[AWS Black Belt Online Seminar]

Amazon Macie

サービスカットシリーズ

Solutions Architect 高橋 悟史
2020/8/12

AWS 公式 Webinar
https://amzn.to/JPWebinar

過去資料
https://amzn.to/JPArchive

自己紹介

名前：髙橋 悟史 （たかはし さとし）
CISSP , CISM

職種：セキュリテイ ソリューション アーキテクト

職務：AWS セキュリティの啓蒙活動、セキュリティ関連サービスに関するお客様のサポート

経歴：外資ITベンダーにて、メインフレーム、HPC、分散系システムの構築運用に関わったあと、システム管理、セキュリティ、特に認証やID管理、SSOに関する技術支援に従事。セキュリティ企業にてサイバーセキュリティソリューション、SIEMを担当。SaaS ベンダーにてクラウド・セキュリティを担当したのちにAWSへ

好きな AWS サービス：AWS Key Management Service

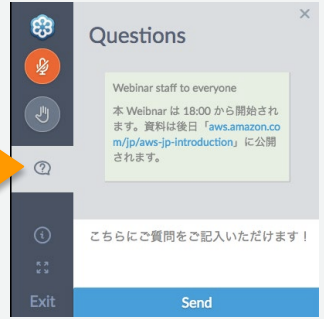
AWS Black Belt Online Seminar とは

「サービス別」「ソリューション別」「業種別」のそれぞれのテーマに分かれて、アマゾン ウェブ サービス ジャパン株式会社が主催するオンラインセミナーシリーズです。

質問を投げることができます！

- 書き込んだ質問は、主催者にしか見えません
- 今後のロードマップに関するご質問は
  お答えできませんのでご了承下さい

① 吹き出しをクリック
② 質問を入力
③ Sendをクリック



Twitter ハッシュタグは以下をご利用ください
#awsblackbelt

内容についての注意点

- 本資料では2020年8月12日時点のサービス内容および価格についてご説明しています。最新の情報はAWS公式ウェブサイト(http://aws.amazon.com)にてご確認ください。

- 資料作成には十分注意しておりますが、資料内の価格とAWS公式ウェブサイト記載の価格に相違があった場合、AWS公式ウェブサイトの価格を優先とさせていただきます。

- 価格は税抜表記となっています。日本居住者のお客様には別途消費税をご請求させていただきます。

- AWS does not offer binding price quotes. AWS pricing is publicly available and is subject to change in accordance with the AWS Customer Agreement available at http://aws.amazon.com/agreement/. Any pricing information included in this document is provided only as an estimate of usage charges for AWS services based on certain information that you have provided. Monthly charges will be based on your actual use of AWS services, and may vary from the estimates provided.

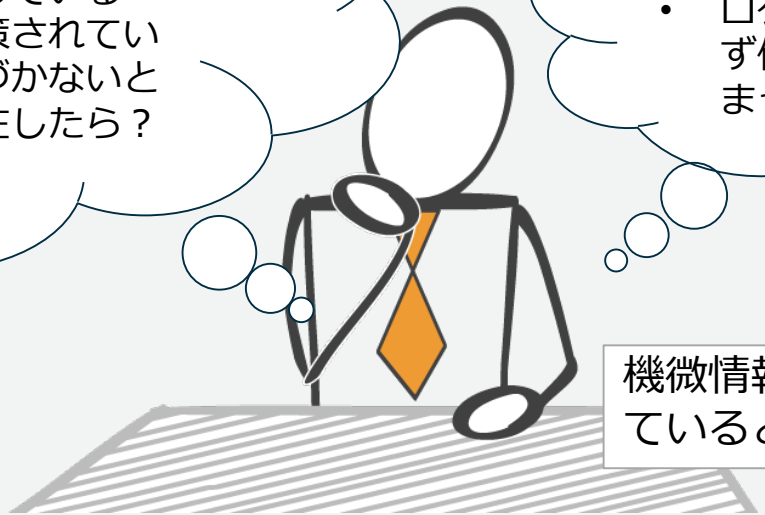
アジェンダ

- 機微情報管理の重要性と課題
- Amazon Macie 概要
- Amazon S3 バケットとオブジェクトの使用状況の可視化
- 機微情報の検出・評価
- 評価・検出結果の参照と、他のサービスとの連携
- 管理・運用・制約

機微情報管理の重要性と難しさ

機微情報とは
個人情報、クレジットカード、銀行口座番号などの保護が必要な情報
特許情報、設計技術など競合他社に知られてはいけない情報

- 機微情報は、既に、認証、アクセス制御、暗号化、ログ管理などで保護している
- 社内の情報規定で対策されているにも関わらず、気づかないところに機微情報が存在したら？



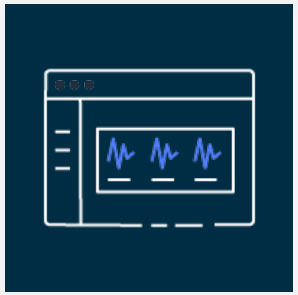
- データレイク上の保管データの中に機微情報が紛れてしまっていたら？
- ログデータの中に意図せず個人情報が入ってしまっていたら？

機微情報が正しく管理され保護されていると自信を持って言えるか？

Amazon Macie が解決する課題

1. アカウント内にある Amazon S3 バケットの利用状況の可視化、格納されている大量のオブジェクトの可視化
2. 設定に基づき、指定されたバケット内の機微情報の評価・検出を効率的に実行
3. 評価・検出結果の参照、他のサービスへの連携

Amazon Macie 概要



Amazon S3 バケットの使用状況の可視化と評価

- バケットの使用状況とバケットポリシーの可視化

機微情報の検出

- 検出ジョブによる検出
- 柔軟な検出範囲と条件の設定
- AWS マネージド定義とカスタム定義の両方を利用可能

スケーラビリティと集中管理の両立

- AWS Organizations 対応

検出結果に対する対応の実行と自動化

- 詳細な結果（Findings）の出力とイベントによる AWS サービスとの連携
- 管理 API の提供

1. Amazon S3 バケットとオブジェクトの使用状況の可視化

Amazon Macie の利用ステップ

- Amazon Macie では、アカウントが持っている S3 バケットを全部スキャンするのではなく、管理者が S3 バケットの状態を確認して、スキャンが必要であるバケットを指定してスキャンする仕組みになっている

