Référence

NetApp Data Classification

NetApp
February 11, 2026

Sommaire

Référence

Types d'instances de NetApp Data Classification pris en charge

Le logiciel de NetApp Data Classification doit s'exécuter sur un hôte qui répond aux exigences spécifiques du système d'exploitation, de la RAM, des logiciels, etc. Lors du déploiement de la classification des données dans le cloud, nous vous recommandons d'utiliser un système doté de caractéristiques « larges » pour bénéficier de toutes les fonctionnalités.

Vous pouvez déployer la classification des données sur un système avec moins de processeurs et moins de RAM, mais il existe certaines limitations lors de l'utilisation de ces systèmes moins puissants. "En savoir plus sur ces limitations" .

Dans les tableaux suivants, si le système marqué comme « par défaut » n'est pas disponible dans la région où vous installez Data Classification, le système suivant dans le tableau sera déployé.

Types d'instances AWS

Taille du système	Spécifications	Type d'instance
Très grand	32 processeurs, 128 Go de RAM, 1 To de SSD gp3	"m6i.8xlarge"(défaut)
Grand	16 processeurs, 64 Go de RAM, 500 Go de SSD	"m6i.4xlarge"(par défaut) m6a.4xlarge m5a.4xlarge m5.4xlarge m4.4xlarge
Moyen	8 processeurs, 32 Go de RAM, 200 Go de SSD	"m6i.2xlarge"(par défaut) m6a.2xlarge m5a.2xlarge m5.2xlarge m4.2xlarge
Petit	8 processeurs, 16 Go de RAM, 100 Go de SSD	"c6a.2xlarge"(par défaut) c5a.2xlarge c5.2xlarge c4.2xlarge

Types d'instances Azure

Taille du système	Spécifications	Type d'instance
Très grand	32 processeurs, 128 Go de RAM, disque système (2 048 Gio, débit minimal de 250 Mo/s) et disque de données (SSD 1 Tio, débit minimal de 750 Mo/s)	"Standard_D32_v3"(défaut)
Grand	16 processeurs, 64 Go de RAM, 500 Go de SSD	"Standard_D16s_v3"(défaut)

Types d'instances GCP

Taille du système	Spécifications	Type d'instance
Grand	16 processeurs, 64 Go de RAM, 500 Go de SSD	"n2-standard-16"(par défaut) n2d-standard-16 n1-standard-16

Métadonnées collectées à partir de sources de données dans la NetApp Data Classification

NetApp Data Classification collecte certaines métadonnées lors de l'exécution d'analyses de classification sur les données de vos sources de données et systèmes. La classification des données peut accéder à la plupart des métadonnées dont nous avons besoin pour classer vos données, mais il existe certaines sources pour lesquelles nous ne pouvons pas accéder aux données dont nous avons besoin.

	Métadonnées	CIFS	NFS
Horodatages	*Heure de création*	disponible	Non disponible (non pris en charge sous Linux)
	Heure du dernier accès	disponible	disponible
	Heure de la dernière modification	disponible	disponible
Autorisations	*Ouvrir les autorisations*	Si le groupe « TOUT LE MONDE » a accès au fichier, il est considéré comme « Ouvert à l'organisation »	Si « Autres » a accès au fichier, il est considéré comme « Ouvert à l'organisation »
	Accès utilisateurs/groupes	Les informations sur les utilisateurs et les groupes sont extraites de LDAP	Non disponible (les utilisateurs NFS sont généralement gérés localement sur le serveur, par conséquent, le même individu peut avoir un UID différent sur chaque serveur)

- La classification des données n'extrait pas la « dernière heure d'accès » des sources de données de la base de données.
- Les anciennes versions du système d'exploitation Windows (par exemple, Windows 7 et Windows 8) désactivent par défaut la collecte de l'attribut « heure du dernier accès » car cela peut avoir un impact sur les performances du système. Lorsque cet attribut n'est pas collecté, les analyses de classification des données basées sur « l'heure du dernier accès » seront affectées. Vous pouvez activer la collecte de l'heure du dernier accès sur ces anciens systèmes Windows si nécessaire.

Horodatage du dernier accès

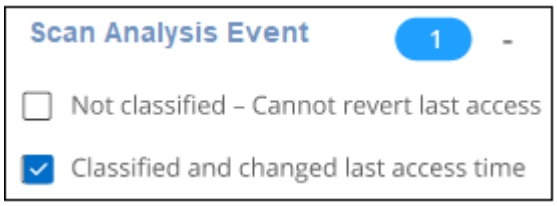
Lorsque Data Classification extrait des données à partir de partages de fichiers, le système d'exploitation considère qu'il accède aux données et modifie l'« heure du dernier accès » en conséquence. Après l'analyse, la classification des données tente de rétablir l'heure du dernier accès à l'horodatage d'origine. Si la classification des données ne dispose pas d'autorisations d'attributs d'écriture dans CIFS ou d'autorisations

d'écriture dans NFS, le système ne peut pas rétablir l'heure du dernier accès à l'horodatage d'origine. Les volumes ONTAP configurés avec SnapLock disposent d'autorisations en lecture seule et ne peuvent pas non plus rétablir l'heure du dernier accès à l'horodatage d'origine.

Par défaut, si Data Classification ne dispose pas de ces autorisations, le système n'analysera pas ces fichiers dans vos volumes car Data Classification ne peut pas rétablir l'« heure du dernier accès » à l'horodatage d'origine. Cependant, si vous ne vous souciez pas de savoir si l'heure du dernier accès est réinitialisée à l'heure d'origine dans vos fichiers, vous pouvez sélectionner le commutateur **Analyser en cas d'absence d'autorisations « attributs d'écriture »** en bas de la page de configuration afin que la classification des données analyse les volumes quelles que soient les autorisations.

