AWS re:Invent

NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

SEC323

Data discovery and classification on AWS

Michael Ingoldby

Sr. Security Specialist SA
AWS

Marshall Jones

Sr. Security Specialist SA
AWS

Agenda

- Amazon Macie overview
- Key features
- Automate and remediate
- Workshop details
- Hands-on workshop

Amazon Macie

Gain visibility and evaluate

- Bucket inventory
- Bucket policies

Discover sensitive data

- Inspection jobs
- Flexible scope

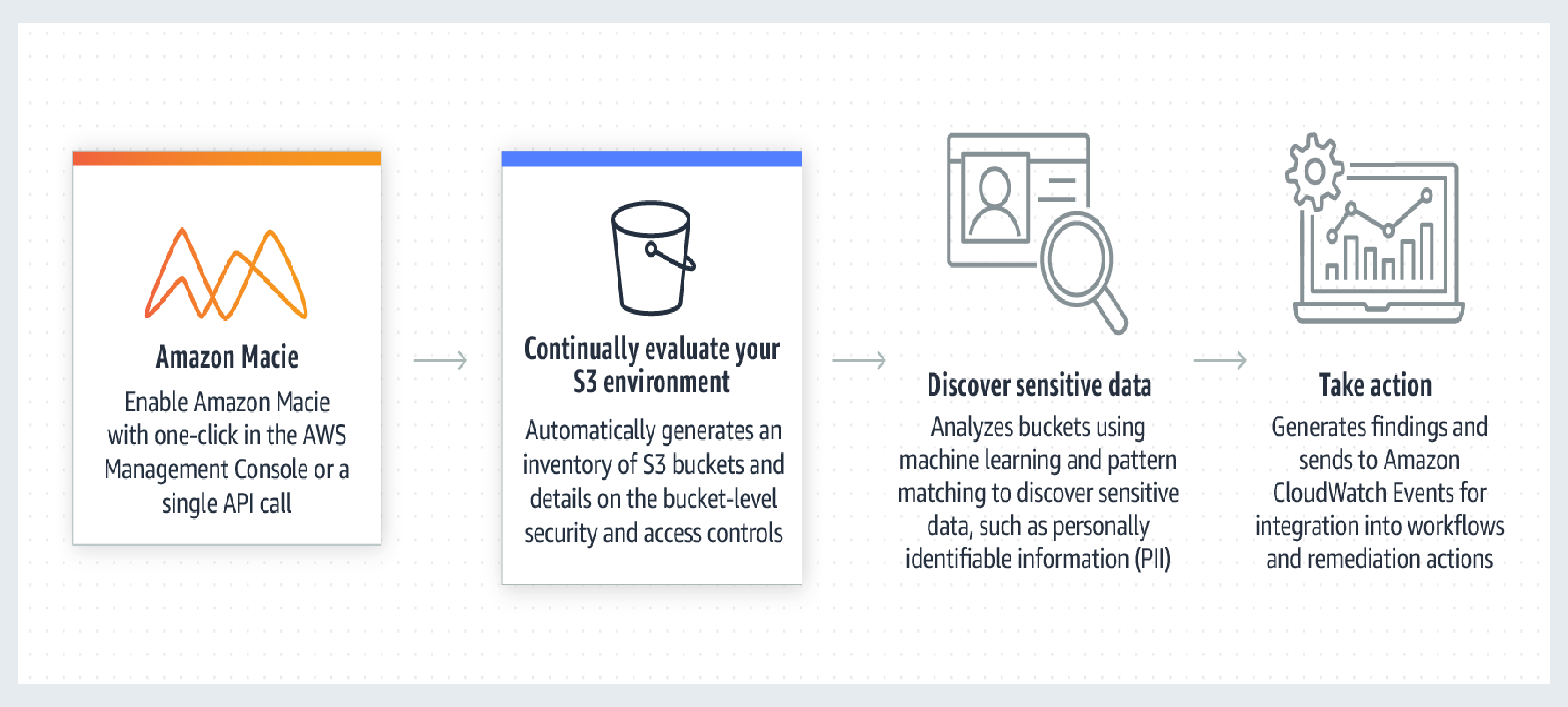
Centrally manage at scale

- AWS Organizations
- Managed and custom data detections

Automate and take actions

- Detailed findings
- Management APIs

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

Macie – Gain visibility and evaluate

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

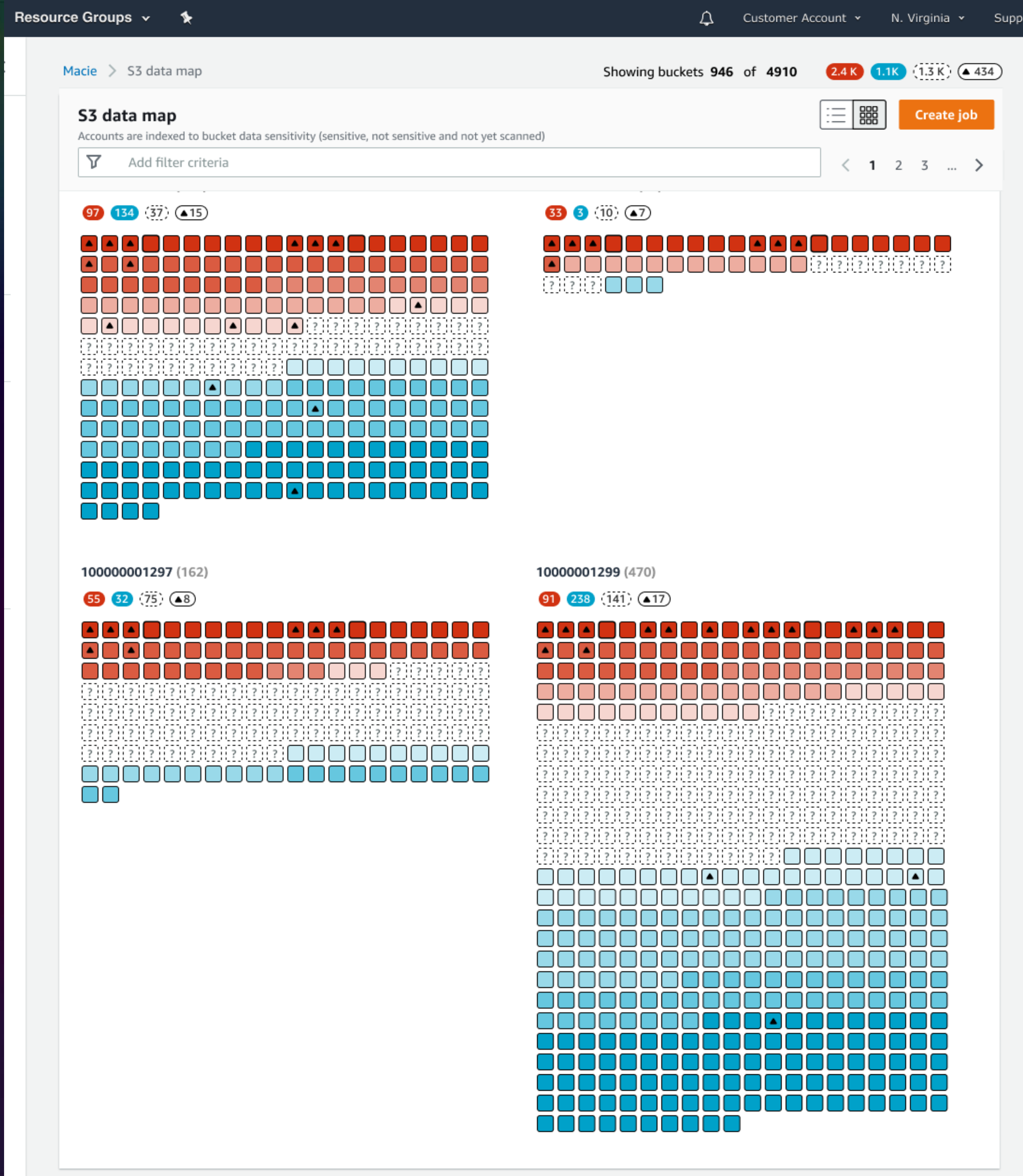
- Monitors changes to Amazon S3 bucket policies
    - Publicly accessible
    - Unencrypted
    - Shared outside of the account
    - Replicated to external accounts

Amazon Macie: Automate sensitive data discovery at scale

- Now gain cost-efficient and broad visibility into sensitive data stored in Amazon S3 across each AWS account

- Macie automatically samples and analyzes objects in your S3 buckets, inspecting them for sensitive data (such as PII), and builds an interactive heat map

- Guide your decisions to perform targeted investigations of specific S3 buckets



Macie > S3 data map

Showing buckets 946 of 4910

S3 data map

Accounts are indexed to bucket data sensitivity (sensitive, not sensitive and not yet scanned)

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
- Schedule frequency (one time, daily, weekly, monthly)
- Object criteria (tags, modified time, extension type, size)

- Review status (complete, canceled, idle)
- Take actions (cancel, copy)

Macie – Discover sensitive data

Fully managed sensitive data types

Custom-defined sensitive data types

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

File formats

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

Data types

- Financial (e.g., card, bank account numbers)
- Personal (e.g., names, address, contact)
- National (e.g., passport, ID, driver's license)
- Medical (e.g., healthcare, drug agency)
- Credentials and secrets

