

AWS RE:INVENT

re:Cap

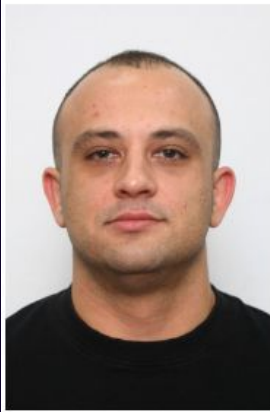
About me

- In IT-u since 2010.
- **Working with AWS since 2011.**
- L1 Support -> System Administrator / DBA -> AWS Consultant -> DevOps Lead -> Production Engineer -> Production Engineer Lead
- Multiple AWS / Database / Security certificates
- **Certified AWS Instructor**
- [AWS Community Bosnia leader](#)
- [AWS Community Hero](#)

Developer / Community / Heroes

Dzenan Dzevlan





Dzenan Dzevlan, Cloud and DevOps Expert at TN-TECH

📍 Mostar, Bosnia and Herzegovina

✓ Hero since 2020

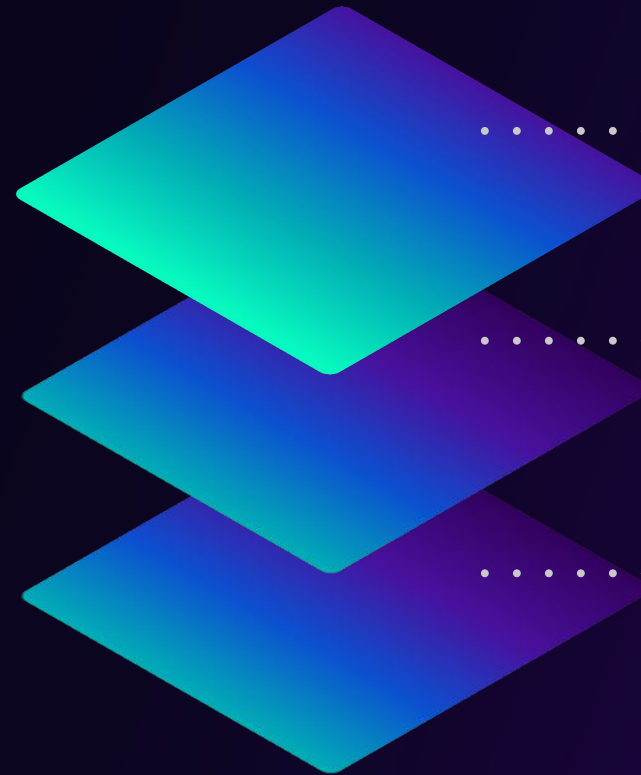
Dzenan is currently working as a part of the Production Engineering team for [Yahoo](#). He is also providing his expertise and services as Cloud and DevOps expert at [TN-TECH](#). Dzenan has been an AWS user since 2011 and has worked on many AWS projects: including migration from on-premises and other public cloud platforms to AWS, and implementing HIPAA compliant infrastructure on AWS with various kinds of companies, from startups to enterprises.

In 2016, Dzenan founded [AWS User Group Bosnia](#) and helped it grow to three user groups with more than 600 members. This AWS community is now the largest IT community in Bosnia. As part of his activities, he runs online meetups, a [YouTube channel](#), and the [sqlheisenberg.com](#) blog (in Bosnian language) to help people in the Balkans region achieve their AWS certification and start working with AWS. In close collaboration with his sister Dzenana, he works with other community members in South-Eastern Europe to help them start and grow AWS user groups in their countries.

Dzenan is passionate about AWS, databases, and security. He holds several AWS, security, and database related certificates and often speaks at various technical meetups and conferences.

[Connect with Dzenan](#)

The Generative AI Stack



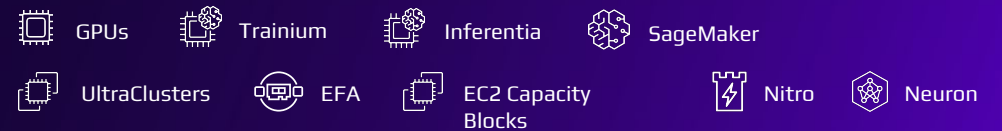
APPLICATIONS THAT LEVERAGE FMs



TOOLS TO BUILD WITH FMS AND LLMS



INFRASTRUCTURE FOR FM TRAINING & INFERENCE



IN PREVIEW



Amazon Q

Trained on 17 years of AWS knowledge

Assists you everywhere you work – in the console, IDE, and documentation

Works with you to build new features, refactor, and troubleshoot on AWS

IN PREVIEW



Amazon Q
... in your IDE

The screenshot displays the Amazon Q IDE interface. On the left, the Explorer pane shows a project named 'SUPPORT-CASES' with files: __pycache__, .aws-sam, preparation, app.py (selected), requirements.txt, samconfig.toml, and template.yaml. The main editor shows the code in app.py, which includes comments and Python code for creating a support case. The code uses the boto3 library to interact with DynamoDB. The terminal at the bottom shows the current directory is 'support-cases'.

```
30 ... # Proceed with creating a case using the  
    provided data  
31  
32 ... # Generate a unique case ID  
33 ... caseid = str(uuid.uuid4())  
34  
35 ... # Put the new item into the DynamoDB table  
36 ... ddbtable.put_item(Item={  
37 ...     'caseid': caseid,  
38 ...     'title': data['title'],  
39 ...     'description': data['description'],  
40 ...     'status': 'open',  
41 ...     # Add any other fields you need  
42 ... })  
43  
44 ... # Return a successful response
```

