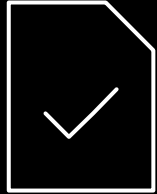


# Build, train and deploy Machine Learning models on Amazon SageMaker

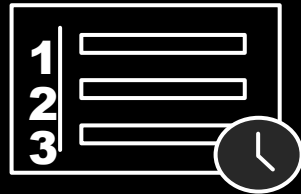
Julien Simon  
Global Evangelist, AI & Machine Learning  
@julsimon

Nils Mohr  
Flight Data Programmer Analyst  
British Airways

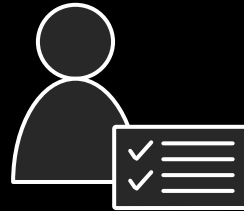
# Amazon SageMaker



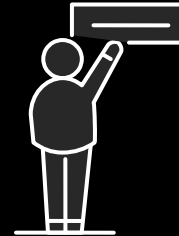
Collect and  
prepare training  
data



Choose and  
optimize your  
ML algorithm



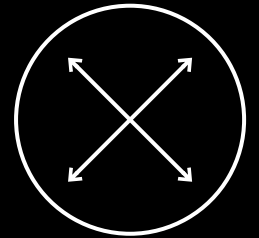
Set up and  
manage  
environments  
for training



Train and  
Tune ML Models



Deploy models  
in production



Scale and manage  
the production  
environment

Same service and SDK from experimentation to production

intuit.



tinder



CONVOY

SIEMENS



DOW JONES



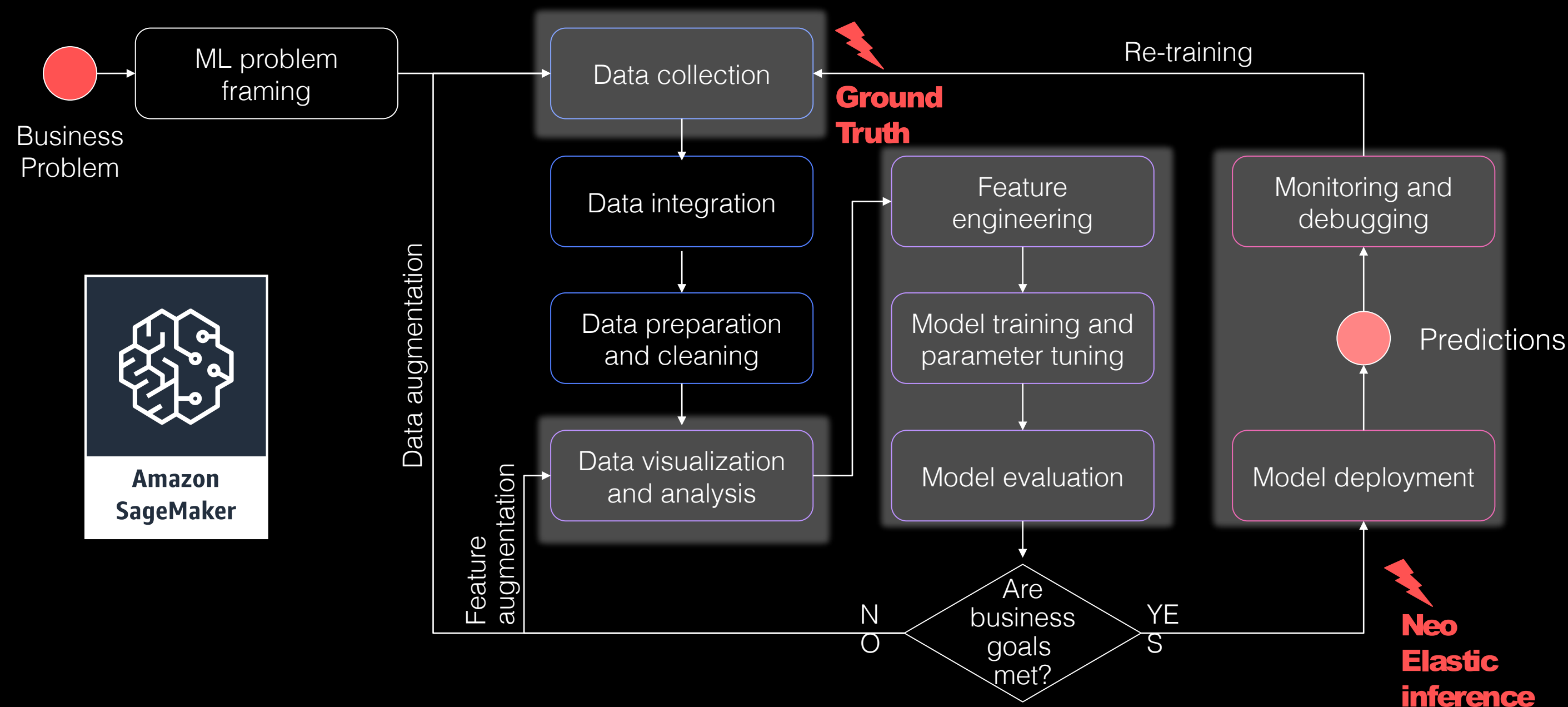
SONY



aws SUMMIT

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Build, train and deploy models using SageMaker



# Ai<sup>2</sup>

Aircraft Information Intelligence  
British Airways Predictive Maintenance Platform

