
AWS
re:Invent

CON305

Running machine learning workflows at enterprise scale using Kubeflow

Peter Dalbhanjan
Container Specialist Solutions Architect
AWS

Mohan Muppidi
ML Cloud Architect
iRobot

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

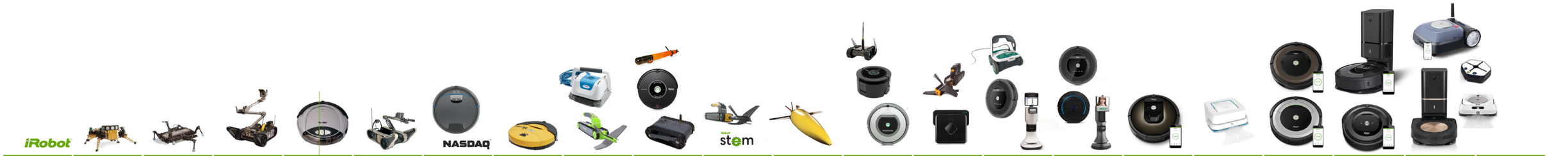


THE CONSUMER ROBOT COMPANY

iRobot is the global home robot leader. **We bring over 30 years of advanced robotics experience** into helping people find smarter ways to clean and accomplish more in their daily lives.

Our history

The building blocks of the world's coolest robots



1990 1991 1996 2001 2002 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020

1990

MIT roboticists Colin Angle, Helen Greiner and Rodney Brooks co-found iRobot.

2001

The iRobot PackBot® is fielded for disaster response and other life-saving missions

2002

iRobot launches the first Roomba® Vacuuming Robot

2010

iRobot celebrates 20 years of making a difference.

2015-2016

iRobot launches the Roomba® 980 Vacuuming Robot and the Braava jet™ Mopping Robot.

2019

iRobot launches the Roomba s9+ Robot Vacuum and Braava jet m6 Robot Mop, announces the Terra Robot Mower, and adds the Root Coding Robot to its product line.

30 YEARS 1990-2020

Agenda

Journey

Evaluation criteria

Use cases

Kubeflow architecture

Best practices



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

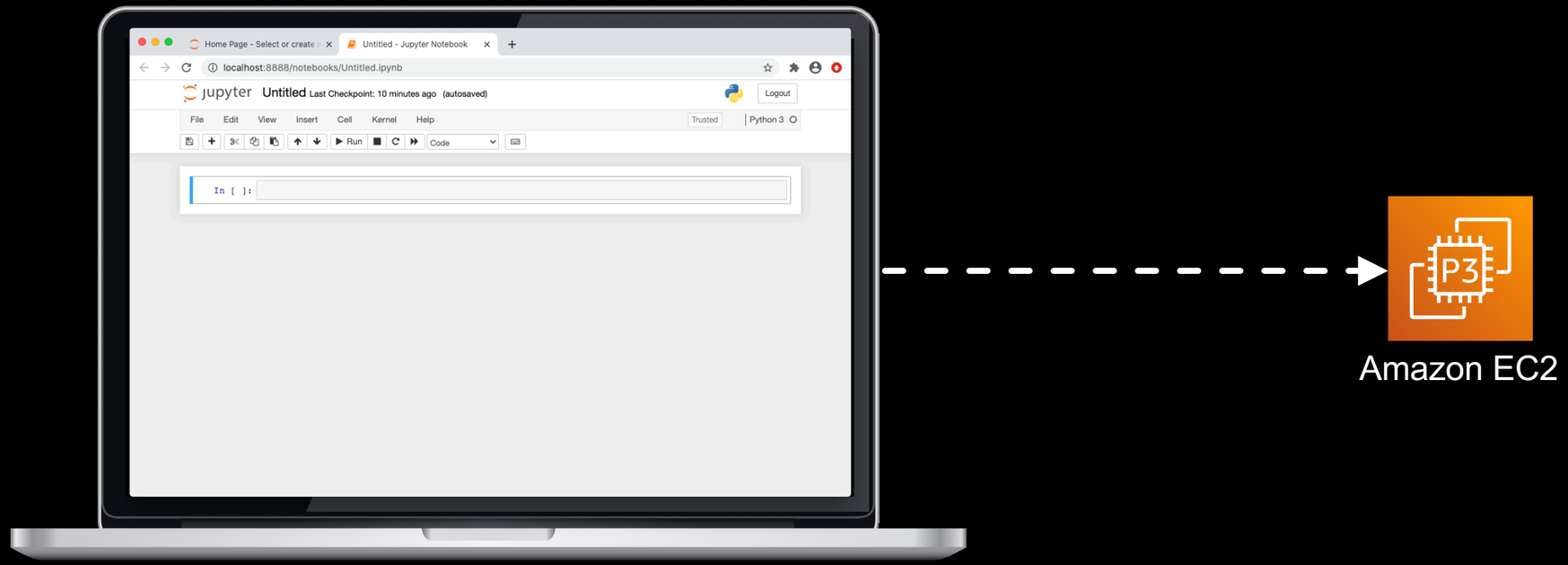
Journey



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Typically start with Amazon EC2



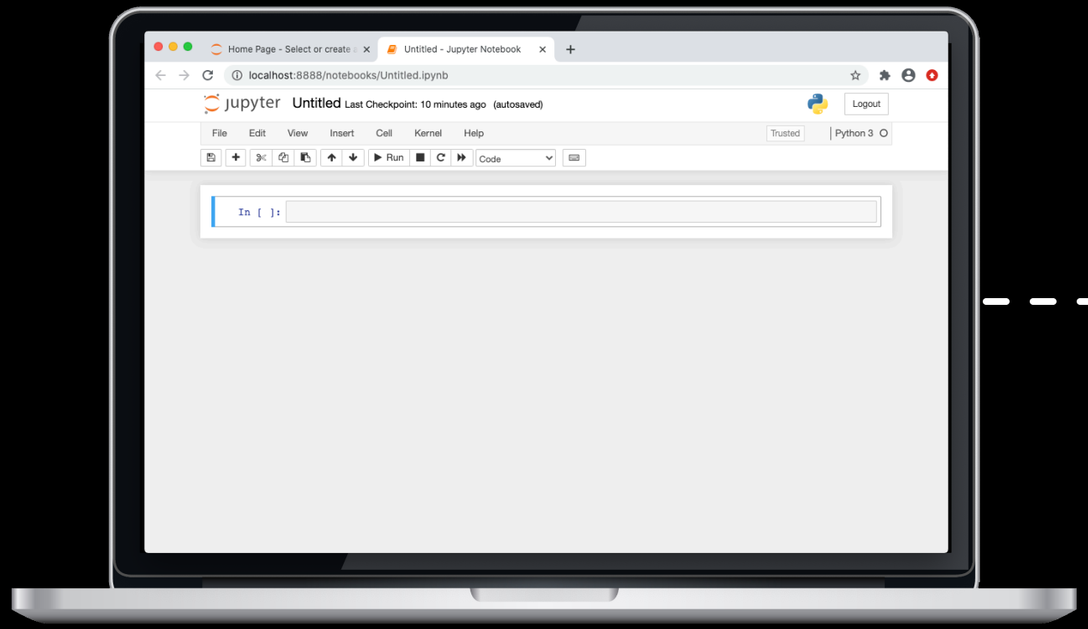
Pros and cons

- Simple use cases
- Easy to get started
- Use infrastructure-as-code to automate

- Difficult to scale
- Requires expertise at hardware level
- Adds maintenance overhead