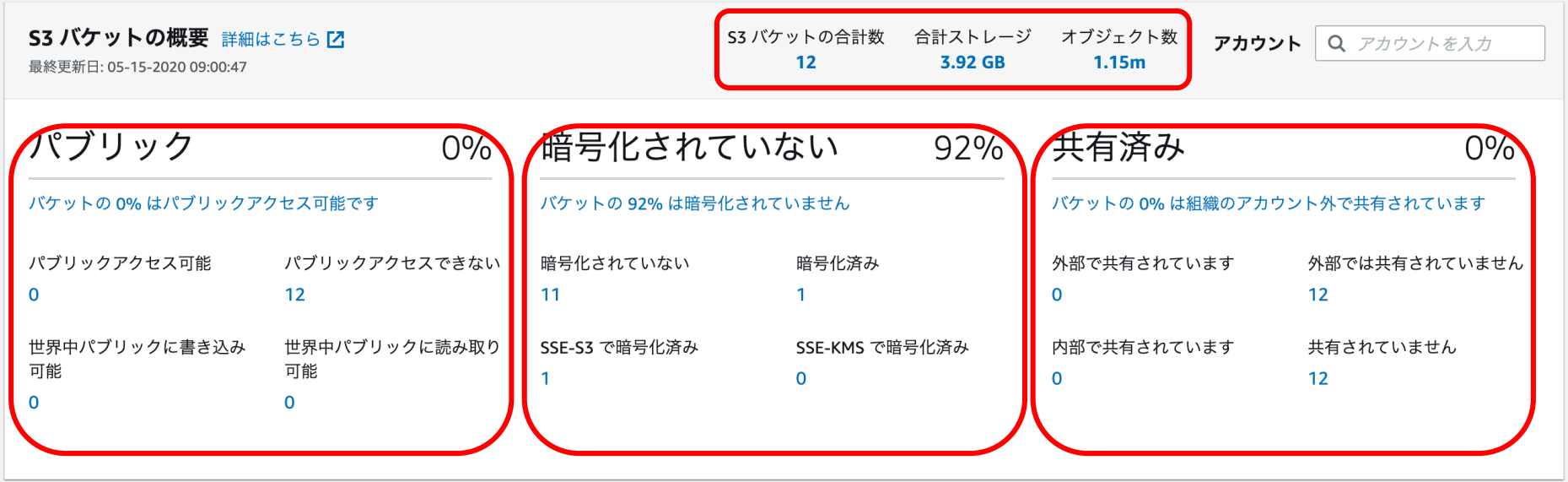
- AWS マネジメントコンソール上の Macie の管理画面で S3 バケットの状況サマリー、それぞれのバケットの詳細を確認して、スキャン対象のバケットを指定する

- スキャンするためにジョブを作成する。ジョブはスキャン対象バケットなどの対象の設定とスケジュールがセットになっている。スケジュールした場合、指定したスケジュールで自動でジョブが実行される

Amazon S3 バケットの概要表示 1

AWS マネジメントコンソール→Amazon Macie→概要
S3 バケットの利用状況を把握可能

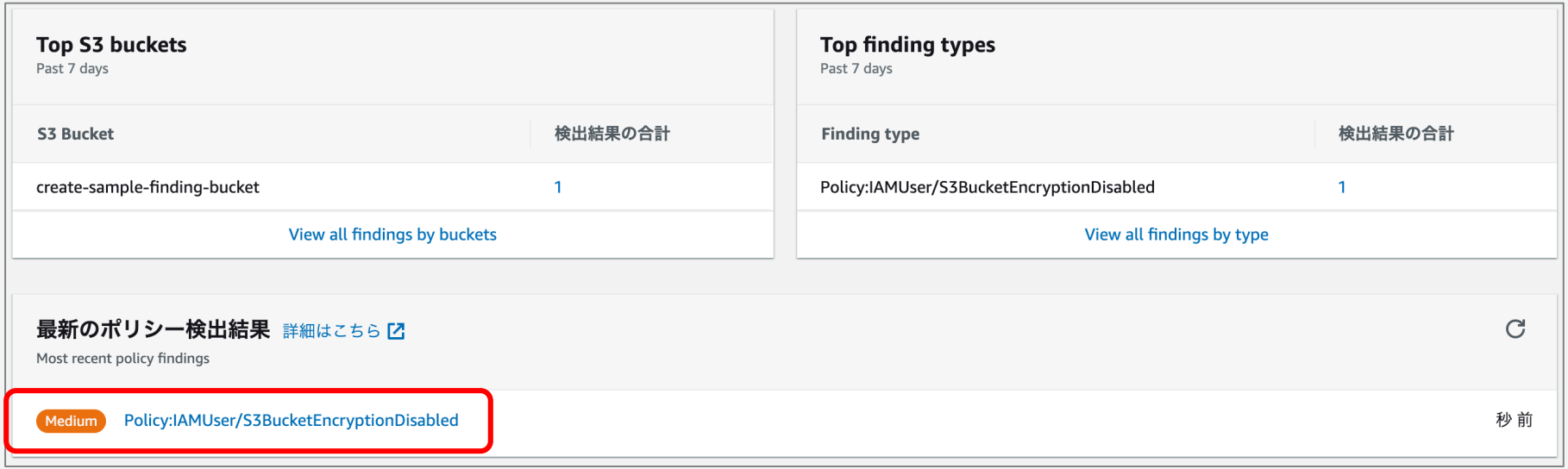
バケット数、ストレージ利用量、オブジェクト数



バケット公開状況　　バケット暗号化状況　　バケット共有状況

Amazon S3 バケットの概要表示 2

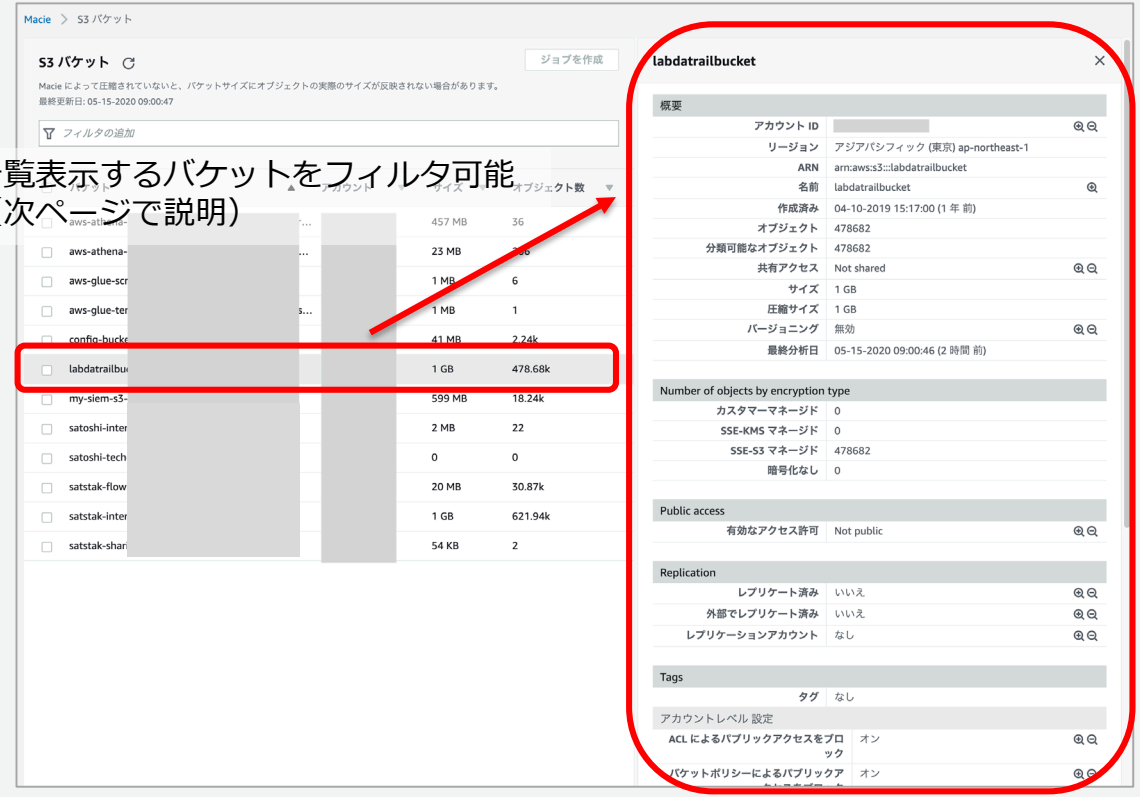
前ページのスクリーンショットの下側の画面
S3 バケットの設定変更を Findings として検出する
これは Macie の機微情報のスキャンとは別機能となる



