Data Discovery and Classification with Amazon Macie

Presenter | Date

Agenda

- Amazon Macie Overview
- Key features & changes
- Pricing and availability

aws

Key Drivers

Productivity trends result in organizations generating growing volumes of data, distributed across multiple shared resources

Increased challenge to comply with data privacy regulations with new additional requirements and expanded definitions of sensitive data

Customers have to take on the expense of monitoring and complexity of continuously updating data classifications

aws

Amazon Macie – How it works

Amazon Macie

Enable Amazon Macie with one-click in the AWS Management Console or a single API call

Continually evaluate your S3 environment

Automatically generates an inventory of S3 buckets and details on the bucket-level security and access controls

Discover sensitive data

Analyzes buckets using machine learning and pattern matching to discover sensitive data, such as personally identifiable information (PII)

Take action

Generates findings and sends to Amazon CloudWatch Events for integration into workflows and remediation actions

aws

Amazon Macie - Gain visibility and evaluate

- Provides customers visibility into S3 bucket inventory
    - Number of buckets
    - Storage size
    - Object count

- Monitors changes to S3 bucket policies
    - Publicly accessible
    - Unencrypted
    - Shared outside of the account
    - Replicated to external accounts

Across multiple accounts and automatically includes new buckets

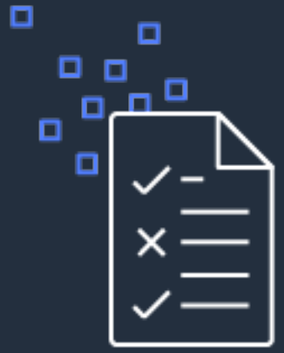
aws

Amazon Macie – Discover sensitive data

- Ongoing evaluation of your Amazon S3 environment and data

- Select target for data discovery
- Create and schedule jobs

- Define the scope
- Scheduled frequency (one-time, daily, weekly, monthly)
- Object criteria (Tags, modified time, extension type, size)



- Review status (complete, cancelled, idle)
- Take actions (Cancel, copy)

aws

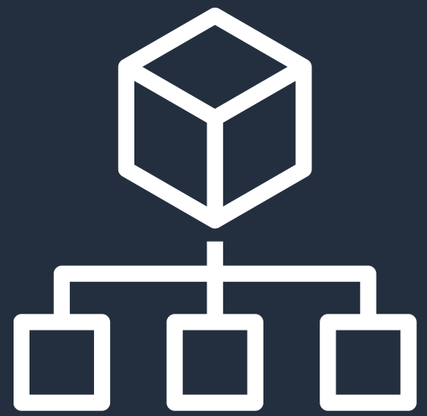
Amazon Macie - Centrally manage at scale

Master/Member setup

- Multi-account with up to 1,000 member accounts
- AWS Organizations support up to 5,000 accounts

Macie master can create jobs on behalf of members

- One-click deployment with no upfront data source integration



With a few more clicks in the console, you can enable Macie across multiple accounts. Once enabled, Macie generates an ongoing Amazon S3 resource summary across accounts that includes bucket and object counts as well as the bucket-level security and access controls.

aws

Amazon Macie - Centrally manage at scale

Fully managed sensitive data types

- Amazon Macie maintains a growing list of sensitive data types that include common personally identifiable information (PII) and other sensitive data types as defined by data privacy regulations, such as GDPR, PCI-DSS, and HIPAA.

File formats

.txt .json .xml Avro
.csv .tsv
.doc .docx .xls .xlsx
.pdf
.tar .zip .gzip
Parquet

Data types

- *Financial (card, bank account numbers…)*
- *Personal (names, address, contact…)*
- *National (passport, ID, driver license…)*
- *Medical (healthcare, drug agency …)*
- *Credentials & secrets*

aws

Amazon Macie - Centrally manage at scale

Custom-defined sensitive data types

- Amazon Macie provides you the ability to add custom-defined data types using regular expressions to enable Macie to discover proprietary or unique sensitive data for your business.

- *Regular expression that defines the pattern to match*
- *Keywords that define specific text to match*
- *Ignore words that define specific text to exclude*

aws

Amazon Macie - Automate and take actions

Finding types

- Bucket policy findings
- Sensitive data discovery findings

Findings categorized by

- By bucket
- By type
- By job

Detailed and actionable security and sensitive data discovery findings

- Findings sent to CloudWatch Events
- Bucket policy findings sent to Security Hub

aws

Amazon Macie - Automate and take actions

Export findings to S3 bucket

- Show classifications

Automated actions on alerts

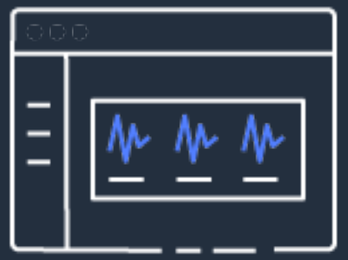
- Simplify with Lambda

Management APIs

- Integrate with additional services
- CloudTrail captures all API calls for Macie as events

