Macie 제대로 사용해보기

근데 이제 IAM을 곁들인...

발표자 소개

천강민 (rex.chun)

2014 ~ 2019 군대 (개발, 점검, 분석 등)

2019 ~ 2021 넷마블 (사고조사 및 클라우드 보안)

2021 ~ 현재 카카오뱅크 (클라우드 엔지니어)

강의

1. 실무에 바로 적용하는 클라우드 보안 프로그래밍 (Python, Terraform) (인프런)

블로그

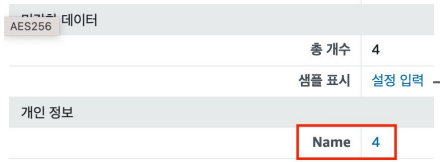
깃허브

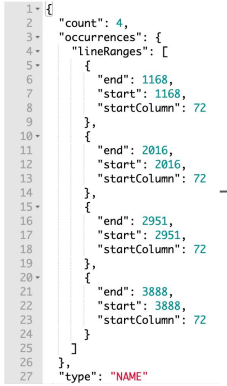
Macie 란?

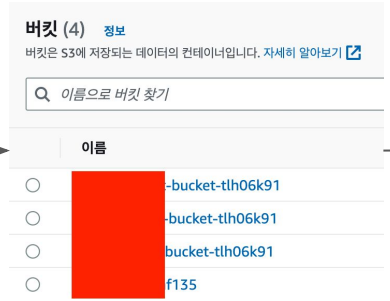

Amazon Macie
민감한 데이터를 대규모로 검색 및 보호
Amazon Macie 시작하기
AWS 프리 티어 가입
S3 버킷 수준 보안 및 액세스 제어 평가 30일 무료 이용
AWS 프리 티어 사용 혜택
민감한 데이터 검색을 규모에 따라 자동화할 수 있습니다.
Amazon S3에 저장된 민감한 데이터에 대해 비용 효율적인 가시성을 확보할 수 있습니다.
Amazon S3 버킷 인벤토리의 보안 및 액세스 제어를 평가할 수 있습니다.
Amazon S3에서 발견된 민감한 데이터에 대한 정보를 실천 가능한 보고 형식으로 제공함으로써 분류 시간을 단축해줍니다.
????

콘솔을 통해 Macie를 사용한다면..

X 파일갯수

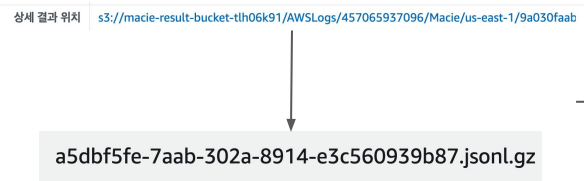


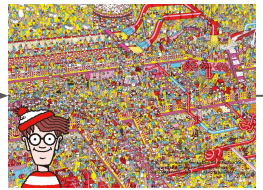




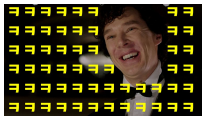
결과

- 민감한 데이터 검색 작업의 실행당 검색 결과: 100,000개 이상 100,(
  이 할당량은 Amazon Macie 콘솔과 Amazon Macie API에만 적용
  각 작업 실행에 대해 생성하는 민감한 데이터 검색 결과 수에 대한 할당
- 민감한 데이터 발견당 탐지 위치: 15개





Amazon S3에서 발견된 민감한 데이터에 대한 정보를 실천 가능한 보고 형식으로 제공함으로써 분류 시간을 단축해줍니다.



처음 Macie를 봤을 때,

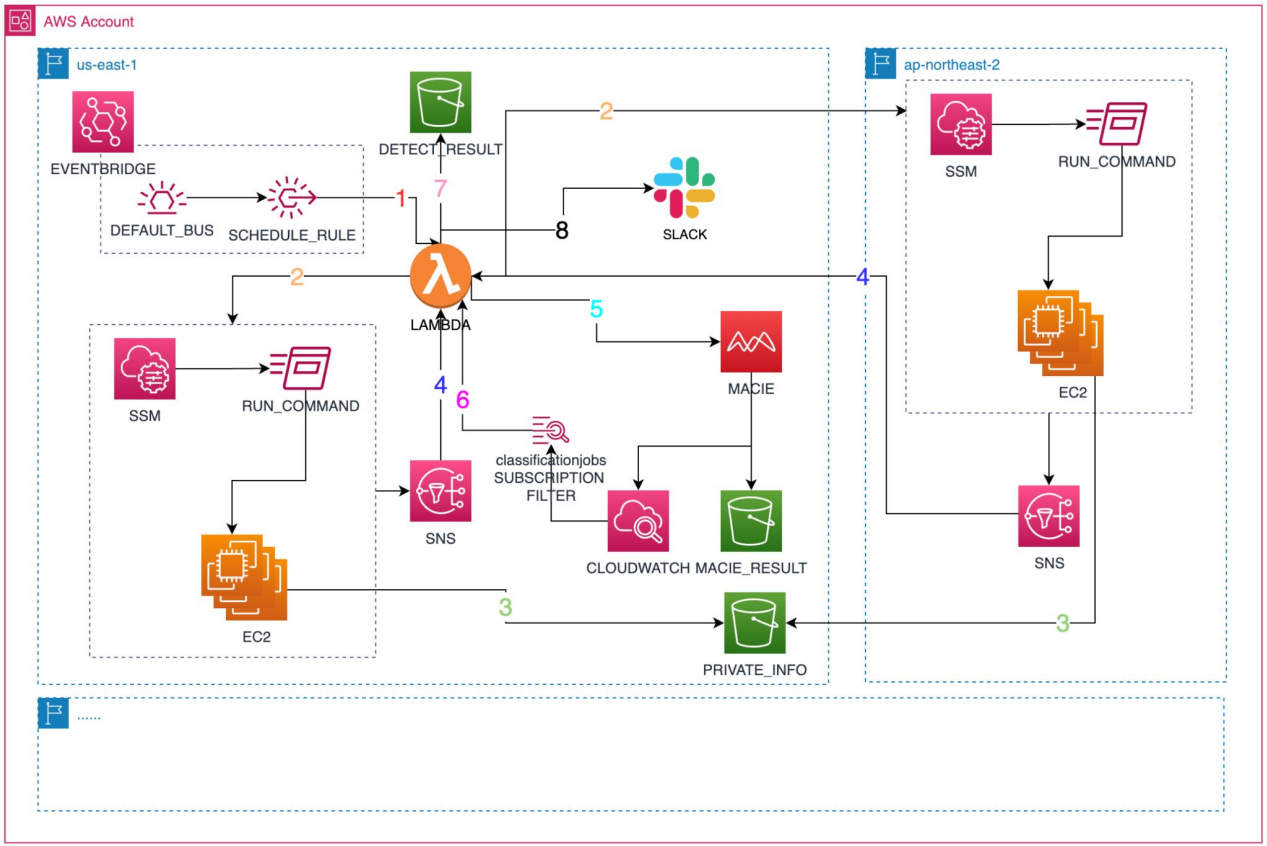
나 : 뭐지? 아니 이런걸 기능이라고 만든건가?

AWS : 뭐지? 저런게 엔지니어라고?

AWS 당신이 옳았어. Macie는 프로그래밍과 함께여야 했다…

~~(사실 대부분의 기능들이…)~~

아키텍처 소개


AWS Account
us-east-1
ap-northeast-2
EVENTBRIDGE
DEFAULT_BUS
SCHEDULE_RULE
DETECT_RESULT
SLACK
LAMBDA
SSM
RUN_COMMAND
EC2
MACIE
classificationjobs
SUBSCRIPTION
FILTER
SNS
CLOUDWATCH
MACIE_RESULT
PRIVATE_INFO
.....