例 : S3 バケットの暗号化が無効化されたことを検出

Amazon S3 バケットごとの状況表示

Amazon Macie → S3 バケット でアカウントが持つバケットの状況を個別に参照出来る



左のペインにバケットの一覧が表示され、バケットをクリックすると右側に詳細が表示される

- オブジェクト数、暗号化の有無と暗号化方式
- 公開（パブリック・アクセス）の有無
- レプリケーション設定の有無

Amazon S3 バケットのフィルタ表示

バケットのフィルタで指定出来る条件

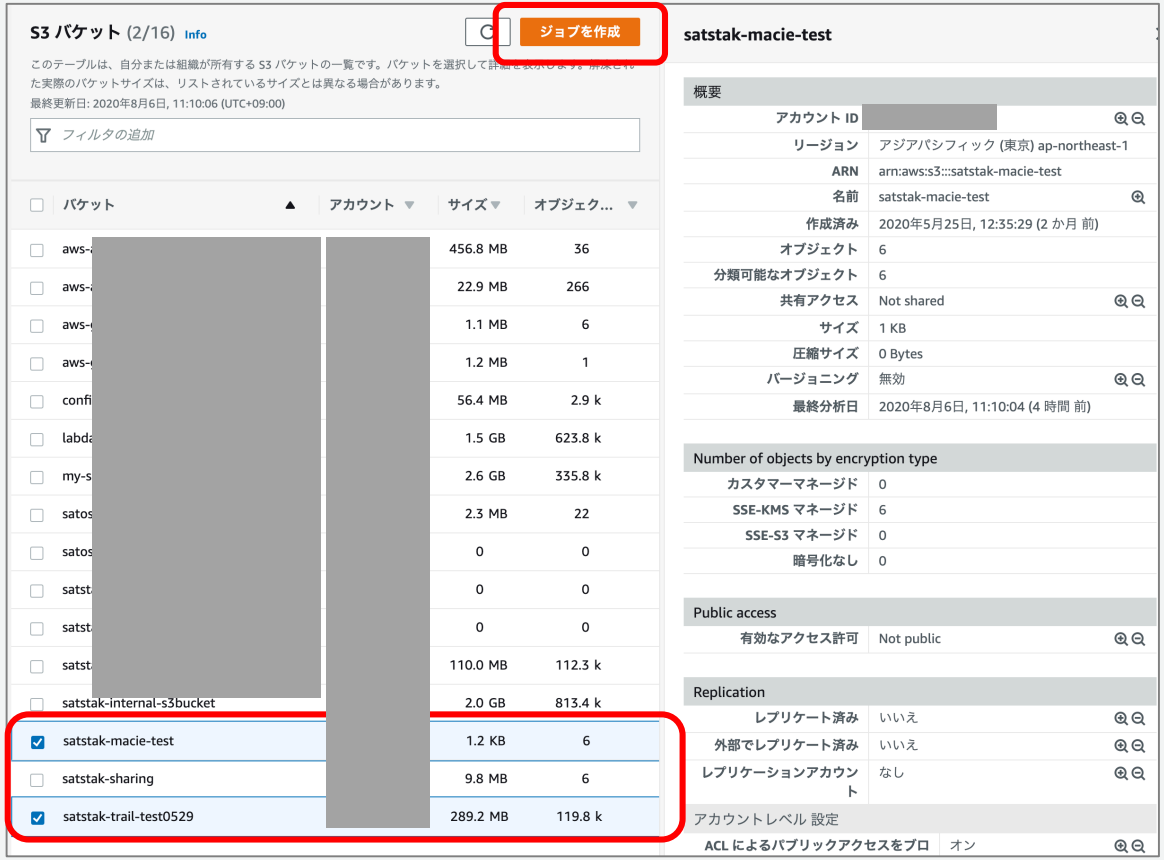
- アカウントID
- オブジェクト数
- ストレージサイズ
- タグキー
- バージョニング有無
- パブリック・アクセス許可の有無（書き込み、読み込み）
- レプリケートの有無
- 暗号化方式別のオブジェクト数（SSE-KMS(AWS), SSE-KMS(Customer), SSE-S3, 暗号化無し）
- バケットの暗号化設定

