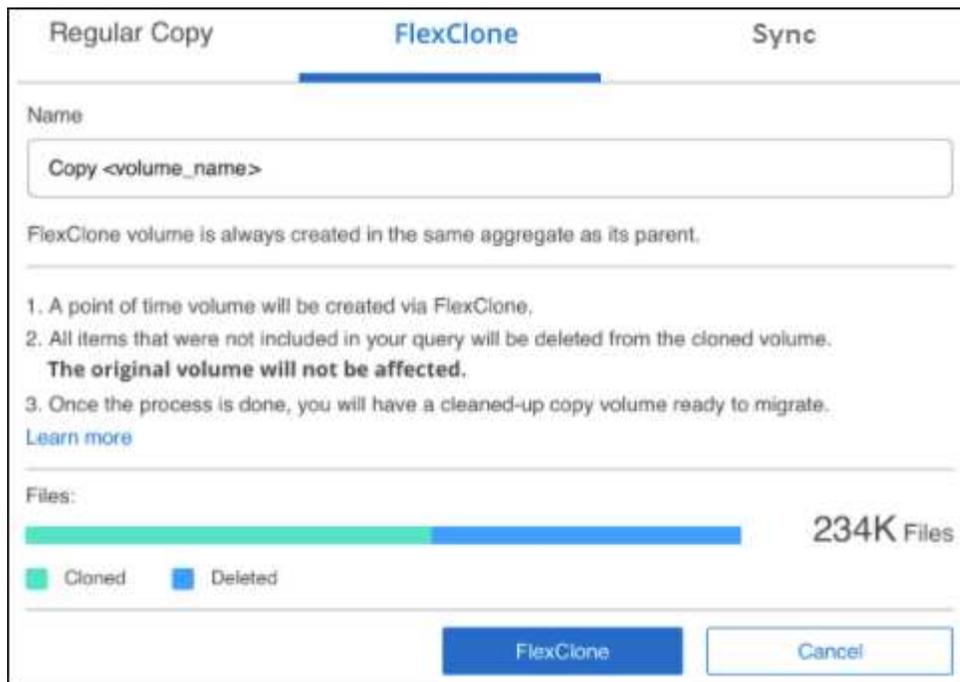
应用任何其他筛选器，以便您只能看到要克隆到新卷的文件。

2. 在调查结果窗格中，选择要克隆的文件，然后单击 \* 复制 \*。



- 要选择单个文件，请选中每个文件 ( Volume\_1)。
- 要选择当前页面上的所有文件，请选中标题行 ( File Name)。
- 要选择所有页面上的所有文件，请选中标题行 ( File Name)，然后显示在弹出消息中 **All 20 Items on this page selected** [Select all Items in list \(63K Items\)](#) 下，单击 \* 选择列表中的所有项 (xxx 项) \*。

3. 在 *Copy Files* 对话框中，选择 \* FlexClone \* 选项卡。此页面显示要从卷克隆的文件总数 (您选择的文件)，以及未从克隆的卷中包含 / 删除的文件数 (您未选择的文件)。



4. 输入新卷的名称，然后单击 \* FlexClone \* 。

此时将显示一个对话框，其中包含克隆操作的状态。

## 结果

新的克隆卷将在与源卷相同的聚合中创建。

您可以在中查看克隆操作的进度 ["操作状态窗格"](#)。

如果在为源卷所在的工作环境启用BlueXP分类时最初选择了\*映射所有卷\*或\*映射并分类所有卷\*、则BlueXP分类将自动扫描新克隆的卷。如果最初未使用上述任一选项，则如果要扫描此新卷，则需要执行以下操作 ["手动对卷启用扫描"](#)。

将源文件复制并同步到目标系统

您可以将BlueXP分类正在扫描的源文件从任何受支持的非结构化数据源复制到特定目标位置的目录 (["BlueXP复制和同步支持的目标位置"](#))。初始复制后，文件中更改的任何数据将根据您配置的计划进行同步。

这对于将数据从一个源系统迁移到另一个源系统的情况非常有用。此操作将使用 ["NetApp BlueXP复制和同步"](#) 用于将数据从源复制并同步到目标的功能。



您不能复制和同步数据库， OneDrive 帐户或 SharePoint 帐户中的文件。

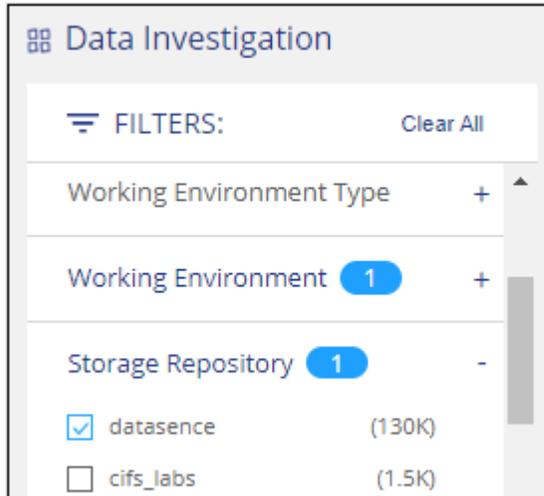
## 要求

- 要复制和同步文件，您必须具有帐户管理员或工作空间管理员角色。
- 您必须至少选择20个文件。
- 所有选定文件都必须来自同一源存储库（ ONTAP 卷， S3 存储分段， NFS 或 CIFS 共享等）。
- 您需要激活BlueXP复制和同步服务、并至少配置一个数据代理、用于在源系统和目标系统之间传输文件。从开始查看BlueXP副本和同步要求 ["快速启动问题描述"](#)。