1. EventBridge to Lambda

macie-lambda-warming-rule

규칙 세부 정보 정보

규칙 이름

macie-lambda-warming-rule

설명

이벤트 일정 | 대상 | 모니터링 | 태그

이벤트 일정 정보

고정 비율:

5 minute

입력 상수

```
{
  "warmer": true
}
```

```
import os
import base64, gzip
import uuid
import json
import boto3
import logging
from datetime import date
from botocore.config import Config
from libs.ec2 import get_running_instances, get_available_regions
from libs.macie2 import get_detect_result_from_sensitive_result
from libs.utils import send_message_to_slack, get_slack_message, json\_gzip_to_jsons


logger = logging.getLogger()
logger.setLevel(logging.INFO)

SLACK_HOOK_URL = os.environ["SLACK_HOOK_URL"]
SLACK_CHANNEL = os.environ["SLACK_CHANNEL"]
SNS_ASSUME_ROLE_ARN = os.getenv("SNS_ASSUME_ROLE_ARN")
SNS_TOPIC_ARN = os.getenv("SNS_TOPIC_ARN")
PRIAVTE_INFO_BUCKET = os.getenv("PRIAVTE_INFO_BUCKET")
RESULT_BUCKET = os.getenv("RESULT_BUCKET")
S3_ACCOUNT_ID = os.getenv("S3_ACCOUNT_ID")
EVENTBRIDGE_WARM_ARN = os.getenv("EVENTBRIDGE_WARM_ARN")

s3 = boto3.client("s3", config=Config(region_name="us-east-1"))
macie2 = boto3.client("macie2", config=Config(region_name="us-east-1"))

COMMAND_COMPLETED_REGIONS: list = []
```

macie-inspect-rule

규칙 세부 정보 정보

규칙 이름

macie-inspect-rule

설명

이벤트 일정 | 대상 | 모니터링 | 태그

이벤트 일정 정보

Cron 식

30 10 * * ? *

2. Lambda to SSM Run Command

```
INSTANCE_ID=`wget -q -O - http://169.254.169.254/
latest/meta-data/instance-id`

find / -type f -mtime -1 ! -executable ! -size 0 !
-path "/proc/*" ! -path "/sys/*" ! -path "/run/*" !
-path "/var/log/dmesg" -exec aws s3 cp {} s3://
private-info-bucket-tlh06k91/$INSTANCE_ID{} \; 2> /
dev/null

exit 0
```

3. EC2 to S3

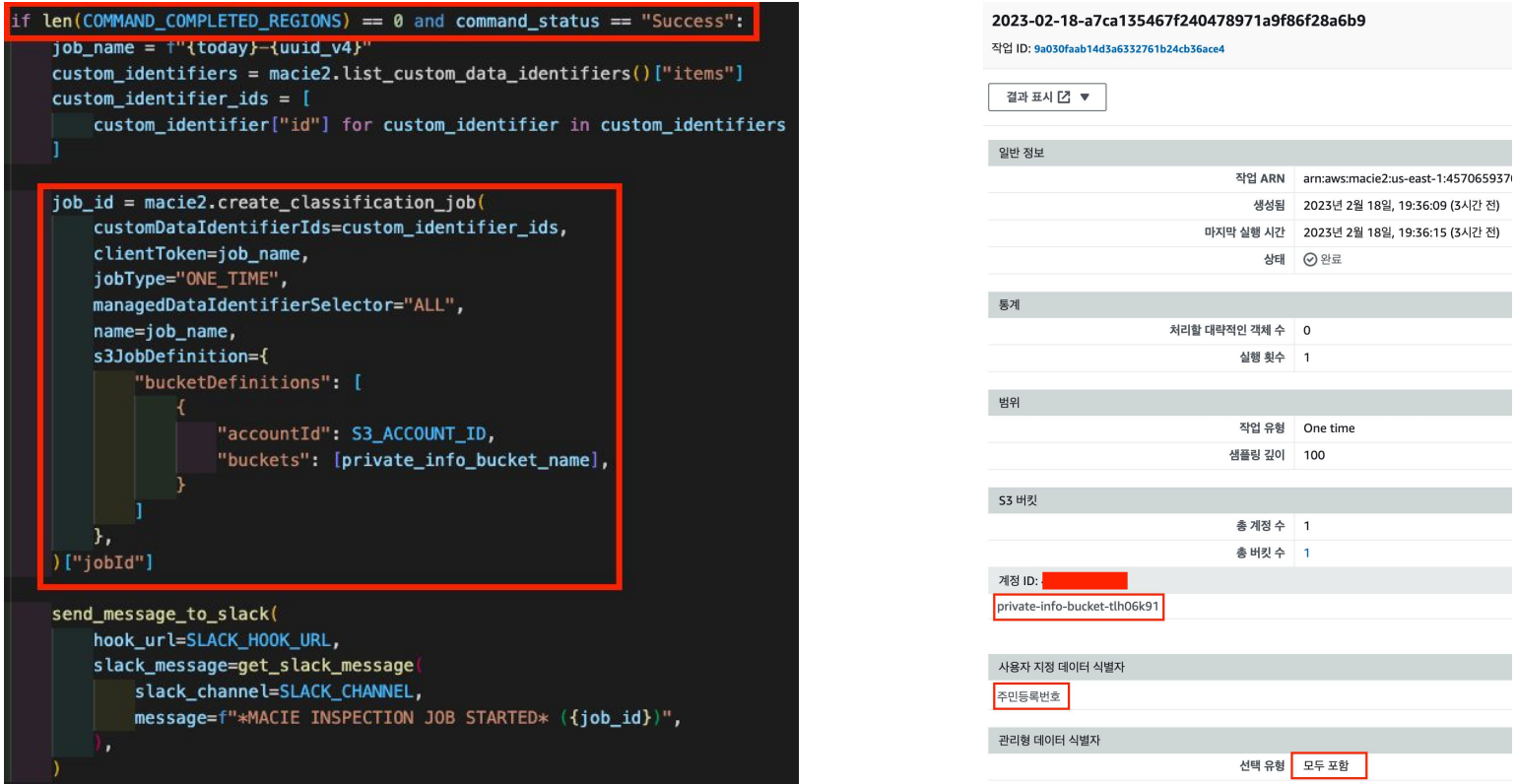
```
Completed 112 Bytes/112 Bytes (1.4 KiB/s) with 1 file(s) remaining upload: ../../etc/resolv.conf to s3://private-info-bucket-tlh06k91/i-0fc900578d3030580/etc/resolv.conf
Completed 158 Bytes/158 Bytes (2.4 KiB/s) with 1 file(s) remaining upload: ../../var/cache/yum/x86_64/2/amzn2-core/mirrorlist.txt to s3://private-info-bucket-tlh06k91/i-0fc900578d3030580/var/cache/yum/x86_64/2/amzn2-core/mirrorlist.txt
Completed 21.4 KiB/21.4 KiB (236.6 KiB/s) with 1 file(s) remaining upload: ../../var/log/wtmp to s3://private-info-bucket-tlh06k91/i-0fc900578d3030580/var/log/wtmp
Completed 15.3 KiB/15.3 KiB (202.0 KiB/s) with 1 file(s) remaining upload: ../../var/log/sa/sa16 to s3://private-info-bucket-tlh06k91/i-0fc900578d3030580/var/log/sa/sa16
```

4/5. Receive run command result and Create Maice job

```python
if len(COMMAND_COMPLETED_REGIONS) == 0 and command_status == "Success":
    job_name = f"{today}-{uuid_v4}"
    custom_identifiers = macie2.list_custom_data_identifiers()["items"]
    custom_identifier_ids = [
        custom_identifier["id"] for custom_identifier in custom_identifiers
    ]

    job_id = macie2.create_classification_job(
        customDataIdentifierIds=custom_identifier_ids,
        clientToken=job_name,
        jobType="ONE_TIME",
        managedDataIdentifierSelector="ALL",
        name=job_name,
        s3JobDefinition={
            "bucketDefinitions": [
                {
                    "accountId": S3_ACCOUNT_ID,
                    "buckets": [private_info_bucket_name],
                }
            ]
        },
    )["jobId"]

    send_message_to_slack(
        hook_url=SLACK_HOOK_URL,
        slack_message=get_slack_message(
            slack_channel=SLACK_CHANNEL,
            message=f"*MACIE INSPECTION JOB STARTED* ({job_id})",
        ),
    )
```

