AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

CON207

Your first container in the cloud

Omar Lari

Principal Product Manager
Amazon Web Service

Miller Fanning

Principal Software Engineer
Cox Automotive

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

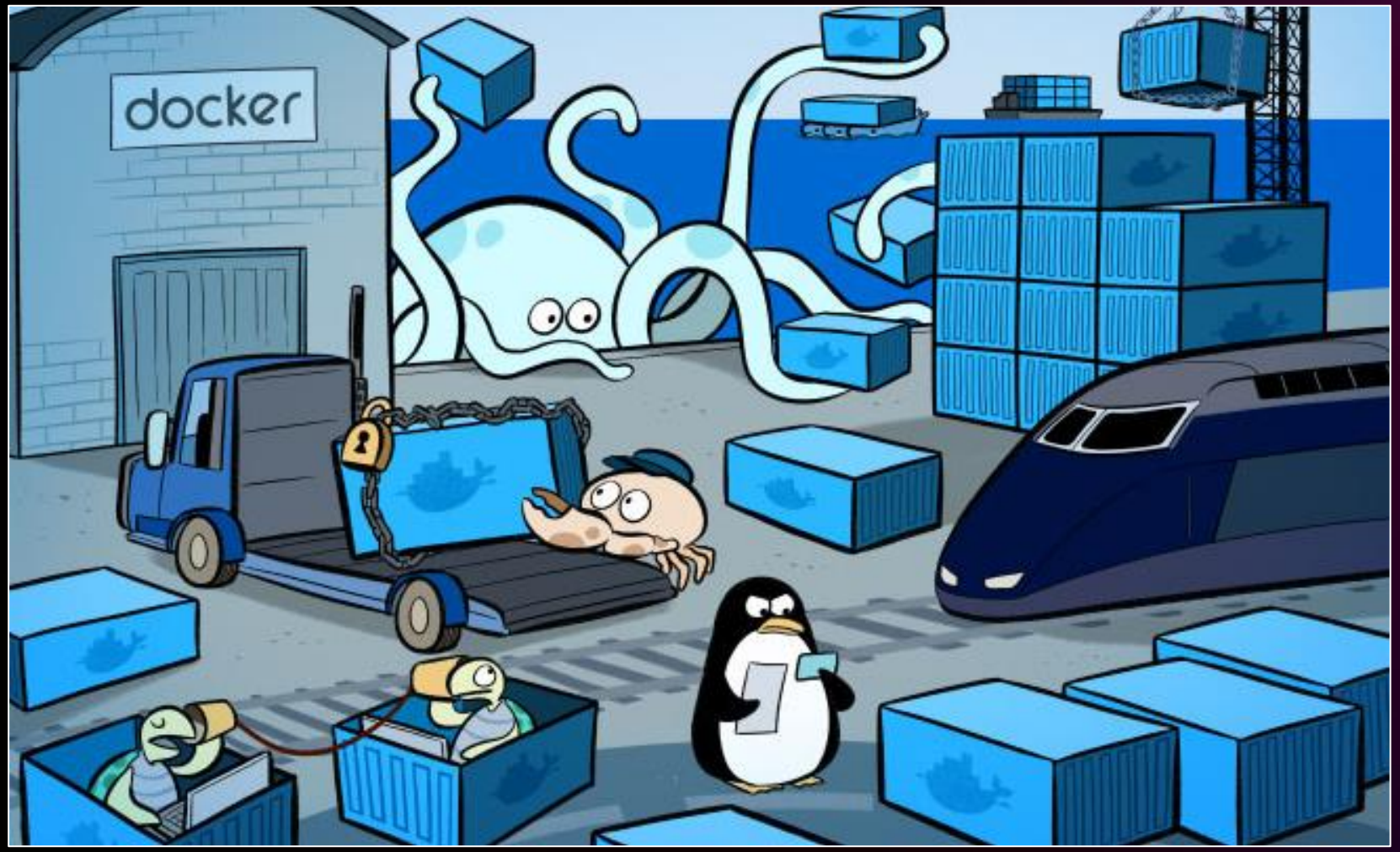
Agenda

One, two, go! – Two easy ways to get started with AWS App Runner

What is App Runner and why did we build it?