aws

Amazon Macie



Gain visibility and evaluate

- Bucket inventory
- Bucket policies

Discover sensitive data

- Inspection jobs
- Flexible scope

Centrally manage at scale

- AWS Organizations
- Managed & custom data detections

Automate and take actions

- Detailed findings
- Management APIs

aws

Workshop

Overview of Workshop

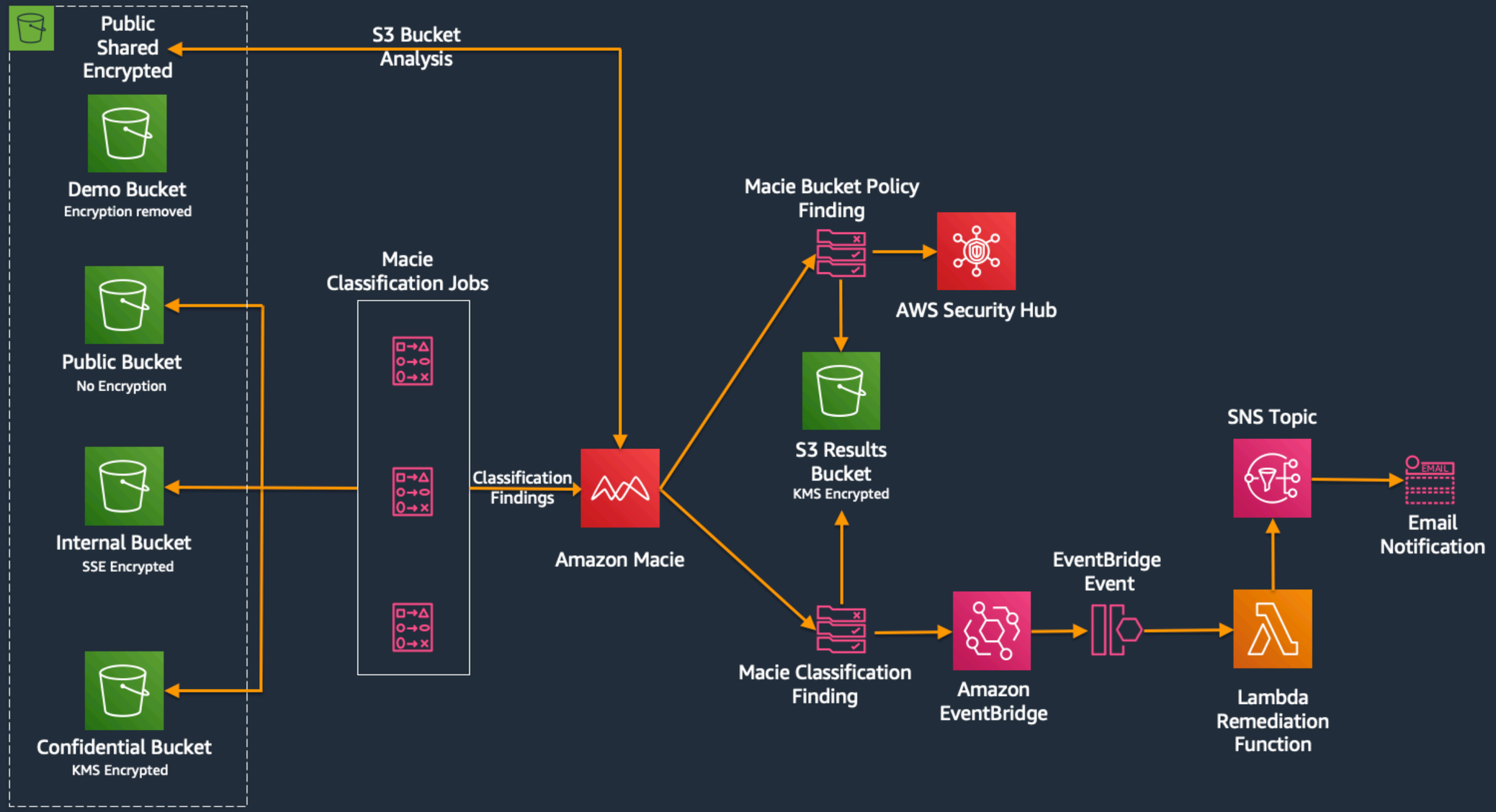
This workshop is designed to help you get familiar with Amazon Macie and learn how to scan and classify data in your S3 buckets.

There are 4 modules that you will complete.

- Setup - 20 mins
- Configure – 50 mins
- Investigate – 50 mins
- Review – 10 mins

aws

Architecture



Module 1 - Setup

1. Enable Macie and Security Hub

   You will learn how to enable Amazon Macie and AWS Security Hub

   Once the services are enabled you will complete the setup of Amazon Macie by configuring an S3 bucket for long term storage of Macie results

2. Macie and KMS CMK

https://docs.aws.amazon.com/macie/latest/user/discovery-results-repository-s3.html

aws

Module 2 - Configure

1. Create a custom data identifier

   Use a regular expression to identify project data stored in S3

2. Create a EventBridge rule

   The EventBridge rule will trigger a Lambda function to remediate incorrectly stored project files

3. Create two data classification jobs

   Use different settings to create two data classification jobs

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

aws

Module 3 - Investigate

1. Review S3 bucket inventory panel
   - Review and understand the components of the S3 bucket inventory panel
2. Review S3 policy alerts in Security Hub
   - Pivot to Security Hub
   - Review details about Macie S3 bucket policy findings
3. Investigate findings from data discovery and classification jobs
   - Create filters to isolate interesting findings
   - Learn to create a saved filter
   - Create a suppression rule to automatically archive findings

https://docs.aws.amazon.com/macie/latest/user/findings.html

aws

Module 4 - Review

1. What did you learn?
   Review skills you have learned
2. What questions could you answer about the data?
   Using filters to query the findings what did you learn?
3. What other questions could you ask and answer about the data?

4. Clean up environment if required

aws

Questions?

aws