TERMINAL OUTPUT
zsh + v [] [X] ... ^ X
o → support-cases

IN PREVIEW



Amazon Q

... in CodeCatalyst



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

 Amazon
CodeCatalyst

reinvent-2023

reinvent2023application





#10 Created 7 days ago by @wvogels

Updated 5 days ago by @wvogels

 Comments (2)

 Amazon Q 5 days ago

Good news! I've completed my work on this issue and created a pull request for review:
PR 9 Change the get_all_mysfits() API to return mysfits sorted by the Age attribute

- ✓ Evaluate prerequisites
- Read repository
- ✓ Generate background
- ▶ ✓ Generate approach
- ▼ ✓

You're right, no need to change the model unit tests. Here is a revised approach:

```
1. src/app.py
```

```
1. src/app.py
```

```
1. Modify get_all_mysfits() to sort the response before returning:
```

- 1. Call MysfitsModel.list_items() to get all unsorted items
- 2. Sort the items list using Python's built-in sorted() function, passing 'Age' as the key
- 3. Return the sorted list

IN PREVIEW



Amazon Q
Code Transformation

Upgrade Java applications in a fraction of the time

Enhance the security and performance of your code

Accelerate the migration from Windows to Linux (coming soon)



Amazon CodeWhisperer

New enhancements

Customization capabilities

Support for CloudFormation (YAML, JSON), AWS CDK (TypeScript, Python), and Hashicorp Terraform (HCL)

Available in Visual Studio (in preview)



Amazon Bedrock

The easiest way to build and scale
generative AI applications with
LLMs and other foundation models

Broad choice of foundation models

Customize foundation models using
your organization's data

Enterprise-grade security and privacy

Amazon Bedrock

AI21 labs

Jurassic

Contextual answers,
summarization,
paraphrasing

ANTHROPIC

Claude

Summarization, complex
reasoning, writing,
coding

Claude 2.1

200K token context window
2x reduction in hallucination rate
25% cost reduction on Bedrock

NEW

cohere

Command + Embed

Text generation,
search,
classification

Meta

Llama 2

Q&A and reading
comprehension

Llama 2 70B

Supports large-scale
tasks,
like language-modeling
and dialogue systems
NEW

stability.ai

Stable Diffusion

High-quality images
and art

amazon

Amazon Titan

Text summarization,
generation, Q&A,
search

Amazon Titan Foundation Models



TITAN TEXT EMBEDDINGS

Translates text into numerical representations



TITAN TEXT LITE

Summarization, copywriting, fine-tuning

NEW



TITAN TEXT EXPRESS

Open-ended text generation, conversational chat, RAG support

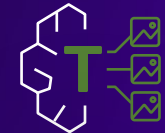
NEW



TITAN MULTIMODAL EMBEDDINGS

Search, recommendation, personalization

NEW



TITAN IMAGE GENERATOR

Generate realistic, studio-quality images

IN PREVIEW

Amazon Bedrock customization capabilities

Fine-tuning Cohere Command Lite,
Meta Llama 2, Amazon Titan Text
Lite & Express generally available

Fine-tuning Anthropic Claude
coming soon

Continued pre-training for Amazon
Titan Text Lite & Express in
preview

GENERALLY AVAILABLE

Knowledge Bases for **Amazon Bedrock**

Fully-managed native support for
retrieval augmented generation (RAG)

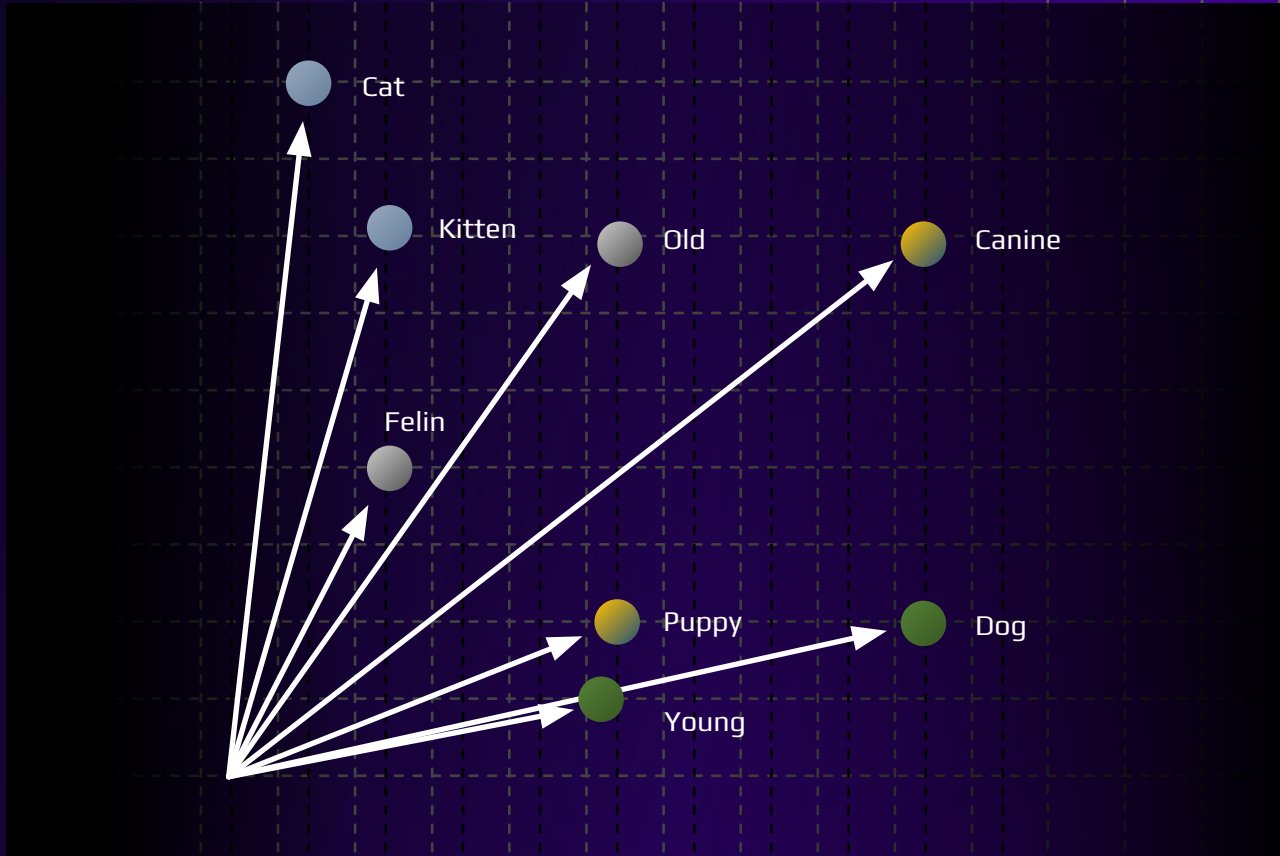
Automatically convert text
documents into embeddings

