AI-Powered Anomaly Detection:
Spotting Anomalies in Real-Time

Joe Khazen
Principal, WW GTM Specialist, Data Streaming (AWS)

Austin Groeneveld
Assoc. WW SSA Data Streaming (AWS)

Ali Alemi
Sr. WW SSA Data Streaming (AWS)

Goals for the Day

- Get hands-on experience with multiple AWS services to identify nefarious behavior and anomalies in network flow logs in real-time
- Diagnose and recommend actions
- Use Amazon Bedrock for integrating Generative AI
- Send e-mail notification to interested parties
- Network
- Have fun

Services used today

- **Amazon MSK- Amazon Managed Service for Apache Kafka**
- **ADF-Amazon Data Firehose**
- **Amazon MSF-Amazon Managed Service for Apache Flink**
- **Amazon OpenSearch**
- **Amazon SageMaker**
- **Amazon Bedrock**
- **Amazon SNS-Simple Notification Service**

Agenda

Introduction

- Discuss the business case and AWS Services Overview

Ingest and Examine Real-time Data

- Produce and validate simulated data.

Processing Data in Real-time and Create Smart Alerts with Generative AI

- Validate model in Sagemaker, start a durable MSF Application with state that queries Sagemaker in real-time, validate the output of the MSF application in MSF Studio, all while interacting with MSK Serverless. Explore LLMs on Amazon Bedrock and the Amazon SNS topic for pushing alerts.

Configure Amazon Data Firehose and OpenSearch Ingestion

- Deliver data to S3 using Amazon Data Firehose and to Amazon OpenSearch for further exploration and analysis.

Guidelines for the session

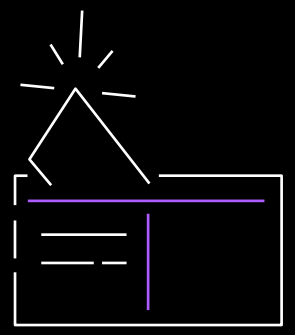
- **Stay on Mute:** Keep your microphone muted when not speaking to avoid background noise.
- **Raise Your Hand for Help in Chime:** If you need assistance, please use the “Raise Hand” feature, or type “Help” in the chat. A facilitator will assist you as soon as possible.
- **Ask Questions in the Chime Chat**: For general questions, please use the chat box. This allows us to address questions without interrupting the flow of the lab.
- **Stay for the Full Lab:** Please commit to staying for the entire session. Hopping in and out disrupts the experience for others and may cause you to miss key steps.
- **Follow Along and Participate:** This is a hands-on lab, so we encourage you to actively follow along with the exercises and avoid multitasking.

Data Streaming and Anomaly Detection

- Why? What problems are we solving?
- Architecture
- Overview of Solution Components
- Workshop (with Checkpoints)

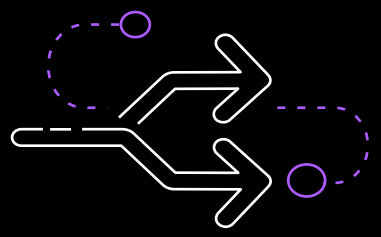
Streaming data technologies can unlock unprecedented value for organizations today

Ingest from various sources



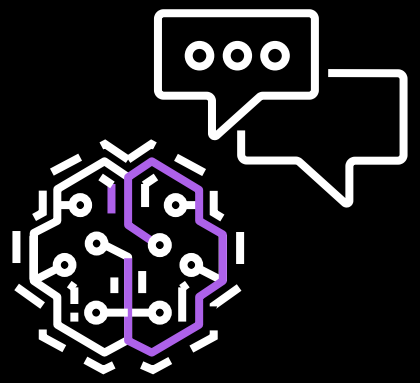
Optimize and reduce costs

Faster insights and better analytics

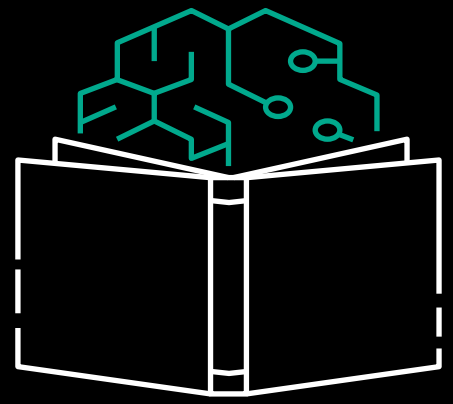


Keep your systems updated and synchronized

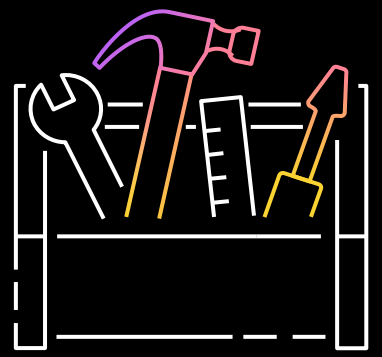
Value of Streaming Data for Anomaly Detection