Macie – Automate and remediate

- Detailed and actionable security and sensitive data discovery findings
  - All findings sent to Amazon EventBridge, Amazon S3, and AWS Security Hub
- Export findings to Amazon S3 bucket
  - Show classifications
- Automated actions on alerts
  - Simplify with AWS Lambda

Macie – Centrally manage at scale

Administrator/member setup

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

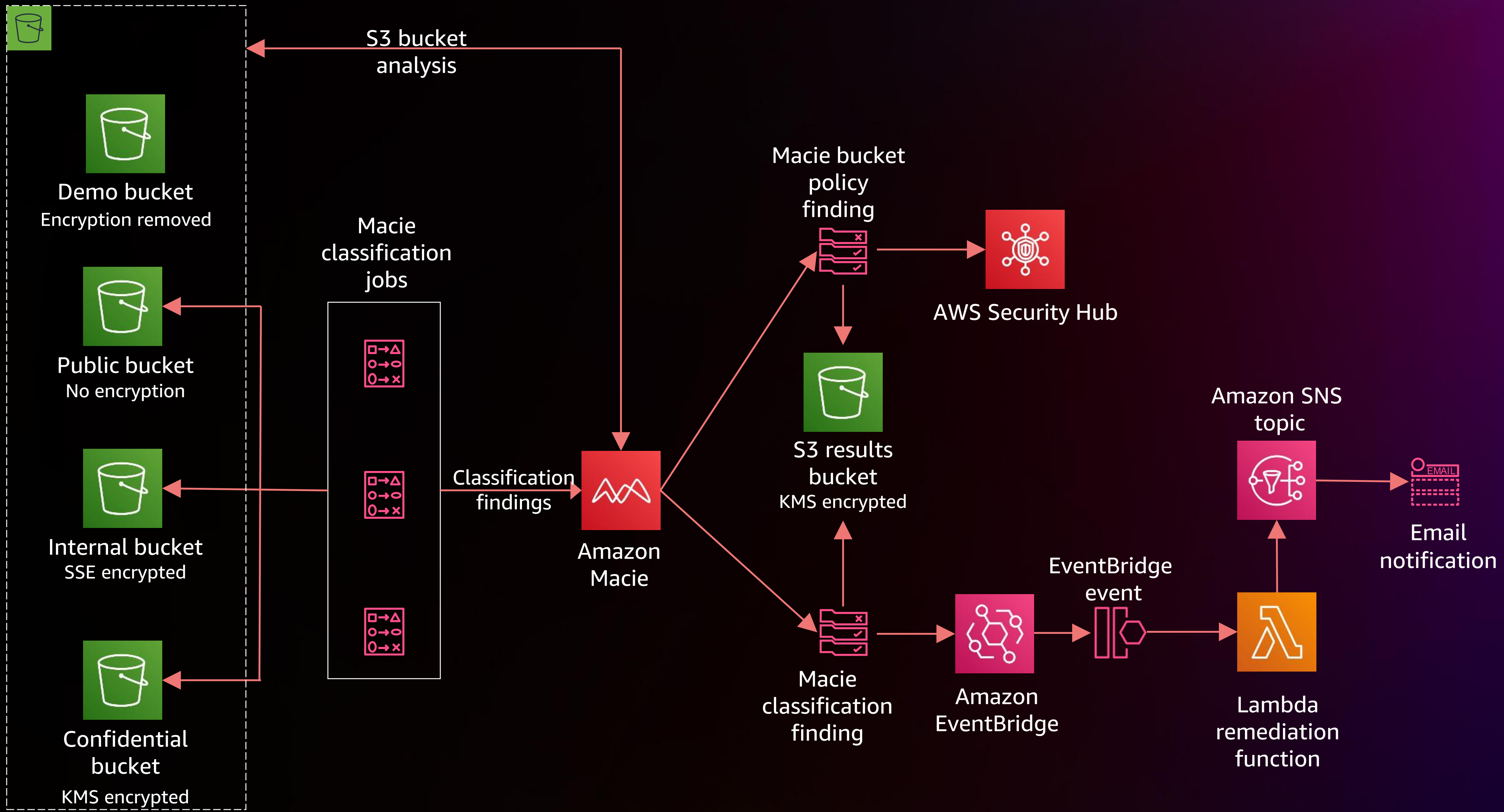
Macie administrator can create jobs on behalf of members

- 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.