Nils Mohr  
Flight Data Programmer Analyst  
British Airways





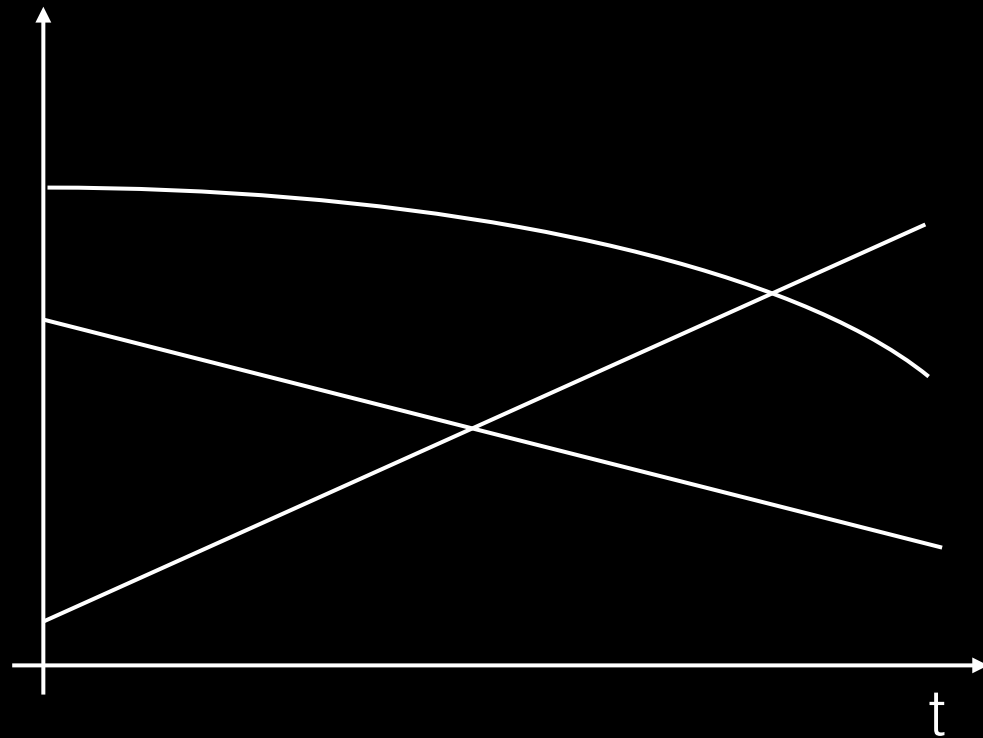
# British Airways

- 290+ aircraft
- 46 million+ passengers per year
- 200+ destinations in 75+ countries

