AWS re:Invent

AIM410-R1

Deep learning applications using TensorFlow, featuring Fannie Mae

Julien Simon
Global Evangelist AI/ML
Amazon Web Services

Bin Lu
Senior Director of Risk Modeling & Analytics
Fannie Mae

Vindhan Sahayam
Lead Architect
Fannie Mae

Agenda

TensorFlow on AWS

Customer case study: Fannie Mae

Demo: TensorFlow on Amazon SageMaker

Getting started

TensorFlow

https://www.tensorflow.org

- Main API in Python, with support for Javascript, Java, C++
- TensorFlow 1.x: symbolic execution
  - ‘Define then run’: build a graph, optimize it, feed data, and compute
  - Low-level API: variables, placeholders, tensor operations
  - High-level API: *tf.estimator.**
  - Keras library: *Sequential* and *Functional* API, predefined layers
- TensorFlow 2.0: imperative execution (aka eager execution)
  - ‘Define by run’: normal Python code, similar to numpy
  - Run it, inspect it, debug it
  - Keras is the preferred API

AWS: The platform of choice for TensorFlow

https://aws.amazon.com/tensorflow/

89% of all deep learning workloads in the cloud run on AWS

85% of all TensorFlow workloads in the cloud run on AWS

Source: Nucleus Research, T147, October 2019

TensorFlow: a first-class citizen on Amazon SageMaker

- Built-in TensorFlow containers for training and prediction
  - Code available on Github: https://github.com/aws/sagemaker-tensorflow-containers
  - Build it, run it on your own machine, customize it, etc.
  - Versions : 1.4.1 → 1.15 (2.0 coming soon)

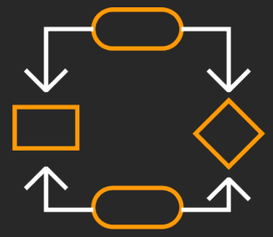
- Not just TensorFlow
  - Standard tools: TensorBoard, TensorFlow Serving
  - SageMaker features: Local Mode, Script Mode, Model Tuning, Spot Training, Pipe Mode, Amazon EFS & Amazon FSx for Lustre, Amazon Elastic Inference, etc.
  - Performance optimizations: GPUs and CPUs (AWS, Intel MKL-DNN library)
  - Distributed training: Parameter Server and Horovod

Amazon SageMaker

re:Invent 2019 announcements

First fully integrated development environment (IDE) for machine learning
SageMaker Studio

Enhanced notebook experience with quick-start & easy collaboration
SageMaker Notebooks (preview)



Automatic debugging, analysis, and alerting
SageMaker Debugger

Experiment management system to organize, track, & compare thousands of experiments
SageMaker Experiments

Model monitoring to detect deviation in quality & take corrective actions
SageMaker Model Monitor

Automatic generation of ML models with full visibility & control
SageMaker Autopilot