请注意、BlueXP复制和同步服务会为您的同步关系单独收取服务费用、如果您在云中部署数据代理、则会产生资源费用。

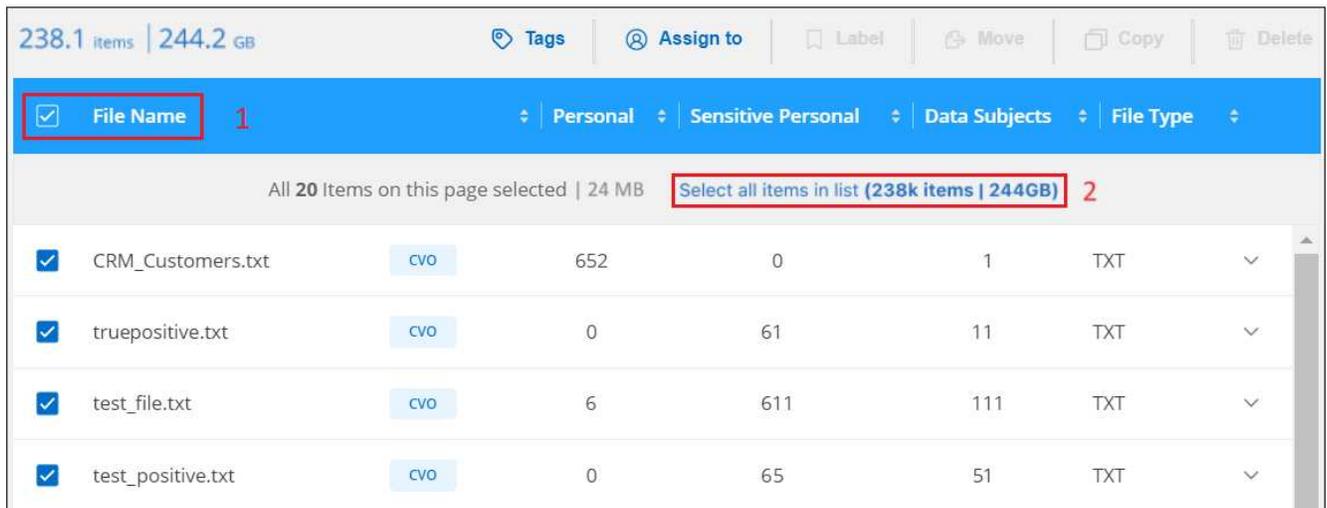
## 步骤

1. 在数据调查窗格中，选择一个 \* 工作环境 \* 和一个 \* 存储库 \* 来创建筛选器，以确保所有文件都来自同一个存储库。

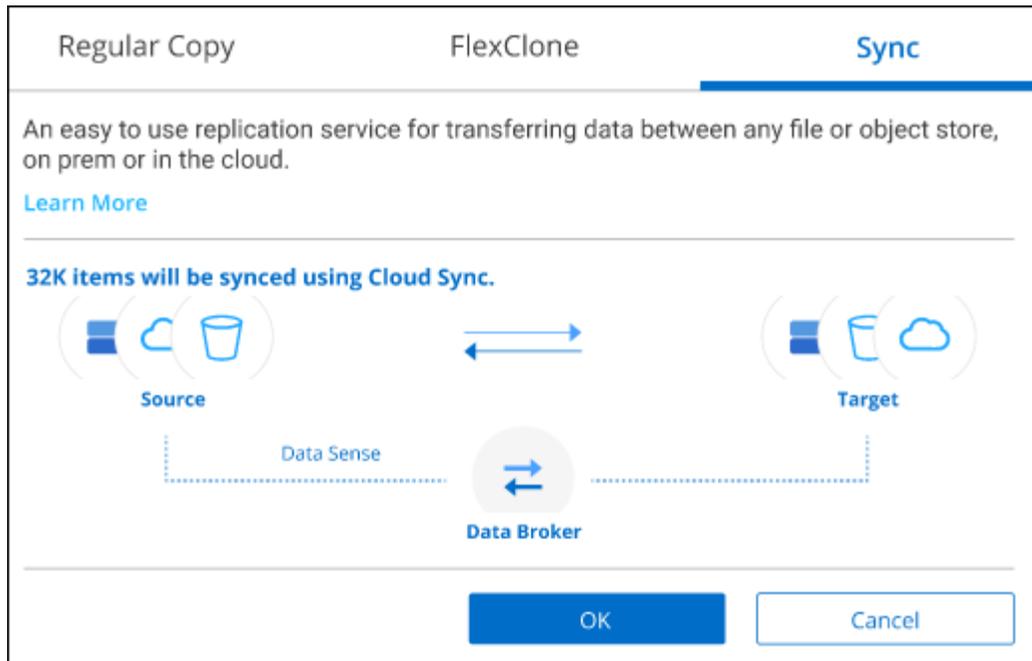


应用任何其他筛选器，以便您仅看到要复制并同步到目标系统的文件。

2. 在调查结果窗格中，选中标题行 ( File Name)，然后显示在弹出消息中  
All 20 Items on this page selected Select all Items in list (63K Items) 单击 \* 选择列表中的所有项 (xxx 项) \*，然后单击 \* 复制 \*。



3. 在 Copy Files 对话框中，选择 \* 同步 \* 选项卡。



4. 如果确实要将选定文件同步到目标位置，请单击 \* 确定 \*。

BlueXP副本和同步UI将在BlueXP中打开。

系统将提示您定义同步关系。源系统会根据您在BlueXP分类中选择的存储库和文件进行预先填充。

5. 您需要选择目标系统，然后选择（或创建）计划使用的数据代理。从开始查看BlueXP副本和同步要求 "[快速启动问题描述](#)"。

## 结果

这些文件将复制到目标系统，并根据您定义的计划进行同步。如果选择一次性同步，则文件只会复制和同步一次。如果选择定期同步，则会根据计划同步文件。请注意，如果源系统添加的新文件与您使用筛选器创建的查询匹配，这些 *new* 文件将复制到目标并在将来进行同步。

请注意、从BlueXP分类调用BlueXP时、某些常见的BlueXP复制和同步操作会被禁用：

- 不能使用 \* 删除源上的文件 \* 或 \* 删除目标上的文件 \* 按钮。
- 已禁用运行报告。

## 将源文件移动到NFS共享

您可以将BlueXP分类正在扫描的源文件移动到任何NFS共享。NFS共享不需要与BlueXP分类集成。

或者、您也可以将痕迹文件保留在移动文件的位置。痕迹文件可帮助用户了解文件从其原始位置移动的原因。对于每个移动的文件、系统会在源位置`<filename>-breadcrumb-<date>.txt`创建一个痕迹文件。您可以在对话框中添加要添加到痕迹文件的文本、以指示文件移动的位置以及文件移动的用户。

请注意、移动文件时、源文件中的子目录结构会在目标共享上重新创建、以便于了解文件的移动位置。如果目标位置存在同名文件、则不会移动该文件。



您无法移动驻留在数据库中的文件。

## 要求

- 要移动文件，您必须具有帐户管理员或工作空间管理员角色。
- 源文件可以位于以下数据源中：内部ONTAP、Cloud Volumes ONTAP、Azure NetApp Files、文件共享和SharePoint Online。
- 一次最多可以移动1、500万个文件。
- 仅移动小于或等于50 MB的文件。
- 目标NFS共享必须允许从BlueXP分类实例IP地址进行访问。

## 步骤

1. 在数据调查结果窗格中，选择要移动的一个或多个文件。



<input type="checkbox"/>	File Name	Personal	Sensitive Personal	Data Subjects	File Type	
<input checked="" type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	16	PDF
<input checked="" type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	6	PDF
<input type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	6	PDF
<input type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	6	PDF

- 要选择单个文件，请选中每个文件（ Volume\_1）。
- 要选择当前页面上的所有文件，请选中标题行（ File Name）。
- 要选择所有页面上的所有文件，请选中标题行（ File Name），然后显示在弹出消息中 **All 20 items on this page selected** [Select all items in list \(63K items\)](#) 下，单击 \* 选择列表中的所有项（xxx 项）\*。

2. 在按钮栏中，单击 \* 移动 \*。

### Move Files (63)

The files will be moved to the destination folder you provide and will no longer be available at their current location.

Moving files is supported only to destination folders in NFS Shares. Any NFS Share is supported, no matter where it is hosted, as long as the share's export policy allows access from the data connector instance IP address.

---

The status of this action will appear in the Action Status.

---

#### Enter the NFS destination folder path to continue

---

Leave breadcrumb

A breadcrumb file helps your users understand why a file was moved from its original location. For each moved file, the system creates a breadcrumb file in the source location named `<filename>-breadcrumb-<date>.txt`.

 Max length should be maximum 400 characters

---

3. 在 `_move Files_` 对话框中、输入要移动所有选定文件的NFS共享的名称、格式为 `'<host_name> : /<share_path>'`。
4. 如果要保留痕迹文件、请选中 `_leave痕迹_` 框。您可以在对话框中输入文本、以指示文件的移动位置、移动文件的用户以及任何其他信息、例如文件移动的原因。
5. 单击 `*移动文件*`。

请注意，在查看文件的元数据详细信息时，您也可以移动单个文件。只需单击 `* 移动文件 *`。



### 删除源文件

您可以永久删除看似不安全或风险太大，无法留在存储系统中的源文件，或者已确定为重复的源文件。此操作为永久操作，不会撤消或还原。

您可以从"调查"窗格或手动删除文件 "自动使用策略"。



您不能删除数据库中的文件。支持所有其他数据源。

删除文件需要以下权限：

- 对于NFS数据—需要使用写入权限定义导出策略。
- 对于CIFS数据—CIFS凭据需要具有写入权限。
- 对于 S3 数据 - IAM 角色必须包括以下权限： `s3 : DeleteObject`

### 手动删除源文件

#### 要求

- 要删除文件，您必须具有帐户管理员或工作空间管理员角色。
- 一次最多可以删除 100,000 个文件。

#### 步骤

1. 在数据调查结果窗格中，选择要删除的一个或多个文件。



- 要选择单个文件，请选中每个文件（ Volume\_1）。
- 要选择当前页面上的所有文件，请选中标题行（ File Name）。
- 要选择所有页面上的所有文件，请选中标题行（ File Name），然后显示在弹出消息中 **All 20 Items on this page selected** [Select all Items in list \(63K Items\)](#) 下，单击 \* 选择列表中的所有项（xxx 项）\*。

2. 在按钮栏中，单击 \* 删除 \*。
3. 由于删除操作是永久性的，因此您必须在后续的 *Delete File* 对话框中键入 "\* 永久删除 \*"，然后单击 \* 删除文件 \*。

您可以在中查看删除操作的进度 "操作状态窗格"。

请注意，您也可以查看文件的元数据详细信息时删除单个文件。只需单击 \* 删除文件 \*。



## 将个人数据标识符添加到BlueXP分类扫描中

BlueXP分类为您提供了许多方法来添加BlueXP分类将在未来扫描中标识的自定义"个人数据"列表、从而让您全面了解组织的\_all\_文件中可能敏感的数据所在的位置。

