AWS re:Invent

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

OPN203-R

Get started with OpenSearch

Arun Lakshmanan (he/him)

OpenSearch Specialist Solutions Architect
AWS

Jon Handler (he/him)

Sr Principal Solutions Architect
AWS

Agenda

- What is OpenSearch?
- Why OpenSearch?
- Search with OpenSearch
- Operational analytics with OpenSearch
- Hands-on lab – Dashboarding and monitoring

What is OpenSearch?

OpenSearch is a fork

Community-driven, open-source search and analytics suite derived from Apache 2.0–licensed Elasticsearch 7.10.2 and Kibana 7.10.2

It consists of

- A search engine – OpenSearch
- A visualization and user interface – OpenSearch Dashboards
- Tools and plugins adding series functionality

Why OpenSearch?

100% open source

Use, modify, monetize

Built in the open

No contributor
license agreement

Community driven

Made with your input

Public road map

Contributions welcome

Full featured

Full text search

Analytics

Fast ingest

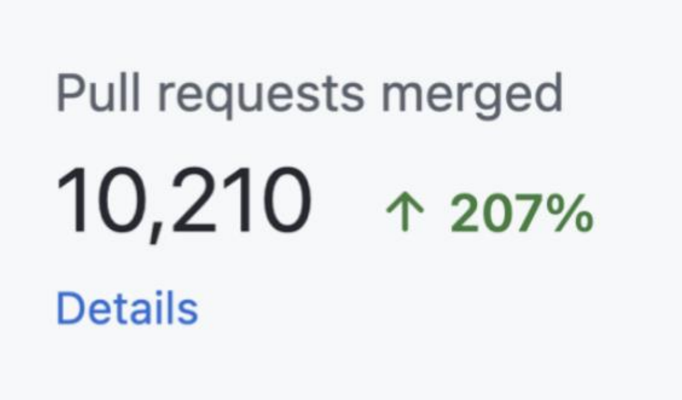
Machine learning

Security

And more . . .

Community driven

- Over 40 partners and growing
- Available with multiple service providers on multiple clouds (Oracle, Aiven-Azure, AWS, Bonsai-GCP)
- Hundreds of contributors
- Thousands of pull requests
- 400+ contributions outside of Amazon in 2022
- 100M+ project downloads
- Top 4 search engine on DB-Engines
- Clients (JS, Java, Rust, Go, .Net, Python, Spring, Hadoop & more)
- Growing number of community projects (for example, K8s operator, Terraform module, and more)



Pull requests merged

10,210 ↑ 207%

Details

OpenSearch: 7 major releases, 90+ features

OpenSearch 1.2

- Observability Interface
- Feature attribution in anomaly detection
- Shard-level indexing back-pressure
- More efficient k-NN dense vectors

OpenSearch 2.0

- Lucene 9.1 performance optimizations
- Document-level alerting
- RPM Package Manager

OpenSearch 2.2

- Logistic regression
- Lucene HNSW implementation
- Custom GeoJSON
- RCFSummarize algorithm
- Support in region maps

OpenSearch 2.4

- Flat data type support
- Security analytics MVP
- Windows support for k-NN plugin
- Spring Data and Hadoop clients

Q1’22 Q2’22 Q3’22 Q4’22

OpenSearch 1.3

- App Analytics, Trace ID correlation, and Live Tail in Observability
- PPL runtime support
- K-means and Random Cut Forest algorithm support

OpenSearch 2.1

- Automate snapshots with Snapshot Management
- Dedicated resources for ML workloads
- **Multi-terms aggregation**

OpenSearch 2.3

- Segment replication
- Remote-backed storage
- Drag-and-drop visualization
- OpenSearch Playground

How does OpenSearch work?

OpenSearch is a database

1

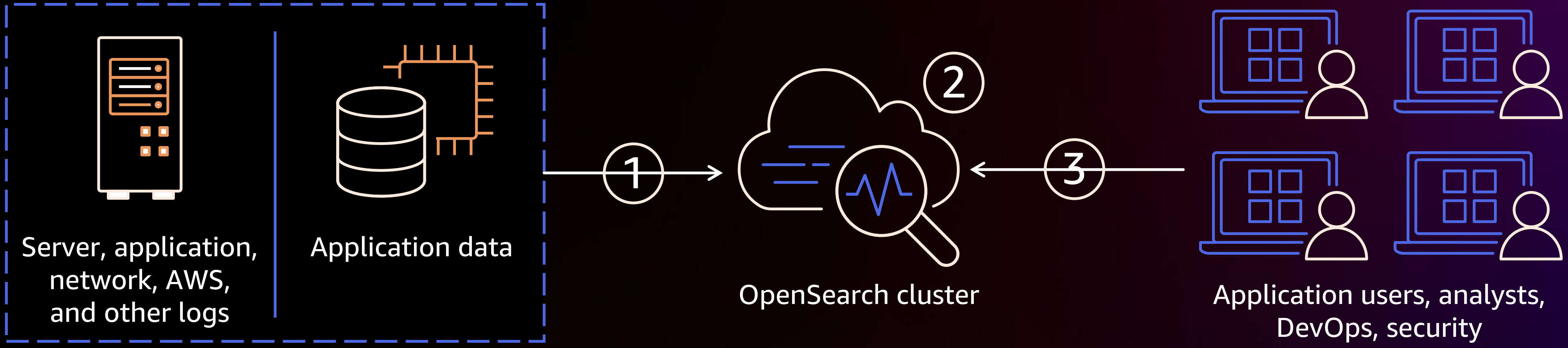
Send data as JSON via REST APIs

2

Data is indexed – all fields searchable, including nested JSON

3

REST APIs, for fielded matching, Boolean expressions, sorting, and analysis


Server, application, network, AWS, and other logs
Application data
1
2
OpenSearch cluster
3
Application users, analysts, DevOps, security

Operational Analytics

IMPROVE INTERNAL BUSINESS EFFICIENCY AND/OR CUSTOMER EXPERIENCE AND VALUE

- Traditional analytics vs operational analytics
  - Operational analytics
    - Descriptive analytics – Observability
    - Predictive analytics
    - Prescriptive analytics

Observability: data drives decisions

Logs

Amazon OpenSearch Service

Metrics

Amazon Managed Service for Prometheus
Amazon Managed Service for Grafana

Traces

AWS Distro for OpenTelemetry
Fluent Bit

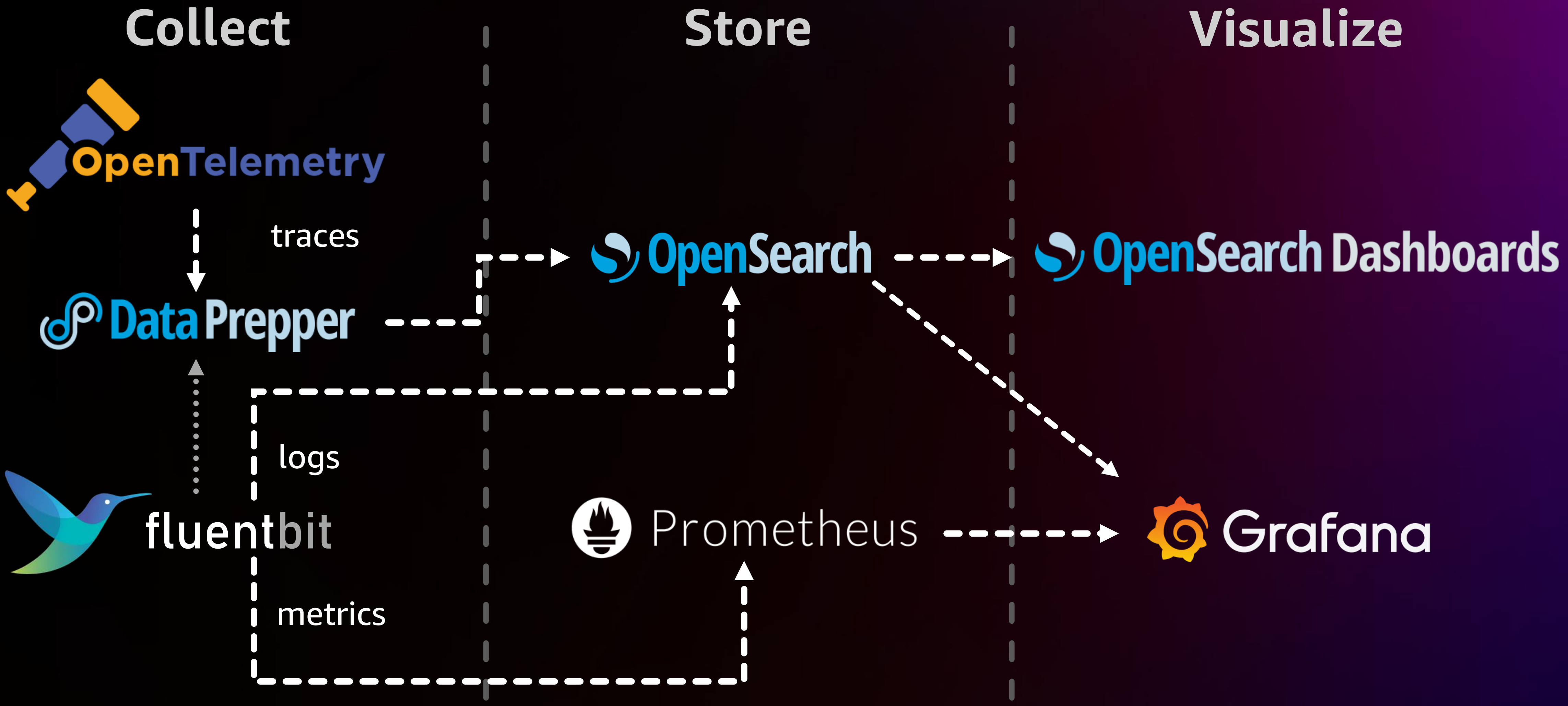
AWS monitoring and observability services help you maintain SLAs by **detecting, investigating, and remediating** problems to achieve