Use Streaming and AI to reduce time to detect & time to respond

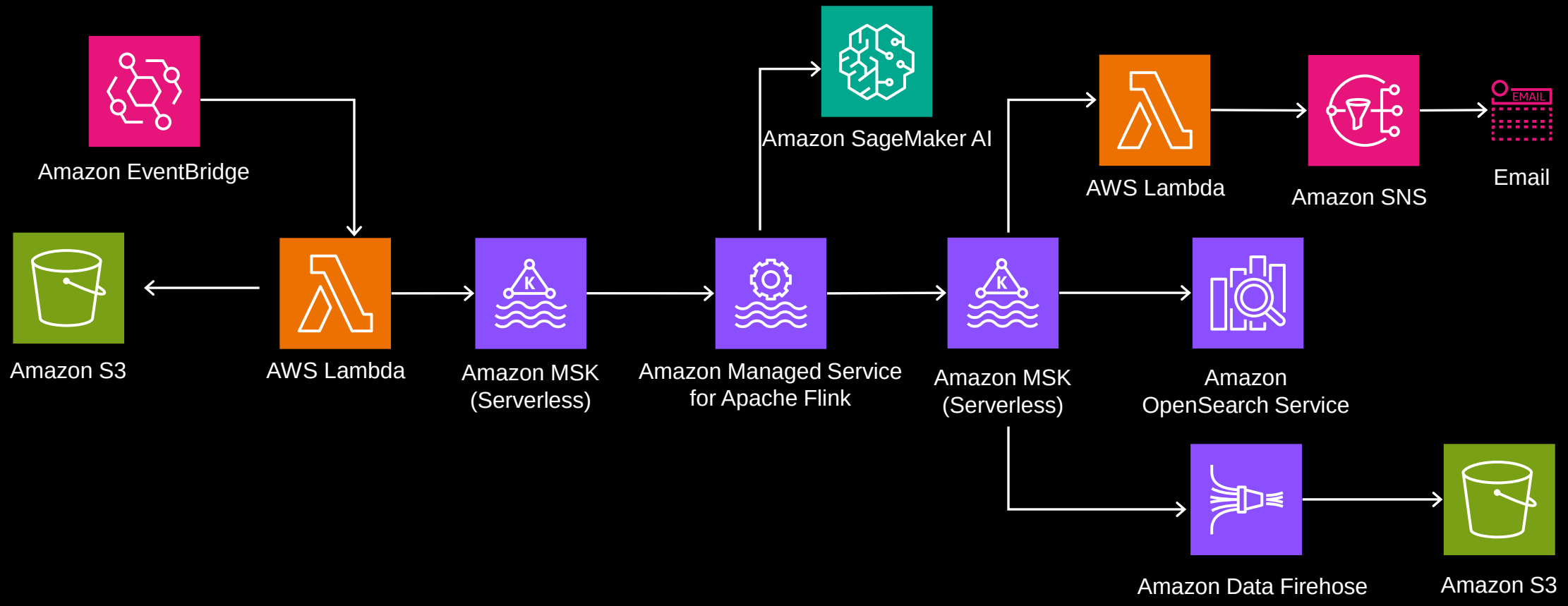


Continuous training the model new traffic patterns

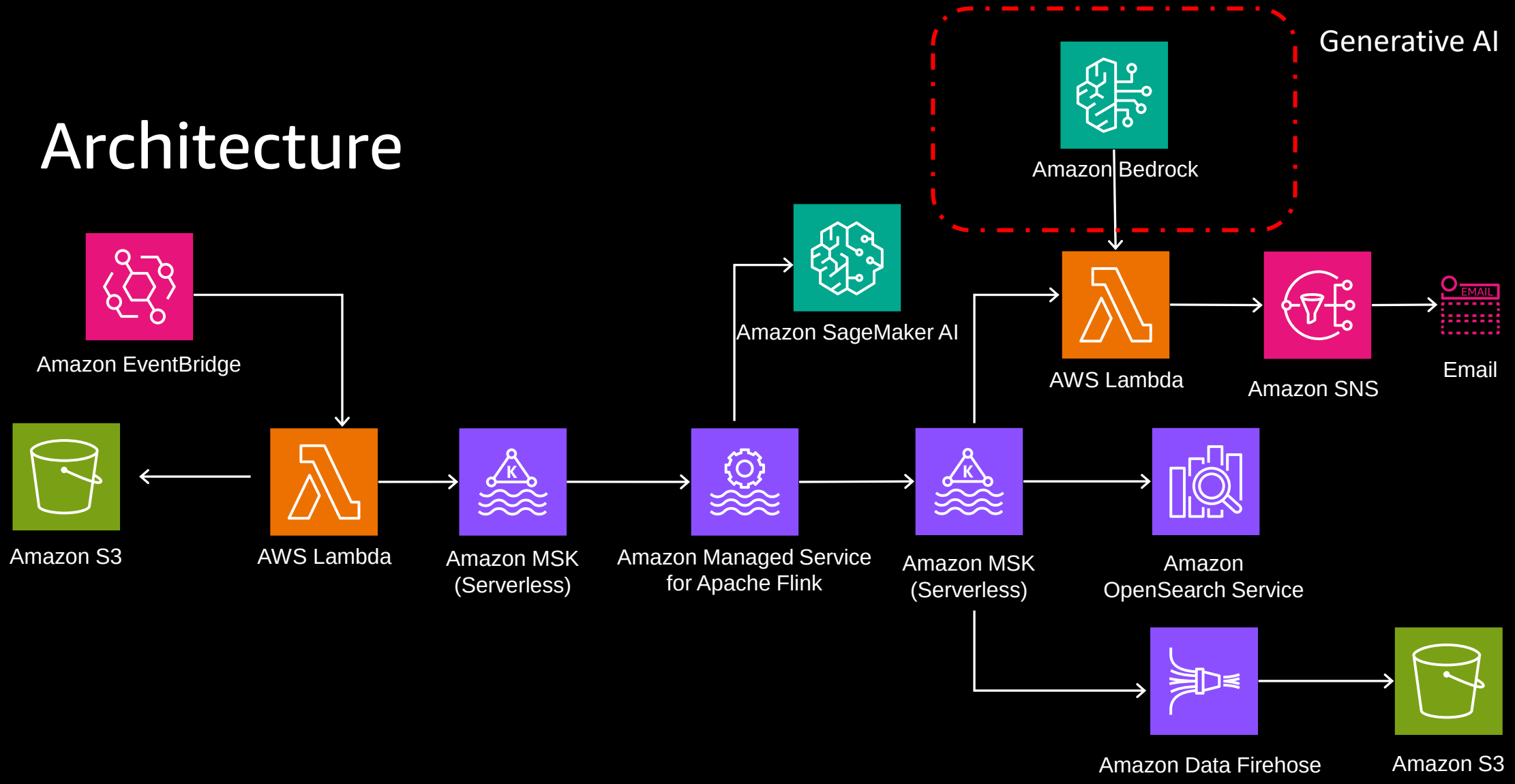


Simplifying Integrations across various components

Architecture



Architecture



Challenge Statement

Businesses would like to leverage their existing network flow logs to identify nefarious behavior and attacks on their network

Your objectives:

- Identify nefarious behavior
- Diagnose and recommend actions
- Send e-mail notification to interested parties

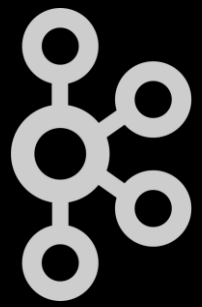
… in real-time.

How?

- Streaming Analytics leveraging open source software
  - Apache Kafka (Amazon Managed Streaming for Apache Kafka)
  - Apache Flink (Amazon Managed Service for Apache Flink)
- Detect Anomalies using Amazon Sagemaker AI
- Identify and recommend next steps using Amazon Bedrock
- And more – Lambda, SNS, S3, OpenSearch, if you are interested

Informal Poll

Familiar with Apache Kafka?



Familiar with Apache Flink?

KAFKA is a registered trademark of The Apache Software Foundation and has been licensed for use by AWS. AWS has no affiliation with and is not endorsed by The Apache Software Foundation.

Apache Kafka 101

Apache Kafka is a “Distributed Streaming Platform”

Apache Kafka has 3 core capabilities:

- Allows applications to publish and subscribe to streams of records
- Store records in the same way they were received
- Allows applications to process records as they occurred

What is Apache Flink ?