Tip: Use Deep Learning AMI

Containers



Containers

Amazon Elastic Container Service

Amazon Elastic Kubernetes Service

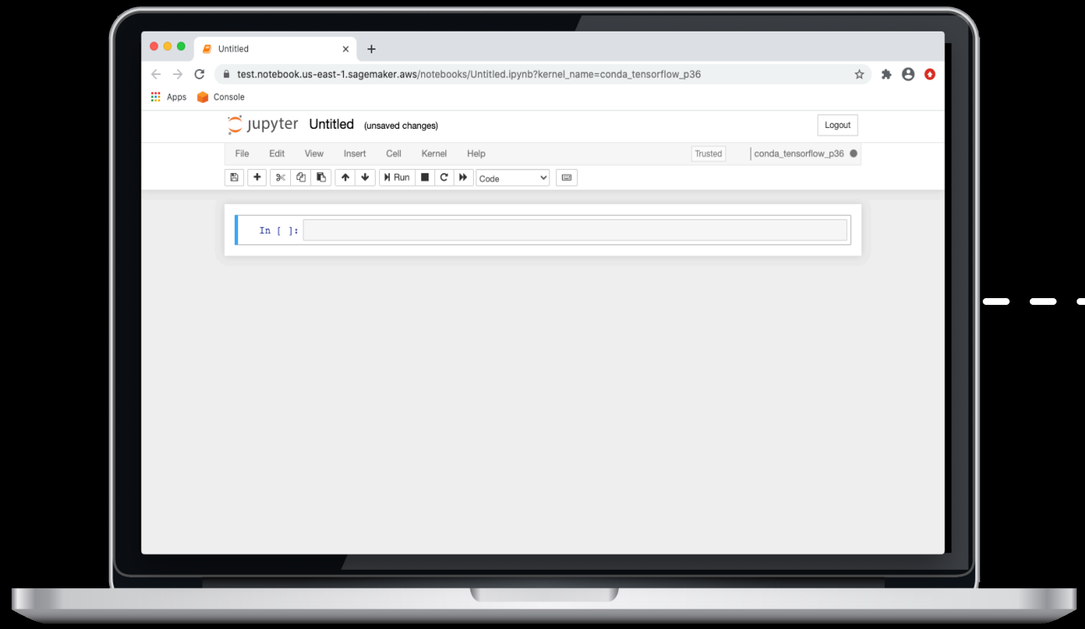
Pros and cons

- Standardization
- Portability
- Shared infrastructure

- Need to build abstractions
- Increases learning curve
- Adds maintenance overhead

Tip: Use Deep Learning Containers

Amazon SageMaker and higher level services



Amazon SageMaker

Text
- Amazon Translate
- Amazon Textract
- Amazon Comprehend

Vision
- Amazon Rekognition

Search
- Amazon Kendra

Fraud
- Amazon Fraud Detector

Chatbots
- Amazon Lex

Personalization
- Amazon Personalize

Speech
- Amazon Polly
- Amazon Transcribe

Development
- Amazon CodeGuru

Forecasting
- Amazon Forecast

Pros and cons

- Fully managed infrastructure
- Built-in optimized algorithms
- Managed spot training

- Hard to maintain consistency
- Requires different toolset for on premises
- Need tooling for workflow management

Evaluation criteria



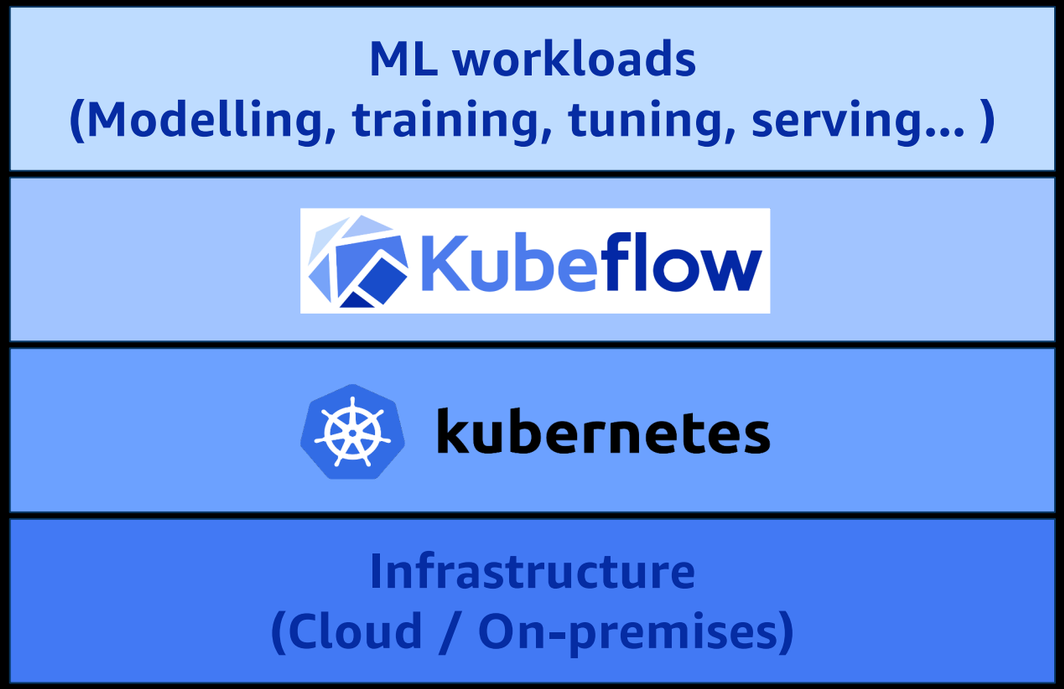
© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Evaluation criteria

- Ability to chain jobs together
- Abstract infrastructure from end users
  - Infrastructure as code
  - Docker for preserving environment
- Deeper integration with AWS platform
  - Put guardrails in place for security
- Repeatable and reproducible environments
- Smaller learning curve for ML engineers
  - Works with python
- Scalable and economical
  - Auto scalable compute
  - Spot instance support

