データ分類を使用する
NetApp Data Classification

NetApp
February 02, 2026

目次

データ分類を使用する 1
NetApp Data Classificationを使用して、組織に保存されているデータのガバナンスの詳細を表示します。 1
ガバナンスダッシュボードを確認する 1
データ検出評価レポートを作成する 3
データマッピングの概要レポートを作成する 4
NetApp Data Classification を使用して、組織内に保存されている個人データのコンプライアンスの詳細を表示します。 6
個人データを含むファイルを表示する 7
機密性の高い個人データを含むファイルを表示する 11
NetApp Data Classificationにおけるプライベートデータのカテゴリ 14
個人データの種類 14
機密個人データの種類 18
カテゴリーの種類 18
ファイルの種類 20
発見された情報の正確性 20
NetApp Data Classificationでカスタム分類を作成する 21
カスタム個人識別子を作成する 21
カスタムカテゴリを作成する 25
カスタム分類子を編集する 26
カスタム分類子を削除する 27
次のステップ 27
NetApp Data Classificationを使用して組織内に保存されているデータを調査します 27
データ調査構造 27
データフィルター 27
ファイルのメタデータを表示 30
ファイルとディレクトリのユーザー権限を表示する 32
ストレージシステム内の重複ファイルをチェックする 32
レポートをダウンロードする 33
選択したフィルターに基づいて保存されたクエリを作成する 36
NetApp Data Classificationで保存したクエリを管理する 38
調査ページで保存したクエリの結果を表示する 39
保存されたクエリとポリシーを作成する 39
保存したクエリまたはポリシーを編集する 41
保存したクエリを削除する 42
デフォルトのクエリ 42
リポジトリのNetApp Data Classificationスキャン設定を変更する 43
リポジトリのスキャンステータスを表示する 43
リポジトリのスキャンの種類を変更する 44

スキャンを優先する . . . 45
リポジトリのスキャンを停止する . . . 46
リポジトリのスキャンを一時停止して再開する . . . 47
NetApp Data Classificationコンプライアンスレポートを表示 . . . 47
レポートのシステムを選択する . . . 48
データ主体アクセス要求レポート . . . 49
医療保険の携行性と責任に関する法律（HIPAA）に関する報告書 . . . 51
ペイメントカード業界データセキュリティ基準（PCI DSS）レポート . . . 52
プライバシーリスク評価レポート . . . 53
NetApp Data Classificationの健全性を監視する . . . 55
ヘルスモニターの洞察 . . . 55
ヘルスモニターダッシュボードにアクセスする . . . 56

データ分類を使用する

NetApp Data Classificationを使用して、組織に保存されているデータのガバナンスの詳細を表示します。

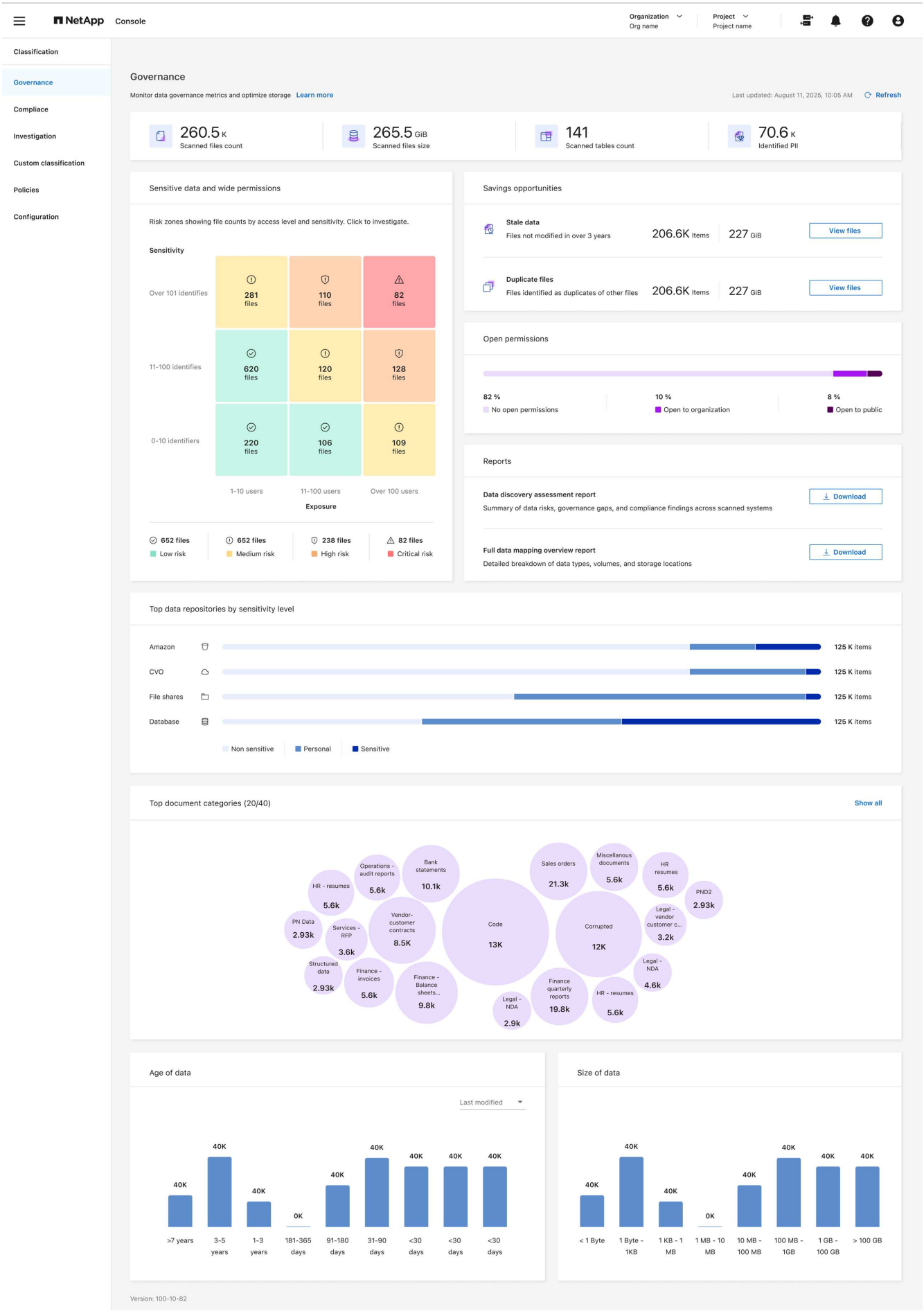
組織のストレージ リソース上のデータに関連するコストを管理します。 NetApp Data Classification は、システム内の古いデータ、重複ファイル、非常に大きなファイルの量を識別するので、一部のファイルを削除するか、より安価なオブジェクト ストレージに階層化するかを決定できます。

ここから研究を始めるべきです。ガバナンス ダッシュボードから、さらに調査する領域を選択できます。

さらに、オンプレミスの場所からクラウドにデータを移行する予定の場合は、データを移動する前に、データのサイズや、データに機密情報が含まれているかどうかを確認できます。

ガバナンスダッシュボードを確認する

ガバナンス ダッシュボードは、ストレージ リソースに保存されているデータに関連する効率を高め、コストを制御できるようにするための情報を提供します。


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

手順

1. NetApp Consoleメニューから、ガバナンス > 分類 を選択します。
2. *ガバナンス*を選択します。

   ガバナンス ダッシュボードが表示されます。

節約の機会を確認する

Saving Opportunities コンポーネントには、削除したり、より安価なオブジェクト ストレージに階層化したりできるデータが表示されます。 *Saving Opportunities* のデータは 2 時間ごとに更新されます。データを手動で更新することもできます。

手順

1. データ分類メニューから、*ガバナンス*を選択します。
2. ガバナンス ダッシュボードの各節約機会タイル内で、ストレージの最適化 を選択して、調査ページでフィルター処理された結果を表示します。削除する必要があるデータや、より安価なストレージに移動する必要があるデータを見つけるには、「*Saving Opportunities*」を調べてください。
   - 古いデータ - デフォルトでは、データが最後に変更されてから 3 年以上経過している場合、そのデータは古いものと見なされます。[古いデータの定義をカスタマイズ](task-stale-data.html)できます。
   - 重複ファイル - スキャンしているデータ ソース内の他の場所に重複しているファイル。"表示される重複ファイルの種類を確認する" 。

> いずれかのデータ ソースでデータ階層化が実装されている場合、オブジェクト ストレージに既に存在する古いデータは、古いデータ カテゴリで識別できます。

データ検出評価レポートを作成する

データ検出評価レポートでは、スキャンされた環境の高レベルの分析が提供され、懸念事項と潜在的な修復手順が示されます。結果はデータのマッピングと分類の両方に基づいています。このレポートの目的は、データセットの 3 つの重要な側面についての認識を高めることです。

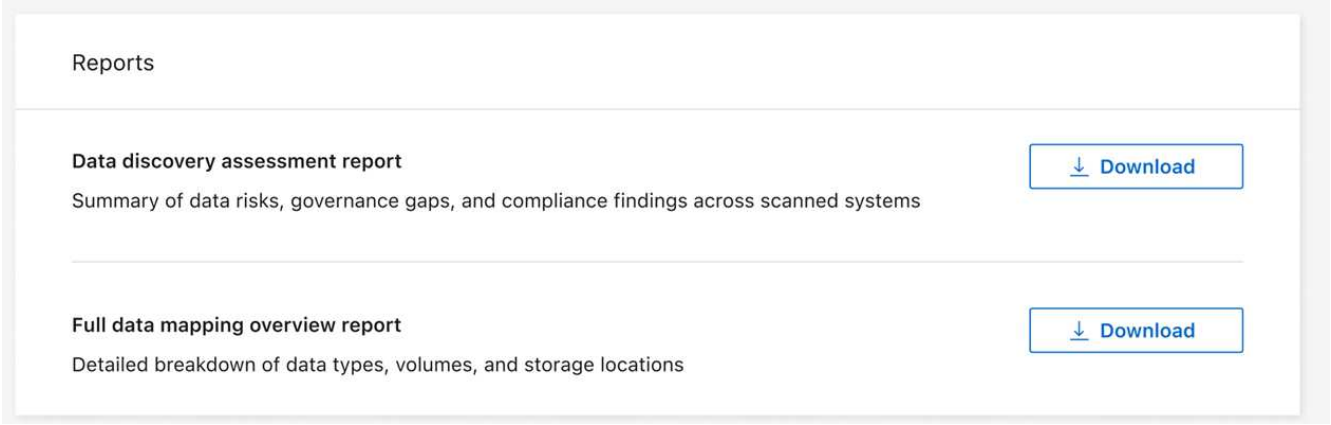
特徴	説明
データガバナンスの懸念	所有するすべてのデータの詳細な概要と、コストを節約するためにデータ量を削減できる可能性のある領域。
データセキュリティの露出	広範なアクセス権限により、内部または外部からの攻撃によってデータがアクセス可能になる領域。
データコンプライアンスのギャップ	セキュリティと DSAR (データ主体によるアクセス要求) の両方のために、個人情報または機密性の高い個人情報が保存される場所。

レポートを使用すると、次のアクションを実行できます。

- 保持ポリシーを変更するか、特定のデータ (古いデータや重複データ) を移動または削除することで、ストレージ コストを削減します。
- グローバル グループ管理ポリシーを改訂して、広範な権限を持つデータを保護します。
- PII をより安全なデータ ストアに移動することで、個人情報や機密性の高い個人情報が含まれるデータを保護します。

手順

1. データ分類から、*ガバナンス*を選択します。
2. レポート タイルで、データ検出評価レポート を選択します。



結果

データ分類では、確認および共有できる PDF レポートが生成されます。

データマッピングの概要レポートを作成する

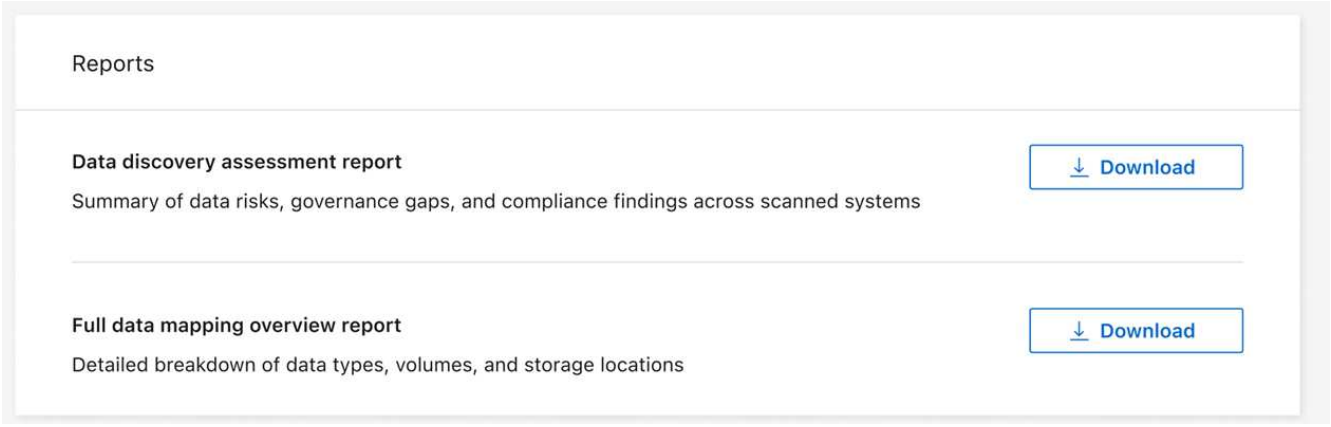
データ マッピングの概要レポートは、企業のデータ ソースに保存されているデータの概要を提供し、移行、バックアップ、セキュリティ、コンプライアンスのプロセスに関する意思決定に役立ちます。レポートでは、すべてのシステムとデータ ソースを要約します。各システムの分析も提供します。

レポートには次の情報が含まれます。

カテゴリ	説明
使用容量	すべてのシステム: 各システムのファイル数と使用容量を一覧表示します。単一システムの場合: 最も多くの容量を使用しているファイルを一覧表示します。
データの時代	ファイルが作成された日時、最後に変更された日時、最後にアクセスされた日時を示す 3 つのチャートとグラフを提供します。特定の日付範囲に基づいて、ファイルの数と使用容量を一覧表示します。
データのサイズ	システム内の特定のサイズ範囲内に存在するファイルの数を一覧表示します。

手順

1. データ分類から、*ガバナンス*を選択します。
2. レポート タイルで、*完全なデータ マッピングの概要レポート*を選択します。



結果

データ分類では、必要に応じて確認したり他のグループに送信したりできる PDF レポートが生成されます。

レポートが 1 MB より大きい場合、PDF ファイルはデータ分類インスタンスに保持され、正確な場所に関するポップアップ メッセージが表示されます。 Data Classification がオンプレミスの Linux マシンまたはクラウドに展開した Linux マシンにインストールされている場合は、PDF ファイルに直接移動できます。 Data Classification がクラウドにデプロイされている場合、PDF ファイルをダウンロードするには、Data Classification インスタンスへの SSH 認証が必要です。

データの機密性に基づいてリストされた上位のデータリポジトリを確認する

データ マッピングの概要レポートの「機密レベル別の上位データ リポジトリ」領域には、最も機密性の高い項目を含む上位 4 つのデータ リポジトリ (システムとデータ ソース) が一覧表示されます。各システムの棒グラフは次のように分かれています。

- 非機密データ
- 個人データ
- 機密性の高い個人データ

このデータは 2 時間ごとに更新され、手動で更新することもできます。

手順

1. 各カテゴリのアイテムの合計数を確認するには、バーの各セクションにカーソルを置きます。
2. 調査ページに表示される結果をフィルタリングするには、バー内の各領域を選択し、さらに調査します。

機密データと幅広い権限を確認する

ガバナンス ダッシュボードの [機密データと幅広い権限] 領域には、機密データが含まれており、幅広い権限を持つファイルの数が表示されます。表には次の種類の権限が示されています。

- 水平軸では、最も制限の厳しい権限から最も許容度の高い制限までを示します。
- 垂直軸では、最も機密性の低いデータから最も機密性の高いデータまで表示されます。

手順

1. 各カテゴリのファイルの合計数を確認するには、各ボックスの上にカーソルを置きます。

2. 調査ページに表示される結果をフィルタリングするには、ボックスを選択してさらに調査します。

オープン権限の種類別にリストされたデータを確認する

データ マッピングの概要レポートの [オープン アクセス許可] 領域には、スキャン対象のすべてのファイルに存在する各アクセス許可の種類の割合が表示されます。この表には、次の種類の権限が表示されます。

- 開く権限がありません
- 組織に開放
- 一般公開
- 不明なアクセス

手順

1. 各カテゴリのファイルの合計数を確認するには、各ボックスの上にカーソルを置きます。
2. 調査ページに表示される結果をフィルタリングするには、ボックスを選択してさらに調査します。

データの古さとサイズを確認する

データ マッピングの概要レポートの *Age* グラフと *Size* グラフの項目を調査して、削除する必要があるデータや、より安価なオブジェクト ストレージに階層化する必要があるデータがあるかどうかを確認できます。

手順

1. データの年齢グラフでデータの年齢に関する詳細を表示するには、グラフ内のポイントの上にカーソルを置きます。
2. 年齢またはサイズの範囲でフィルタリングするには、その年齢またはサイズを選択します。
    - データの年齢グラフ - データが作成された時刻、最後にアクセスされた時刻、または最後に変更された時刻に基づいてデータを分類します。
    - データ グラフのサイズ - サイズに基づいてデータを分類します。

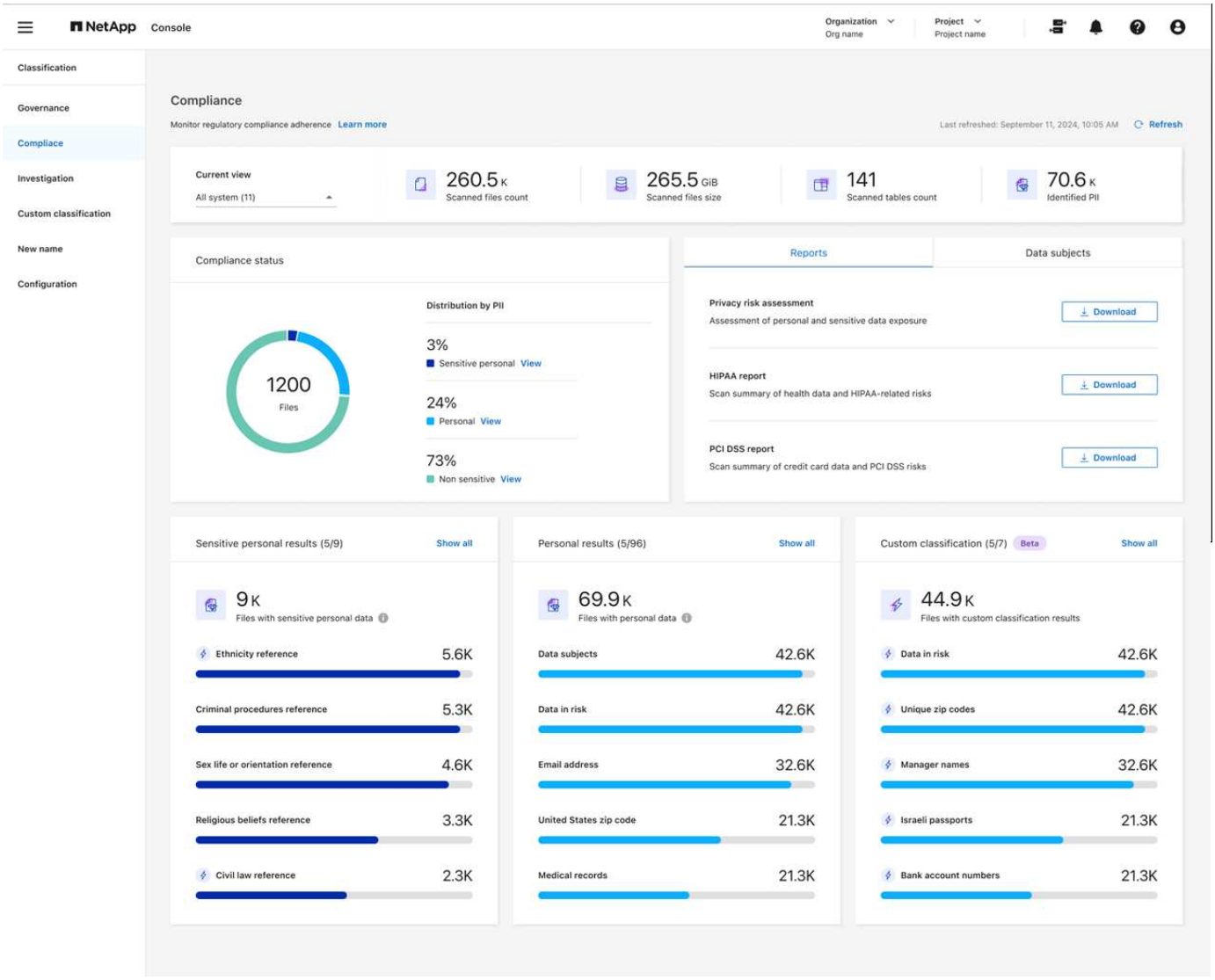
> いずれかのデータ ソースでデータ階層化が実装されている場合、オブジェクト ストレージに既に存在する古いデータが「データの年齢」グラフで識別される可能性があります。