Store embeddings in your vector
database

Retrieve embeddings and
augment prompts



Vectors are critical for
**customizing generative AI
applications**



Vector support for Amazon **DocumentDB**

GENERALLY AVAILABLE

Vector support for Amazon **DynamoDB**

GENERALLY AVAILABLE



Vector databases for **Amazon Bedrock**



Vector Engine For
Amazon OpenSearch
Serverless



Redis Enterprise
Cloud



Pinecone

COMING SOON



Amazon
Aurora

COMING SOON



MongoDB

GENERALLY AVAILABLE

Agents for Amazon Bedrock

Execute multi-step tasks across
company systems and data sources

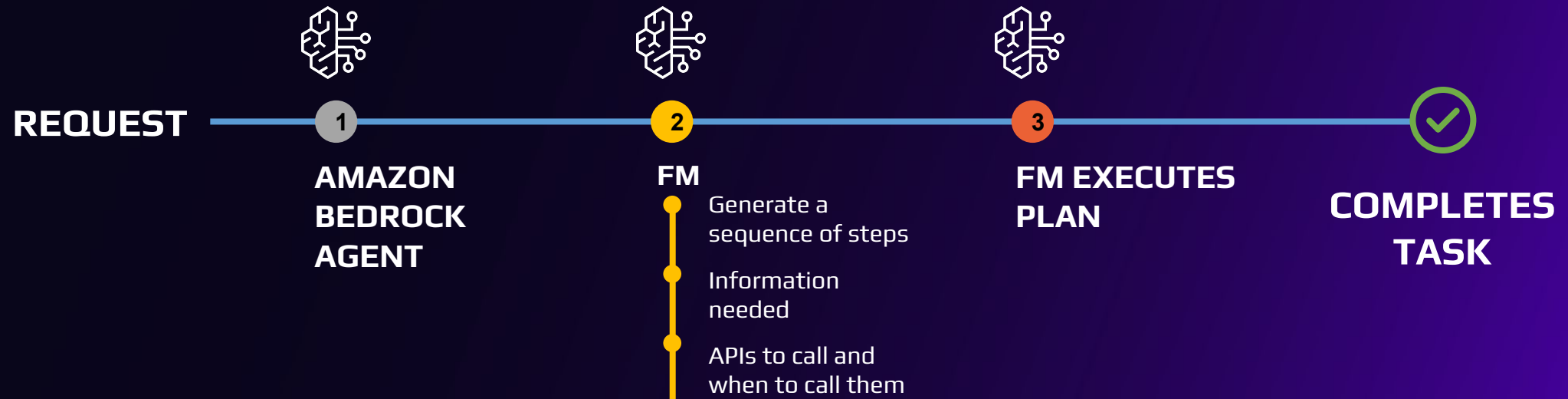
Enable generative AI applications to
take actions in just a few clicks

Break down and orchestrates tasks
and executes API calls on your
behalf

Securely access and retrieve
company data



Resolving tasks with Agents for Amazon Bedrock



IN PREVIEW

Model Evaluation on **Amazon Bedrock**

Evaluate, compare, and select the
best foundation model for any use
case



IN PREVIEW

Guardrails for Amazon Bedrock

Safeguard your generative AI applications with your custom responsible AI policies

Easily configure harmful content filtering

Apply your guardrails to any foundation model or agent

Redact personally identifiable information (PII) in model responses (coming soon)



GENERALLY AVAILABLE

AWS Application Composer in VS Code

Visually compose infrastructure as code directly in your IDE

Drag, drop, and connect AWS resources using a visual canvas

Use generative AI-based code suggestions to use any of 1,000+ CloudFormation resources

AWS Application Composer is now part of the AWS Toolkit for VS Code



AMAZON Q (PREVIEW): CHAT

Chat

Hi, I am Amazon Q (Preview). I can answer your software development questions. Ask me to explain, debug, or optimize your code. You can enter / to see a list of quick actions.