Kubeflow

Kubeflow is *the machine learning toolkit for Kubernetes*



Solution



Kubeflow

+



Amazon Elastic
Kubernetes Service

Kubeflow on EKS

Kubeflow integrations with AWS platform



Amazon CloudWatch logs & metrics



Amazon Elastic Kubernetes Service



Amazon Cognito



AWS Certificate Manager



Amazon SageMaker



AWS Identity and Access Management



Amazon EC2 Auto Scaling





Amazon RDS



Elastic Load Balancing



Amazon FSx



Amazon Elastic File System



Amazon Simple Storage Service



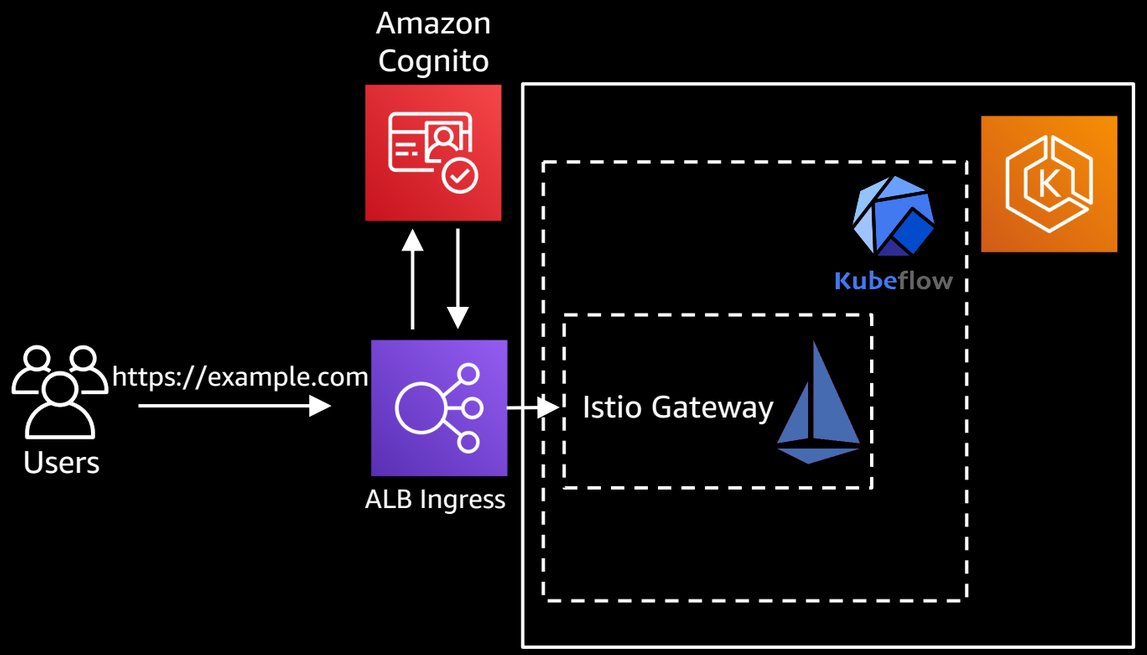
Amazon Athena



Amazon EMR

Kubeflow authentication

- Securely access Kubeflow dashboard using Cognito or basic auth (Dex)
- TLS certificate managed by ACM
- Built-in multi-tenancy for notebooks sharing and pipelines

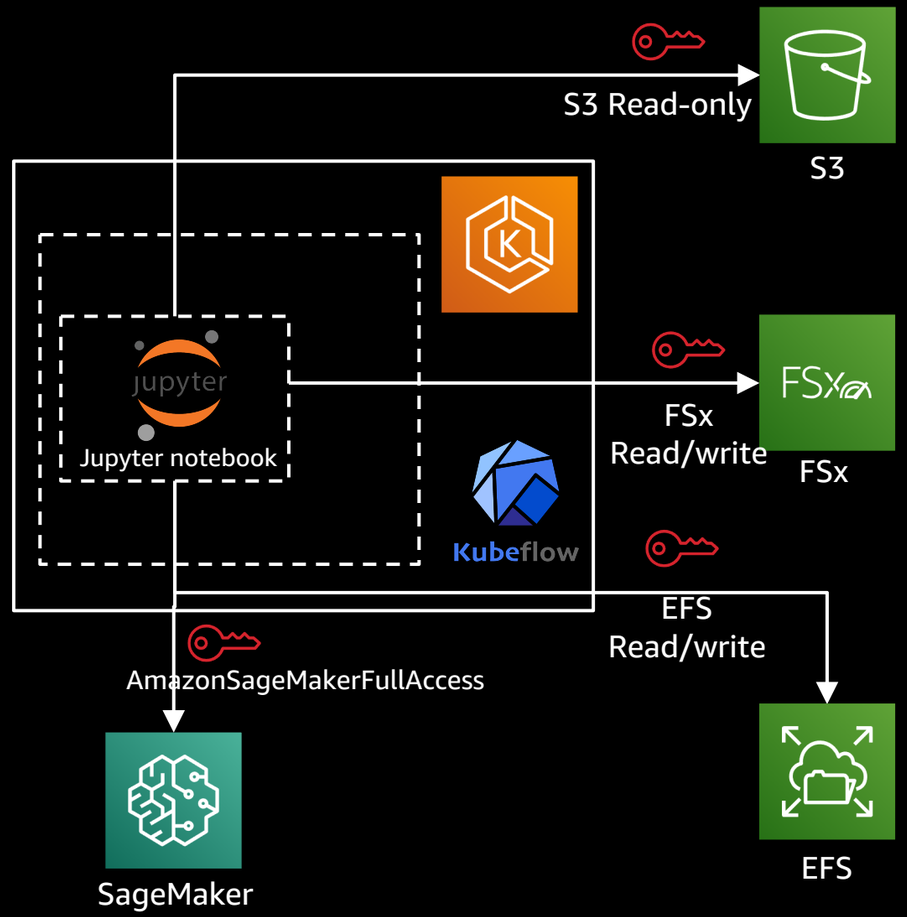


Kubeflow security

- IAM role for service account for fine-grained access control

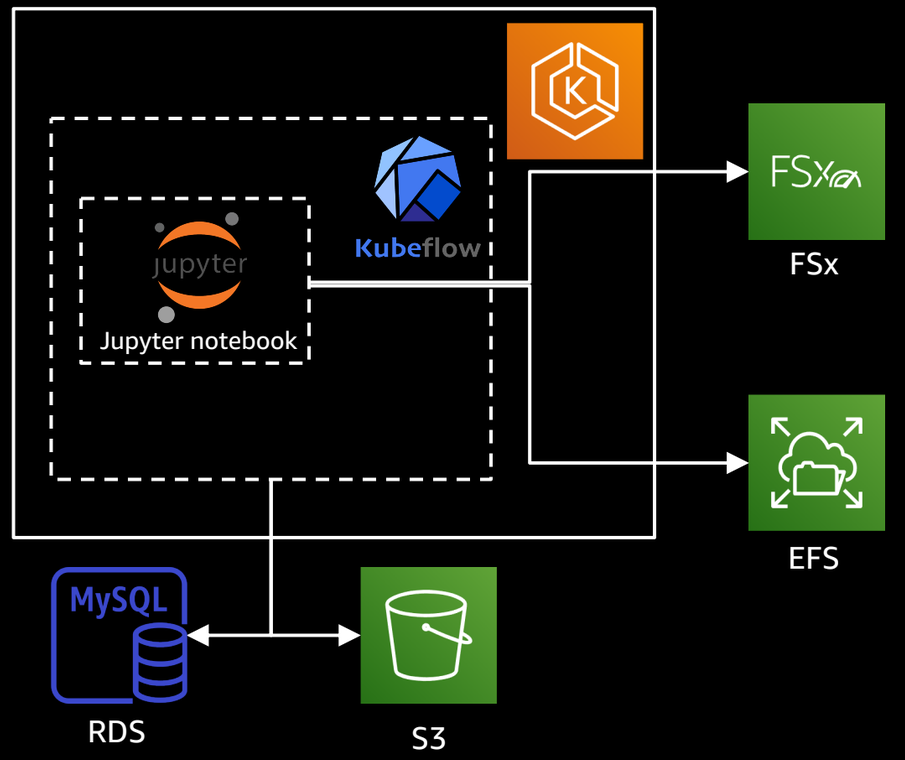
- Streamline IAM controls to AWS resources by assigning permissions to Kubernetes service accounts

```
plugins:
  - kind: KfAwsPlugin
  metadata:
    name: aws
  spec:
    Region: us-west-2
    enablePodIamPolicy: true
```



