Riferimento

NetApp Data Classification

NetApp
January 14, 2026

Sommario

Riferimento

Tipi di istanza NetApp Data Classification supportati

Il software NetApp Data Classification deve essere eseguito su un host che soddisfi specifici requisiti del sistema operativo, requisiti RAM, requisiti software e così via. Quando si distribuisce la classificazione dei dati nel cloud, si consiglia di utilizzare un sistema con caratteristiche "large" per una funzionalità completa.

È possibile implementare Data Classification su un sistema con meno CPU e meno RAM, ma ci sono alcune limitazioni quando si utilizzano questi sistemi meno potenti. "Scopri di più su queste limitazioni" .

Nelle tabelle seguenti, se il sistema contrassegnato come "predefinito" non è disponibile nella regione in cui si sta installando Data Classification, verrà distribuito il sistema successivo nella tabella.

Tipi di istanza AWS

Dimensioni del sistema	Specifiche	Tipo di istanza
Extra Large	32 CPU, 128 GB di RAM, 1 TiB gp3 SSD	"m6i.8xlarge"(predefinito)
Grande	16 CPU, 64 GB di RAM, SSD da 500 GiB	"m6i.4xlarge"(predefinito) m6a.4xlarge m5a.4xlarge m5.4xlarge m4.4xlarge
Medio	8 CPU, 32 GB di RAM, SSD da 200 GiB	"m6i.2xlarge"(predefinito) m6a.2xlarge m5a.2xlarge m5.2xlarge m4.2xlarge
Piccolo	8 CPU, 16 GB di RAM, SSD da 100 GiB	"c6a.2xlarge"(predefinito) c5a.2xlarge c5.2xlarge c4.2xlarge

Tipi di istanza di Azure

Dimensioni del sistema	Specifiche	Tipo di istanza
Extra Large	32 CPU, 128 GB di RAM, disco del sistema operativo (2.048 GiB, velocità di trasmissione minima 250 MB/s) e disco dati (SSD da 1 TiB, velocità di trasmissione minima 750 MB/s)	"Standard_D32_v3"(predefinito)
Grande	16 CPU, 64 GB di RAM, SSD da 500 GiB	"Standard_D16s_v3"(predefinito)

Tipi di istanza GCP

Dimensioni del sistema	Specifiche	Tipo di istanza
Grande	16 CPU, 64 GB di RAM, SSD da 500 GiB	"n2-standard-16"(predefinito) n2d-standard-16 n1-standard-16

Metadati raccolti da fonti di dati in NetApp Data Classification

NetApp Data Classification raccoglie determinati metadati durante l'esecuzione di scansioni di classificazione sui dati provenienti dalle fonti dati e dai sistemi. Data Classification può accedere alla maggior parte dei metadati di cui abbiamo bisogno per classificare i tuoi dati, ma ci sono alcune fonti da cui non siamo in grado di accedere ai dati di cui abbiamo bisogno.

	Metadati	CIFS	Non è vero
Timbri temporali	*Ora di creazione*	Disponibile	Non disponibile (non supportato in Linux)
	Ultimo orario di accesso	Disponibile	Disponibile
	Ora dell'ultima modifica	Disponibile	Disponibile
Autorizzazioni	*Apri permessi*	Se il gruppo "TUTTI" ha accesso al file, questo è considerato "Aperto all'organizzazione"	Se "Altri" ha accesso al file, questo è considerato "Aperto all'organizzazione"
	Accesso utenti/gruppi	Le informazioni sugli utenti e sui gruppi vengono prese da LDAP	Non disponibile (gli utenti NFS sono solitamente gestiti localmente sul server, pertanto lo stesso individuo può avere un UID diverso in ogni server)

- La classificazione dei dati non estrae l'"orario dell'ultimo accesso" dalle fonti dati del database.
- Le versioni precedenti del sistema operativo Windows (ad esempio, Windows 7 e Windows 8) disabilitano per impostazione predefinita la raccolta dell'attributo "ora dell'ultimo accesso" perché può influire sulle prestazioni del sistema. Se questo attributo non viene raccolto, le analisi di classificazione dei dati basate sull'"orario dell'ultimo accesso" saranno interessate. Se necessario, è possibile abilitare la raccolta dell'orario dell'ultimo accesso su questi vecchi sistemi Windows.

Timestamp dell'ultimo accesso

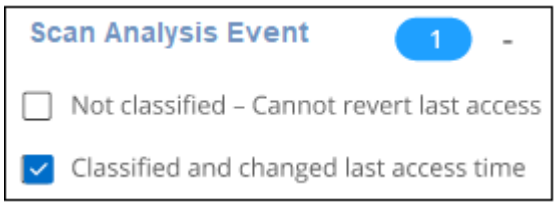
Quando Data Classification estrae dati dalle condivisioni di file, il sistema operativo lo considera come se stesse accedendo ai dati e modifica di conseguenza l'"orario dell'ultimo accesso". Dopo la scansione, la classificazione dei dati tenta di ripristinare l'orario dell'ultimo accesso al timestamp originale. Se Data Classification non dispone di autorizzazioni di scrittura degli attributi in CIFS o di autorizzazioni di scrittura in NFS, il sistema non può ripristinare l'orario dell'ultimo accesso al timestamp originale. I volumi ONTAP configurati con SnapLock hanno autorizzazioni di sola lettura e non possono ripristinare l'orario dell'ultimo accesso al timestamp originale.