A powerful open-source framework and engine for processing data streams

Diverse use-cases

- Process both batch and streaming data
- Transform and act on streaming data
- Extract value from streaming data
- Event-driven applications
- Streaming Analytics &ETL
- Batch analytics

Amazon SageMaker AI

"End-to-End Machine Learning Platform”

- **Data Preparation**: Label, Prepare, and Process your machine learning data
- **Model Training**: Train models using built-in algorithms or custom containers.
- **Model Deployment**: Deploy models to production with auto-scaling and real-time inference.
- **MLOps**: Manage the entire machine learning lifecycle, including monitoring and governance, with integrated tools.

Amazon Bedrock

"Foundation Model Service for Generative AI"

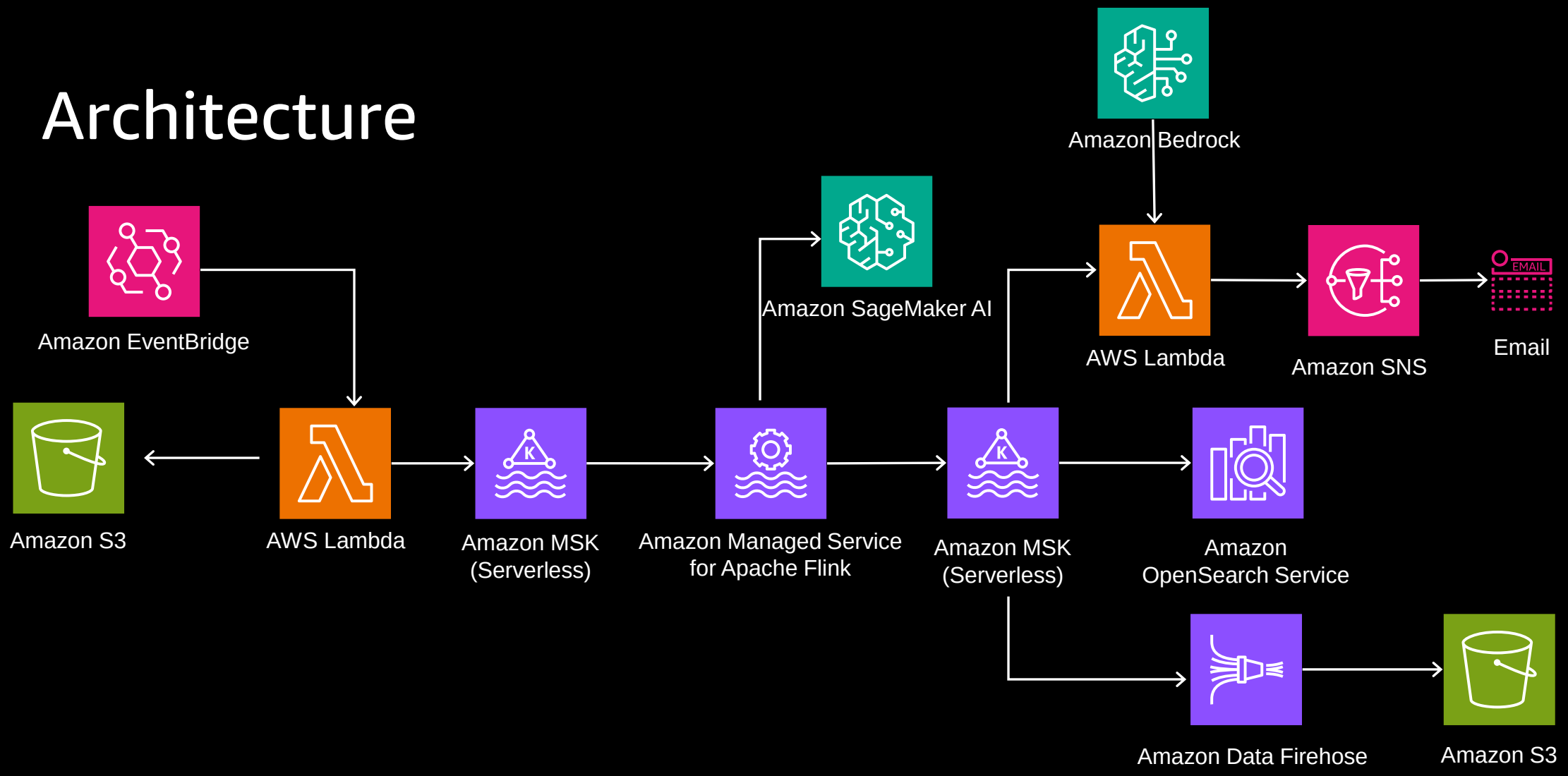
Pre-built Foundation Models: Access pre-trained foundation models from AWS, AI21 Labs, Anthropic, Cohere, Meta, Mistral AI, Stability AI

Custom Model Fine-tuning: Customize foundation models with your own data

Scalable Deployment: Deploy generative AI applications at scale

This workshop specifically uses Anthropic *Claude 3 Sonnet v1*

Architecture



Collaboration

Stuck? Don’t be afraid to ask your table-mates! Work together to understand and complete the workshop.

Don’t hesitate to raise your hand and ask support staff for assistance.

Getting started with Workshop Studio

You have access to an AWS account with everything needed to complete this workshop.

The AWS account is only available for the duration of this session. **You will lose access to the account once the session is complete.**

The workshop environment is deployed to a specific AWS Region. Make sure that you are working in this Region; other Regions are blocked.

Review the terms and conditions of the event. **Do not upload any personal or confidential information to the account.**