ユースケース

- オブジェクト数が多いバケットを選択する
- 暗号化されていないオブジェクトがあるバケットを選択する
- リスクの高いバケットを優先的にスキャンすることが可能

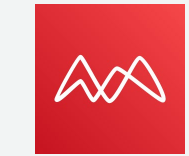
ジョブの作成

対象とするバケットにチェックを入れて右上のジョブを作成ボタンを押すと、
ジョブ作成の画面が表示される



2. 機微情報の検出・評価

Macie のスキャンはジョブを作成して定義する



Amazon Macie

ジョブ定義

ジョブ名

対象 S3 バケット（複数指定可能）

機密データ 検出オプション

- 実行スケジュール
- サンプリング深度
- オブジェクト条件

カスタムデータ識別子オプション

Amazon S3

Bucket-A

Bucket-B

Bucket-C

aws

参考：S3 バケット以外のデータをスキャンしたい場合は？

- Amazon Macie は現在 S3 バケットのスキャンのみをサポートしている

- ワークアラウンド
- Amazon RDS を Parquet 形式で S3 バケットにエクスポートして、それをスキャンする

- 他の AWS サービスやデータストアで CSV でエクスポート出来る場合には、それを S3 バケットの上にエクスポートしてスキャンする

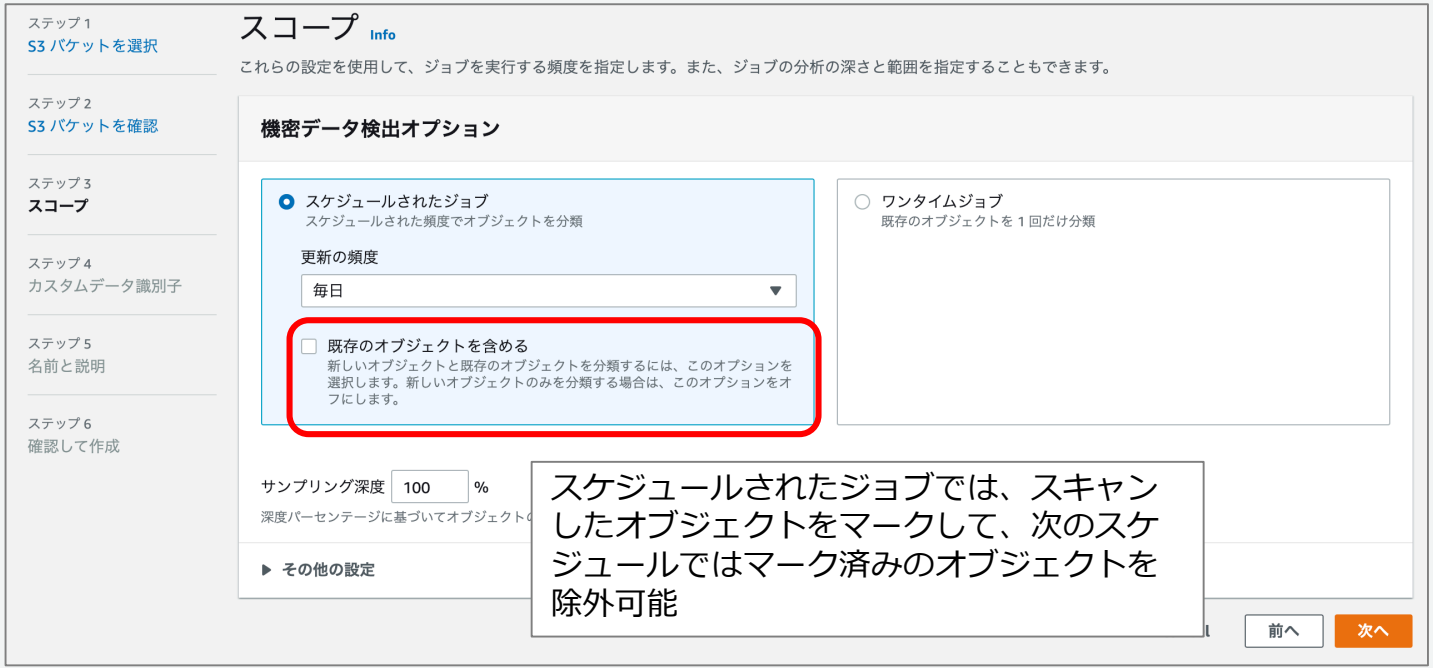
Macie の機微情報検出方法

- Macie は機械学習とパターンマッチングを組み合わせて機微情報を検出する

1. キーワードが存在するかチェック
※ キーワード不要なケースあり
2. キーワードから近い位置（距離が短い）に機微情報があるかをチェック
3. 機微情報に該当する文字列があるかチェック

スキャンスコープの設定 1

- ワンタイムの実行とスケジュール実行の選択が可能
- スケジュールされたジョブでは、どこまでスキャンをしたかを保存しておくことで、差分スキャンを行わせることが可能



設定可能なスケジュール

- 毎日、毎週、毎月を選択可能
- ジョブが実行される時間帯や日付を指定することは出来ない

サンプリング深度とは？

- ジョブの中でサンプリング深度を指定（デフォルトは 100）
- 100（＝デフォルト）を指定すると、全てのオブジェクトをスキャンする
- 100 未満を指定すると、ランダムに指定された比率のオブジェクトをスキャンする。100オブジェクト存在した場合に、90を指定すると90オブジェクトをランダムに選択してスキャンする

aws

スキャンスコープの設定 2

- スキャン対象にする条件、除外する条件を設定することで、スキャン対象を絞り込む設定が可能
- 条件には、タグ、最終改訂日、ファイル拡張子、オブジェクトサイズを利用可能
- 除外条件は対象にする条件を上書きする（両方マッチした場合は除外される）

JSON ファイルのみ対象と設定した例

Macie のデータ識別子

AWS マネージド識別子

- AWS がマネージして提供する機微情報検出用の定義
- お客様によるカスタマイズは出来ない
- Macie のジョブを実行すると自動で検出が行われる

カスタマーマネージド識別子

- お客様が検出定義を作成して独自の検出を行うことが可能
- キーワードと正規表現のパターンマッチング定義を作成する必要がある
- ジョブ単位に指定を行う必要がある（作成した識別子定義は再利用可能）

AWS マネージド識別子の対象情報 1 （抜粋）

カテゴリ	データタイプ（抜粋）	キーワード	補足情報	対応している国とリージョン
個人特定情報（PII）	生年月日	Bday, b-day, birth date, birthday, date of birth, dob		全て
	姓名	必要なし	ラテン言語の文字セットのみ	全て
	住所	必要なし	ラテン言語の文字セットのみ	全て
	パスポート番号	https://docs.aws.amazon.com/macie/latest/user/managed-data-identifiers.html#managed-data-identifiers-pii-passport-keywords		カナダ、フランス、ドイツ、イタリア、スペイン、UK、US
	電話番号	cel, cell, celular, contact, fone, mobile, móvel, número residencial, numero residencial, phone, tel, telephone, telephone number		ブラジル、カナダ、フランス、ドイツ、イタリア、スペイン、UK、US

参考 : https://docs.aws.amazon.com/macie/latest/user/managed-data-identifiers.html

AWS マネージド識別子の対象情報 2（抜粋）

カテゴリ	データタイプ（抜粋）	キーワード	補足情報	対応している国とリージョン
金融関連情報	銀行口座番号	https://docs.aws.amazon.com/macie/latest/user/managed-data-identifiers.html#managed-data-identifiers-financial-bankacct-keywords		カナダ、フランス、ドイツ、イタリア、スペイン、UK、米国
	クレジットカード失効日	expiration, expiry	サポートされるフォーマットは MM/YY もしくは YY/MM	全て
	クレジットカードの磁気ストライプデータ	ard data, iso7813, mag, magstripe, stripe, swipe	Tracks 1 と 2を含む	全て
	クレジットカード番号	なし	キーワードは不要だが、ルーンアルゴリズムによるチェックディジットを含む13桁から19桁のフォーマットが必要	全て

参考：https://docs.aws.amazon.com/macie/latest/user/managed-data-identifiers.html

AWS マネージド識別子の対象情報 3（抜粋）

カテゴリ	データタイプ（抜粋）	キーワード	補足情報	対応している国とリージョン
認証情報	AWS シークレットキー	aws_secret_access_key, credentials, secret access key, secret key, set-awscredential		全て
	OpenSSH 秘密鍵	なし		全て
	PGP 秘密鍵	なし		全て
	PKCS 秘密鍵	なし		全て
	PuTTY 秘密鍵	なし		全て

参考 : https://docs.aws.amazon.com/macie/latest/user/managed-data-identifiers.html

日本語対応

- サポートされている文字コード
  - UTF-8 のみがサポートされている
  - 他の文字コード（Shift_JIS、EUC）の文字が入ったファイルはスキャン対象外となる（スキップされる）

- AWS マネージドの識別子では、日本語のキーワードが登録されていない
  - 例 Credit Card というキーワードがあると検出するが “クレジットカード” というキーワードは検出しない

- カスタマーマネージドの識別子では、日本語のキーワードを指定可能
  - マネージメントコンソールで日本語のキーワードを指定した場合、UTF-8として解釈される、正規表現で日本語の文字とのマッチングを指定した場合もUTF-8の場合はマッチングが行われる

