



AWS
re:Invent

AIM 368

Experience the real-world ML lifecycle with Amazon SageMaker

Rick Nayar

Solutions Architect
Amazon Web Services

Mike Calder

Software Engineer
Amazon Robotics

Today's Agenda

- Introduction to Amazon SageMaker
- Amazon Robotics Machine Learning Use Case
- Build a Real-World ML Solution via Hands-On Labs

Our mission at AWS

Put machine learning in the
hands of every developer


THE AWS ML STACK

Broadest and deepest set of capabilities















AI Services

VISION			SPEECH		LANGUAGE		CHATBOTS	FORECASTING	RECOMMENDATIONS
 AMAZON REKOGNITION IMAGE	 AMAZON REKOGNITION VIDEO	 AMAZON TEXTRACT	 AMAZON POLLY	 AMAZON TRANSCRIBE	 AMAZON TRANSLATE	 AMAZON COMPREHEND & AMAZON COMPREHEND MEDICAL	 AMAZON LEX	 AMAZON FORECAST	 AMAZON PERSONALIZE

ML Services

 Amazon SageMaker	Ground Truth	Notebooks	Algorithms + Marketplace	Reinforcement Learning	Training	Optimization	Deployment	Hosting










ML Frameworks + Infrastructure

FRAMEWORKS	INTERFACES	INFRASTRUCTURE								
 TensorFlow  mxnet  PYTORCH	 GLUON  K Keras	 EC2 P3 & P3DN	 EC2 G4 EC2 C5	 FPGAS	 DL CONTAINERS & AMIs	 AMAZON ELASTIC CONTAINER SERVICE	 AMAZON ELASTIC KUBERNETES SERVICE	 AWS IoT GREENGRASS	 AMAZON ELASTIC INFERENCE	 INFERENCE


THE AWS ML STACK

Broadest and deepest set of capabilities















AI Services

VISION			SPEECH		LANGUAGE		CHATBOTS	FORECASTING	RECOMMENDATIONS
 AMAZON REKOGNITION IMAGE	 AMAZON REKOGNITION VIDEO	 AMAZON TEXTRACT	 AMAZON POLLY	 AMAZON TRANSCRIBE	 AMAZON TRANSLATE	 AMAZON COMPREHEND & AMAZON COMPREHEND MEDICAL	 AMAZON LEX	 AMAZON FORECAST	 AMAZON PERSONALIZE

ML Services

 Amazon SageMaker	Ground Truth	Notebooks	Algorithms + Marketplace	Reinforcement Learning	Training	Optimization	Deployment	Hosting
---	--------------	-----------	--------------------------	------------------------	----------	--------------	------------	---------











ML Frameworks + Infrastructure

FRAMEWORKS	INTERFACES	INFRASTRUCTURE								
 TensorFlow  mxnet  PYTORCH	 GLUON  K Keras	 EC2 P3 & P3DN	 EC2 G4 EC2 C5	 FPGA	 DL CONTAINERS & AMIs	 ELASTIC CONTAINER SERVICE	 ELASTIC KUBERNETES SERVICE	 GREENGRASS	 ELASTIC INFERENCE	 INFENTIA


THE AWS ML STACK

Broadest and deepest set of capabilities















AI Services

VISION			SPEECH		LANGUAGE		CHATBOTS	FORECASTING	RECOMMENDATIONS
 REKOGNITION IMAGE	 REKOGNITION VIDEO	 TEXTRACT	 POLLY	 TRANSCRIBE	 TRANSLATE	 COMPREHEND & COMPREHEND MEDICAL	 LEX	 FORECAST	 PERSONALIZE

ML Services

 Amazon SageMaker	Ground Truth	Notebooks	Algorithms + Marketplace	Reinforcement Learning	Training	Optimization	Deployment	Hosting
---	--------------	-----------	--------------------------	------------------------	----------	--------------	------------	---------

ML Frameworks + Infrastructure

FRAMEWORKS	INTERFACES	INFRASTRUCTURE								
 TensorFlow  mxnet  PYTORCH	 GLUON  Keras	 EC2 P3 & P3DN	 EC2 G4 EC2 C5	 FPGAS	 DL CONTAINERS & AMIs	 ELASTIC CONTAINER SERVICE	 ELASTIC KUBERNETES SERVICE	 GREENGRASS	 ELASTIC INFERENCE	 INFERENCE

Amazon Robotics

Amazon Robotics



Amazon Robotics



Amazon Robotics



Amazon Robotics

- Warehouse associates interact with hundreds of robots daily
- Our cloud computing coordinates hundreds of thousands of robots



Machine Learning Lifecycle

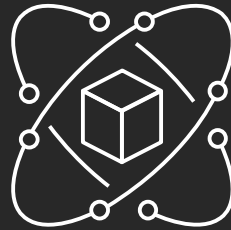
Machine Learning Lifecycle



Feature
Selection



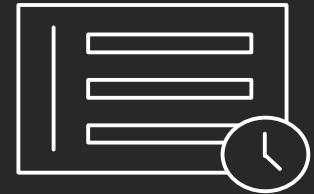
Sample
Collection



Model
Training



Live
Inference



Automated
Retraining

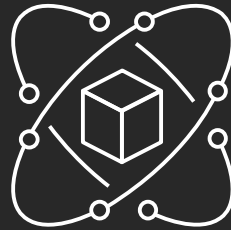
Machine Learning Lifecycle



Feature
Selection



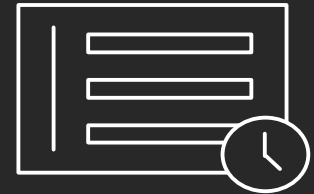
Sample
Collection



Model
Training



Live
Inference



Automated
Retraining

Problem Definition

- Robots bring pods full of inventory to warehouse associates for ICQA
- The associates verify the quantity of a specific product in a specific bin
- How long will each of these human-robot interactions take?





amazon
fulfillment




amazonfulfillment

Machine Learning Features

- Average Duration
- Product Dimensions
- Product Quantity
- Product Weight
- Bin Location
- Bin Dimensions
- Bin Fullness
- Bin Clutter