NetApp Data Classification を使用して、組織内に保存されている個人データのコンプライアンスの詳細を表示します。

組織内の個人データ (PII) と機密個人データ (SPII) の詳細を表示して、プライベート データを制御します。また、 NetApp Data Classification がデータ内で検出したカテゴリとファイル タイプを確認することで、可視性を高めることもできます。

> ファイル レベルのコンプライアンスの詳細は、完全な分類スキャンを実行した場合にのみ利用できます。マッピングのみのスキャンでは、ファイルレベルの詳細は得られません。

デフォルトでは、データ分類ダッシュボードには、すべてのシステムとデータベースのコンプライアンス データが表示されます。一部のシステムのデータのみを表示するには、そのシステムを選択します。



データ調査ページから結果をフィルタリングし、結果のレポートを CSV ファイルとしてダウンロードできます。見る"データ調査ページでのデータのフィルタリング"詳細については。

個人データを含むファイルを表示する

データ分類は、データ内の特定の単語、文字列、パターン (正規表現) を自動的に識別します。"たとえば、クレジットカード番号、社会保障番号、銀行口座番号、パスワードなどです。"データ分類では、個々のファイル、ディレクトリ (共有とフォルダ) 内のファイル、およびデータベース テーブル内のこの種類の情報を識別します。

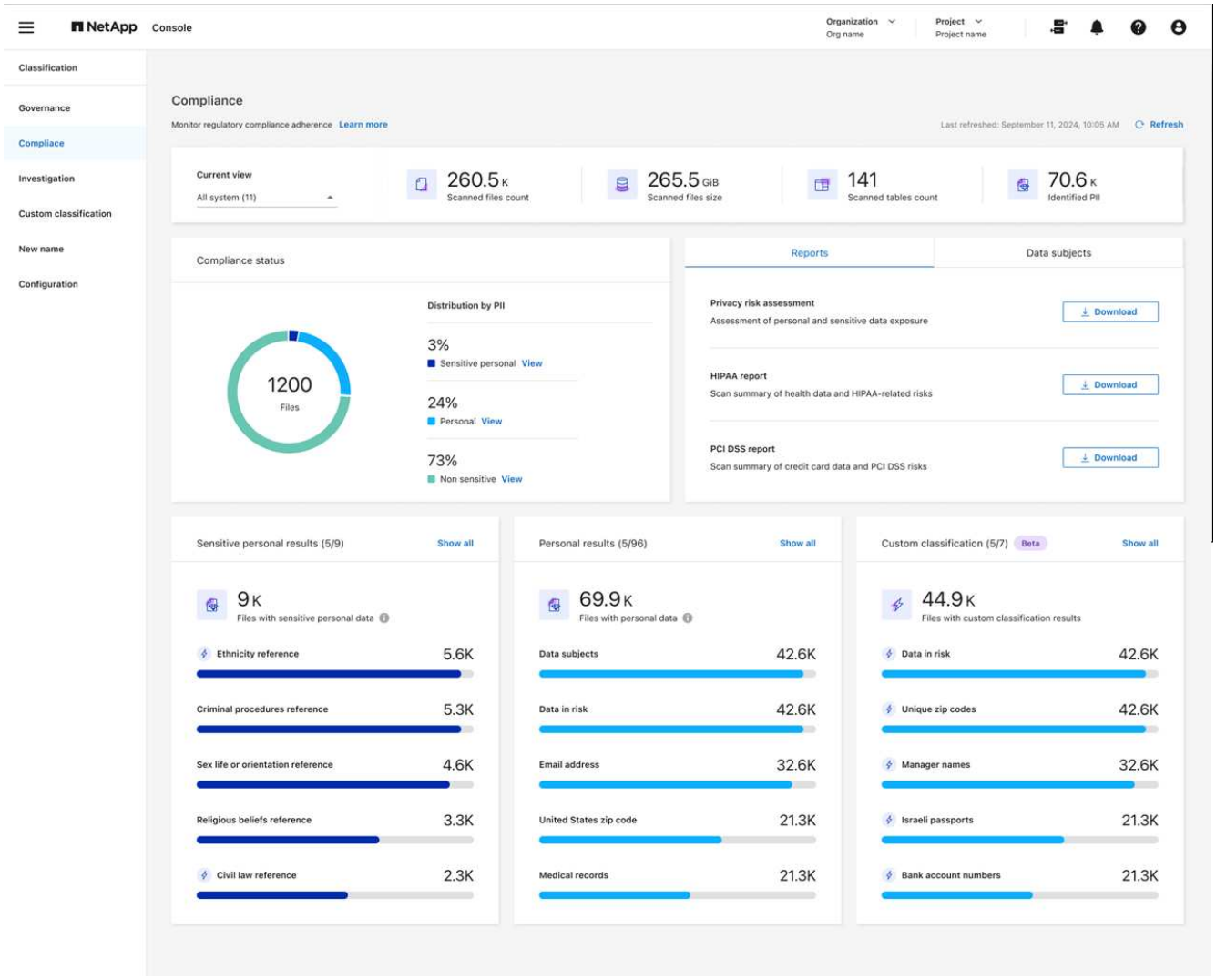
組織固有の個人データを識別するためのカスタム検索用語を作成することもできます。詳細については、以下を参照してください。 "カスタム分類を作成する" 。

一部の種類の個人データについては、データ分類は近接検証を使用して結果を検証します。検証は、見つかった個人データの近くにある 1 つ以上の定義済みキーワードを探すことによって行われます。たとえば、データ分類では、米国の社会保障番号 (SSN) の横に近接語 (たとえば、*SSN* または *social security*) がある場合、SSN として識別します。"個人データの表"データ分類で近接検証が使用されるタイミングを示します。

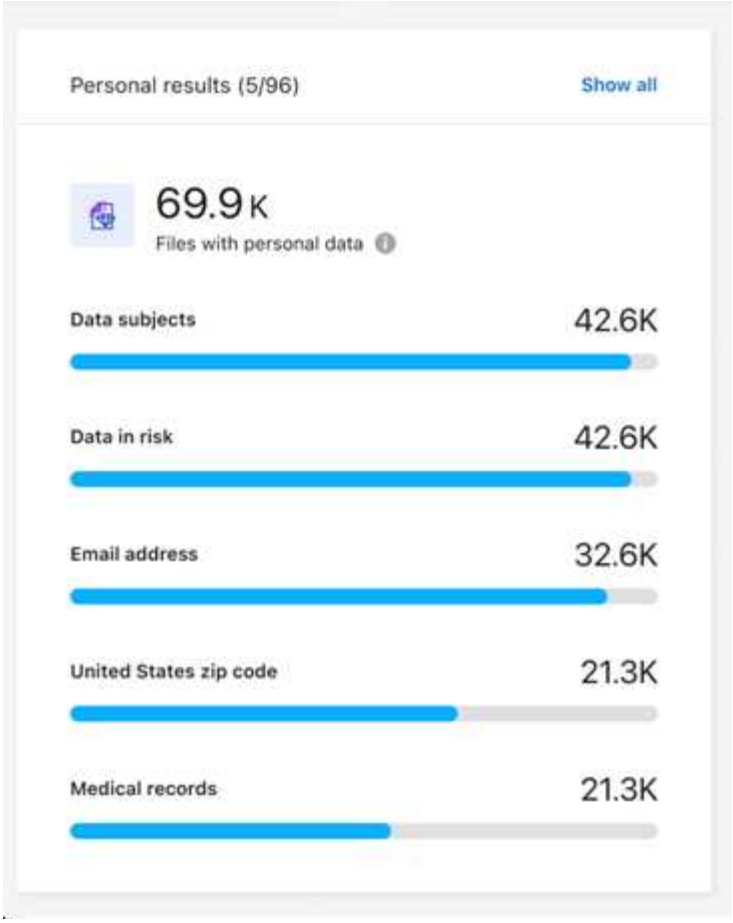
手順

1. データ分類メニューから、*コンプライアンス*タブを選択します。

2. すべての個人データの詳細を調査するには、個人データの割合の横にあるアイコンを選択します。



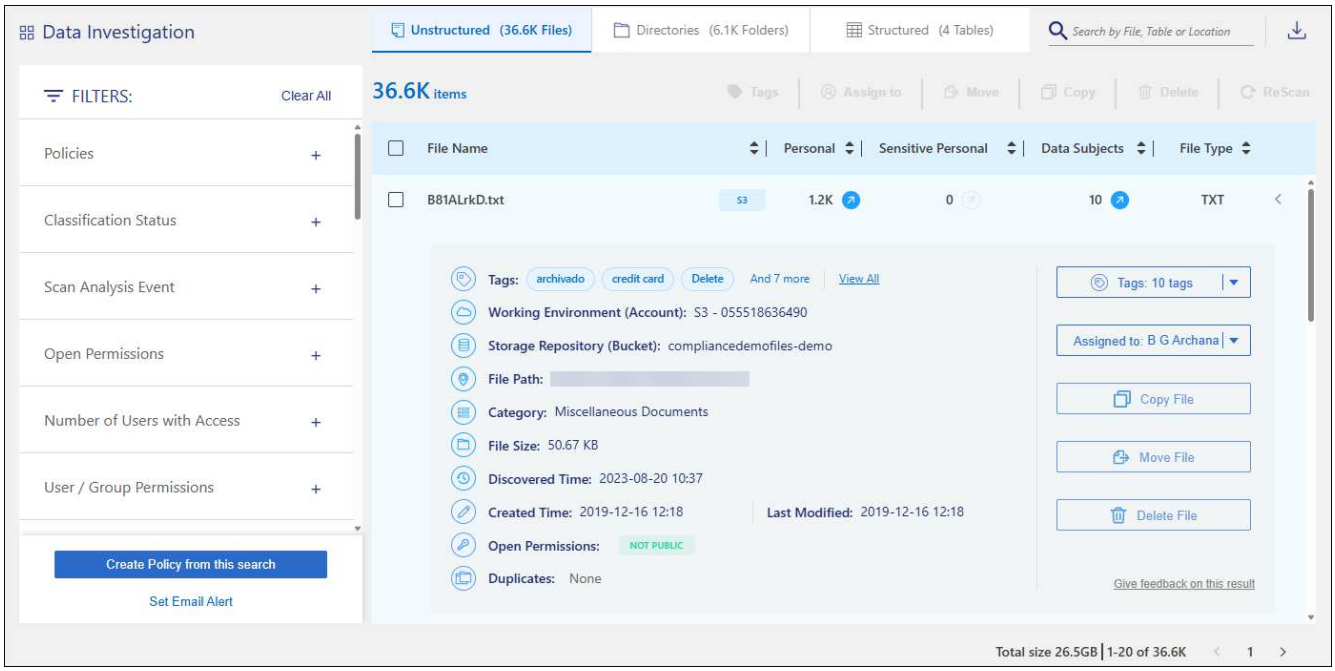
3. 特定の種類の個人データの詳細を調査するには、[すべて表示] を選択し、電子メール アドレスなどの特定の種類の個人データの [結果の調査] 矢印アイコンを選択します。

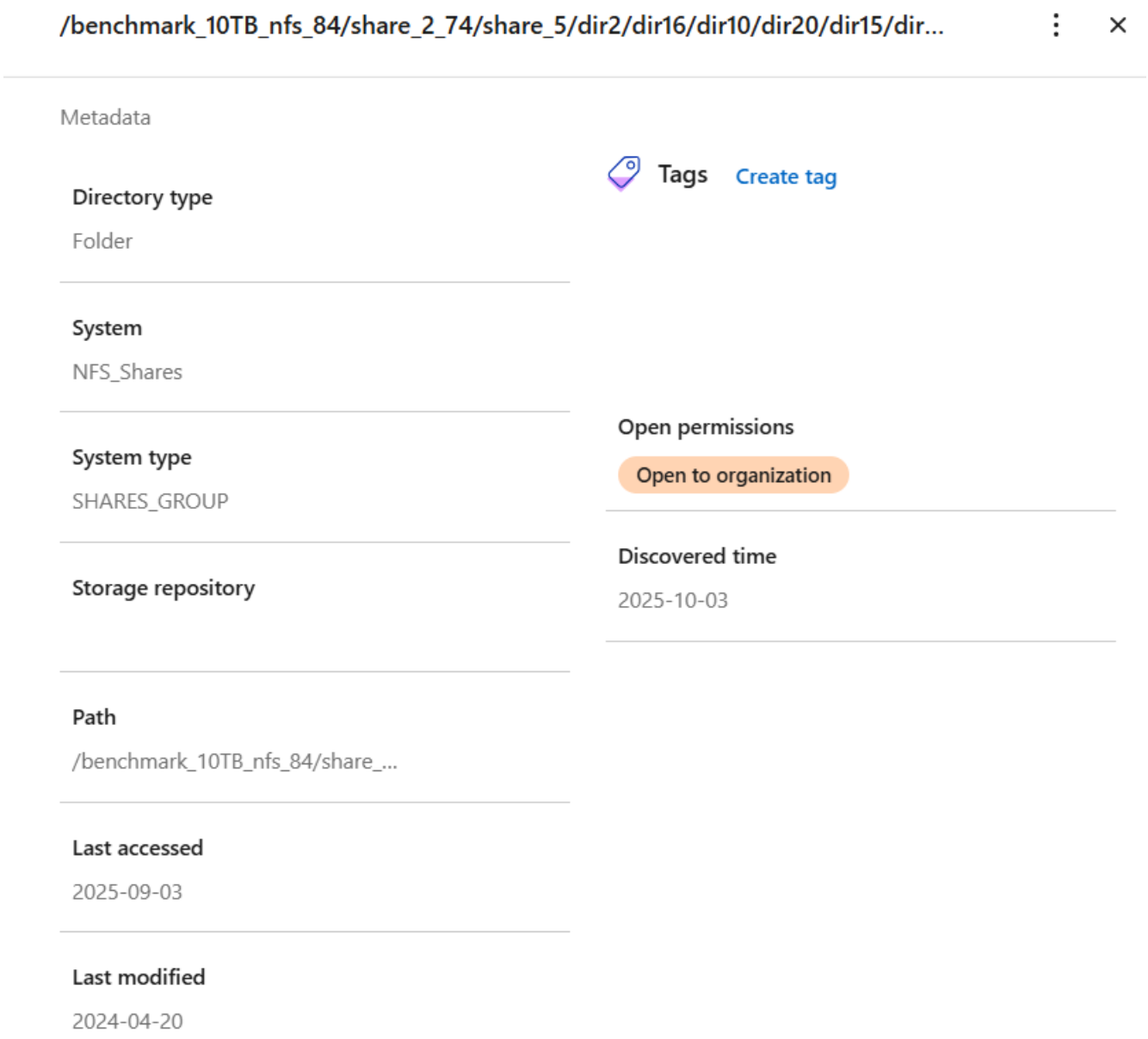


アイコンを選択できる個人ファイル ダイアログ ボックスのスクリーンショット。"]

4. 特定のファイルの詳細を検索、並べ替え、展開したり、*結果の調査*矢印を選択してマスクされた情報を表示したり、ファイル リストをダウンロードしたりして、データを調査します。

   次の画像は、ディレクトリ (共有およびフォルダー) で見つかった個人データを示しています。*構造化*タブでは、データベース内にある個人データを表示します。*非構造化*タブでは、ファイルレベルのデータを表示できます。


Data Investigation
Unstructured (36.6K Files)
Directories (6.1K Folders)
Structured (4 Tables)
Search by File, Table or Location
FILTERS:
Clear All
Policies
Classification Status
Scan Analysis Event
Open Permissions
Number of Users with Access
User / Group Permissions
Create Policy from this search
Set Email Alert
36.6K items
Tags
Assign to
Move
Copy
Delete
ReScan
File Name
Personal
Sensitive Personal
Data Subjects
File Type
B81ALrkD.txt
S3
1.2K
0
10
TXT
Tags: archivado credit card Delete And 7 more View All
Working Environment (Account): S3 - 055518636490
Storage Repository (Bucket): compliancedemofiles-demo
File Path:
Category: Miscellaneous Documents
File Size: 50.67 KB
Discovered Time: 2023-08-20 10:37
Created Time: 2019-12-16 12:18
Last Modified: 2019-12-16 12:18
Open Permissions: NOT PUBLIC
Duplicates: None
Tags: 10 tags
Assigned to: B G Archana
Copy File
Move File
Delete File
Give feedback on this result
Total size 26.5GB | 1-20 of 36.6K
1



機密性の高い個人データを含むファイルを表示する

データ分類は、プライバシー規制で定義されている特別な種類の機密個人情報を自動的に識別します。"GDPR第9条および第10条" 。たとえば、個人の健康、民族的起源、性的指向に関する情報などです。"全リストを見る" 。データ分類では、個々のファイル、ディレクトリ (共有とフォルダ) 内のファイル、およびデータベース テーブル内のこの種類の情報を識別します。

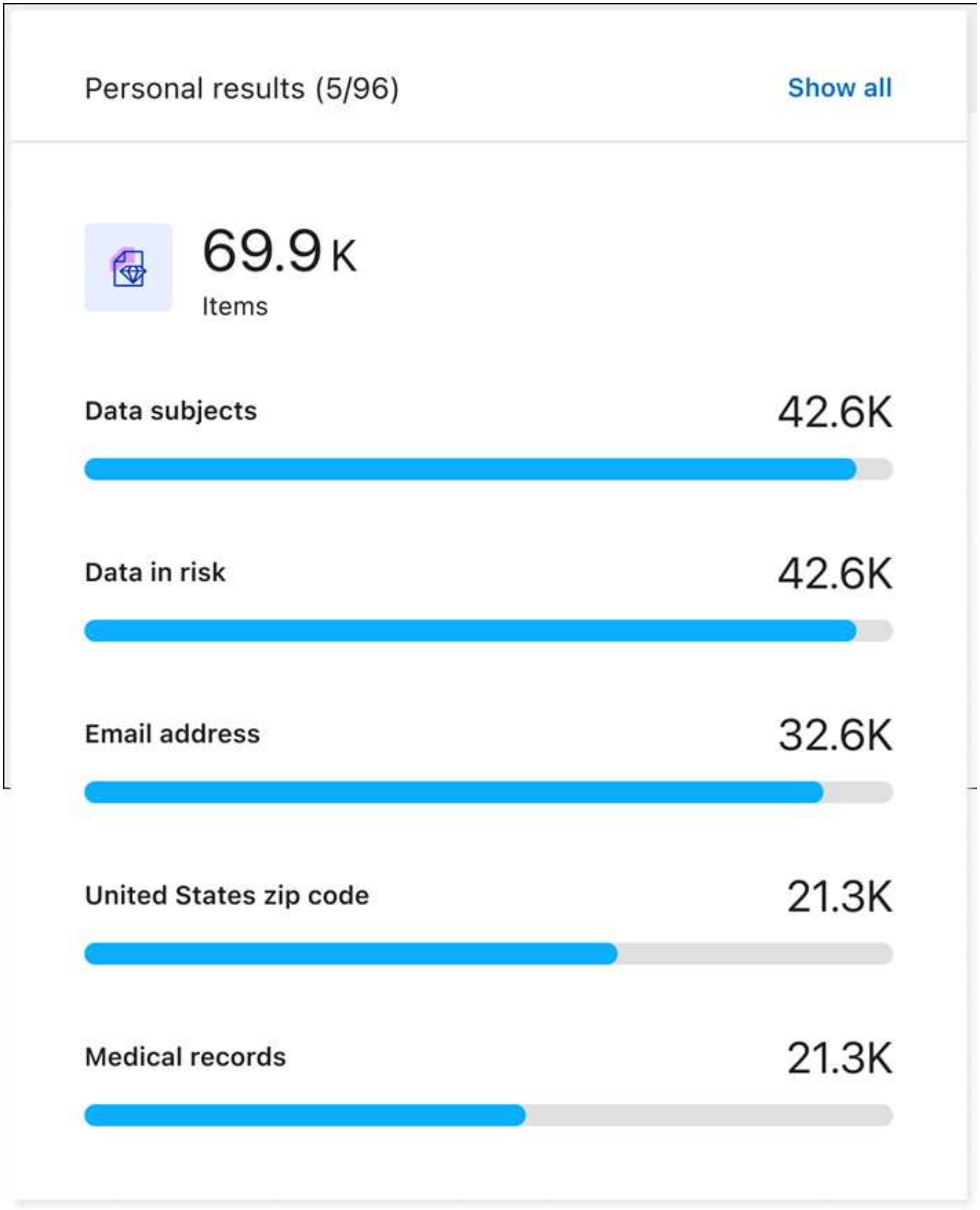
データ分類では、AI、自然言語処理 (NLP)、機械学習 (ML)、認知コンピューティング (CC) を使用して、スキャンしたコンテンツの意味を理解し、エンティティを抽出してそれに応じて分類します。

たとえば、GDPR の機密データ カテゴリの 1 つは民族的起源です。データ分類では、NLP 機能により、「George はメキシコ人です」(GDPR の第 9 条に指定されている機密データを示します) という文と、「George はメキシコ料理を食べています」という文を区別できます。

機密性の高い個人データをスキャンする場合は、英語のみがサポートされます。今後、さらに多くの言語のサポートが追加される予定です。

手順

1. データ分類メニューから、*コンプライアンス*を選択します。
2. すべての機密個人データの詳細を調査するには、[機密性の高い個人データの結果] カードを見つけて、[すべて表示] を選択します。