Operating App Runner in production

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Bryan Cantrill

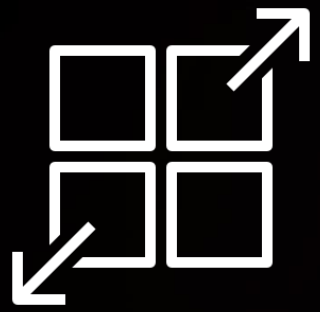
@bcantrill

Follow

Question I've been posing at #KubeCon: are we near Peak Confusion in the container space? Consensus: no -- confusion still accelerating!

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS App Runner



Your web applications

MANAGED BY AWS

No EC2 instances, clusters, load balancers that you need to provision, scale, or manage.

ELASTIC

Scale up and down seamlessly. Pay only for what you use.

INTEGRATED

With the AWS portfolio of services: VPC Networking, Amazon CloudWatch, AWS Certificate Manager (ACM), and more.

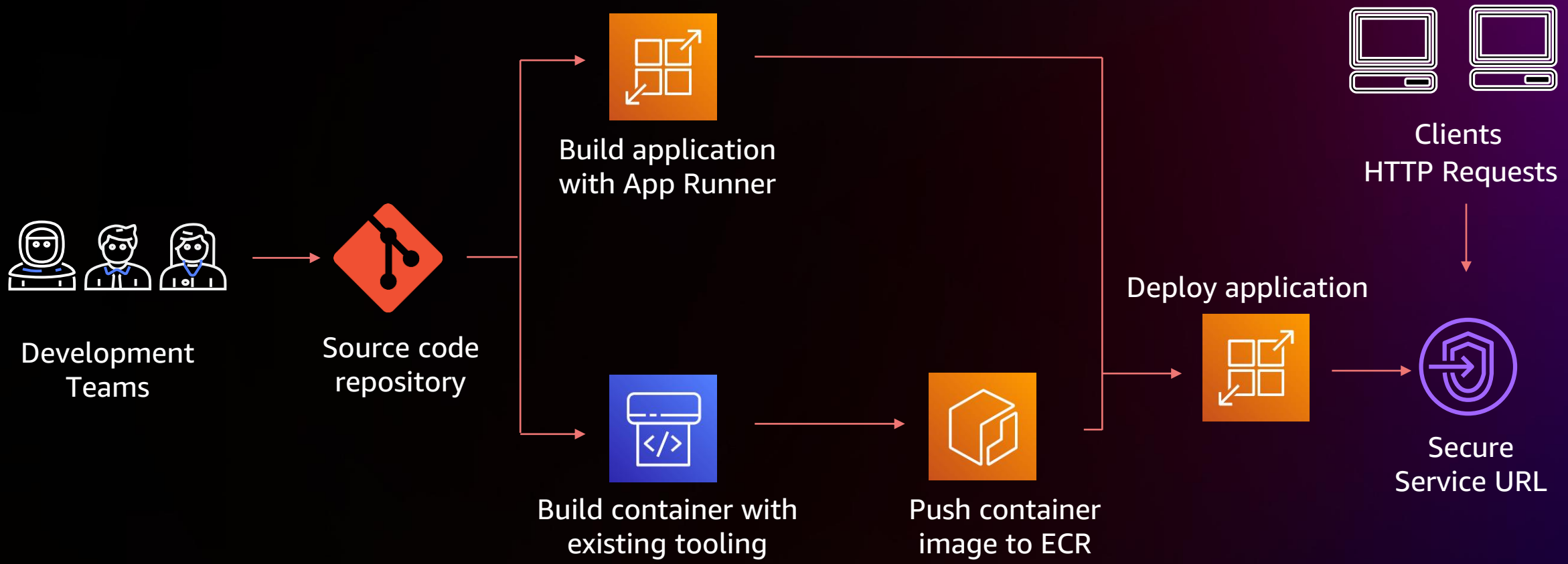
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Idea, Code, Launch