Per impostazione predefinita, se Data Classification non dispone di queste autorizzazioni, il sistema non analizzerà i file nei volumi perché Data Classification non può ripristinare l'"ultimo orario di accesso" al timestamp originale. Tuttavia, se non ti interessa che l'orario dell'ultimo accesso venga reimpostato sull'orario originale nei tuoi file, puoi selezionare l'opzione **Scansiona quando mancano le autorizzazioni "attributi di**

scrittura" nella parte inferiore della pagina Configurazione, in modo che Data Classification esegua la scansione dei volumi indipendentemente dalle autorizzazioni.

Questa funzionalità è applicabile ai sistemi ONTAP locali, Cloud Volumes ONTAP, Azure NetApp Files, Amazon FSx for NetApp ONTAP e condivisioni file di terze parti.

Nella pagina Indagine è presente un filtro denominato *Evento analisi scansione* che consente di visualizzare i file che non sono stati classificati perché la classificazione dei dati non è riuscita a ripristinare l'orario dell'ultimo accesso oppure i file che sono stati classificati anche se la classificazione dei dati non è riuscita a ripristinare l'orario dell'ultimo accesso.



Le selezioni del filtro sono:

- "Non classificato - Impossibile ripristinare l'ultimo orario di accesso": mostra i file che non sono stati classificati a causa della mancanza di autorizzazioni di scrittura.
- "Ultimo orario di accesso classificato e aggiornato": mostra i file che sono stati classificati e la classificazione dei dati non è riuscita a reimpostare l'ultimo orario di accesso alla data originale. Questo filtro è rilevante solo per gli ambienti in cui è stata attivata l'opzione **Scansione quando mancano le autorizzazioni "attributi di scrittura"**.

Se necessario, è possibile esportare questi risultati in un report, in modo da poter vedere quali file vengono o non vengono analizzati in base alle autorizzazioni. "Scopri di più sui report di Data Investigation".

Accedi al sistema NetApp Data Classification

È necessario accedere al sistema NetApp Data Classification per poter accedere ai file di registro o modificare i file di configurazione.

Quando Data Classification è installato su una macchina Linux in sede o su una macchina Linux distribuita nel cloud, è possibile accedere direttamente al file di configurazione e allo script.

Quando Data Classification viene distribuito nel cloud, è necessario connettersi tramite SSH all'istanza di Data Classification. È possibile accedere al sistema tramite SSH immettendo nome utente e password oppure utilizzando la chiave SSH fornita durante l'installazione dell'agente Console. Il comando SSH è:

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

- `<path_to_the_ssh_key>`= posizione delle chiavi di autenticazione ssh
- `<machine_user>`:
  - Per AWS: utilizzare <ec2-user>
  - Per Azure: utilizzare l'utente creato per l'istanza della console
  - Per GCP: utilizzare l'utente creato per l'istanza della console
- `<datasense_ip>`= Indirizzo IP dell'istanza della macchina virtuale

Per accedere al sistema nel cloud è necessario modificare le regole in entrata del gruppo di sicurezza. Per maggiori dettagli, vedere:

- "Regole del gruppo di sicurezza in AWS"
- "Regole del gruppo di sicurezza in Azure"
- "Regole del firewall in Google Cloud"

API NetApp Data Classification

Le funzionalità NetApp Data Classification disponibili tramite l'interfaccia utente Web sono disponibili anche tramite l'API REST.

Nella Classificazione dei dati sono definite quattro categorie che corrispondono alle schede nell'interfaccia utente:

- Indagine
- Conformità
- Governance
- Configurazione

Le API nella documentazione di Swagger consentono di cercare, aggregare dati, monitorare le scansioni ed eseguire azioni tra cui copia, spostamento ed eliminazione.

Panoramica

L'API consente di eseguire le seguenti funzioni:

- Informazioni sull'esportazione
  - Tutto ciò che è disponibile nell'interfaccia utente può essere esportato tramite l'API (ad eccezione dei report)
  - I dati vengono esportati in formato JSON (facile da analizzare e inviare ad applicazioni di terze parti, come Splunk)
- Crea query utilizzando le istruzioni "AND" e "OR", includi ed escludi informazioni e molto altro.

  Ad esempio, è possibile individuare file *senza* informazioni personali identificabili (PII) specifiche (funzionalità non disponibile nell'interfaccia utente). È anche possibile escludere campi specifici dall'operazione di esportazione.