。

3. 特定の種類の機密個人データの詳細を調査するには、[すべて表示] を選択し、特定の種類の機密個人データの [結果の調査] 矢印アイコンを選択します。
4. 特定のファイルの詳細を検索、並べ替え、展開したり、「結果の調査」をクリックしてマスクされた情報を表示したり、ファイル リストをダウンロードしたりして、データを調査します。

NetApp Data Classificationにおけるプライベートデータのカテゴリ

NetApp Data Classification がボリュームやデータベース内で識別できるプライベート データには多くの種類があります。

データ分類では、次の 2 種類の個人データを識別します。

- 個人を特定できる情報（**PII**）
- 機密個人情報（**SPII**）

追加の国民 ID 番号や医療 ID 番号など、他のプライベート データの種類を識別するためにデータ分類が必要な場合は、アカウント マネージャーにお問い合わせください。

個人データの種類

ファイル内に含まれる個人データ、つまり個人を特定できる情報 (PII) は、一般的な個人データまたは国民識別子である場合があります。下の表の3番目の列は、データ分類が使用するかどうかを示しています。"近接検証"識別子に対する調査結果を検証するため。

これらの項目を認識できる言語は表に示されています。

タイプ	識別子	近接検証？	英語	ドイツ語	スペイン語	フランス語	日本語
全般	クレジットカード番号	はい	✓	✓	✓		✓
	データ主体	いいえ	✓	✓	✓		
	E メール アドレス	いいえ	✓	✓	✓		✓
	IBAN番号（国際銀行口座番号）	いいえ	✓	✓	✓		✓
	IP アドレス	いいえ	✓	✓	✓		✓
	パスワード	はい	✓	✓	✓		✓

タイプ	識別子	近接検証?	英語	ドイツ語	スペイン語	フランス語	日本語
国民識別子							

タイプ	識別子	近接検証?	英語	ドイツ語	スペイン語	フランス語	日本語

タイプ	識別子	近接検証?	英語	ドイツ語	スペイン語	フランス語	日本語

	（NRIC）						
	スロベニアID（EMSO）	はい	✓	✓	✓		
	南アフリカのID	はい	✓	✓	✓		
タイプ	識別子	近接検証	英語	ドイツ語	スペイン語	フランス語	日本語
	スペイン納税者番号	はい	✓	✓	✓		
	スウェーデンのID	はい	✓	✓	✓		
	英国ID（NINO）	はい	✓	✓	✓		
	米国カリフォルニア州運転免許証	はい	✓	✓	✓		
	米国インディアナ州運転免許証	はい	✓	✓	✓		
	米国ニューヨーク州運転免許証	はい	✓	✓	✓		
	米国テキサス州運転免許証	はい	✓	✓	✓		
	米国社会保障番号（SSN）	はい	✓	✓	✓		

機密個人データの種類

データ分類では、ファイル内の次の機密個人情報 (SPII) を見つけることができます。

以下の SPII は現在英語でのみ認識できます。

- 刑事手続きリファレンス: 自然人の刑事上の有罪判決および犯罪に関するデータ。
- 民族参照: 自然人の人種または民族的起源に関するデータ。
- 健康参考: 自然人の健康に関するデータ。
- **ICD-9-CM** 医療コード: 医療および健康業界で使用されるコード。
- **ICD-10-CM** 医療コード: 医療および健康業界で使用されるコード。
- 哲学的信念の参照: 自然人の哲学的信念に関するデータ。
- 政治的意見参照: 自然人の政治的意見に関するデータ。
- 宗教的信念の参照: 自然人の宗教的信念に関するデータ。
- 性生活または性的指向の参照: 自然人の性生活または性的指向に関するデータ。

カテゴリーの種類

データ分類では、データを次のように分類します。

これらのカテゴリのほとんどは、英語、ドイツ語、スペイン語で認識できます。

カテゴリ	タイプ	英語	ドイツ語	スペイン語
Finance	貸借対照表	✓	✓	✓
	発注書	✓	✓	✓
	請求書	✓	✓	✓
	四半期報告書	✓	✓	✓

カテゴリ	タイプ	英語	ドイツ語	スペイン語
人事	身元調査	✓		✓
	報酬プラン	✓	✓	✓
	従業員契約	✓		✓
	従業員レビュー	✓		✓
	健康	✓		✓
	履歴書	✓	✓	✓
法律上の	NDA	✓	✓	✓
	ベンダーと顧客の契約	✓	✓	✓
マーケティング	キャンペーン	✓	✓	✓
	会議	✓	✓	✓
オペレーション	監査報告書	✓	✓	✓
売り上げ	販売注文	✓	✓	
サービス	情報提供依頼	✓		✓
	提案依頼書	✓		✓
	種をまく	✓	✓	✓
	トレーニング	✓	✓	✓
サポート	苦情とチケット	✓	✓	✓

次のメタデータも同じサポートされている言語で分類および識別されます。

- アプリケーションデータ
- アーカイブファイル
- オーディオ
- データ分類ビジネスアプリケーションデータからのパンくずリスト
- CADファイル
- コード
- 破損した
- データベースとインデックスファイル
- デザインファイル
- 電子メールアプリケーションデータ
- 暗号化されたファイル（エントロピースコアの高いファイル）
- 実行可能ファイル
- 金融アプリケーションデータ
- 健康アプリケーションデータ

- イメージ
- Logs
- その他の文書
- その他のプレゼンテーション
- その他のスプレッドシート
- その他「不明」
- パスワード保護されたファイル
- 構造化データ
- ビデオ
- ゼロバイトファイル

ファイルの種類

データ分類では、すべてのファイルをスキャンしてカテゴリとメタデータの分析情報を取得し、ダッシュボードのファイル タイプ セクションにすべてのファイル タイプを表示します。データ分類が個人識別情報 (PII) を検出する場合、または DSAR 検索を実行する場合は、次のファイル形式のみがサポートされます。

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

発見された情報の正確性

NetApp は、データ分類によって識別される個人データおよび機密個人データの 100% の正確性を保証することはできません。常にデータを確認して情報を検証する必要があります。

当社のテストに基づき、以下の表はデータ分類が検出した情報の正確性を示しています。これを 精度 と 再現率 で分類します。

精度
: データ分類によって検出されたものが正しく識別された確率。たとえば、個人データの精度率が 90% の場合、個人情報が含まれていると識別された 10 個のファイルのうち 9 個に実際に個人情報が含まれていることを意味します。 10 個のファイルのうち 1 個は誤検出となります。

想起
: データ分類が必要なものを見つける確率。たとえば、個人データのリコール率が 70% の場合、データ分類では組織内の個人情報が実際に含まれているファイル 10 個のうち 7 個を識別できることになります。データ分類ではデータの 30% が失われ、ダッシュボードに表示されません。

当社は結果の精度を継続的に向上させています。これらの改善は、将来のデータ分類リリースで自動的に利用できるようになります。

タイプ	精度	想起
個人データ - 一般	90%～95%	60%～80%
個人データ - 国識別子	30%～60%	40%～60%
機密性の高い個人データ	80%～95%	20%～30%

タイプ	精度	想起
カテゴリ	90%～97%	60%～80%

NetApp Data Classificationでカスタム分類を作成する

NetApp Data Classification を使用すると、カスタム カテゴリまたは個人識別子を作成して、組織の規制およびコンプライアンス要件に固有のデータを識別できます。

データ分類では、カテゴリと個人識別子の 2 種類のカスタム分類子がサポートされています。カスタム カテゴリは、アップロードした一連のファイルに基づいて作成され、データ分類によって組織内の類似データを識別する AI モデルが作成されます (たとえば、健康研究会社では臨床分析カテゴリを作成する場合があります)。カスタム個人識別子は、キーワードリストまたは正規表現 (regex) を使用して作成され、コンプライアンスリスクをもたらす可能性のある組織固有の情報を識別します。

すべてのカスタム分類は、カスタム分類ダッシュボードで利用できます。

カスタム個人識別子を作成する

データ分類を使用すると、コンテキスト キーワードまたは正規表現を使用してカスタム個人識別子を作成し、組織固有のデータを識別できます。

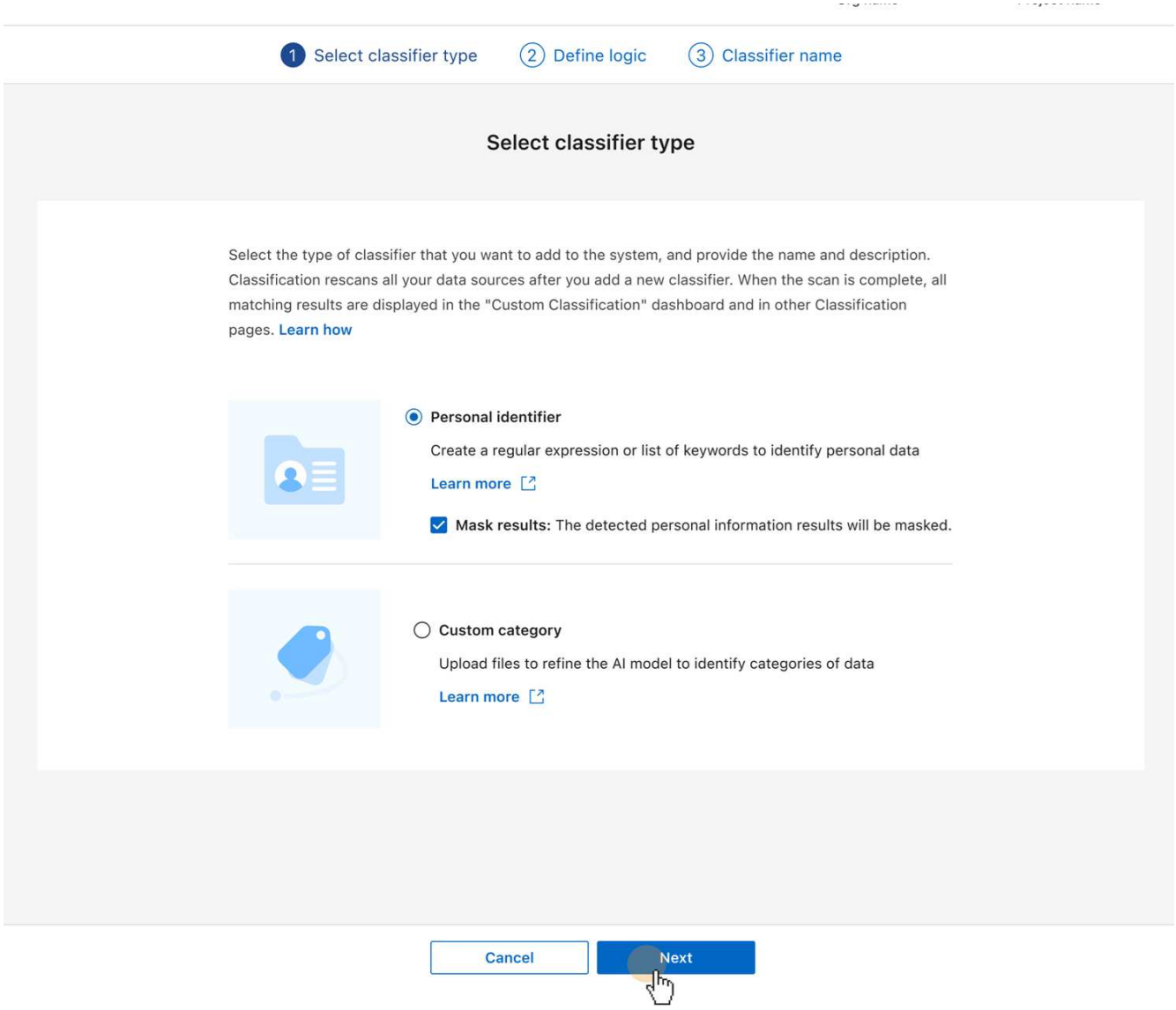
キーワードの要件

キーワードリストを使用して個人識別子を作成する場合、リストは次の要件を満たしている必要があります。

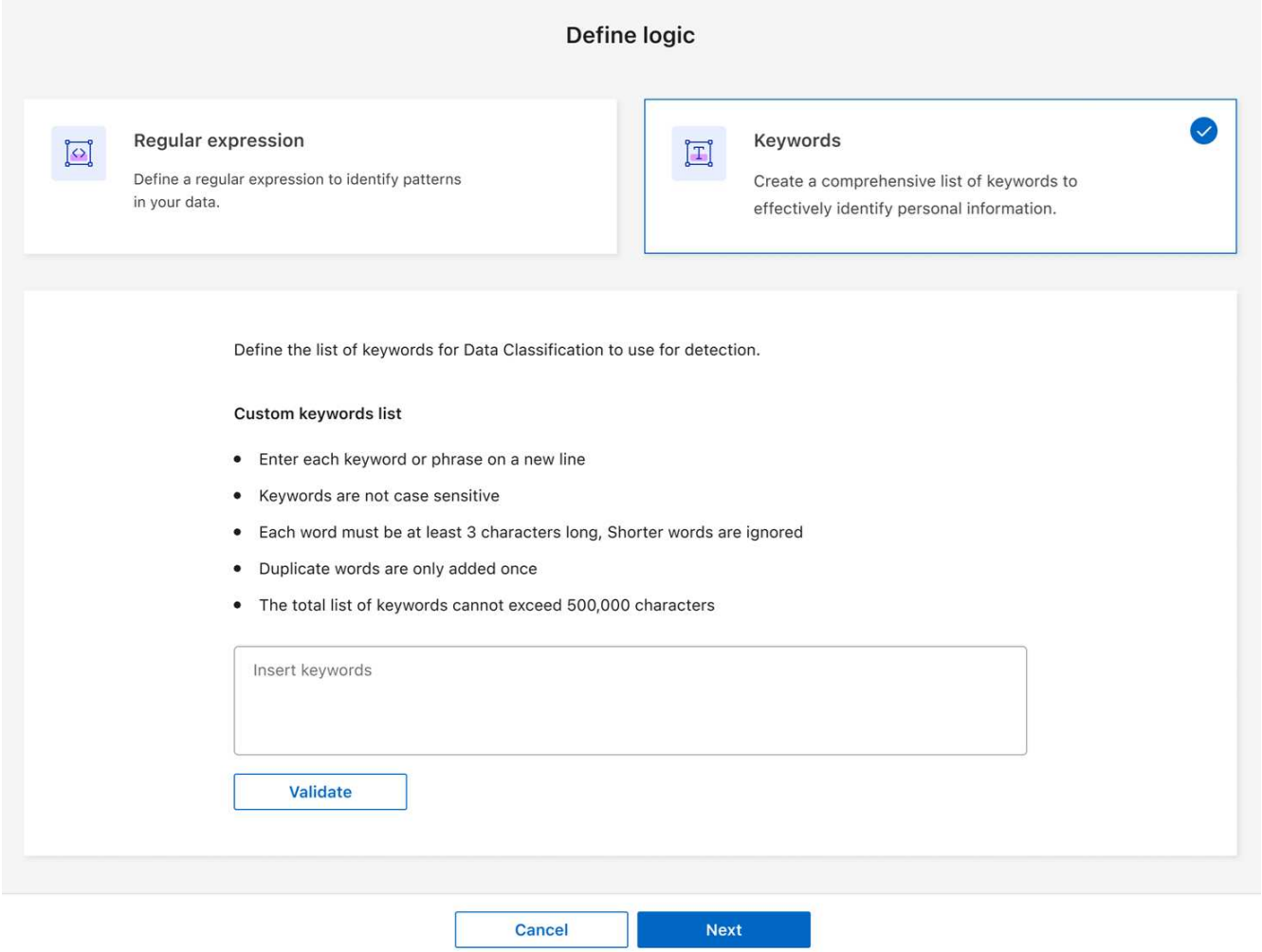
- キーワードの入力では大文字と小文字は区別されません。
- キーワードは 3 文字以上である必要があります。3 文字未満の単語は無視されます。
- 重複する単語は 1 回だけ追加されます。
- キーワードの合計リストは 500,000 文字を超えることはできません。リストには少なくとも 1 つのキーワードが含まれている必要があります。

手順

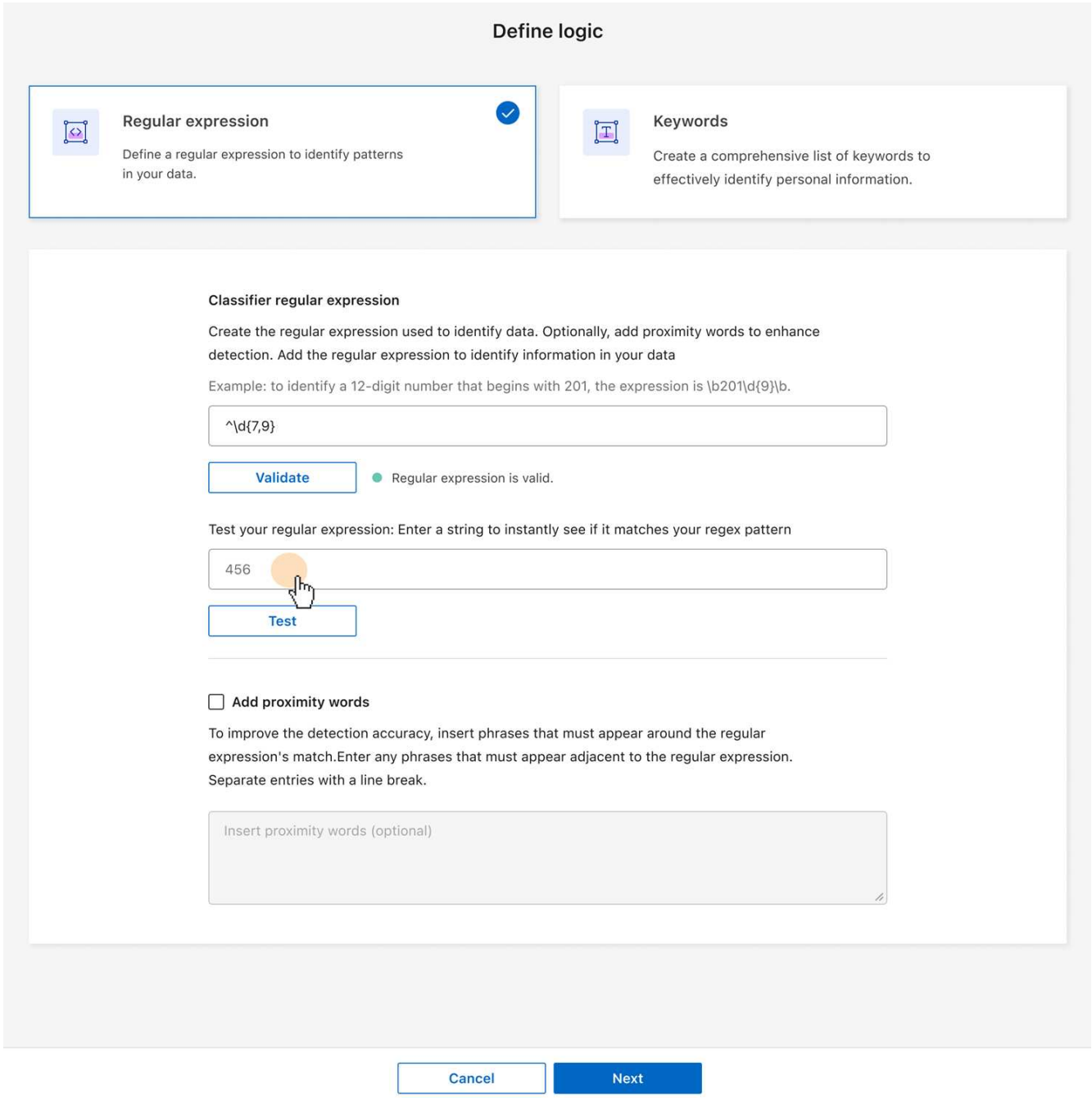
1. カスタム分類 タブを選択します。
2. カスタム分類子を作成するには、[**+** 新しい分類子] を選択します。
3. *個人識別子*を選択します。必要に応じて、結果をマスク を選択して、検出された個人データをマスクします。
4. 次へを選択します。



5. キーワードを含む分類子を追加するには、キーワード を選択します。キーワードのリストを入力します。各エントリは別々の行に入力します。キーワードが要件に準拠していることを確認します。



分類子を正規表現として追加するには、正規表現 を選択し、データの特定の情報を検出するためのパターンを追加します。入力した構文を確認するには、[検証] を選択します。



a. 必要に応じて、正規表現パターンに一致するサンプル文字列を入力し、[テスト] を選択して確認します。

b. 必要に応じて、近接語を追加します。近接単語を追加すると、データ分類では、近接単語が一致する文字列に隣接している場合にのみ正規表現パターンにフラグが付けられます。

6. 次へを選択します。

7. ダッシュボードでカスタム カテゴリを識別するには、分類子名 と 概要 を入力します。

8. **[保存]** を選択して、カスタム個人識別子を作成します。

カスタム個人識別子を作成すると、その結果は次回のスケジュールされたスキャンでキャプチャされます。より早く結果を取得するには、オンデマンドスキャンを実行します。結果を見るには、 コンプライアンスレポートを生成する。