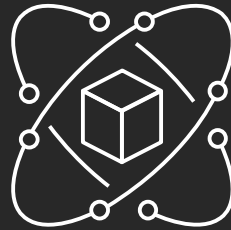
Machine Learning Lifecycle



Feature
Selection



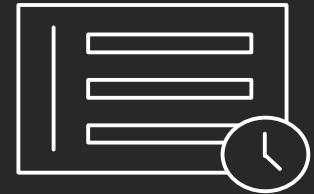
Sample
Collection



Model
Training

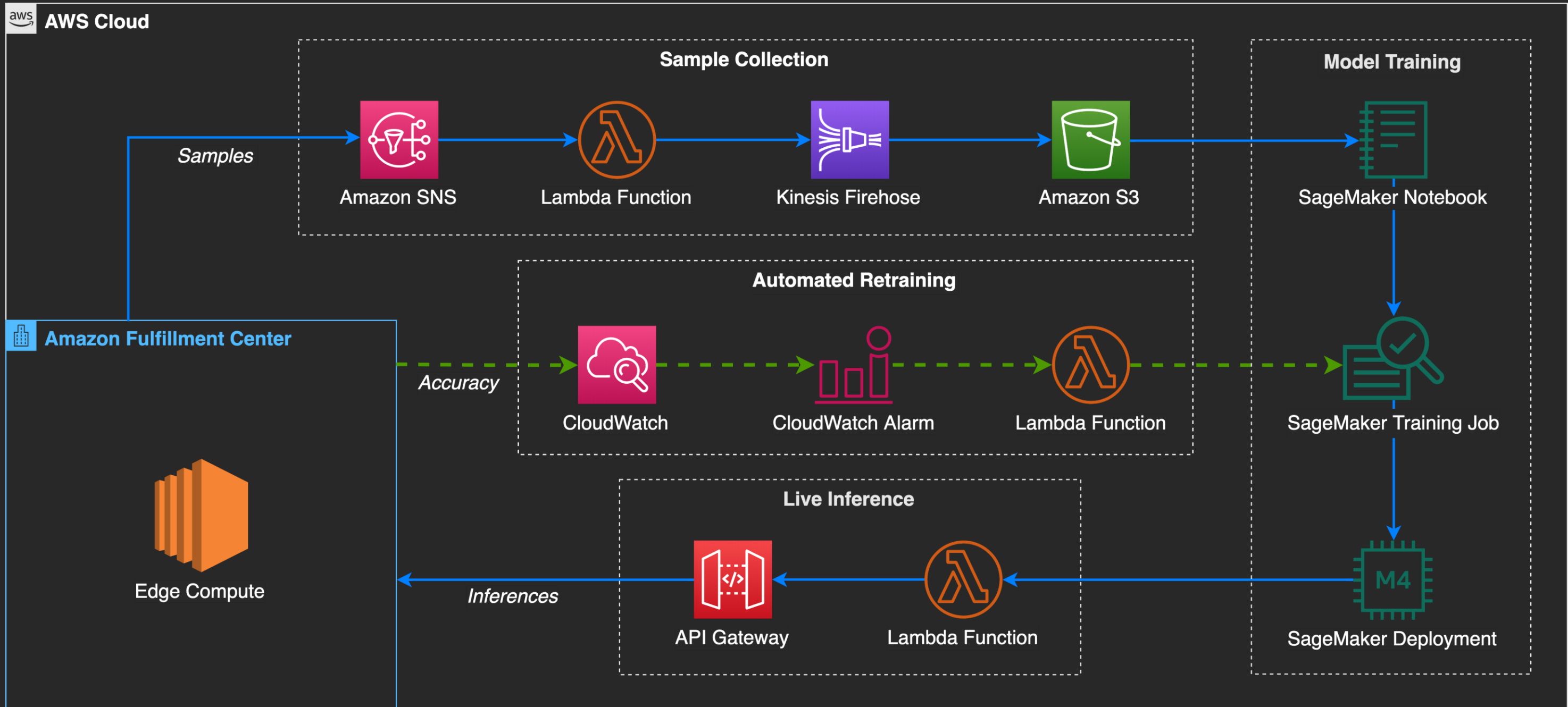


Live
Inference



Automated
Retraining

End-to-End Architecture



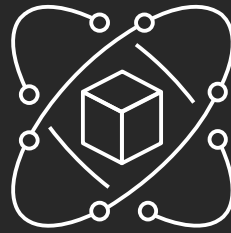
Machine Learning Lifecycle



Feature
Selection



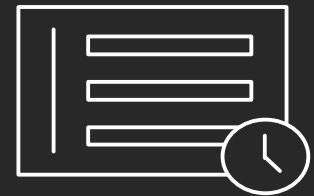
Sample