Amazon SageMaker at Fannie Mae

Bin Lu

Senior Director of Risk Modeling and Analytics
Fannie Mae

Vindhan Sahayam

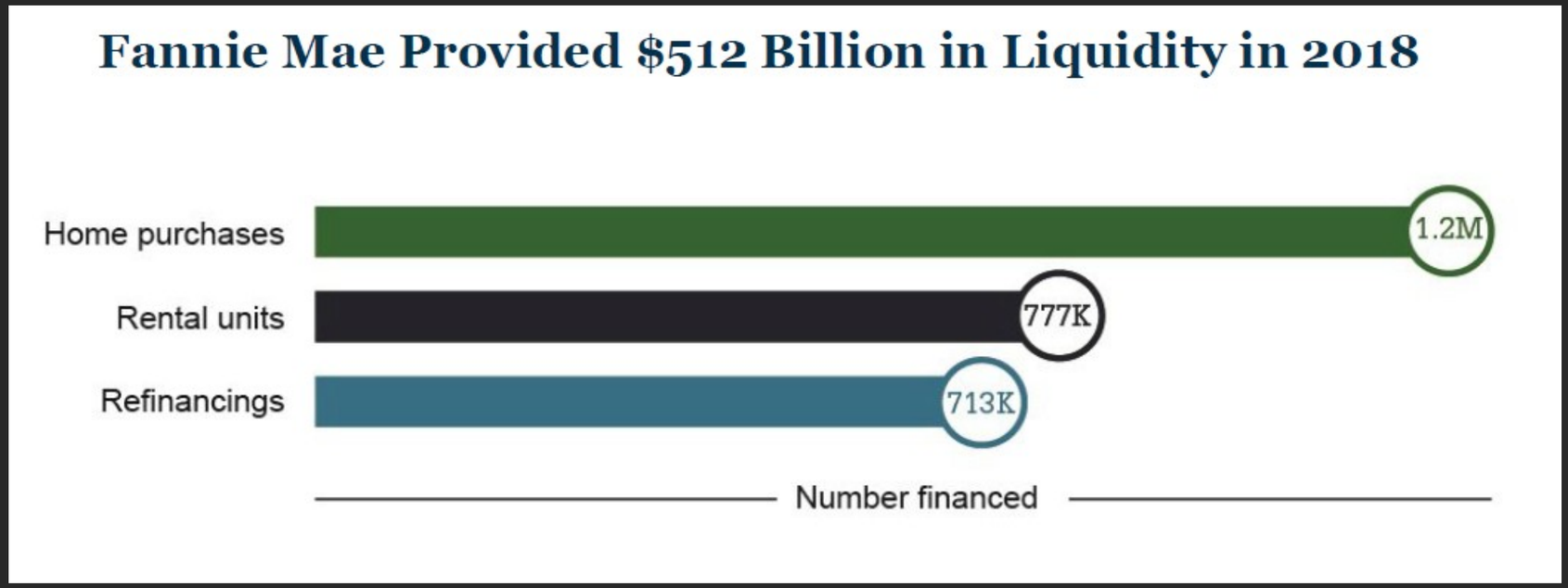
Lead Architect
Fannie Mae

AWS re:Invent

aws

Fannie Mae is a leading source of financing for mortgage lenders

- Provide access to affordable mortgage financing in all markets at all times
- Effectively manage and reduce risk to our business, taxpayers, and the housing finance system



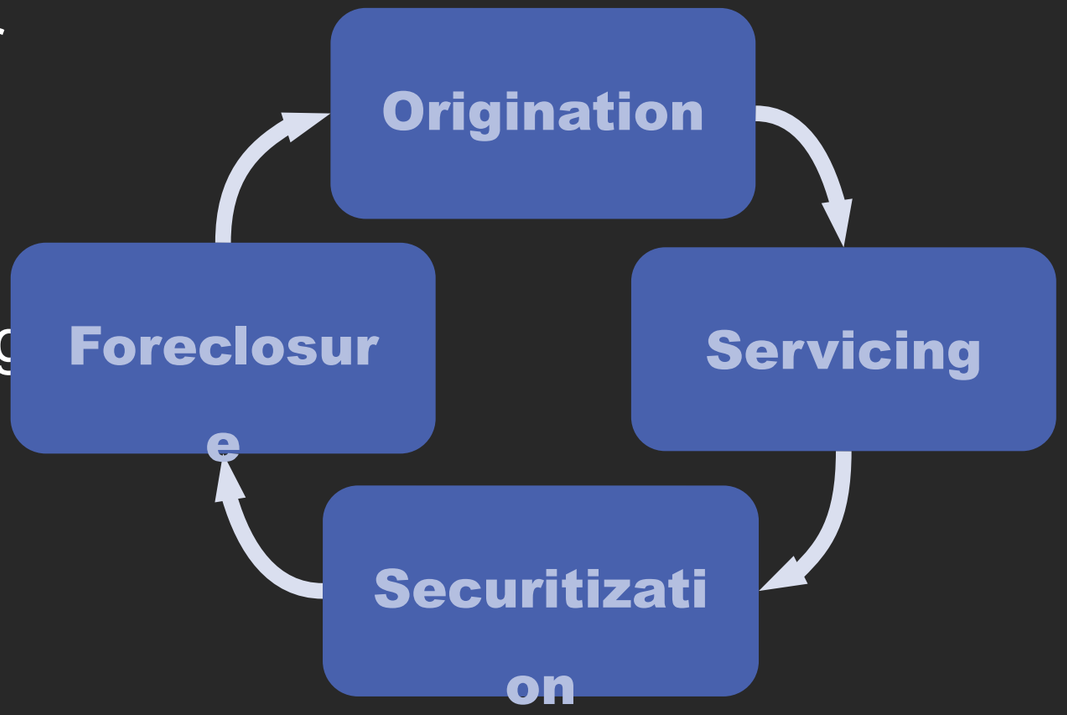
Accurate property valuation reduces mortgage risk

It is used in all stages of the loan lifecycle:

- Origination and underwriting, where a lender determines whether a borrower's loan application is an acceptable risk
- Post-purchase quality control
- Portfolio risk management, financial reporting and regulatory reporting
- Loss mitigation

Fannie Mae credit portfolio is ~$3 trillion

Mortgage lifecycle



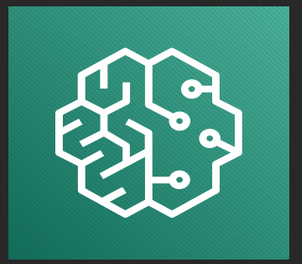
Machine learning example: Property valuation

Property appraisal by certified/licensed appraiser

- Quantitative valuation based on comparable property sale prices and market trends
- Adjustments for unobservable inputs