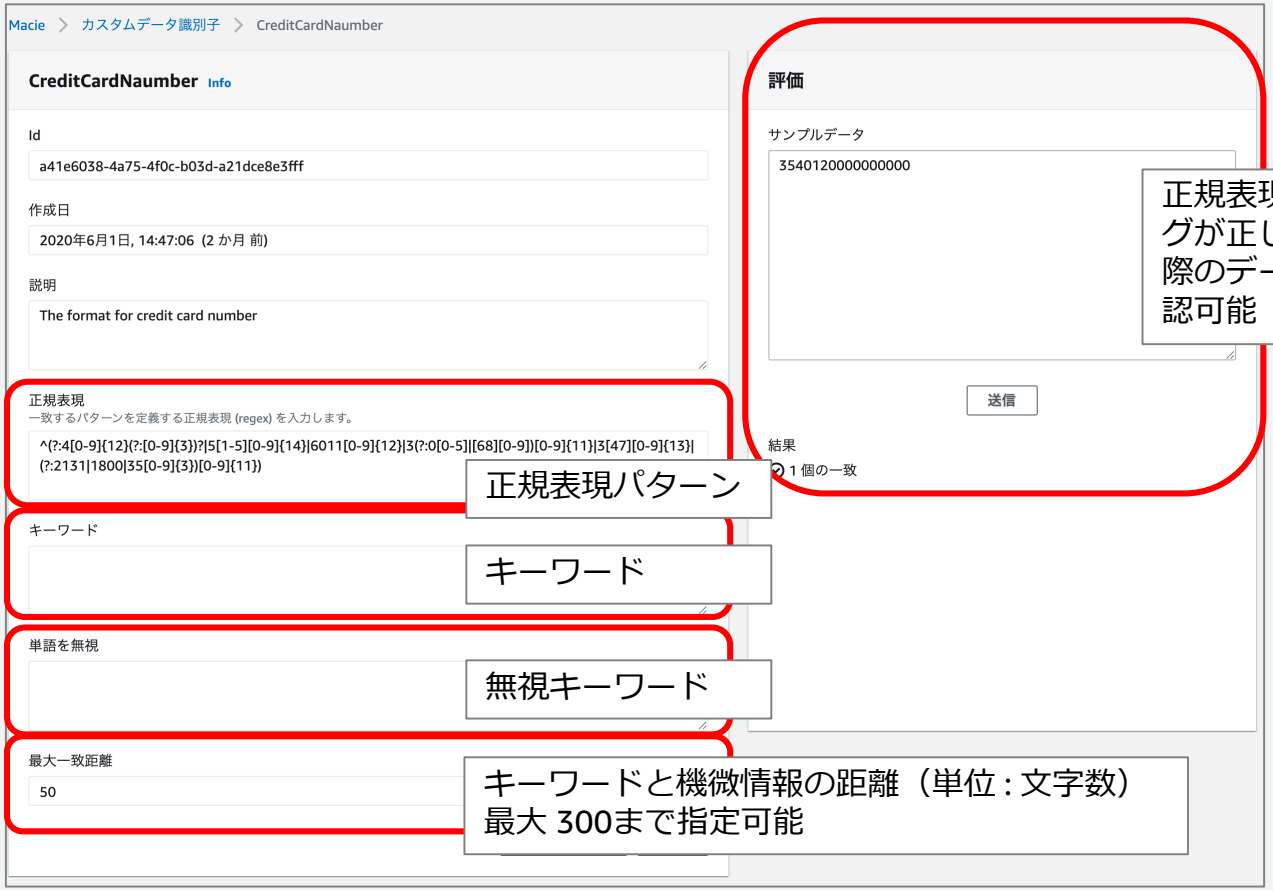
カスタムデータ識別子

- Macie は PCRE（Perl Compatible Regular Expressions）を利用した正規表現によるパターンマッチングを利用してお客様のユースケースに応じたカスタム検出をサポート

- カスタムデータ識別子で指定可能なパターン
    - キーワード
    - 除外キーワード
    - 最大一致距離
    - 正規表現パターン

参照 https://docs.aws.amazon.com/macie/latest/user/custom-data-identifiers.html

カスタムデータ識別子の定義



正規表現による機微情報のマッチング例

クレジットカード番号

VISA の例、4で始まる16桁の数字
^4[0-9]{12}(?:[0-9]{3})?$

IP アドレス

例 192.0.2.5
^(?:(?:25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?)¥.){3}(?:25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?)$

※ Macie がサポートしない正規表現要素
- 前方参照
- キャプチャグループ
- 条件つきパターン
- 埋め込みコード
- /i, /m, /x のようなグローバルフラグ
- 再帰的パターン
- ?=, ?!, ?<=, ?<! などの肯定先読み、否定後読み

暗号化された S3 バケットのスキャン

- SSE-S3, AWS-KMS, SSE-KMS による暗号化が行われている場合、 Macie はバケットをスキャンすることが可能
    - SSE-S3, AWS-KMS の場合は追加設定不要でスキャン可能
    - SSE-KMS の場合には、Macie のサービスリンクロール AWSServiceRoleForAmazonMacie に CMK へのアクセス権限の付与が必要

- クライアントサイド暗号化、カスタマー提供型のサーバーサイド暗号化（SSE-C）の場合には、復号する手段が無いため、Macie はスキャンが出来ない
    - バケットの使用状況の可視化は S3 のメタデータを使っているので、上記のタイプの暗号化がされていても実行可能

サポートされるファイルタイプ

ファイル/ ストレージタイプ	説明	ファイル拡張子
ビッグデータ	Apache Avro オブジェクトコンテナ、Apache Parquet ファイル	N/A
圧縮・アーカイブ	GNU Zip 圧縮アーカイブ、Tar アーカイブ、 ZIP 圧縮アーカイブ	.gz, .gzip, .tar, .zip
文書	Adobe PDF ファイル、Microsoft エクセルワークブック、Microsoft Word 文書	.doc, .docx, .pdf, .xls, .xlsx
テキスト	CSV ファイル、HTMLファイル、JSONファイル、プレーンテキスト文書、タブ区切りファイル（TSV）、XMLファイル	.csv, .htm, .html, .json, .tsv, .txt, .xml

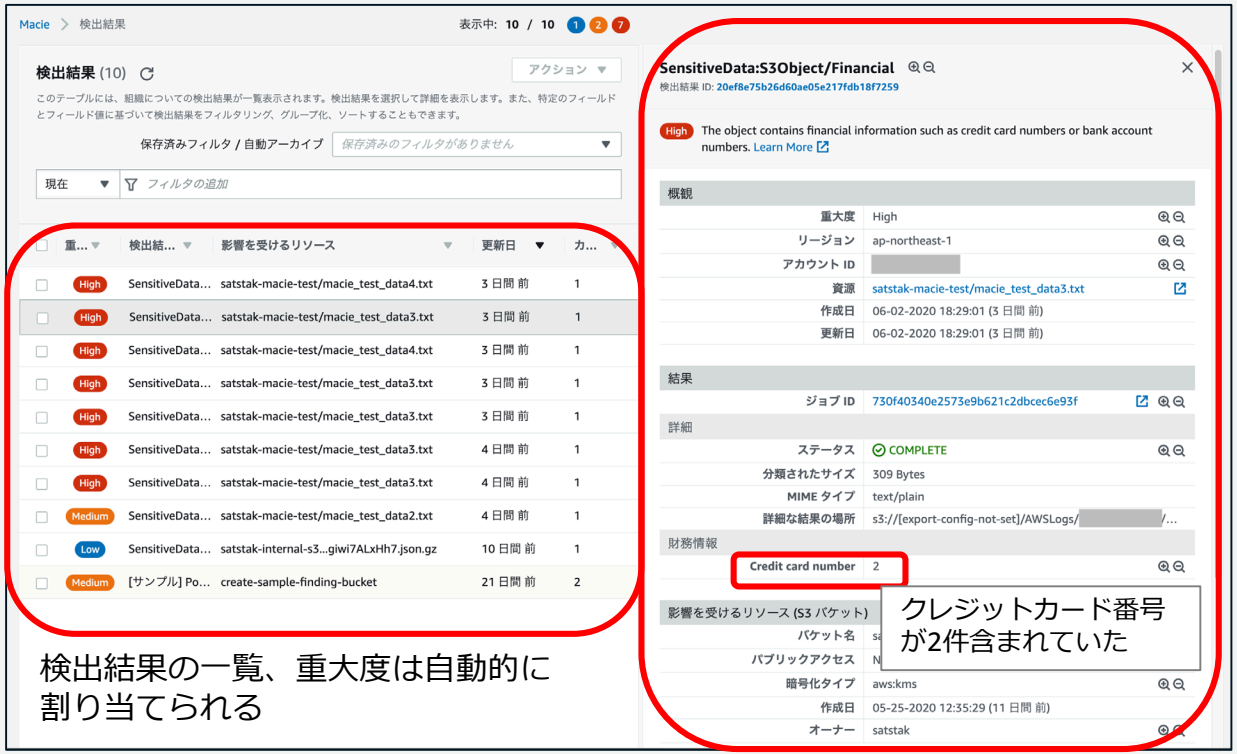
※ Macie は画像、オーディオ、ビデオなどのメディアファイルは分析しない、また、文字コードやサイズの制限があるので注意する

