참조

NetApp Data Classification

NetApp
January 14, 2026

목차

참조

지원되는 NetApp Data Classification 인스턴스 유형

NetApp Data Classification 소프트웨어는 특정 운영 체제 요구 사항, RAM 요구 사항, 소프트웨어 요구 사항 등을 충족하는 호스트에서 실행되어야 합니다. 클라우드에서 데이터 분류를 배포할 때 모든 기능을 활용하려면 "대규모" 특성을 갖춘 시스템을 사용하는 것이 좋습니다.

CPU와 RAM이 적은 시스템에도 데이터 분류를 배포할 수 있지만, 이러한 덜 강력한 시스템을 사용할 경우 몇 가지 제한 사항이 있습니다. "이러한 제한 사항에 대해 알아보세요" .

다음 표에서 "기본"으로 표시된 시스템을 데이터 분류를 설치하는 지역에서 사용할 수 없는 경우 표의 다음 시스템이 배포됩니다.

AWS 인스턴스 유형

시스템 크기	명세서	인스턴스 유형
특대	32개 CPU, 128GB RAM, 1TiB gp3 SSD	"m6i.8xlarge"(기본)
크기가 큰	CPU 16개, 64GB RAM, 500GiB SSD	"m6i.4xlarge"(기본값) m6a.4xlarge m5a.4xlarge m5.4xlarge m4.4xlarge
중간	CPU 8개, 32GB RAM, 200GiB SSD	"m6i.2xlarge"(기본값) m6a.2xlarge m5a.2xlarge m5.2xlarge m4.2xlarge
작은	CPU 8개, 16GB RAM, 100GiB SSD	"c6a.2xlarge"(기본값) c5a.2xlarge c5.2xlarge c4.2xlarge

Azure 인스턴스 유형

시스템 크기	명세서	인스턴스 유형
특대	32개 CPU, 128GB RAM, OS 디스크(2,048GiB, 최소 250MB/s 처리량), 데이터 디스크(1TiB SSD, 최소 750MB/s 처리량)	"Standard_D32_v3"(기본)
크기가 큰	CPU 16개, 64GB RAM, 500GiB SSD	"Standard_D16s_v3"(기본)

GCP 인스턴스 유형

시스템 크기	명세서	인스턴스 유형
크기가 큰	CPU 16개, 64GB RAM, 500GiB SSD	"n2-표준-16"(기본값) n2d-standard-16 n1-standard-16

NetApp Data Classification 데이터 소스에서 수집된 메타데이터

NetApp Data Classification 데이터 소스와 시스템의 데이터에 대한 분류 스캔을 수행할 때 특정 메타데이터를 수집합니다. 데이터 분류는 데이터 분류에 필요한 대부분의 메타데이터에 접근할 수 있지만, 필요한 데이터에 접근할 수 없는 일부 소스도 있습니다.

	메타데이터	CIFS	NFS
타임스탬프	생성 시간	사용 가능	사용할 수 없음(Linux에서는 지원되지 않음)
	마지막 접속 시간	사용 가능	사용 가능
	마지막 수정 시간	사용 가능	사용 가능
권한	열기 권한	"EVERYONE" 그룹이 파일에 액세스할 수 있는 경우 해당 파일은 "조직에 공개"로 간주됩니다.	"기타"가 파일에 액세스할 수 있는 경우 해당 파일은 "조직에 공개됨"으로 간주됩니다.
	사용자/그룹 액세스	사용자 및 그룹 정보는 LDAP에서 가져옵니다.	사용할 수 없음(NFS 사용자는 일반적으로 서버에서 로컬로 관리되므로 동일한 개인이 각 서버에서 다른 UID를 가질 수 있음)

- 데이터 분류는 데이터베이스 데이터 소스에서 "마지막 액세스 시간"을 추출하지 않습니다.
- 이전 버전의 Windows OS(예: Windows 7 및 Windows 8)는 시스템 성능에 영향을 줄 수 있으므로 기본적으로 "마지막 액세스 시간" 특성 수집을 비활성화합니다. 이 속성이 수집되지 않으면 "마지막 액세스 시간"을 기반으로 하는 데이터 분류 분석에 영향을 미칩니다. 필요한 경우 이러한 이전 Windows 시스템에서 마지막 액세스 시간 수집을 활성화할 수 있습니다.