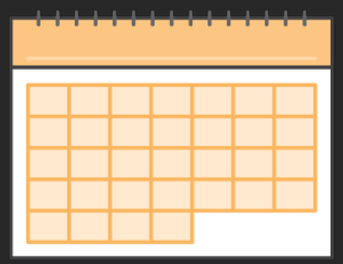
Fannie Mae is leveraging machine learning

- Automated home price valuation model based on observables (XGBoost, KNN)
- Automated review of the adjustment based on visual inspection (TensorFlow – CNN)

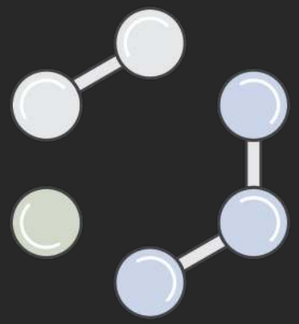


Fannie Mae receives ~40,000 appraisal reports,
with 500,000+ property images every day

Technology challenges in machine learning

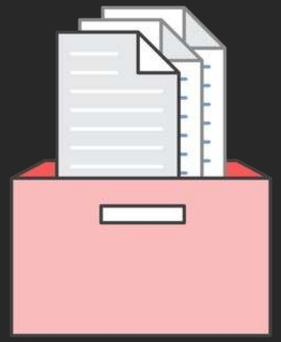


Limited CPU/GPU resources to train and run models



Difficult to connect machine learning and analytics tools to data

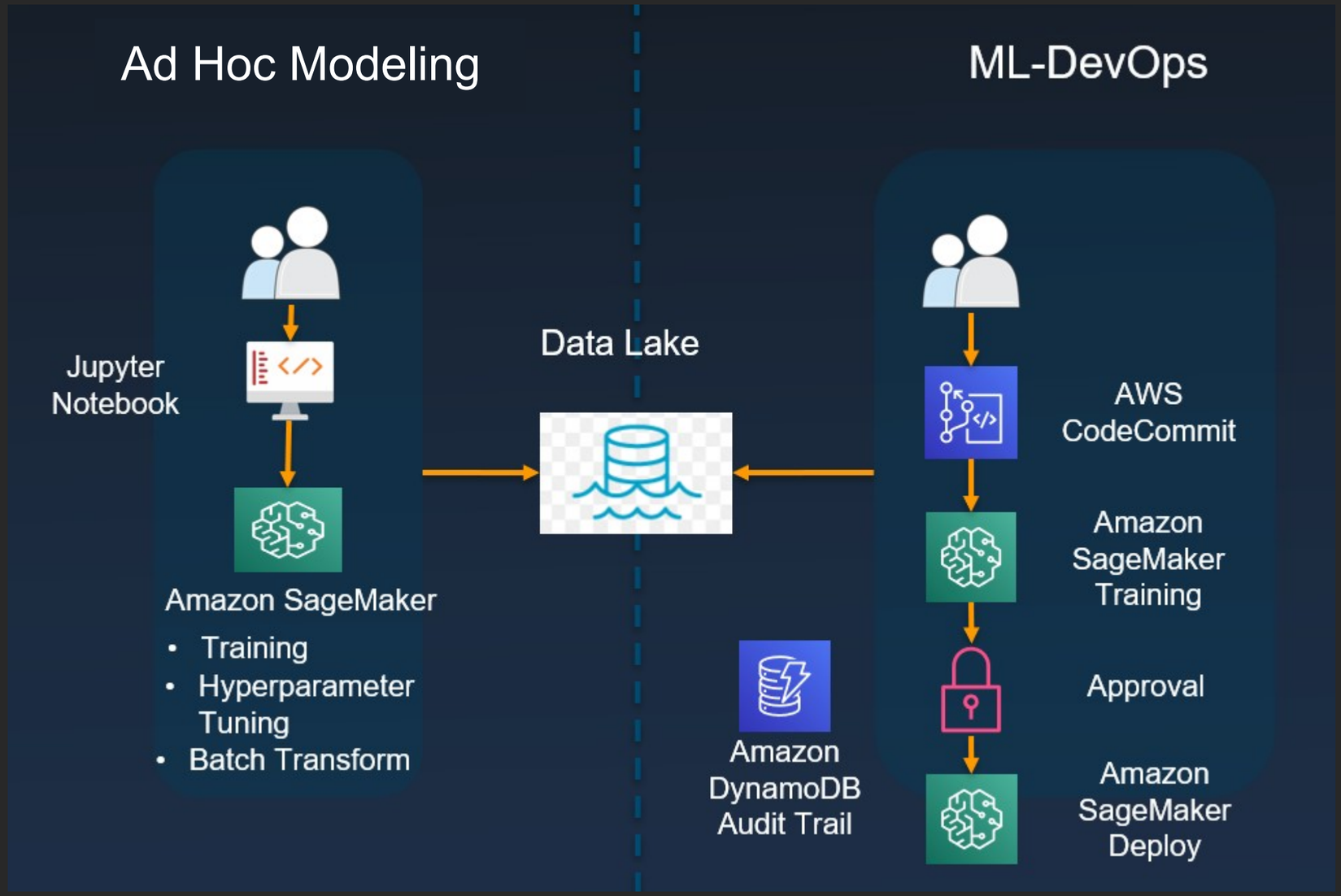
No streamlined approach for model development