カスタムカテゴリを作成する

カスタム カテゴリを使用すると、組織固有のデータを分類できます。カスタム カテゴリは、アップロードしたテキスト ファイルに基づいて作成されます。データ分類では、このテキスト ファイルから AI モデルを作成し、他のファイル内の同様の情報を識別します。

トレーニングデータ要件

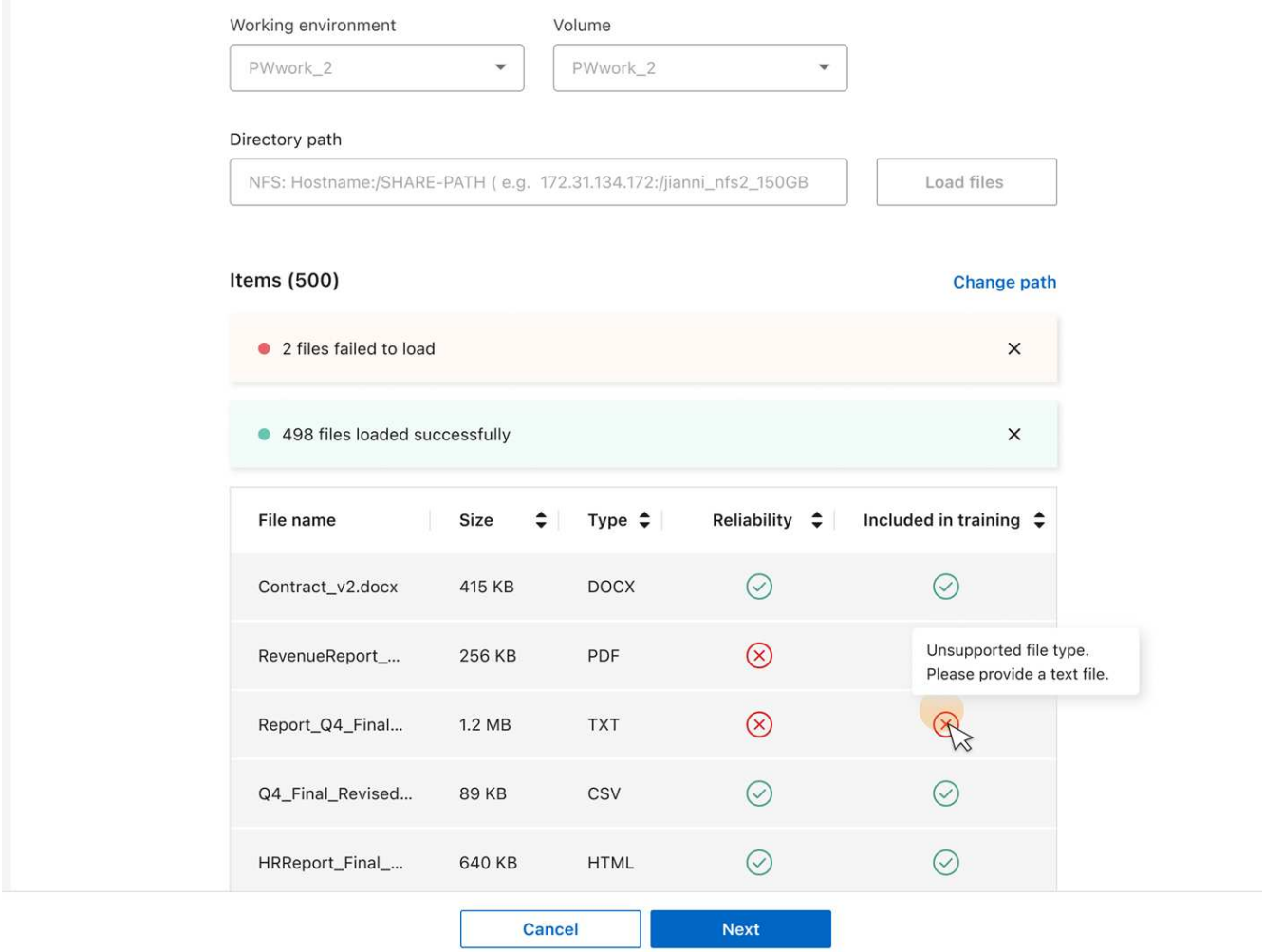
- トレーニング データセットには少なくとも 25 個のファイルが含まれている必要があります。最大ファイル数は 1,000 です。
- すべてのファイルは、指定したファイル パスに直接配置する必要があります。
- すべてのファイルは 100 バイトより大きくなければなりません。
- データ分類トレーニング データは、CSV、DOCX、DOC、GZ、JSON、PDF、PPTX、TXT、RTT、XLS、または XLSX のいずれかのファイル タイプである必要があります。サポートされているすべてのファイルタイプを組み合わせてアップロードできます。

手順

1. NetApp Data Classificationで、*カスタム分類*を選択します。
2. *+ 新しい分類子*を選択します。
3. 分類子の種類として「カスタム カテゴリ」を選択し、「次へ」をクリックします。
4. テキストベースのファイルのコレクションを使用して、カスタム カテゴリのロジックを定義します。作業アドレス*の **IP** アドレスを入力し、ドロップダウン メニューから *ボリューム を選択します。

   トレーニング データが含まれているディレクトリの ディレクトリ パス を入力します。

5. ファイルのチェックを実行するには、データ分類の [ファイルの読み込み] を選択します。ファイルの概要を確認することができます。そこには、ファイル名、サイズ、タイプ、およびファイルがトレーニングに適していると判断されたかどうかのメモが一覧表示されます。



a. ファイル パスを変更したり、ファイルを再アップロードしたりするには、[パスの変更] を選択し、データを入力してファイルを再度ロードします。

6. アップロードしたファイルに問題がなければ、[次へ] を選択します。
7. ダッシュボードでカスタム カテゴリを識別するには、分類子名 と 概要 を入力します。
8. **[保存]** を選択してカスタム カテゴリを作成します。

結果

カスタム カテゴリを作成すると、その結果は次回のスケジュールされたスキャンでキャプチャされます。より早く結果を取得するには、手動でスキャンを開始します。

カスタム分類子を編集する

個人識別子を作成した後で、そのロジックを変更できます。個人識別子のタイプまたはロジック タイプを変更することはできません。たとえば、カスタム カテゴリをカスタム個人識別子に変更することはできません。また、キーワードベースのカスタム識別子を正規表現ベースのカスタム識別子に変更することもできません。

手順

1. NetApp Data Classificationで、*カスタム分類*を選択します。
2. 削除したい分類子を特定し、アクションメニューを選択します ... 列の末尾にあります。

3. ロジックの編集を選択します。
4. キーワードを変更する場合は、適切なキーワードを追加、削除、または編集します。正規表現を変更する場合は、新しい正規表現を入力して検証します。必要に応じて、近接キーワードを追加します。
5. 変更を適用するには、[保存] を選択します。

カスタム分類子を削除する

1. NetApp Data Classificationで、*カスタム分類*を選択します。
2. 削除したい分類子を識別し、アクションメニューを選択します ... 列の末尾にあります。
3. 分類子を削除 を選択します。

次のステップ

- コンプライアンスレポートを生成する

NetApp Data Classificationを使用して組織内に保存されているデータを調査します

データ調査ダッシュボードには、データに関するファイルおよびディレクトリ レベルの分析情報が表示され、結果を並べ替えたりフィルター処理したりできます。データ調査ページでは、ファイルとディレクトリのメタデータと権限に関する詳細情報が表示され、重複したファイルも識別されます。ファイル、ディレクトリ、データベース レベルの分析情報を活用することで、組織のコンプライアンスを向上させ、ストレージ スペースを節約するためのアクションを実行できます。データ調査ページでは、ファイルの移動、コピー、削除もサポートされています。

調査ページから洞察を得るには、データ ソースに対して完全な分類スキャンを実行する必要があります。マッピングのみのスキャンが行われたデータ ソースでは、ファイル レベルの詳細は表示されません。

データ調査構造

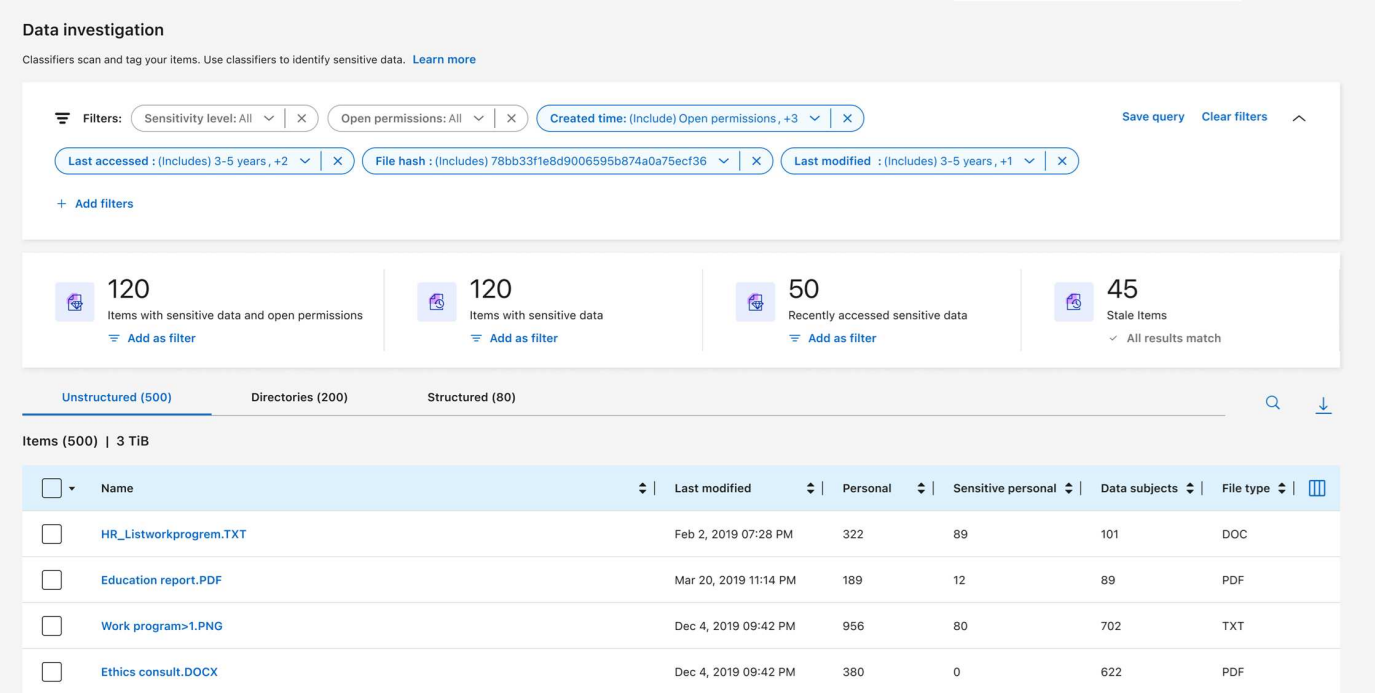
データ調査ページでは、データが 3 つのタブに分類されます。

- 非構造化データ: ファイルデータ
- ディレクトリ: フォルダとファイル共有
- 構造化: データベース

データフィルター

データ調査ページには、必要なものだけを見つけるためにデータを並べ替えるためのさまざまなフィルターが用意されています。複数のフィルターを組み合わせて使うこともできます。

フィルターを追加するには、[フィルターを追加] ボタンを選択します。



フィルターの感度とコンテンツ

次のフィルターを使用して、データに含まれる機密情報の量を確認します。

フィルタ	詳細
カテゴリ	を選択する"カテゴリーの種類"。
感度レベル	機密レベルを選択します: 個人、機密個人、または非機密。
識別子の数	ファイルごとに検出された機密識別子の範囲を選択します。個人データおよび機密性の高い個人データが含まれます。ディレクトリでフィルタリングする場合、データ分類は各フォルダー (およびサブフォルダー) 内のすべてのファイルからの一致を合計します。注: 2023 年 12 月 (バージョン 1.26.6) のリリースでは、ディレクトリ別に個人識別情報 (PII) データの数を計算するオプションが削除されました。
個人データ	を選択する"個人データの種類"。
機密性の高い個人データ	を選択する"機密個人データの種類"。
データ主体	データ主体のフルネームまたは既知の識別子を入力します。"データ主体についての詳細はこちらをご覧ください" 。

ユーザー所有者とユーザー権限をフィルタリング

次のフィルターを使用して、ファイルの所有者とデータへのアクセス許可を表示します。

フィルタ	詳細
オープン権限	データ内およびフォルダー/共有内の権限の種類を選択します。
ユーザー/グループの権限	1 つまたは複数のユーザー名またはグループ名を選択するか、名前の一部を入力します。
ファイル所有者	ファイルの所有者名を入力します。

フィルタ	詳細
アクセス権を持つユーザーの数	1 つまたは複数のカテゴリ範囲を選択して、特定の数のユーザーに公開されているファイルとフォルダーを表示します。

時系列でフィルタリング

時間基準に基づいてデータを表示するには、次のフィルターを使用します。

フィルタ	詳細
作成時刻	ファイルが作成された時間の範囲を選択します。検索結果をさらに絞り込むために、カスタムの時間範囲を指定することもできます。
発見された時間	データ分類がファイルを検出した時間範囲を選択します。検索結果をさらに絞り込むために、カスタムの時間範囲を指定することもできます。
最終変更日時	ファイルが最後に変更された時間の範囲を選択します。検索結果をさらに絞り込むために、カスタムの時間範囲を指定することもできます。
最終アクセス日時	ファイルまたはディレクトリ* が最後にアクセスされた時間の範囲を選択します。検索結果をさらに絞り込むために、カスタムの時間範囲を指定することもできます。データ分類がスキャンするファイルの種類の場合、これはデータ分類がファイルを最後にスキャンした時刻です。

{アスタリスク} ディレクトリの最終アクセス時刻は、NFS または CIFS 共有でのみ使用できます。

メタデータをフィルタリング

場所、サイズ、ディレクトリまたはファイルの種類に基づいてデータを表示するには、次のフィルターを使用します。

フィルタ	詳細
ファイルパス	クエリに含めるか除外する部分パスまたは完全パスを最大 20 個入力します。含めるパスと除外するパスの両方を入力すると、データ分類はまず含めるパス内のすべてのファイルを見つけ、次に除外するパスからファイルを削除して、結果を表示します。このフィルターで「*」を使用しても効果はなく、特定のフォルダーをスキャンから除外することはできないことに注意してください。構成された共有の下にあるすべてのディレクトリとファイルがスキャンされます。
ディレクトリタイプ	ディレクトリの種類として「共有」または「フォルダー」を選択します。
ファイル タイプ	を選択する"ファイルの種類"。
ファイル サイズ	ファイルサイズの範囲を選択します。
ファイルハッシュ	名前が異なっていても、特定のファイルを見つけるには、ファイルのハッシュを入力します。

フィルター収納タイプ

ストレージ タイプ別にデータを表示するには、次のフィルターを使用します。

フィルタ	詳細
システムタイプ	システムの種類を選択します。
システム環境名	特定のシステムを選択します。
ストレージリポジトリ	ボリュームやスキーマなどのストレージ リポジトリを選択します。

フィルタークエリ

保存されたクエリ別にデータを表示するには、次のフィルターを使用します。

フィルタ	詳細
保存されたクエリ	保存したクエリを 1 つまたは複数選択します。に行く"保存されたクエリタブ"既存の保存済みクエリのリストを表示し、新しいクエリを作成します。
タグ	選択"タグ"ファイルに割り当てられているもの。

フィルター分析ステータス

次のフィルターを使用して、データ分類スキャンのステータス別にデータを表示します。

フィルタ	詳細
分析ステータス	最初のスキャンが保留中、スキャンが完了、再スキャンが保留中、またはスキャンに失敗したファイルのリストを表示するには、オプションを選択します。
スキャン分析イベント	データ分類が最終アクセス時間を元に戻すことができなかったために分類されなかったファイルを表示するか、データ分類が最終アクセス時間を元に戻すことができなかったにもかかわらず分類されたファイルを表示するかを選択します。

"「最終アクセス時刻」のタイムスタンプの詳細を見る"スキャン分析イベントを使用してフィルタリングするときに調査ページに表示される項目の詳細については、こちらをご覧ください。

重複データによるフィルタリング

ストレージ内に重複しているファイルを表示するには、次のフィルターを使用します。

フィルタ	詳細
重複	リポジトリ内でファイルを複製するかどうかを選択します。

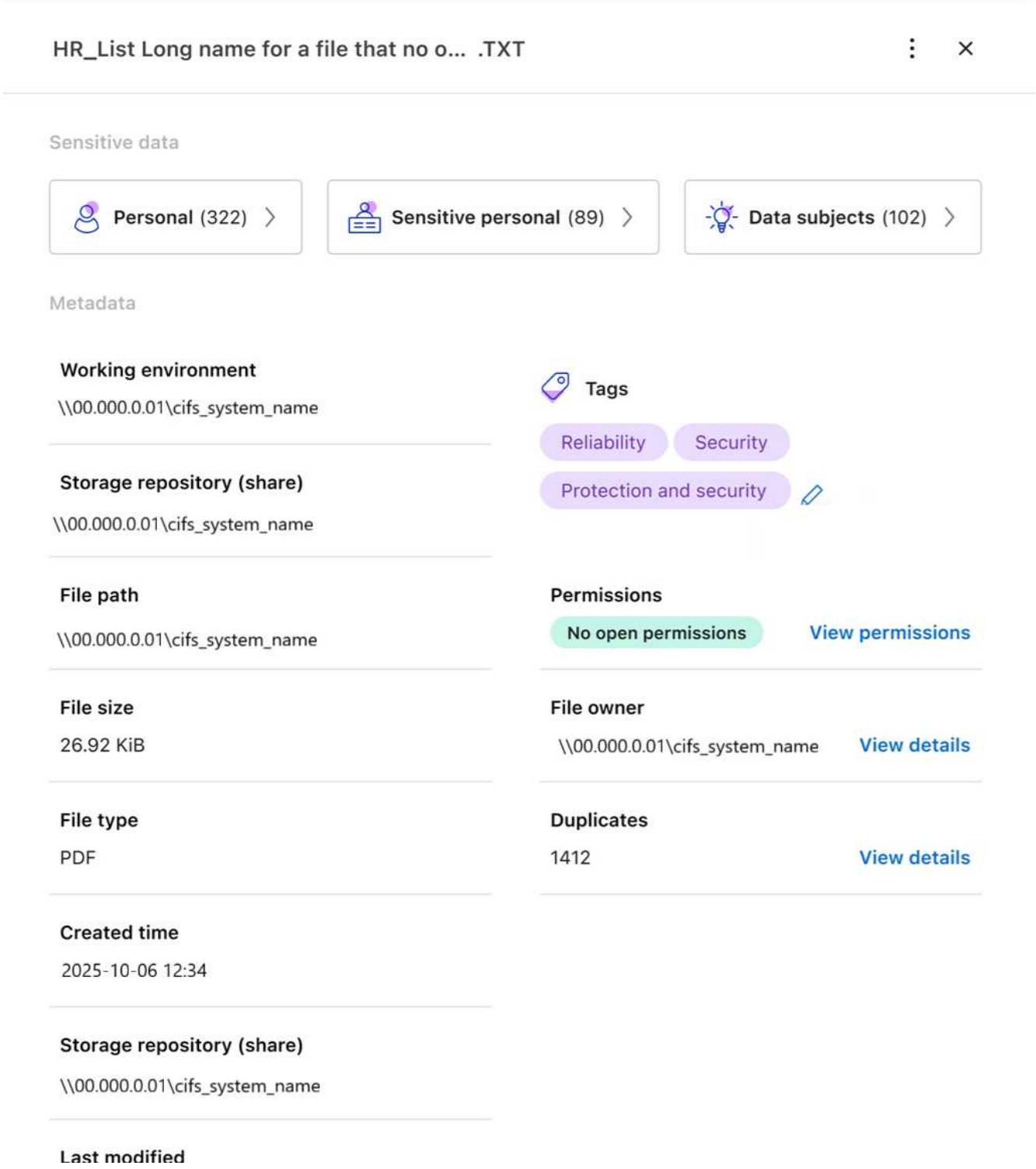
ファイルのメタデータを表示

メタデータには、ファイルが存在するシステムとボリュームが表示されるだけでなく、ファイルの権限、ファイルの所有者、このファイルの重複があるかどうかなど、さらに多くの情報が表示されます。この情報は、"保存したクエリを作成する"データをフィルタリングするために使用できるすべての情報を確認できるためです。

情報の可用性はデータ ソースによって異なります。たとえば、データベース ファイルのボリューム名とアクセス許可は共有されません。

手順

1. データ分類メニューから、*調査*を選択します。
2. 右側のデータ調査リストで、下向き矢印を選択します。 ファイルのメタデータを表示するには、任意のファイルの右側にある をクリックします。



3. オプションで、[タグを作成] ボタンを使用してファイルにタグを作成または追加できます。ドロップダウン メニューから既存のタグを選択するか、[+ 追加] ボタンを使用して新しいタグを追加します。タグを使用してデータをフィルタリングできます。

ファイルとディレクトリのユーザー権限を表示する

ファイルまたはディレクトリへのアクセス権を持つすべてのユーザーまたはグループと、それらの権限の種類のリストを表示するには、[すべての権限を表示] を選択します。このオプションは、CIFS 共有内のデータに対してのみ使用できます。