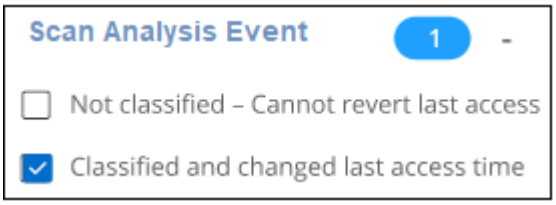
마지막 액세스 시간 타임스탬프

데이터 분류가 파일 공유에서 데이터를 추출할 때, 운영 체제는 이를 데이터에 액세스하는 것으로 간주하고 그에 따라 "마지막 액세스 시간"을 변경합니다. 스캐닝 후, 데이터 분류는 마지막 액세스 시간을 원래 타임스탬프로 되돌리려고 시도합니다. 데이터 분류에 CIFS의 쓰기 속성 권한이 없거나 NFS의 쓰기 권한이 없는 경우 시스템은 마지막 액세스 시간을 원래 타임스탬프로 되돌릴 수 없습니다. SnapLock 으로 구성된 ONTAP 볼륨은 읽기 전용 권한을 가지며 마지막 액세스 시간을 원래 타임스탬프로 되돌릴 수 없습니다.

기본적으로 데이터 분류에 이러한 권한이 없으면 시스템은 볼륨에서 해당 파일을 검사하지 않습니다. 데이터 분류는 "마지막 액세스 시간"을 원래 타임스탬프로 되돌릴 수 없기 때문입니다. 하지만 파일의 마지막 액세스 시간이 원래 시간으로 재설정되는 것이 문제가 되지 않는다면 구성 페이지 하단의 **"쓰기 속성"** 권한이 없는 경우 검사 스위치를 선택하면 데이터 분류가 권한에 관계없이 볼륨을 검사합니다.

이 기능은 온프레미스 ONTAP 시스템, Cloud Volumes ONTAP, Azure NetApp Files, Amazon FSx for NetApp ONTAP 관리 및 타사 파일 공유에 적용할 수 있습니다.

조사 페이지에는 _스캔 분석 이벤트_라는 필터가 있는데, 이를 사용하면 데이터 분류가 마지막 액세스 시간을 되돌릴 수 없어 분류되지 않은 파일이나 데이터 분류가 마지막 액세스 시간을 되돌릴 수 없어도 분류된 파일을 표시할 수 있습니다.



필터 선택은 다음과 같습니다.

- "분류되지 않음 - 마지막 액세스 시간을 되돌릴 수 없음" - 쓰기 권한이 없어 분류되지 않은 파일을 표시합니다.
- "분류 및 업데이트된 마지막 액세스 시간" - 이는 분류된 파일을 보여주며, 데이터 분류는 마지막 액세스 시간을 원래 날짜로 재설정하지 못했습니다. 이 필터는 *"쓰기 속성" 권한이 없는 경우 검사*를 켜둔 환경에만 적용됩니다.

필요한 경우 이러한 결과를 보고서로 내보내어 권한 때문에 어떤 파일이 검사되고 있는지, 검사되지 않고 있는지 확인할 수 있습니다. "데이터 조사 보고서에 대해 자세히 알아보세요".

NetApp Data Classification 시스템에 로그인하세요

로그 파일에 액세스하거나 구성 파일을 편집하려면 NetApp Data Classification 시스템에 로그인해야 합니다.

데이터 분류가 사내 Linux 머신이나 클라우드에 배포한 Linux 머신에 설치된 경우 구성 파일과 스크립트에 직접 액세스할 수 있습니다.