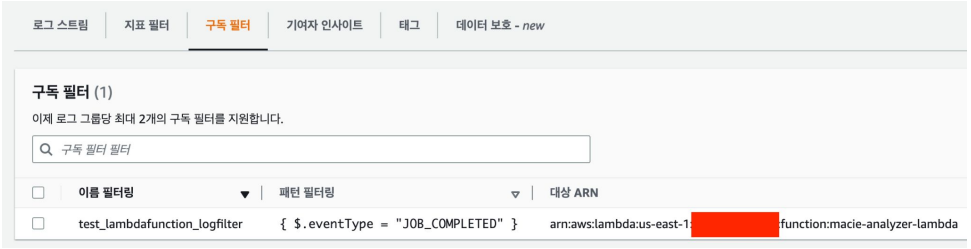


6. Receive job completed event

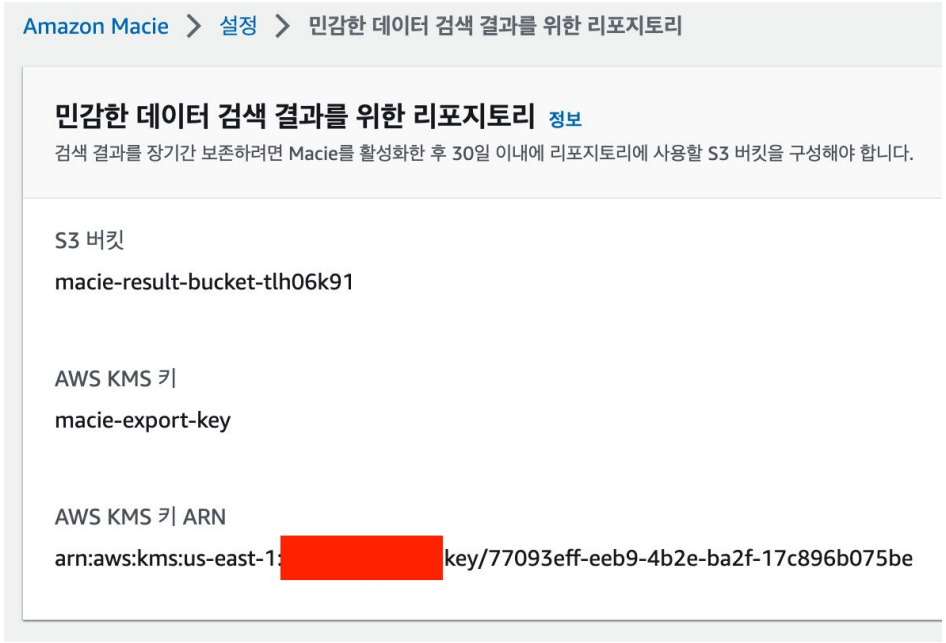

```
{
    "adminAccountId": "[redacted]",
    "jobId": "9a030faab14d3a6332761b24cb36ace4",
    "eventType": "JOB_CREATED",
    "occurredAt": "2023-02-18T10:36:09.965999Z",
    "description": "The job was created.",
    "jobName": "2023-02-18-a7ca135467f240478971a9f86f28a6b9"
}
```

```
{
    "adminAccountId": "[redacted]",
    "jobId": "9a030faab14d3a6332761b24cb36ace4",
    "eventType": "ONE_TIME_JOB_STARTED",
    "occurredAt": "2023-02-18T10:36:15.032159Z",
    "description": "The job started running.",
    "jobName": "2023-02-18-a7ca135467f240478971a9f86f28a6b9",
    "runDate": "2023-02-18T10:36:09.737023Z"
}
```

```
{
    "adminAccountId": "[redacted]",
    "jobId": "9a030faab14d3a6332761b24cb36ace4",
    "eventType": "JOB_COMPLETED",
    "occurredAt": "2023-02-18T10:48:57.642815Z",
    "description": "The job finished running.",
    "jobName": "2023-02-18-a7ca135467f240478971a9f86f28a6b9",
    "runDate": "2023-02-18T10:36:09.737023Z"
}
```

7. Get detailed detect results and Parse objects



private-info-bucket-tlh06k91/i-0f092927d5829b09a/etc/ssh/ssh_host_rsa_key
private-info-bucket-tlh06k91/i-0fae5291351c8c8aa/var/log/messages
private-info-bucket-tlh06k91/i-0fae5291351c8c8aa/var/log/messages
private-info-bucket-tlh06k91/i-0fae5291351c8c8aa/var/log/messages
private-info-bucket-tlh06k91/i-0fae5291351c8c8aa/var/log/messages
private-info-bucket-tlh06k91/i-0fae5291351c8c8aa/var/log/messages
private-info-bucket-tlh06k91/i-0fae5291351c8c8aa/var/log/messages
private-info-bucket-tlh06k91/i-078e738ed5fa3a031/etc/ssh/ssh_host_rsa_key
private-info-bucket-tlh06k91/i-078e738ed5fa3a031/etc/ssh/ssh_host_ed25519_key
private-info-bucket-tlh06k91/i-078e738ed5fa3a031/etc/ssh/ssh_host_ecdsa_key
private-info-bucket-tlh06k91/i-078e738ed5fa3a031/var/lib/yum/rpmdb-indexes/conflicts

SensitiveData:S3Object/Personal

결과 ID: ae1706d7e32d0bb2231776106a42fbdf

Medium The object contains personal information such as first or last names, addresses, or identification numbers. Learn More

개요	
심각도	Medium
리전	us-east-1
계정 ID	[redacted]
리소스	private-info-bucket-tlh06k91/i-0fc900578d3030580/var/log/messages
생성 날짜	2023년 2월 18일, 19:41:17 (3시간 전)
업데이트 날짜	2023년 2월 18일, 19:41:17 (3시간 전)
오리진 유형	SENSITIVE_DATA_DISCOVERY_JOB

결과	
작업 ID	9a030faab14d3a6332761b24cb36ace4
세부 정보	
상태	COMPLETE
분류된 크기	551 KB
MIME 유형	text/plain
상세 결과 위치	s3://macie-result-bucket-tlh06k91/AWSLogs[redacted]Macie/us-e
민감한 데이터	
총 개수	6
샘플 표시	설정 입력
개인 정보	
Name	6

영향을 받는 리소스(S3 버킷)	
버킷 이름	private-info-bucket-tlh06k91
퍼블릭 액세스	NOT_PUBLIC
버킷 정책에 필요한 암호화	아니요
기본 암호화	AES256
생성 날짜	2023년 2월 11일, 16:24:13 (7일 전)
소유자	[redacted]

영향을 받는 리소스(S3 객체)	
키	i-0fc900578d3030580/var/log/messages

```
"sensitiveData": [
    {
      "category": "PERSONAL_INFORMATION",
      "detections": [
         {
            "count": 30,
            "occurrences": {
               "cells": [
                  {
                     "cellReference": null,
                     "column": 1,
                     "columnName": "SSN",
                     "row": 2
                  },
                  {
                     "cellReference": null,
                     "column": 1,
                     "columnName": "SSN",
                     "row": 3
                  },
                  {
                     "cellReference": null,
                     "column": 1,
                     "columnName": "SSN",
                     "row": 4
                  }
               ]
            },
            "type": "USA_SOCIAL_SECURITY_NUMBER"
         }
```


- `cells` 배열 — 이 배열은 Microsoft Excel 통합 문서, CSV 파일 및 TSV 파일에 적용됩니다. 이 배열의 객체는 Macie가 민감한 데이터의 발생을 감지한 셀이나 필드를 지정합니다.
- `lineRanges` 배열 — 이 배열은 이메일 메시지 (EML) 파일과 CSV, JSON, JSON 라인 및 TSV 파일 이외의 바이너리가 아닌 텍스트 파일 (예: HTML, TXT 및 XML 파일) 에 적용됩니다. 이 배열의 객체는 Macie가 민감한 데이터의 발생을 감지한 라인 또는 포함 라인 범위와 지정된 라인 또는 여러 라인에서의 데이터 위치를 지정합니다.

  경우에 따라 배열의 객체는 다른 유형의 `lineRanges` 배열에서 지원하는 파일 형식 또는 저장소 형식으로 중요한 데이터 탐지 위치를 지정합니다. 이러한 경우는 파일 내 주석과 같이 구조화되지 않은 파일의 구조화되지 않은 부분에서의 탐지, Macie가 일반 텍스트로 분석하는 잘못된 형식의 파일에서의 탐지, Macie가 민감한 데이터를 감지한 열 이름이 하나 이상 있는 CSV 또는 TSV 파일에서의 탐지 등입니다.
- `offsetRanges array` — 이 배열은 future 사용할 수 있도록 예약되어 있습니다. 이 배열이 있는 경우 해당 배열의 값은 null입니다.
- `pages` 배열 — 이 배열은 Adobe 휴대용 문서 형식 (PDF) 파일에 적용됩니다. 이 배열의 객체는 Macie가 민감한 데이터의 발생을 감지한 페이지를 지정합니다.
- `records` 배열 — 이 배열은 아파치 아브로 객체 컨테이너, 아파치 파켓 파일, JSON 파일 및 JSON 라인 파일에 적용됩니다. Avro 개체 컨테이너 및 Parquet 파일의 경우 이 배열의 개체는 레코드 색인과 Macie가 민감한 데이터의 발생을 감지한 레코드의 필드 경로를 지정합니다. JSON 및 JSON 라인 파일의 경우 이 배열의 객체는 Macie가 민감한 데이터의 발생을 감지한 필드 또는 배열의 경로를 지정합니다. JSON Lines 파일의 경우 데이터가 포함된 행의 인덱스도 지정합니다.

셀형 배열

적용 대상: 마이크로소프트 엑셀 워크북, CSV 파일 및 TSV 파일

`cells` 배열에서 `Cell` 객체는 Macie가 민감한 데이터의 발생을 감지한 셀이나 필드를 지정합니다. 다음 표에서는 `Cell` 객체에 있는 각 필드의 용도를 설명합니다.