ユーザー名やグループ名の代わりにセキュリティ識別子 (SID) を使用する場合は、Active Directory をデータ分類に統合する必要があります。詳細については、以下を参照してください。 "データ分類にActive Directoryを追加する" 。

手順

1. データ分類メニューから、*調査*を選択します。
2. 右側のデータ調査リストで、下向き矢印を選択します。 ファイルのメタデータを表示するには、任意のファイルの右側にある をクリックします。
3. ファイルまたはディレクトリへのアクセス権を持つすべてのユーザーまたはグループのリストと、それらの権限の種類を表示するには、[開く権限] フィールドで [すべての権限を表示] を選択します。

   データ分類では、リストに最大 100 人のユーザーが表示されます。

4. 下向き矢印を選択 任意のグループのボタンをクリックすると、そのグループに属しているユーザーのリストが表示されます。

   グループの 1 レベルを展開すると、グループに属しているユーザーを表示できます。

5. ユーザーまたはグループの名前を選択して調査ページを更新し、そのユーザーまたはグループがアクセスできるすべてのファイルとディレクトリを表示できるようにします。

ストレージシステム内の重複ファイルをチェックする

ストレージ システムに重複したファイルが保存されているかどうかを確認できます。これは、ストレージスペースを節約できる領域を特定する場合に便利です。また、特定の権限や機密情報を持つ特定のファイルが、ストレージ システム内で不必要に重複しないようにすることも重要です。

データ分類では、次の場合にすべてのファイル (データベースを除く) の重複を比較します。

- 1 MB以上
- または個人情報や機密性の高い個人情報が含まれている

データ分類では、ハッシュ テクノロジを使用して重複ファイルを判別します。いずれかのファイルに別のファイルと同じハッシュ コードがある場合、ファイル名が異なっていてもファイルは完全に重複しています。

手順

1. データ分類メニューから、*調査*を選択します。
2. フィルター ペインで、「ファイル サイズ」と「重複」(「重複あり」) を選択して、環境内で重複している特定のサイズ範囲のファイルを確認します。
3. オプションで、重複ファイルのリストをダウンロードしてストレージ管理者に送信し、削除できるファイルがあるかどうかを管理者が判断できるようにします。

4. 必要に応じて、重複ファイルを削除、タグ付け、または移動することもできます。アクションを実行するファイルを選択し、適切なアクションを選択します。

特定のファイルが重複しているかどうかを確認する

1 つのファイルに重複があるかどうかを確認できます。

手順

1. データ分類メニューから、*調査*を選択します。
2. データ調査リストで、ファイルのメタデータを表示するには、任意のファイルの右側にある をクリックします。

   ファイルに重複が存在する場合、この情報は [重複] フィールドの横に表示されます。

3. 重複ファイルのリストとその保存場所を表示するには、[詳細の表示] を選択します。
4. 次のページで「重複を表示」を選択し、調査ページでファイルを表示します。
5. 必要に応じて、重複ファイルを削除、タグ付け、または移動することもできます。アクションを実行するファイルを選択し、適切なアクションを選択します。

> このページで提供されている「ファイル ハッシュ」値を使用して、調査ページに直接入力し、いつでも特定の重複ファイルを検索できます。また、保存したクエリで使用することもできます。

レポートをダウンロードする

フィルタリングされた結果を CSV または JSON 形式でダウンロードできます。

データ分類がファイル (非構造化データ)、ディレクトリ (フォルダーとファイル共有)、およびデータベース (構造化データ) をスキャンしている場合、最大 3 つのレポート ファイルをダウンロードできます。

ファイルは、固定数の行またはレコードを持つファイルに分割されます。

- JSON: レポートあたり10万件のレコード。生成には約5分かかります。
- CSV: レポートあたり20万件のレコード、生成に約4分かかります

> このブラウザで表示するには、CSV ファイルのバージョンをダウンロードできます。このバージョンは 10,000 件のレコードに制限されています。

ダウンロード可能なレポートに含まれるもの

*非構造化ファイル データ レポート*には、ファイルに関する次の情報が含まれます。

- ファイル名
- 場所の種類
- システム名
- ストレージリポジトリ（ボリューム、バケット、共有など）
- リポジトリの種類

- ファイル パス
- ファイル タイプ
- ファイルサイズ（MB）
- 作成時間
- 最終更新日
- 最終アクセス
- ファイルの所有者
  - ファイル所有者データには、Active Directory が構成されている場合のアカウント名、SAM アカウント名、および電子メール アドレスが含まれます。
- カテゴリ
- 個人情報
- 機密個人情報
- オープン権限
- スキャン分析エラー
- 削除検出日

  削除検出日は、ファイルが削除または移動された日付を識別します。これにより、機密ファイルが移動された時期を識別できるようになります。削除されたファイルは、ダッシュボードまたは調査ページに表示されるファイル数には含まれません。ファイルは CSV レポートにのみ表示されます。

*非構造化ディレクトリ データ レポート*には、フォルダーとファイル共有に関する次の情報が含まれます。

- システムタイプ
- システム名
- ディレクトリ名
- ストレージリポジトリ（フォルダやファイル共有など）
- ディレクトリ所有者
- 作成時間
- 発見された時間
- 最終更新日
- 最終アクセス
- オープン権限
- ディレクトリタイプ

*構造化データ レポート*には、データベース テーブルに関する次の情報が含まれます。

- DBテーブル名
- 場所の種類
- システム名

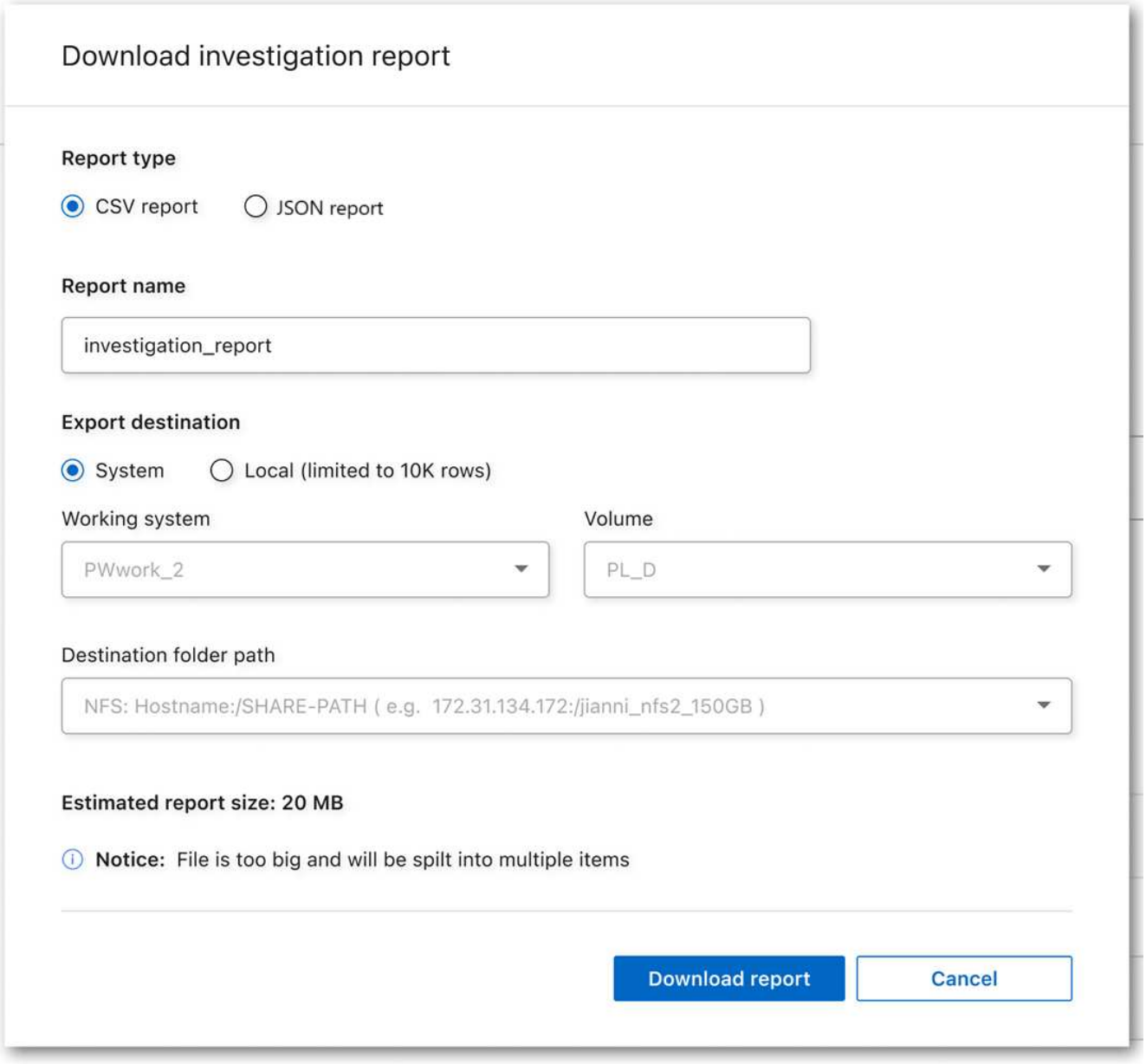
- ストレージリポジトリ（スキーマなど）
- 列数
- 行数
- 個人情報
- 機密個人情報

レポートを生成する手順

1. データ調査ページから、ページの右上にあるボタンをクリックします。
2. レポートタイプ（CSV または JSON）を選択します。
3. レポート名 を入力します。
4. 完全なレポートをダウンロードするには、[システム] を選択し、それぞれのドロップダウン メニューから [システム] と [ボリューム] を選択します。 宛先フォルダーのパス を指定します。

   ブラウザでレポートをダウンロードするには、[ローカル] を選択します。このオプションでは、レポートが最初の 10,000 行に制限され、**CSV** 形式に制限されることに注意してください。 ローカル を選択した場合は、他のフィールドを入力する必要はありません。

5. レポートのダウンロードを選択します。



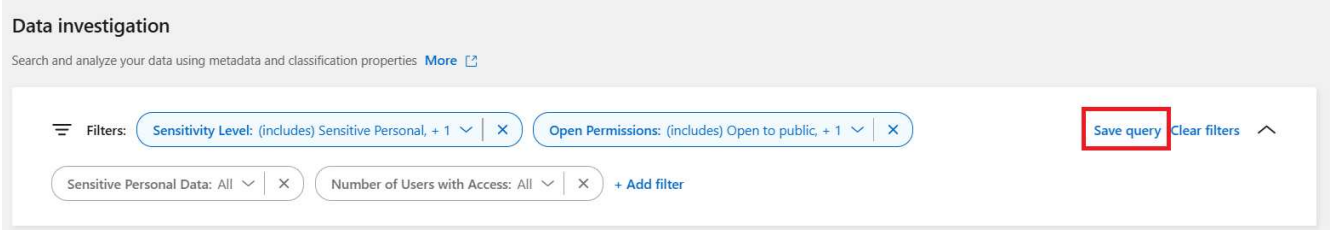
結果

レポートをダウンロード中であることを示すメッセージがダイアログに表示されます。

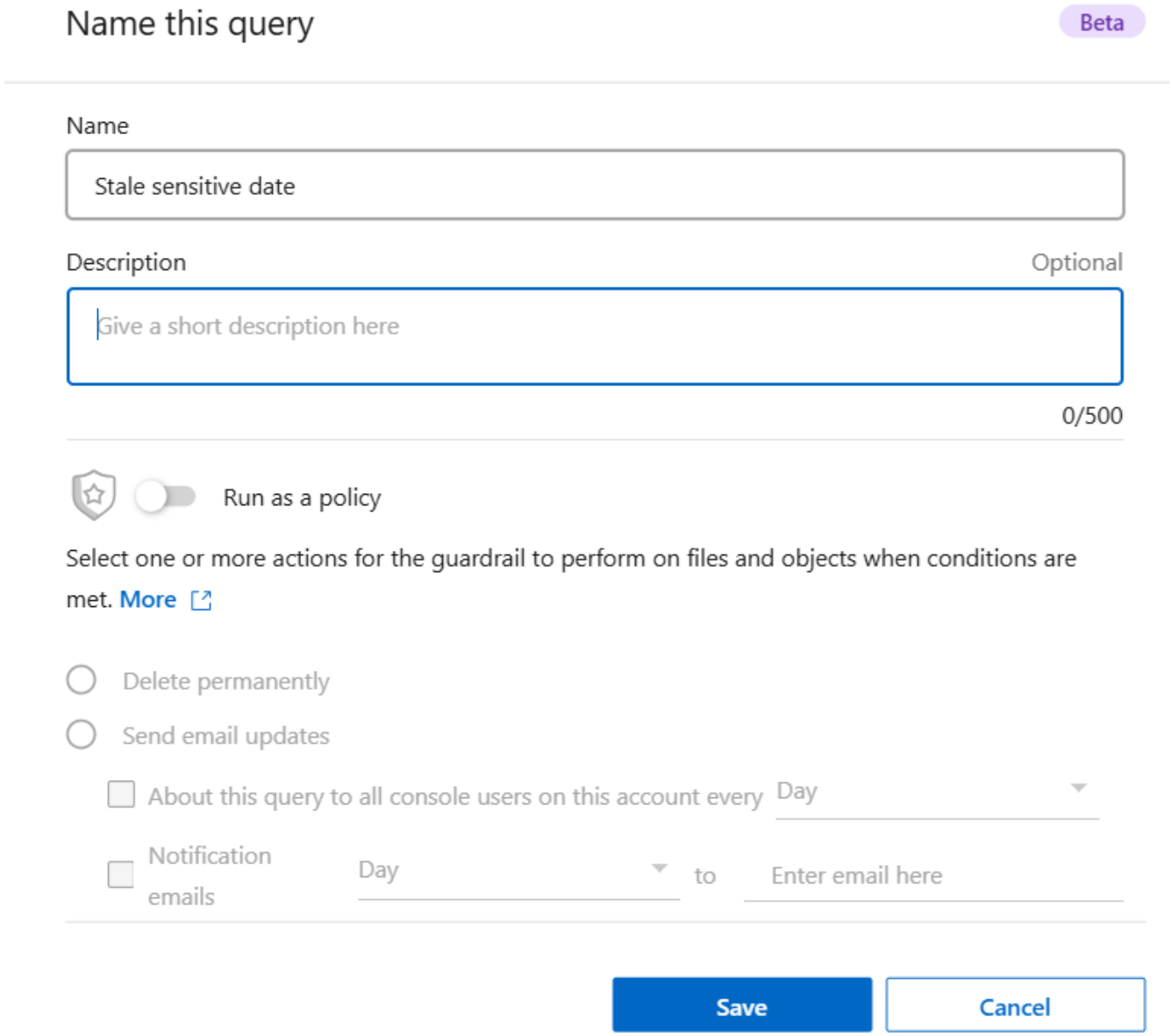
選択したフィルターに基づいて保存されたクエリを作成する

手順

1. 「調査」タブで、使用するフィルターを選択して検索を定義します。見る"調査ページでのデータのフィルタリング"詳細については。
2. すべてのフィルター特性を好みに合わせて設定したら、「クエリを保存」を選択します。



3. 保存したクエリに名前を付け、説明を追加します。名前は一意である必要があります。
4. オプションでクエリをポリシーとして保存できます。
   a. クエリをポリシーとして保存するには、[ポリシーとして実行] トグルを切り替えます。
   b. 完全に削除 または 電子メールで更新を送信 を選択します。電子メールによる更新を選択した場合は、クエリ結果を毎日、毎週、または毎月、すべてのコンソール ユーザーに電子メールで送信できます。あるいは、同じ頻度で特定の電子メール アドレスに通知を送信することもできます。
5. *保存*を選択します。



検索またはポリシーを作成したら、[保存したクエリ] タブで表示できます。

結果が「保存されたクエリ」ページに表示されるまで、最大 15 分かかる場合があります。

NetApp Data Classificationで保存したクエリを管理する

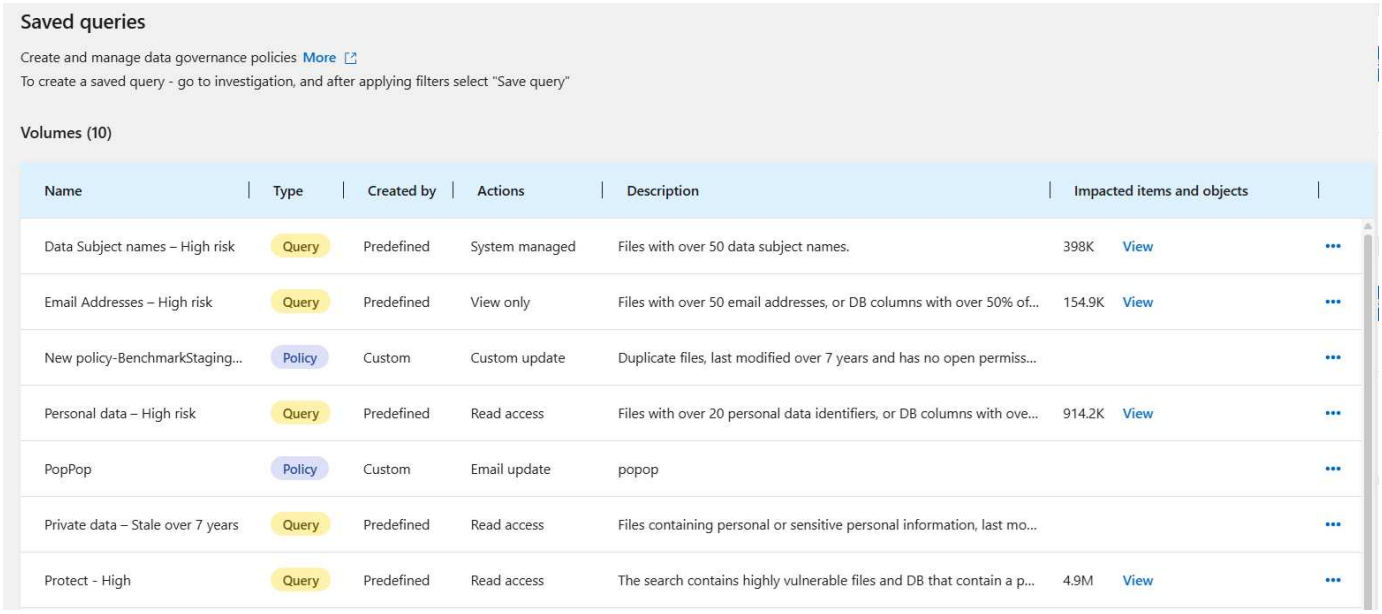
NetaApp Data Classification は、検索クエリの保存をサポートしています。保存したクエリを使用すると、カスタム フィルターを作成して、データ調査ページで頻繁に実行されるクエリを並べ替えることができます。データ分類には、一般的なリクエストに基づいて事前定義された保存済みクエリも含まれます。

コンプライアンス ダッシュボードの 保存済みクエリ タブには、データ分類のこのインスタンスで使用できる

すべての定義済みクエリとカスタム保存済みクエリが一覧表示されます。

保存されたクエリはポリシーとして保存することもできます。クエリはデータをフィルタリングしますが、ポリシーを使用するとデータに基づいて操作を行うことができます。ポリシーを使用すると、検出されたデータを削除したり、検出されたデータに関する電子メールの更新を送信したりできます。

保存されたクエリは、調査ページのフィルターのリストにも表示されます。



調査ページで保存したクエリの結果を表示する

調査ページに保存したクエリの結果を表示するには、特定の検索のボタンをクリックし、[結果の調査] を選択します。

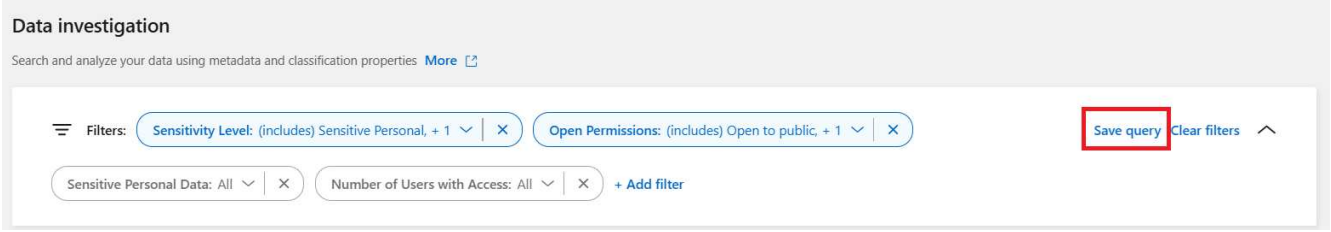
タブから特定の検索の [結果の調査] を選択するスクリーンショット。"]