Process of packaging and hosting models is complex and time consuming

Amazon SageMaker fits our needs

- Flexible and self-service machine learning platform
- Easy access to compute resources and data
- Streamlined model training and deployment
- Built-in governance procedure and audit trail



Automated property image classification

Three multi-layer convolutional neural network models with transferred learning

1st layer fixes image orientation

```
Predicted_class:  90
Predicted_class Probability:  0.999998
```

2nd layer identifies room type

```
Predicted_class:  kitchen
Predicted_class Probability:  0.995813
```

3rd layer predicts marketabili

```
Predicted_class:  upscaleLuxuryViable
Predicted_class Probability:  0.997592
```

Benefits of Amazon SageMaker

Effective cost management

- Never pay for idle; the cost is based on actual vCPU/GPU usage, not the maximum processing capacity of the infrastructure
- Designed to enable performance improvement at zero cost

Rapid time to market

- Instant access to dedicated computing resources
- Ability to focus on business needs; no server to manage and no complex code to write for distributed model training, hyperparameter tuning, or model deployment

AWS breadth and depth

- Streamlined integration with big data analytics platform
- Automated version controls, governance, audit trails, and secured workload
- Business resiliency

Consideration for provisioning Amazon SageMaker

Implementation of governance is as important as developing business capabilities

- InfoSec risk management
- Data governance
- Model governance
- Technology risk management

Establish guiding principles at the start

- Technology and software
- Models and analytics

Consider data gravity

- Co-locate machine learning platform with data sources

We engaged with the Amazon SageMaker team early

A special shout-out to the Fannie Mae Digital Incubator team for developing the property image classification machine learning model:

Hamid Reza Khakpour, Timur Fatykhov, and Felix Meale

Enabling Amazon SageMaker for Enterprise

Three very important goals

+ Non-negotiable data security
+ Self-service access
+ End-to-end governance with traceability

Realistic to achieve all the above with a fully-managed service such as Amazon SageMaker?

Given these conditions …

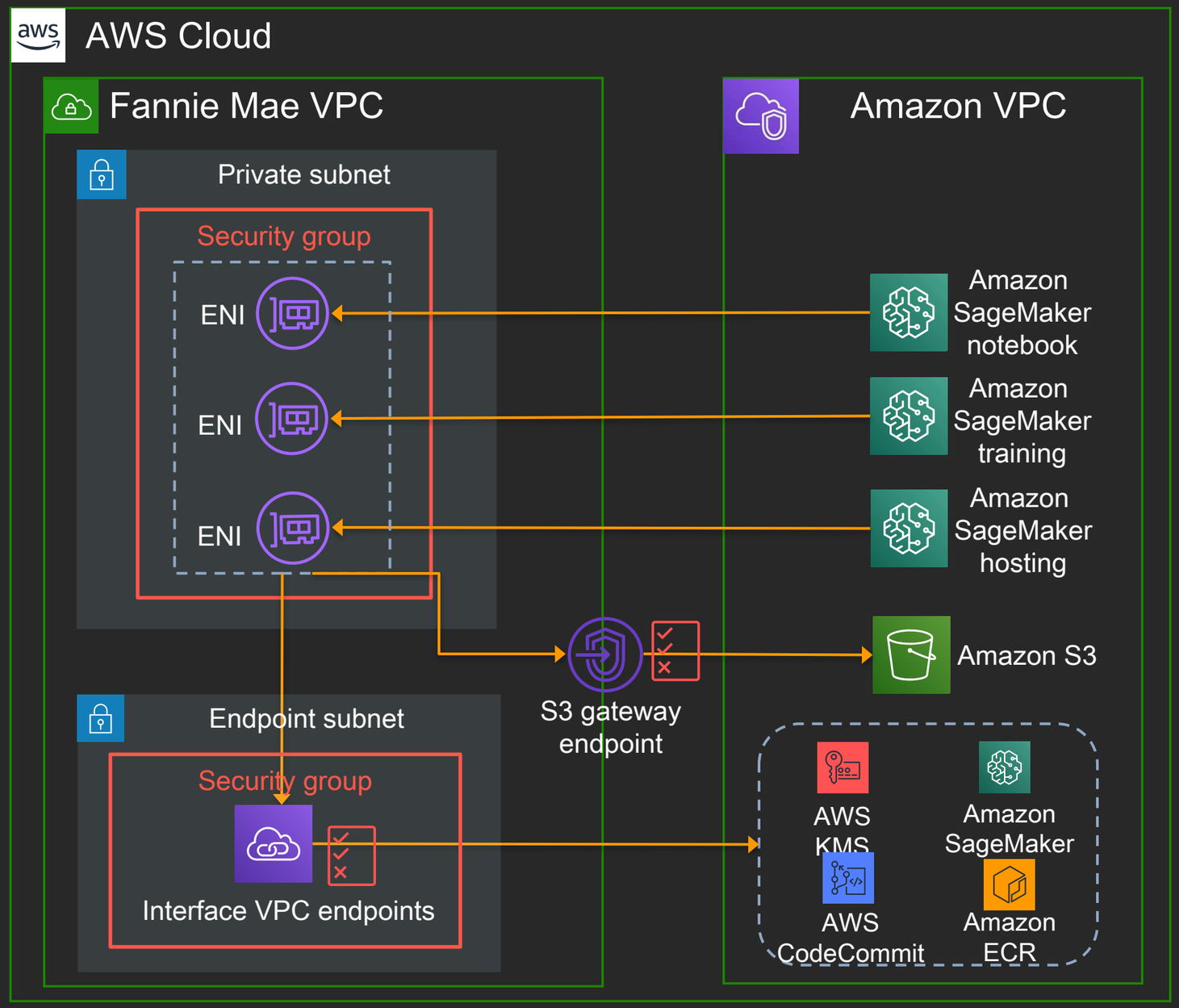
- Amazon SageMaker infrastructure is deployed in AWS-managed, multi-tenant VPCs and subnets
- Data scientists work with highly sensitive data using powerful dev tools