Collection



Model
Training

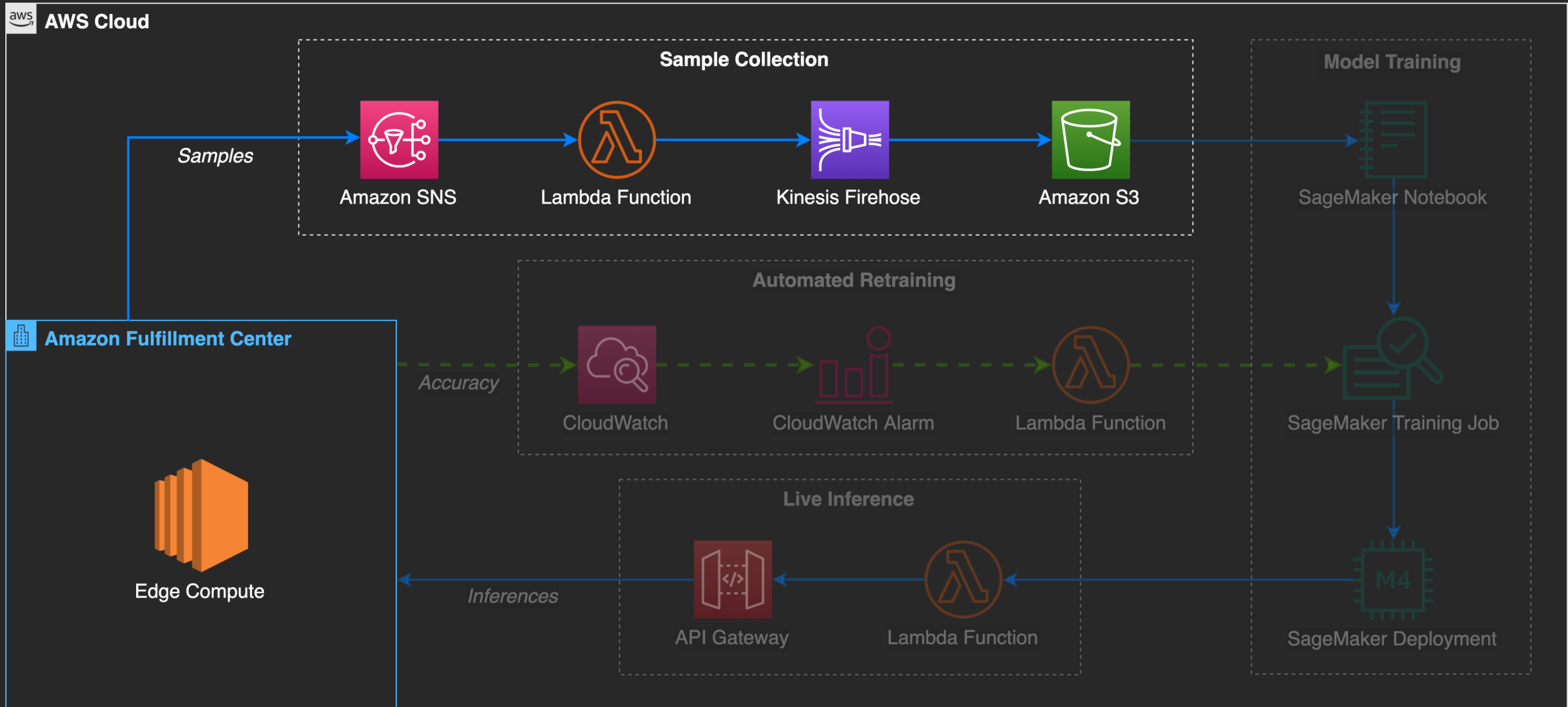


Live
Inference



Automated
Retraining

Sample Collection



AWS Technologies



SNS – publisher/subscriber message service



Lambda – trigger-based compute service

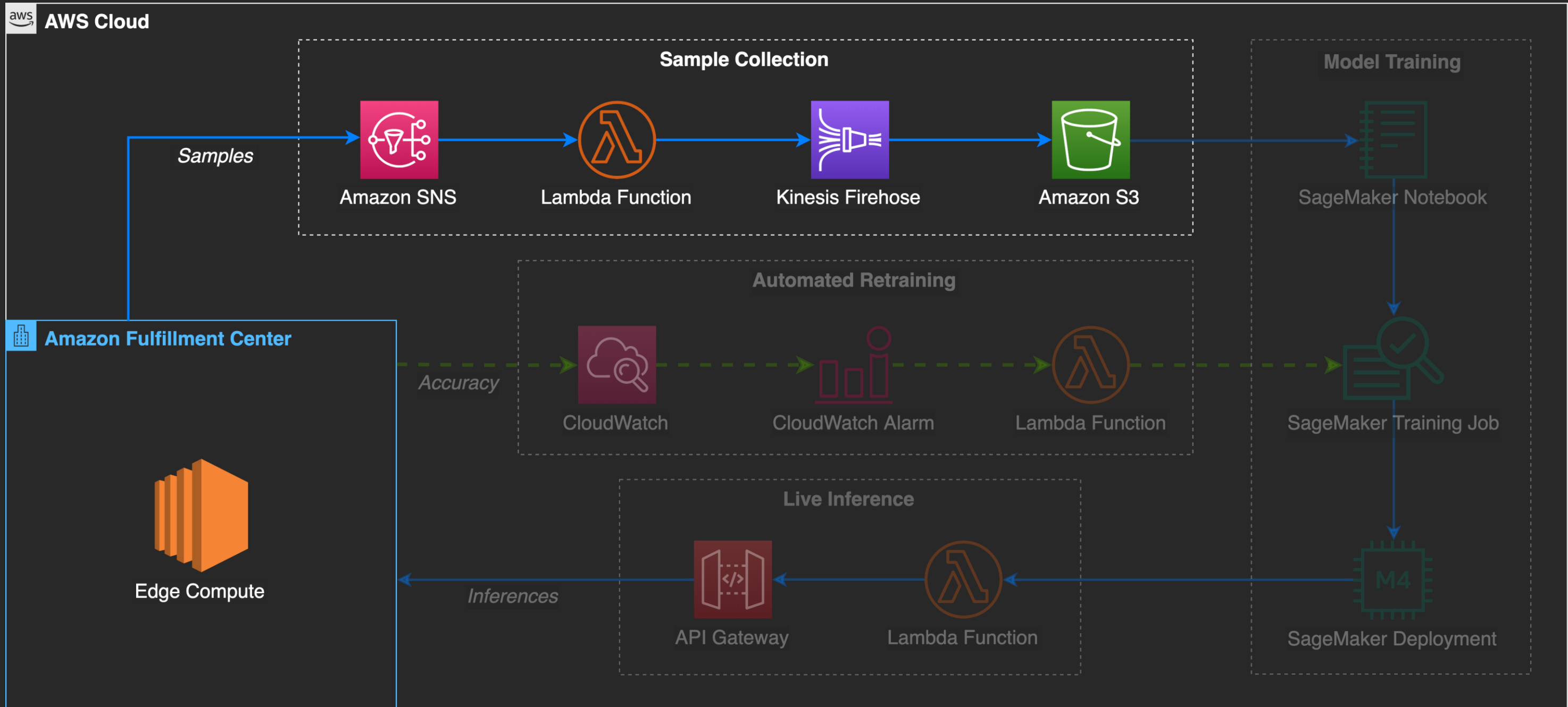


Kinesis Firehose – data batching service



S3 – unlimited object storage service

Sample Collection



Hands-On Lab 1

(20 Minutes)