保存されたクエリとポリシーを作成する

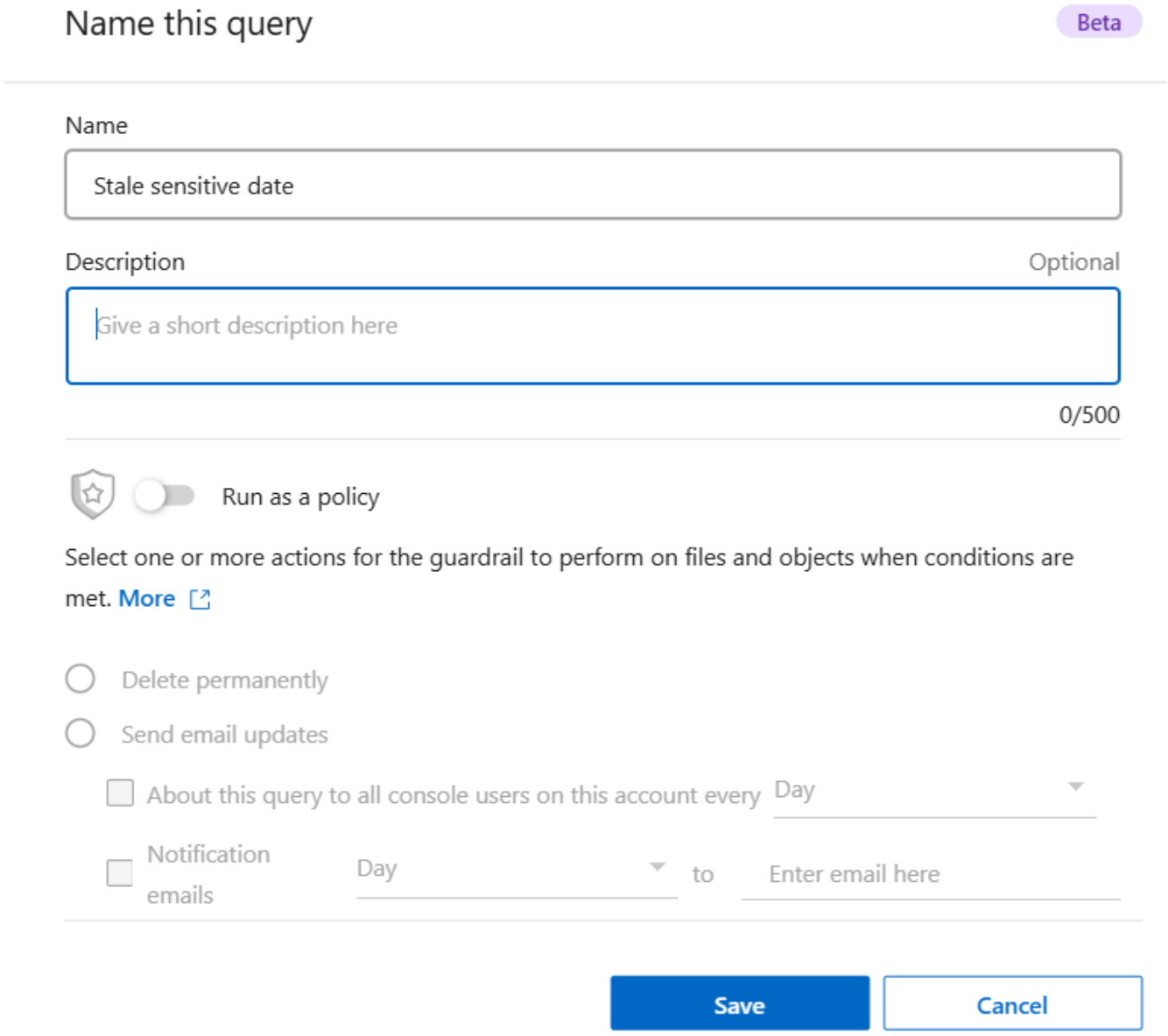
組織固有のクエリの結果を提供する独自のカスタム保存クエリを作成できます。検索条件に一致するすべてのファイルとディレクトリ (共有とフォルダー) の結果が返されます。

手順

1. 「調査」タブで、使用するフィルターを選択して検索を定義します。見る"調査ページでのデータのフィルタリング"詳細については。
2. すべてのフィルター特性を好みに合わせて設定したら、「クエリを保存」を選択します。



3. 保存したクエリに名前を付け、説明を追加します。名前は一意である必要があります。
4. オプションでクエリをポリシーとして保存できます。
   a. クエリをポリシーとして保存するには、[ポリシーとして実行] トグルを切り替えます。
   b. 完全に削除 または 電子メールで更新を送信 を選択します。電子メールによる更新を選択した場合は、クエリ結果を毎日、毎週、または毎月、すべてのコンソール ユーザーに電子メールで送信できます。あるいは、同じ頻度で特定の電子メール アドレスに通知を送信することもできます。
5. *保存*を選択します。



検索またはポリシーを作成したら、[保存したクエリ] タブで表示できます。

保存したクエリまたはポリシーを編集する

保存したクエリの名前と説明を変更できます。クエリをポリシーに変換したり、その逆を行ったりすることもできます。

デフォルトの保存済みクエリを変更することはできません。保存されたクエリのフィルターを変更することはできません。保存したクエリの調査結果を表示したり、フィルターを変更したり、修正したりして、新しいクエリまたはポリシーとして保存することもできます。

手順

1. [保存されたクエリ] ページで、変更する検索の [検索の編集] を選択します。

2. 名前と説明のフィールドを変更します。名前と説明のフィールドのみを変更します。

   オプションで、クエリをポリシーに変換したり、ポリシーを保存されたクエリに変換したりできます。必要に応じて、[ポリシーとして実行] トグルを切り替えます。..クエリをポリシーに変換する場合は、[完全に削除] または [電子メールの更新を送信] を選択します。電子メールによる更新を選択した場合は、クエリ結果を毎日、毎週、または毎月、すべてのコンソール ユーザーに電子メールで送信できます。あるいは、同じ頻度で特定の電子メール アドレスに通知を送信することもできます。

3. 変更を完了するには、[保存] を選択します。

保存したクエリを削除する

不要になった場合は、保存したカスタムクエリまたはポリシーを削除できます。デフォルトで保存されたクエリは削除できません。

保存したクエリを削除するには、特定の検索のボタンをクリックし、[クエリの削除] を選択してから、確認ダイアログでもう一度 [クエリの削除] を選択します。

デフォルトのクエリ

データ分類では、次のシステム定義の検索クエリが提供されます。

- データ主体名 - 高リスク

  50 人以上のデータ主体名を含むファイル

- メールアドレス - 高リスク

  50 個を超える電子メール アドレスを含むファイル、またはデータベース列の 50% を超える行に電子メール アドレスが含まれているファイル

- 個人データ - 高リスク

  20 個を超える個人データ識別子を含むファイル、またはデータベース列の 50% を超える行に個人データ識別子が含まれているファイル

- 個人データ - 7年以上古い

  個人情報または機密性の高い個人情報を含むファイル（最終更新日が 7 年以上前）

- 保護 - 高

  パスワード、クレジットカード情報、IBAN 番号、社会保障番号を含むファイルまたはデータベースの列

- 保護 - 低

3年以上アクセスされていないファイル

- 保護 - 中

  ID番号、納税者番号、運転免許証番号、医薬品ID、パスポート番号などの個人データ識別子を含むファイルまたはデータベース列を含むファイル

- 機密個人データ - 高リスク

  20 個を超える機密個人データ識別子を持つファイル、または行の 50% 以上に機密個人データが含まれるデータベース列

リポジトリのNetApp Data Classificationスキャン設定を変更する

各システムおよびデータ ソースでデータがスキャンされる方法を管理できます。変更は「リポジトリ」ベースで行うことができます。つまり、スキャンするデータ ソースの種類に応じて、ボリューム、スキーマ、ユーザーなどごとに変更を加えることができます。

変更できる項目としては、リポジトリをスキャンするかどうか、 NetApp Data Classificationが"マッピングスキャンまたはマッピング＆分類スキャン"。また、一定期間ボリュームのスキャンを停止する必要がある場合など、スキャンを一時停止したり再開したりすることもできます。

リポジトリのスキャンステータスを表示する

NetApp Data Classificationがスキャンしている個々のリポジトリ (ボリューム、バケットなど) を、システムおよびデータ ソースごとに表示できます。また、いくつが「マップ」され、いくつが「分類」されたかを確認することもできます。すべてのデータに対して完全な AI 識別が実行されるため、分類には時間がかかります。

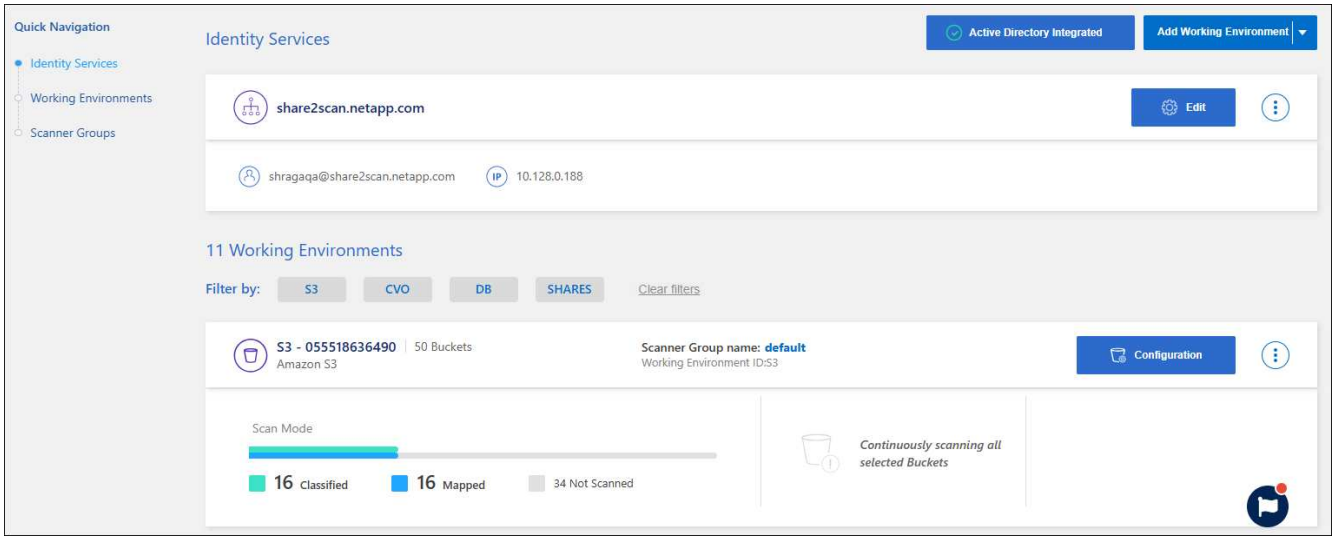
各作業環境のスキャンステータスは、構成ページで確認できます。

- 初期化中 (水色の点): マップまたは分類構成がアクティブ化されています。これは、「保留中のキュー」ステータスに移行する前に短時間表示されます。
- 保留中のキュー (オレンジ色の点): スキャン タスクはスキャン キューにリストされるのを待機しています。
- キューに追加 (オレンジ色の点): タスクがスキャン キューに正常に追加されました。システムは、キュー内の順番が来ると、ボリュームのマッピングまたは分類を開始します。
- 実行中 (緑のドット): キューにあったスキャン タスクが、選択したストレージ リポジトリでアクティブに進行中です。
- 完了 (緑の点): ストレージ リポジトリのスキャンが完了しました。
- 一時停止 (灰色の点): スキャンを一時停止しました。ボリュームの変化はシステムには表示されませんが、スキャンされたインサイトは引き続き利用できます。
- エラー (赤い点): 問題が発生したため、スキャンを完了できません。アクションを完了する必要がある場合は、「必要なアクション」列の下のツールチップにエラーが表示されます。それ以外の場合、システムは「エラー」ステータスを表示し、回復を試みます。終了するとステータスが変わります。

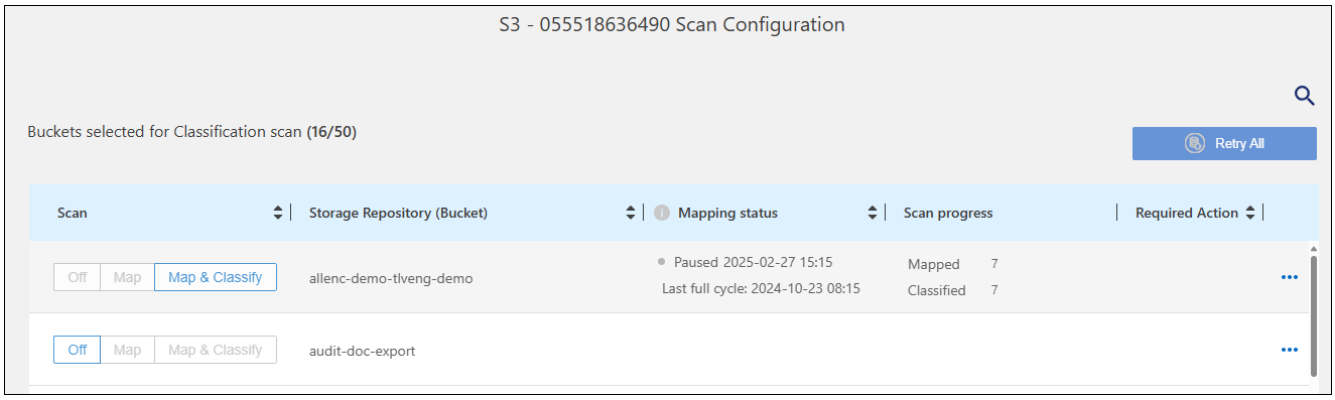
- スキャンしていません: ボリューム構成で「オフ」が選択されており、システムはボリュームをスキャンしていません。

手順

1. データ分類メニューから、*構成*を選択します。



2. [構成] タブから、システムの **[構成]** ボタンを選択します。
3. 「スキャン構成」ページで、すべてのリポジトリのスキャン設定を表示します。



4. スキャン中に、[マッピング ステータス] 列の進行状況バーにカーソルを合わせると、そのリポジトリにマッピングまたは分類されるキュー内のファイルの数が表示されます。

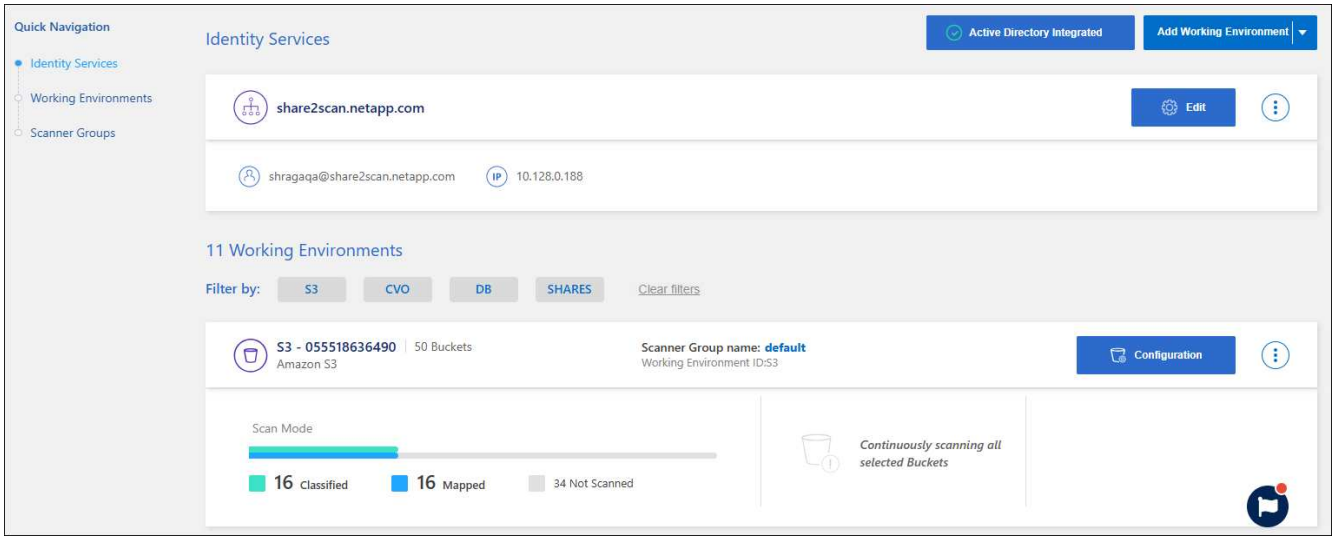
リポジトリのスキャンの種類を変更する

システム内のマッピングのみのスキャン、またはマッピングと分類のスキャンは、いつでも [構成] ページから開始または停止できます。マッピングのみのスキャンからマッピングと分類のスキャンに、またはその逆に変更することもできます。

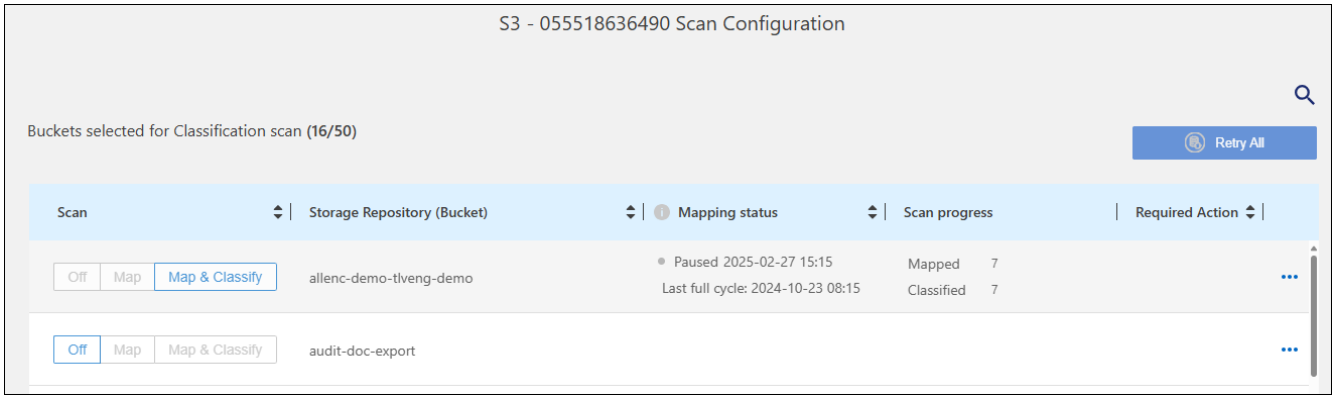
データベースをマッピングのみのスキャンに設定することはできません。データベース スキャンはオフまたはオンにすることができます。オンはマップと分類と同じです。

手順

1. データ分類メニューから、*構成*を選択します。
2. [構成] タブから、システムの [構成] ボタンを選択します。

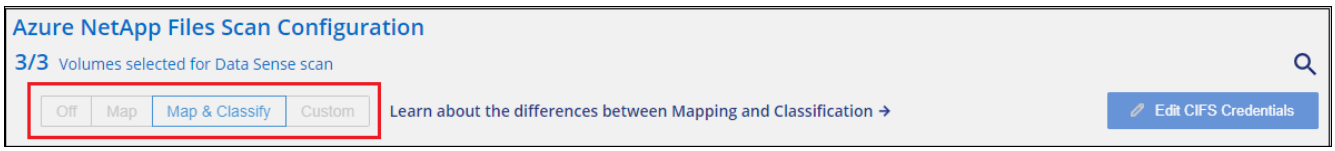


3. [スキャン構成] ページで、いずれかのリポジトリ (この例ではバケット) を変更して、マップ スキャンまたは マップと分類 スキャンを実行します。



特定の種類のシステムでは、ページ上部のボタン バーを使用して、すべてのリポジトリのスキャンの種類をグローバルに変更できます。これは、 Cloud Volumes ONTAP、オンプレミスのONTAP、 Azure NetApp Files、およびAmazon FSx for ONTAPシステムに有効です。

以下の例は、Azure NetApp Filesシステムのこのボタン バーを示しています。



スキャンを優先する

最も重要なマッピングのみのスキャンまたはマッピングと分類のスキャンを優先して、優先度の高いスキャンが最初に完了するようにすることができます。

デフォルトでは、スキャンは開始された順序に基づいてキューに入れられます。スキャンを優先順位付けする

機能を使用すると、スキャンをキューの先頭に移動できます。複数のスキャンを優先できます。優先順位は先入先出順で指定されます。つまり、最初に優先順位を指定したスキャンがキューの先頭に移動し、2 番目に優先順位を指定したスキャンはキューの 2 番目になり、以下同様に続きます。