- Eseguire azioni

- Aggiorna le credenziali CIFS
- Visualizza e annulla le azioni
- Ripeti la scansione delle directory
- Esporta dati

L'API è sicura e utilizza lo stesso metodo di autenticazione dell'interfaccia utente. Puoi trovare informazioni sull'autenticazione nel"Documentazione REST API" .

Accesso al riferimento API Swagger

Per accedere a Swagger ti servirà l'indirizzo IP della tua istanza di classificazione dei dati. Nel caso di un'implementazione cloud, utilizzerai l'indirizzo IP pubblico. Quindi dovrai accedere a questo endpoint:

https://<ip_classificazione>/documentazione

Esempio utilizzando le API

L'esempio seguente mostra una chiamata API per copiare i file.

Richiesta API

Inizialmente sarà necessario ottenere tutti i campi e le opzioni rilevanti affinché un sistema possa visualizzare tutti i filtri nella scheda di indagine.

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

Risposta

```
{
  "options": [
    {
      "active_directory_affected": false,
      "data_mode": "ALL_SCANNED",
      "field": "string",
      "is_rulable": true,
      "name": "string",
      "operators": [
        "EQUALS"
      ],
      "optional_values": [
        {}
      ],
      "secondary": {},
      "server_data": false,
      "type": "TEXT"
```

```
    }
  ]
}
{
  "options": [
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "POLICIES",
      "name": "Policies",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "EXTRACTION_STATUS_RANGE",
      "name": "Scan Analysis Status",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "SCAN_ANALYSIS_ERROR",
      "name": "Scan Analysis Event",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "PUBLIC_ACCESS",
      "name": "Open Permissions",
      "operators": [
        "IN",
```

```
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": true,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "USERS_PERMISSIONS_COUNT_RANGE",
      "name": "Number of Users with Access",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": true,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "USER_GROUP_PERMISSIONS",
      "name": "User / Group Permissions",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "FILE_OWNER",
      "name": "File Owner",
      "operators": [
        "EQUALS",
        "CONTAINS"
      ],
      "server_data": true,
      "type": "TEXT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "ENVIRONMENT_TYPE",
      "name": "system-type",
      "operators": [
```

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

```
      "name": "Category",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "PATTERN_SENSITIVITY_LEVEL",
      "name": "Sensitivity Level",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "NUMBER_OF_IDENTIFIERS",
      "name": "Number of identifiers",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "PATTERN_PERSONAL",
      "name": "Personal Data",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
```

```
    "field": "PATTERN_SENSITIVE",
    "name": "Sensitive Personal Data",
    "operators": [
      "IN",
      "NOT_IN"
    ],
    "server_data": true,
    "type": "SELECT"
  },
  {
    "active_directory_affected": false,
    "data_mode": "ALL_EXTRACTABLE",
    "field": "DATA_SUBJECT",
    "name": "Data Subject",
    "operators": [
      "EQUALS",
      "CONTAINS"
    ],
    "server_data": true,
    "type": "TEXT"
  },
  {
    "active_directory_affected": false,
    "data_mode": "DIRECTORIES",
    "field": "DIRECTORY_TYPE",
    "name": "Directory Type",
    "operators": [
      "IN",
      "NOT_IN"
    ],
    "server_data": true,
    "type": "SELECT"
  },
  {
    "active_directory_affected": false,
    "data_mode": "ALL_EXTRACTABLE",
    "field": "FILE_TYPE",
    "name": "File Type",
    "operators": [
      "IN",
      "NOT_IN"
    ],
    "server_data": true,
    "type": "SELECT"
  },
  {
```

```
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "FILE_SIZE_RANGE",
      "name": "File Size",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "FILE_CREATION_RANGE_RETENTION",
      "name": "Created Time",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "DISCOVERED_TIME_RANGE",
      "name": "Discovered Time",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "FILE_LAST_MODIFICATION_RETENTION",
      "name": "Last Modified",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
```

```
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "FILE_LAST_ACCESS_RANGE_RETENTION",
      "name": "Last Accessed",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "FILES",
      "field": "IS_DUPLICATE",
      "name": "Duplicates",
      "operators": [
        "EQUALS",
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "FILES",
      "field": "FILE_HASH",
      "name": "File Hash",
      "operators": [
        "EQUALS",
        "IN"
      ],
      "server_data": true,
      "type": "TEXT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "USER_DEFINED_STATUS",
      "name": "Tags",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
```

```
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "ASSIGNED_TO",
      "name": "Assigned to",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    }
  ]
}
```

Utilizzeremo questa risposta nei parametri della nostra richiesta per filtrare i file desiderati che vogliamo copiare.

È possibile applicare un'azione a più elementi. I tipi di azioni supportati includono: sposta, elimina e copia.

Creeremo l'azione di copia:

Richiesta API

La prossima API è l'API di azione e consente di creare più azioni.

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

Risposta

La risposta restituirà l'oggetto azione, quindi è possibile utilizzare le API get ed delete per ottenere lo stato dell'azione o per annullarla.

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