Availability

Reliability

Performance

Open source observability architecture

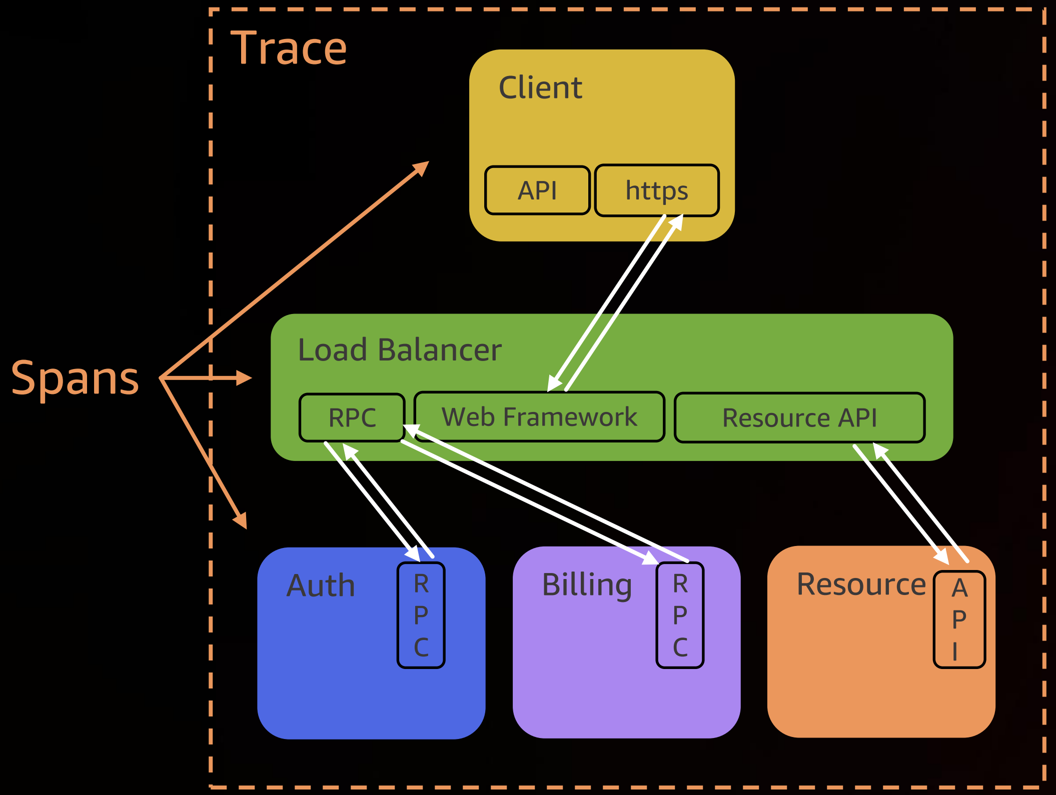


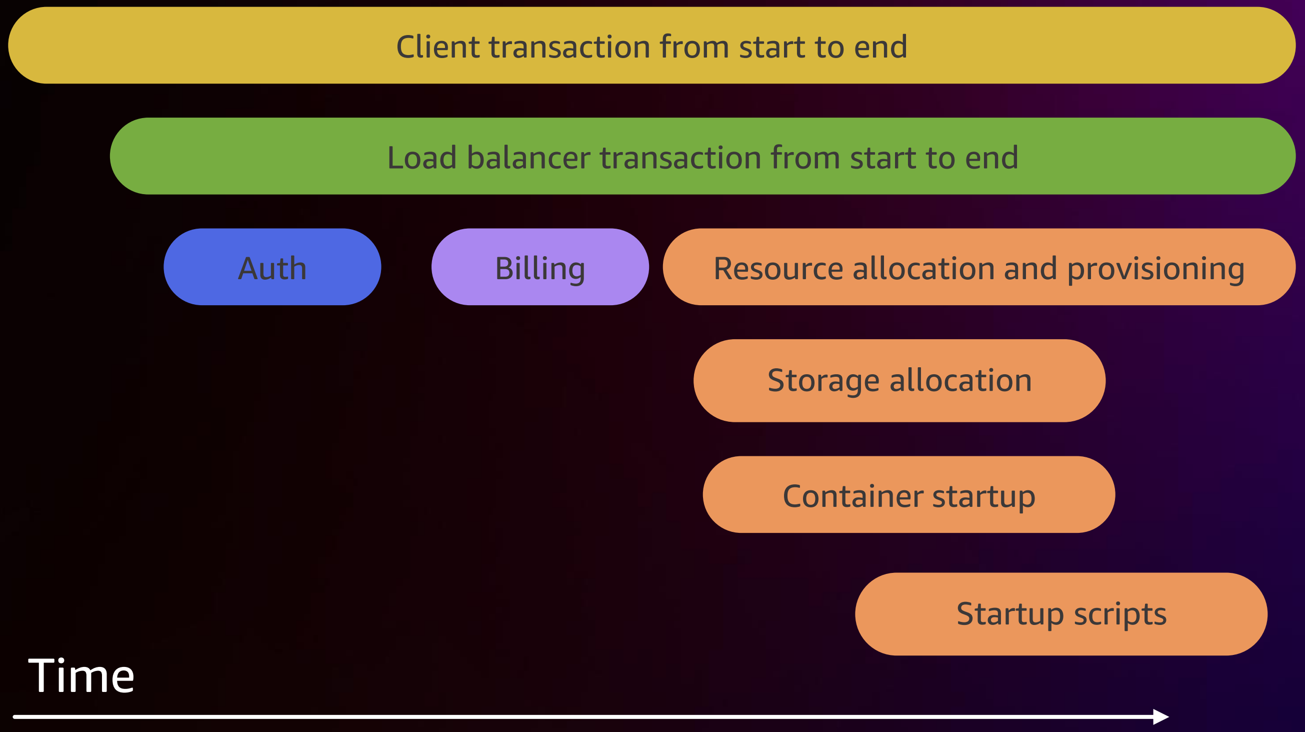
What is distributed tracing?

Identifying problems in cloud applications

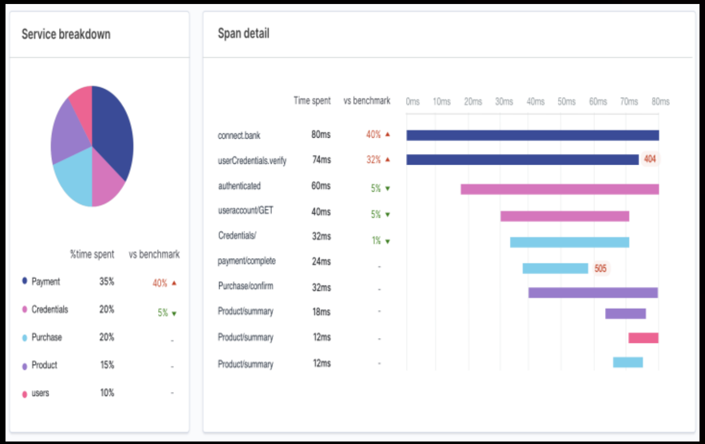
A method of observing requests as they propagate through distributed systems

Trace: Hierarchical, end-to-end record of processing a request




Client transaction from start to end
Load balancer transaction from start to end
Auth
Billing
Resource allocation and provisioning
Storage allocation
Container startup
Startup scripts
Time

What is trace analytics?