How do we keep data absolutely secure?

Keeping data secure: Harden network security

+ How do we prevent data exfiltration?

+ How do we avoid exposure to internet?



Interface endpoint enforcement: Example

```
{
  "Effect": ”Allow",
  "Action": "sagemaker:CreatePresignedNotebookInstanceUrl",
  "Resource": "*"
  "Principal": "*",
  "Condition": {
    "IpAddress": {
      "aws:VpcSourceIp": [
        "x.x.x.x/a",
        "y.y.y.y/b"
      ]
    }
  }
}
```

Interface endpoint policy

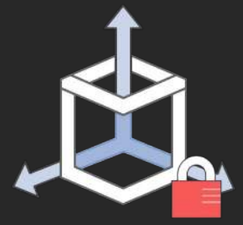
Identity policy

```
{
  "Effect": "Allow",
  "Action": "sagemaker:CreatePresignedNotebookInstanceUrl",
  "Resource": "*",
  "Condition": {
    "StringEquals": {
      "aws:SourceVpce": "vpce-x"
    }
  }
}
```

Keeping data secure: Encrypt everywhere

Volumes
encryption

Bucket encryption
&
deny policies



Inter-container
traffic
encryption

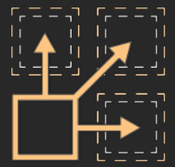
Enable Amazon S3 default encryption. Additionally use deny policies to prevent unencrypted uploads

Use customer managed CMK for volumes and S3 encryption

With the greater flexibility of self-service access …

- How do we ensure users comply with security controls?
- How do we ensure users do not step into each other?

Amazon SageMaker
IAM context
keys



AWS Service
Catalog
provisioning

Tagging
& resource
grouping

Access controls enforcement: Examples

New

```
{
  "Effect": "Allow",
  "Action": "sagemaker:CreateTrainingJob" ,
  "Resource": "arn:aws:sagemaker:x:x:*/app1-*",
  "Condition": {
    "StringEquals": {
      "aws:RequestTag/CostCenter": "x",
      "sagemaker:VolumeKmsKey": "arn:aws:kms:x:x:key/x"
    },
    "Bool": {
      "sagemaker:InterContainerTrafficEncryption": "true",
      "sagemaker:NetworkIsolation": "true"
    },
    "ForAllValues:StringEquals": {
      "sagemaker:VpcSubnets": [
        "subnet-a",
        "subnet-b"
      ],
      "sagemaker:VpcSecurityGroupIds": [
        "sg-x",
        "sg-y"
      ]
    },
    "Null": {
      "sagemaker:VpcSubnets": "false"
    }
  }
}
```