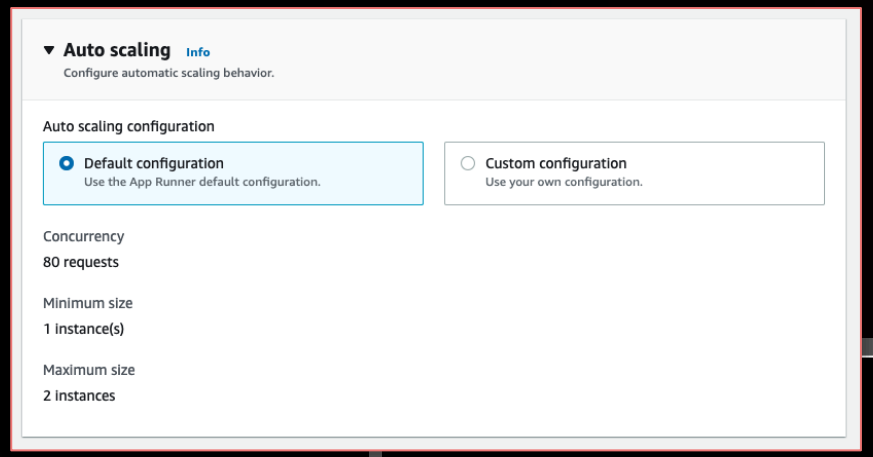
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Deploy from source or container image



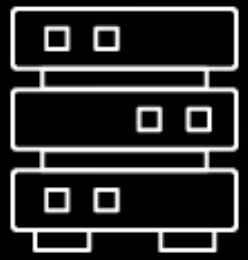
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Pay for compute that scales with requests



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Pricing dimensions



Compute pricing
$0.064/vCPU-hour
$0.007/GB-hour

Automatic deployments
$1/application, per month

Build fee
$0.005/build minute

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Pricing scenarios

Development or test app

Built by individual developers, these apps serve test or development traffic and use manual deployments. Developers often pause these services to manage costs.

$5/month

Lightweight, latency-sensitive API

Lightweight commercial API's or web services see low request volume and do not undergo any peak hour activity.

$25/month

High-volume production app

Production app powers your most latency-sensitive business logic. They see several peak traffic hours daily.

$102/month

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cox Automotive

Autotrader

Cox Automotive Mobility

Dealer.com

Dealertrack

F&I Express

FYUSION

Kelley Blue Book

Manheim

NextGear Capital

SNT Spiers New Technologies Cox Automotive / Mobility

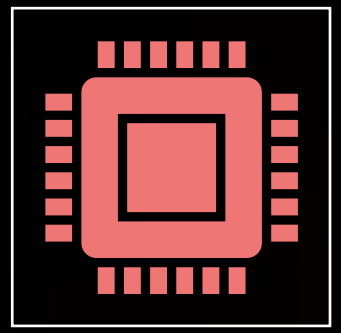
vAuto Live Market View

VinSolutions

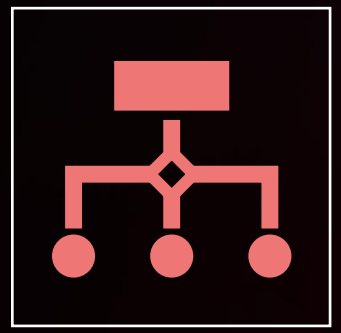
xtime

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

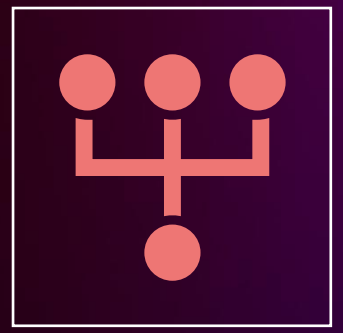
Model delivery At Cox Automotive



Our goal is to do for data science what DevOps did for software development



One of the problems that we had to solve was how to run models as APIs at scale

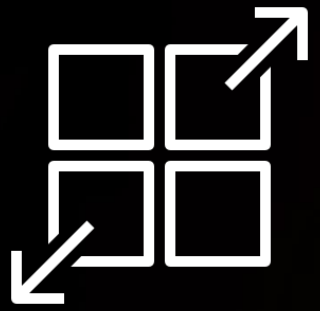


Can we build a managed service–driven platform to deliver a wide range of models?

aws

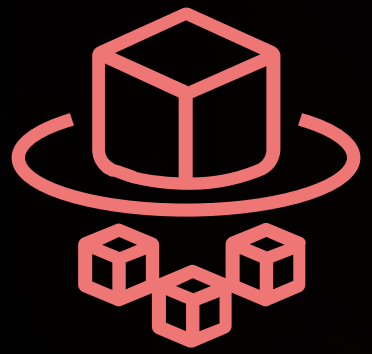
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Containers in the cloud