Service breakdown

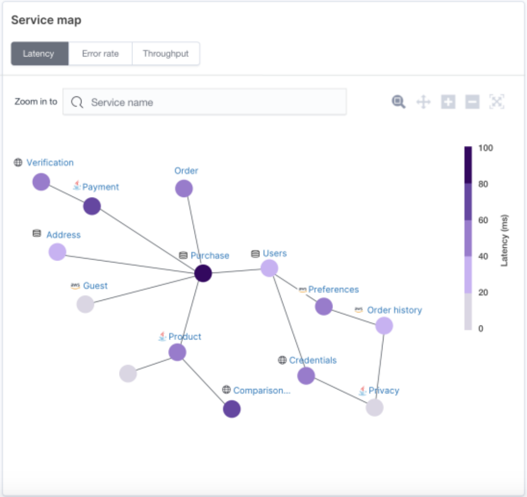
	%time spent	vs benchmark
● Payment	35%	40% ▲
● Credentials	20%	5% ▼
● Purchase	20%	-
● Product	15%	-
● users	10%	-

Span detail

	Time spent	vs benchmark
connect.bank	80ms	40% ▲
userCredentials.verify	74ms	32% ▲
authenticated	60ms	5% ▼
useraccount/GET	40ms	5% ▼
Credentials/	32ms	1% ▼
payment/complete	24ms	-
Purchase/confirm	32ms	-
Product/summary	18ms	-
Product/summary	12ms	-
Product/summary	12ms	-

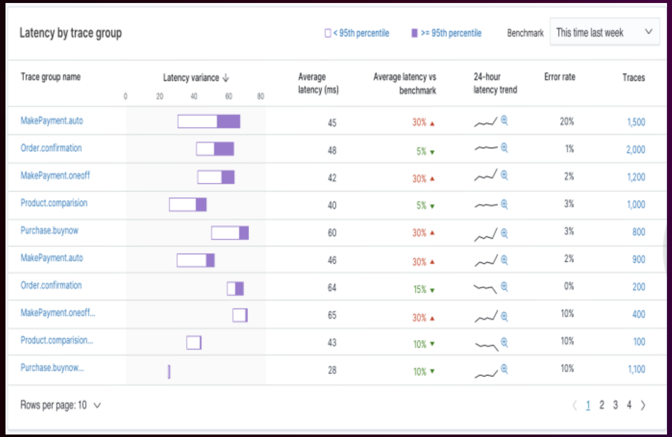
Trace-span details

- Single request performance
- Diagnose root cause



Service maps

- End-to-end view
- Isolate issues to services



Latency by trace group

□ < 95th percentile ■ >= 95th percentile Benchmark This time last week

Trace group name	Latency variance ↓	Average latency (ms)	Average latency vs benchmark	24-hour latency trend	Error rate	Traces
MakePayment.auto		45	30% ▲		20%	1,500
Order.confirmation		48	5% ▼		1%	2,000
MakePayment.oneoff		42	30% ▲		2%	1,200
Product.comparision		40	5% ▼		3%	1,000
Purchase.buynow		60	30% ▲		3%	800
MakePayment.auto		46	30% ▲		2%	900
Order.confirmation		64	15% ▼		0%	200
MakePayment.oneoff...		65	30% ▲		10%	400
Product.comparision...		43	10% ▼		10%	100
Purchase.buynow...		28	10% ▼		10%	1,100

Rows per page: 10

‹ 1 2 3 4 ›

Trace groups

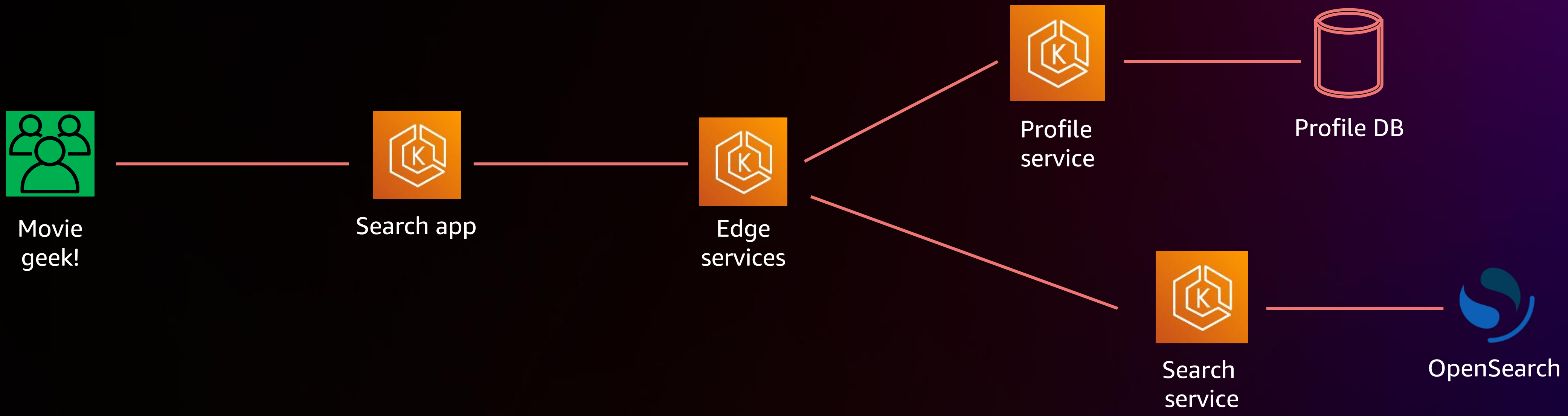
- Monitor performance
- Identify issues early

What are you building?

Lab 1: Search application

Prebuilt set of microservices coordinate searching in OpenSearch

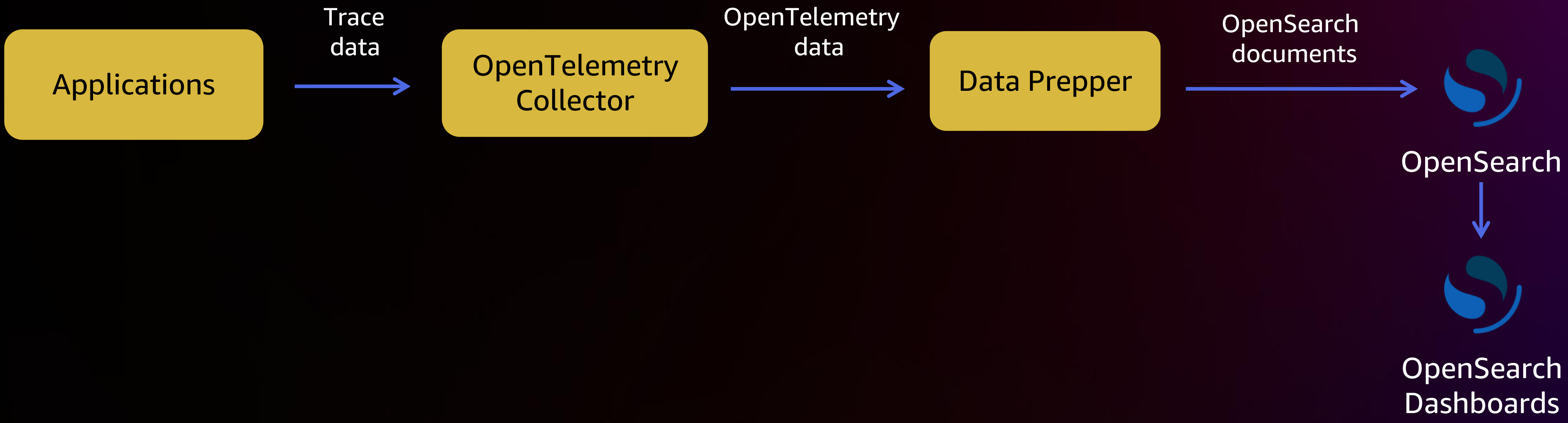
- Provision application data (IMDb Top 5000 titles)
- Publish search templates
- Review application monitoring with prebuilt dashboards