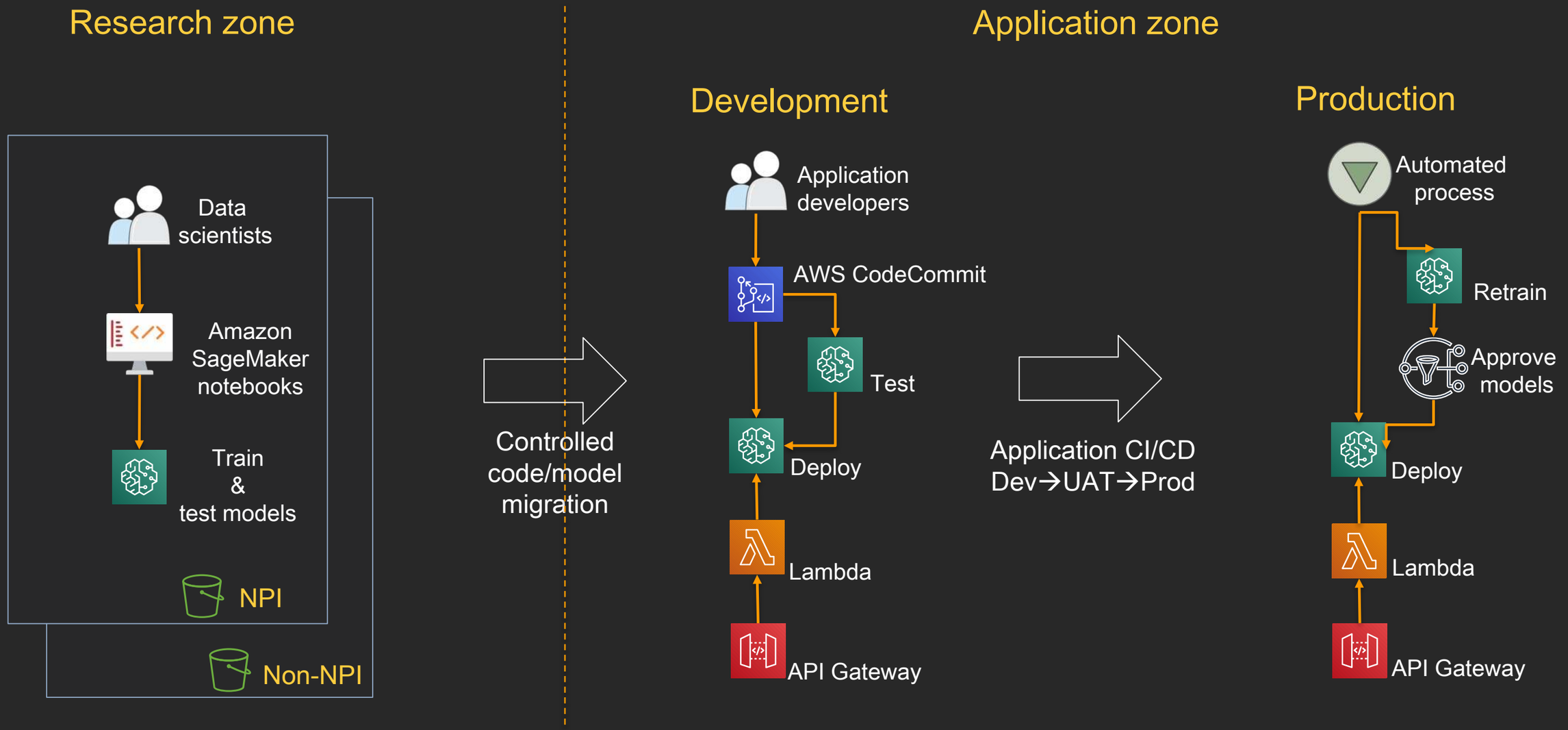
Network and encryption enforcement

Notebook access control

```
{
  "Effect": "Deny",
  "Action": "sagemaker:CreatePresignedNotebookInstanceUrl",
  "Resource": "*",
  "Condition": {
    "ForAllValues:StringNotEquals": {
      "sagemaker:ResourceTag/creatorUserId": "${aws:userId}"
    }
  }
}
```