Cette fonctionnalité s'applique aux systèmes ONTAP sur site, à Cloud Volumes ONTAP, à Azure NetApp Files, à Amazon FSx for NetApp ONTAP et aux partages de fichiers tiers.

Il existe un filtre dans la page Investigation appelé *Événement d'analyse d'analyse* qui vous permet d'afficher soit les fichiers qui n'ont pas été classés parce que la classification des données n'a pas pu revenir à l'heure du dernier accès, soit les fichiers qui ont été classés même si la classification des données n'a pas pu revenir à l'heure du dernier accès.



Les sélections de filtres sont :

- « Non classé – Impossible de revenir à l'heure du dernier accès » – Ceci affiche les fichiers qui n'ont pas été classés en raison d'autorisations d'écriture manquantes.
- « Heure du dernier accès classé et mis à jour » : cela affiche les fichiers qui ont été classés et la classification des données n'a pas pu réinitialiser l'heure du dernier accès à la date d'origine. Ce filtre n'est pertinent que pour les environnements dans lesquels vous avez activé **Analyser en cas d'absence d'autorisations « attributs d'écriture »**.

Si nécessaire, vous pouvez exporter ces résultats vers un rapport afin de voir quels fichiers sont ou ne sont pas analysés en raison des autorisations. "En savoir plus sur les rapports d'enquête sur les données".

Connectez-vous au système de NetApp Data Classification

Vous devez vous connecter au système de NetApp Data Classification pour pouvoir accéder aux fichiers journaux ou modifier les fichiers de configuration.

Lorsque Data Classification est installé sur une machine Linux dans vos locaux ou sur une machine Linux que vous avez déployée dans le cloud, vous pouvez accéder directement au fichier de configuration et au script.

Lorsque la classification des données est déployée dans le cloud, vous devez vous connecter en SSH à l'instance de classification des données. Vous vous connectez au système via SSH en saisissant le nom d'utilisateur et le mot de passe, ou en utilisant la clé SSH que vous avez fournie lors de l'installation de l'agent de console. La commande SSH est :

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

- `<path_to_the_ssh_key>`= emplacement des clés d'authentification ssh
- `<machine_user>`:
  - Pour AWS : utilisez <ec2-user>
  - Pour Azure : utilisez l'utilisateur créé pour l'instance de la console
  - Pour GCP : utilisez l'utilisateur créé pour l'instance de la console
- `<datasense_ip>`= Adresse IP de l'instance de la machine virtuelle

Vous devez modifier les règles entrantes du groupe de sécurité pour accéder au système dans le cloud. Pour plus de détails, voir :

- "Règles de groupe de sécurité dans AWS"
- "Règles de groupe de sécurité dans Azure"
- "Règles de pare-feu dans Google Cloud"

API de NetApp Data Classification

Les fonctionnalités de NetApp Data Classification disponibles via l'interface utilisateur Web sont également disponibles via l'API REST.

Il existe quatre catégories définies dans la classification des données qui correspondent aux onglets de l'interface utilisateur :

- Enquête
- Conformité
- Gouvernance
- Configuration

Les API de la documentation Swagger vous permettent de rechercher, d'agréger des données, de suivre vos analyses et d'effectuer des actions telles que copier, déplacer et supprimer.

Aperçu

L'API vous permet d'exécuter les fonctions suivantes :

- Informations sur l'exportation
  - Tout ce qui est disponible dans l'interface utilisateur peut être exporté via l'API (à l'exception des rapports)

    - Les données sont exportées au format JSON (faciles à analyser et à transmettre à des applications tierces, comme Splunk)
- Créez des requêtes à l'aide d'instructions « AND » et « OR », incluez et excluez des informations, et bien plus encore.

  Par exemple, vous pouvez localiser des fichiers *sans* informations personnelles identifiables (PII) spécifiques (fonctionnalité non disponible dans l'interface utilisateur). Vous pouvez également exclure des champs spécifiques de l'opération d'exportation.

- Effectuer des actions
    - Mettre à jour les informations d'identification CIFS
    - Afficher et annuler les actions
    - Réanalyser les répertoires
    - Exporter des données

L'API est sécurisée et utilise la même méthode d'authentification que l'interface utilisateur. Vous trouverez des informations sur l'authentification dans le"Documentation REST API" .

Accéder à la référence de l'API Swagger

Pour accéder à Swagger, vous aurez besoin de l'adresse IP de votre instance de classification des données. Dans le cas d'un déploiement cloud, vous utiliserez l'adresse IP publique. Ensuite, vous devrez accéder à ce point de terminaison :

https://<classification_ip>/documentation

Exemple utilisant les API

L'exemple suivant montre un appel d'API pour copier des fichiers.

Demande d'API

Vous devrez d'abord obtenir tous les champs et options pertinents pour qu'un système puisse afficher tous les filtres dans l'onglet d'enquête.

```
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"
```

Réponse

```
{
  "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"
    }
  ]
}
```

Nous utiliserons cette réponse dans nos paramètres de requête pour filtrer les fichiers souhaités que nous souhaitons copier.

Vous pouvez appliquer une action sur plusieurs éléments. Les types d'actions pris en charge incluent : déplacer, supprimer et copier.

Nous allons créer l'action de copie :

Demande d'API

Cette API suivante est cette API d'action et elle vous permet de créer plusieurs actions.

```
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"]}]}}"
```

Réponse

La réponse renverra l'objet d'action, vous pouvez donc utiliser les API get et delete pour obtenir l'état de l'action ou pour l'annuler.


```
{
  "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"
}
```