\*注\*此信息仅与BlueXP分类的旧版版本1.3及更早版本相关。

- 您可以根据要扫描的数据库中的特定列添加唯一标识符。
- 您可以从文本文件中添加自定义关键字—这些字在数据中标识。
- 您可以使用正则表达式(regex)添加个人模式—正则表达式将添加到现有预定义模式中。
- 您可以添加自定义类别以确定数据中的特定信息类别。

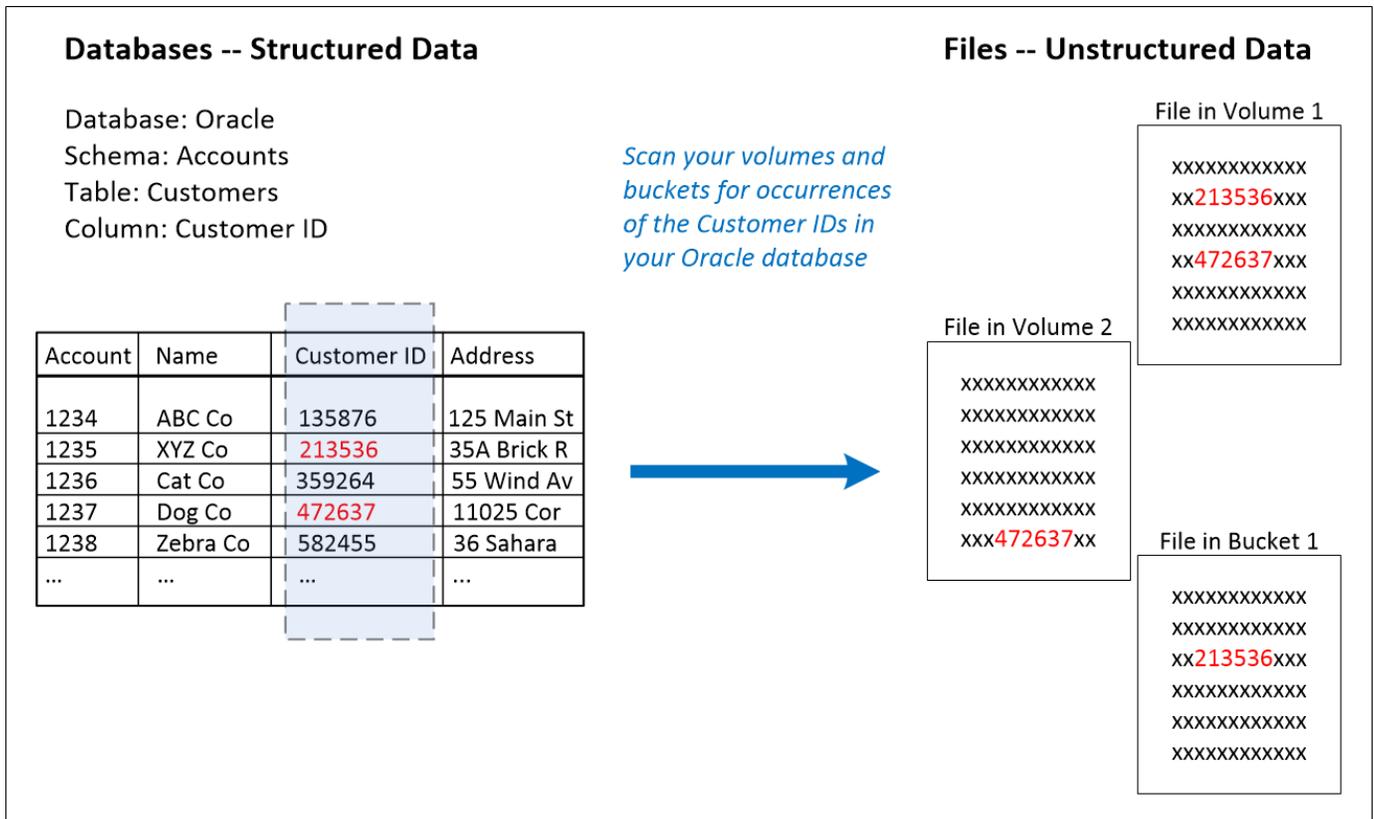
所有语言均支持所有这些添加自定义扫描条件的机制。



只有在选择对数据源执行完整分类扫描后，才可以使用本节所述的功能。已执行仅映射扫描的数据源不会显示文件级详细信息。

## 从数据库中添加自定义个人数据标识符

我们称为 `Data Fusion` 的一项功能可用于扫描组织的数据、以确定您的数据库中的唯一标识符是否位于任何其他数据源中。您可以通过选择数据库表中的一个或多个特定列来选择 BlueXP 分类在其扫描中查找的其他标识符。例如，下图显示了如何使用 `Data Fusion` 扫描卷，分段和数据库，以确定 Oracle 数据库中所有客户 ID 是否存在。



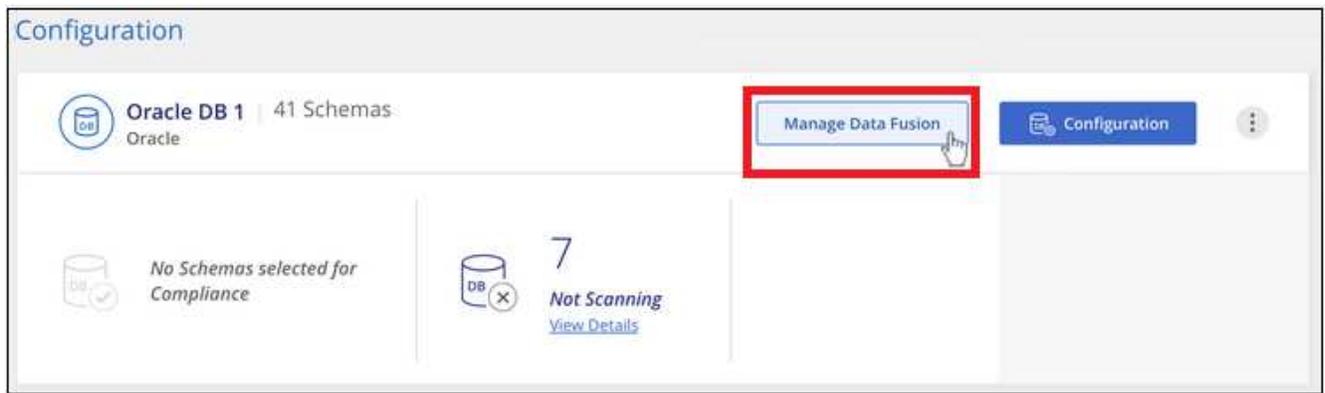
如您所见，在两个卷和一个 S3 存储分段中发现了两个唯一的客户 ID。此外，还将确定数据库表中的任何匹配项。

请注意、由于您正在扫描自己的数据库、因此在未来的 BlueXP 分类扫描中、将使用存储数据的任何语言来标识数据。

### 步骤

您必须拥有 ["已至少添加一个数据库服务器"](#) 到 BlueXP 分类、然后才能添加数据 Fusion 源。

1. 在配置页面中，单击源数据所在数据库中的 `* 管理 Data Fusion *`。



2. 单击下一页的 \* 添加 Data Fusion 源 \*。
3. 在 *Add Data Fusion Sources* 页面中：
  - a. 从下拉菜单中选择数据库架构。
  - b. 输入该模式中的表名称。
  - c. 输入包含要使用的唯一标识符的 Column 或 Columns。