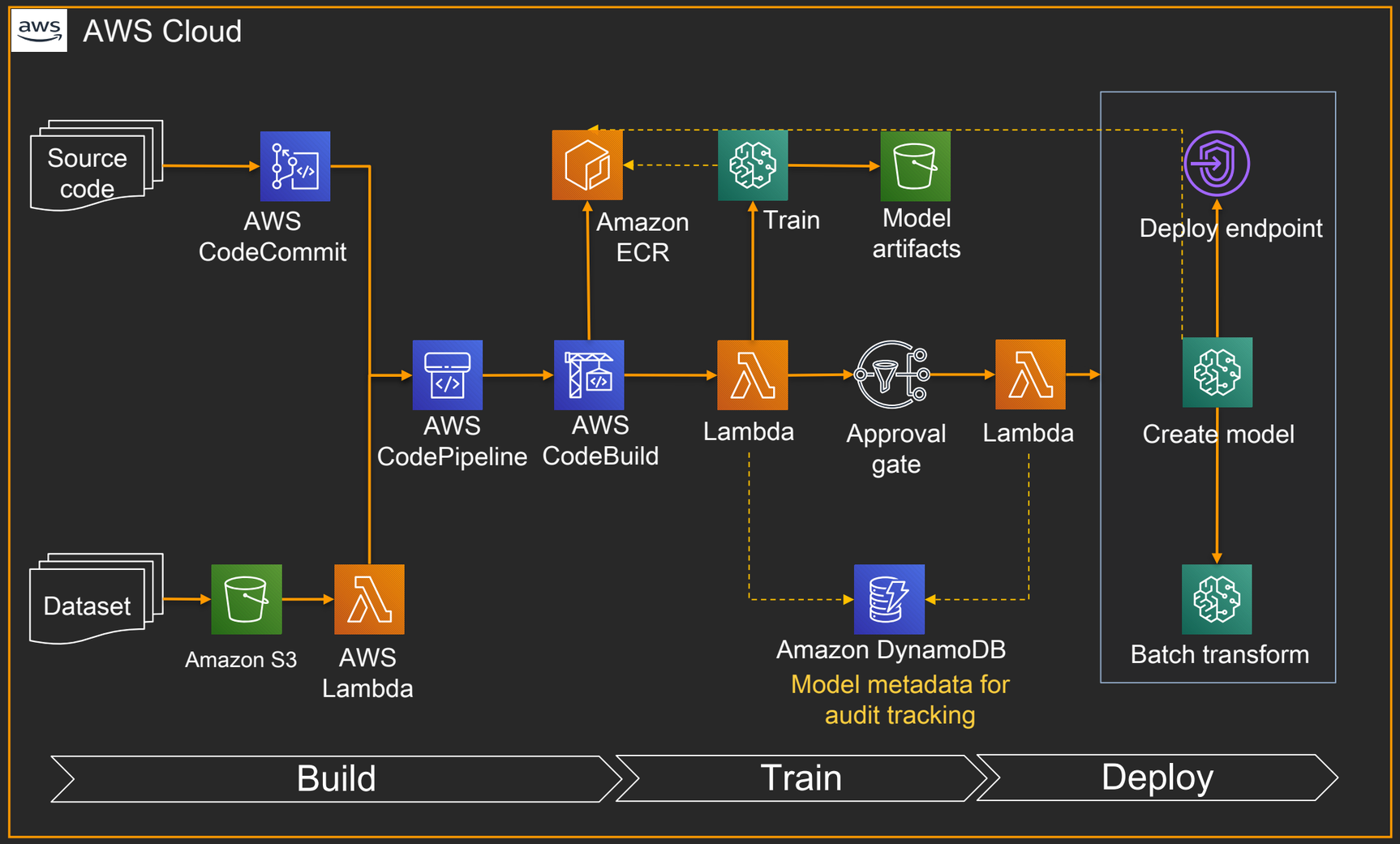
Enabling governance: Operating zones

Create guardrails early: Establish zones to manage ML lifecycle



Machine learning orchestration with auditing: Example

+ Reproducible and reusable pipeline
+ Built-in audit tracking capability
+ Other options: AWS Step Functions, Apache Airflow