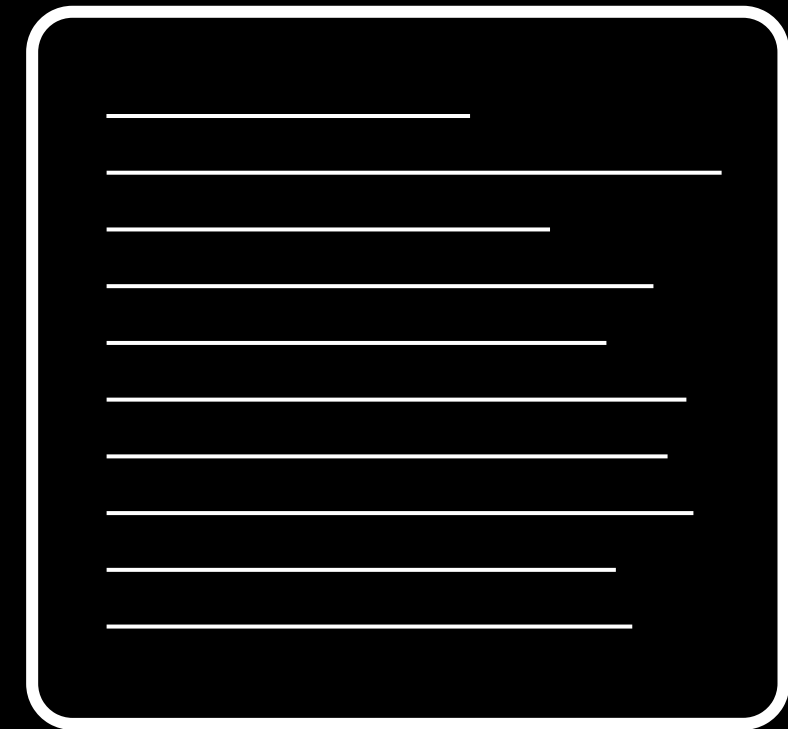


# Flight data

# What is flight data?



Timeseries data



Reports



# Timeseries data

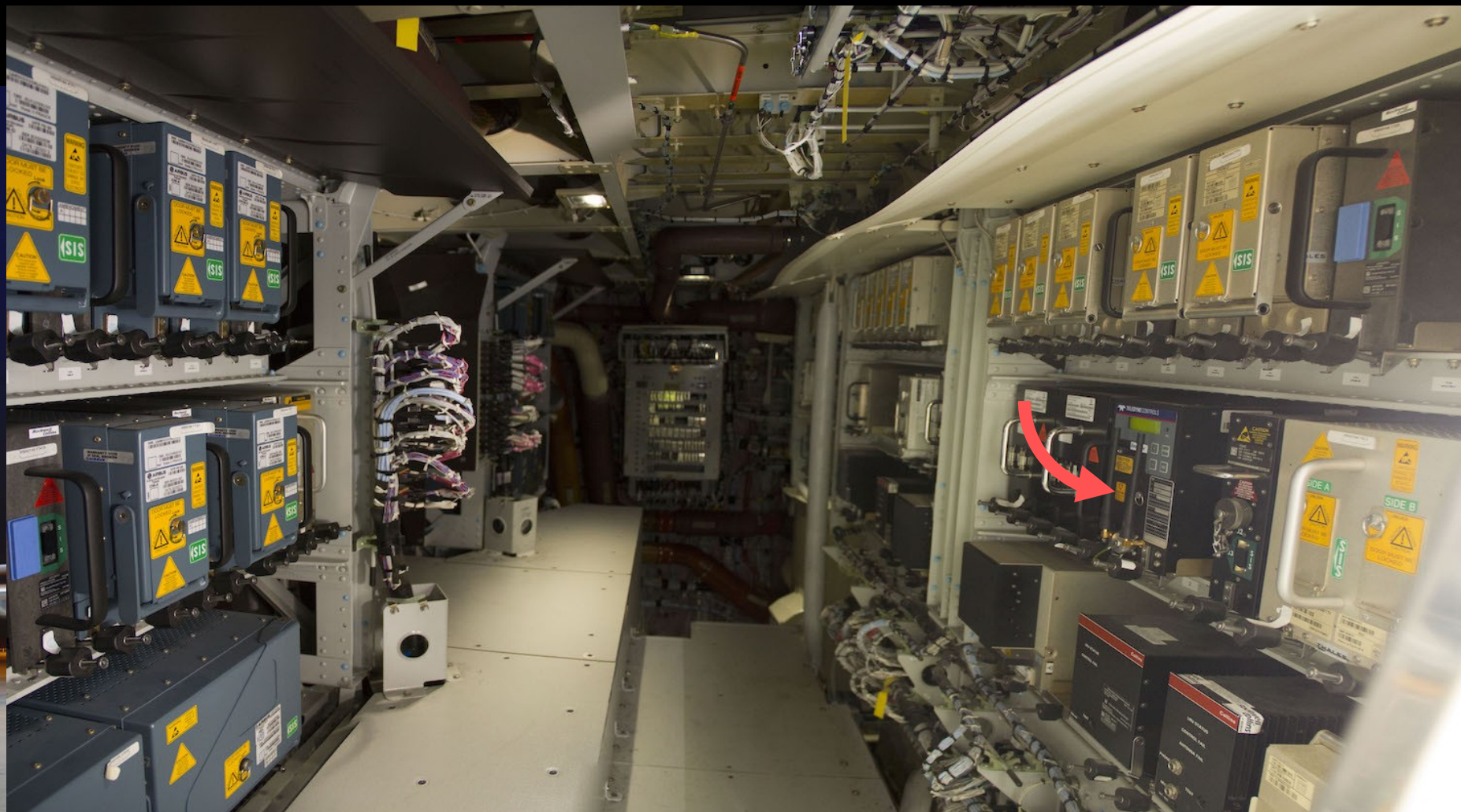