Lab 2: Distributed tracing

Review prebuilt trace data ingestion pipeline

- Simulate errors in services
- Debug and fix them



Applications

Trace data

OpenTelemetry Collector

OpenTelemetry data

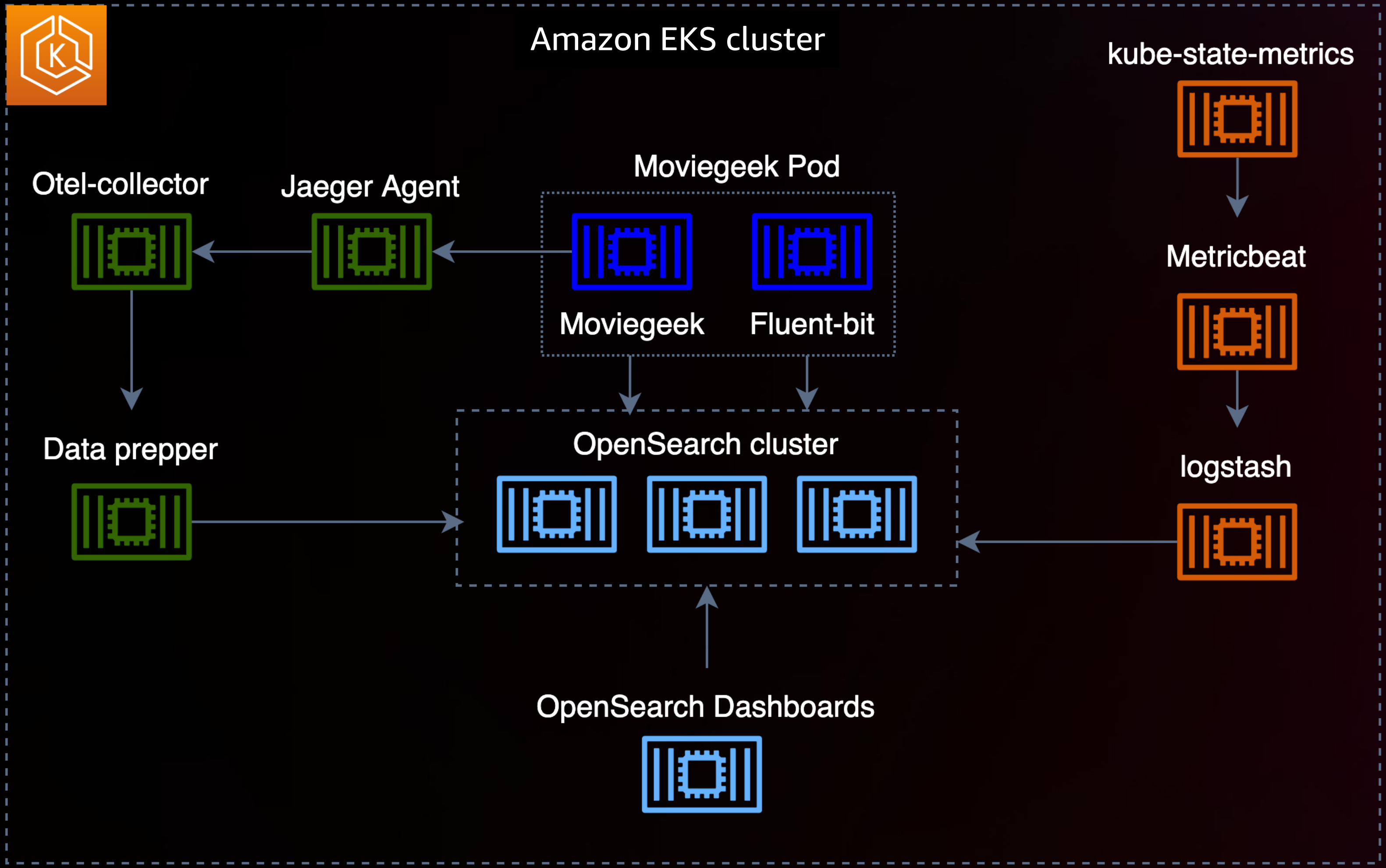
Data Prepper

OpenSearch documents

OpenSearch

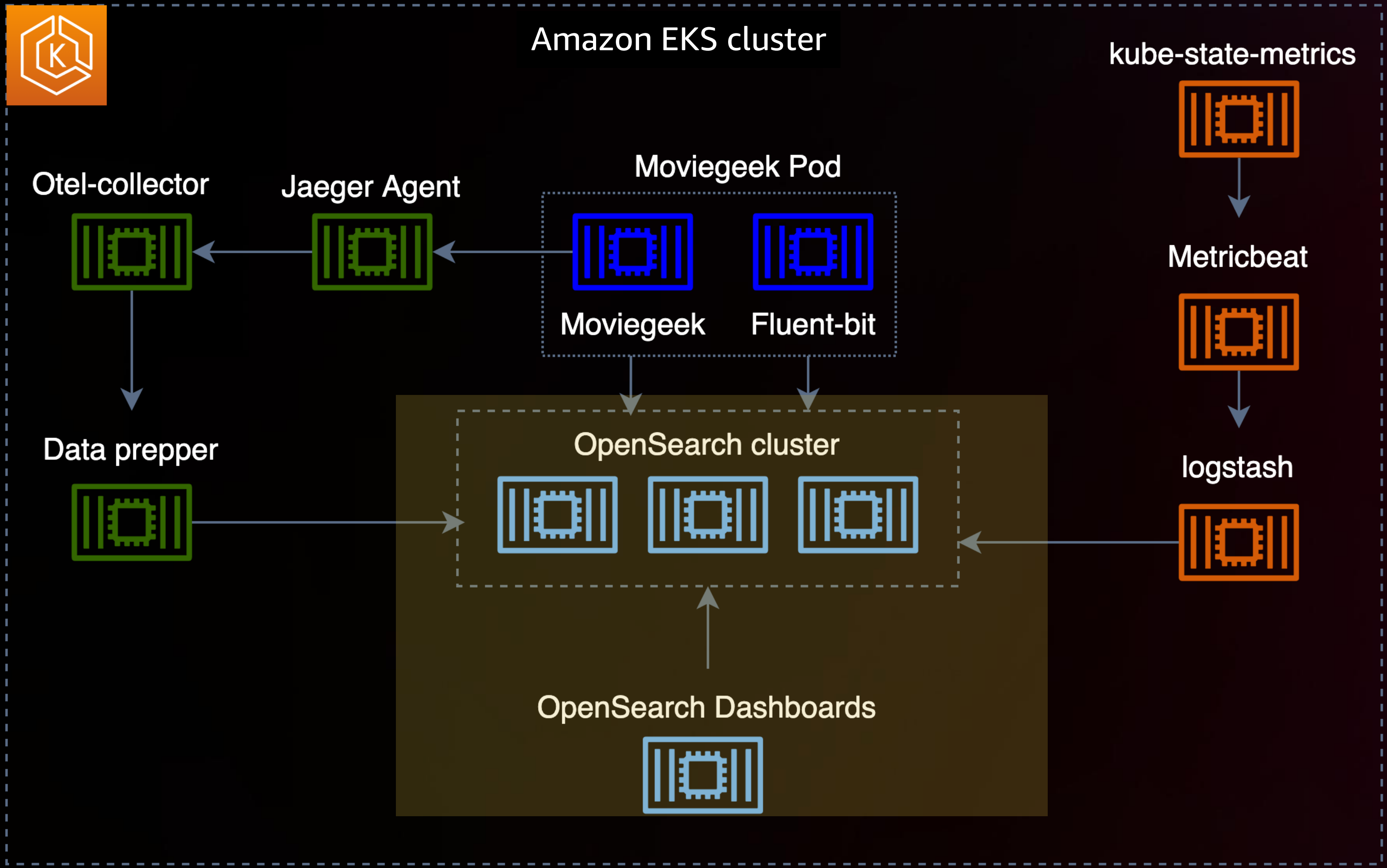
OpenSearch Dashboards

Lab 3: Monitoring Kubernetes cluster