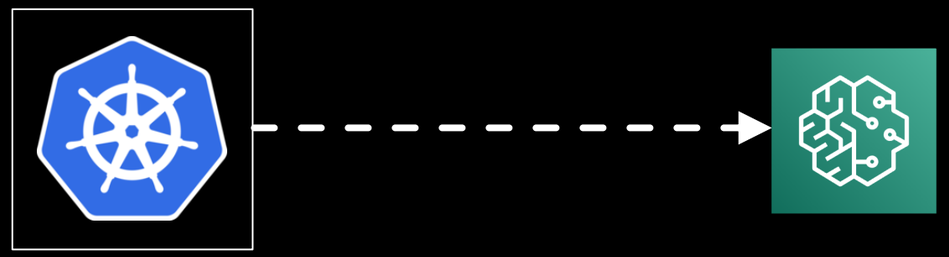
Kubeflow storage options

- Use RDS MySQL to store metadata from pipelines and Katib (experiments, jobs, runs, etc.)
- Use Amazon S3 to store pipeline artifacts (packages, metrics, views, etc.), and to store, process data
- Use Amazon EFS to share Jupyter notebooks, and to store, process data using NFS in your pipeline
- Use Amazon FSx for Lustre to store, process data and as a caching layer for data stored in S3

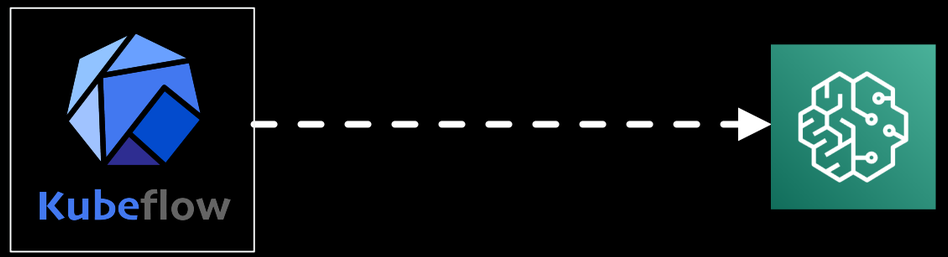


Kubeflow + SageMaker

- Amazon SageMaker operators for Kubernetes



- Amazon SageMaker components for Kubeflow pipelines

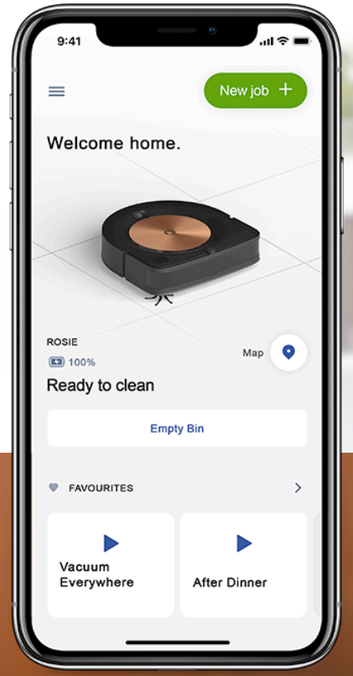


Use cases



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

iRobot Genius™ Home Intelligence



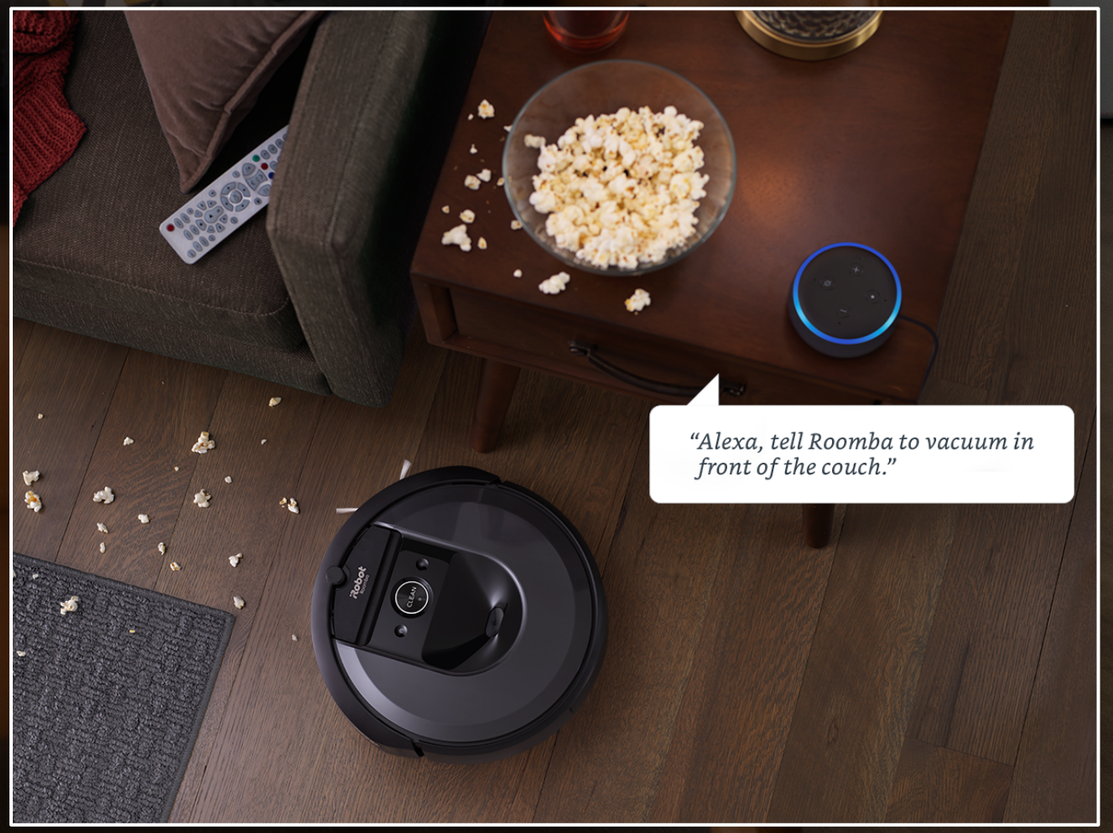
greater control of **where, when** & **how** their robots clean