Flight Data Recorder Quick Access Recorder



Cockpit Voice Recorder

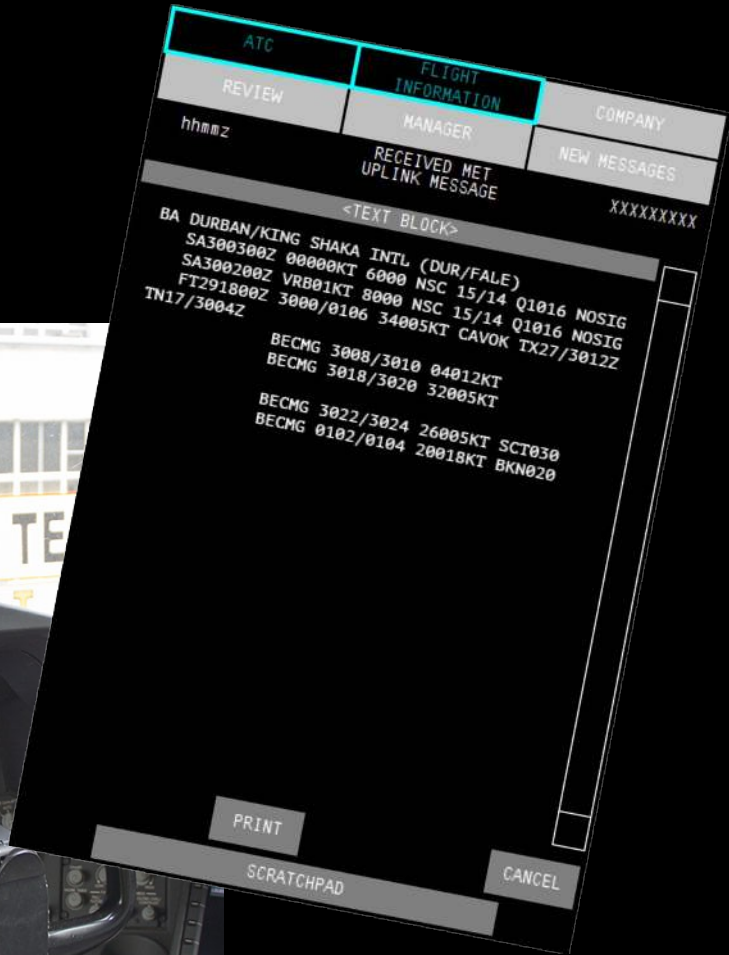
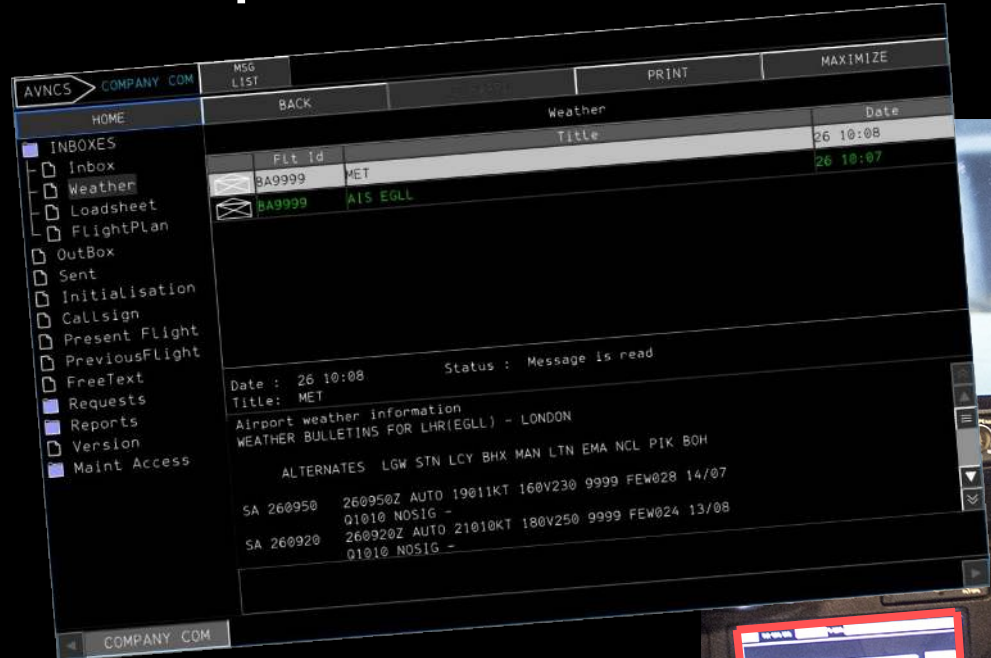