優先権は 1 回限り付与されます。マッピング データの自動再スキャンはデフォルトの順序で実行されます。

手順

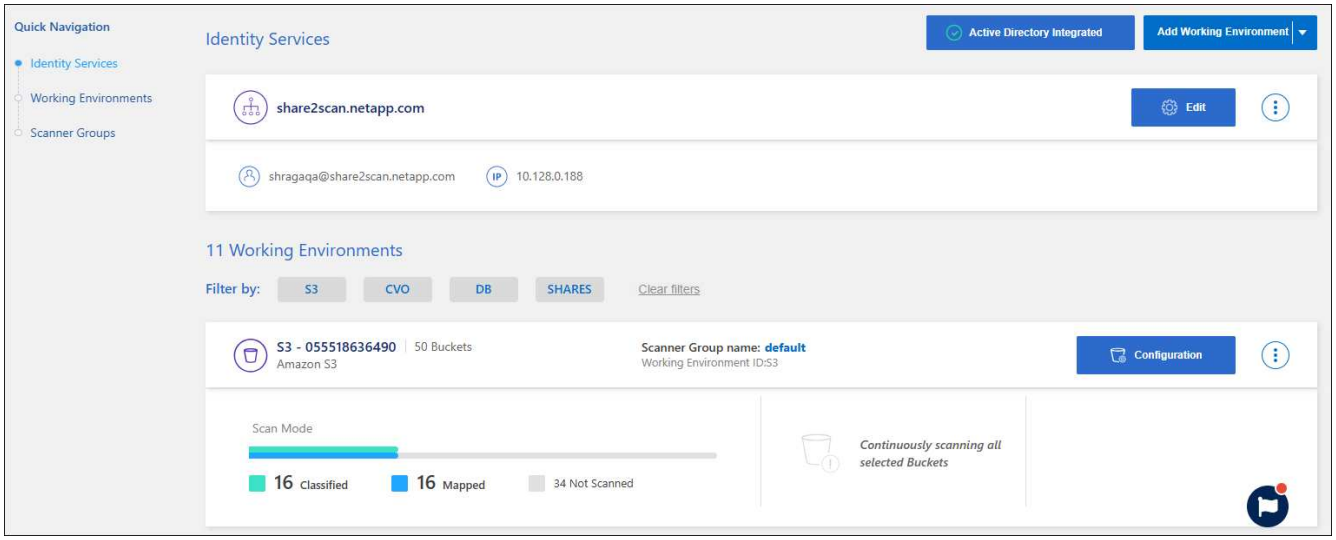
1. データ分類メニューから、*構成*を選択します。
2. 優先したいリソースを選択します。
3. 行動から `...`オプションで、[スキャンを優先] を選択します。

リポジトリのスキャンを停止する

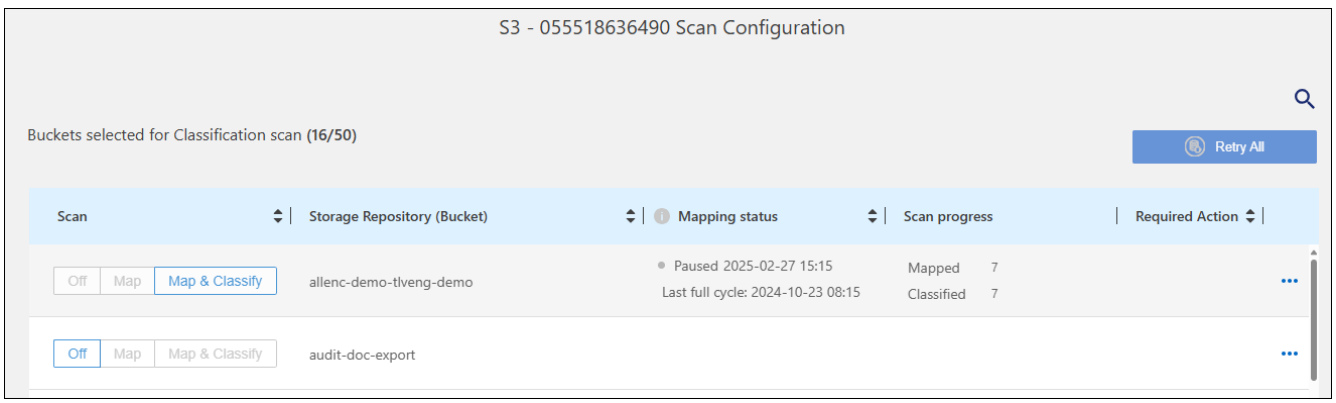
コンプライアンスを監視する必要がなくなった場合は、リポジトリ (ボリュームなど) のスキャンを停止できます。これを行うには、スキャンを「オフ」にします。スキャンをオフにすると、そのボリュームに関するすべてのインデックスと情報がシステムから削除され、データのスキャンに対する課金が停止されます。

手順

1. データ分類メニューから、*構成*を選択します。
2. [構成] タブから、システムの **[構成]** ボタンを選択します。



3. スキャン構成ページで オフ を選択して、特定のバケットのスキャンを停止します。



リポジトリのスキャンを一時停止して再開する

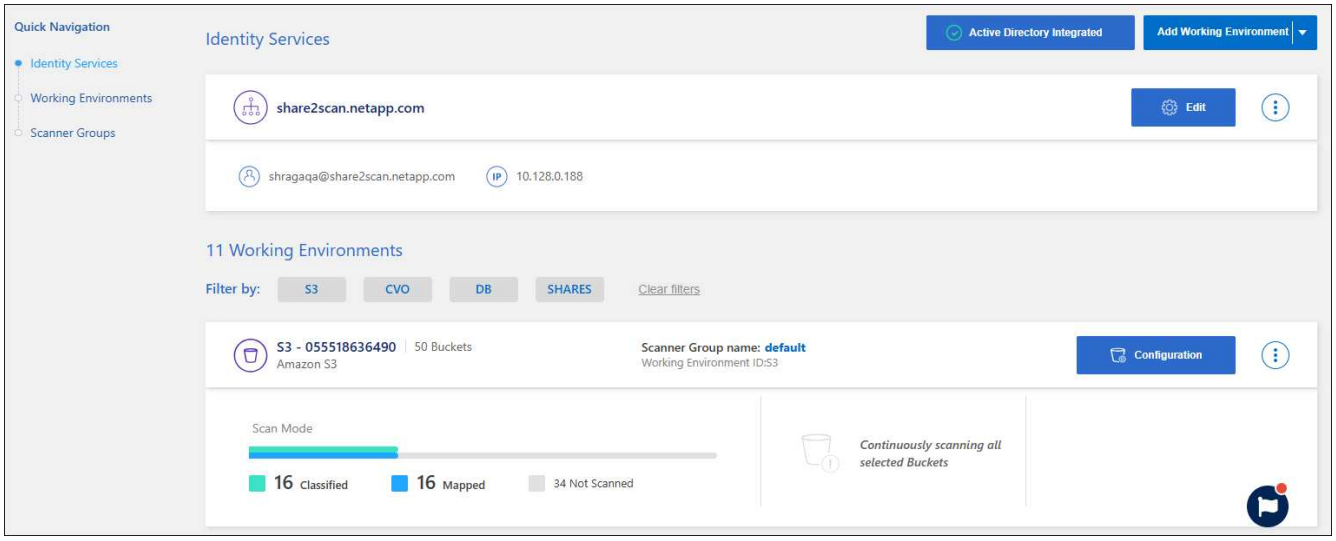
特定のコンテンツのスキャンを一時的に停止したい場合は、リポジトリのスキャンを「一時停止」することができます。スキャンを一時停止すると、データ分類はリポジトリの変更や追加について今後スキャンを実行しません。現在のすべてのスキャン結果は、データ分類で引き続きアクセスできます。

スキャンを一時停止しても、データはまだシステム上に残っているため、課金は発生しません。

いつでもスキャンを再開できます。

手順

1. データ分類メニューから、*構成*を選択します。
2. [構成] タブから、システムの [構成] ボタンを選択します。



3. スキャン設定ページで、アクションを選択します。 ••• アイコン。
4. ボリュームのスキャンを一時停止するには「一時停止」を選択し、一時停止していたボリュームのスキャンを再開するには「再開」を選択します。

NetApp Data Classificationコンプライアンスレポートを表示

NetApp Data Classification は、組織のデータ プライバシー プログラムの状態をより深く理解するために使用できるレポートを提供します。

デフォルトでは、データ分類ダッシュボードには、すべてのシステム、データベース、およびデータ ソースのコンプライアンスとガバナンスのデータが表示されます。一部のシステムのデータのみを含むレポートを表示する場合は、フィルターしてそれらのシステムのみを表示できます。

- コンプライアンス レポートは、データ ソースに対して完全な分類スキャンを実行した場合にのみ利用できます。マッピングのみのスキャンが行われたデータ ソースでは、データ マッピング レポートのみを生成できます。
- NetApp は、データ分類によって識別される個人データおよび機密個人データの 100% の正確性を保証することはできません。常にデータを確認して情報を検証する必要があります。

データ分類では次のレポートが利用可能です。

- データ検出評価レポート: スキャンされた環境の高レベルの分析を提供し、システムの検出結果を強調し、懸念事項と潜在的な修復手順を示します。このレポートはガバナンス ダッシュボードで利用できます。
- 完全なデータ マッピングの概要レポート: システム内のファイルのサイズと数に関する情報を提供します。これには、使用容量、データの古さ、データのサイズ、ファイルの種類が含まれます。このレポートはガバナンス ダッシュボードで利用できます。
- データ主体アクセス要求レポート: データ主体の特定の名前または個人識別子に関する情報が含まれるすべてのファイルのレポートを抽出できます。このレポートはコンプライアンス ダッシュボードで利用できます。
- **HIPAA** レポート: ファイル全体にわたる健康情報の分布を識別するのに役立ちます。このレポートはコンプライアンス ダッシュボードで利用できます。
- **PCI DSS** レポート: ファイル全体にわたるクレジットカード情報の分布を識別するのに役立ちます。このレポートはコンプライアンス ダッシュボードで利用できます。
- プライバシー リスク評価レポート: データから得られたプライバシーの分析情報とプライバシー リスク スコアを提供します。このレポートはコンプライアンス ダッシュボードで利用できます。
- 特定の情報タイプに関するレポート: 個人データや機密性の高い個人データを含む、特定されたファイルに関する詳細を含むレポートが利用可能です。カテゴリやファイルタイプ別に分類されたファイルを表示することもできます。

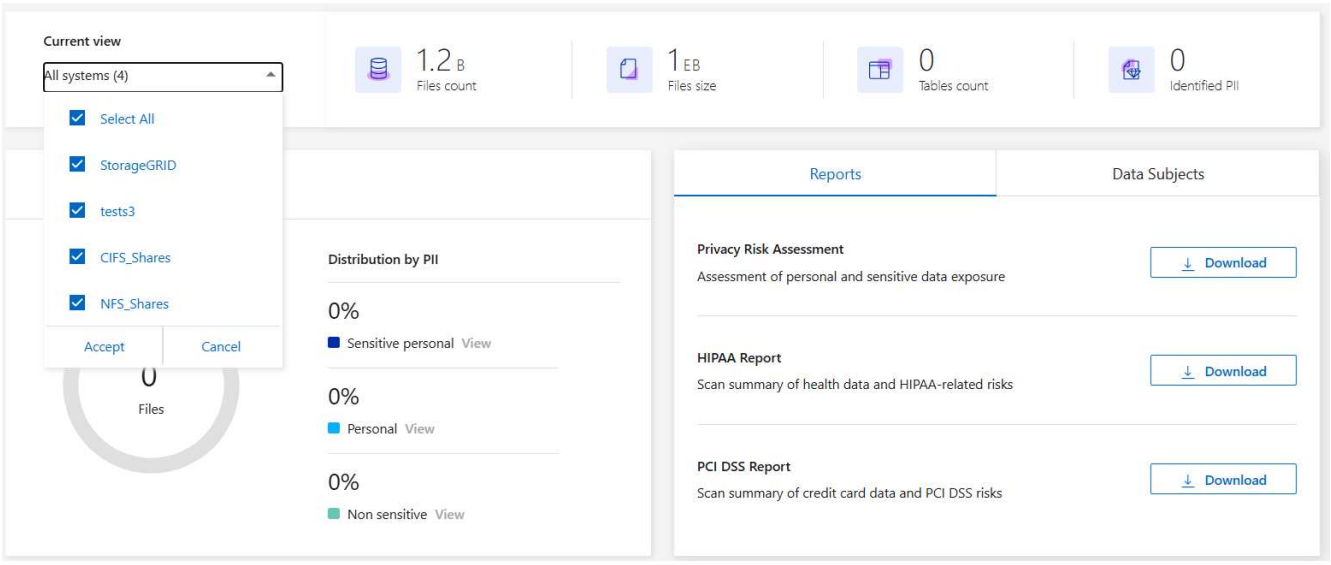
レポートのシステムを選択する

データ分類コンプライアンス ダッシュボードの内容をフィルタリングして、すべてのシステムとデータベース、または特定のシステムのコンプライアンス データを表示できます。

ダッシュボードをフィルターすると、データ分類によってコンプライアンス データの範囲が限定され、選択したシステムのみがレポートされます。

手順

1. データ分類メニューから、*コンプライアンス*を選択します。
2. システム フィルターのドロップダウンを選択し、システムを選択します。
3. 選択内容を確認するには、[承認] を選択します。



データ主体アクセス要求レポート

欧州の GDPR などのプライバシー規制では、データ主体 (顧客や従業員など) に個人データにアクセスする権利が付与されます。データ主体がこの情報を要求する場合、これは DSAR (データ主体アクセス要求) と呼ばれます。組織はこれらの要求に対して「不当な遅延なく」、遅くとも受領後 1 か月以内に応答する必要があります。

DSAR に応答するには、対象のフルネームまたは既知の識別子 (電子メール アドレスなど) を検索し、レポートをダウンロードします。このレポートは、GDPR または同様のデータ プライバシー法に準拠するという組織の要件を支援するために設計されています。

データ分類は **DSAR** への対応にどのように役立ちますか**?**

データ主体の検索を実行すると、データ分類によって、その人物の名前または識別子が含まれるすべてのファイルが検索されます。データ分類では、事前にインデックス付けされた最新のデータで名前または識別子を確認します。新しいスキャンは開始されません。

検索が完了したら、データ主体アクセス要求レポートのファイル リストをダウンロードできます。レポートはデータから得られた洞察を集約し、それを法的用語にまとめて相手に返送できるようにします。

現在、データベース内ではデータ主体の検索はサポートされていません。

データ主体を検索し、レポートをダウンロード

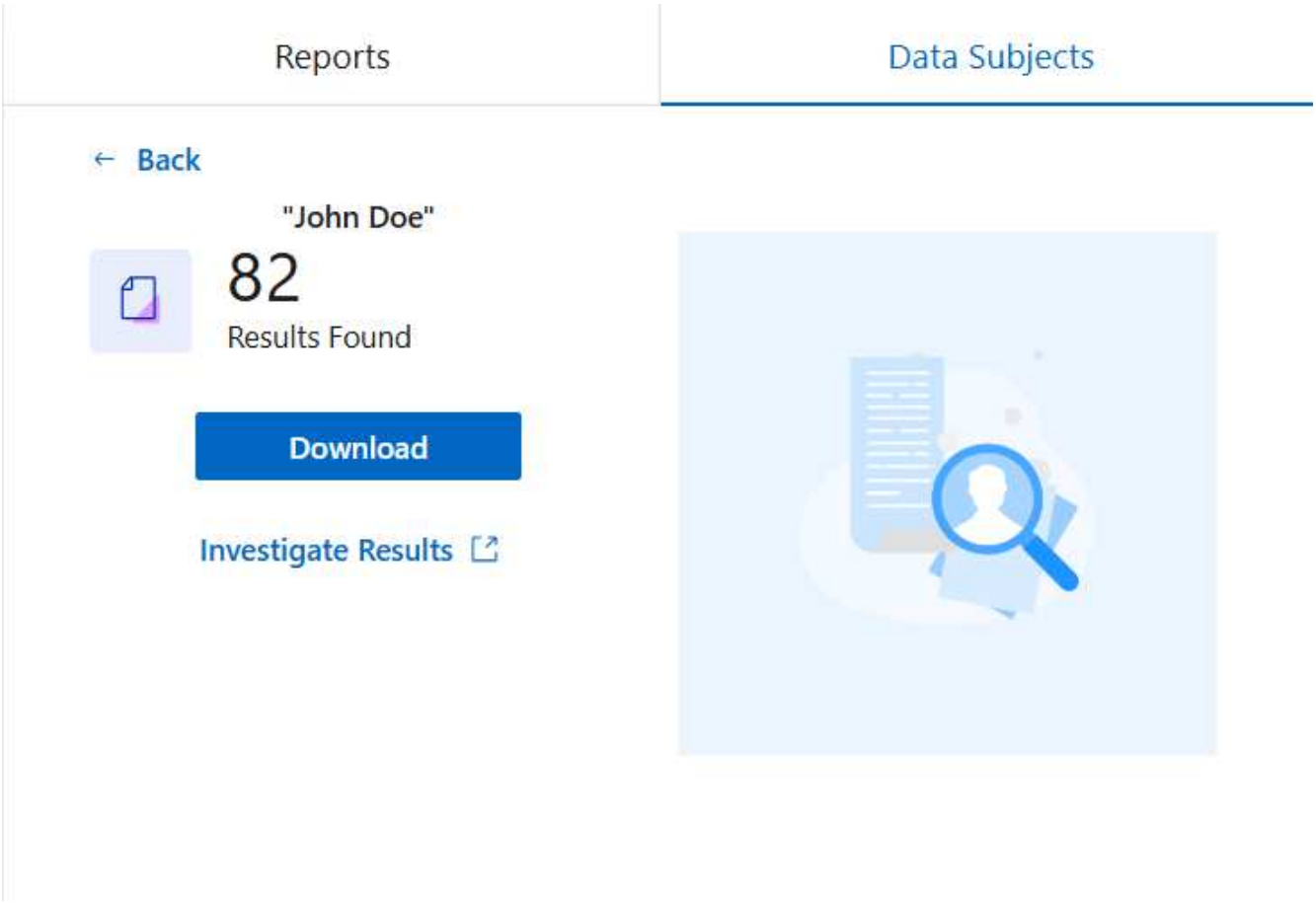
データ主体のフルネームまたは既知の識別子を検索し、ファイル リスト レポートまたは DSAR レポートをダウンロードします。検索条件"あらゆる個人情報の種類"。

データ主体の名前を検索する際には、英語、ドイツ語、日本語、スペイン語がサポートされています。今後、さらに多くの言語のサポートが追加される予定です。

手順

1. データ分類メニューから、*コンプライアンス*を選択します。
2. コンプライアンス ページで、データ主体 タブを見つけます。

3. *データ主体*セクションで、名前または既知の識別子を入力し、検索を選択します。
4. 検索が完了したら、[ダウンロード] を選択して、データ主体のアクセス要求応答にアクセスします。データ調査ページで詳細情報を表示するには、結果の調査 を選択しま
す。



5. データ分類で結果を確認するか、ダウンロード アイコンを選択してレポートとしてダウンロードします。

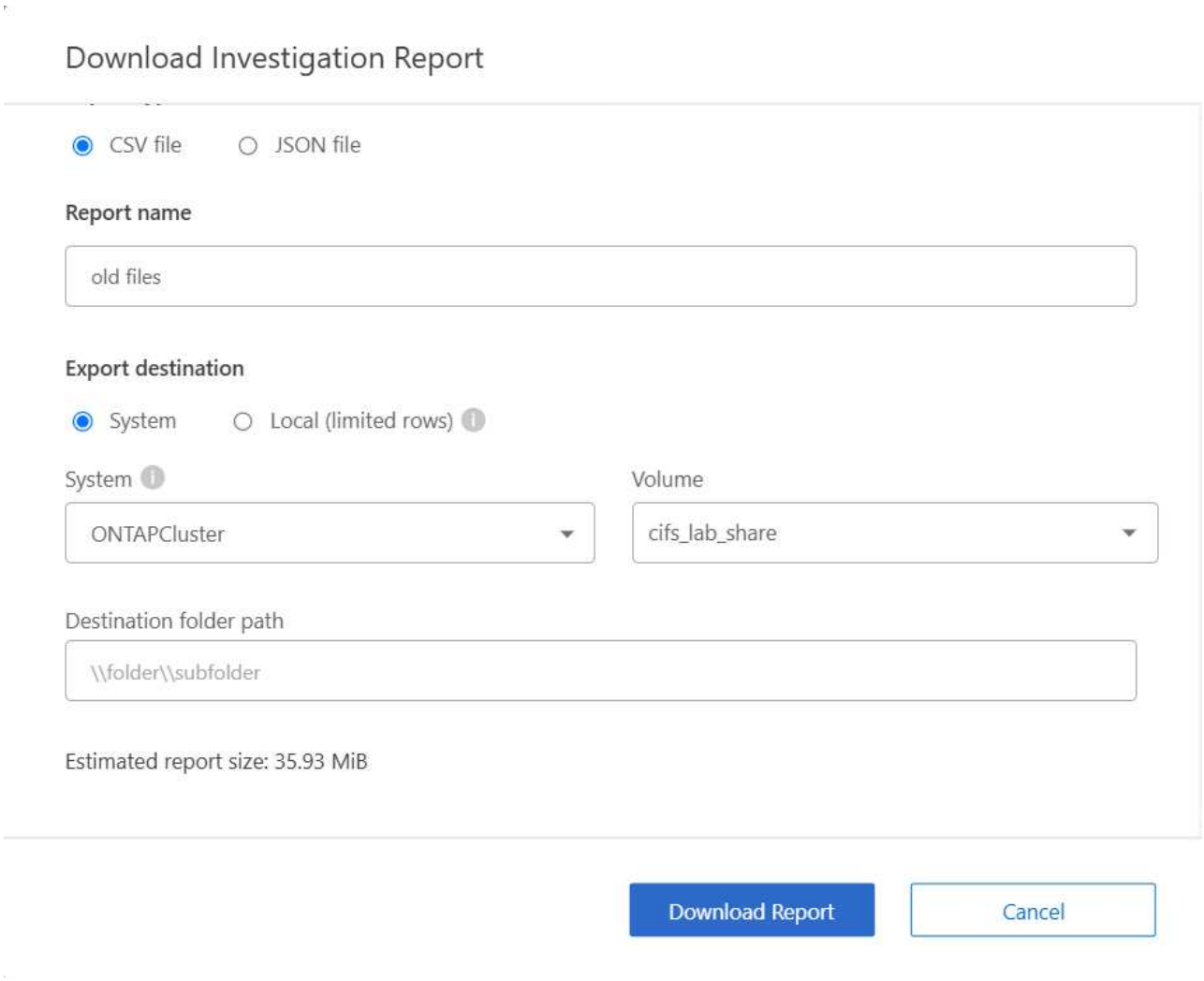
   a. ダウンロード アイコンを選択したら、ダウンロード設定を構成します。

      - フィルムフォーマットを選択してください: CSV または JSON
      - *レポート名*を入力してください
      - エクスポート先として「システム」または「ローカル」マシンを選択します。