필드	유형	설명
cellReference	문자열	절대 셀 참조로서 발생 항목이 포함된 셀의 위치입니다. 이 필드는 Excel 통합 문서에만 적용됩니다. CSV 및 TSV 파일의 경우 이 값은 null입니다.
column	Integer	발생이 들어 있는 열의 열번호입니다. Excel 통합 문서의 경우 이 값은 열 식별자의 알파벳 문자 (예: A열,1 B열 등2) 와 상관 관계가 있습니다.
columnName	문자열	발생이 들어 있는 열의 이름입니다 (가능한 경우).
row	Integer	해당 이벤트가 포함된 행의 행 번호입니다.

페이지 배열

적용 대상: Adobe 휴대용 문서 형식 (PDF) 파일

`pages` 배열에서 `Page` 객체는 Macie가 민감한 데이터의 발생을 감지한 페이지를 지정합니다. 객체에는 `pageNumber` 필드가 있습니다. `pageNumber` 필드에는 해당 항목이 포함된 페이지의 페이지 번호를 지정하는 정수가 저장됩니다.

다음 예제는 Macie가 PDF 파일에서 감지한 중요한 데이터의 발생 위치를 지정하는 `Page` 객체의 구조를 보여 줍니다.

```
"pages": [
   {
      "pageNumber": 10
   }
]
```

LineRanges배열

적용 대상: 이메일 메시지 (EML) 파일 및 CSV, JSON, JSON 라인 및 TSV 파일을 제외한 바이너리가 아닌 텍스트 파일 (예: HTML, TXT 및 XML 파일)

`lineRanges` 배열에서 `Range` 객체는 Macie가 민감한 데이터의 발생을 감지한 라인 또는 포함 라인 범위와 지정된 라인 또는 여러 라인에서의 데이터 위치를 지정합니다.

이 객체는 객체의 다른 유형의 배열에서 `occurrences` 지원하는 파일 유형의 경우 비어 있는 경우가 많습니다. 예외는 다음과 같습니다.

- 구조화되지 않은 파일의 구조화되지 않은 섹션에 있는 데이터 (예: 파일의 주석)
- Macie가 일반 텍스트로 분석하는 형식이 잘못된 파일의 데이터입니다.
- Macie가 민감한 데이터를 감지한 열 이름이 하나 이상 있는 CSV 또는 TSV 파일입니다.

다음 표는 `lineRanges` 배열 `Range` 객체에 있는 각 필드의 용도를 설명합니다.

필드	유형	설명
end	Integer	파일의 시작부터 해당 항목의 끝까지의 줄 수입니다.
start	Integer	파일의 시작 부분부터 발생 시작 시점까지의 줄 수입니다.
startColumn	Integer	발생 () 이 포함된 첫 번째 줄의 시작 부분부터 발생 시작 부분까지의 문자 수 (공백 포함`start`) 는 1부터 시작합니다.

레코드 어레이

적용 대상: 아파치 아브로 오브젝트 컨테이너, 아파치 파켓 파일, JSON 파일 및 JSON 라인 파일

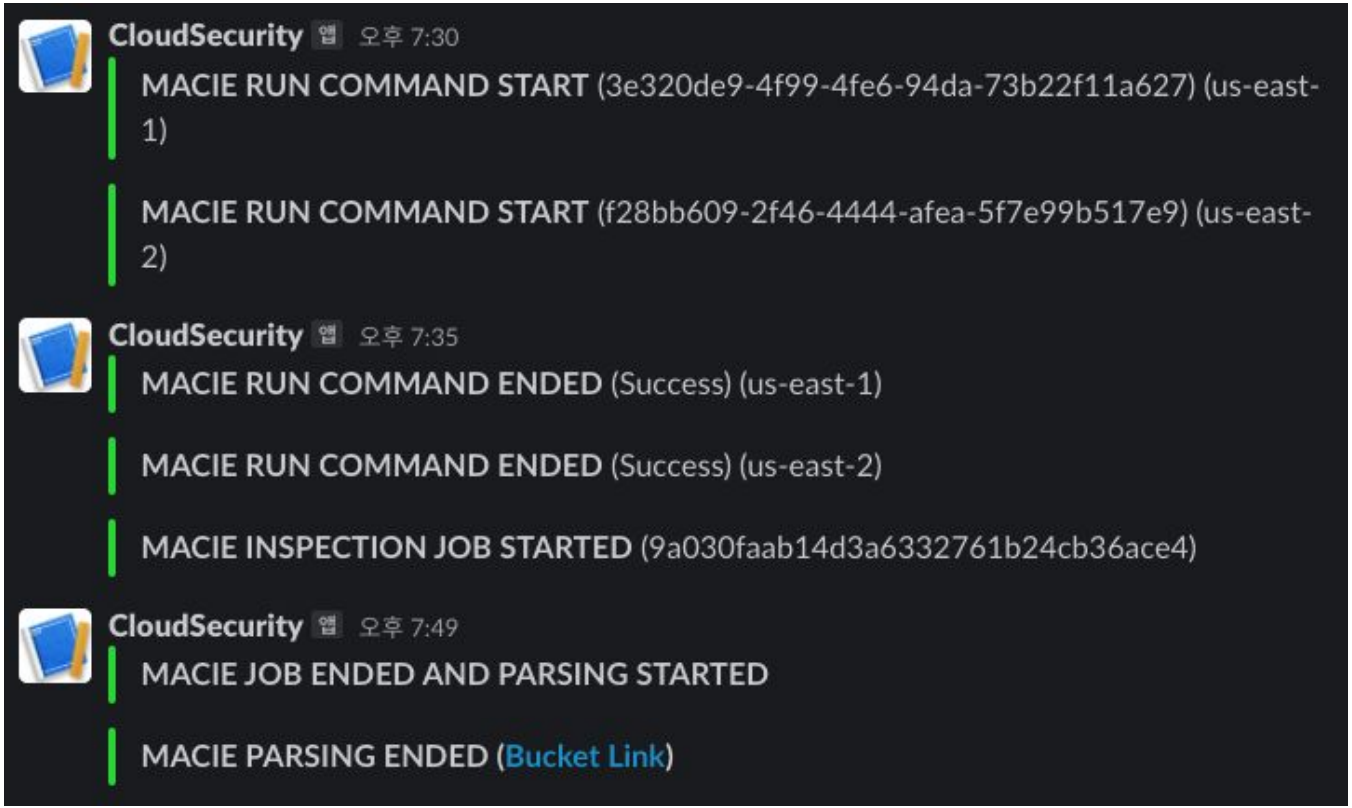
Avro 객체 컨테이너 또는 Parquet 파일의 경우 `records` 배열의 `Record` 객체는 레코드 색인과 Macie가 민감한 데이터의 발생을 감지한 레코드의 필드 경로를 지정합니다. JSON 및 JSON 라인 파일의 경우 `Record` 객체는 Macie가 민감한 데이터의 발생을 감지한 필드 또는 배열의 경로를 지정합니다. JSON Lines 파일의 경우 해당 항목이 포함된 줄의 인덱스도 지정합니다.

다음 표에서는 `Record` 객체에 있는 각 필드의 용도를 설명합니다.

필드	유형	설명
jsonPath	문자열	발생 경로인 JSONPath 표현식입니다. Avro 오브젝트 컨테이너 또는 Parquet 파일의 경우 해당 항목이 포함된 레코드 (`recordIndex`) 의 필드 경로입니다. JSON 또는 JSON 라인 파일의 경우 해당 항목이 포함된 필드 또는 배열의 경로입니다. 데이터가 배열의 값인 경우 경로는 해당 항목이 포함된 값도 나타냅니다. Macie가 경로에 있는 요소 이름에서 민감한 데이터를 감지하면 Macie는 `Record` 객체에서 해당 `jsonPath` 필드를 생략합니다. 경로 요소 이름이 20자를 초과하는 경우 Macie는 이름 시작 부분의 문자를 제거하여 이름을 잘라냅니다. 결과로 생성되는 전체 경로가 250자를 초과하는 경우 Macie는 경로의 첫 번째 요소부터 시작하여 경로에 250자 이하가 될 때까지 경로를 잘라냅니다.
recordIndex	Integer	Avro 객체 컨테이너 또는 Parquet 파일의 경우 해당 항목이 포함된 레코드의 레코드 색인 (0부터 시작) JSON 라인 파일의 경우 해당 항목이 포함된 라인의 라인 인덱스는 0부터 시작합니다. 이 값은 항상 JSON0 파일용입니다.

```
if occurrences.get("cells"):
    results[result_key] = occurrences["cells"]
if occurrences.get("lineRanges"):
    results[result_key] = []
    for line_range in occurrences["lineRanges"]:
        [start, end, start_column] = [
            (int(line_range[key]) - 1) for key in ["start", "end", "startColumn"]
        ]
        sensitive_columns = "".join(body.split("\n")[start : end + 1])
        results[result_key] += [
            convert_center_string_to_asterisk(sensitive_columns[start_column:start_column + 100])
        ]
elif occurrences.get("pages"):
    results[result_key] = occurrences["pages"]
elif occurrences.get("records"):
    results[result_key] = occurrences["records"]
elif occurrences.get("offsetRanges"):
    #  This array is reserved for future use. If this array is present, the value for it is always null.
    pass
```