How do I use Amazon EventBridge with GameLift FlexMatch to trigger Lambda when a potential match is created? I want CloudFormation for this.

To use Amazon EventBridge with GameLift FlexMatch to trigger a Lambda function when a potential match is created, you can create a CloudFormation template that sets this up. Here are the basic steps [1]:

1. Create an Amazon EventBridge rule that matches the "PotentialMatchCreated" event from your GameLift FlexMatch configuration [2] (https://docs.aws.amazon.com/gamelift/latest/events.html). The rule would target a Lambda function.

2. Create an AWS Lambda function that will be invoked when the event is matched. This function could process the match data and take actions like notifying players.

3. Use CloudFormation to define the Lambda function resource, an IAM role for it to assume, and the EventBridge rule resource linking the two.

4. Deploy the CloudFormation stack. EventBridge will then start routing "PotentialMatchCreated" events to invoke your Lambda function.

The AWS documentation has more details on setting up FlexMatch event notification and using EventBridge with CloudFormation. Let me know if you have any other questions [3]!

Sources

[1] Set up FlexMatch event notifications - Amazon GameLift

docs.aws.amazon.com/gamelift/latest> match>

Show more

Suggested follow up questions:

How do I define the EventBridge rule?

What permissions does the Lambda role...

What does the Lambda function do?

Ask a question or enter "/" for quick commands

4000/4000

Amazon Q may produce inaccurate information. Verify all provided

template.yaml (Application Composer)

Arrange

Sync

API Gateway

PlayerAPI

GET /playerdata

POST /playeravatar

POST /playersession

Lambda Function

PlayerDataManager

Lambda Function

AvatarProcessing

Lambda Function

GameSessionManager

DynamoDB Table

PlayerDataTable

S3 Bucket

PlayerAvatar

Standard Component

Fleets

Standard Component

West

template.yaml

! template.yaml > CloudFormation Linter > {} Resources >

171 GameSessionManager:

172 Type: AWS::Serverless::Function

173 Properties:

174 Handler: index.handler

175 Runtime: nodejs18.x

176 Policies:

177 - AWSLambdaBasicExecutionRole

178 Environment:

179 Variables:

180 FLEET_ID: !Ref Fleets

181 CodeUri: src/GameSessionManager

182 MemorySize: 3008

183 Timeout: 30

184 AWS: Add API Debug Configuration

185 Events:

186 PlayerAPIPOSTplayersession:

187 Type: Api

188 Properties:

189 Path: /playersession

190 Method: POST

RestApiId: !Ref PlayerAPI

NEW

Generative AI CDK Constructs

[github.com/aws-labs/
generative-ai-cdk-constructs](https://github.com/aws-labs/generative-ai-cdk-constructs)



github.com/aws-labs/generative-ai-cdk-constructs

Public

<> Code Issues Pull requests Discussions Actions Projects Security Insights

main 2 branches 89 tags Go to file Code

github-actions and krokoko chore(deps): upgrade dependencies 1b11e6b 2 hours ago 239 commits

.github	fix(bandit): Correct GitHub workflow for Bandit	2 weeks ago
.projen	added cdk-nag dependency	yesterday
docs	fix architecture diagrams	2 weeks ago
lambda	fix issue where bytes and not str are passed to response in case of e...	last week
layers	chore(deps): bump langchain in /layers/langchain-common-deps	2 weeks ago
projenrc	fix(bandit): Correct GitHub workflow for Bandit	2 weeks ago
resources/gen-ai	added valition on qa schema for qainput	last week
src	Updated port to 8686 for testing.	yesterday
test/patterns/gen-ai	cdk nag fixes	last week
.eslintrc.json	feat: Feature/ingestion construct (#8)	last month
.gitattributes	feat(bandit): Add Bandit (#108)	2 weeks ago
.gitignore	feat(bandit): Add Bandit (#108)	2 weeks ago
.mergify.yml	feat(construct): enable pypi (#46)	last month