AWS App Runner

Amazon ECS



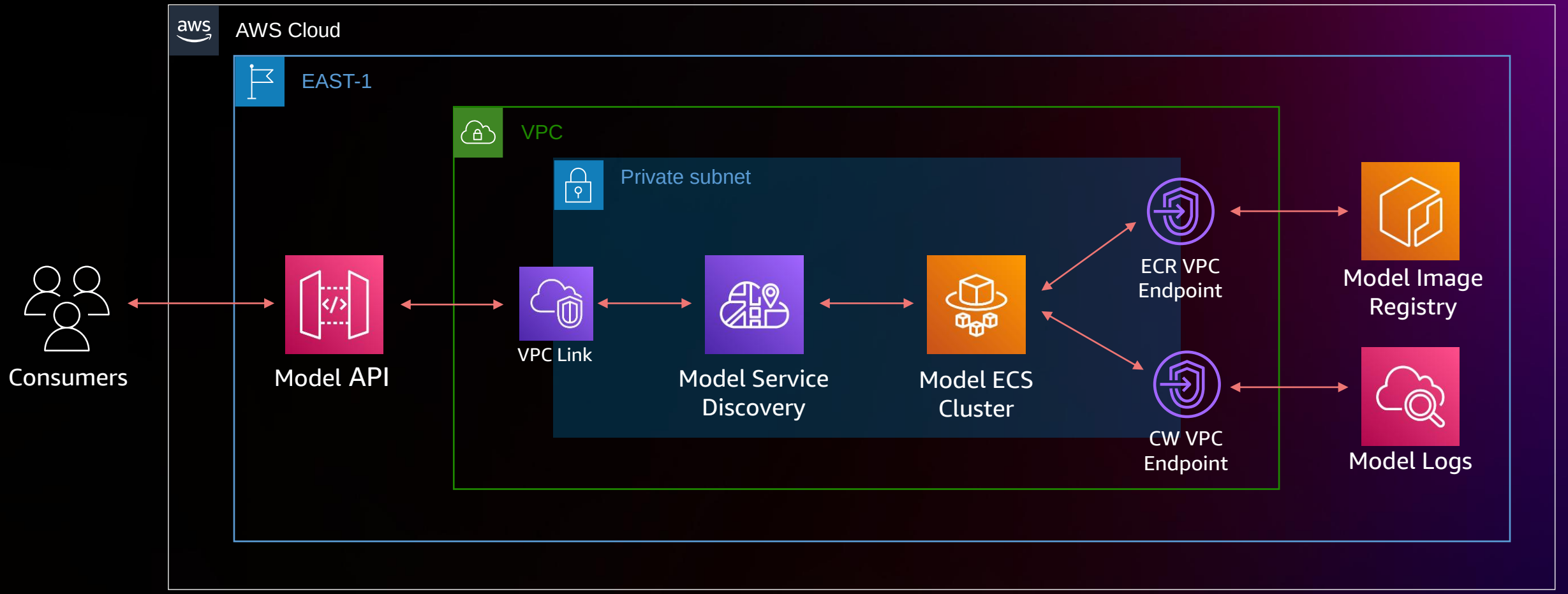
AWS Fargate

Amazon EKS

aws

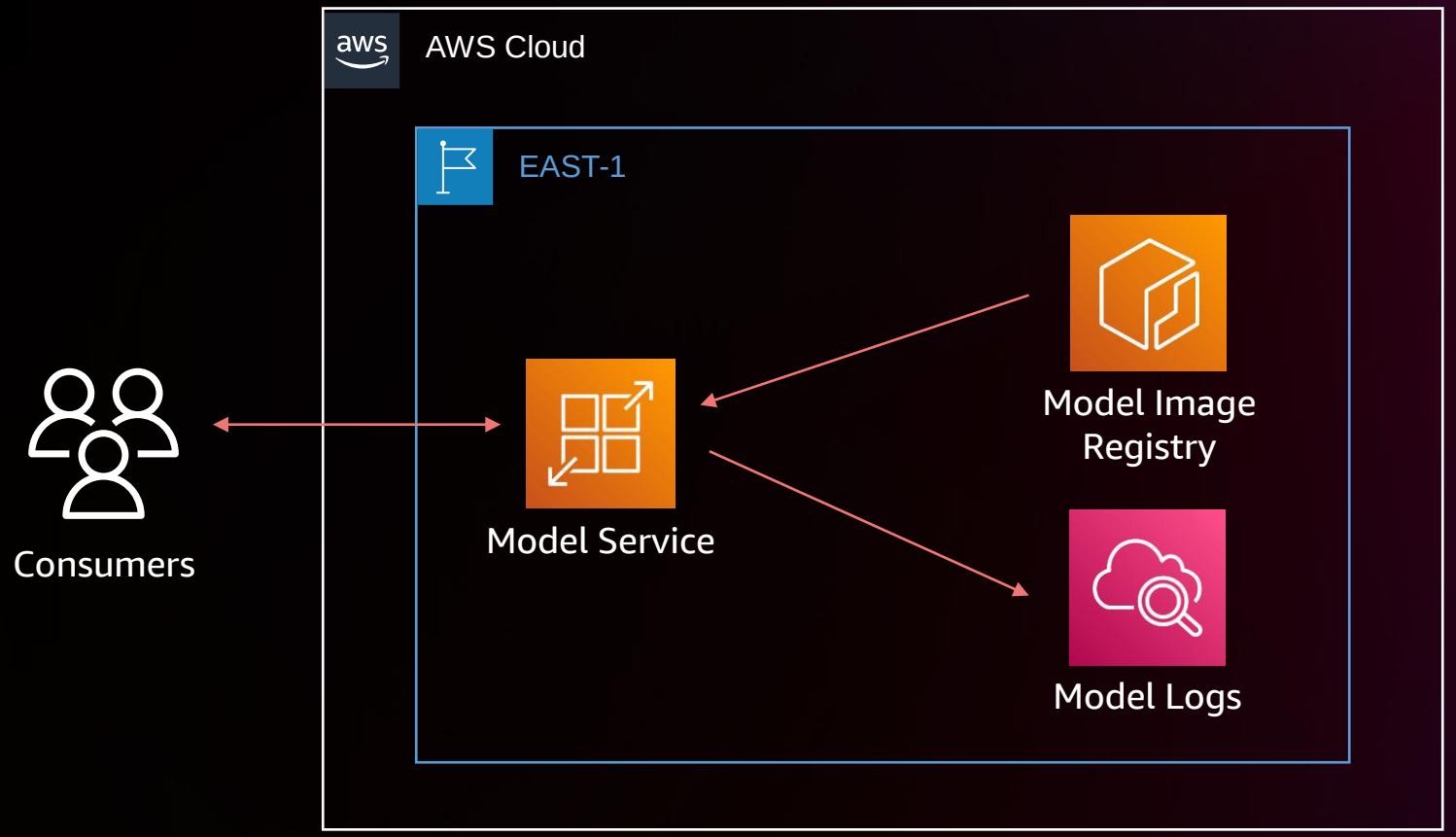
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Known ECS pattern



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Fitting into the App Runner pattern



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Simplified operations

Over-provisioning

- Solve for burst traffic
- Previous model versions scaled
- Lower environments mirror production

Avoiding scale to 0

- Keep lower environments warm for rapid testing
- Minimal cost for infrequently used models