A level of **personalization and control** for unique

homes
schedules
cleaning preferences
smart home integrations

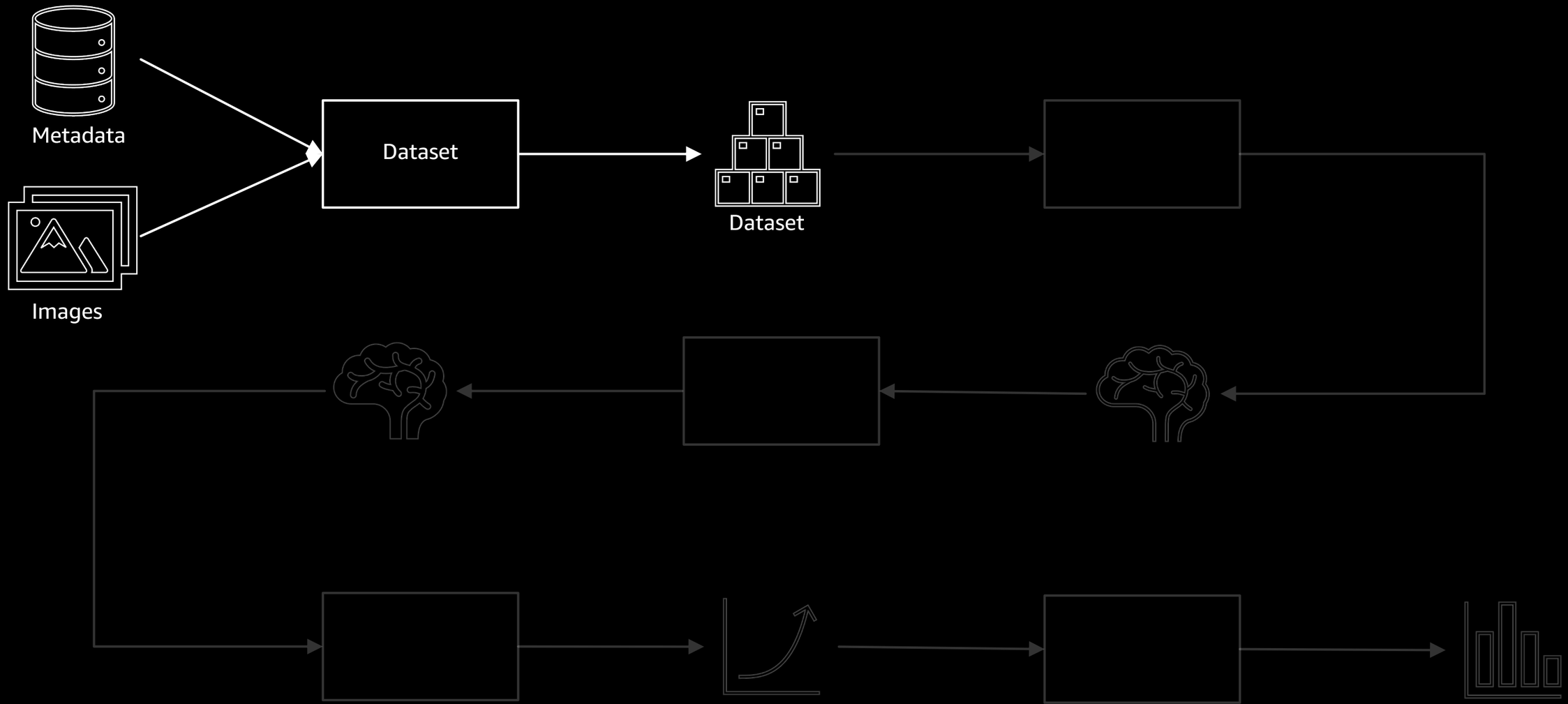

iRobot
Cleaning **where** you need it
Precision **Clean Zones**



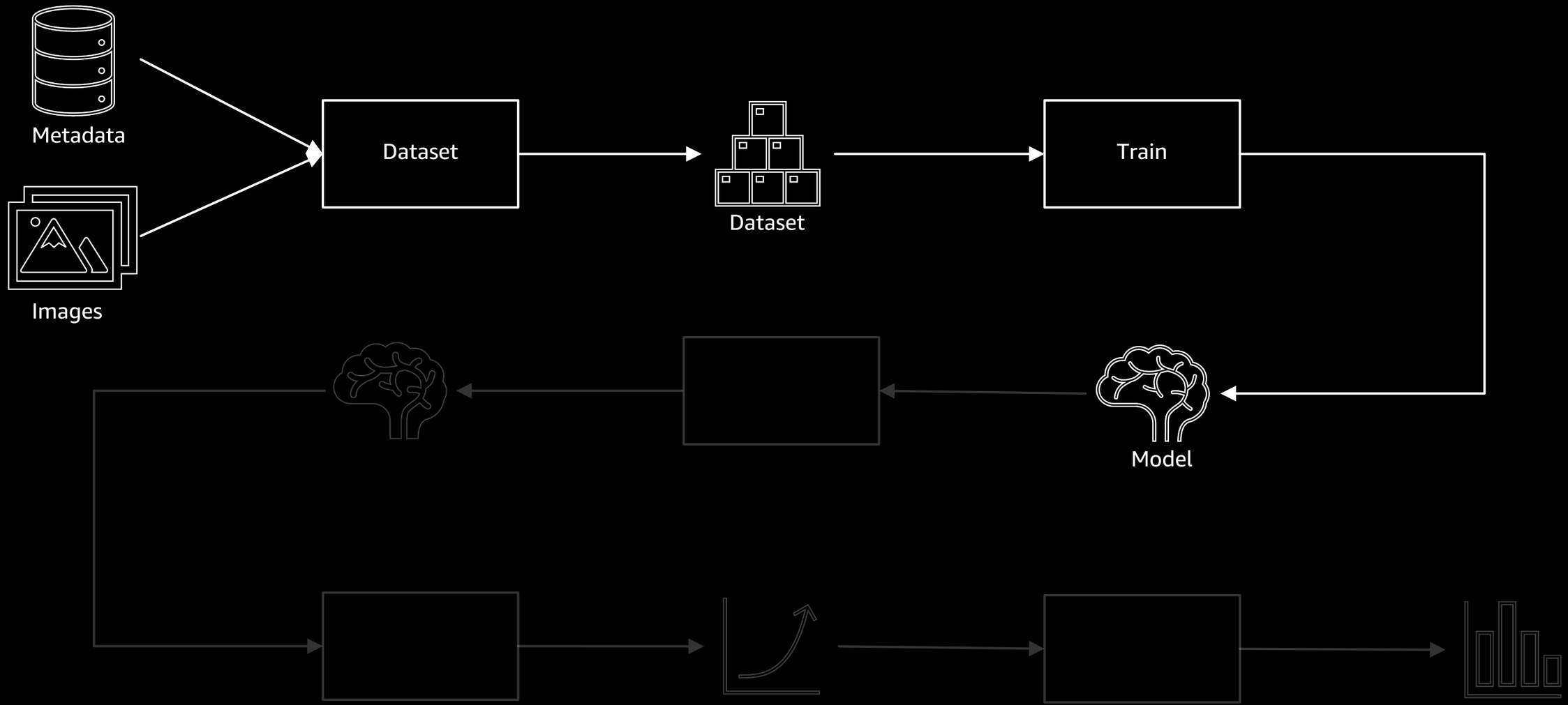


Precision **Clean Zones**