Fannie Mae's Enterprise Data Lake (EDL) at a glance

3,000+ datasets

1,000+ users

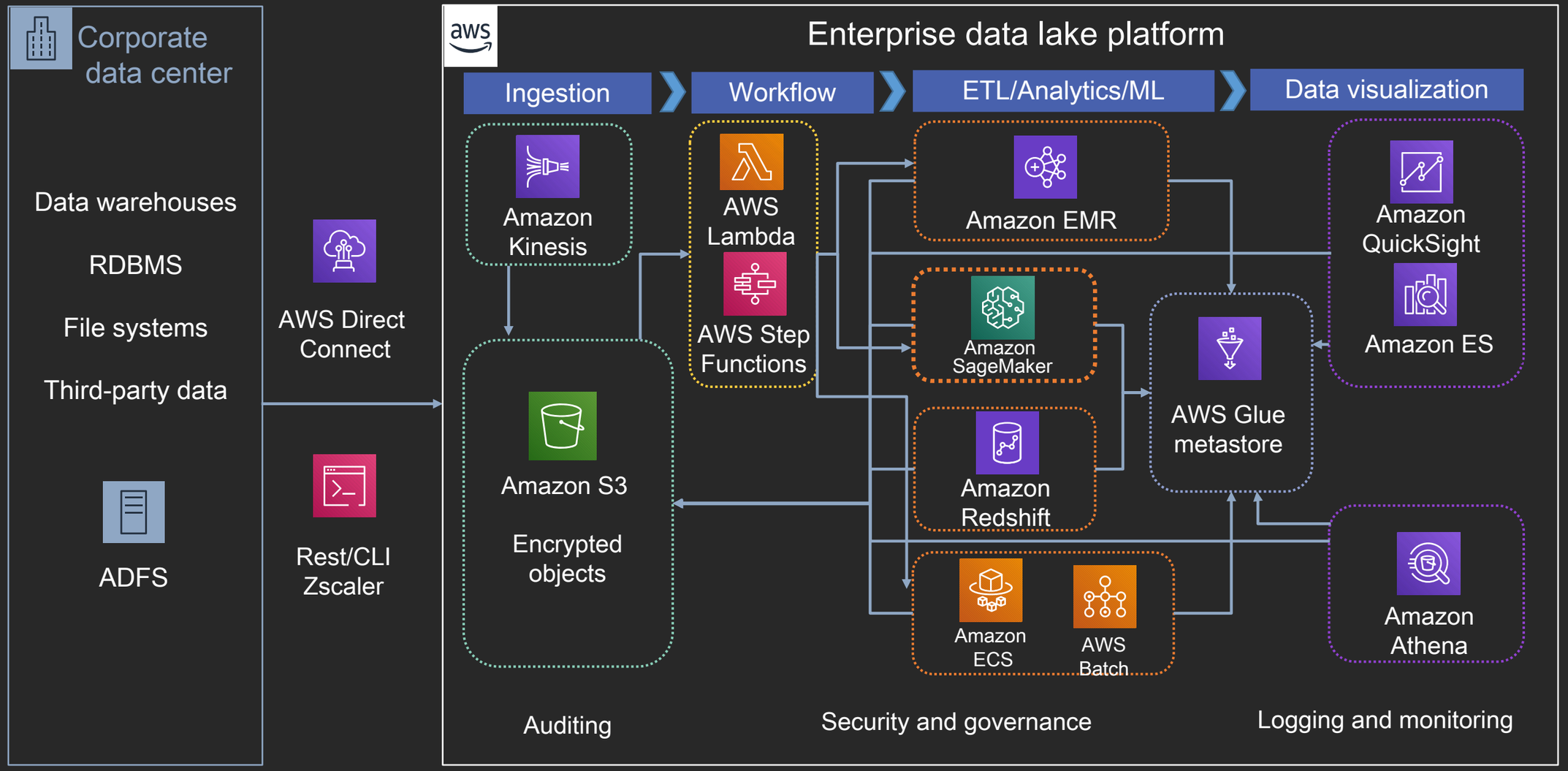
500+ AWS Glue Data Catalog databases

100+ applications

... and growing

Build machine learning capability with a fully functional data lake as a foundation

Amazon SageMaker in EDL: Reference architecture



Platform built with 100% native AWS services => less integration challenges

Key takeaways

- New IAM context keys are valuable
- Restrict access to buckets, utilize S3 endpoint policy
- Amazon SageMaker has full support for PrivateLink endpoints; Enabling and enforcing those is crucial
- Data is a first-class primitive in ML workflows; keep track of data collection and preparation
- Make predictions traceable to original training record
- Introduce segregation of duties; establish operating zones
- Leverage data lake pattern

Build a highly-secure, self-service & end-to-end traceable ML capability with Amazon SageMaker

Amazon SageMaker at Fannie Mae

Bin Lu

Senior Director of Risk Modeling and Analytics
Fannie Mae

Vindhan Sahayam

Lead Architect
Fannie Mae

AWS re:Invent

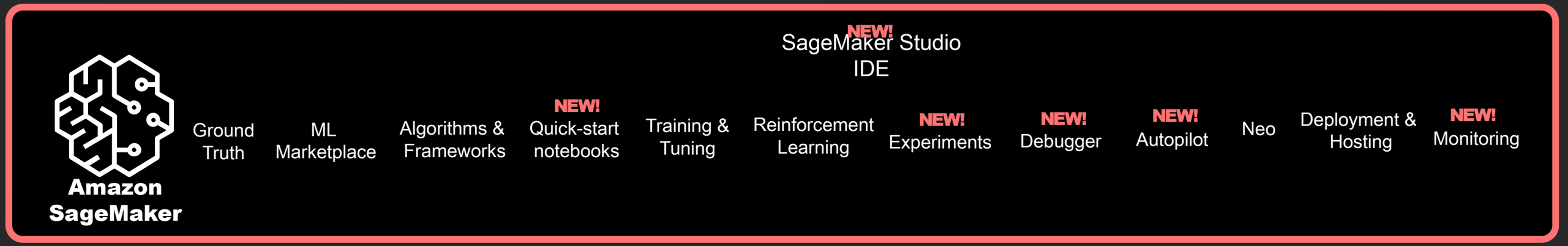
aws

Demo

+ Script Mode
+ Managed Spot Training
+ Elastic Inference

Amazon SageMaker

Build, train, deploy machine learning models quickly at scale



Getting started

http://aws.amazon.com/free

https://aws.amazon.com/tensorflow/

https://aws.amazon.com/sagemaker

https://github.com/aws/sagemaker-python-sdk

https://sagemaker.readthedocs.io/en/stable/using_tf.html

https://github.com/awslabs/amazon-sagemaker-examples

https://gitlab.com/juliensimon/dlnotebooks

Thank you!

Julien Simon
Global Evangelist AI/ML
Amazon Web Services

Bin Lu
Senior Director of Risk Modeling & Analytics
Fannie Mae

Vindhan Sahayam
Lead Architect
Fannie Mae

AWS re:Invent

aws

Please complete the session survey in the mobile app.