AWS SDKs for Rust & Kotlin



GENERALLY AVAILABLE

AWS Step Functions

HTTP endpoint integration

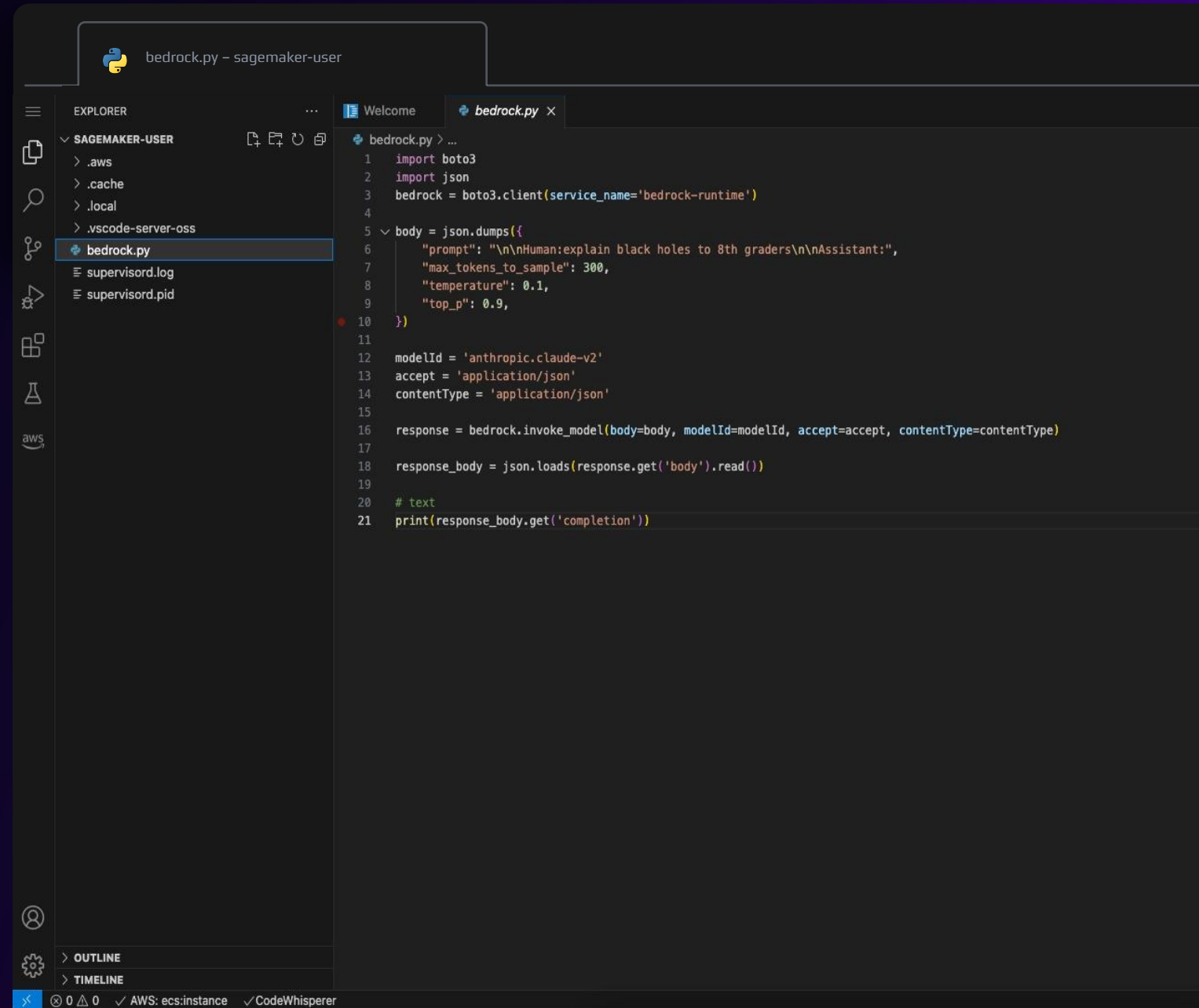
Native Amazon Bedrock integration

New Test States API

GENERALLY AVAILABLE

Amazon SageMaker Studio Code Editor

A new IDE option based on Code-OSS



GENERALLY AVAILABLE

Amazon CloudWatch

Pattern Analysis and Anomaly Detection

Patterns view visualizes recurring patterns while querying your logs

Compare mode helps you quickly find changes over time

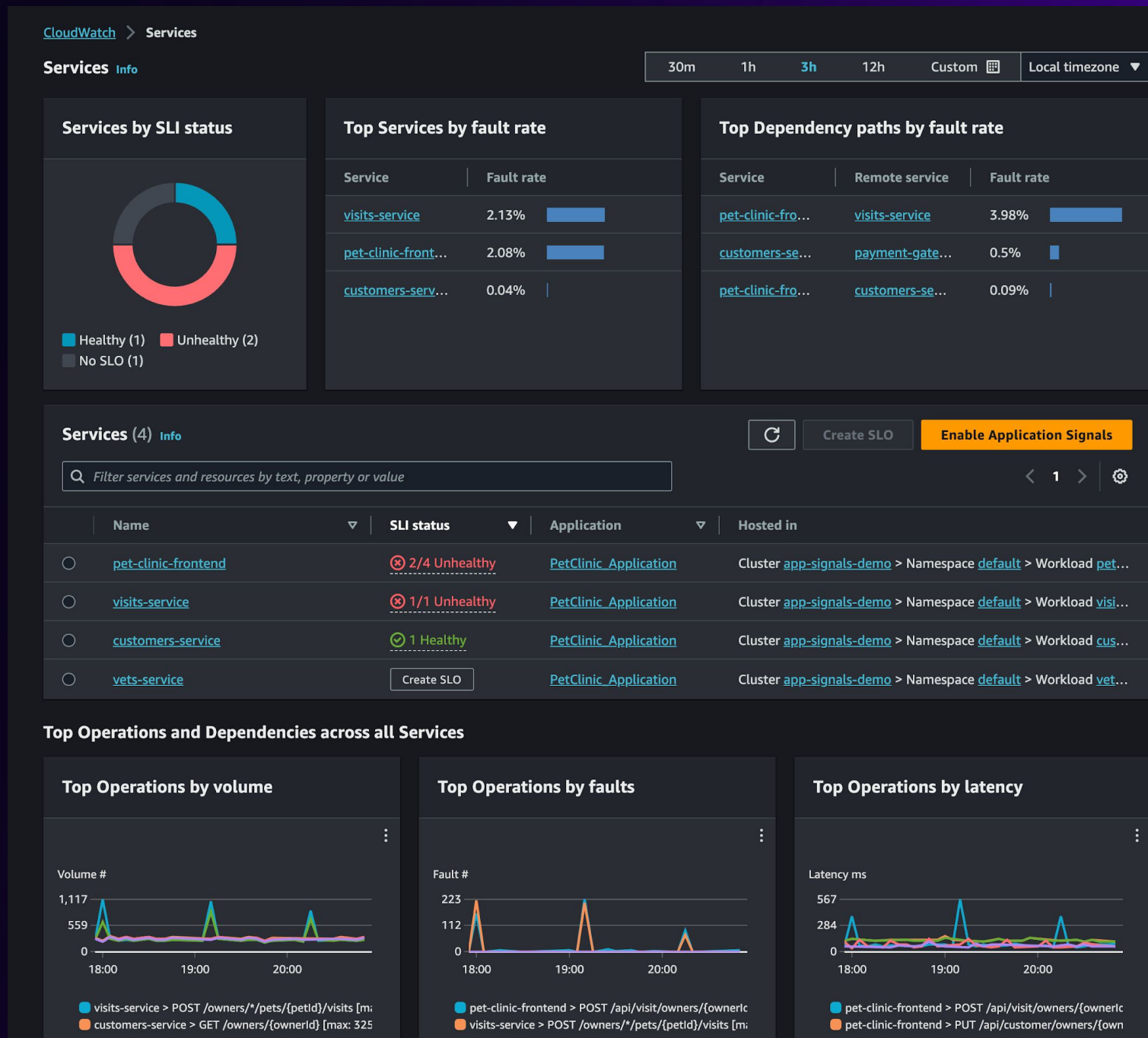
Anomaly detection evaluates incoming logs against historical baselines