# Timeseries data





# Report data



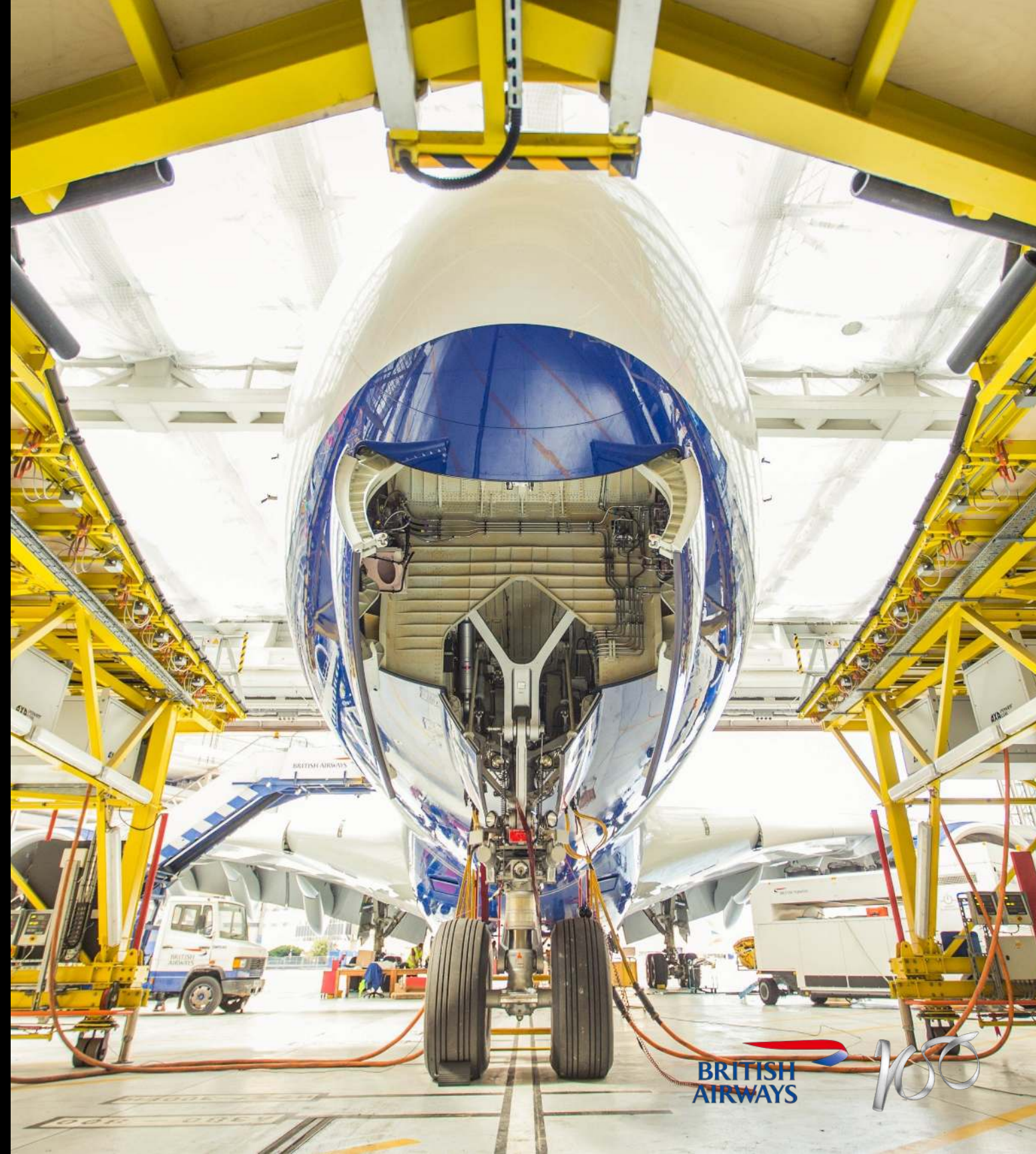


# Our dataset

3500 parameters

15 million reports

5 million flights





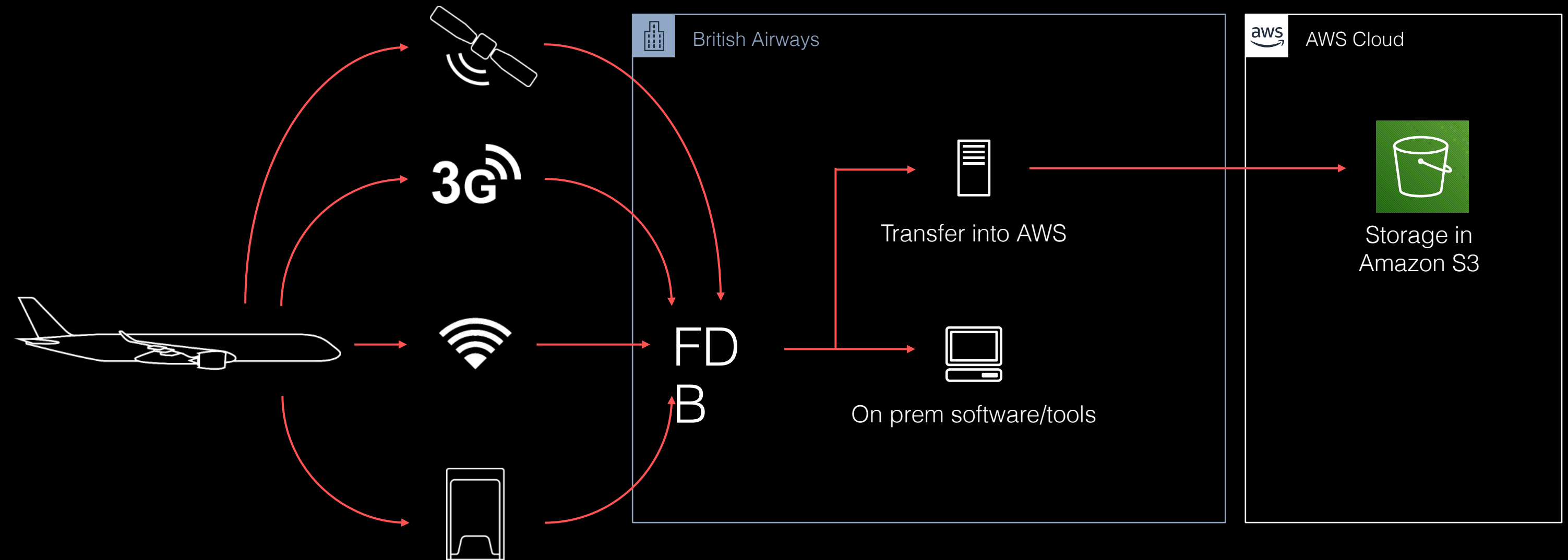
# Reasons for the development



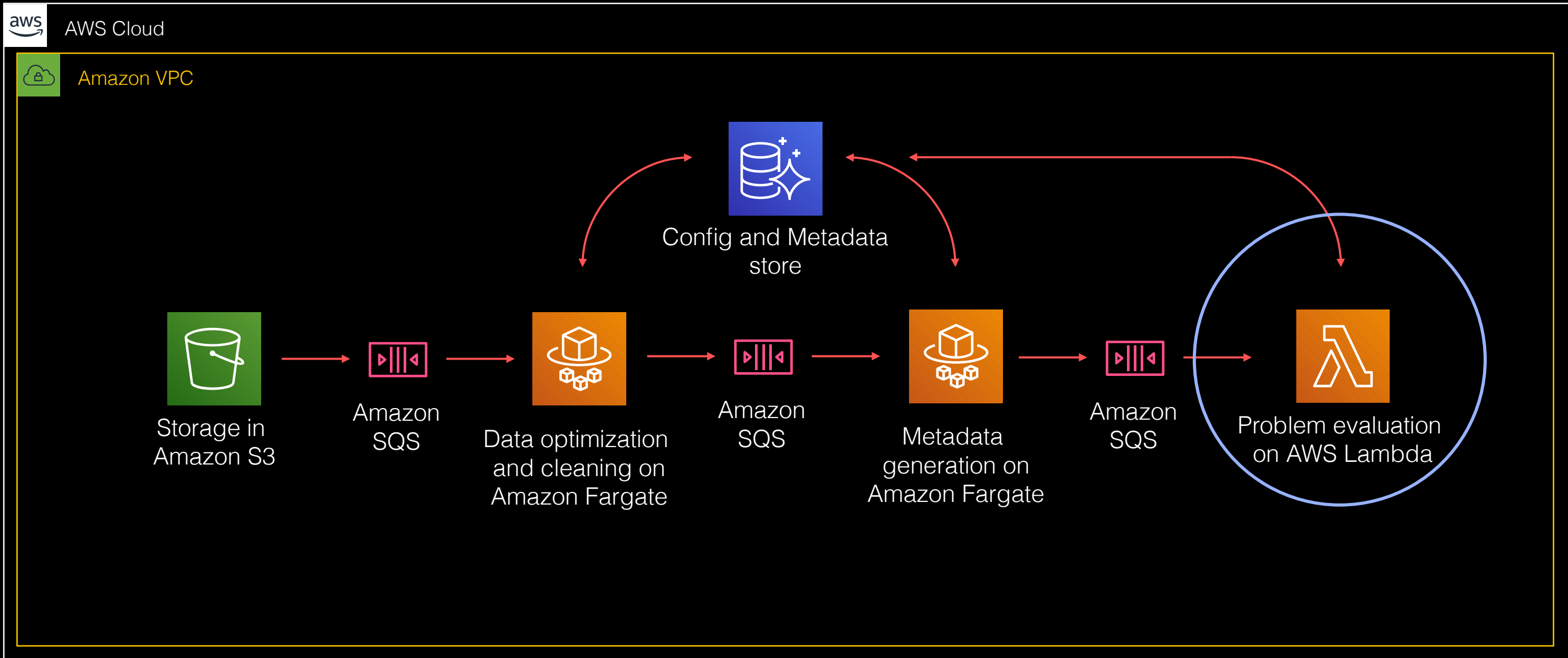


# Architecture

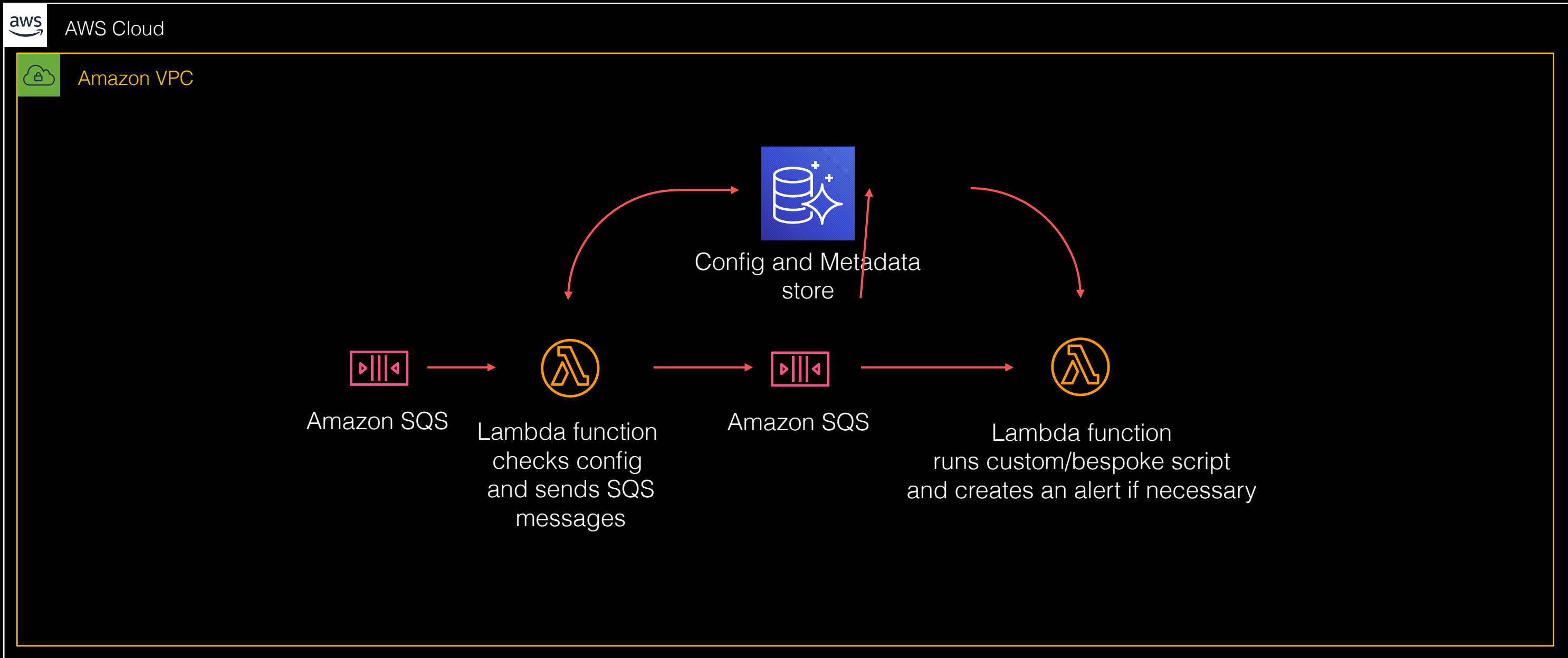
# High level architecture



# Our AWS Architecture

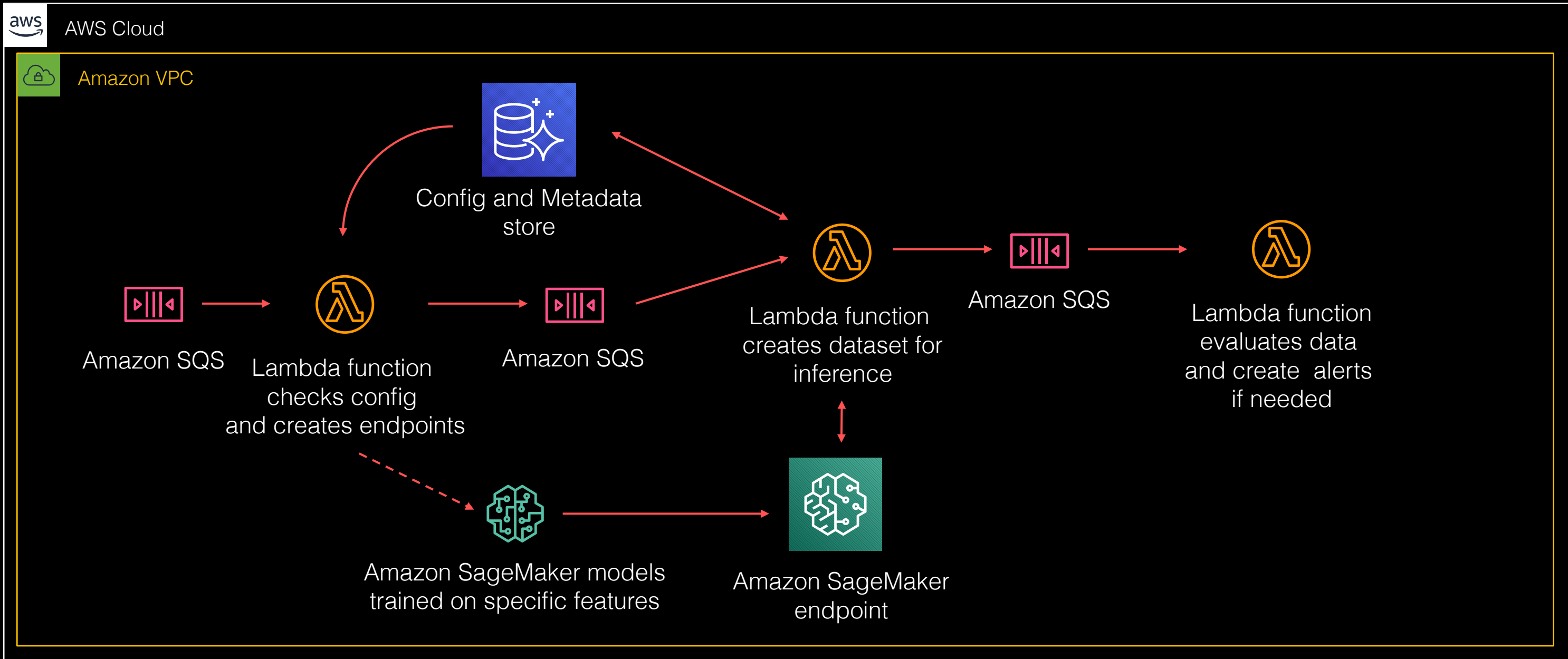


# Traditional Event detection





# Event detection with Amazon SageMaker



# Lessons learned

Understand the problem

Ensure data quality

Ask questions

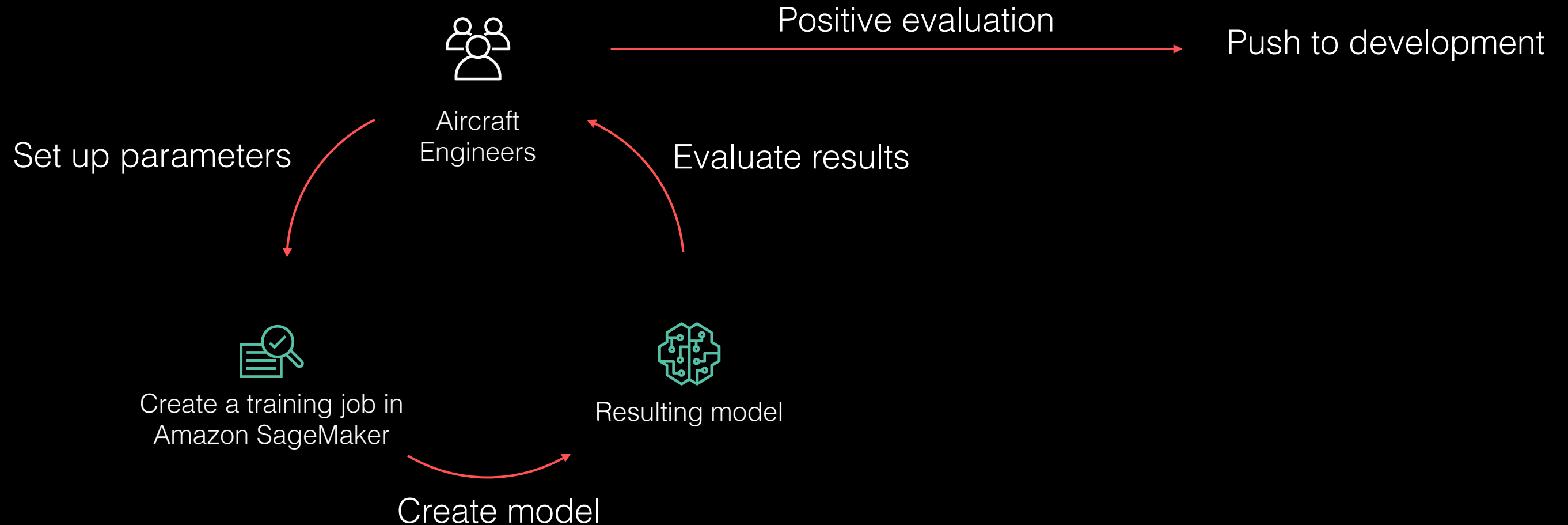
Apply the easiest solution





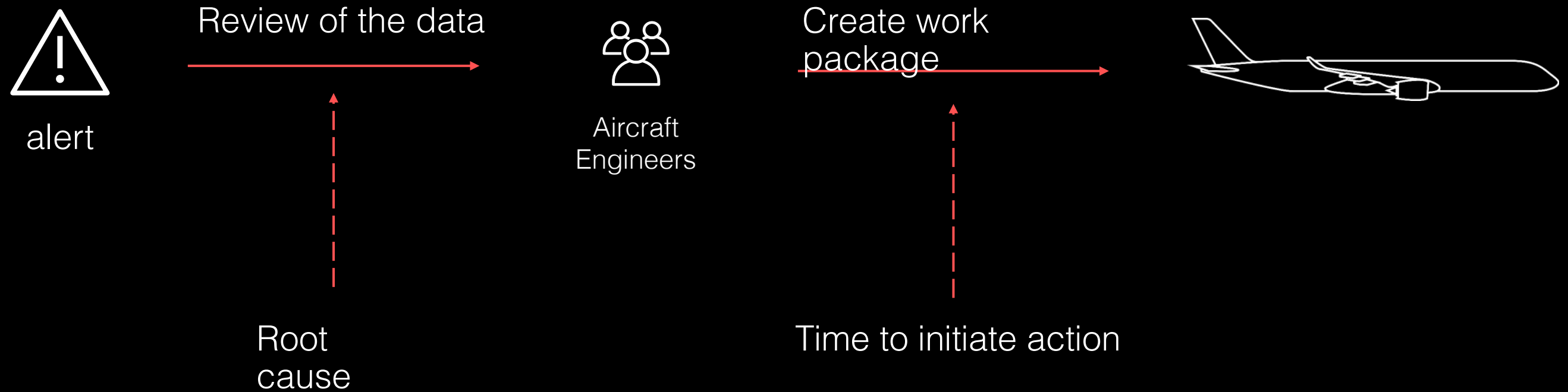
# What's next?

# Where do we go from here?





# Where do we go from here?



# Deep Learning on Amazon SageMaker

# Model options



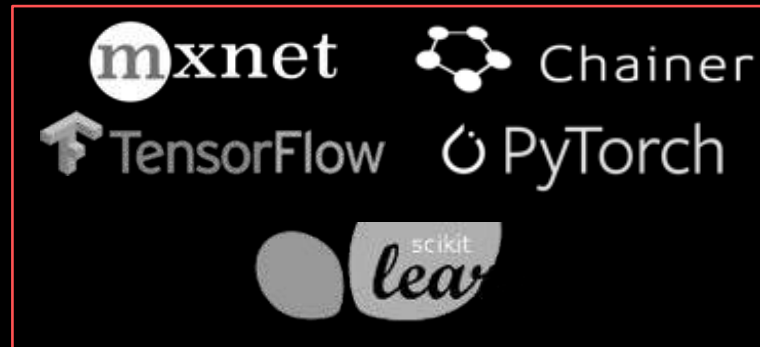
Training code