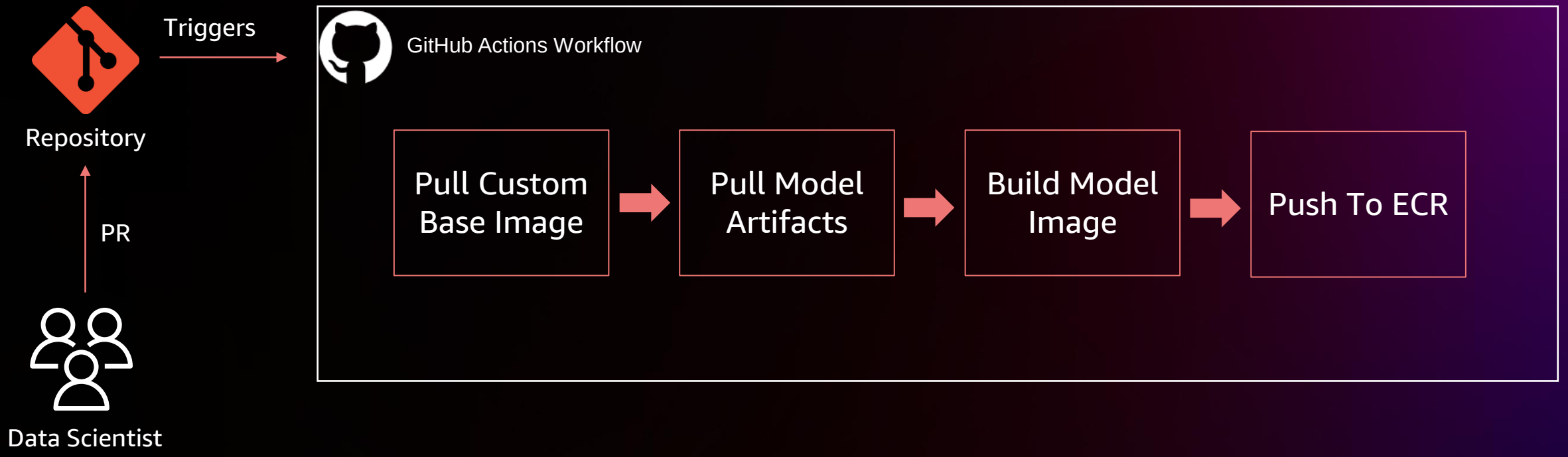
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Taking *YOUR* first container to the cloud

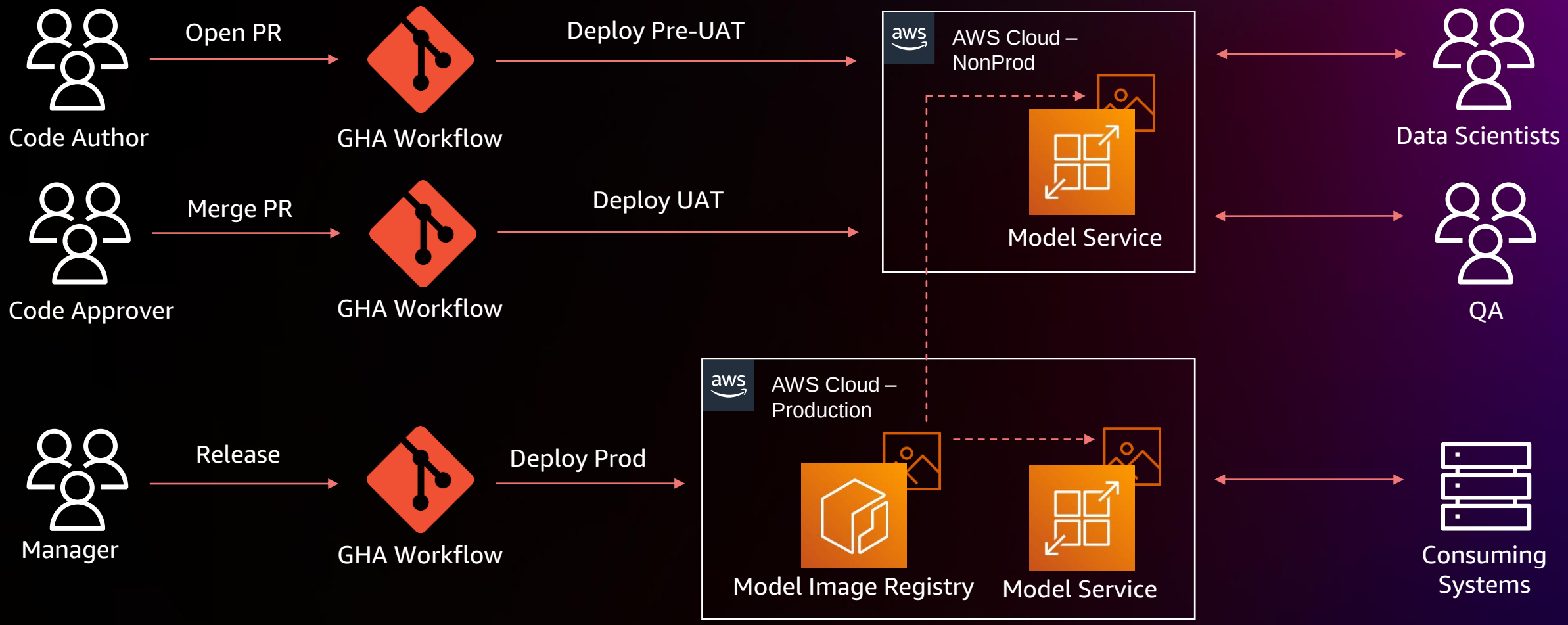
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Building the container image