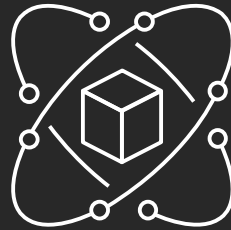
Machine Learning Lifecycle



Feature
Selection



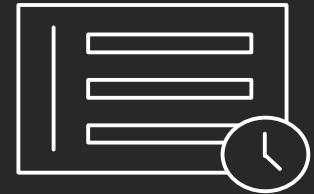
Sample
Collection



Model
Training



Live
Inference



Automated
Retraining

Machine Learning Lifecycle



Feature
Selection



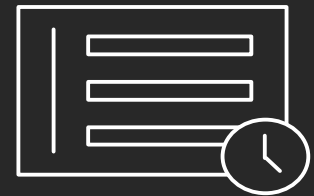
Sample
Collection



Model
Training

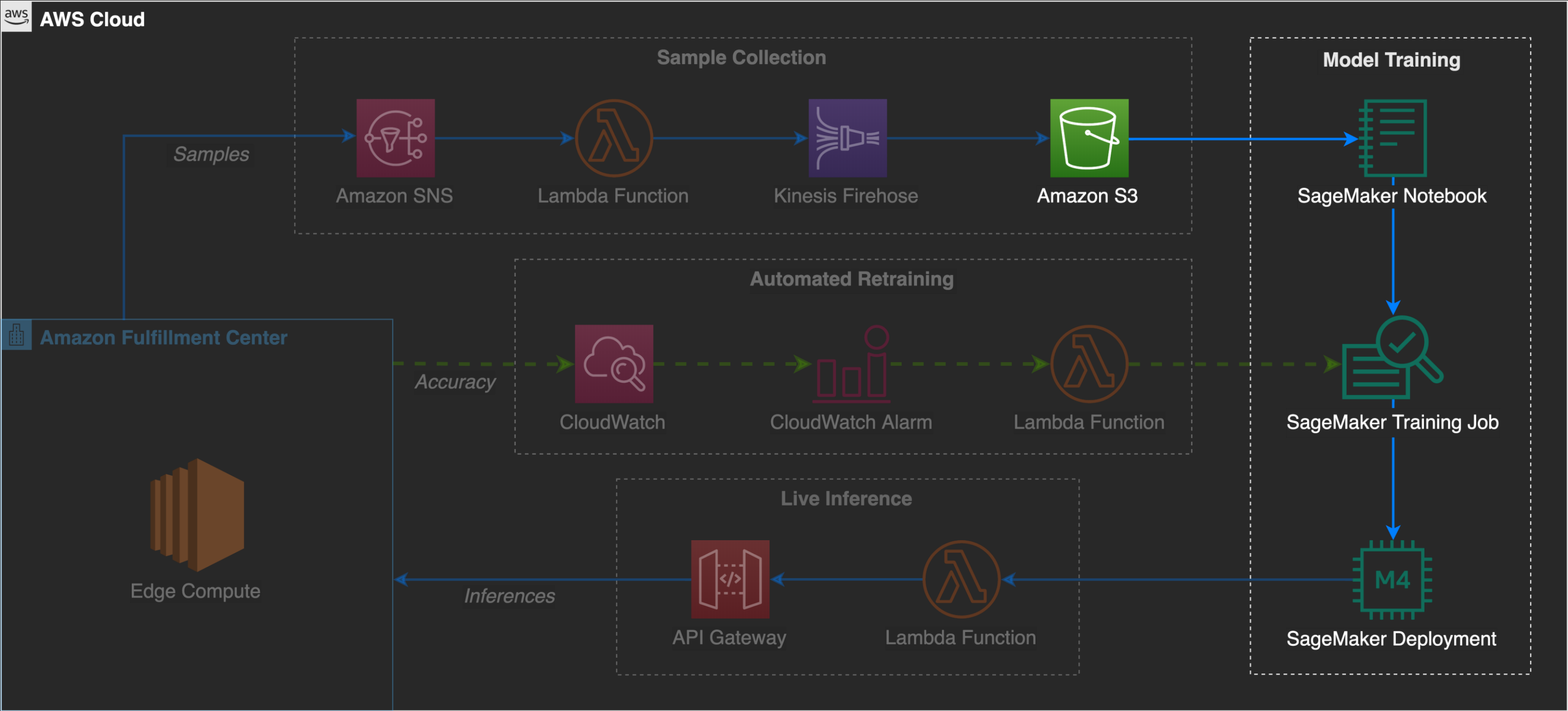


Live
Inference



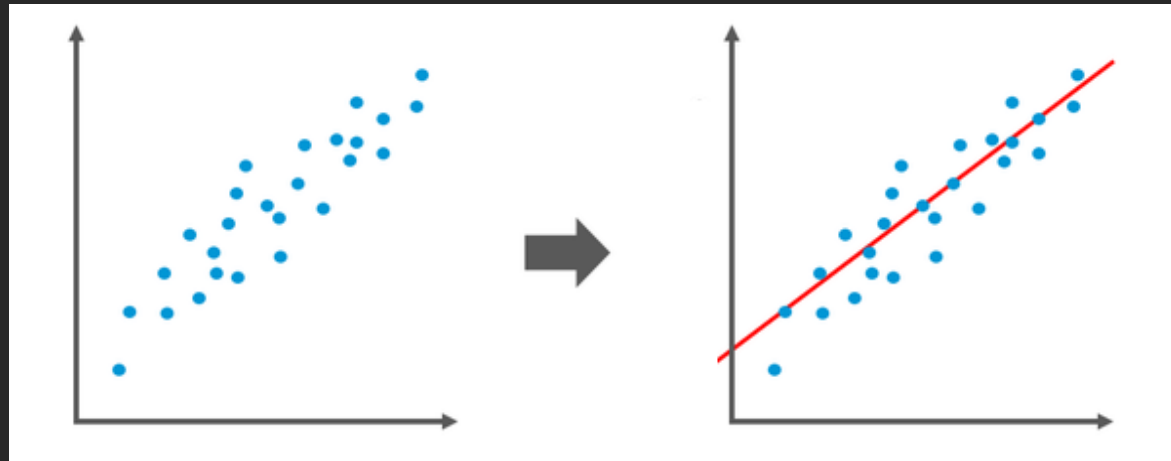
Automated
Retraining

Model Training

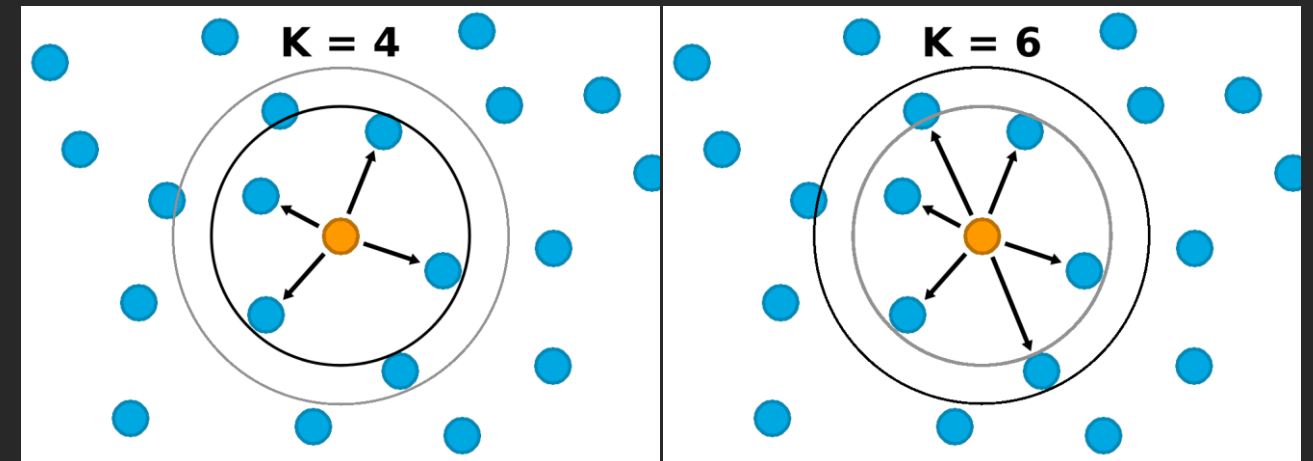


Built-In Algorithms

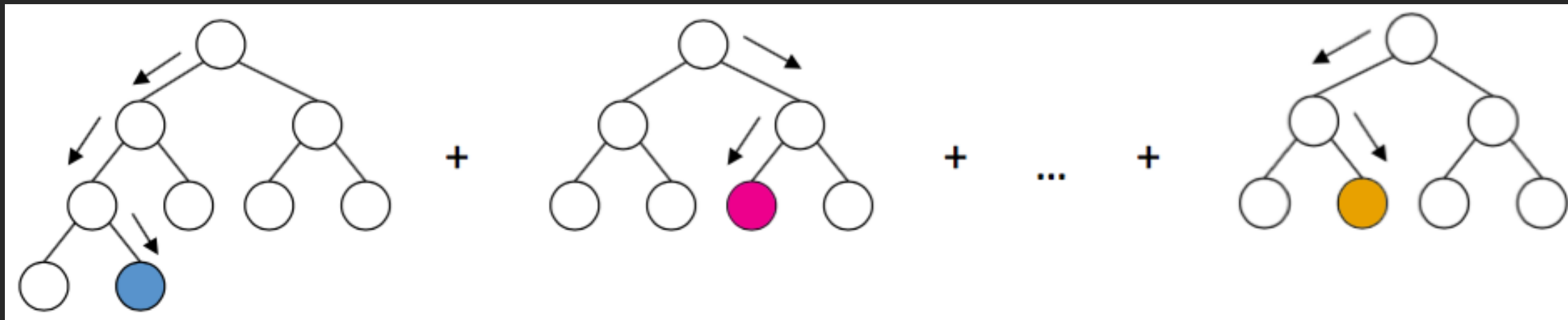
Linear Learner



K-Nearest Neighbors



XGBoost Decision Trees



Hands-On Lab 2

(40 Minutes)