8. Notify results to Slack

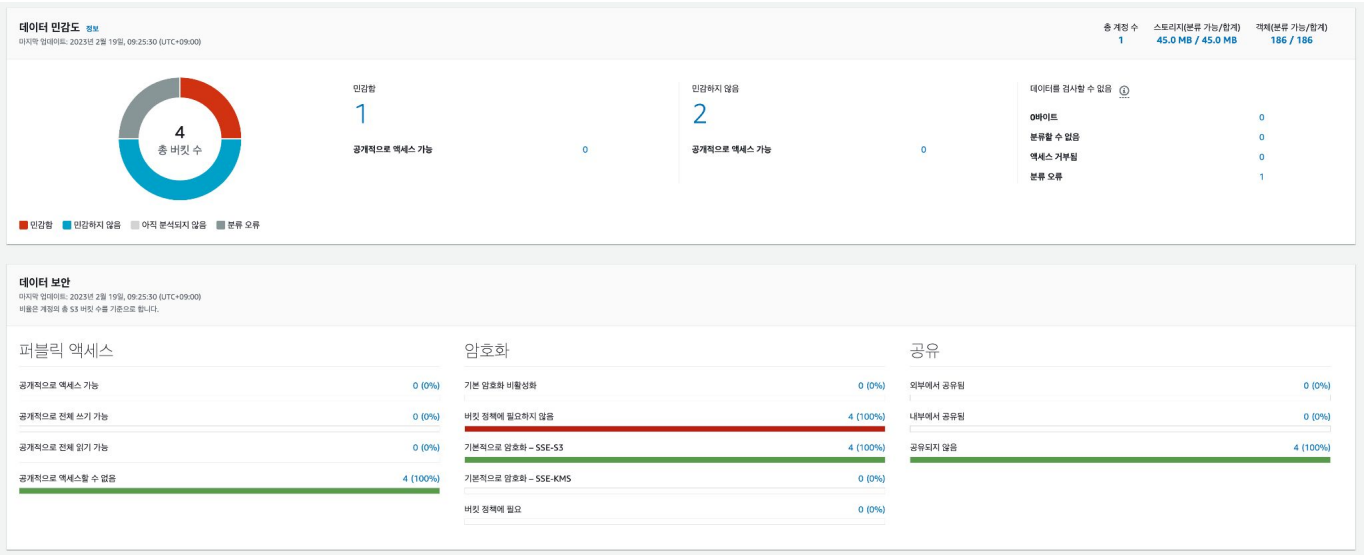


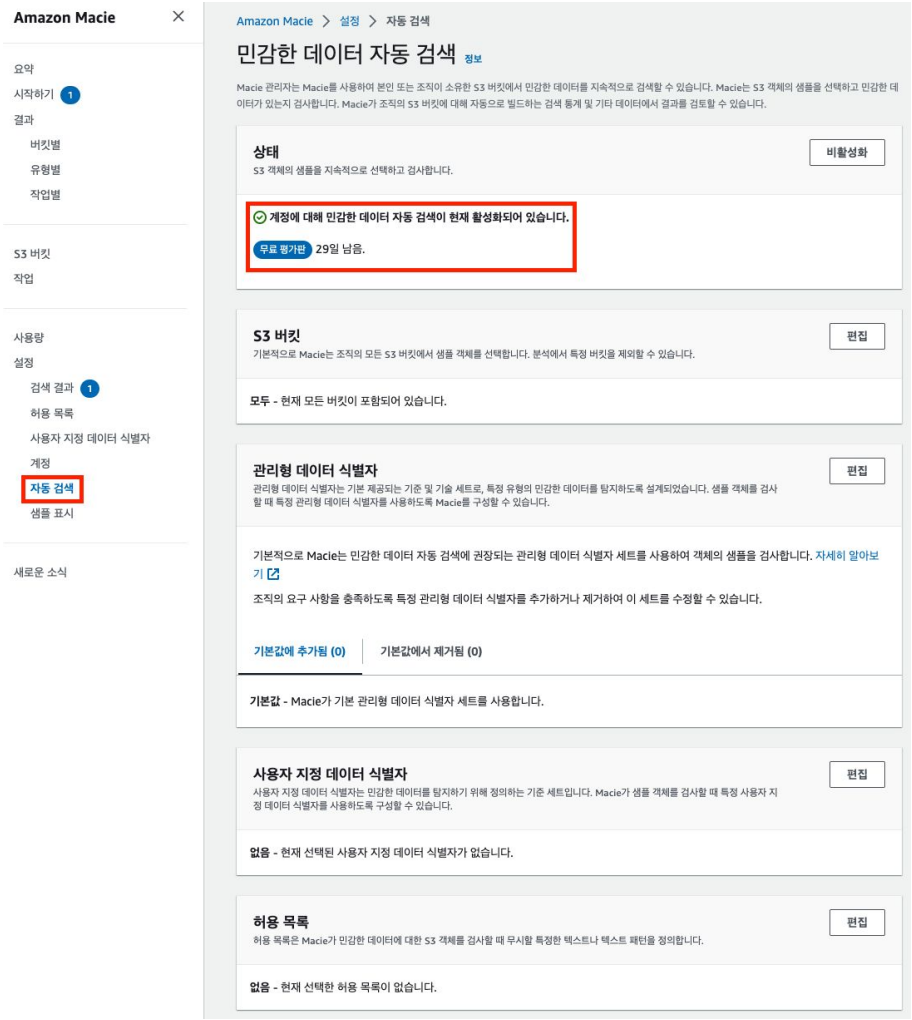
CATEGORY	TYPE	OBJECT	VALUE
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0fc900578d3030580/var/log/messages	Jiri***sina
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0fc900578d3030580/var/log/messages	Jiri***sina
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0fc900578d3030580/var/log/messages	Jiri***sina
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0fc900578d3030580/var/log/messages	Jiri***sina
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0fc900578d3030580/var/log/messages	Jiri***sina
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0fc900578d3030580/var/log/messages	Jiri***sina
ADDRESS	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0f092927d5829b09a/var/lib/yum/rpmdb-indexes/conflicts	x86_6400.23.221.amzn2.0.1p************************3.amzn2.0.1plymout
OPENSSH_PRIVATE_KEY	CREDENTIALS	private-info-bucket-tlh06k91/i-0f092927d5829b09a/etc/ssh/ssh_host_ed25519_key	-----BEGIN OPENSSH PRIVATE KEY----********************************AAA
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0f092927d5829b09a/var/log/messages	Jiri***sina
OPENSSH_PRIVATE_KEY	CREDENTIALS	private-info-bucket-tlh06k91/i-0f092927d5829b09a/etc/ssh/ssh_host_ecdsa_key	-----BEGIN EC PRIVATE KEY-----MHcC********************************vhX
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/var/log/messages	Jiri***sina
OPENSSH_PRIVATE_KEY	CREDENTIALS	private-info-bucket-tlh06k91/i-0f092927d5829b09a/etc/ssh/ssh_host_rsa_key	-----BEGIN RSA PRIVATE KEY-----MII********************************r6BKP
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0fae5291351c8c8aa/var/log/messages	Jiri***sina
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0fae5291351c8c8aa/var/log/messages	Jiri***sina
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0fae5291351c8c8aa/var/log/messages	Jiri***sina
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0fae5291351c8c8aa/var/log/messages	Jiri***sina
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0fae5291351c8c8aa/var/log/messages	Jiri***sina
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0fae5291351c8c8aa/var/log/messages	Jiri***sina
OPENSSH_PRIVATE_KEY	CREDENTIALS	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/etc/ssh/ssh_host_rsa_key	-----BEGIN RSA PRIVATE KEY-----MII********************************46qxb
OPENSSH_PRIVATE_KEY	CREDENTIALS	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/etc/ssh/ssh_host_ed25519_key	-----BEGIN OPENSSH PRIVATE KEY----********************************AAA
OPENSSH_PRIVATE_KEY	CREDENTIALS	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/etc/ssh/ssh_host_ecdsa_key	-----BEGIN EC PRIVATE KEY-----MHcC********************************DbU
ADDRESS	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/var/lib/yum/rpmdb-indexes/conflicts	x86_6400.23.221.amzn2.0.1p************************3.amzn2.0.1plymout