데이터 분류가 클라우드에 배포되는 경우 데이터 분류 인스턴스에 SSH를 사용해야 합니다. 사용자 이름과 비밀번호를 입력하거나 콘솔 에이전트 설치 중에 제공한 SSH 키를 사용하여 시스템에 SSH를 실행합니다. SSH 명령은 다음과 같습니다.

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

- `<path_to_the_ssh_key>`= ssh 인증 키의 위치
- `<machine_user>`:
  - AWS의 경우: <ec2-user>를 사용하세요.

- Azure의 경우: 콘솔 인스턴스에 대해 생성된 사용자를 사용합니다.
- GCP의 경우: 콘솔 인스턴스에 대해 생성된 사용자를 사용합니다.

- `<datasense_ip>`= 가상 머신 인스턴스의 IP 주소

클라우드 시스템에 액세스하려면 보안 그룹 인바운드 규칙을 수정해야 합니다. 자세한 내용은 다음을 참조하세요.

- "AWS의 보안 그룹 규칙"
- "Azure의 보안 그룹 규칙"
- "Google Cloud의 방화벽 규칙"

NetApp Data Classification API

웹 UI를 통해 제공되는 NetApp Data Classification 기능은 REST API를 통해서도 사용할 수 있습니다.

UI의 탭에 해당하는 데이터 분류에는 4가지 범주가 정의되어 있습니다.

- 조사
- 규정 준수
- 통치
- 구성

Swagger 문서의 API를 사용하면 검색, 데이터 집계, 스캔 추적이 가능하며 복사, 이동, 삭제 등의 작업을 수행할 수 있습니다.

개요

API를 사용하면 다음 기능을 수행할 수 있습니다.

- 수출 정보
  - UI에서 사용 가능한 모든 것은 API를 통해 내보낼 수 있습니다(보고서 제외)
  - 데이터는 JSON 형식으로 내보내집니다(Splunk와 같은 타사 애플리케이션에 쉽게 구문 분석하고 푸시할 수 있음)
- "AND" 및 "OR" 문을 사용하여 쿼리를 만들고, 정보를 포함하거나 제외하는 등의 작업을 수행합니다.

  예를 들어, 특정 개인 식별 정보(PII)가 없는 파일을 찾을 수 있습니다(이 기능은 UI에서 사용할 수 없습니다). 내보내기 작업에서 특정 필드를 제외할 수도 있습니다.

- 작업 수행
  - CIFS 자격 증명 업데이트
  - 작업 보기 및 취소
  - 디렉토리 재스캔
  - 데이터 내보내기

API는 안전하며 UI와 동일한 인증 방법을 사용합니다. 인증에 대한 정보는 다음에서 찾을 수 있습니다."REST API 설명서" .

Swagger API 참조에 액세스하기

Swagger에 들어가려면 데이터 분류 인스턴스의 IP 주소가 필요합니다. 클라우드에 배포하는 경우 공용 IP 주소를 사용합니다. 그러면 다음 엔드포인트로 이동해야 합니다.

https://<분류_IP>/문서

API를 사용한 예

다음 예제는 파일을 복사하는 API 호출을 보여줍니다.

API 요청

조사 탭의 모든 필터를 보려면 먼저 시스템에 필요한 모든 관련 필드와 옵션을 확보해야 합니다.

```
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"
```

응답

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

우리는 복사하려는 원하는 파일을 필터링하기 위해 요청 매개변수에서 해당 응답을 사용할 것입니다.

여러 항목에 작업을 적용할 수 있습니다. 지원되는 작업 유형에는 이동, 삭제, 복사가 있습니다.

복사 작업을 생성합니다.

API 요청

다음 API는 액션 API이며 이를 통해 여러 액션을 생성할 수 있습니다.

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

응답

응답에서는 작업 객체가 반환되므로 get 및 delete API를 사용하여 작업에 대한 상태를 얻거나 작업을 취소할 수 있습니다.

```
{
  "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"
}
```