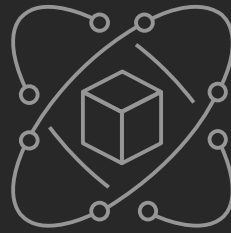
Machine Learning Lifecycle



Feature
Selection



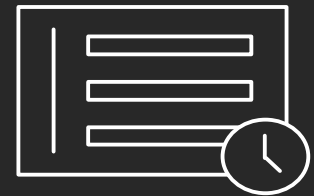
Sample
Collection



Model
Training



Live
Inference



Automated
Retraining

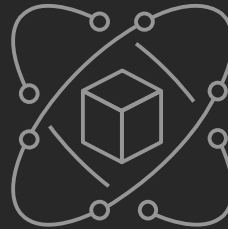
Machine Learning Lifecycle



Feature
Selection



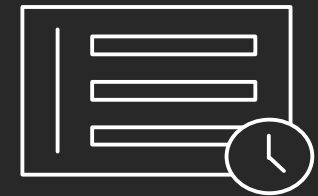
Sample
Collection



Model
Training

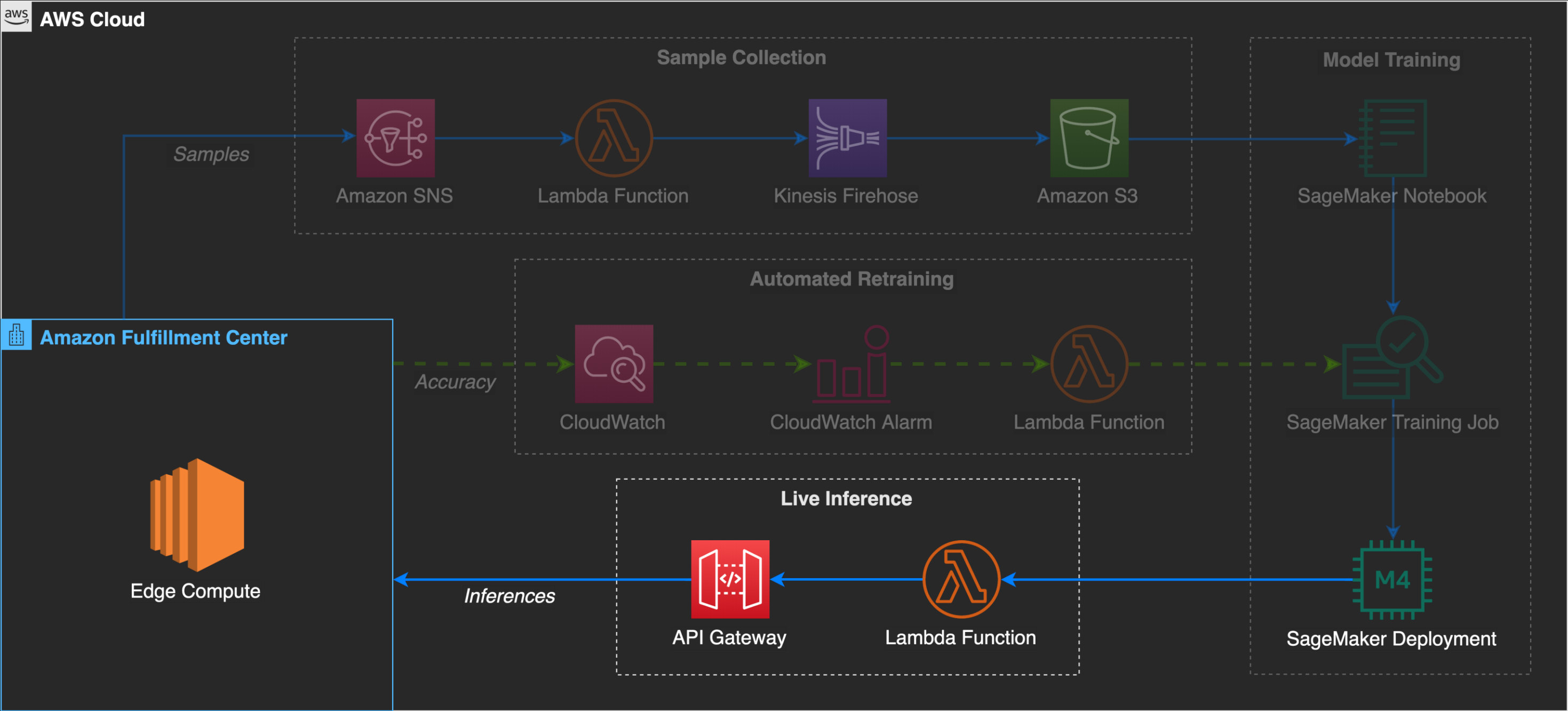


Live
Inference

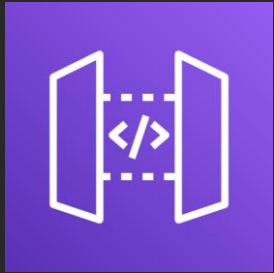


Automated
Retraining

Live Inference



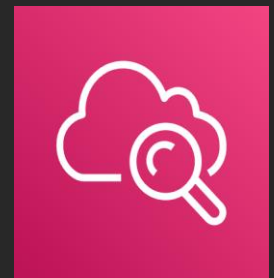
AWS Technologies



API Gateway – endpoint API service

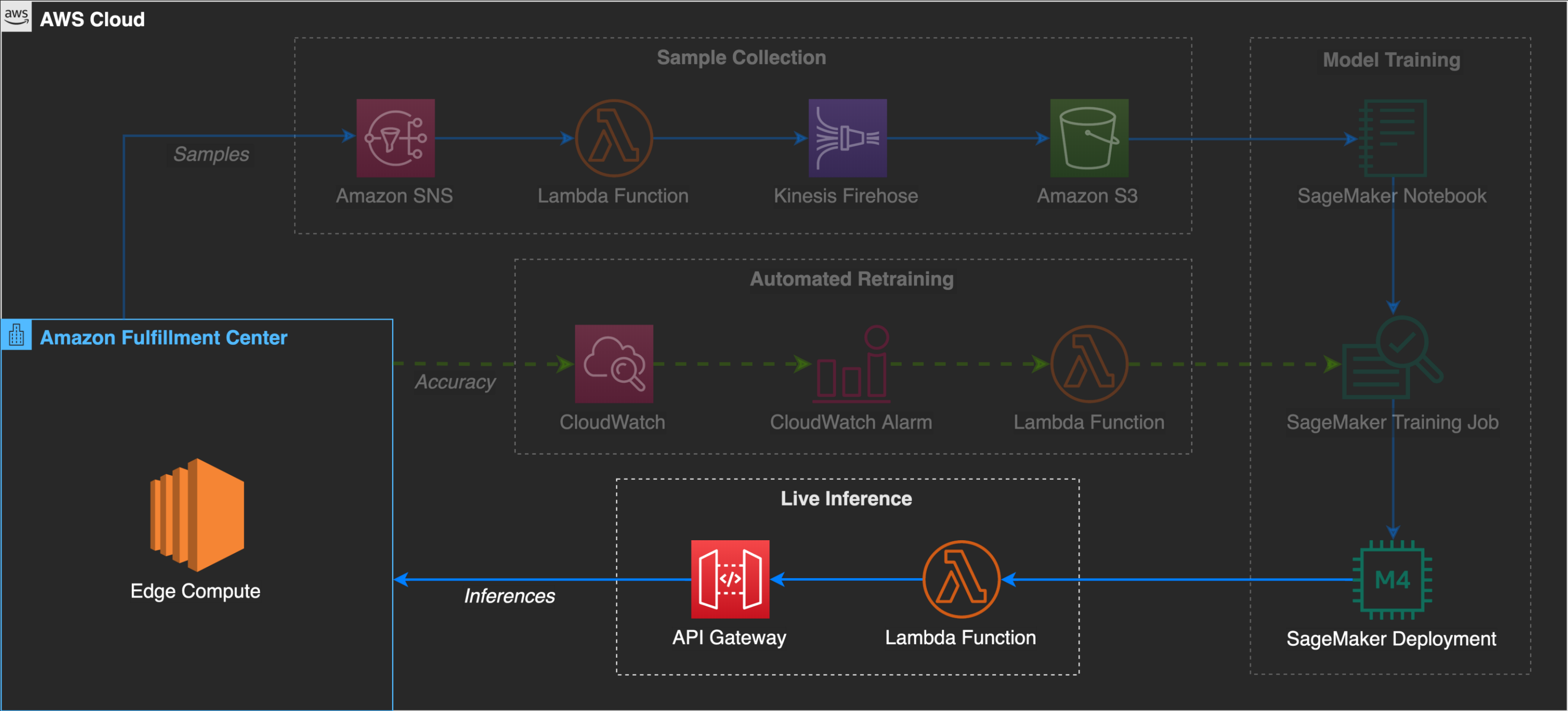


Lambda – trigger-based compute service



CloudWatch – monitoring platform

Live Inference



Hands-On Lab 3

(20 Minutes)