CATEGORY	TYPE	OBJECT	VALUE
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0f092927d5829b09a/var/log/messages	Jiri***sina
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/var/log/messages	Jiri***sina
ADDRESS	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/var/lib/yum/rpmdb-indexes/conflicts	x86_6400.23.221.amzn2.0.1p**********3.amzn2.0.1plymouthx86_64
ADDRESS	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0f092927d5829b09a/var/lib/yum/rpmdb-indexes/conflicts	x86_6400.23.221.amzn2.0.1p**********3.amzn2.0.1plymouthx86_64
OPENSSH_PRIVATE_KEY	CREDENTIALS	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/etc/ssh/ssh_host_ed25519_key	-----BEGIN OPENSSH PRIVATE KEY----********** AAAEbm9uZ(
OPENSSH_PRIVATE_KEY	CREDENTIALS	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/etc/ssh/ssh_host_rsa_key	-----BEGIN RSA PRIVATE KEY-----MII**********46qxbBbw7HZH
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/var/log/messages-20230219	Jiri***sina
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/var/log/messages-20230219	Jiri***sina
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/var/log/messages-20230219	Jiri***sina
AWS_CREDENTIALS	CREDENTIALS	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/home/ssm-user/.aws/credentials	=2zlCNBIFMtqY1**********9BD/lBlcRddfa4
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0f092927d5829b09a/var/log/messages-20230219	Jiri***sina
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-0f092927d5829b09a/var/log/messages-20230219	Jiri***sina
AWS_CREDENTIALS	CREDENTIALS	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/home/ssm-user/credentials	=2zlCNBIFMtqY1**********9BD/lBlcRddfa4
OPENSSH_PRIVATE_KEY	CREDENTIALS	private-info-bucket-tlh06k91/i-0f092927d5829b09a/etc/ssh/ssh_host_ed25519_key	-----BEGIN OPENSSH PRIVATE KEY----********** AAAEbm9uZ(
OPENSSH_PRIVATE_KEY	CREDENTIALS	private-info-bucket-tlh06k91/i-0f092927d5829b09a/etc/ssh/ssh_host_ecdsa_key	-----BEGIN EC PRIVATE KEY-----MHcC**********vhX+8+ehRAu
주민등록번호	CUSTOM	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/home/ssm-user/test.txt	930822*****1234
NAME	PERSONAL_INFORMATION	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/var/log/dmesg.old	Jiri***sina
OPENSSH_PRIVATE_KEY	CREDENTIALS	private-info-bucket-tlh06k91/i-0f092927d5829b09a/etc/ssh/ssh_host_rsa_key	-----BEGIN RSA PRIVATE KEY-----MII**********r6BKPgVl3Xyhn
OPENSSH_PRIVATE_KEY	CREDENTIALS	private-info-bucket-tlh06k91/i-078e738ed5fa3a031/etc/ssh/ssh_host_ecdsa_key	-----BEGIN EC PRIVATE KEY-----MHcC**********DbUjfu48RBF

Automated Data Discovery / 설정

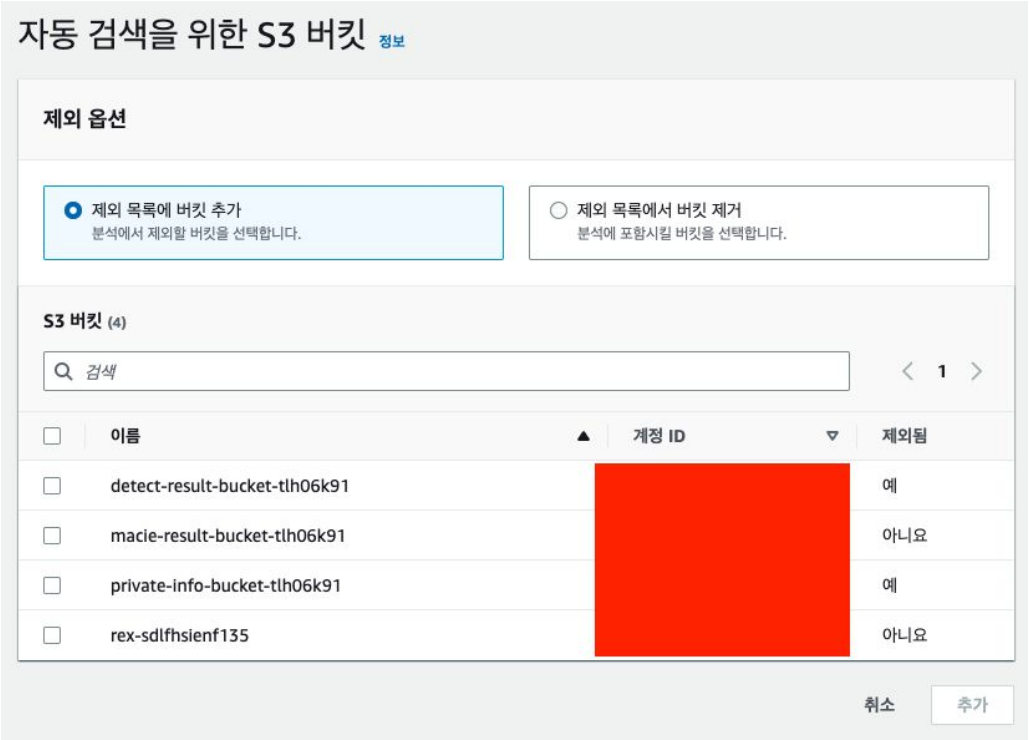
1. 2022년 11월 28일에 공개된 **자동 데이터 검색** 기능
2. 기능을 켠 후 최대 **48**시간 이후부터 결과 수신 가능
3. 현재 103개의 관리형 데이터 식별자 제공 중 (23년 2월 기준)




Amazon Macie
요약
시작하기 1
결과
버킷별
유형별
작업별
S3 버킷
작업
사용량
설정
검색 결과 1
허용 목록
사용자 지정 데이터 식별자
계정
자동 검색
샘플 표시
새로운 소식
Amazon Macie > 설정 > 자동 검색
민감한 데이터 자동 검색 정보
Macie 관리자는 Macie를 사용하여 본인 또는 조직이 소유한 S3 버킷에서 민감한 데이터를 지속적으로 검색할 수 있습니다. Macie는 S3 객체의 샘플을 선택하고 민감한 데이터가 있는지 검사합니다. Macie가 조직의 S3 버킷에 대해 자동으로 빌드하는 검색 통계 및 기타 데이터에서 결과를 검토할 수 있습니다.
상태
S3 객체의 샘플을 지속적으로 선택하고 검사합니다.
비활성화
계정에 대해 민감한 데이터 자동 검색이 현재 활성화되어 있습니다.
무료 평가판 29일 남음.
S3 버킷
기본적으로 Macie는 조직의 모든 S3 버킷에서 샘플 객체를 선택합니다. 분석에서 특정 버킷을 제외할 수 있습니다.
편집
모두 - 현재 모든 버킷이 포함되어 있습니다.
관리형 데이터 식별자
관리형 데이터 식별자는 기본 제공되는 기준 및 기술 세트로, 특정 유형의 민감한 데이터를 탐지하도록 설계되었습니다. 샘플 객체를 검사할 때 특정 관리형 데이터 식별자를 사용하도록 Macie를 구성할 수 있습니다.
편집
기본적으로 Macie는 민감한 데이터 자동 검색에 권장되는 관리형 데이터 식별자 세트를 사용하여 객체의 샘플을 검사합니다. 자세히 알아보기
조직의 요구 사항을 충족하도록 특정 관리형 데이터 식별자를 추가하거나 제거하여 이 세트를 수정할 수 있습니다.
기본값에 추가됨 (0)
기본값에서 제거됨 (0)
기본값 - Macie가 기본 관리형 데이터 식별자 세트를 사용합니다.
사용자 지정 데이터 식별자
사용자 지정 데이터 식별자는 민감한 데이터를 탐지하기 위해 정의하는 기준 세트입니다. Macie가 샘플 객체를 검사할 때 특정 사용자 지정 데이터 식별자를 사용하도록 구성할 수 있습니다.
편집
없음 - 현재 선택된 사용자 지정 데이터 식별자가 없습니다.
허용 목록
허용 목록은 Macie가 민감한 데이터에 대한 S3 객체를 검사할 때 무시할 특정한 텍스트나 텍스트 패턴을 정의합니다.
편집
없음 - 현재 선택한 허용 목록이 없습니다.