添加多个列时，请在单独的行中输入每个列名称或表视图名称。

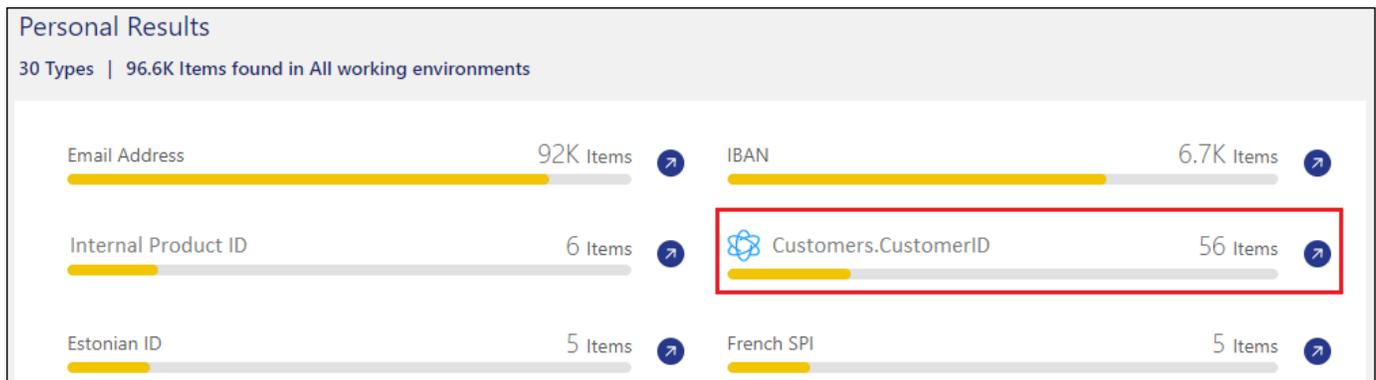
The screenshot shows the 'Add Data Fusion Source' form. It includes instructions: 'To add a Data Fusion source reference, specify one or more columns which contain your organization's unique identifiers, such as a column used to store customer IDs. Note that adding a Data Fusion Source will initiate an additional scan of your data stores.' The form has two input fields: 'Database Schema' with 'Oracle1,Accounts' and 'Table' with 'Customers'. Below these is a section 'Columns Containing Identifiers' with a list box containing 'Customer ID'. At the bottom are 'Add Data Fusion Source' and 'Cancel' buttons.

4. 单击 \* 添加 Data Fusion 源 \*。

Oracle DB 1 Data Fusion			+ Add Data Fusion source
Database Schema	Table	Data Fusion Source Columns	
Schema1	Table 1	Column 12, Column 4, Column 18	...
Schema2	Table 2	Column 2, Column 14, Column 8	...

## 结果

下次扫描后、结果将在"个人结果"部分的"合规性信息板"中以及"个人数据"筛选器的"调查"页面中包含此新信息。例如，您用于此分层器的名称将显示在筛选器列表中 Customers.CustomerID。



## 删除Data Fusion 源

如果您在某个时刻决定不使用某个 Data Fusion 源扫描文件，则可以从 Data Fusion 清单页面中选择源行，然后单击 \* 删除 Data Fusion 源 \*。



## 从词列表中添加自定义关键字

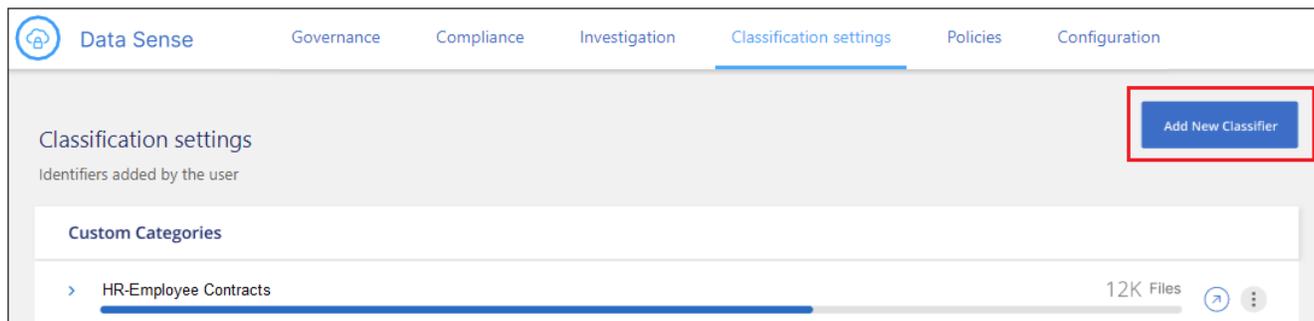
您可以将自定义关键字添加到BlueXP分类中、以便标识该信息在数据中的位置。您只需输入希望BlueXP分类识别的每个单词即可添加关键字。关键字将添加到BlueXP分类已使用的现有预定义关键字中、结果将显示在个人模式部分下。

例如、您可能希望查看所有文件中提及内部产品名称的位置、以确保这些名称在不安全的位置不可访问。

更新自定义关键字后、BlueXP分类将重新开始扫描所有数据源。扫描完成后、新结果将显示在BlueXP分类合规性信息板中的"个人结果"部分下、以及调查页面中的"个人数据"筛选器中。

## 步骤

1. 在\_Classification settings\_选项卡中、单击\*添加新的类元\*以启动\_Add Custom Classifier\_向导。



2. 在 \_Select type\_ 页面中、输入分类器的名称、提供简短的问题描述、选择\*个人标识符\*、然后单击\*下一步\*。

您输入的名称将在BlueXP分类UI中显示为符合分类器要求的已扫描文件的标题、并在"调查"页面中显示为筛选器的名称。

您还可以选中"屏蔽系统中检测到的结果"复选框、以便完整结果不会显示在UI中。例如、您可能需要执行此操作以隐藏完整的信用卡号或类似的个人数据(掩码将显示在UI中、如下所示: "passe: [\$\$\$] pass: [\$\$\$] pass: [\$\$\$] passe: [\$\$\$]"3434)。

1 Select type   2 Select tool   3 Create Logic

## Select type

Select the type of classifier that you want to add to the system, and provide the name and description. Data Sense re-scans all your data sources after you add a new classifier. When the scan is complete, all matching results are displayed in the "Classification Settings" dashboard and in other Data Sense pages.

---

Classifier name

Description

**Personal identifier**

The classifier will be added to the system as a new personal identifier. Any matches are considered "personal data", and they are added to the results that are displayed in the Personal Results page and in the Investigation page. [See the list of personal data that Data Sense identifies by default.](#)

Mask detected results in the system

**Category**

The classifier will be added to the system as a new Category. Any matches are added to the results that are displayed in the Categories page and in the Investigation page. [See the list of categories that Data Sense identifies by default.](#)

3. 在\_Select数据分析工具\_页面中、选择\*自定义关键字\*作为要用于定义分类器的方法、然后单击\*下一步\*。

## Select Data Analysis Tool

Select the tool that will be used to build the list of words, or patterns, that Data Sense will attempt to match in your data sources.

**Custom keywords** ⓘ  
Create a custom personal pattern based on a list of keywords that you provide.

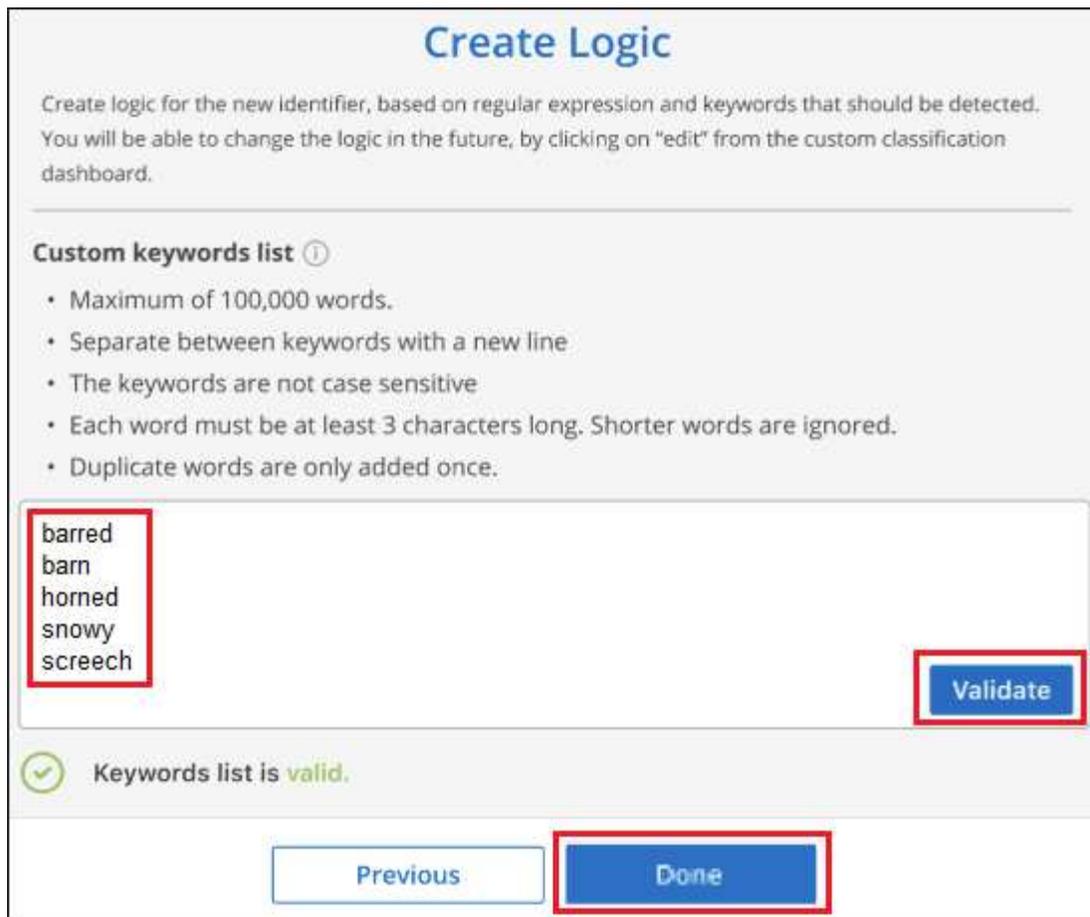
**Custom regular expression** ⓘ  
Create a custom personal pattern based on a regular expression that you define.