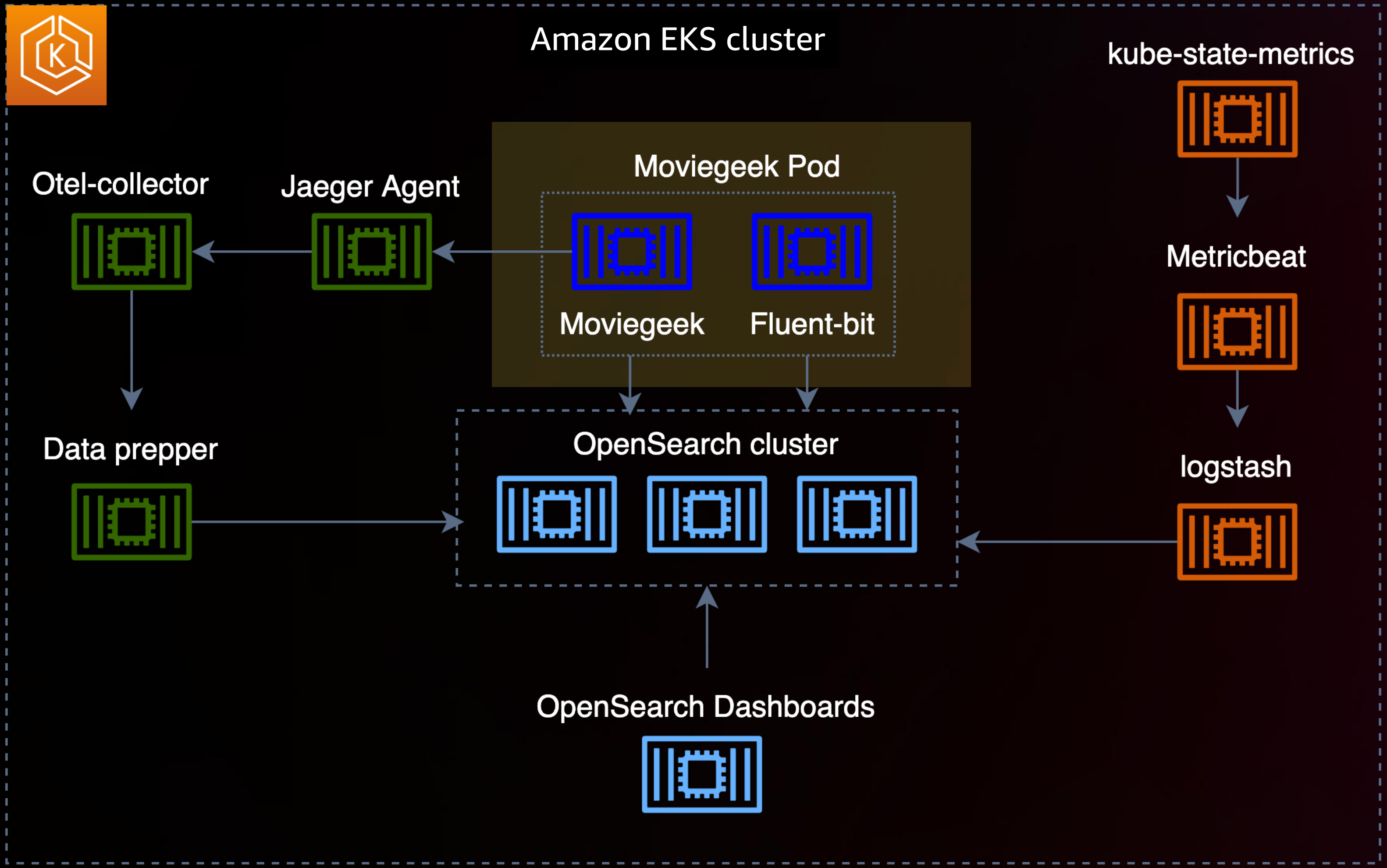
- 3 m5.large worker nodes

Lab 3: Monitoring Kubernetes cluster



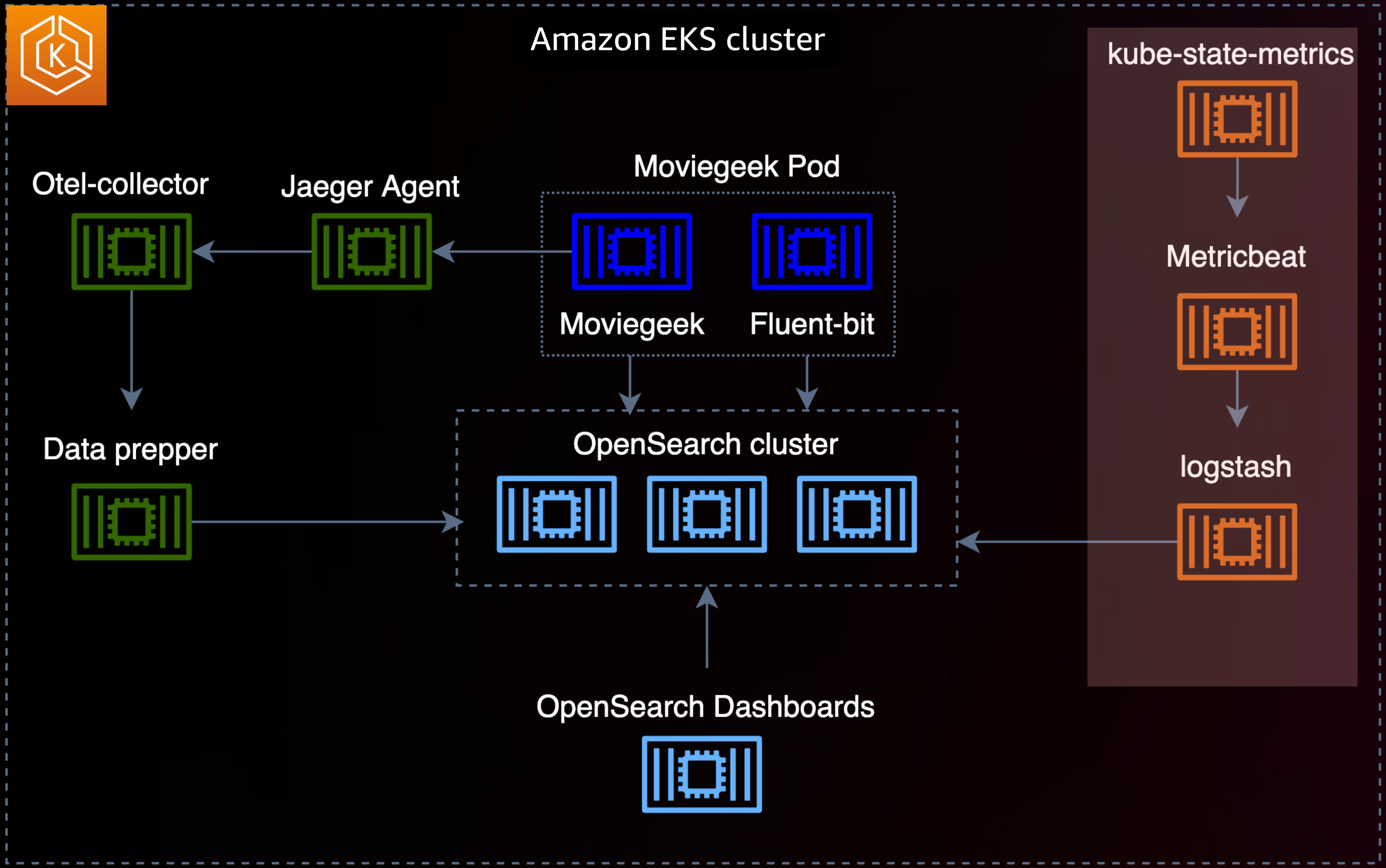
- 3 m5.large worker nodes
- OpenSearch Cluster and Dashboards
  - 4 Pods

Lab 3: Monitoring Kubernetes cluster



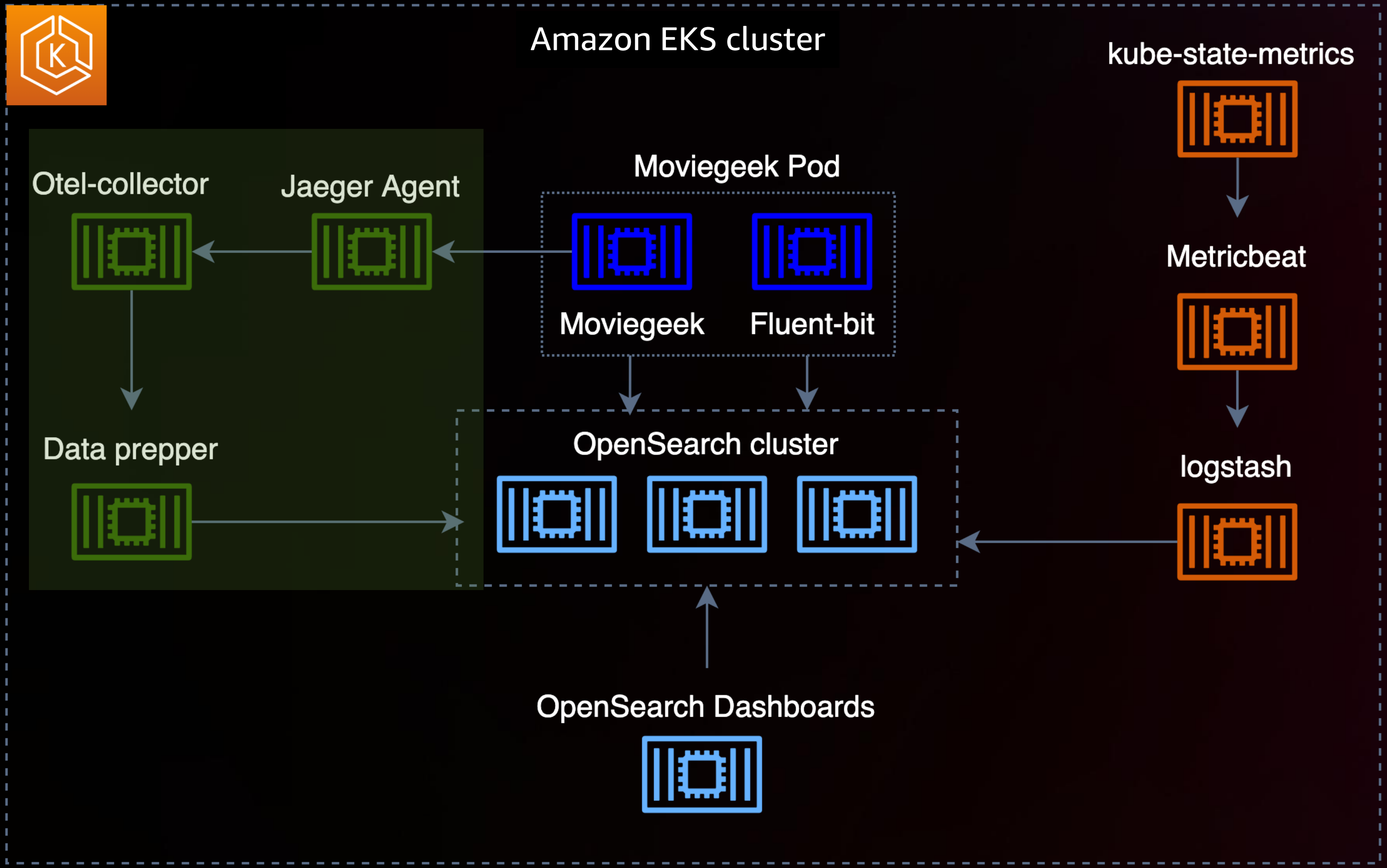
- 3 m5.large worker nodes
- OpenSearch Cluster and Dashboards
  - 4 Pods
- Application with Fluent-bit sidecar