S3 버킷

1. 제외할 버킷 선택
2. CloudTrail
3. Config
4. Vpc Flow Logs
5. 수동 작업(Job) 대상 버킷
6. 등 AWS 기본 버킷 제외 검토



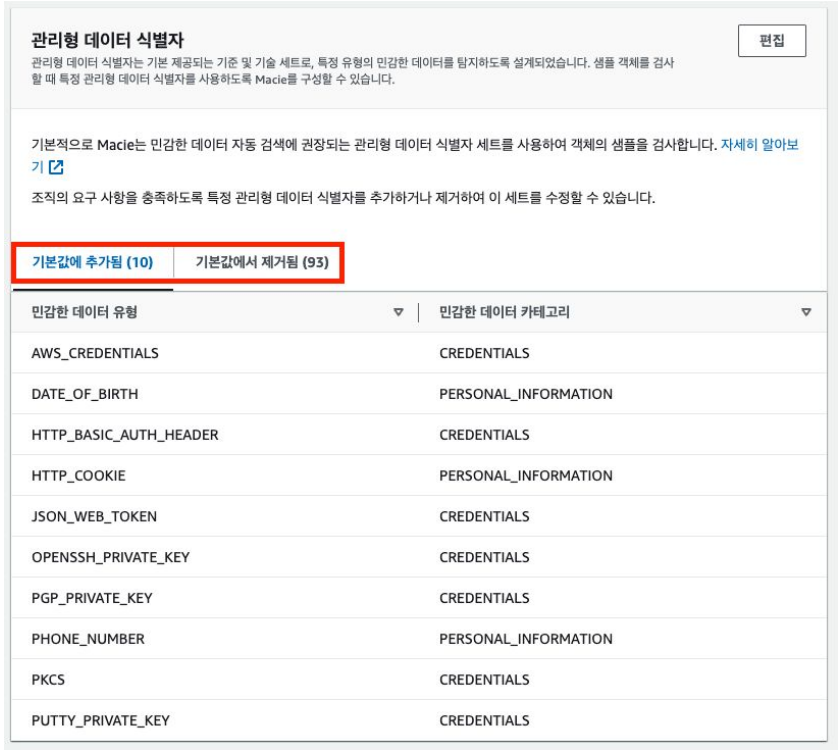
관리형 데이터 식별자

1. 리전별 필요한 식별자 선택
2. 선택 시 아래 내용을 주의

3. 탐색 창의 **설정에서 자동 검색을** 선택합니다.

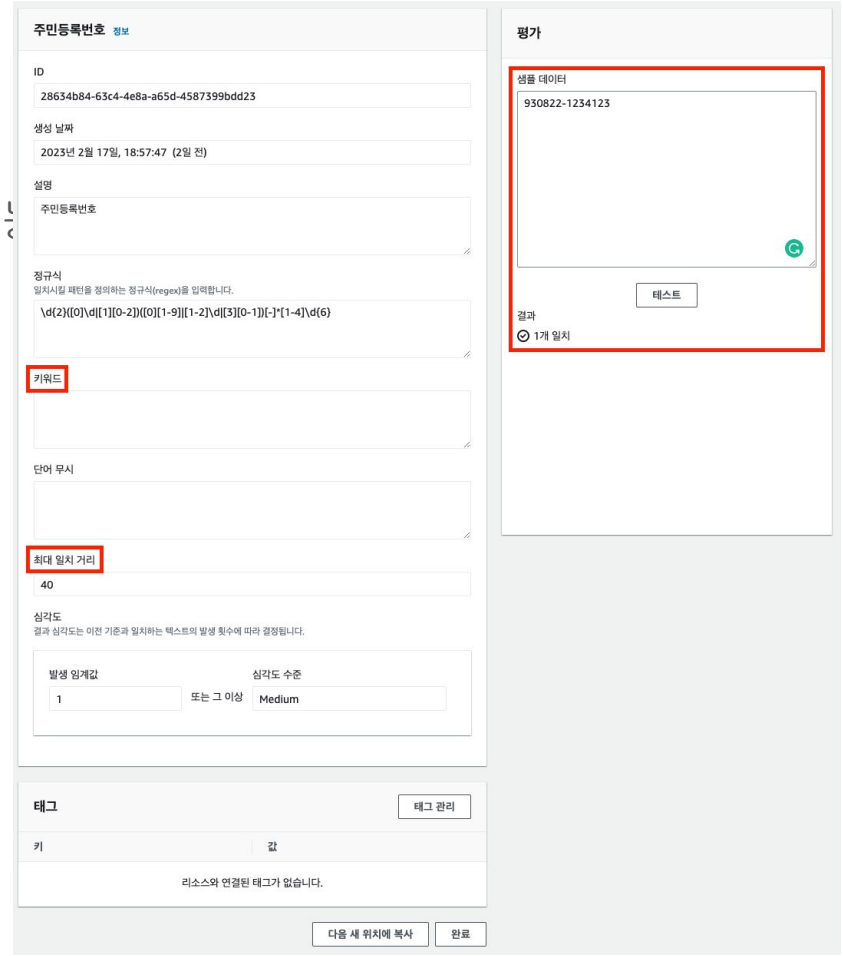
   자동 중요 데이터 검색 페이지의 **관리 대상 데이터 식별자** 섹션에는 현재 설정이 두 개의 탭으로 구성되어 표시됩니다.

   - **기본값으로 추가됨** — 이 탭에는 명시적으로 추가한 관리 데이터 식별자가 나열됩니다. Macie는 기본 세트에 포함되어 있지만 사용자가 명시적으로 제거하지 않은 관리 데이터 식별자와 함께 이러한 관리 데이터 식별자를 사용합니다.
   - **기본값에서 제거됨** — 이 탭에는 명시적으로 제거된 관리 데이터 식별자가 나열됩니다. Macie는 이러한 관리형 데이터 식별자를 사용하지 않습니다.



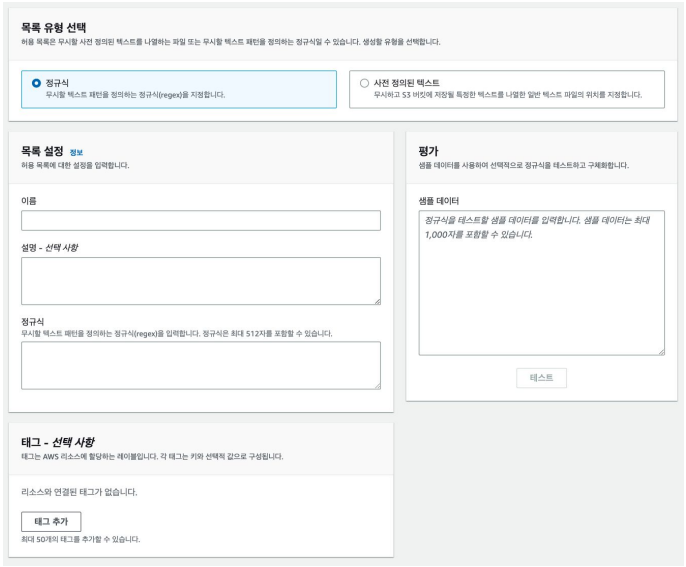
사용자 데이터 식별자

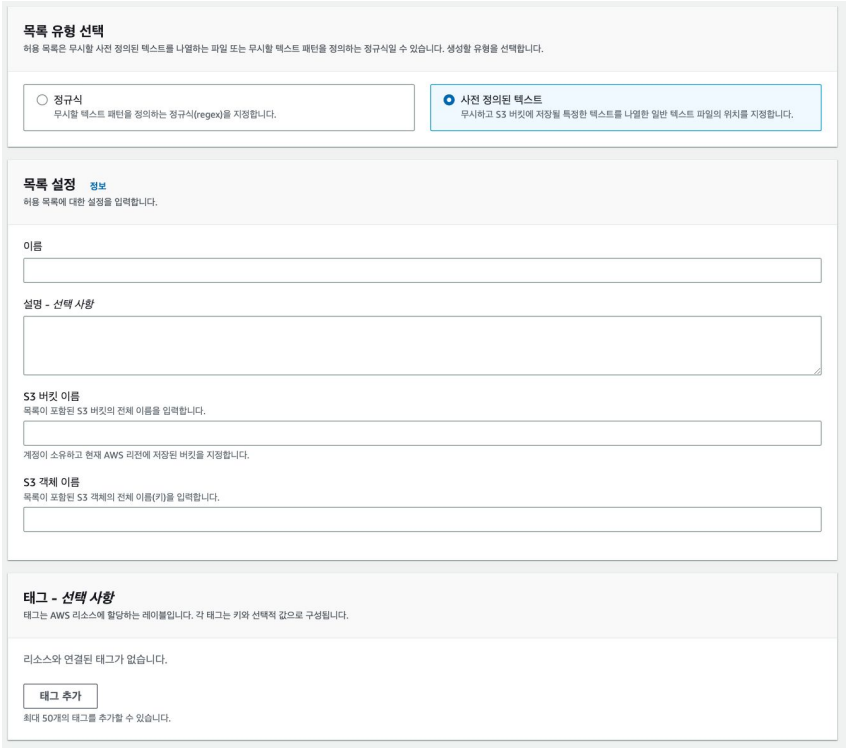
1. 정규표현식을 이용해 탐지 규칙 생성 가능
2. 회사에서 규칙 있는 데이터가 있다면,
   a. 키워드 : 특정 단어 (REX)
   b. 최대 일치 거리 (40 글자 이내)