Accessing the workshop environment for this session

01 Sign in using your preferred method – **https://join.workshops.aws**

02 Enter the event access code – **[XXXX-XXXXXX-XX]**

03 Review terms and join event

04 Select **Get Started** to begin the workshop

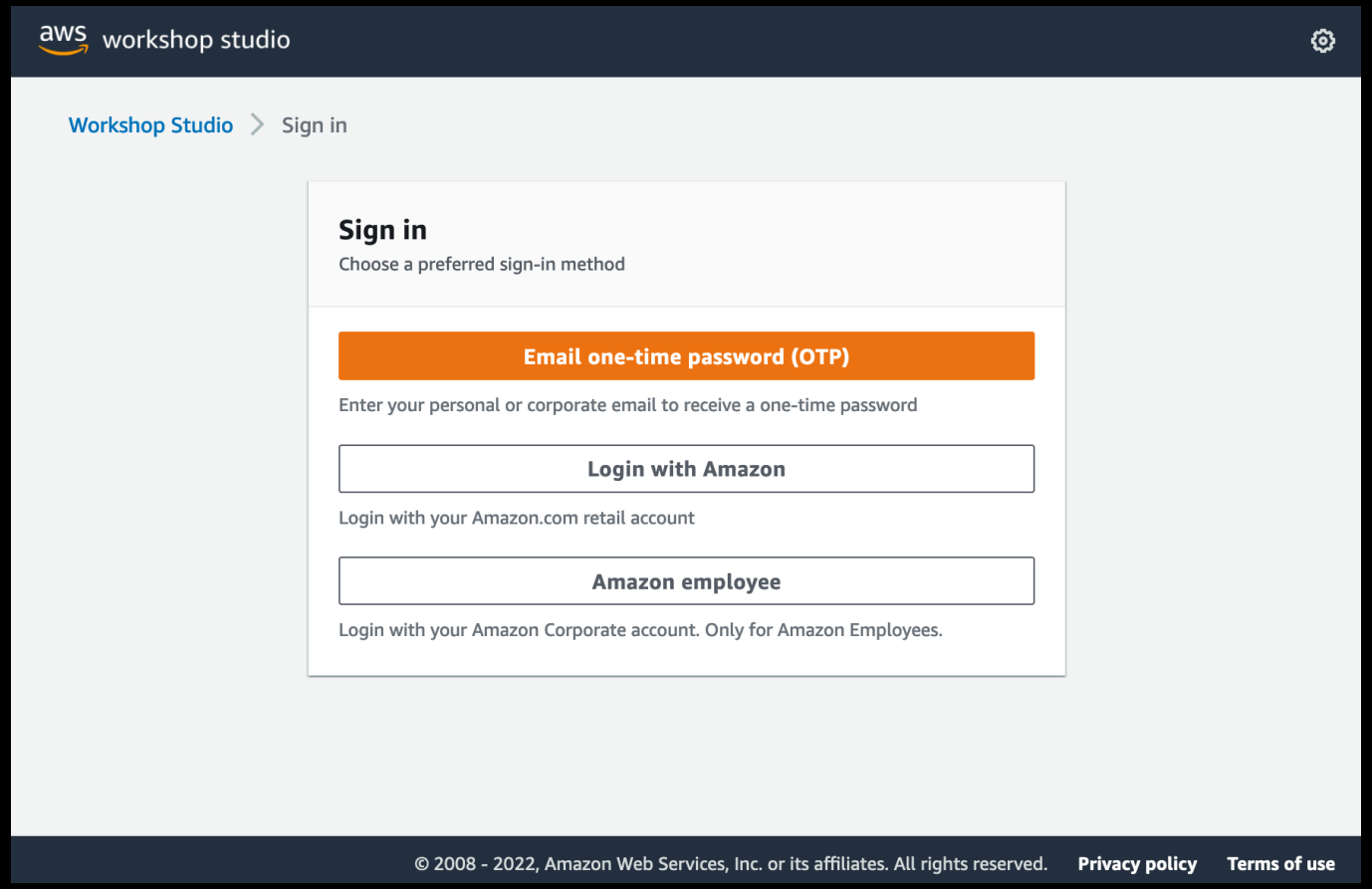
05 Access AWS accounts – access the AWS Management Console or generate AWS CLI credentials as needed

Next up, hands on Workshop!

- We are going to organize into sections called “checkpoints”
- Before and after each checkpoint, we will review objectives, lessons learned and Q&A
- Feel free to complete at your own pace if you don’t want to follow the Checkpoint approach

Hands-On

Workshop Join URL:



Review terms and join event

Workshop Join URL:

aws workshop studio

Workshop Studio > Join event

Step 1
Enter event access code

Step 2
Review and join

Review and join

Event details

Name	Start time	Duration	Level
AWS General Immersion Day	9/23/2022 01:13 AM	12 hours	-

Description
AWS General Immersion Day

Terms and Conditions
Read and accept before joining the event

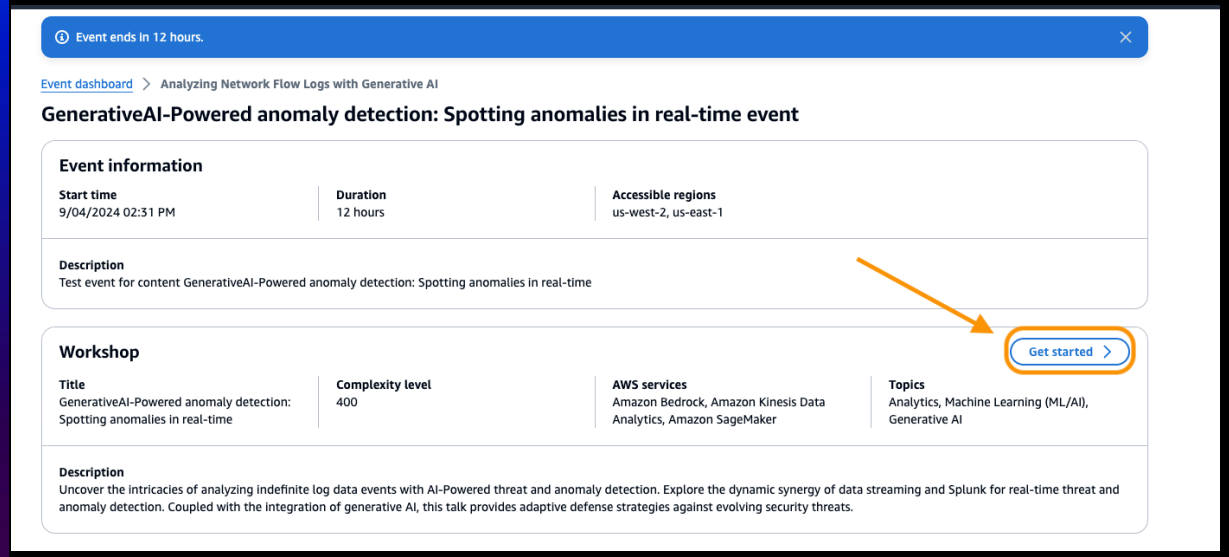
1. By using AWS Workshop Studio for the relevant event, you agree to the AWS Event Terms and Conditions and the AWS Acceptable Use Policy. You acknowledge and agree that are using an AWS-owned account that you can only access for the duration of the relevant event. If you find residual resources or materials in the AWS-owned account, you will make us aware and cease use of the account. AWS reserves the right to terminate the account and delete the contents at any time.
2. You will not: (a) process or run any operation on any data other than test data sets or lab-approved materials by AWS, and (b) copy, import, export or otherwise create derivate works of materials provided by AWS, including but not limited to, data sets.
3. AWS is under no obligation to enable the transmission of your materials through Event Engine and may, in its discretion, edit, block, refuse to post, or remove your materials at any time.
4. Your use of AWS Workshop Studio will comply with these terms and all applicable laws, and your access to AWS Workshop Studio will immediately and automatically terminate if you do not comply with any of these terms or conditions.