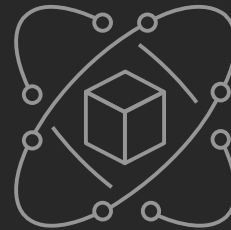
Machine Learning Lifecycle



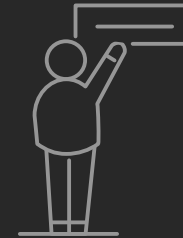
Feature
Selection



Sample
Collection



Model
Training



Live
Inference



Automated
Retraining

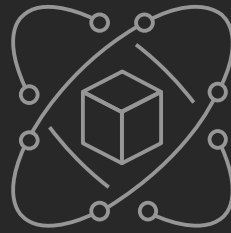
Machine Learning Lifecycle



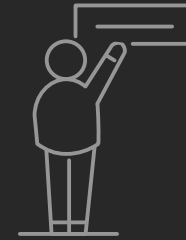
Feature
Selection



Sample
Collection



Model
Training

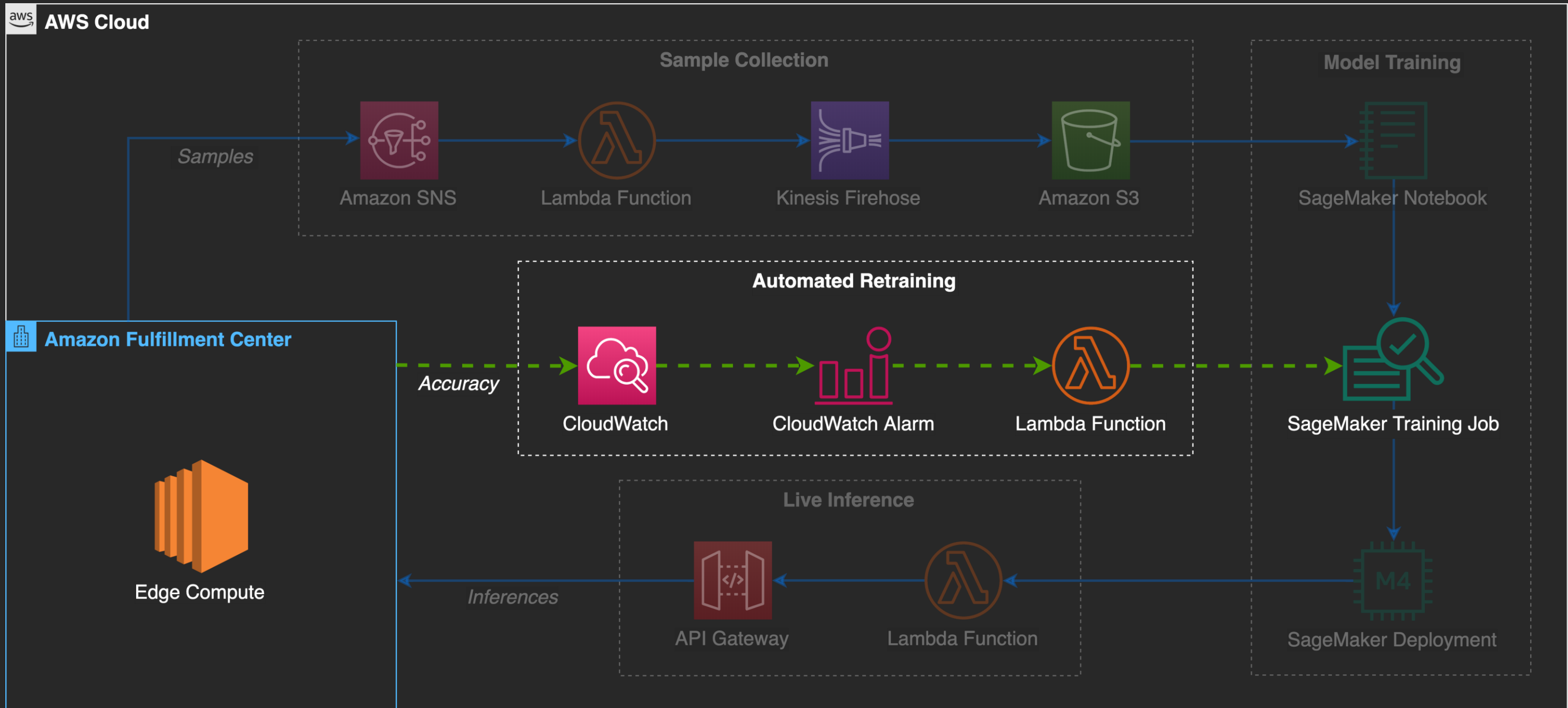


Live
Inference



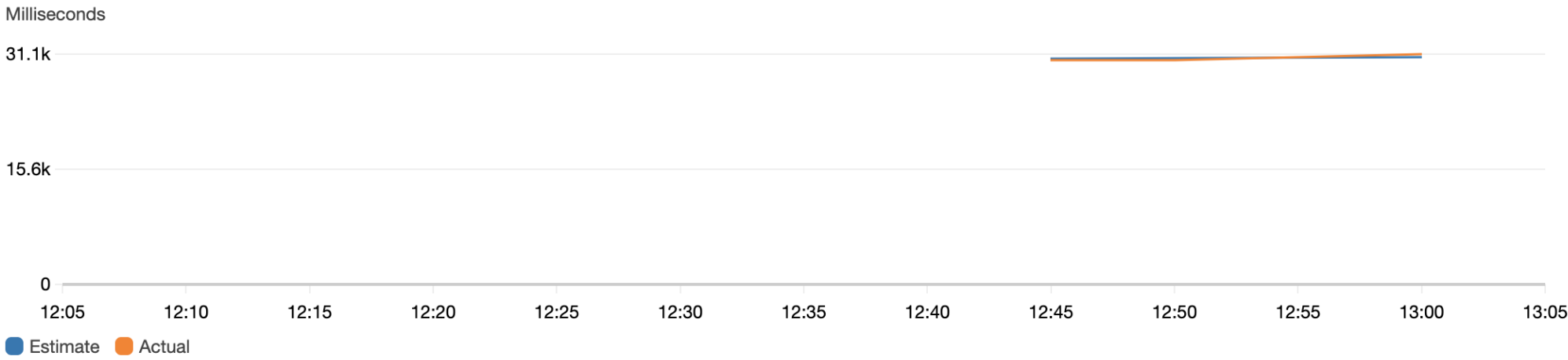
Automated
Retraining

Automated Retraining



CloudWatch Dashboard

Estimated Duration vs Actual Duration



Average Signed Error

-0.05

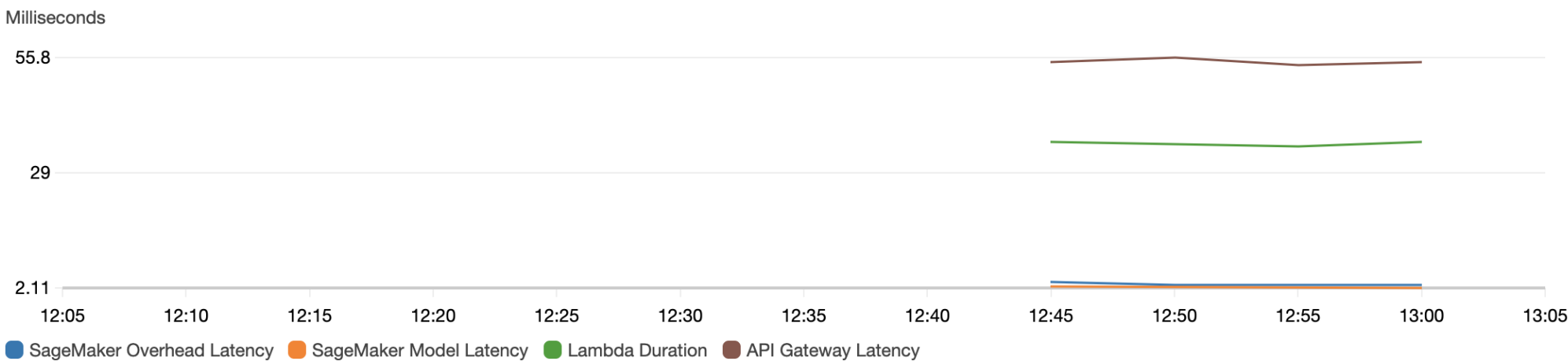
Signed Error

Average Absolute Error

7.36_s

AbsoluteError

Latency per Inference