Lab 3: Monitoring Kubernetes cluster



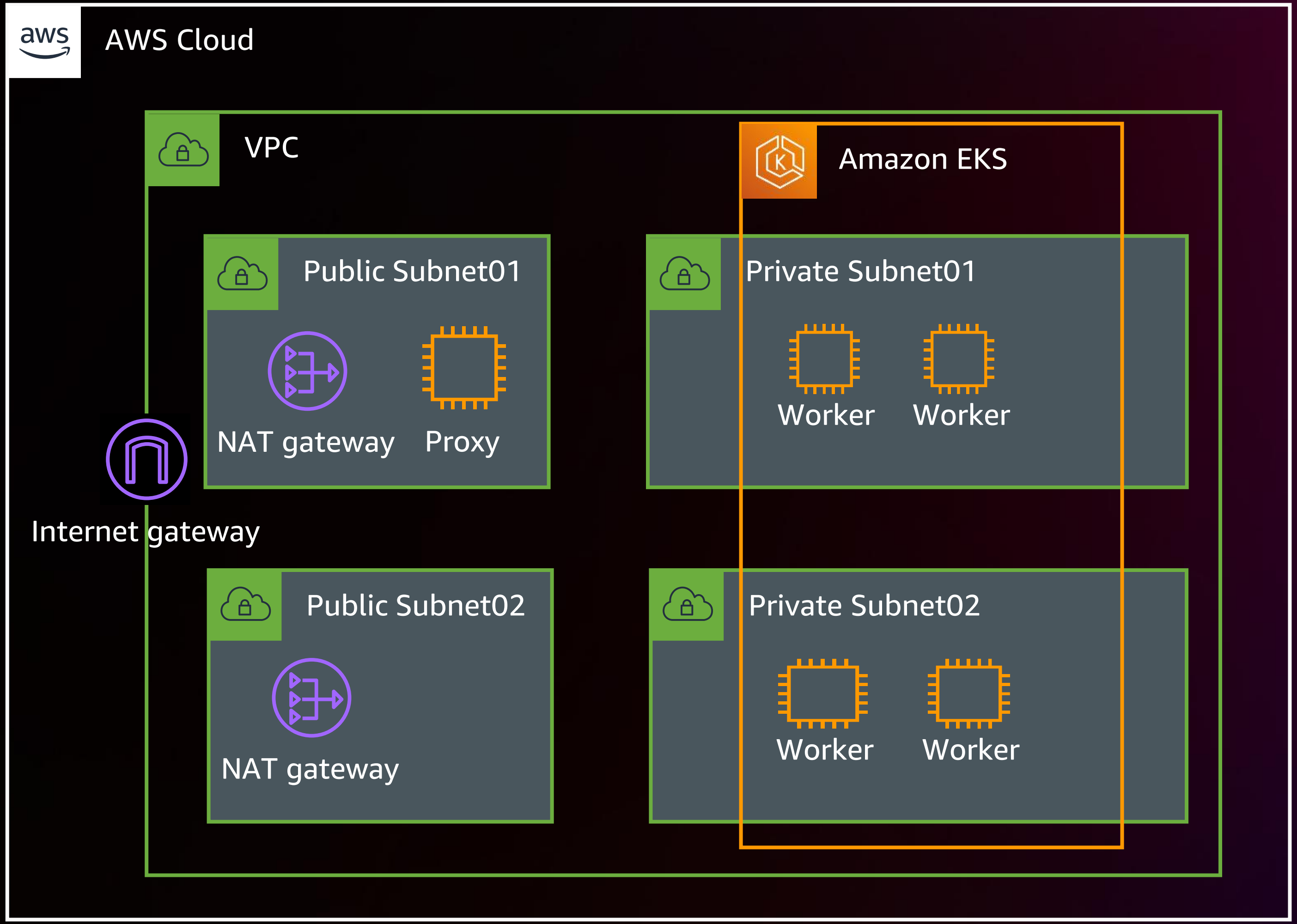
- 3 m5.large worker nodes
- OpenSearch Cluster and Dashboards
  - 4 Pods
- Application with Fluent-bit sidecar
- Cluster monitoring – 6 pods
  - Logstash
  - Metricbeat – daemonset
  - Kube-state-metrics

Lab 3: Monitoring Kubernetes cluster



- 3 m5.large worker nodes
- OpenSearch cluster and dashboards
  - 4 Pods
- Application with Fluent-bit sidecar
- Cluster monitoring – 6 pods
  - Logstash
  - Metricbeat – daemonset
  - Kube-state-metrics
- Trace analytics – 3 pods
  - Jaeger Agent
  - Otel-collector
  - Data prepper

Lab architecture

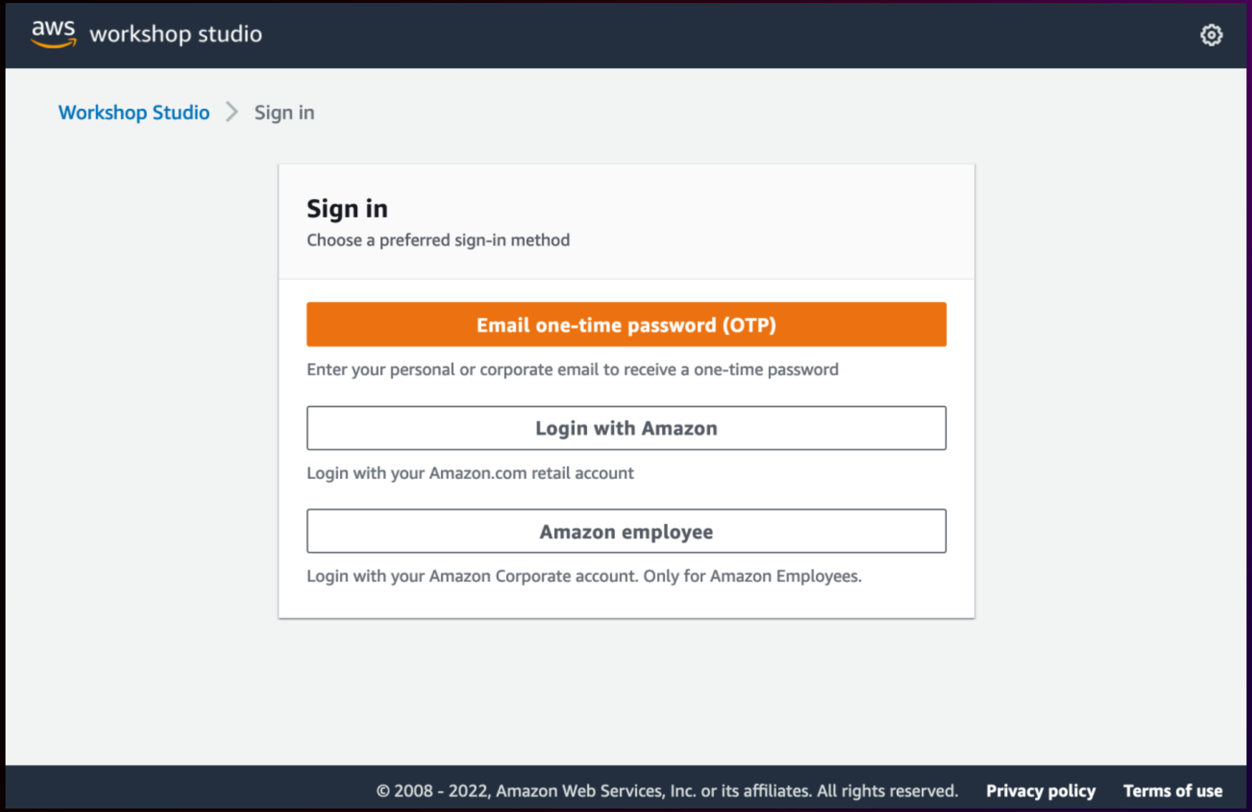


Get started with this workshop

- 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 pre-provisioned infrastructure are deployed to the **us-east-1** Region
- Be sure to review the terms and conditions of the event; do not upload any personal or confidential information in the account