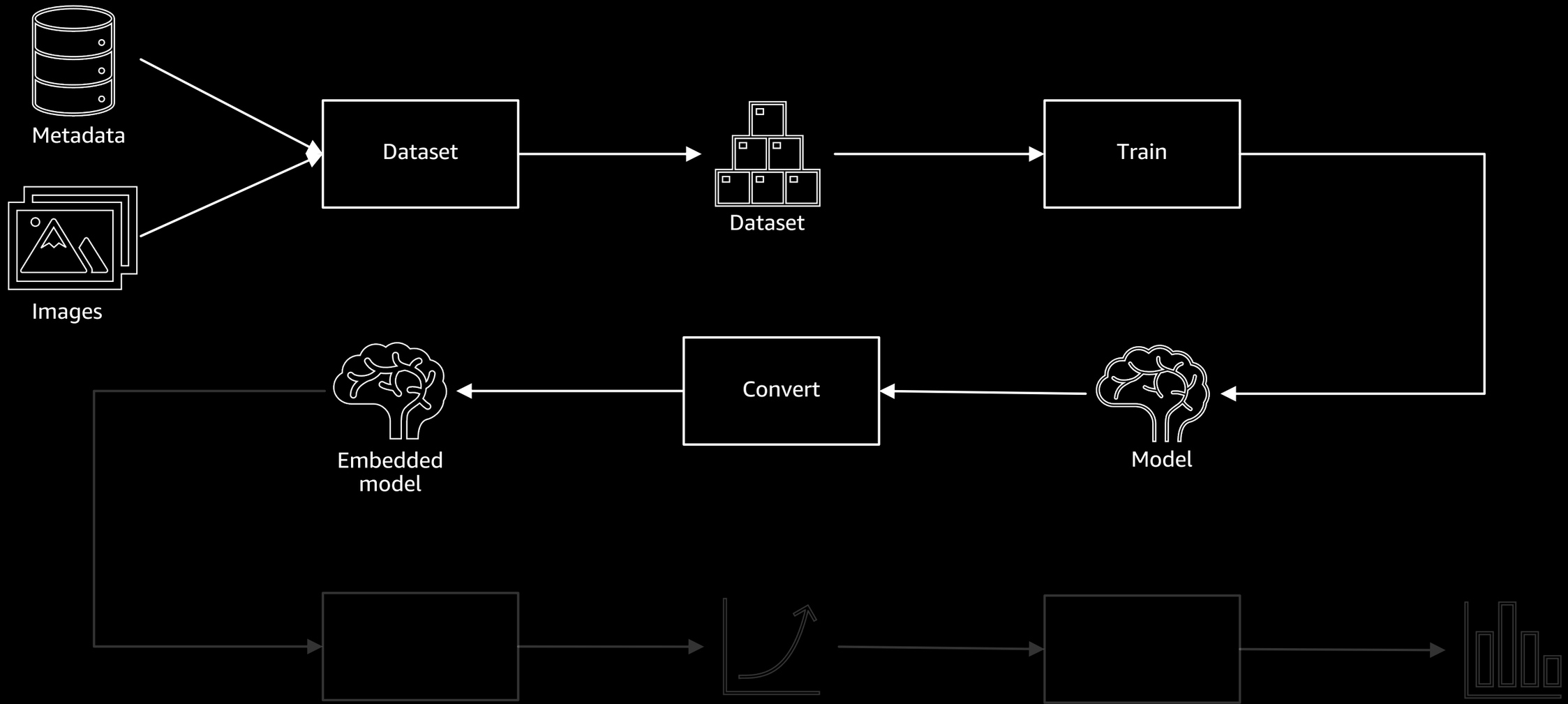
End-to-end model development



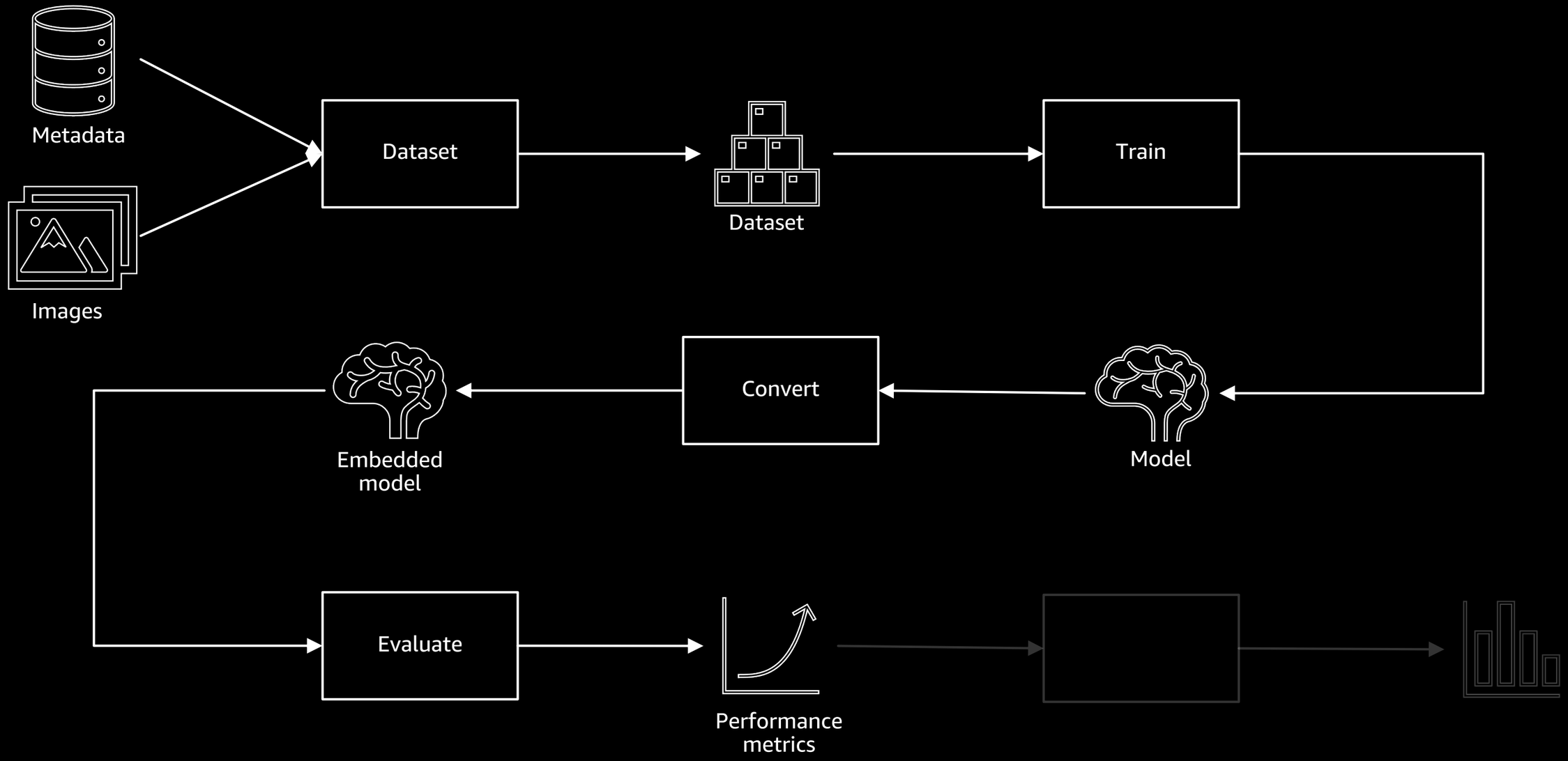
End-to-end model development



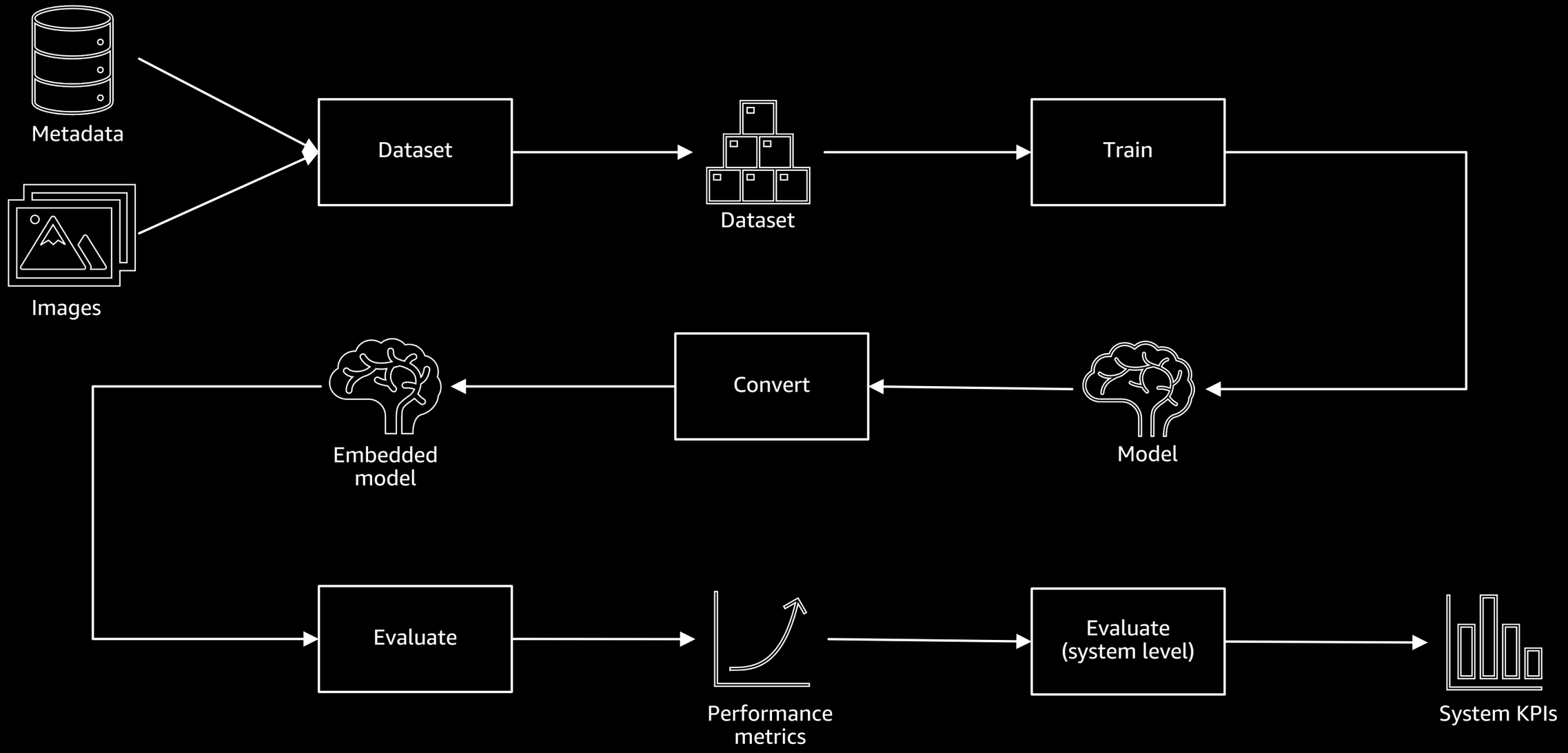
End-to-end model development



End-to-end model development



End-to-end model development



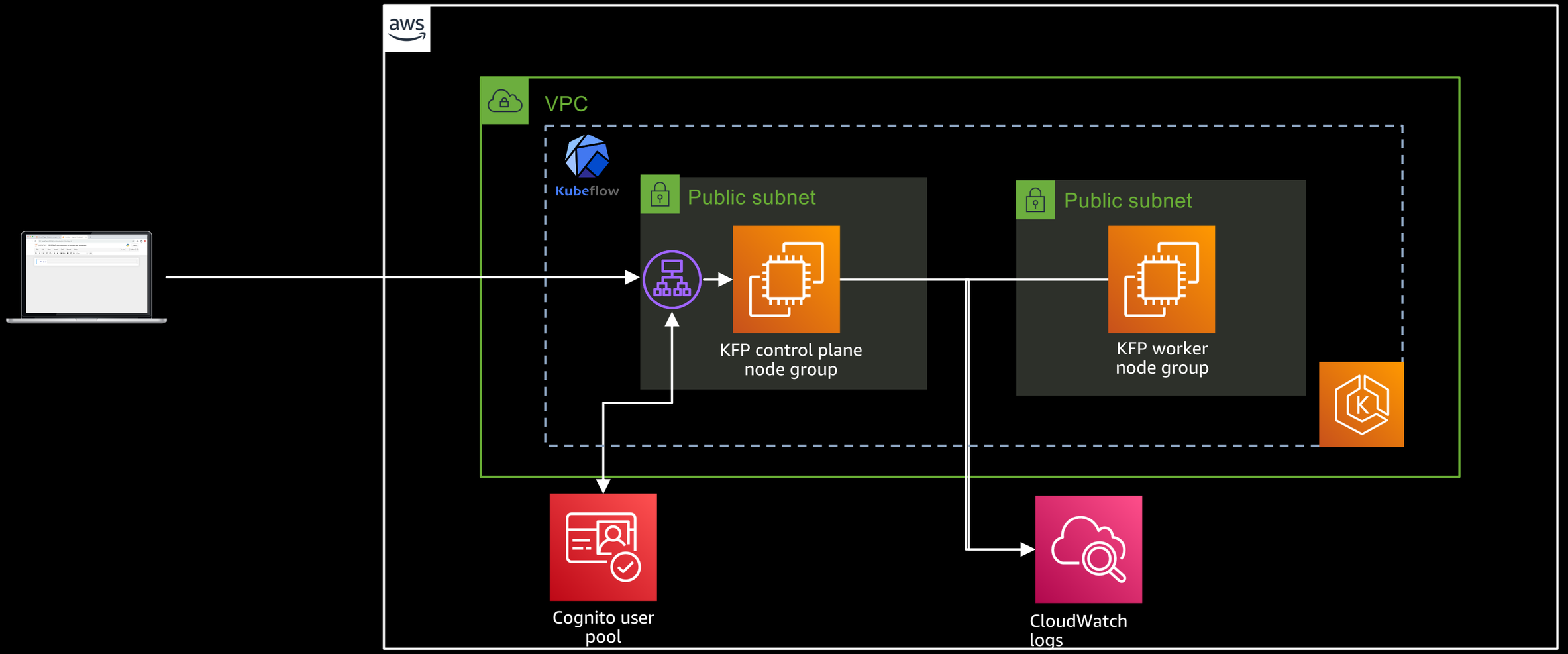
Kubeflow architecture



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Initial deployment

ARCHITECTURE



Initial deployment

- Manual deployment
- Manual upgrades
- Harder to troubleshoot
- Ended up maintaining several versions of Kubeflow (started with Kubeflow 0.5 → Kubeflow 0.7 → Kubeflow 1.0)