参考 : https://docs.aws.amazon.com/ja_jp/macie/latest/user/discovery-supported-formats.html

3. 評価・検出結果の参照、他のサービスとの連携

マネジメントコンソール上での検出結果の確認

検出ジョブの実行が完了すると、マネジメントコンソール上で検出結果の確認が出来る



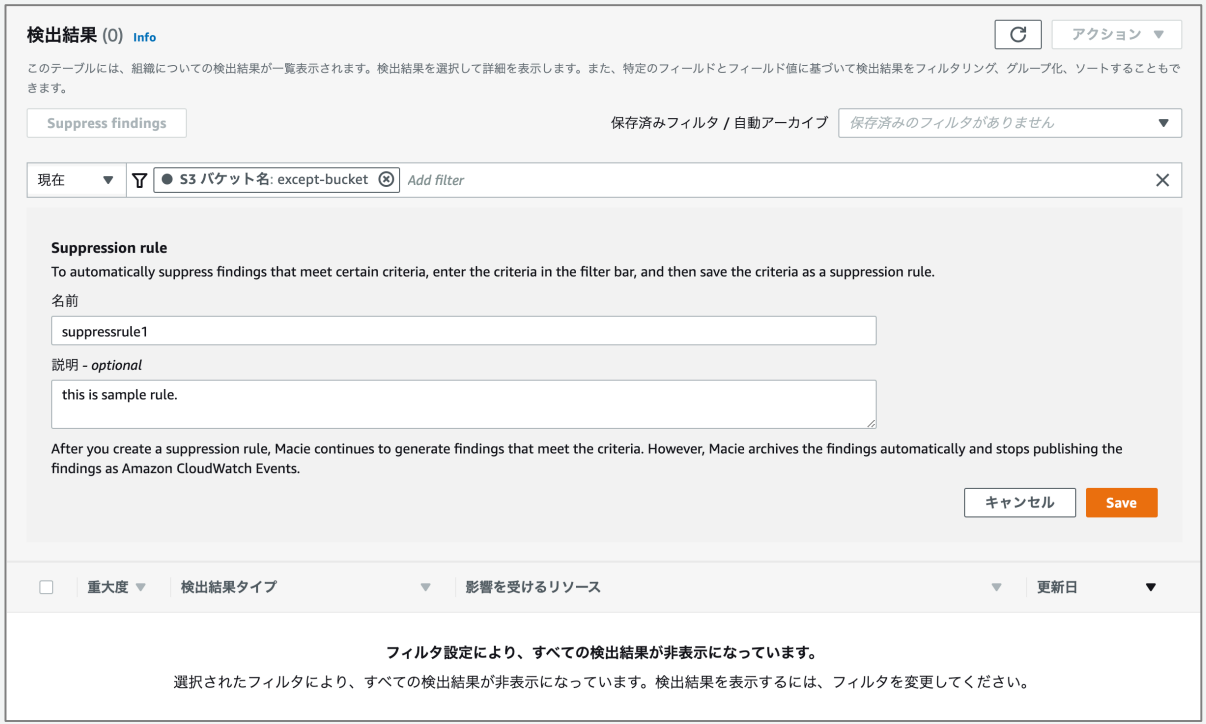
検出結果の一覧、重大度は自動的に割り当てられる

検出結果を選択すると詳細が表示される

- ファイル単位に検出結果が出力される
- 何も検出されなかった場合には出力されない
- 1つのファイルに複数の機微情報が入っていた場合には１つの結果が検出結果が出力され、その中にいくつのデータがあったかが表示される
- 何行目に含まれていたという情報は検出結果に含まれない

検出結果の非表示化

検出結果をフィルター表示する画面で、
指定した条件の Findings を自動アーカイブする Suppression Rule を作成して適用することが可能
不要な表示を減らすことで、管理を容易にする

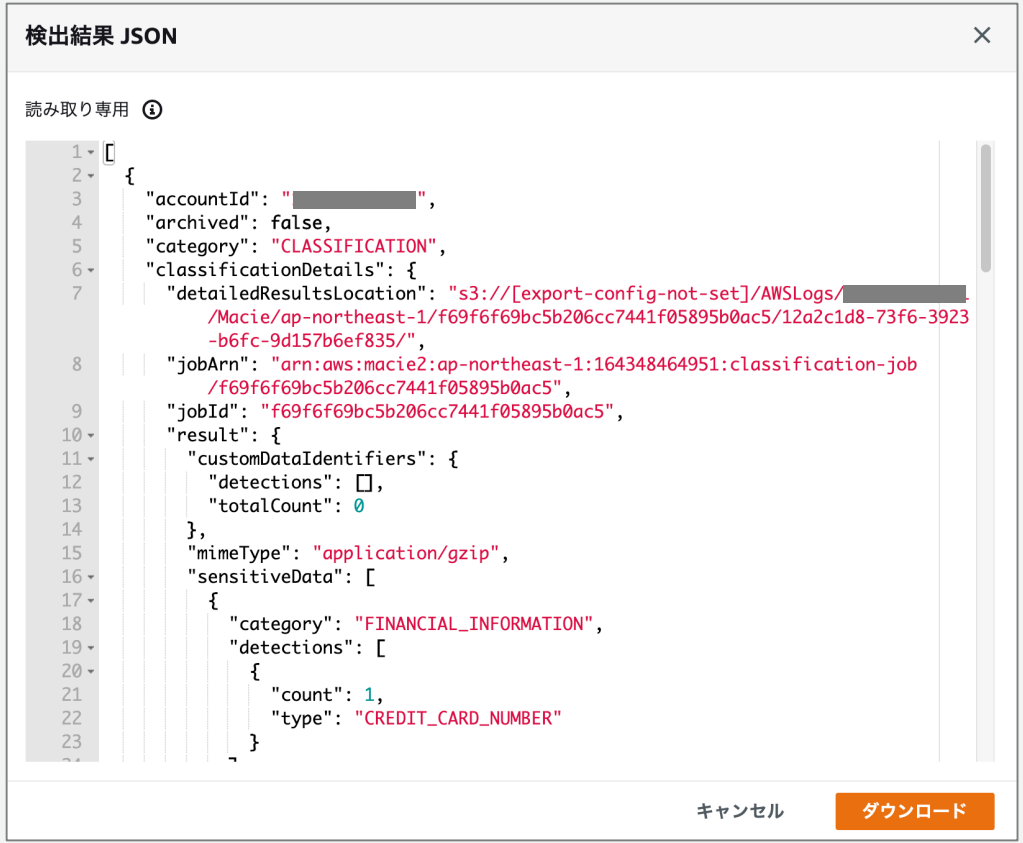


適用出来る条件（抜粋）

- アカウントID
- カテゴリ
- 検出結果タイプ
- 重大度
- S3 オブジェクトタグキー、タグ値
- S3 バケットタグキー、タグ値
- S3 バケット名