Step 1: Sign-in via OTP

https://catalog.workshops.aws/join



Workshop Studio > Sign in

Sign in

Choose a preferred sign-in method

Enter your personal or corporate email to receive a one-time password

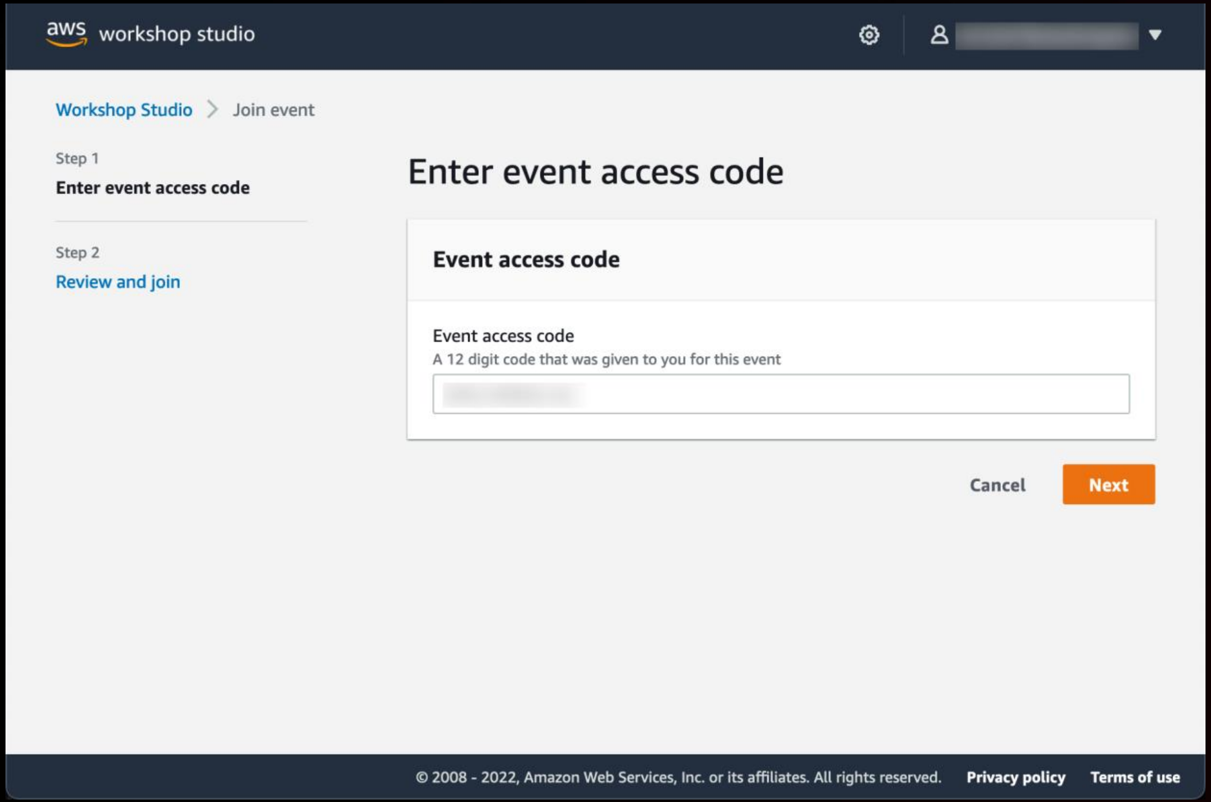
Login with your Amazon.com retail account

Login with your Amazon Corporate account. Only for Amazon Employees.

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

Step 2: Enter event access code

Enter 12-digit event access code



aws workshop studio [redacted]

Workshop Studio > Join event

Step 1
Enter event access code

Step 2
Review and join

Enter event access code

Event access code

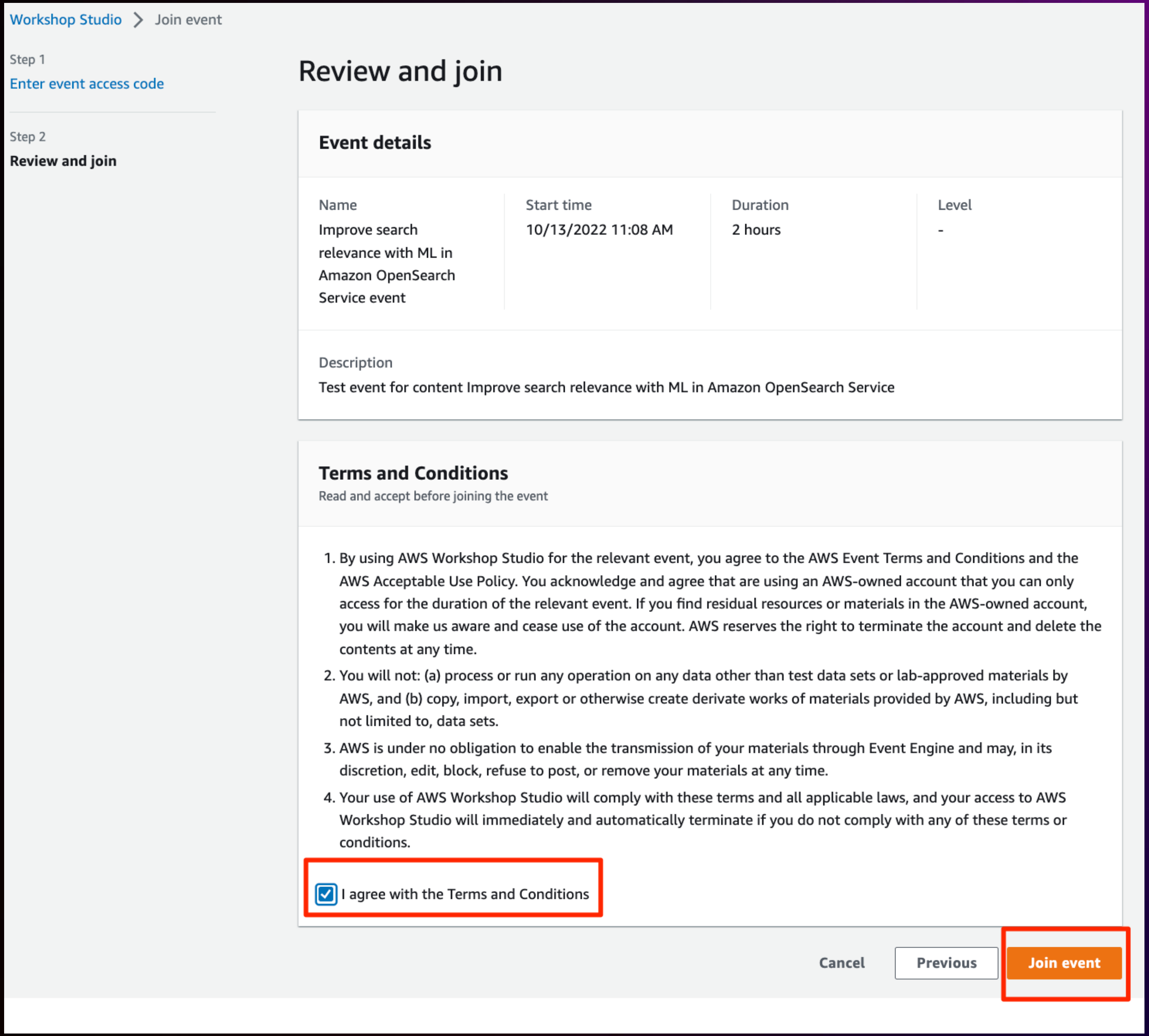
Event access code
A 12 digit code that was given to you for this event
[redacted]

Cancel | Next

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

767a-037d58-84

Step 3: Review terms and join event



Review and join

Event details

Name	Start time	Duration	Level
Improve search relevance with ML in Amazon OpenSearch Service event	10/13/2022 11:08 AM	2 hours	-

Description

Test event for content Improve search relevance with ML in Amazon OpenSearch Service