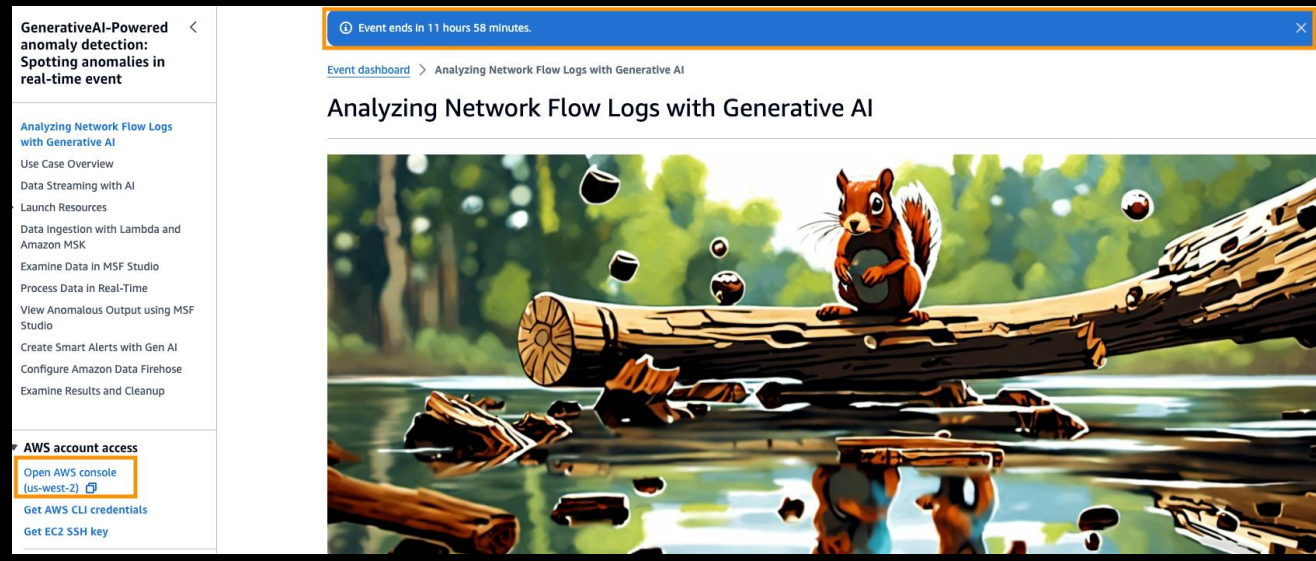
☑ I agree with the Terms and Conditions

Cancel | Previous | Join event

Access AWS account

Access the AWS Console, or generate AWS CLI credentials as needed





Checkpoint 1

Analyzing Network Flow Logs with Generative AI

- Launching Resources
- Use Case Overview
- Data Streaming with AI
- Data Ingestion with Lambda and Amazon MSK
- Examine Data in MSF Studio
- Process Data in Real-Time
- View Anomalous Output using MSF Studio
- Create Smart Alerts with Gen AI
- Configure your MSK Serverless Consumers
- Where to go from here?
- Examine Results and Cleanup