**DB fusion** ⓘ  
Create a custom personal pattern based on the values found in specific columns in your scanned database tables. This allows you to identify whether unique identifiers from your databases are found in any of your other data sources.

[Previous](#) [Next](#)

4. 在\_Create Logic\_页面中、输入要识别的关键字-每一个词位于一条单独的行上-然后单击\*验证\*。

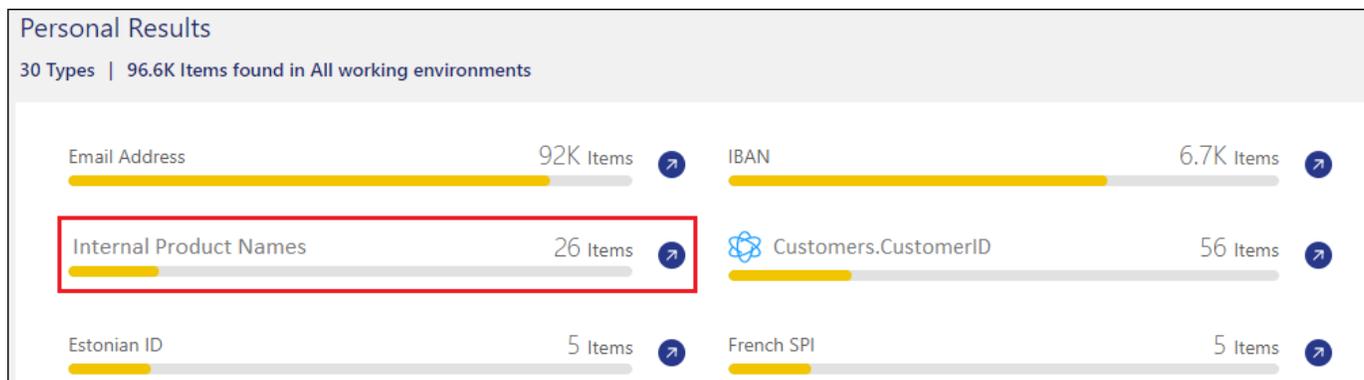
以下屏幕截图显示了内部产品名称(不同类型的所有权)。这些项目的BlueXP分类搜索不区分大小写。



5. 单击\*Done (完成)\*、BlueXP分类将开始重新扫描您的数据。

### 结果

扫描完成后、结果将在"个人结果"部分下的"合规性信息板"中以及"个人数据"筛选器的"调查"页面中包含此新信息。



如您所见、此类元的名称将在"个人结果"面板中用作名称。通过这种方式、您可以激活多组不同的关键字、并查看每个组的结果。

### 使用正则表达式添加自定义个人数据标识符

您可以使用自定义正则表达式(regex)添加个人模式来标识数据中的特定信息。这样、您就可以创建一个新的自定义regex、以确定系统中尚不存在的新个人信息元素。正则表达式将添加到BlueXP分类已使用的现有预定义模式中、结果将显示在个人模式部分下。

例如、您可能希望查看所有文件中提及内部产品ID的位置。如果产品ID结构清晰、例如、它是一个以201开头的12位数、则可以使用自定义正则表达式功能在文件中搜索它。此示例的正则表达式为`*.b201\d {9} \b*`。

添加正则表达式后、BlueXP分类将重新开始扫描所有数据源。扫描完成后、新结果将显示在BlueXP分类合规性信息板中的"个人结果"部分下、以及调查页面中的"个人数据"筛选器中。

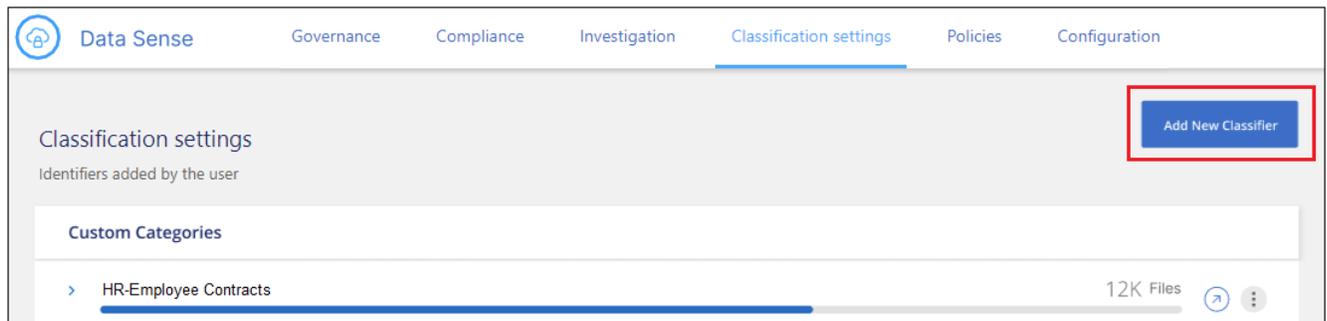
如果在构建正则表达式时需要帮助、请参见 ["正则表达式101"](#)。选择\*PYTHYTH\*作为风味、以查看BlueXP分类将与正则表达式匹配的结果类型。。 ["Python Regex测试程序页面"](#) 通过以图形方式显示模式也很有用。



目前、我们不允许在创建正则表达式时使用模式标志-这意味着您不应使用"/"。

## 步骤

1. 在 `_Classification settings_` 选项卡中、单击\*添加新的类元\*以启动 `_Add Custom Classifier_` 向导。



2. 在 `_Select type_` 页面中、输入分类器的名称、提供简短的问题描述、选择\*个人标识符\*、然后单击\*下一步\*。

您输入的名称将在BlueXP分类UI中显示为符合分类器要求的已扫描文件的标题、并在"调查"页面中显示为筛选器的名称。您还可以选中"屏蔽系统中检测到的结果"复选框、以便完整结果不会显示在UI中。例如、您可能希望执行此操作以隐藏完整的信用卡号或类似的个人数据。

1 Select type      2 Select tool      3 Create Logic

## Select type

Select the type of classifier that you want to add to the system, and provide the name and description. Data Sense re-scans all your data sources after you add a new classifier. When the scan is complete, all matching results are displayed in the "Classification Settings" dashboard and in other Data Sense pages.

Classifier name

Description

**Personal identifier**

The classifier will be added to the system as a new personal identifier. Any matches are considered "personal data", and they are added to the results that are displayed in the Personal Results page and in the Investigation page. [See the list of personal data that Data Sense identifies by default.](#)

Mask detected results in the system

**Category**

The classifier will be added to the system as a new Category. Any matches are added to the results that are displayed in the Categories page and in the Investigation page. [See the list of categories that Data Sense identifies by default.](#)

3. 在\_Select数据分析工具\_页面中、选择\*自定义正则表达式\*作为要用于定义分类器的方法、然后单击\*下一步\*。

## Select Data Analysis Tool

Select the tool that will be used to build the list of words, or patterns, that Data Sense will attempt to match in your data sources.

**Custom keywords** ⓘ  
Create a custom personal pattern based on a list of keywords that you provide.

**Custom regular expression** ⓘ  
Create a custom personal pattern based on a regular expression that you define.

**DB fusion** ⓘ  
Create a custom personal pattern based on the values found in specific columns in your scanned database tables. This allows you to identify whether unique identifiers from your databases are found in any of your other data sources.