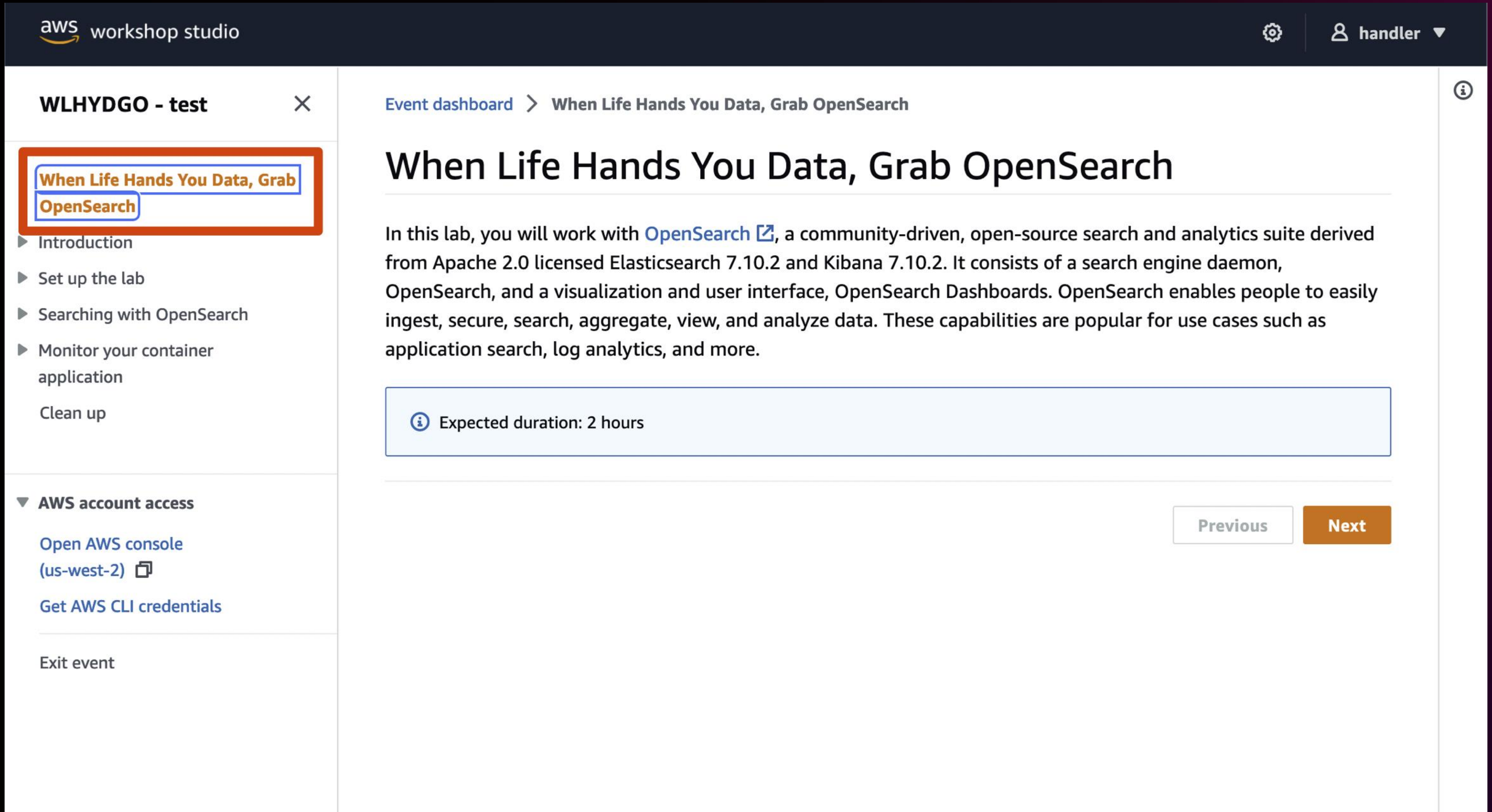
Terms and Conditions

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

Step 4: Get started with the workshop



aws workshop studio
handler ▼
WLHYDGO - test
When Life Hands You Data, Grab OpenSearch
▶ Introduction
▶ Set up the lab
▶ Searching with OpenSearch
▶ Monitor your container application
Clean up
▼ AWS account access
Open AWS console (us-west-2)
Get AWS CLI credentials
Exit event
Event dashboard > When Life Hands You Data, Grab OpenSearch

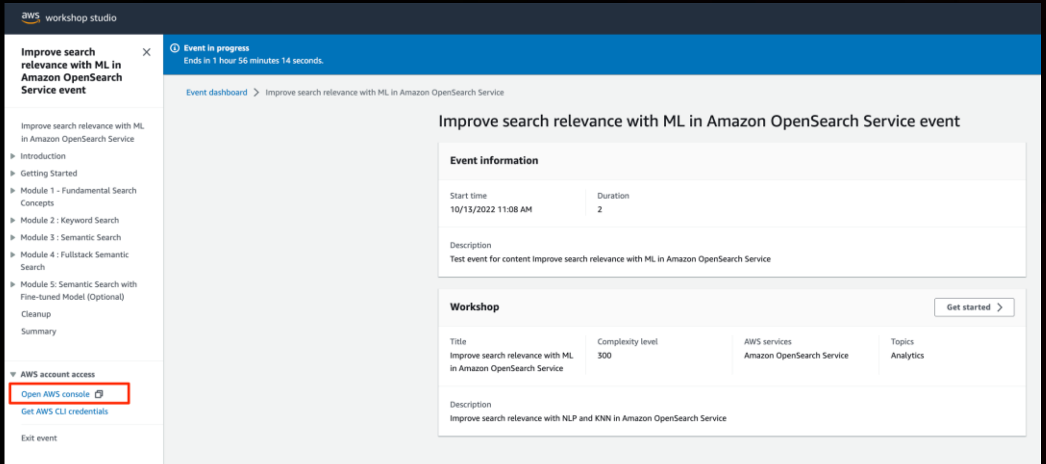
When Life Hands You Data, Grab OpenSearch

In this lab, you will work with OpenSearch, a community-driven, open-source search and analytics suite derived from Apache 2.0 licensed Elasticsearch 7.10.2 and Kibana 7.10.2. It consists of a search engine daemon, OpenSearch, and a visualization and user interface, OpenSearch Dashboards. OpenSearch enables people to easily ingest, secure, search, aggregate, view, and analyze data. These capabilities are popular for use cases such as application search, log analytics, and more.

ⓘ Expected duration: 2 hours

Access AWS account

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



Improve search relevance with ML in Amazon OpenSearch Service event

Improve search relevance with ML in Amazon OpenSearch Service
- Introduction
- Getting Started
- Module 1 - Fundamental Search Concepts
- Module 2 : Keyword Search
- Module 3 : Semantic Search
- Module 4 : Fullstack Semantic Search
- Module 5: Semantic Search with Fine-tuned Model (Optional)
- Cleanup
- Summary

AWS account access
- Open AWS console
- Get AWS CLI credentials

Exit event

Event in progress
Ends in 1 hour 56 minutes 14 seconds.

Event dashboard > Improve search relevance with ML in Amazon OpenSearch Service

Improve search relevance with ML in Amazon OpenSearch Service event

Event information

Start time
10/13/2022 11:08 AM

Duration
2

Description
Test event for content Improve search relevance with ML in Amazon OpenSearch Service

Workshop

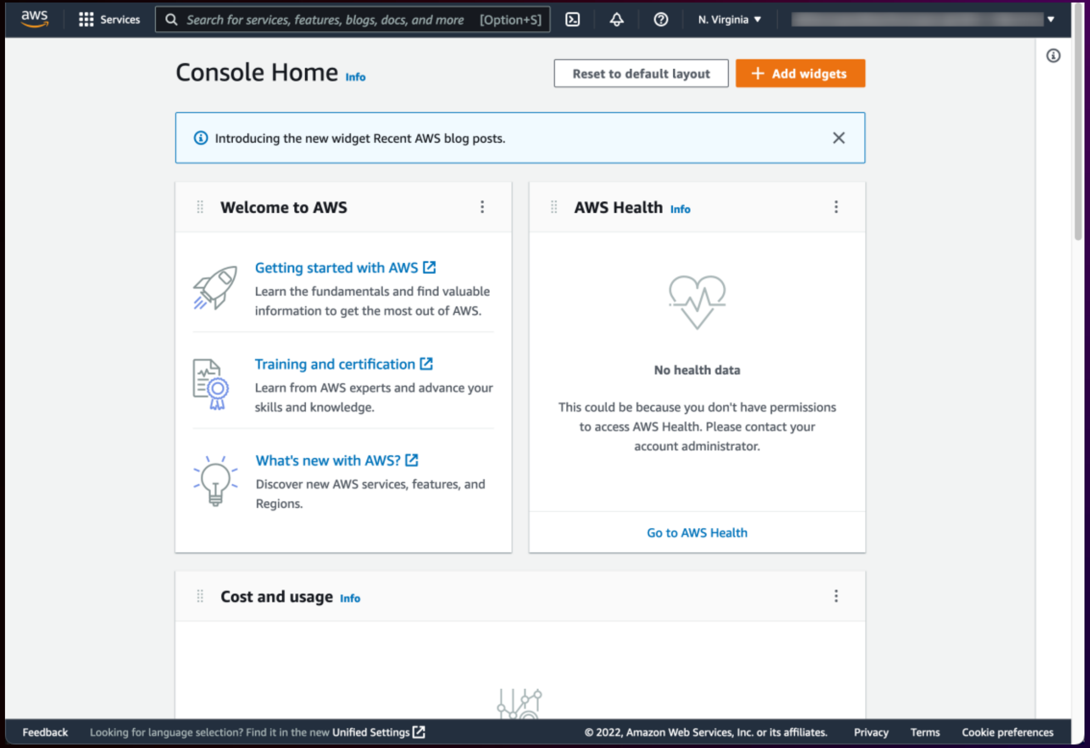
Title
Improve search relevance with ML in Amazon OpenSearch Service

Complexity level
300

AWS services
Amazon OpenSearch Service

Topics
Analytics

Description
Improve search relevance with NLP and KNN in Amazon OpenSearch Service


aws
Services
Search for services, features, blogs, docs, and more [Option+S]
N. Virginia ▾

[redacted]

Console Home Info

Reset to default layout
+ Add widgets

Introducing the new widget Recent AWS blog posts.

Welcome to AWS

Getting started with AWS
Learn the fundamentals and find valuable information to get the most out of AWS.

Training and certification
Learn from AWS experts and advance your skills and knowledge.

What's new with AWS?
Discover new AWS services, features, and Regions.

AWS Health Info

No health data

This could be because you don't have permissions to access AWS Health. Please contact your account administrator.

Go to AWS Health

Cost and usage Info

Feedback
Looking for language selection? Find it in the new Unified Settings
© 2022, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

Get started

Event access for today

https://tinyurl.com/26f4j4cd

Event access code

767a-037d58-84

Workshop permanent link

https://tinyurl.com/2mwmx9n2

Thank you!

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