Workshop details

Architecture



Overview of workshop

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

There are 5 modules that you will complete

- Setup – 20 minutes
- Configure – 50 minutes
- Investigate – 50 minutes
- AWS Security Hub – 25 minutes
- Review – 10 minutes

Module 1 – Setup

1. Enable Macie and Security Hub

   You will learn how to enable Macie and Security Hub

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

2. Run AWS CloudFormation template

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

Module 2 – Configure

1. Create a custom data identifier
   Use a regular expression to identify project data stored in S3

2. Create an 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

Module 3 – Investigate

1. Review Amazon S3 bucket inventory panel
   Review and understand the components of the Amazon S3 bucket inventory panel

2. Review Amazon S3 policy alerts in Security Hub
   Pivot to Security Hub
   Review details about Macie Amazon 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

Module 4 – Security Hub custom action

1. Create a Security Hub custom action

2. Link the custom action to a Lambda function

3. Trigger custom action to remediate

https://docs.aws.amazon.com/securityhub/latest/userguide/securityhub-cwe-custom-actions.html

Module 5 – 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?

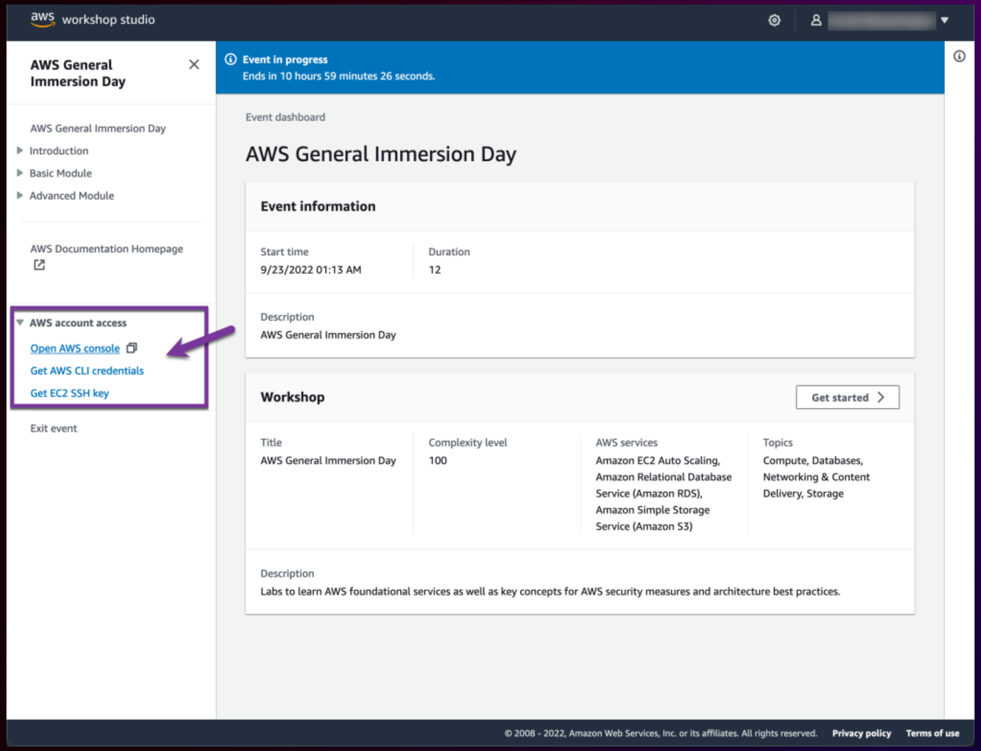
4. Clean up environment if required

Getting started with this workshop

- As a participant, you will have access to an AWS account with any optional pre-provisioned infrastructure and IAM policies needed to complete this workshop.
- The AWS account will only be available for the duration of this workshop. You will lose access to the account thereafter.
- The optional pre-provisioned infrastructure will be deployed to a specific Region. Check your workshop content to determine whether other Regions will be used.
- Be sure to review the terms and conditions of the event. Do not upload any personal or confidential information in the account.

Workshop logistics

1. Go to https://catalog.workshops.aws/join
2. Enter 12 digit event access code: b369-04491a-ce
3. Review terms and join the event
4. Access the AWS Console.
5. Get started with the workshop.



aws workshop studio [redacted]

AWS General Immersion Day ✕

AWS General Immersion Day

- Introduction
- Basic Module
- Advanced Module

AWS Documentation Homepage

AWS account access

Open AWS console

Get AWS CLI credentials

Get EC2 SSH key

Exit event

Event in progress
Ends in 10 hours 59 minutes 26 seconds.

Event dashboard

AWS General Immersion Day

Event information

Start time
9/23/2022 01:13 AM

Duration
12

Description
AWS General Immersion Day

Workshop

Get started >

Title	Complexity level	AWS services	Topics
AWS General Immersion Day	100	Amazon EC2 Auto Scaling, Amazon Relational Database Service (Amazon RDS), Amazon Simple Storage Service (Amazon S3)	Compute, Databases, Networking & Content Delivery, Storage

Description
Labs to learn AWS foundational services as well as key concepts for AWS security measures and architecture best practices.

© 2008 - 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy policy Terms of use

Thank you!

Please complete the session survey in the **mobile app**