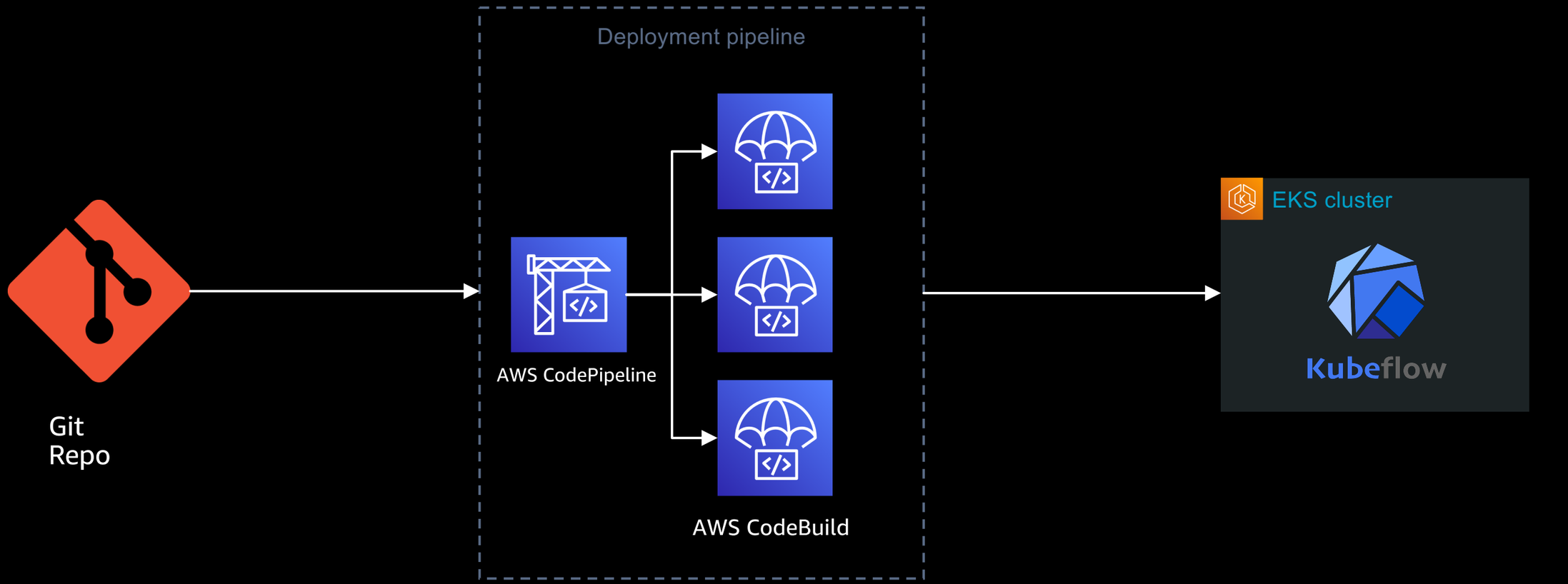
Initial deployment

PAIN POINTS

- Cluster configs were all over the place
- Babysitting deployments that sometimes took hours
- Data loss between deployments / upgrades
- Manual changes to configs, roles, policies, etc., at times
- Basic VPC setup with public subnets, no access to corporate network
- Cognito-based auth, different identities from corporate identities
- Basic monitoring in place

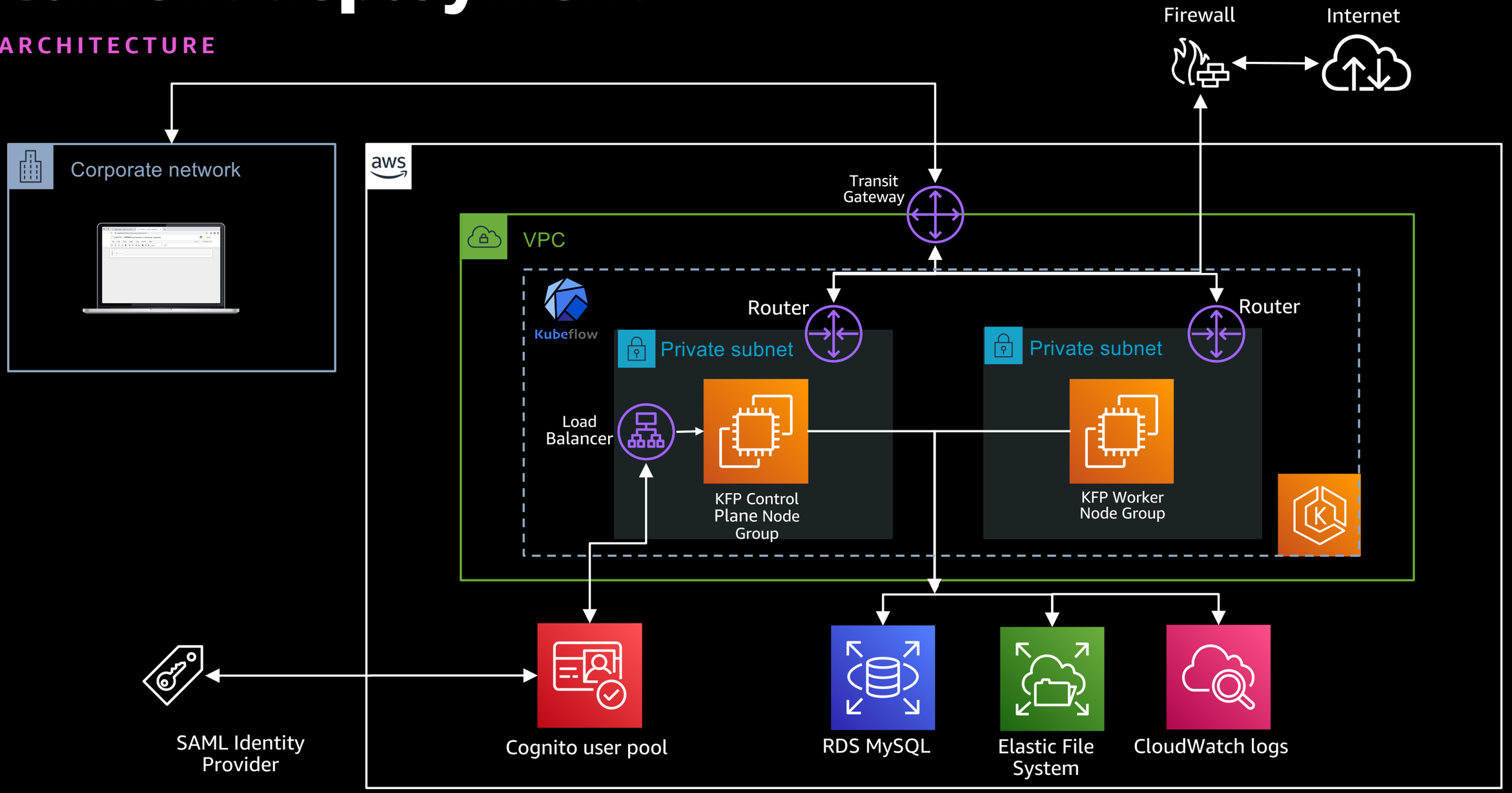
Current deployment

GITOPS



Current deployment

ARCHITECTURE



Current deployment

- Infrastructure as code tracked with Git
- Consistent and repeatable dev / test / prod environments
- SAML-based auth with corporate identities
- VPC with private subnets and connection to corporate network
- Data persisted across deployments / updates
- CD pipeline to take care of deployments
- Ability to revert changes

Best practices and “gotchas”



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Best practices and “gotchas”



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Best practices and “gotchas”



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Best practices and “gotchas”

INFRASTRUCTURE

Gotcha	Best practice
During scale-in operation, control plane pods might prevent node groups from going down	Use taints and tolerations to restrict control plane nodes
KFP component pods running on control plane nodes can take over control plane resources	Use labels to run components on targeted nodes groups (for example GPU / non-GPU)

Best practices and “gotchas”

INFRASTRUCTURE

Gotcha	Best practice
Logs of components run on auto-scalable nodes can vanish with scale-down	Use FluentD to persist the logs in CloudWatch
Streaming all K8s logs / metrics using FluentD can cost you more than your cluster	Use filters on FluentD to stream a subset of logs that are of interest

Best practices

USAGE

- EBS volumes works best for sharing data within pipeline run
- EFS volumes works best for sharing data between pipelines / pipeline runs
- Invest in repository of common components that your team can reuse
- Use SageMaker components to run parts of pipeline on SageMaker services
- Don’t forget KFP SDK transformers, which can be very handy

Resources

Kubeflow on AWS docs: https://www.kubeflow.org/docs/aws/

Getting started EKS workshop tutorials: https://www.eksworkshop.com/advanced/420_kubeflow/

Sample Jupyter notebooks: https://github.com/aws-samples/eks-kubeflow-workshop

Thank you!



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.