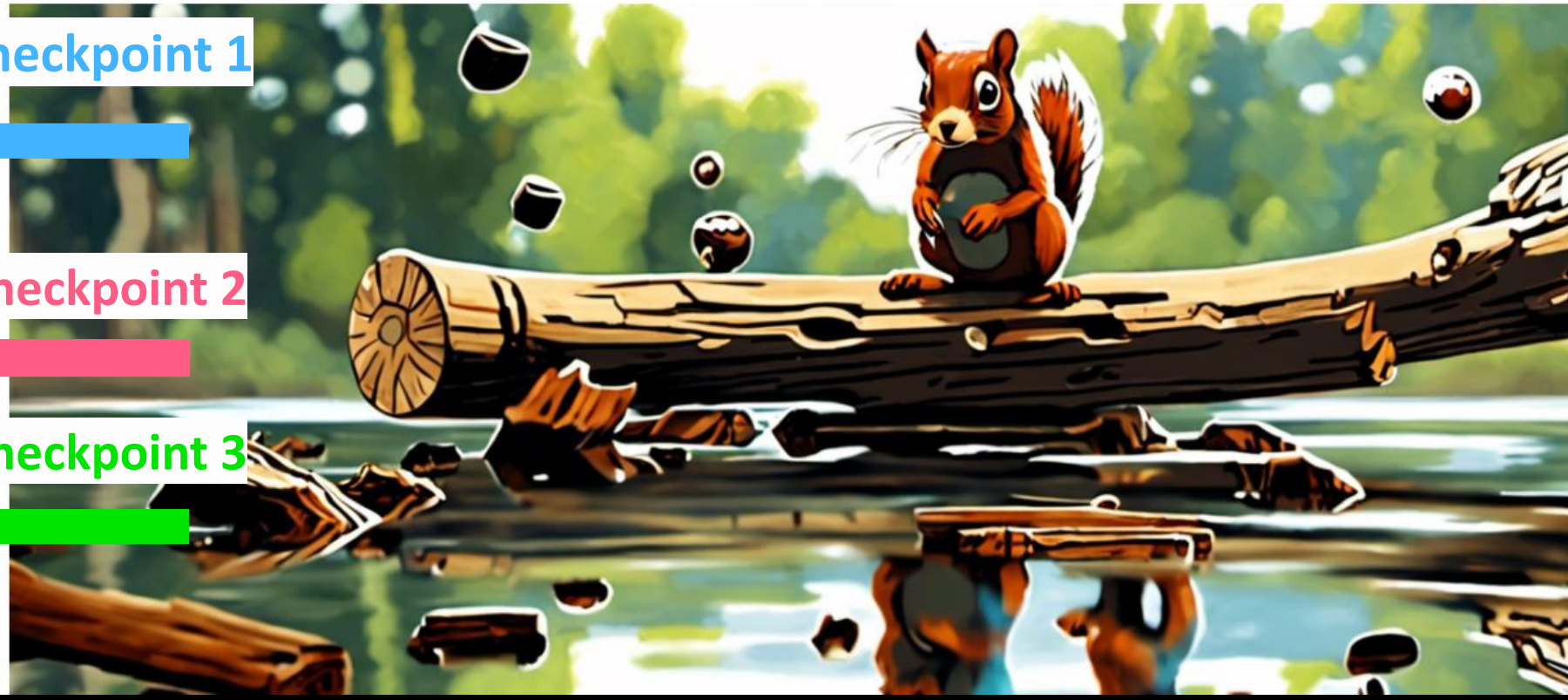
Checkpoint 1

Checkpoint 2

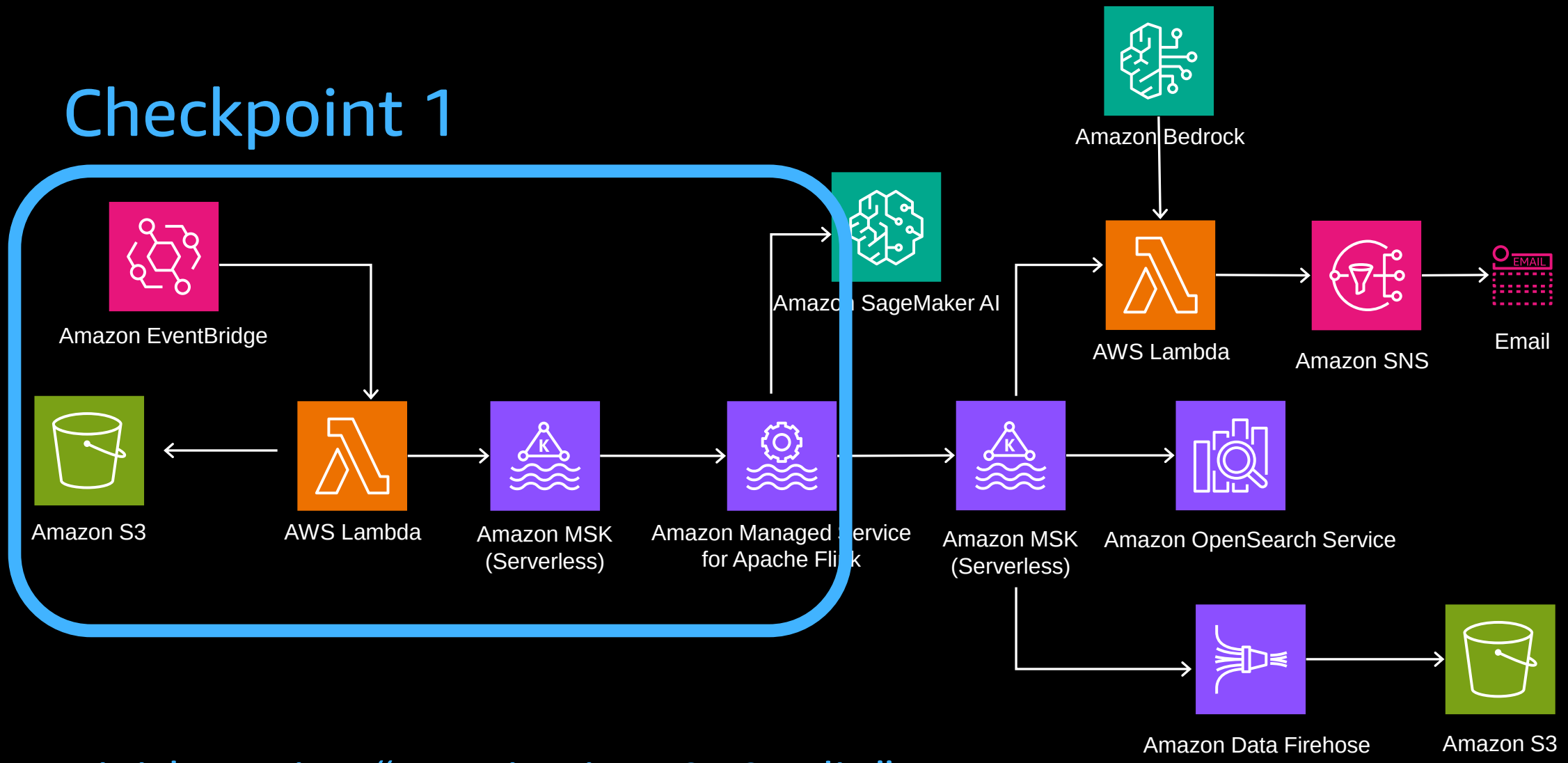
Checkpoint 3

Analyzing Network Flow Logs with Generative AI

Analyzing Network Flow Logs with Generative AI



Checkpoint 1



Finish section “Examine in MSF Studio”

Checkpoint 1 Review

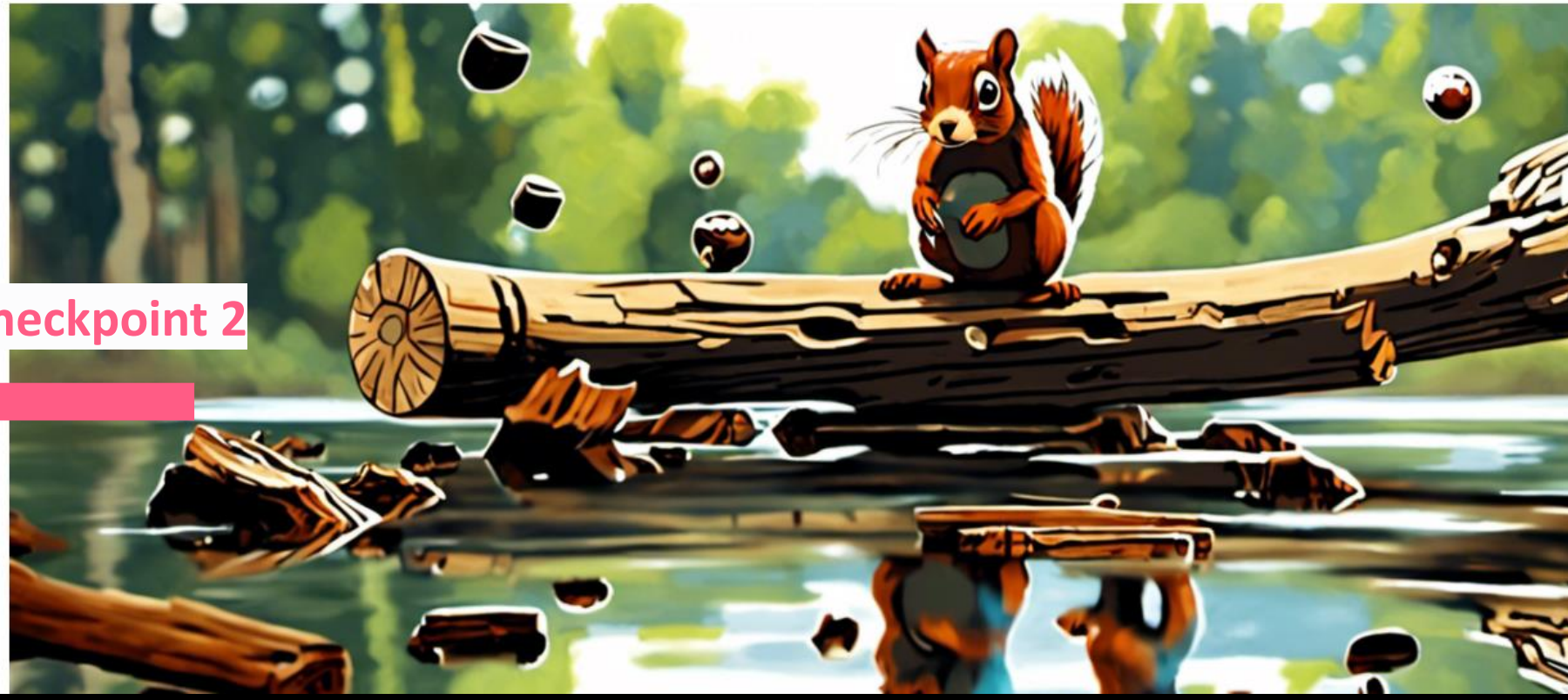
- Producing Simulated Data to MSK Cluster via EventBridge Rule with Lambda
- We can validate / see that data in MSF Studio in real-time

