Reference

NetApp Data Classification

NetApp
February 11, 2026

Table of Contents

Reference

Supported NetApp Data Classification instance types

NetApp Data Classification software must run on a host that meets specific operating system requirements, RAM requirements, software requirements, and so on. When deploying Data Classification in the cloud, we recommend that you use a system with the "large" characteristics for full functionality.

You can deploy Data Classification on a system with fewer CPUs and less RAM, but there are some limitations when using these less powerful systems. Learn about these limitations.

In the following tables, if the system marked as "default" is not available in the region where you are installing Data Classification, the next system in the table will be deployed.

AWS instance types

System size	Specs	Instance type
Extra Large	32 CPUs, 128 GB RAM, 1 TiB gp3 SSD	m6i.8xlarge (default)
Large	16 CPUs, 64 GB RAM, 500 GiB SSD	m6i.4xlarge (default) m6a.4xlarge m5a.4xlarge m5.4xlarge m4.4xlarge
Medium	8 CPUs, 32 GB RAM, 200 GiB SSD	m6i.2xlarge (default) m6a.2xlarge m5a.2xlarge m5.2xlarge m4.2xlarge
Small	8 CPUs, 16 GB RAM, 100 GiB SSD	c6a.2xlarge (default) c5a.2xlarge c5.2xlarge c4.2xlarge

Azure instance types

System size	Specs	Instance type
Extra Large	32 CPUs, 128 GB RAM, OS Disk (2,048 GiB, min 250 MB/s throughput), and Data Disk (1 TiB SSD, min 750 MB/s throughput)	Standard_D32_v3 (default)
Large	16 CPUs, 64 GB RAM, 500 GiB SSD	Standard_D16s_v3 (default)

GCP instance types

System size	Specs	Instance type
Large	16 CPUs, 64 GB RAM, 500 GiB SSD	n2-standard-16 (default) n2d-standard-16 n1-standard-16

Metadata collected from data sources in NetApp Data Classification

NetApp Data Classification collects certain metadata when performing classification scans on the data from your data sources and systems. Data Classification can access most of the metadata we need to classify your data, but there are some sources where we are unable to access the data we need.

	Metadata	CIFS	NFS
Time stamps	*Creation time*	Available	Not available (Unsupported in Linux)
	Last access time	Available	Available
	Last modify time	Available	Available
Permissions	*Open permissions*	If "EVERYONE" group has access to the file, it is considered "Open to organization"	If "Others" has access to the file, it is considered "Open to organization"
	Users/group access	Users and group information is taken from LDAP	Not available (NFS users are usually managed locally on the server, therefore, the same individual can have a different UID in each server)

- Data Classification does not extract the "last accessed time" from the database data sources.
- Older versions of the Windows OS (for example, Windows 7 and Windows 8) disable the collection of the "last accessed time" attribute by default because it can impact system performance. When this attribute is not collected, Data Classification analytics that are based on "last accessed time" will be impacted. You can enable the collection of the last access time on these older Windows systems if needed.

Last access time timestamp

When Data Classification extracts data from file shares, the operating system considers it as accessing the data and it changes the "last access time" accordingly. After scanning, Data Classification attempts to revert the last access time to the original timestamp. If Data Classification doesn’t have write attributes permissions in CIFS, or write permissions in NFS, the system can’t revert the last access time to the original timestamp.

ONTAP volumes configured with SnapLock have read-only permissions and also can't revert the last access time to the original timestamp.

By default, if Data Classification doesn't have these permissions, the system won't scan those files in your volumes because Data Classification can't revert the "last access time" to the original timestamp. However, if you don't care if the last access time is reset to the original time in your files, you can select the **Scan when missing "write attributes" permissions** switch at the bottom of the Configuration page so that Data Classification will scan the volumes regardless of permissions.

This functionality is applicable to on-premises ONTAP systems, Cloud Volumes ONTAP, Azure NetApp Files, Amazon FSx for NetApp ONTAP management, and third-party file shares.

There's a filter in the Investigation page called *Scan Analysis Event* that enables you to display either the files that were not classified because Data Classification couldn't revert the last accessed time, or the files that were classified even though Data Classification couldn't revert the last access time.

The filter selections are:

- "Not classified — Cannot revert last access time" - This shows the files that were not classified due to missing write permissions.
- "Classified and updated last access time" - This shows the files that were classified and Data Classification was unable to reset the last access time back to the original date. This filter is relevant only for environments where you turned **Scan when missing "write attributes" permissions** ON.

If needed, you can export these results to a report so you can see which files are, or aren't, being scanned because of permissions. Learn more about Data Investigation reports.

Log in to the NetApp Data Classification system

You need to log into NetApp Data Classification system so you can access log files or edit configuration files.

When Data Classification is installed on a Linux machine on your premises, or on a Linux machine you deployed in the cloud, you can access the configuration file and script directly.

When Data Classification is deployed in the cloud, you need to SSH to the Data Classification instance. You SSH to the system by entering the user and password, or by using the SSH key you provided during the Console agent installation. The SSH command is:

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

- `<path_to_the_ssh_key>` = location of ssh authentication keys
- `<machine_user>:`
  - For AWS: use the <ec2-user>
  - For Azure: use the user created for the Console instance
  - For GCP: use the user created for the Console instance
- `<datasense_ip>` = IP address of the virtual machine instance

You need to modify the security group inbound rules to access the system in the cloud. For details, see:

- Security group rules in AWS
- Security group rules in Azure
- Firewall rules in Google Cloud

NetApp Data Classification APIs

The NetApp Data Classification capabilities available through the web UI are also available through the REST API.

There are four categories defined within Data Classification that correspond to the tabs in the UI:

- Investigation
- Compliance
- Governance
- Configuration

The APIs in the Swagger documentation allow you to search, aggregate data, track your scans, and perform actions including copy, move, and delete.

Overview

The API enables you to perform the following functions:

- Export information
  - Everything that is available in the UI can be exported via the API (with the exception of reports)
  - Data is exported in a JSON format (easy to parse and push to 3rd party applications, like Splunk)
- Create queries using "AND" and "OR" statements, include and exclude information, and more.

  For example, you can locate files *without* specific Personal Identifiable Information (PII) (functionality not available in the UI). You can also exclude specific fields for the export operation.

- Perform actions
  - Update CIFS credentials
  - View and cancel actions
  - Re-scan directories
  - Export data

The API is secure and it uses the same authentication method as the UI. You can find information on the authentication in the REST API documentation.

Accessing the Swagger API reference

To get into Swagger you'll need the IP address of the your Data Classification instance. In the case of a cloud deployment you'll use the public IP address. Then you'll need to get into this endpoint:

https://<classification_ip>/documentation

Example using the APIs

The following example shows an API call to copy files.

API Request

You'll initially need to get all the relevant fields and options for a system to view all of the filters in the investigation tab.

```
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"
```

Response

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

We will use that response in our request parameters to filter the desired files we want to copy.

You can apply an action on multiple items. Supported action types include: move, delete, and copy.

We will create the copy action:

API Request

This next API is that action API and it allows you to create multiple 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"]}]}}"
```

Response

The response will return the action object, so you can use the get and delete APIs to get status about the action, or to cancel it.

```
{
  "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"
}
```

Copyright information

Trademark information