IN PREVIEW

Amazon CloudWatch Application Signals



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.



GENERALLY AVAILABLE

AWS Secrets Manager batch-retrieval of secrets

BatchGetSecretValue identifies and retrieves secret groups with a single API call

Use with list of secret names or ARNs, along with filter criteria, such as tags

GENERALLY AVAILABLE

Amazon Inspector Lambda code scanning with generative AI

Assess function code for security issues like injection risks, data leaks, and weak or missing encryption

Actionable security findings including affected code snippets and remediation suggestions

GENERALLY AVAILABLE

Amazon Inspector container image security

Scans container images for security issues from within the CI/CD pipeline

Works with CI/CD tools hosted in AWS, on-premises, or hybrid clouds



IN PREVIEW

Amazon Inspector agentless vulnerability assessments for EC2

Continuous EC2 instance monitoring
without the need for an agent

Takes EBS volume snapshots to
perform vulnerability analysis



ECS: GA / EC2: IN PREVIEW

AWS Guard Duty

ECS and EC2 runtime monitoring

[GuardDuty](#) > Runtime Monitoring

Runtime Monitoring [Info](#)

Monitor runtime activity to detect threats to your compute workloads on Amazon ECS (including AWS Fargate), Amazon EKS, and Amazon EC2. To use this feature, you need to manage a GuardDuty security agent that will ingest runtime activity for your resource.

Configuration | **Runtime coverage**

Coverage statistics [Info](#)

Healthy EKS clusters

0/0 (100%)

Healthy ECS clusters

1/1 (100.0%)

Healthy EC2 instances

0/0 (100%)