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Path to production

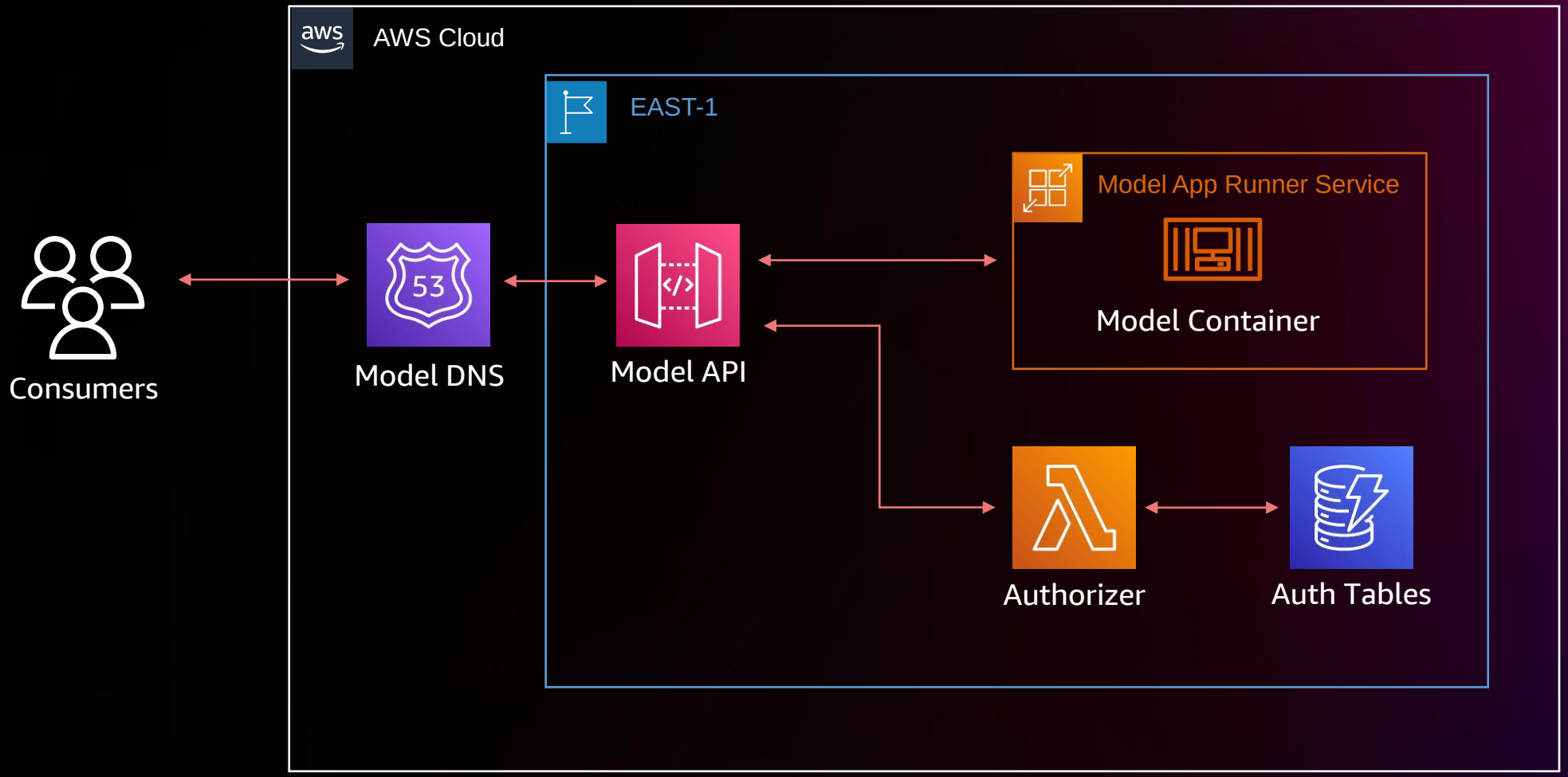


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Extending *YOUR* first container to the cloud

