Referencia

NetApp Data Classification

NetApp
January 14, 2026

Tabla de contenidos

Referencia

Tipos de instancias de NetApp Data Classification compatibles

El software de NetApp Data Classification debe ejecutarse en un host que cumpla con los requisitos específicos del sistema operativo, requisitos de RAM, requisitos de software, etc. Al implementar la clasificación de datos en la nube, recomendamos utilizar un sistema con características "grandes" para obtener una funcionalidad completa.

Puede implementar la clasificación de datos en un sistema con menos CPU y menos RAM, pero existen algunas limitaciones al utilizar estos sistemas menos potentes. "Conozca estas limitaciones" .

En las siguientes tablas, si el sistema marcado como "predeterminado" no está disponible en la región donde está instalando Data Classification, se implementará el siguiente sistema de la tabla.

Tipos de instancias de AWS

Tamaño del sistema	Especificaciones	Tipo de instancia
Extra grande	32 CPU, 128 GB de RAM, 1 TiB gp3 SSD	"m6i.8xlarge"(por defecto)
Grande	16 CPU, 64 GB de RAM, SSD de 500 GiB	"m6i.4xlarge"(predeterminado) m6a.4xlarge m5a.4xlarge m5.4xlarge m4.4xlarge
Medio	8 CPU, 32 GB de RAM, SSD de 200 GiB	"m6i.2xlarge"(predeterminado) m6a.2xlarge m5a.2xlarge m5.2xlarge m4.2xlarge
Pequeño	8 CPU, 16 GB de RAM, SSD de 100 GiB	"c6a.2xlarge"(predeterminado) c5a.2xlarge c5.2xlarge c4.2xlarge

Tipos de instancias de Azure

Tamaño del sistema	Especificaciones	Tipo de instancia
Extra grande	32 CPU, 128 GB de RAM, disco de SO (2048 GiB, rendimiento mínimo de 250 MB/s) y disco de datos (SSD de 1 TiB, rendimiento mínimo de 750 MB/s)	"Standard_D32_v3"(por defecto)
Grande	16 CPU, 64 GB de RAM, SSD de 500 GiB	"Standard_D16s_v3"(por defecto)

Tipos de instancias de GCP

Tamaño del sistema	Especificaciones	Tipo de instancia
Grande	16 CPU, 64 GB de RAM, SSD de 500 GiB	"n2-estándar-16"(predeterminado) n2d-standard-16 n1-standard-16

Metadatos recopilados de fuentes de datos en NetApp Data Classification

NetApp Data Classification recopila ciertos metadatos al realizar escaneos de clasificación en los datos de sus fuentes de datos y sistemas. La clasificación de datos puede acceder a la mayoría de los metadatos que necesitamos para clasificar sus datos, pero hay algunas fuentes en las que no podemos acceder a los datos que necesitamos.

	Metadatos	CIFS	No sé
Marcas de tiempo	*Hora de creación*	Disponible	No disponible (no compatible con Linux)
	Hora del último acceso	Disponible	Disponible
	Hora de la última modificación	Disponible	Disponible
Permisos	*Permisos abiertos*	Si el grupo "TODOS" tiene acceso al archivo, se considera "Abierto a la organización".	Si "Otros" tiene acceso al archivo, se considera "Abierto a la organización".
	Acceso de usuarios/grupos	La información de usuarios y grupos se toma de LDAP	No disponible (los usuarios NFS normalmente se administran localmente en el servidor, por lo tanto, el mismo individuo puede tener un UID diferente en cada servidor)

- La clasificación de datos no extrae la "hora del último acceso" de las fuentes de datos de la base de datos.
- Las versiones anteriores del sistema operativo Windows (por ejemplo, Windows 7 y Windows 8) deshabilitan la recopilación del atributo "hora del último acceso" de forma predeterminada porque puede afectar el rendimiento del sistema. Cuando no se recopila este atributo, los análisis de clasificación de datos que se basan en la "hora del último acceso" se verán afectados. Puede habilitar la recopilación de la hora del último acceso en estos sistemas Windows más antiguos si es necesario.

Marca de tiempo del último acceso

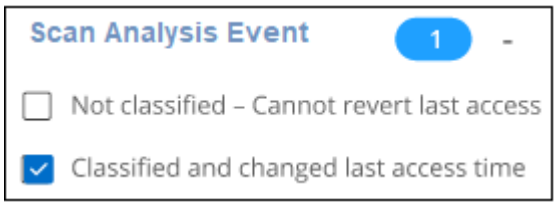
Cuando la clasificación de datos extrae datos de recursos compartidos de archivos, el sistema operativo considera que está accediendo a los datos y cambia la "hora del último acceso" en consecuencia. Después del escaneo, la clasificación de datos intenta revertir la última hora de acceso a la marca de tiempo original. Si la clasificación de datos no tiene permisos de escritura de atributos en CIFS o permisos de escritura en NFS, el sistema no puede revertir la última hora de acceso a la marca de tiempo original. Los volúmenes ONTAP configurados con SnapLock tienen permisos de solo lectura y tampoco pueden revertir la última hora de acceso a la marca de tiempo original.