EKS clusters runtime coverage

ECS clusters runtime coverage

EC2 instance runtime coverage

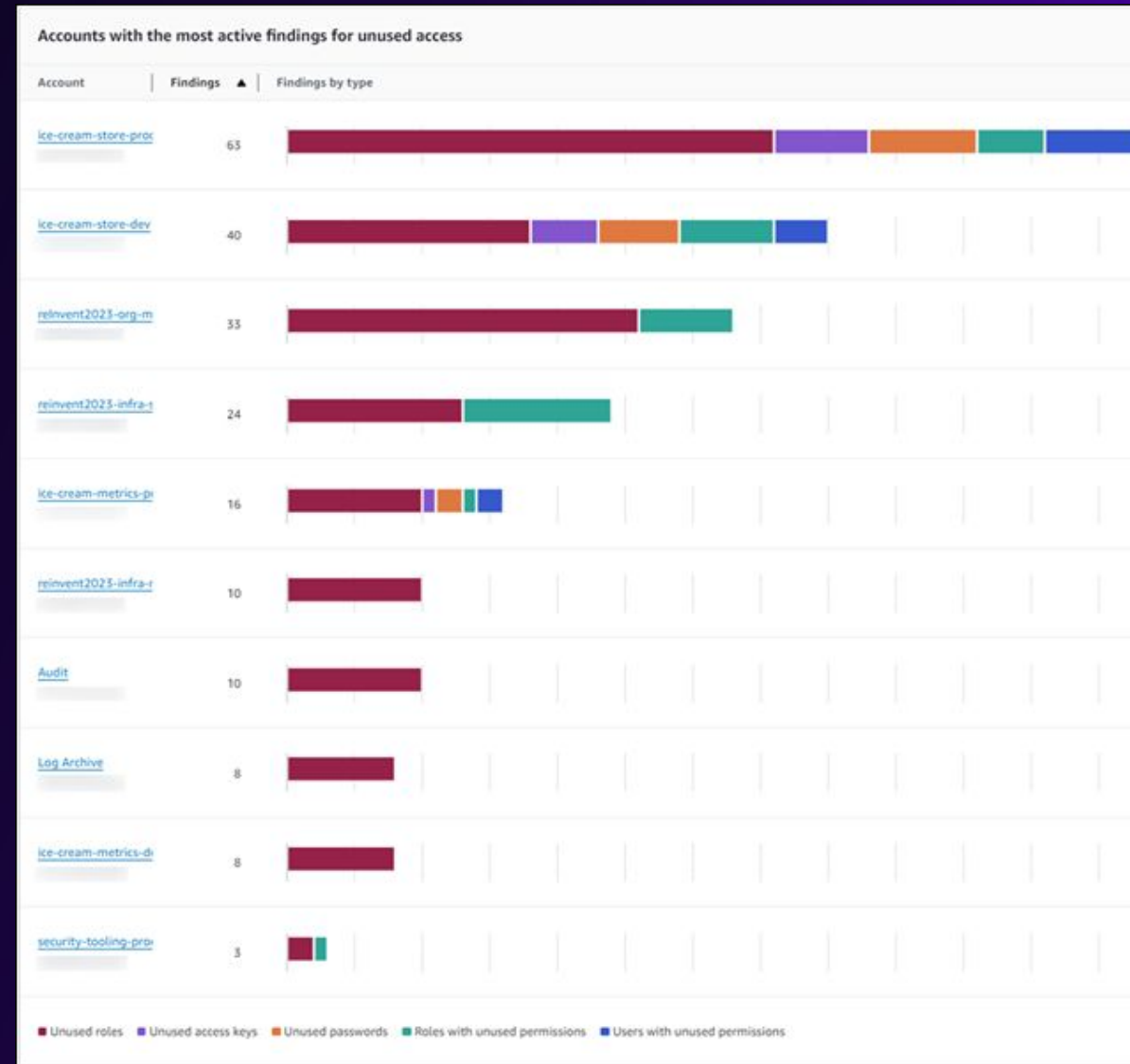
Clusters list (1)

Cluster name	Account ID	Agent management type (Fargate)	Coverage status
SpaceTwin-Infra-SpaceTwinCluster3D77526A-5D6rTXjgdSb2	636183562745	Auto-managed	✓ Healthy

GENERALLY AVAILABLE

IAM Access Analyzer

simplified access inspection



GENERALLY AVAILABLE

Unified Billing and Cost Management

Unified console with recommendations and insights to help understand your spend, identify issues, and save money



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Billing and Cost Management home [Info](#)

[Reset layout](#)

Cost summary [Info](#)

Month-to-date cost

\$5.87

↓ 2% compared to last month for same period

Last month's cost for same time period

\$6.00

Nov 1 – 5

Total forecasted cost for current month

\$11.66

↓ 2% compared to last month's total costs

Last month's total cost

\$11.90

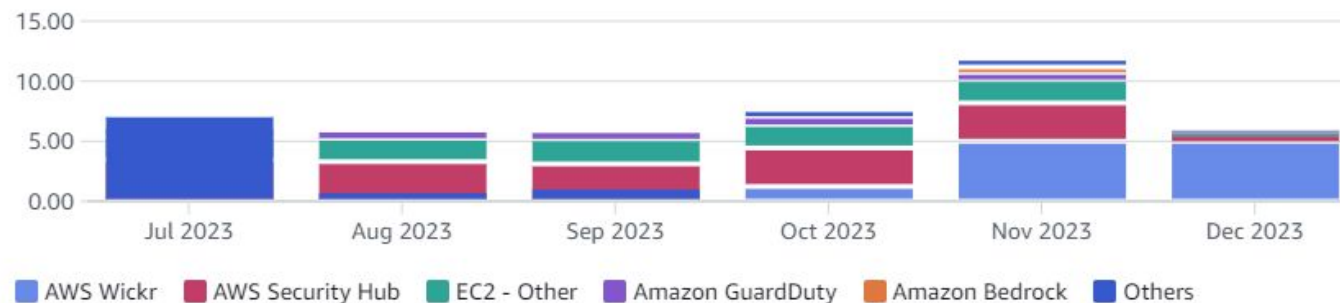
[View bill](#)

Cost breakdown [Info](#)

Group costs by

Service ▼

Costs (\$)

[Analyze your costs in Cost Explorer](#)