        システムを選択した場合は、すべてのデータがダウンロードされます。 システム、ボリューム、*宛先フォルダーのパス*も選択する必要があります。

   ローカル を選択した場合、レポートは非構造化データの最初の 10,000 行、非構造化データの 5,000 行、および構造化データの 1,000 行に制限されます。

   a. ダウンロードを開始するには、[レポートのダウンロード] を選択しま
   す。



医療保険の携行性と責任に関する法律（HIPAA）に関する報告書

医療保険の携行性と責任に関する法律 (HIPAA) レポートは、健康情報を含むファイルを識別するのに役立ちます。これは、組織の HIPAA データ プライバシー法の遵守要件を支援するために設計されています。データ分類が探す情報には次のものが含まれます。

- 健康参照パターン
- ICD-10-CM医療コード
- ICD-9-CM医療コード
- HR - 健康カテゴリ
- 健康アプリケーションデータカテゴリ

レポートには次の情報が含まれます。

- 概要: 健康情報を含むファイルの数と、そのシステム。
- 暗号化: 暗号化されたシステムまたは暗号化されていないシステム上にある健康情報を含むファイルの割合。この情報はCloud Volumes ONTAPに固有のものです。
- ランサムウェア保護: ランサムウェア保護が有効になっているシステム、または有効になっていないシステム上にある健康情報を含むファイルの割合。この情報はCloud Volumes ONTAPに固有のものです。

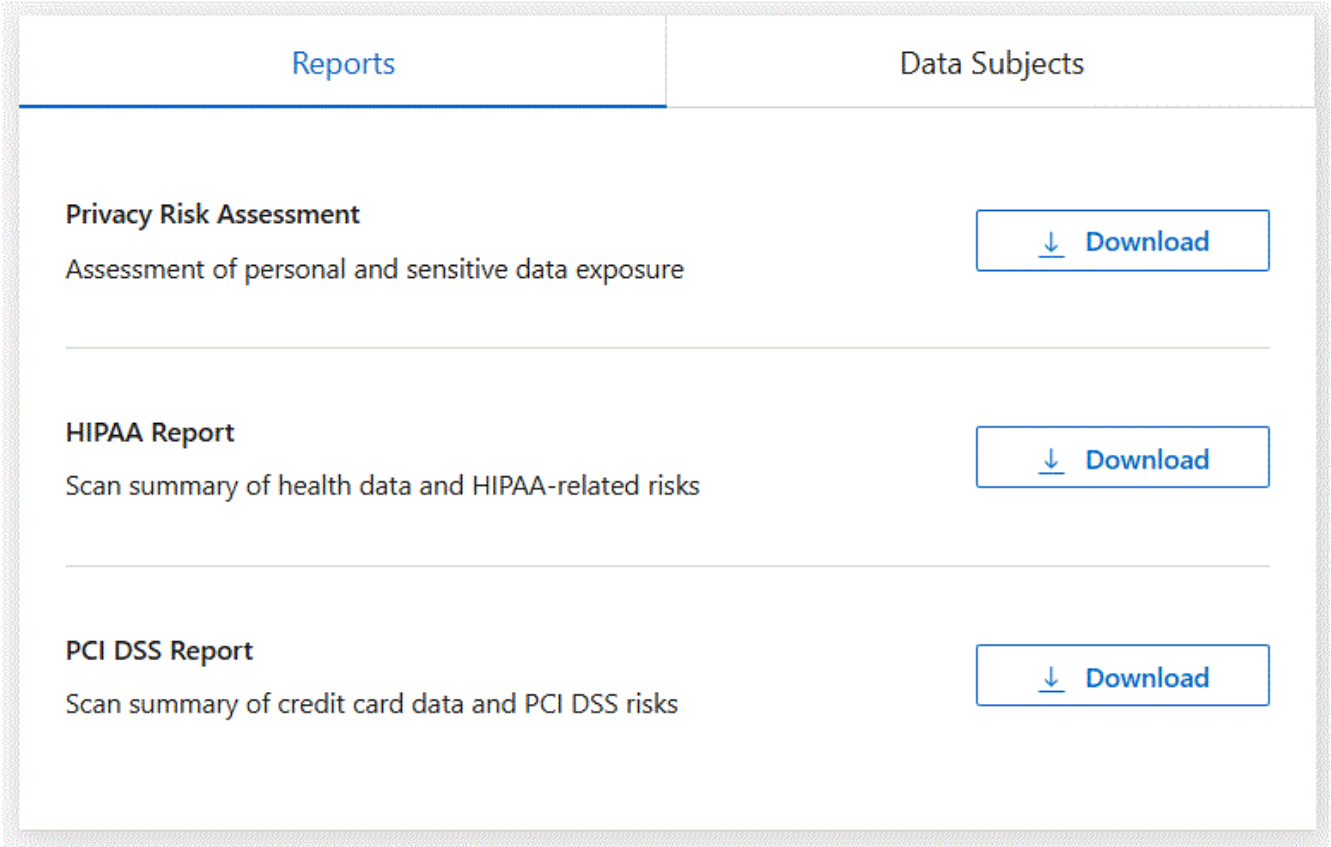
- 保持期間: ファイルが最後に変更された期間。これは、健康情報を処理に必要な期間以上保存すべきではないため、役立ちます。
- 健康情報の配布: 健康情報が見つかったシステムと、暗号化およびランサムウェア保護が有効になっているかどうか。

HIPAAレポートを生成する

レポートを生成するには、「コンプライアンス」タブに移動します。

手順

1. データ分類メニューから、*コンプライアンス*を選択します。
2. レポート ペイン を見つけます。 **HIPAA** レポート の横にあるダウンロード アイコンを選択します。



結果

データ分類により PDF レポートが生成されます。

ペイメントカード業界データセキュリティ基準（**PCI DSS**）レポート

ペイメント カード 業界データ セキュリティ標準 (PCI DSS) レポートは、ファイル全体にわたるクレジットカード情報の分布を識別するのに役立ちます。

レポートには次の情報が含まれます。

- 概要: クレジットカード情報が含まれているファイルの数と、そのシステム。
- 暗号化: 暗号化されたシステムまたは暗号化されていないシステム上にあるクレジットカード情報を含む

ファイルの割合。この情報はCloud Volumes ONTAPに固有のものです。

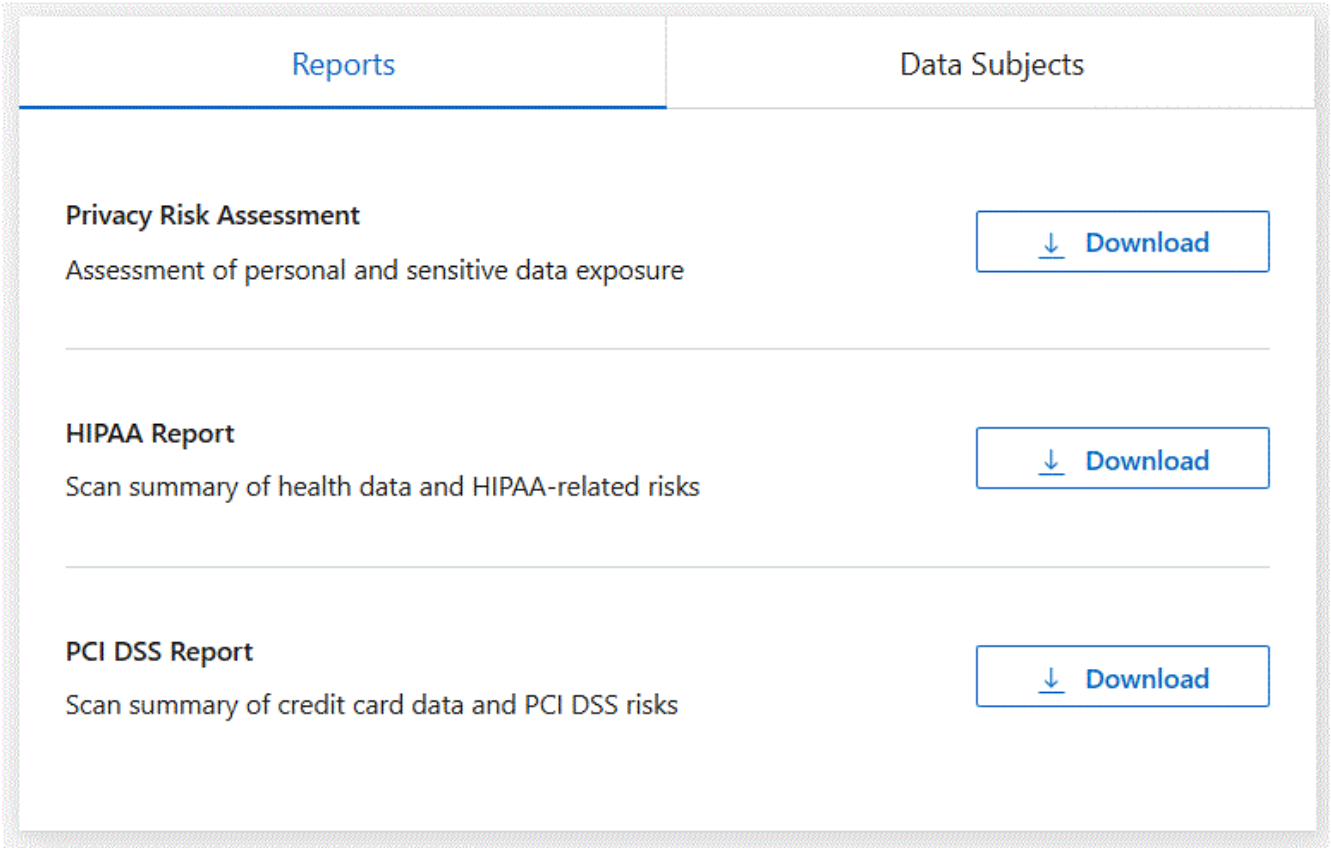
- ランサムウェア保護: ランサムウェア保護が有効になっているシステム、または有効になっていないシステム上にある、クレジットカード情報を含むファイルの割合。この情報はCloud Volumes ONTAPに固有のものです。
- 保持期間: ファイルが最後に変更された期間。これは、クレジットカード情報を処理に必要な期間以上保持するべきではないため、役立ちます。
- クレジットカード情報の配布: クレジットカード情報が見つかったシステムと、暗号化およびランサムウェア保護が有効になっているかどうか。

PCI DSSレポートを生成する

レポートを生成するには、「コンプライアンス」タブに移動します。

手順

1. データ分類メニューから、*コンプライアンス*を選択します。
2. レポート ペイン を見つけます。 **PCI DSS** レポート の横にあるダウンロード アイコンを選択します。



結果

データ分類では、必要に応じて確認したり他のグループに送信したりできる PDF レポートが生成されます。

プライバシーリスク評価レポート

プライバシー リスク評価レポートでは、GDPR や CCPA などのプライバシー規制の要件に従って、組織のプライバシー リスク状態の概要が提供されます。

レポートには次の情報が含まれます。

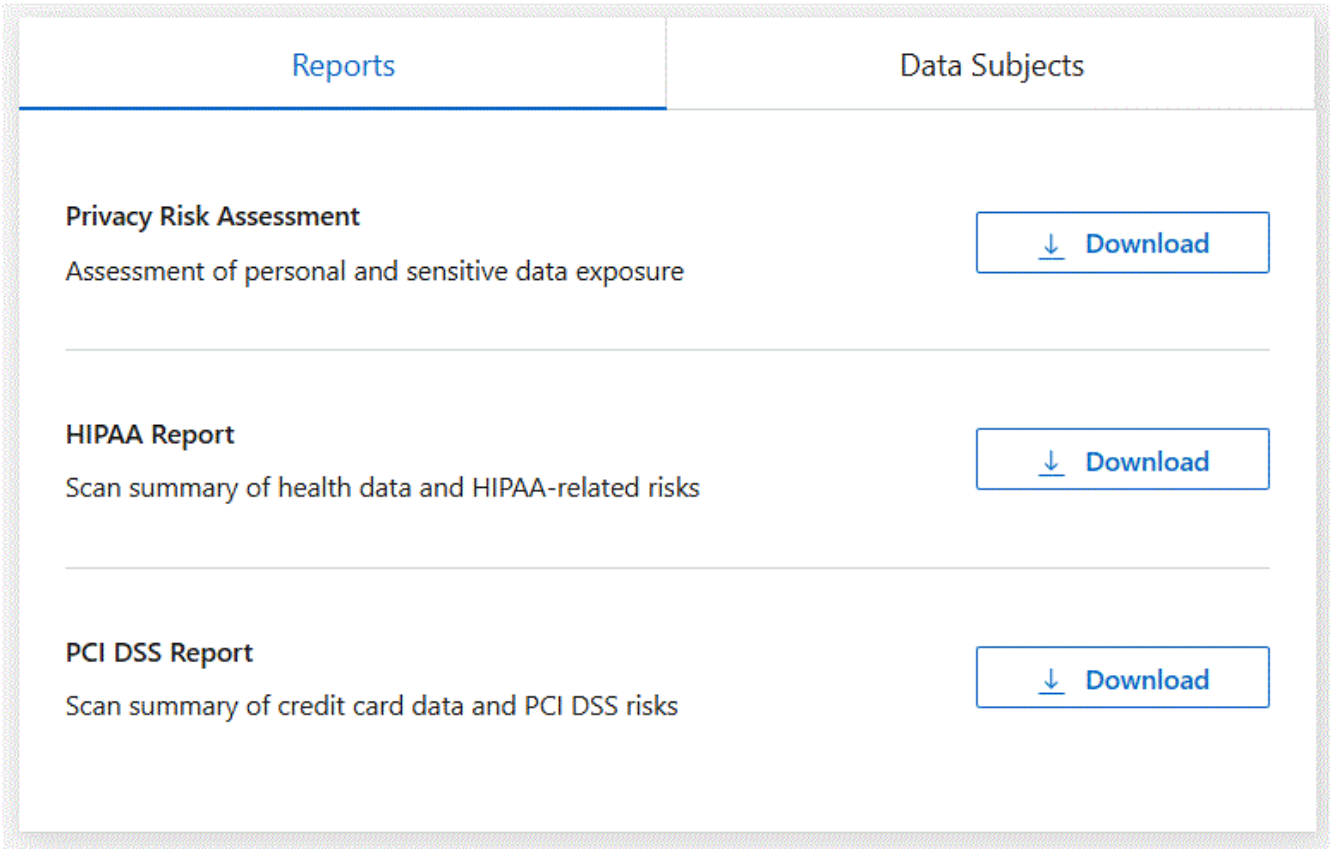
- コンプライアンス ステータス: 重大度スコアと、データが非機密データ、個人情報、または機密個人情報であるかどうかの分布。
- 評価の概要: 見つかった個人データの種類とデータのカテゴリの内訳。
- この評価におけるデータ主体: 国民識別子が見つかった場所別の人数。

プライバシーリスク評価レポートを生成する

レポートを生成するには、「コンプライアンス」タブに移動します。

手順

1. データ分類メニューから、*コンプライアンス*を選択します。
2. レポート ペイン を見つけます。 プライバシー リスク評価レポート の横にあるダウンロード アイコンを選択します。



結果

データ分類では、必要に応じて確認したり他のグループに送信したりできる PDF レポートが生成されます。

重症度スコア

データ分類では、次の 3 つの変数に基づいてプライバシー リスク評価レポートの重大度スコアを計算します。

- すべてのデータのうち個人データが占める割合。

- すべてのデータのうちの機密個人データの割合。
- 国民 ID、社会保障番号、納税者番号などの国民識別子によって決定されるデータ主体を含むファイルの割合。

スコアを決定するために使用されるロジックは次のとおりです。

重症度スコア	論理
0	3つの変数はすべて0%
1	変数の1つが0%より大きい
2	変数の1つが3%より大きい
3	変数のうち2つは3%より大きい
4	変数のうち3つは3%より大きい
5	変数の1つが6%より大きい
6	変数のうち2つは6%より大きい
7	変数のうち3つは6%より大きい
8	変数の1つが15%より大きい
9	変数のうち2つは15%より大きい
10	変数のうち3つは15%より大きい

NetApp Data Classificationの健全性を監視する

NetApp Data Classification Health Monitor ダッシュボードは、パフォーマンスのリアルタイム監視と分析情報を提供します。ヘルス モニターは、データ分類インフラストラクチャ、システムの健全性、使用状況メトリック、使用率データに関する情報を取得し、問題を特定して修復できるようにします。

ヘルスモニターの洞察

ヘルス モニター ダッシュボードには、4 つのカテゴリで情報が表示されます。

- インフラストラクチャの状態

  バージョンの状態、システムの安定性、展開の種類、マシンのスケールなどの情報を表示します。

- 問題のあるコンテナ

  問題のあるコンテナ フィールドを確認して、頻繁に停止または再起動されるコンテナに関する情報を取得します。この情報を使用して、特定のコンテナーを調査します。

- システム情報

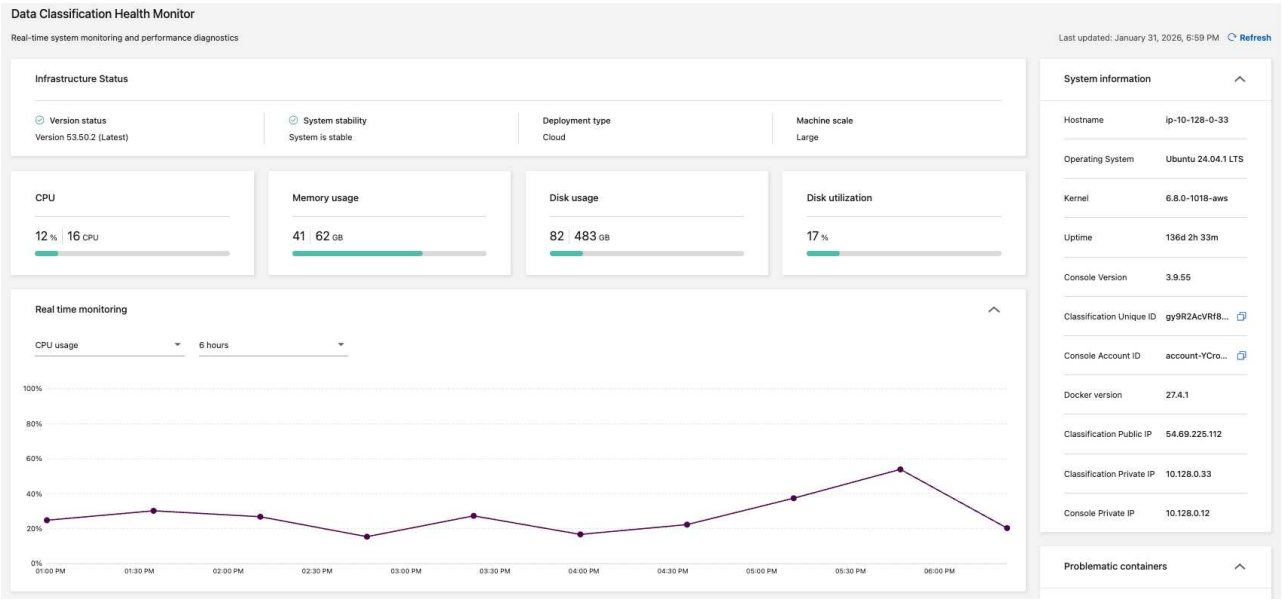
  システム情報パネルには、パブリックおよびプライベート IP アドレス、ホスト名、オペレーティング システム、コンソール バージョン、コンソール ID など、 NetApp Consoleとデータ分類に関する重要な情報が記録されます。

- 使用と活用

  CPU 使用率、ディスク使用率、ディスク使用率、メモリ使用率を確認します。これらの値は、ストレージ単位 (GB) または合計使用量のパーセンテージで表示されます。いずれかのフィールドに警告が表示された場合は、情報と修復の推奨事項については警告を選択してください。

ヘルスモニターダッシュボードにアクセスする

1. データ分類で、構成 を選択します。
2. 構成 の見出しの下で、データ分類ヘルスモニター を選択します。
3. ヘルスモニターダッシュボードでは、次のことができます。
   - 使用状況と利用状況を確認します。使用状況または使用率のメトリックに警告が表示される場合は、問題を解決するための推奨事項の警告を選択してください。
   - グラフを切り替えると、CPU 使用率、ディスク使用率、ディスク使用率、メモリ使用率が表示されます。x 軸を変更して、コンテンツを時間単位 (6、12、または 24) または日単位 (2、7、または 14) で表示できます。
   - 最新のデータ メトリックを表示するには、ダッシュボードを更新します。



著作権に関する情報

商標に関する情報