Checkpoint 2

Analyzing Network Flow Logs with Generative AI

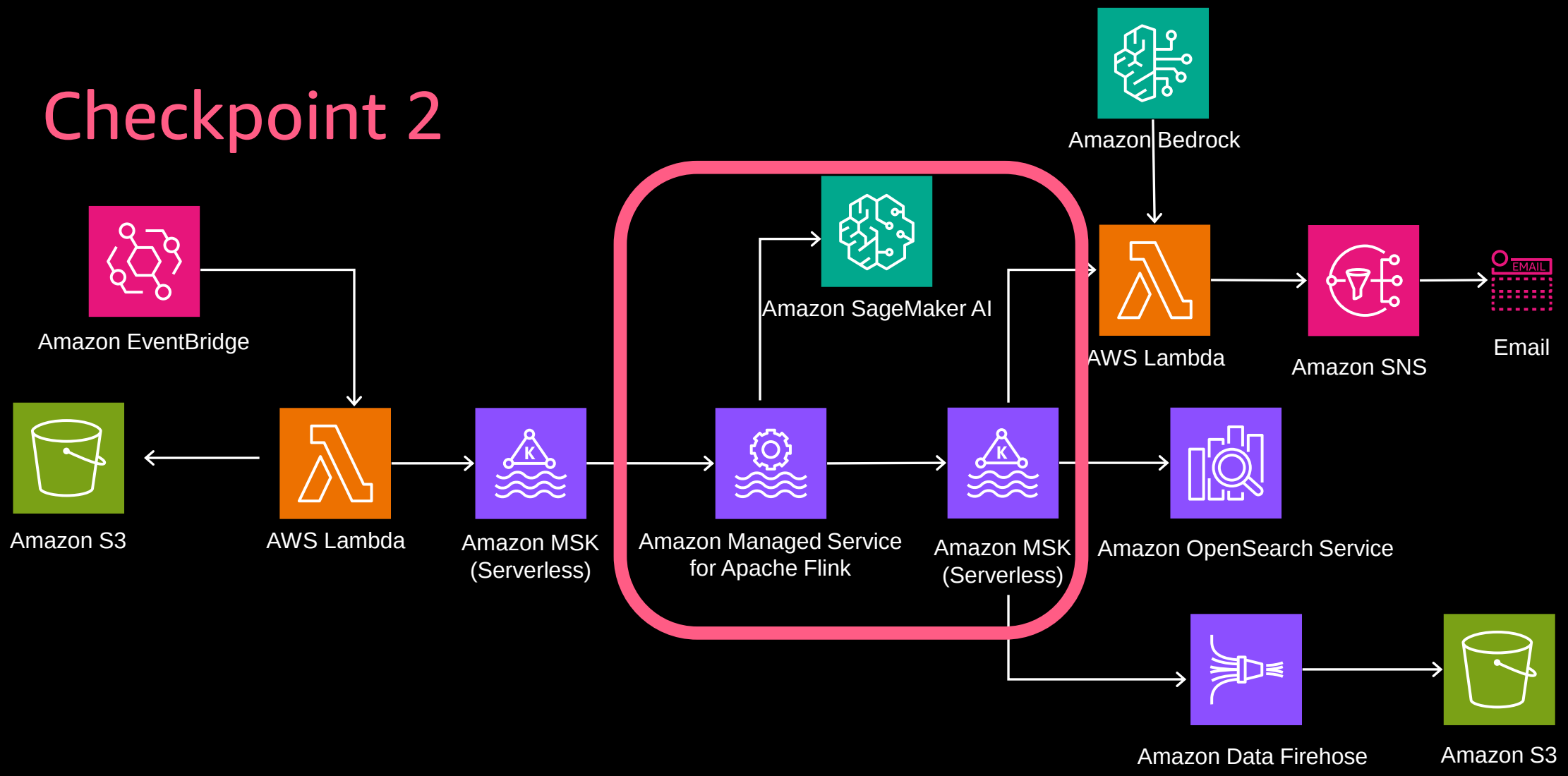
Analyzing Network Flow Logs with Generative AI

Analyzing Network Flow Logs with Generative AI



Checkpoint 2

Checkpoint 2



Finish Section “View Anomalous Output using MSF Studio”

Checkpoint 2 Review

- Validated our model in Sagemaker AI
- Started a durable MSF Application with state that queries SageMaker AI in real-time
- Validated we can view the output of MSF application in MSF Studio
- All the while interacting with MSK Serverless