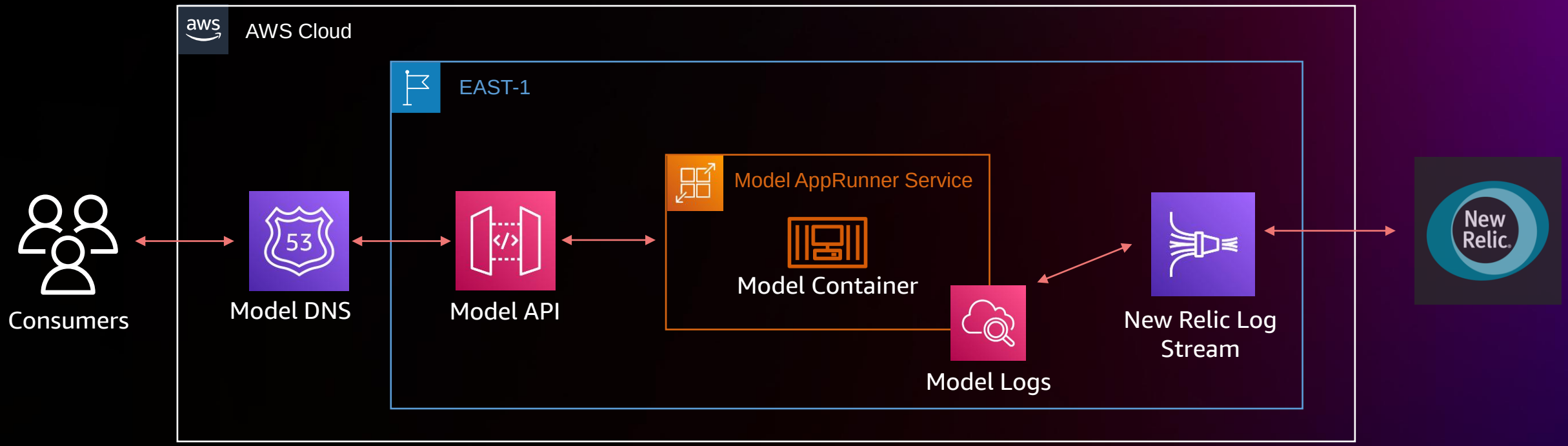
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon API Gateway integration



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Integrating App Runner logging capabilities



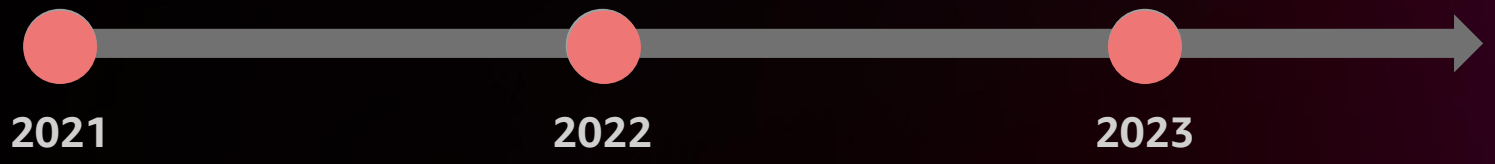
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Opportunities for improvement

- AWS CloudFormation support
- Compute SKUs
- Steady-state optimization

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

App Runner – Launch timeline



AWS App Runner launches in 2022
February – VPC Egress
February – Java run time and Node.js 14
March – FIPS endpoint support
April – AWS X-Ray integration
August – Support for alias records
September – Cross-Region ECR support
September – Node.js 16 runtime
October - Support for .NET core, PHP, Ruby, and Go
October – VPC Ingress (private services)

Please visit our GitHub roadmap to upvote:
https://github.com/aws/apprunner-roadmap/issues

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Want to learn more?

- CON406 – A close look at AWS Fargate and AWS App Runner
- CON305 – Deploy Node.js using App Runner for applications at scale
- CON408 – Deploying containerized applications in minutes with AWS App Runner

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Omar Lari
Twitter: @omarlari

Miller Fanning
LinkedIn: miller-fanning

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

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.