허용목록

1. 정규식 또는 버킷의 객체를 통해 허용 ㄱ
2. 웬만하면 사용하지 말자





참고사항

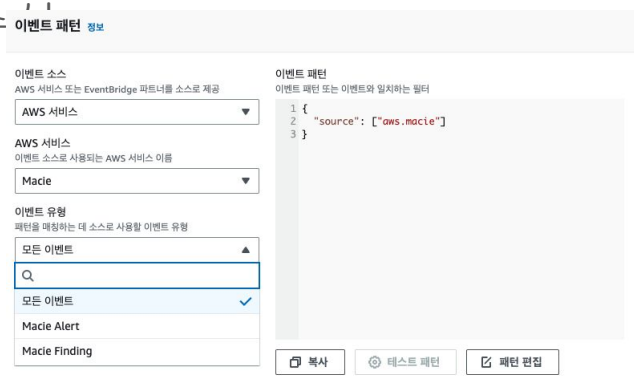
1. 민감정보 무결성 검증 가능 여부 검토 필요 (주민등록번호 등)
2. EventBridge는 기본 이벤트가 아닌 별도의 이벤트 형태를 정의할 수 있음
3. 리전별 리소스 분리 시 더욱 빠른 속도로 처리 가능
4. Macie 요금
    a. 프리티어 활용하여 테스트 추천

주의할점

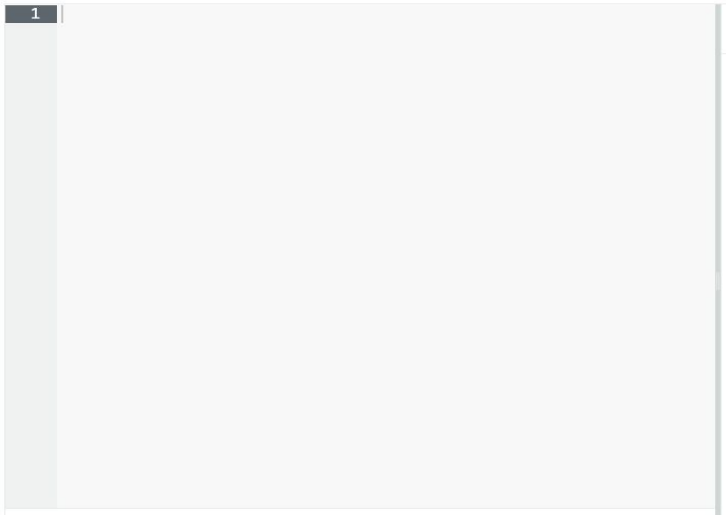
1. 멀티 어카운트에서 Run Command 결과 SNS 알림 사용 시, sts:PassRole 주의
    a. PassRole은 교차 계정 역할 전달 불가
    b. 계정별로 역할을 만들어야 함
2. 인스턴스 프로파일 규격화 및 자동화 필요
    a. 관리형 정책을 사용하여 보안용 정책 관리
    b. Automate Instance Profile 또는 테라폼 모듈 등을 통해 강제화 검토 필요
3. EC2 IMDSv2 사용 시, 토큰 발급하여 메타데이터 요청 필요
4. AWS의 모든 리소스 사용의 시작 할당량

자동 검색. 시작은 어떻게 하는게 좋을까?

1. 자동 검색 기능 활성화
2. 불필요 버킷 검사 제외
3. 리전별 필요한 관리형 데이터 식별자 및 사용자 데이터 식별자 추가
4. “결과" 페이지에서 실제 결과 검토 및 **버킷 보안 정책 강화**
5. EventBridge, Lambda 등을 이용해 파싱 및 알람 수



여러분의 버킷 정책은 어떤 방식인가요?

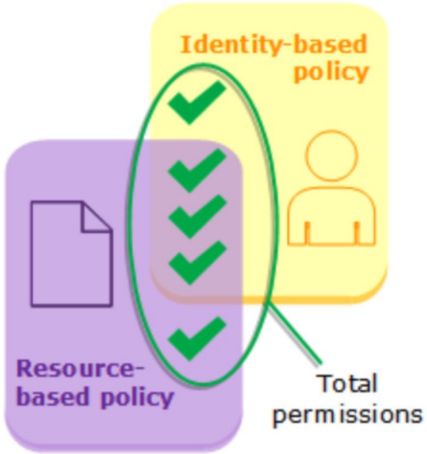


“난 사람을 믿어” 형

“난 이것만 허용될거라고 믿어” 형

리소스 기반 정책과 함께 자격 증명 기반 정책 평가

자격 증명 기반 정책 및 리소스 기반 정책은 연결된 자격 증명이나 리소스에 권한을 부여합니다. IAM 엔터티(사용자 또는 역할)가 동일 계정 내에서 리소스에 대한 액세스를 요청할 경우 AWS는 자격 증명 기반 및 리소스 기반 정책을 통해 부여된 모든 권한을 평가합니다. 결과적으로 두 정책 유형의 모든 권한이 권한으로 부여됩니다. 자격 증명 기반 정책, 리소스 기반 정책 또는 두 정책 모두에 의해 작업이 허용되는 경우 AWS에서는 해당 작업을 허용합니다. 이들 정책 중 하나에 포함된 명시적 거부는 허용을 재정의합니다.



앞으론 이렇게 해보는게 어떨까요?

```
{
    "Sid": "denyOutdatedTLS",
    "Effect": "Deny",
    "Principal": "*",
    "Action": "s3:*",
    "Resource": [
        "arn:aws:s3:::macie-result-bucket-tlh06k91/*",
        "arn:aws:s3:::macie-result-bucket-tlh06k91"
    ],
    "Condition": {
        "NumericLessThan": {
            "s3:TlsVersion": "1.2"
        }
    }
},
```

```
{
    "Sid": "PrincipalArns",
    "Effect": "Deny",
    "Principal": {
        "AWS": "*"
    },
    "Action": [
        "s3:PutObject",
        "s3:GetObject",
        "s3:DeleteObject"
    ],
    "Resource": "arn:aws:s3:::rex-chun-s3-bucket-with-security-policy/*",
    "Condition": {
        "ArnNotEquals": {
            "aws:PrincipalArn": "arn:aws:iam::[redacted]:root"
        }
    }
}
```

다만...

AWSServiceRoleForAmazonMacie 라는 서비스 연결 역할을 항상 포함해야 함.

```
{
   "Version": "2012-10-17",
   "Id":" Policy1415115example ",
   "Statement": [
      {
         "Sid": "Access from specific VPCE and Macie only",
         "Effect": "Deny",
         "Principal": "*",
         "Action": "s3:*",
         "Resource": [
            "arn:aws:s3:::DOC-EXAMPLE-BUCKET",
            "arn:aws:s3:::DOC-EXAMPLE-BUCKET/*"
         ],
         "Condition": {
            "StringNotEquals": {
               "aws:SourceVpce": "vpce-1a2b3c4d"
            },
            "StringNotLike": {
               "aws:PrincipalArn": "arn:aws:iam::123456789012:role/aws-service-role/macie.amazonaws.com/AWSServiceRoleForAmaz
            }
         }
      }
   ]
}
```

실제로 테스트 환경에서 어떻게 데이터를 받고 있을까?

결론적으로 왜 Macie?

1. 서버 및 S3 내 허가되지 않은 SSH, AWS 키 탐지/비교 및 조치
    a. 민감 정보에 대한 회사 차원의 자동화 프로세스 수립 가능
2. 무수히 많은 관리형 데이터 식별자들
    a. 각 국가별 규제에 맞춰 유연하게 적용 가능 (리전 단위)
3. 적은 비용으로 최대의 효과를 누릴 수 있음
4. SSM Agent를 온프렘과 ~~PC~~에도..?
    a. With RolesAnywhere

Q&A

참고자료

1. https://github.com/cjsrkd3321/aws-security-architectures/tree/main/ALL_REGIONS_DETECT_PRIVATE_INFO
2. https://github.com/cjsrkd3321/aws-security-architectures/tree/main/EC2_INSTANCE_PROFILE_AUTOMATION
3. https://inf.run/8DHH
4. https://medium.com/@7424069/aws-detect-sensitive-data-using-macie-with-fully-event-driven-architecture-in-all-regions-89c10c405b50
5. https://medium.com/@7424069/aws-ec2-%EC%9D%B8%EC%8A%A4%ED%84%B4%EC%8A%A4-%ED%94%84%EB%A1%9C%ED%8C%8C%EC%9D%BC-%EC%9E%90%EB%8F%99-%EB%B6%80%EC%97%AC-%EC%95%84%ED%82%A4%ED%85%8D%EC%B2%98-%EA%B5%AC%ED%98%84-8667724dd06b