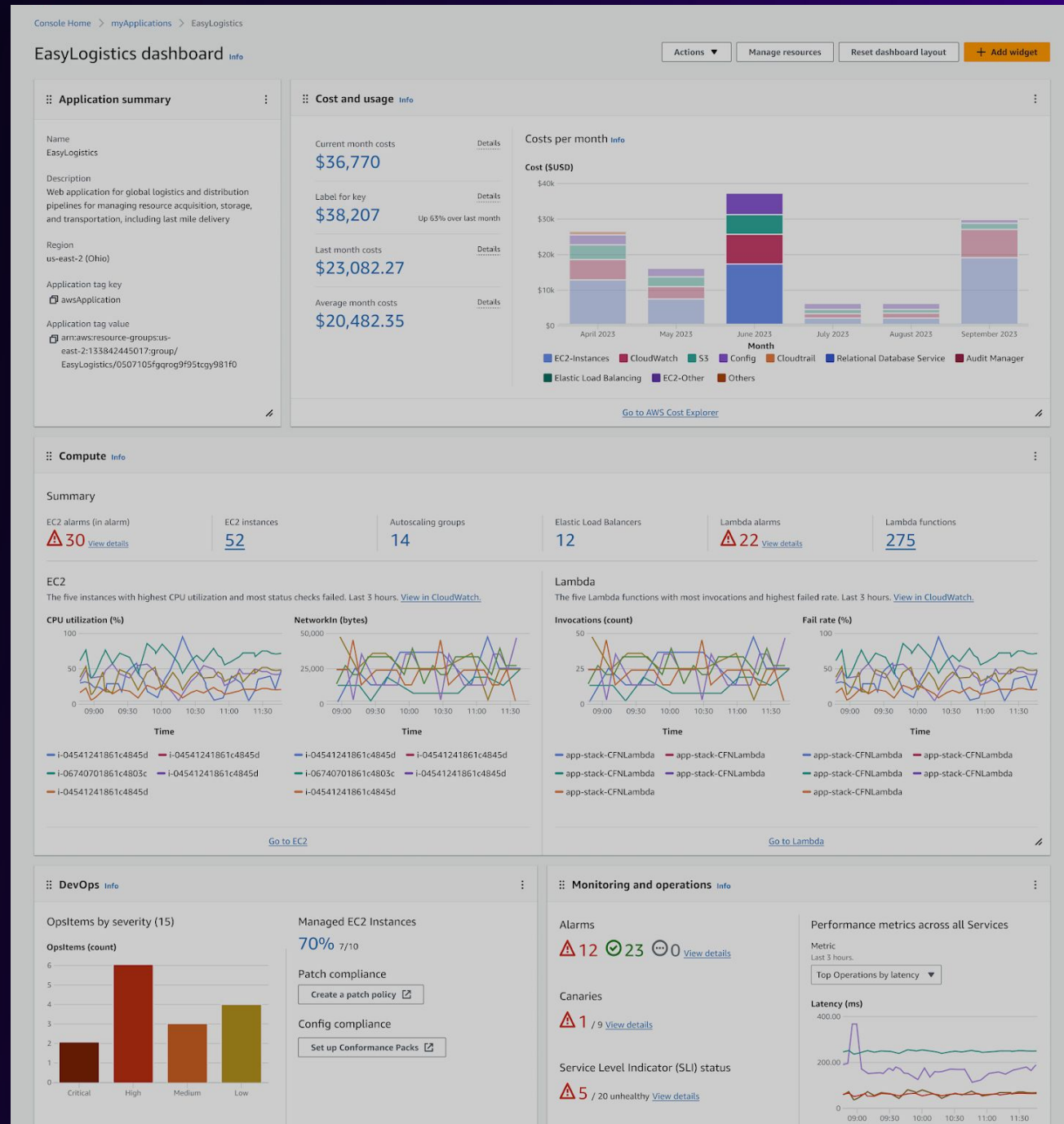
GENERALLY AVAILABLE

AWS Management Console myApplications

Monitor and manage the cost, health,
security posture, and performance of your
applications



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.



GENERALLY AVAILABLE

AWS Free Tier usage API

Compare forecasted usage to the free tier for each service to avoid unintended charges

Programmatically track free tier usage against monthly usage limits

GENERALLY AVAILABLE

Amazon CloudWatch Infrequent Access class

A new log class for cost-effective
log consolidation

Best suited for logs that require
infrequent querying, e.g. for
forensic analysis



GENERALLY AVAILABLE

AWS Systems Manager Automation low-code Runbook design

Drag and drop automation actions
and AWS APIs and connect them as
steps

Features a guided flow, best practice
recommendations, and error
highlighting

Integrates with Amazon CodeGuru
Security for embedded Python scripts

Amazon S3 Express One Zone

The fastest cloud object storage class for performance-critical applications

Millions of access requests per minute with single-digit millisecond latency

New bucket type, pinned to the Availability Zone of your choice

Mountpoint for **Amazon S3**

Container Storage Interface (CSI)
driver for Kubernetes now available

Added support for the new S3
Express One Zone storage class

Amazon EFS

New Archive storage class optimized for long-lived file data that is accessed a few times a year or less

Infrequent Access storage class price reduction by 36%

Now supports up to 250,000 read and up to 50,000 write IOPS per file system

AWS Backup

Now supports automatic archival
of EBS Snapshots into the EBS
Snapshot Archive

Restore testing helps perform
automated and periodic restore
tests of backed up resources

IN PREVIEW

Amazon Aurora Limitless Database

Managed horizontal scale-out beyond
the limits of a single database instance



GENERALLY AVAILABLE

Amazon ElastiCache Serverless

Serverless Amazon ElastiCache
configuration for Redis and
Memcached



GENERALLY AVAILABLE

Amazon Q generative SQL in **Amazon Redshift**

Simplify query authoring by
expressing queries in natural
language



Zero-ETL integrations with Amazon Redshift

**Amazon Aurora
PostgreSQL**

**Amazon RDS
for MySQL**

**Amazon
DynamoDB**

AVAILABLE IN PREVIEW



Zero-ETL integrations

**Amazon
OpenSearch
with
Amazon S3**

**Amazon
DynamoDB
with
Amazon OpenSearch**

**Amazon Athena
with
AWS CloudTrail
DataLake**

AVAILABLE IN PREVIEW



ANNOUNCEMENT

Project Kuiper

Private connectivity to AWS





PartyRock

Join the party!



Thank you!



Please complete the session
survey in the mobile app