**AWS Machine Learning Marketplace: 150+ off-the-shelf models**

Factorization Machines  
Linear Learner  
Principal Component Analysis  
K-Means Clustering  
XGBoost  
And more

Built-in Algorithms (17)

No ML coding required  
No infrastructure work required  
Distributed training  
Pipe mode



Built-in Frameworks

Bring your own code: script mode  
Open source containers  
No infrastructure work required  
Distributed training  
Pipe mode



Bring Your Own Container

Full control, run anything!  
R, C++, etc.  
No infrastructure work required

# Built-in Deep Learning frameworks: just add your code



- Built-in containers for **training** and **prediction**.
  - Code available on Github, e.g. <https://github.com/aws/sagemaker-tensorflow-containers>
  - Build them, run them on your own machine, customize them, etc.
- **Script mode**: use the **same code** as on your laptop

---

**No infrastructure work required**: simply define instance type and instance count

**Distributed training** out of the box: zero setup

**Pipe mode**: stream infinitely large datasets directly from Amazon S3



# AWS: The platform of choice to run TensorFlow



**85%** of all  
TensorFlow  
workloads in the  
cloud runs on AWS

Source: Nucleus Research, November 2018

# Optimizing Tensorflow for Amazon EC2 instances

C5 instances (Intel Skylake)

Training ResNet-50 with the ImageNet dataset using our optimized build of Tensorflow 1.11 on a **c5.18xlarge** instance type is **11x faster** than training on the stock binaries.

P3 instances (NVIDIA V100)

Tensorflow scaling efficiency with 256 GPUs

**65**

Stock version



**90**  
**%**

AWS-optimized version

# Demo: Keras+Tensorflow

Script mode

Automatic model tuning

Elastic inference

<https://gitlab.com/juliensimon/dlnotebooks/tree/master/keras/04-fashion-mnist-sagemaker-advanced>

# Apache MXNet: Deep Learning for enterprise developers



## Start with off-the-shelf models

- Gluon CV and Gluon NLP
- ONNX compatibility

## Fast and scalable training

- Keras-MXNet up to 2x faster than Keras-TensorFlow
- Near-linear scalability up to 256 GPUs
- Dynamic training

## Easy deployment

- Java/Scala APIs
- Model Server

# Demo: Gluon CV

State of the art models in just a few lines of code

<https://gluon-cv.mxnet.io/>



# Getting started

<http://aws.amazon.com/free>

<https://aws.amazon.com/sagemaker>

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

<https://github.com/aws-labs/amazon-sagemaker-examples>

<https://medium.com/@julsimon>

<https://gitlab.com/juliensimon/dlnotebooks>

# Thank you!

Julien Simon  
Global Evangelist, AI and Machine Learning

@julsimon



Please complete the  
session survey.