[Previous](#) [Next](#)

4. 在 `_Create Logic_` 页面中、输入正则表达式和任何邻近词、然后单击\*完成\*。
  - a. 您可以输入任何合法正则表达式。单击\*Validify\*按钮可让BlueXP分类验证正则表达式是否有效，并且不会太宽泛，这意味着它将返回太多结果。
  - b. 或者、您也可以输入一些接近词来帮助细化结果的准确性。这些字词通常位于所搜索模式的300个字符内(在找到的模式之前或之后)。在单独的行中输入每个词或短语。

## Create Logic

Create logic for the new identifier, based on regular expression and keywords that should be detected.

---

**Regular expression** ⓘ

Add the pattern that should be detected to identify specific information in your data, using a custom regular expression.

Validate

✔ **Success:** Regular expression is valid.

**Proximity words** - To improve the detection accuracy, insert phrases that must appear near by the regular expression's match.

Previous
Done

### 结果

此时将添加分类器、BlueXP分类将开始重新扫描所有数据源。此时将返回自定义类元页面、在此页面中、您可以查看与新类元匹配的文件数。扫描所有数据源的结果将需要一段时间、具体取决于需要扫描的文件数量。

Data Sense
Governance
Compliance
Investigation
Classification settings
Policies
Configuration

### Classification settings

Add New Classifier

Identifiers added by the user

**Custom Categories**

> HR - Employee Contracts

7.5K Files

**Personal information**

> Internal Product ID

12K Files

### 添加自定义类别

BlueXP分类采用它扫描的数据并将其划分为不同类型的类别。类别是基于对每个文件的内容和元数据进行人工智能分析的主题。"请参见预定义类别列表"。

类别可以通过向您显示所拥有的信息类型来帮助您了解数据的变化。例如、\_resumes\_或\_empleeee contracts\_等类别可能包含敏感数据。调查结果时，您可能会发现员工合同存储在不安全的位置。然后，您可以更正此问题描述。

您可以将自定义类别添加到BlueXP分类中、以便确定数据资产的唯一信息类别在数据中的位置。您可以通过创建包含要标识的数据类别的"训练"文件来添加每个类别、然后让BlueXP分类扫描这些文件、以便通过AI进行"学习"、以便它可以标识数据源中的数据。这些类别将添加到BlueXP分类已标识的现有预定义类别中、结果将显示在"类别"部分下。

例如、您可能希望查看.gz格式的压缩安装文件在文件中的位置、以便在必要时将其删除。

更新自定义类别后、BlueXP分类将重新开始扫描所有数据源。扫描完成后、新结果将显示在BlueXP分类合规性信息板中的"类别"部分下、以及调查页面中的"类别"筛选器中。 ["请参见如何按类别查看文件"](#)。

### 您需要的内容

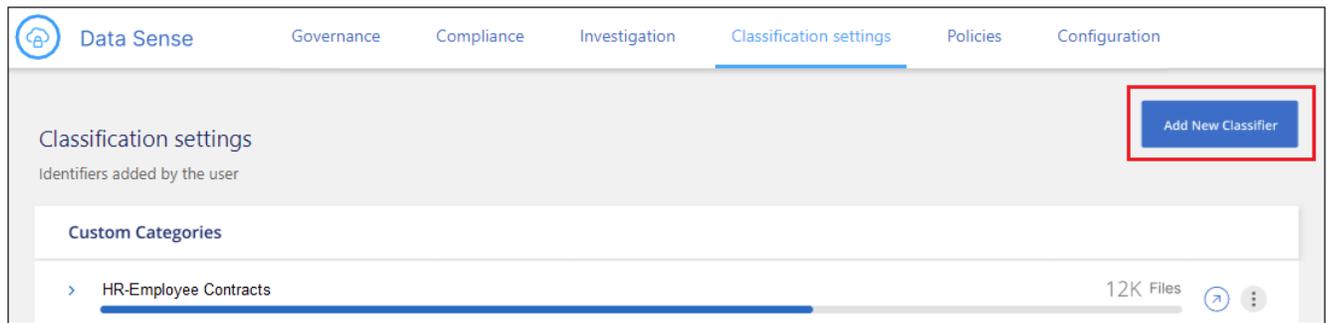
您至少需要创建25个培训文件、这些文件包含您希望BlueXP分类识别的数据类别的示例。支持以下文件类型：

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

这些文件必须至少为100字节、并且必须位于可通过BlueXP分类访问的文件夹中。

### 步骤

1. 在\_Classification settings\_选项卡中、单击\*添加新的类元\*以启动\_Add Custom Classifier\_向导。



2. 在\_Select type\_页面中、输入分类器的名称、提供简短的问题描述、选择\*类别\*、然后单击\*下一步\*。

您输入的名称将在BlueXP分类UI中显示为与您定义的数据类别匹配的已扫描文件的标题、并在调查页面中显示为筛选器的名称。

1 Select type      2 Select tool      3 Create Logic

## Select type

Select the type of classifier that you want to add to the system, and provide the name and description. Data Sense re-scans all your data sources after you add a new classifier. When the scan is complete, all matching results are displayed in the "Classification Settings" dashboard and in other Data Sense pages.

Classifier name

Compressed installer files

Description

Installation files in .GZ format

Personal identifier

The classifier will be added to the system as a new personal identifier. Any matches are considered "personal data", and they are added to the results that are displayed in the Personal Results page and in the Investigation page. [See the list of personal data that Data Sense identifies by default.](#)

Mask detected results in the system

Category

The classifier will be added to the system as a new Category. Any matches are added to the results that are displayed in the Categories page and in the Investigation page. [See the list of categories that Data Sense identifies by default.](#)

Previous      Next

3. 在\_Create Logic\_页面中、确保已准备好学习文件、然后单击\*选择文件\*。

## Create Logic

**AI-based similarity training** ⓘ

- Insert NFS folder path
- The folder should contain minimum 25 files and maximum 1000 files that will be used for the AI training.
- Supported file types: pdf, docx, doc, pptx, xls, xlsx, csv, txt, gz, rtf, docs, sheets, slides, json
- The keywords are not case sensitive
- Minimum file size: 100B

Compressed Installer files

Select Files

4. 输入卷的IP地址以及培训文件所在的路径、然后单击\*添加\*。

**Insert folder path that contains at least 25 files for the training**

Enter the IP address and volume name, along with the path to the location of the training files.

IP:

Training Data - Folder path:

5. 验证训练文件是否已被BlueXP分类识别。单击\*。删除不符合要求的任何培训文件。然后单击\*完成。

### Create Logic

**AI-based similarity training**

- Insert NFS folder path
- The folder should contain minimum 25 files and maximum 1000 files that will be used for the AI training.
- Supported file types: pdf, docx, doc, pptx, xls, xlsx, csv, txt, gz, rtf, docs, sheets, slides, json
- The keywords are not case sensitive
- Minimum file size: 100B

**Compressed Installer files**

Total uploaded files: 54

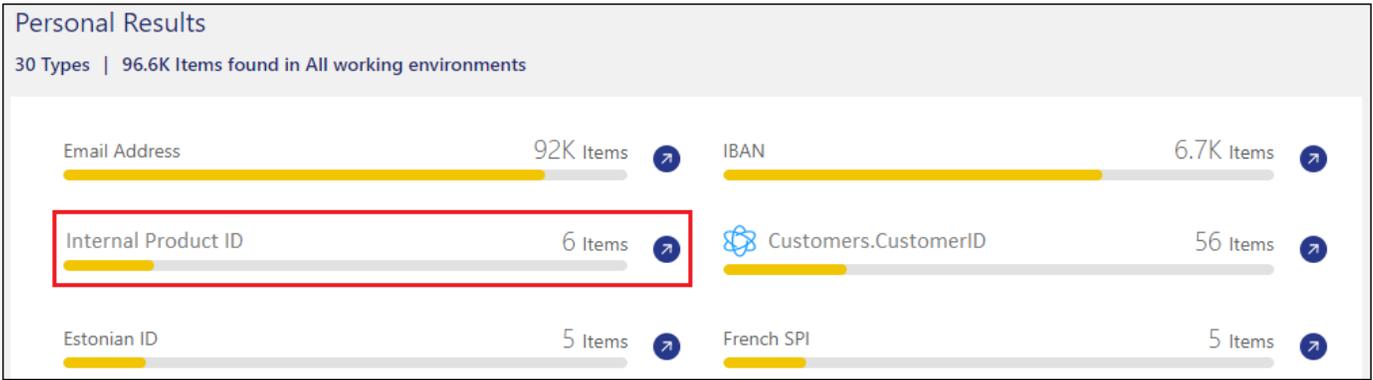
File name	File Size	File Type	Reliability	Included in training
File1	56	File type	Sufficient	×
File2	22	File type	Sufficient	×
File3	43	File type	Sufficient	×
File4	11	File type	Sufficient	×