Inferences per Second

4.9

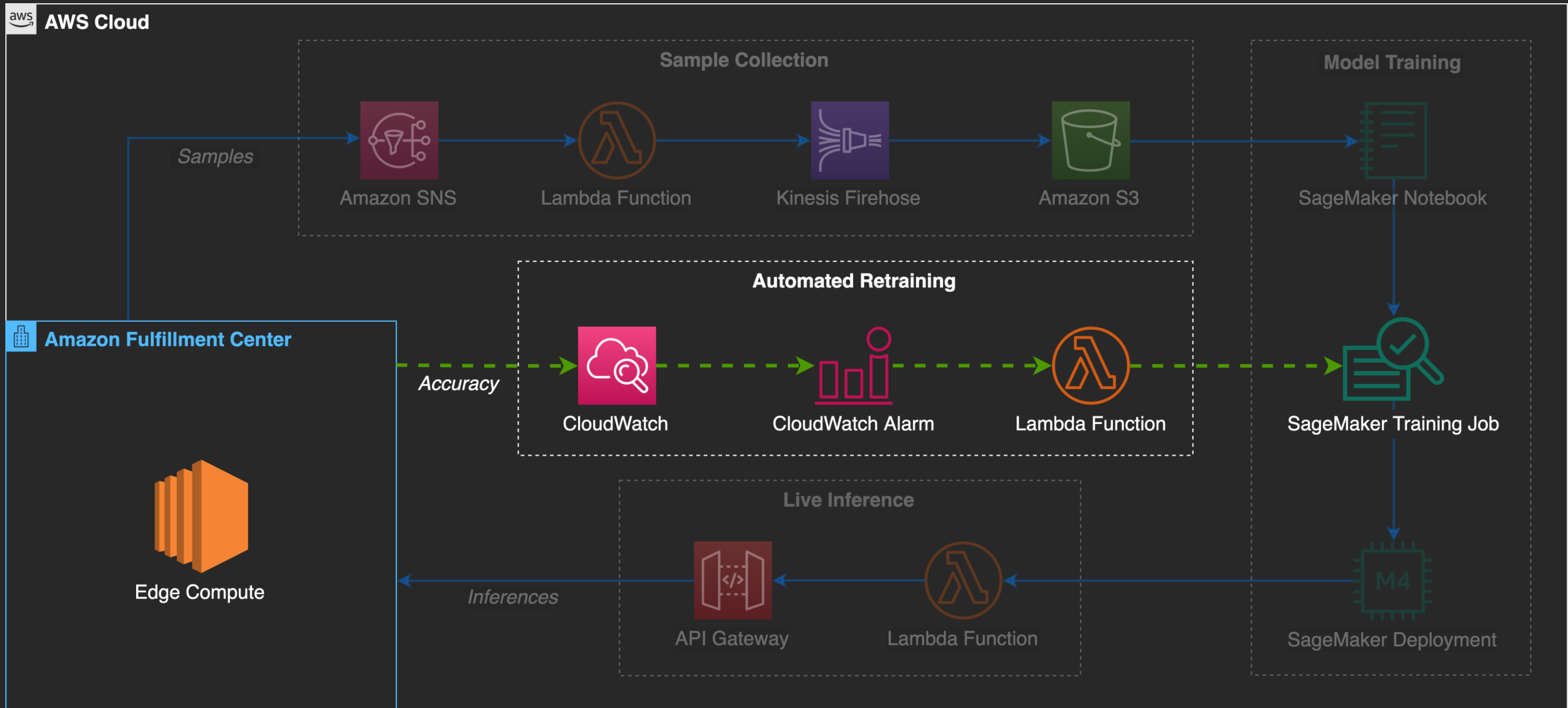
Inferences per Second

Total Inferences Made

5.61_k

Invocations

Automated Retraining



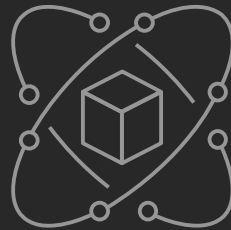
Machine Learning Lifecycle



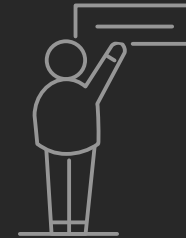
Feature
Selection



Sample
Collection



Model
Training



Live
Inference



Automated
Retraining

Next Steps

- Run the same experiments in SageMaker Studio
- Try using SageMaker Autopilot to select the algorithm
- Try using Model Monitor to kick off automated retraining
- Step through the ML lifecycle for your own use case

Thank you!

Rick Nayar

Solutions Architect
Amazon Web Services

Mike Calder

Software Engineer
Amazon Robotics



Please complete the session survey in the mobile app.