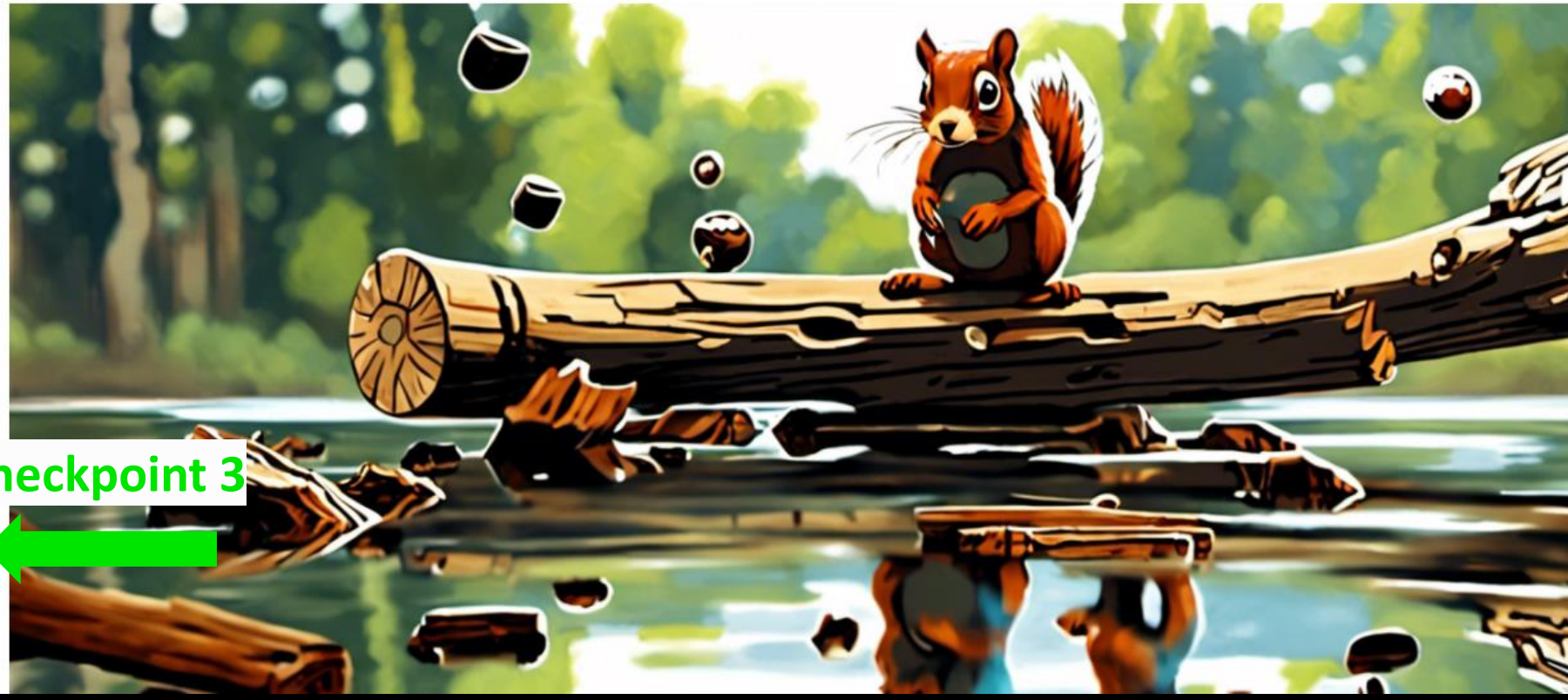
Checkpoint 3

Analyzing Network Flow Logs with Generative AI

- Launching Resources
- Use Case Overview
- Data Streaming with AI
- Data Ingestion with Lambda and Amazon MSK
- Examine Data in MSF Studio
- Process Data in Real-Time
- View Anomalous Output using MSF Studio
- Create Smart Alerts with Gen AI
- Configure your MSK Serverless Consumers
- Where to go from here?
- Examine Results and Cleanup

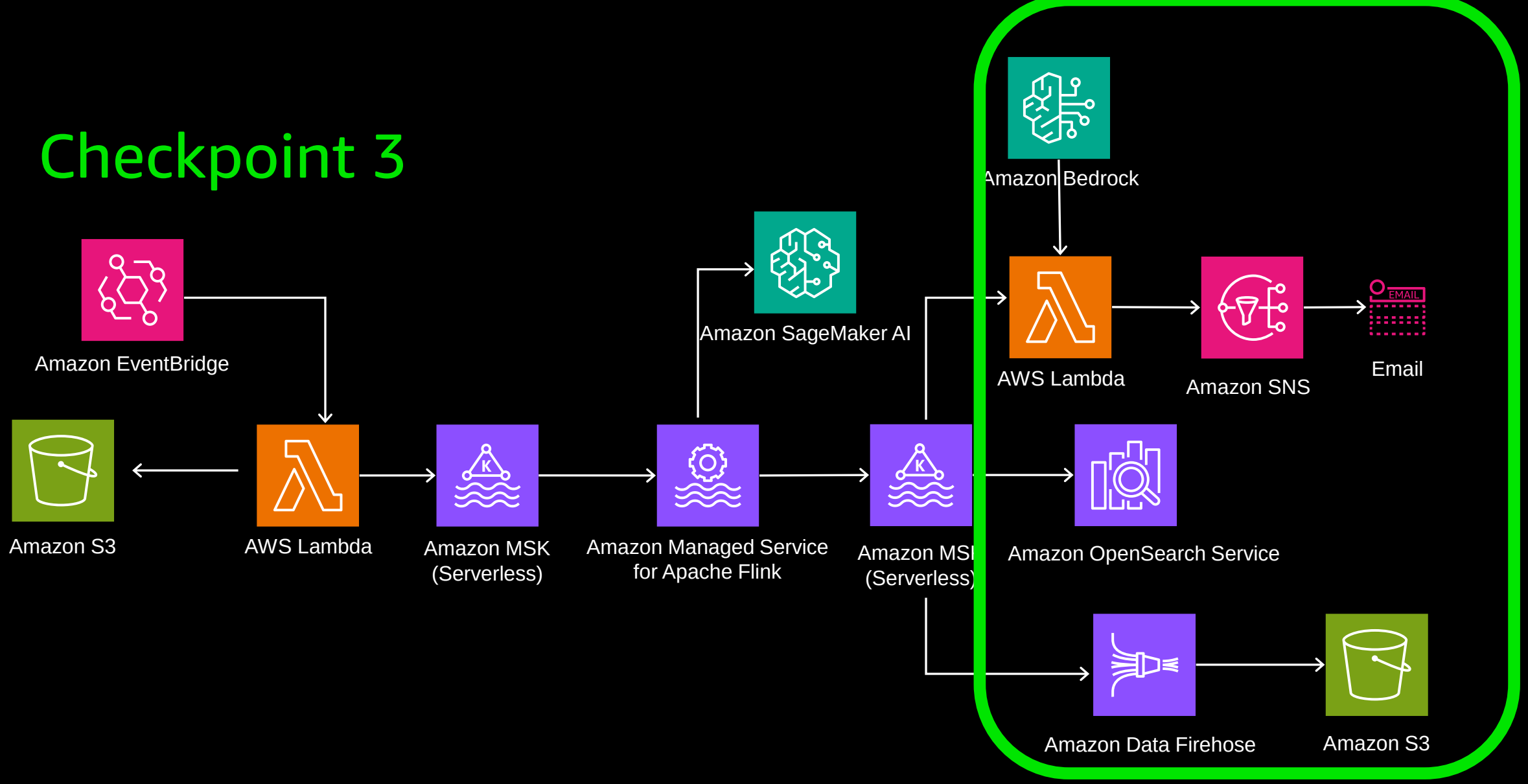
Analyzing Network Flow Logs with Generative AI

Analyzing Network Flow Logs with Generative AI



Checkpoint 3

Checkpoint 3



Finish Section “Configure your MSK Serverless Consumers”

Checkpoint 3 Review

- Explored Amazon Bedrock with anomalous data
- Triggered and received e-mails related to anomalies
- Ingested data into Amazon Opensearch and explored in Dashboard
- Backed data up using Amazon Data Firehose

Extra Credit!

If you've completed Checkpoints 1, 2 and 3, feel free to continue on with the workshop, working through some challenges we've added to dive even deeper

Thank you!

Joe Khazen

Principal, WW GTM Specialist, Data Streaming (AWS)

Austin Groeneveld

Assoc. WW SSA Data Streaming (AWS)

Ali Alemi

Sr. WW SSA Data Streaming (AWS)