## 结果

根据培训文件的定义创建新类别、并将其添加到BlueXP分类中。然后、BlueXP分类开始重新扫描所有数据源、以确定适合此新类别的文件。此时将返回自定义类元页面、在此页面中、您可以查看与新类别匹配的文件数。扫描所有数据源的结果将需要一段时间、具体取决于需要扫描的文件数量。

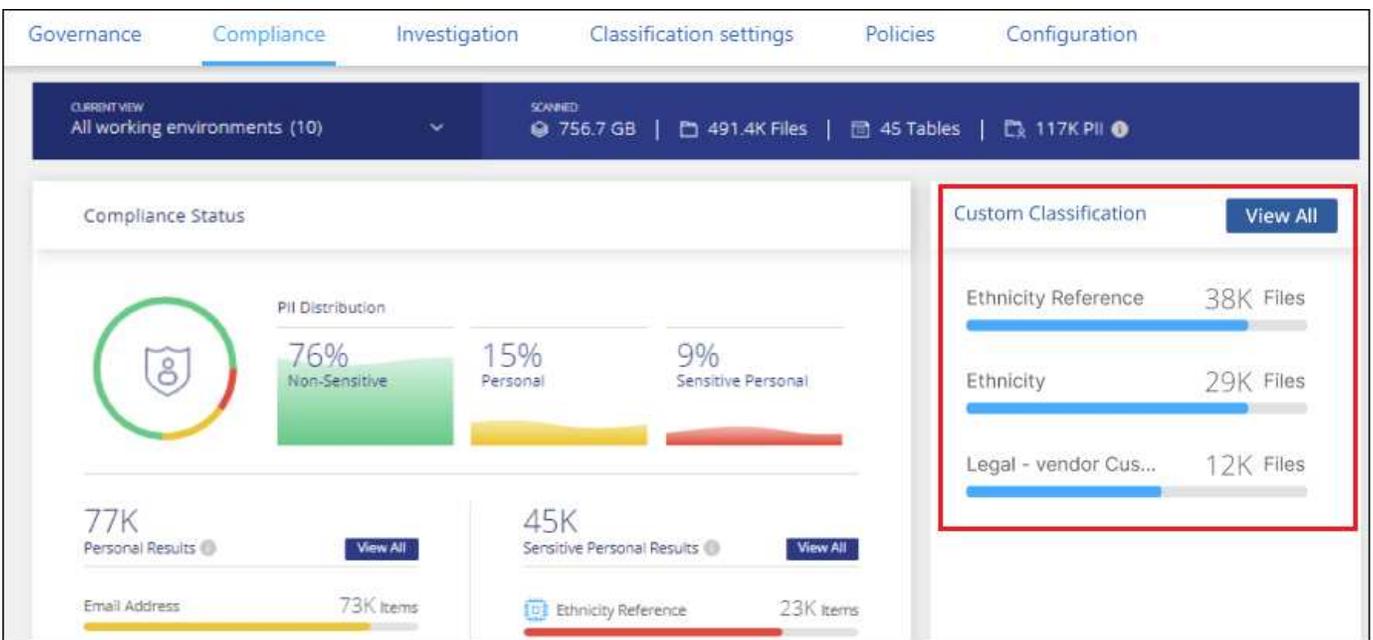
## 查看自定义分类器的结果

您可以在合规性信息板和调查页面中查看任何自定义分类器的结果。例如、此屏幕截图在"个人结果"部分的"合规性信息板"中显示匹配的信息。



单击  按钮以在"调查"页面中查看详细结果。

此外、所有自定义分类器结果都会显示在"Custom Classifiers (自定义分类器)"选项卡中、排名前6位的自定义分类器结果会显示在"Compliance Dashboard"(合规性信息板)中、如下所示。



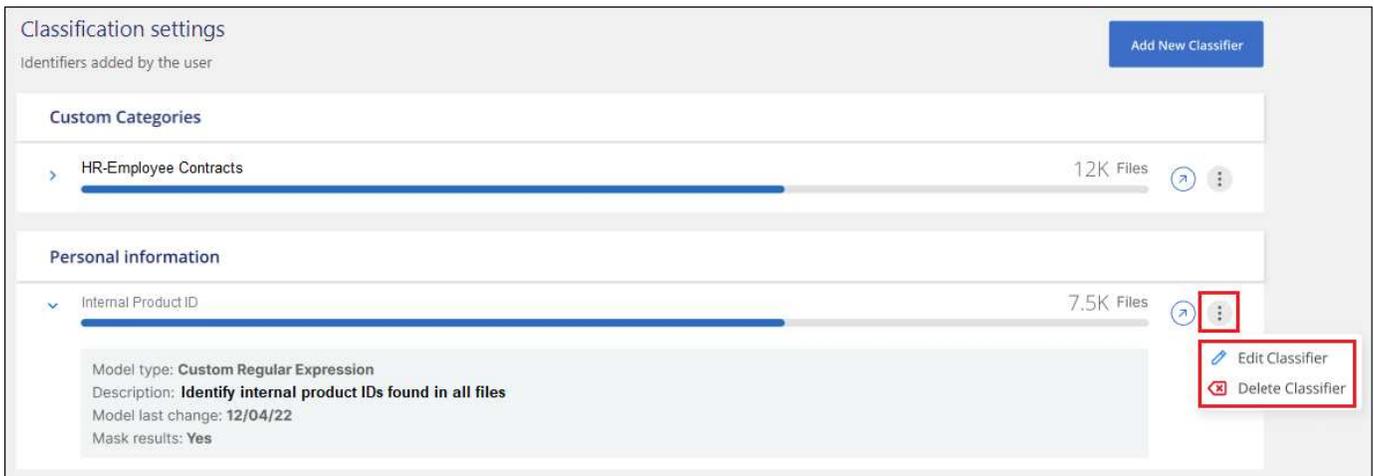
### 管理自定义分类器

您可以使用\*编辑类元\*按钮更改已创建的任何自定义类元。



此时无法编辑Data Fusion类元。

如果您稍后决定不需要BlueXP分类来识别您添加的自定义模式，则可以使用\*删除分类器\*按钮删除每个项目。



## 查看合规性操作的状态

从"调查结果"窗格对多个文件运行异步操作时、例如移动或删除100个文件、此过程可能需要一些时间。您可以在 `_Action Status_` 窗格中监控这些操作的状态、以便了解何时将其应用于所有文件。

这样，您就可以查看已成功完成的操作，当前正在执行的操作以及失败的操作，以便诊断和修复任何问题。请注意、移动单个文件等快速完成的短操作不会显示在操作状态窗格中。

\*注\*此信息仅与BlueXP分类的旧版版本1.3及更早版本相关。

状态可以是：

- 成功—BlueXP分类操作已完成、所有项目均成功。
- 部分成功—BlueXP分类操作已完成、一些项目失败、一些成功。
- In progress—操作仍在进行中。
- Queued—操作尚未启动。
- 已取消-操作已取消。
- failed—操作失败。

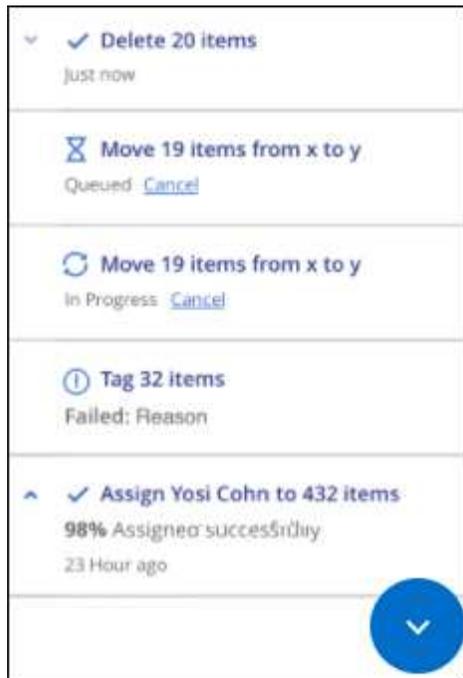
请注意，您可以取消状态为 " 已排队 " 或 " 正在进行 " 的任何操作。

步骤

1.