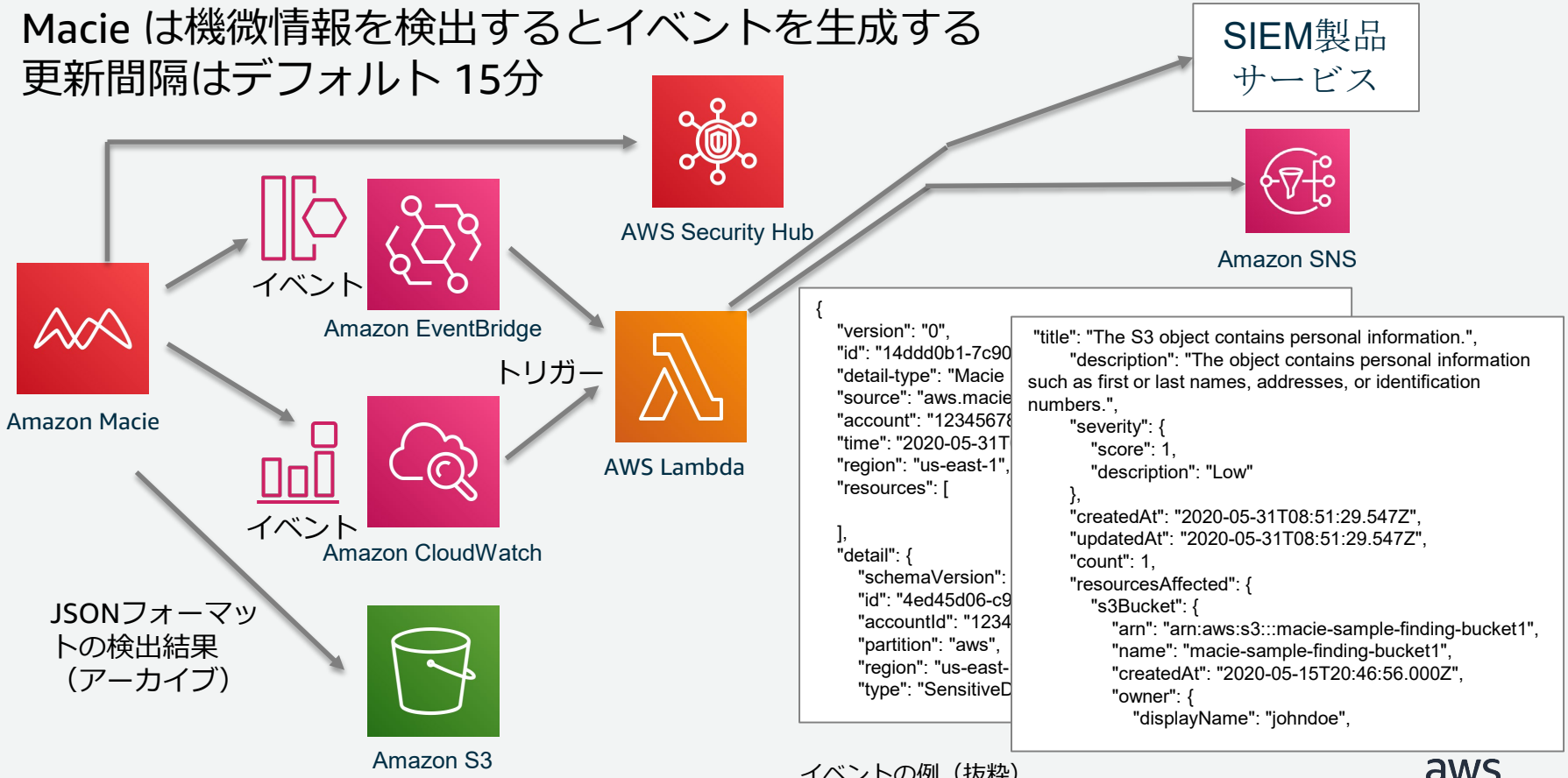
アクションの実行

- 検出結果を選択して、アクションメニューから JSON へのエクスポートが可能
- ブラウザ経由でローカルにダウンロードされる
- 同じ内容を S3 に自動保存する設定をすることが可能
  - Macie の検出結果は、30日間のみ保管されているので、長期間保管したい場合には、S3 に保存することを推奨する



イベントの生成と AWS サービスとの連携

Macie は機微情報を検出するとイベントを生成する
更新間隔はデフォルト 15分

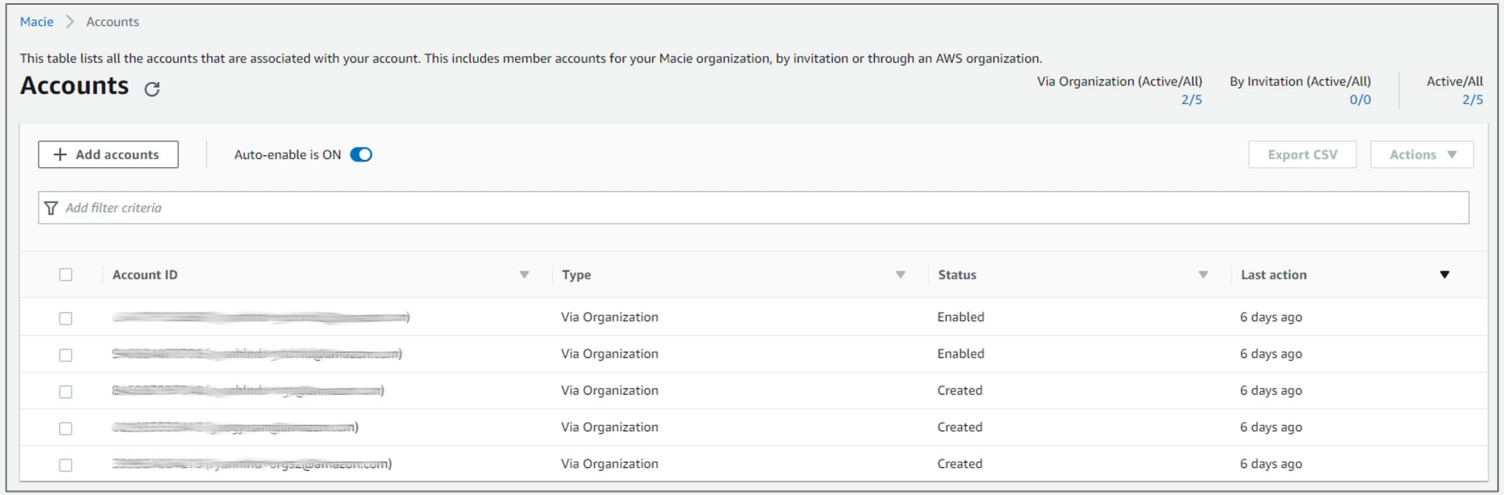


イベントの例（抜粋）

4. 管理、運用、制約

マルチアカウント環境における Macie の管理

- マスター/メンバー構成をとることが可能
- マスターアカウントの Macie でジョブを作成して、メンバーアカウントで実行出来る
- 手動によるメンバーアカウント追加の場合 1000 アカウントまでサポート
- AWS Organizations 構成の場合は 5000 アカウントまでサポート