De forma predeterminada, si la clasificación de datos no tiene estos permisos, el sistema no escaneará esos archivos en sus volúmenes porque la clasificación de datos no puede revertir la "hora del último acceso" a la marca de tiempo original. Sin embargo, si no le importa si la última hora de acceso se restablece a la hora

original en sus archivos, puede seleccionar el interruptor **Escanear cuando faltan permisos de "atributos de escritura"** en la parte inferior de la página de Configuración para que la Clasificación de datos escanee los volúmenes independientemente de los permisos.

Esta funcionalidad es aplicable a sistemas ONTAP locales, Cloud Volumes ONTAP, Azure NetApp Files, Amazon FSx for NetApp ONTAP y recursos compartidos de archivos de terceros.

Hay un filtro en la página de Investigación llamado *Evento de análisis de escaneo* que le permite mostrar los archivos que no se clasificaron porque la Clasificación de datos no pudo revertir la última hora de acceso, o los archivos que se clasificaron aunque la Clasificación de datos no pudo revertir la última hora de acceso.



Las selecciones de filtro son:

- "No clasificado: no se puede revertir la última hora de acceso": esto muestra los archivos que no se clasificaron debido a la falta de permisos de escritura.
- "Hora del último acceso clasificado y actualizado": muestra los archivos que fueron clasificados y la Clasificación de datos no pudo restablecer la hora del último acceso a la fecha original. Este filtro es relevante solo para entornos en los que activó la opción **Escanear cuando faltan permisos de "atributos de escritura"**.

Si es necesario, puede exportar estos resultados a un informe para ver qué archivos se están escaneando o no debido a los permisos. "Obtenga más información sobre los informes de investigación de datos".

Inicie sesión en el sistema de NetApp Data Classification

Debe iniciar sesión en el sistema de NetApp Data Classification para poder acceder a los archivos de registro o editar los archivos de configuración.

Cuando Data Classification está instalado en una máquina Linux en sus instalaciones o en una máquina Linux implementada en la nube, puede acceder directamente al archivo de configuración y al script.

Cuando se implementa la clasificación de datos en la nube, es necesario acceder mediante SSH a la instancia de clasificación de datos. Puede acceder al sistema mediante SSH ingresando el usuario y la contraseña, o utilizando la clave SSH que proporcionó durante la instalación del agente de consola. El comando SSH es:

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

- `<path_to_the_ssh_key>`= ubicación de las claves de autenticación ssh
- `<machine_user>`:
  - Para AWS: utilice <ec2-user>
  - Para Azure: use el usuario creado para la instancia de consola
  - Para GCP: utilice el usuario creado para la instancia de la consola
- `<datasense_ip>`= Dirección IP de la instancia de la máquina virtual

Debe modificar las reglas de entrada del grupo de seguridad para acceder al sistema en la nube. Para más detalles, consulte:

- "Reglas de grupo de seguridad en AWS"
- "Reglas de grupo de seguridad en Azure"
- "Reglas de firewall en Google Cloud"

API de NetApp Data Classification

Las capacidades de NetApp Data Classification disponibles a través de la interfaz de usuario web también están disponibles a través de la API REST.

Hay cuatro categorías definidas dentro de Clasificación de datos que corresponden a las pestañas de la interfaz de usuario:

- Investigación
- Cumplimiento
- Gobernancia
- Configuración

Las API en la documentación de Swagger le permiten buscar, agregar datos, rastrear sus escaneos y realizar acciones que incluyen copiar, mover y eliminar.

Descripción general

La API le permite realizar las siguientes funciones:

- Información de exportación
  - Todo lo que está disponible en la interfaz de usuario se puede exportar a través de la API (con excepción de los informes)
  - Los datos se exportan en formato JSON (fácil de analizar y enviar a aplicaciones de terceros, como Splunk)
- Cree consultas utilizando declaraciones "AND" y "OR", incluya y excluya información, y más.

  Por ejemplo, puede localizar archivos *sin* información personal identificable (PII) específica (funcionalidad no disponible en la interfaz de usuario). También puede excluir campos específicos para la operación de exportación.

- Realizar acciones

- Actualizar las credenciales de CIFS
- Ver y cancelar acciones
- Volver a escanear directorios
- Exportar datos

La API es segura y utiliza el mismo método de autenticación que la UI. Puede encontrar información sobre la autenticación en el"Documentación de REST API" .

Acceder a la referencia de la API de Swagger

Para ingresar a Swagger, necesitará la dirección IP de su instancia de clasificación de datos. En el caso de una implementación en la nube, utilizará la dirección IP pública. Luego tendrás que acceder a este punto final:

https://<ip_de_clasificación>/documentación

Ejemplo de uso de las API

El siguiente ejemplo muestra una llamada API para copiar archivos.

Solicitud de API

Inicialmente necesitará obtener todos los campos y opciones relevantes para que un sistema pueda ver todos los filtros en la pestaña de investigación.

```
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"
```

Respuesta


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

Usaremos esa respuesta en nuestros parámetros de solicitud para filtrar los archivos que queremos copiar.

Puede aplicar una acción a varios elementos. Los tipos de acciones admitidos incluyen: mover, eliminar y copiar.

Crearemos la acción de copia:

Solicitud de API

La siguiente API es la API de acción y le permite crear múltiples acciones.

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

Respuesta

La respuesta devolverá el objeto de acción, por lo que puede utilizar las API de obtención y eliminación para obtener el estado de la acción o cancelarla.

```
{
  "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"
}
```

Información de copyright

Información de la marca comercial