在BlueXP分类UI的右下角可以看到\*Actions Status\*按钮 。

2. 单击此按钮，将列出最近 20 个操作。



您可以单击某个操作的名称以查看与该操作对应的详细信息。

## 审核BlueXP分类操作的历史记录

BlueXP分类可记录对BlueXP分类正在扫描的所有工作环境和数据源中的文件执行的管理活动。BlueXP分类还会记录部署BlueXP分类实例时的活动。

您可以查看BlueXP分类审核日志文件的内容、也可以下载这些文件、以查看发生了哪些文件更改以及更改的时间。例如、您可以查看发出的请求、请求的时间以及详细信息、例如删除文件时的源位置或移动文件时的源和目标位置。

\*注\*此信息仅与BlueXP分类的旧版版本1.3及更早版本相关。

## 日志文件内容

审核日志中的每一行都包含以下格式的信息：

```
<full date> | <status> | ds_audit_logger | <module> | 0 | 0 | File <full file path> deleted from device <device path> - <result>
```

- date and time—事件的完整时间戳
- status—信息、警告
- 操作类型(删除、复制、移动、创建策略、更新策略、重新扫描文件、下载JSON报告等)
- 文件名(如果操作与文件相关)
- 操作的详细信息—已执行的操作：取决于操作
  - Policy name

- for move—源和目标
- 对于copy—源和目标
- 标记-标记名称
- 用于assign to—用户名
- 对于电子邮件警报—电子邮件地址/帐户

例如、日志文件中的以下行显示复制操作成功、复制操作失败。

```
2022-06-06 15:23:08,910 | INFO | ds_audit_logger | es_scanned_file | 237 | 49 | Copy file /CIFS_share/data/dopl/random_positives.tsv from device 10.31.133.183 (type: SMB_SHARE) to device 10.31.130.133:/export_reports (NFS_SHARE) - SUCCESS
2022-06-06 15:23:08,968 | WARNING | ds_audit_logger | es_scanned_file | 239 | 153 | Copy file /CIFS_share/data/compliance-netapp.tar.gz from device 10.31.133.183 (type: SMB_SHARE) to device 10.31.130.133:/export_reports (NFS_SHARE) - FAILURE
```

## 日志文件位置

管理审核日志文件位于BlueXP分类计算机的以下位置： /opt/netapp/audit\_logs/

安装审核日志文件将写入 /opt/netapp/install\_logs/

每个日志文件的大小最多可以为10 MB。达到此限制后、将启动一个新的日志文件。这些日志文件名为"DataSICE\_AUDIST.LOG\_"、"DataSICE\_AUDULT\_LOG\_1"、"DataSICE\_AUDIST\_LOG.2"等。系统中最多保留100个日志文件-达到最大值后、较早的日志文件将自动删除。

## 访问日志文件

您需要登录BlueXP分类系统才能访问日志文件。请参见操作说明 ["登录到BlueXP分类系统"](#) 具体取决于您是在Linux计算机上手动安装软件、还是在云中部署了实例。

## 降低BlueXP分类扫描速度

数据扫描对存储系统和数据的影响可以忽略不计。但是、如果您担心的影响非常小、则可以将BlueXP分类配置为执行"慢速"扫描。

启用后，所有数据源都会使用慢速扫描—您不能为单个工作环境或数据源配置慢速扫描。

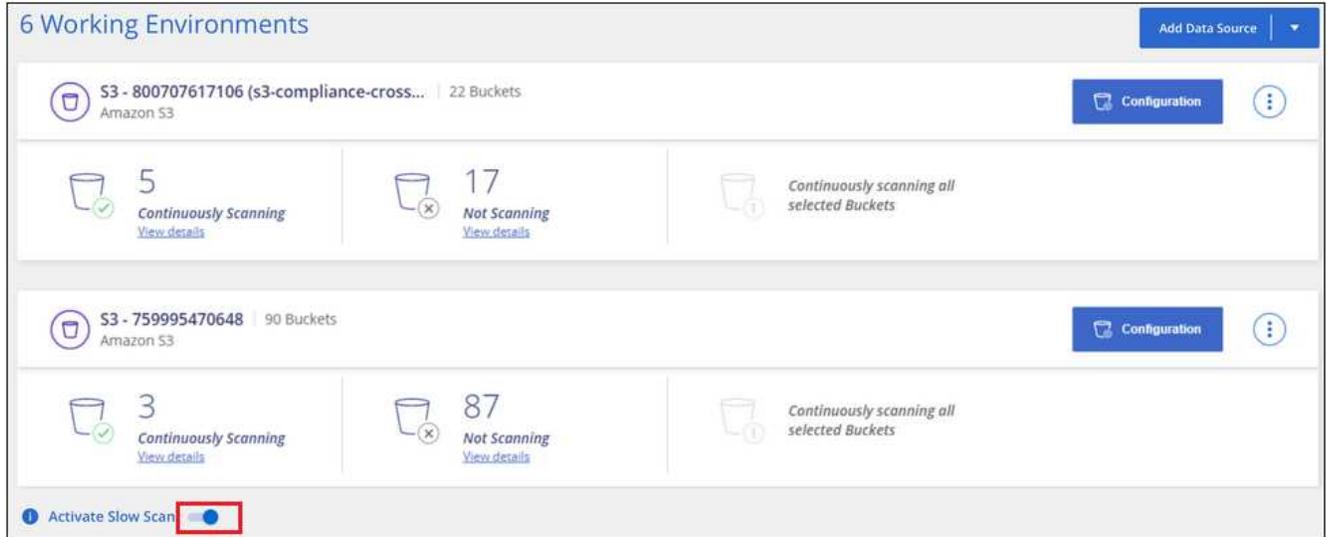


扫描数据库时，扫描速度不能降低。

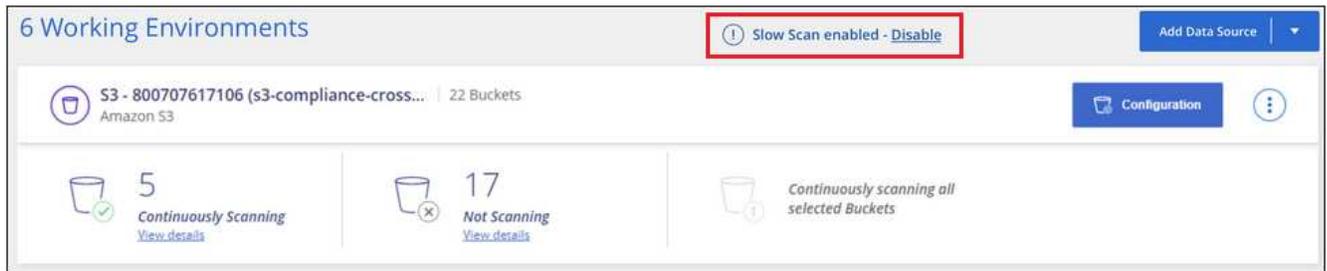
\*注\*此信息仅与BlueXP分类的旧版版本1.3及更早版本相关。

## 步骤

1. 从 *Configuration* 页面底部，向右移动滑块以激活慢速扫描。



配置页面顶部指示已启用慢速扫描。



2. 您可以通过单击此消息中的 \* 禁用 \* 来禁用慢速扫描。

## 从BlueXP分类中删除OneDrive、SharePoint或Google Drive帐户

如果您不再需要扫描特定OneDrive帐户、特定SharePoint帐户或Google Drive帐户中的用户文件、则可以从BlueXP分类界面中删除该帐户并停止所有扫描。

步骤

1. 从 *Configuration* 页面中，单击  按钮、然后单击\*删除OneDrive帐户\*、删除SharePoint帐户\*或\*删除Google Drive帐户。



2. 单击确认对话框中的 \* 删除帐户 \*。

## 版权信息

版权所有 © 2024 NetApp, Inc.。保留所有权利。中国印刷。未经版权所有者事先书面许可，本档中受版权保护的任何部分不得以任何形式或通过任何手段（图片、电子或机械方式，包括影印、录音、录像或存储在电子检索系统中）进行复制。

从受版权保护的 NetApp 资料派生的软件受以下许可和免责声明的约束：

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

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

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

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

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

## 商标信息

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