Macie API コールのロギング

- CloudTrail への API コール記録をサポートしている


```
{
    "eventVersion": "1.05",
    "userIdentity": {
        "type": "IAMUser",
        "principalId": "123456789012",
        "arn": "arn:aws:iam::123456789012:user/Mary",
        "accountId": "123456789012",
        "accessKeyId": "AKIAIOSFODNN7EXAMPLE",
        "userName": "Mary",
    },
    "eventTime": "2020-05-22T16:09:56Z",
    "eventSource": "macie2.amazonaws.com",
    "eventName": "ListFindings",
    "awsRegion": "us-east-1",
    "sourceIPAddress": "198.51.100.1",
    "userAgent": "console.amazonaws.com",
    "requestParameters": {
        "sortCriteria": {
            "attributeName": "updatedAt",
            "orderBy": "DESC"
        },
```

```
  "findingCriteria": {
            "criterion": {
                "archived": {
                    "eq": [
                        "false"
                    ]
                },
                "category": {
                    "eq": [
                        "POLICY"
                    ]
                }
            }
        },
        "maxResults": 10
    },
    "responseElements": null,
    "requestID": "d58af6be-1115-4a41-91f8-ace03EXAMPLE",
    "eventID": "ad97fac5-f7cf-4ff9-9cf2-d0676EXAMPLE",
    "eventType": "AwsApiCall",
    "recipientAccountId": "123456789012"
}
```

API, AWSCLI からの呼び出し

新しい Macie は サービスとして macie2 を指定して API や AWS CLI から呼び出しが可能

※ 東京リージョンで利用出来なかった 前のバージョンの Macie（v1）があるため、2 がついている

```
$ aws macie2 list-findings
{
    "findingIds": [
        "495eb82347841ba75f56f82cc35ba5bd",
        "c8c99a7b6617b43c1e7ad459555120b0",
        "b0e0cd61f868a825307b7f69c04e8e56",
        "f939afe6c315636b44d6cd40620e777a",
        "7e7c1318d2815e122ebd92243460c264",
        "ce64ea47a88a25558e7b6a39e7271f14",
        "58f8487fd7091a9949cc715804715d14"
    ]
}
```

```
$ aws macie2 get-findings --finding-ids 495eb82347841ba75f56f82cc35ba5bd
{
    "findings": [
        {
            "accountId": ”000000000000",
            "archived": false,
            "category": "CLASSIFICATION",
            "classificationDetails": {
                "detailedResultsLocation": "s3://[export-config-not-
set]/AWSLogs/164348464951/Macie/ap-northeast-
1/f69f6f69bc5b206cc7441f05895b0ac5/12a2c1d8-73f6-3923-b6fc-9d157b6ef835/",
                "jobArn": "arn:aws:macie2:ap-northeast-1:164348464951:classification-
job/f69f6f69bc5b206cc7441f05895b0ac5",
                "jobId": "f69f6f69bc5b206cc7441f05895b0ac5",
                "result": {
 以下省略
```

参照 : https://docs.aws.amazon.com/cli/latest/reference/macie2/index.html

Macie のセキュリティ

- Macie は Amazon S3 のデータをスキャンするが、データそのものや、スキャン結果が勝手に利用されることはないか？
  - Macie によるスキャンデータは解析のためにメモリー上で一時的に保持されるが、他のストレージにコピーされることはない
  - スキャン以外の目的で Macie のサービスは、お客様の S3 のデータにアクセスしたり、スキャン結果を利用することはない
  - Macie によるスキャン、分類結果は、暗号化されたストレージに保管される

- Amazon S3 と Macie の間の通信は暗号化される

参考 : https://docs.aws.amazon.com/macie/latest/user/data-protection.html

サポートされるリージョン （2020年8月現在）

- Macie はリージョン別のサービスで、リージョンごとに有効化する必要がある

- Macie を利用可能なリージョン
    - 米国（GovCloudを除く）、サンパウロ、東京、シンガポール、シドニー、ソウル、ムンバイ、香港、フランクフルト、アイルランド、ロンドン、パリ、ストックホルム

サポートされるリージョンの情報
https://aws.amazon.com/jp/about-aws/global-infrastructure/regional-product-services/

制約、クオータ

制約

- メンバーアカウントの数 : 手動の場合 1,000 アカウント、 AWS Organizations 経由の場合 5,000 アカウント
- 検出ジョブ単位の Findings の数 100,000 + 検出対象のオブジェクト数の 5% が限度
- 機微情報の検出と分類
    - 月あたりの検出と分類の上限 5TB、サービスクオータコンソールから25TB までの増加をリクエスト可能。さらに増やす場合にはサポートにご相談ください
    - データ検出結果中の検出ロケーション : 検出タイプごとに 1,000
    - データ検出ジョブに設定出来るカスタムデータ識別子の数 : 30
    - 圧縮されたりアーカイブされているファイルの解凍の上限 :
      ネストされたアーカイブの深さの制限 : 10 、解凍するファイル数 1,000,000
    - ファイルサイズの制限 : PDF 1024MB, テキストファイル 20GB, Microsoft Excel /Word 512MB（一部を抜粋、詳細は下記URL参照）
      https://docs.aws.amazon.com/macie/latest/user/macie-quotas.html

利用料金

- 課金対象
  - S3 バケットの数（月ごと）
  - 機微情報検出対象となったデータ処理量（月ごと）
- 東京リージョンの例
  - S3 バケットあたり 0.10 USD
  - データ処理量
    - 50,000GB /月までGBあたり 1.25 USD
    - 450,000GB /月までGBあたり 0.63 USD
    - 500,000DB /月を超えた場合、GBあたり0.31USD

- Macie が S3 にアクセスする際の、GET および LIST について S3側の料金もかかる

- 無料試用期間 30日
  - バケットの可視化のみが対象で、スキャンに関しては下記のルールが適用される
  - スキャンに関して、1ヶ月 1GBまで無料（30日間限定ではない）

まとめ

Amazon Macie を利用することで

S3 バケット上の大量の機微情報の検出を効率的に実行可能

容易な設定と管理が実現可能

AWS マネージド識別子と、カスタム識別子の両方を
活用することで効率よく検出が可能

スケジュールして定期的にスキャンを行うことで、S3 バケット上に
意図しない機微情報が含まれていることがないことを可視化可能

参考資料

- Amazon Macie ドキュメント
  https://docs.aws.amazon.com/ja_jp/macie/index.html
  ※ 2020年8月現在 英語のみの提供

AWS の日本語資料の場所「AWS 資料」で検索

https://amzn.to/JPArchive

Q&A

お答えできなかったご質問については
AWS Japan Blog 「https://aws.amazon.com/jp/blogs/news/」にて
後日掲載します。

AWS Well-Architected 個別技術相談会

毎週”W-A個別技術相談会”を実施中

- AWSのソリューションアーキテクト(SA)に対策などを相談することも可能
- 申込みはイベント告知サイトから

(https://aws.amazon.com/jp/about-aws/events/)

AWS イベント で[検索]

ご視聴ありがとうございました

AWS 公式 Webinar
https://amzn.to/JPWebinar

過去資料
https://amzn.to/JPArchive