

How We Used Databricks, MLeap, and Kubernetes to Productionize Spark ML Faster



Edward Kent

Spark Summit Europe 2018

#SAISEnt9

2,489 cars found

★ Save search ⚙️ Reset

Within 20 miles ▼

m154fn

Make Audi ▶

Model Any ▶

Model Variant Any ▶

Price Any ▼

Total price Monthly price

From (any) ▼

To (any) ▼

Finance examples for HP and PCP will be returned if available, based on 36 month term, £1,000 deposit and 10,000 miles per year. Finance subject to status.

Year Any ▼

Mileage Any ▼

Body type Any ▶

Fuel type Any ▶

All

New

Used

Sort by: Sponsored adverts first ▼

Featured listing

38

**Audi A3 Hatchback S Line 1.4 3dr**

NAV+PARKING SENSORS+BLUETOOTH

2015 (65 reg) | Hatchback | 7,639 miles | 1.4L | 148bhp | Manual | Petrol

SAT NAV! AIR CON! ALLOY WHEELS! 12 MONTHS WARRANTY! 12 MONTHS RAC COVER! FINANCE ...

Trade seller - See all 156 cars
Wigan - 17 miles away**£16,995**

GREAT PRICE

Based on similar cars

£349per month (PCP)
Finance example

CHORLEY GROUP

[Get an insurance quote](#) | [Check its history](#)

20

**Audi A3 1.6 5dr**

Full Service History, 2 Keys

2001 (51 reg) | Hatchback | 128,000 miles | 1.6L | 101bhp | Manual | Petrol

MOT November 2018, Vehicle History & Maintenance Receipts, 2 Former Keepers, in ...

Trade seller - See all 256 cars
Newton-Le-Willows - 16 miles away**£885**

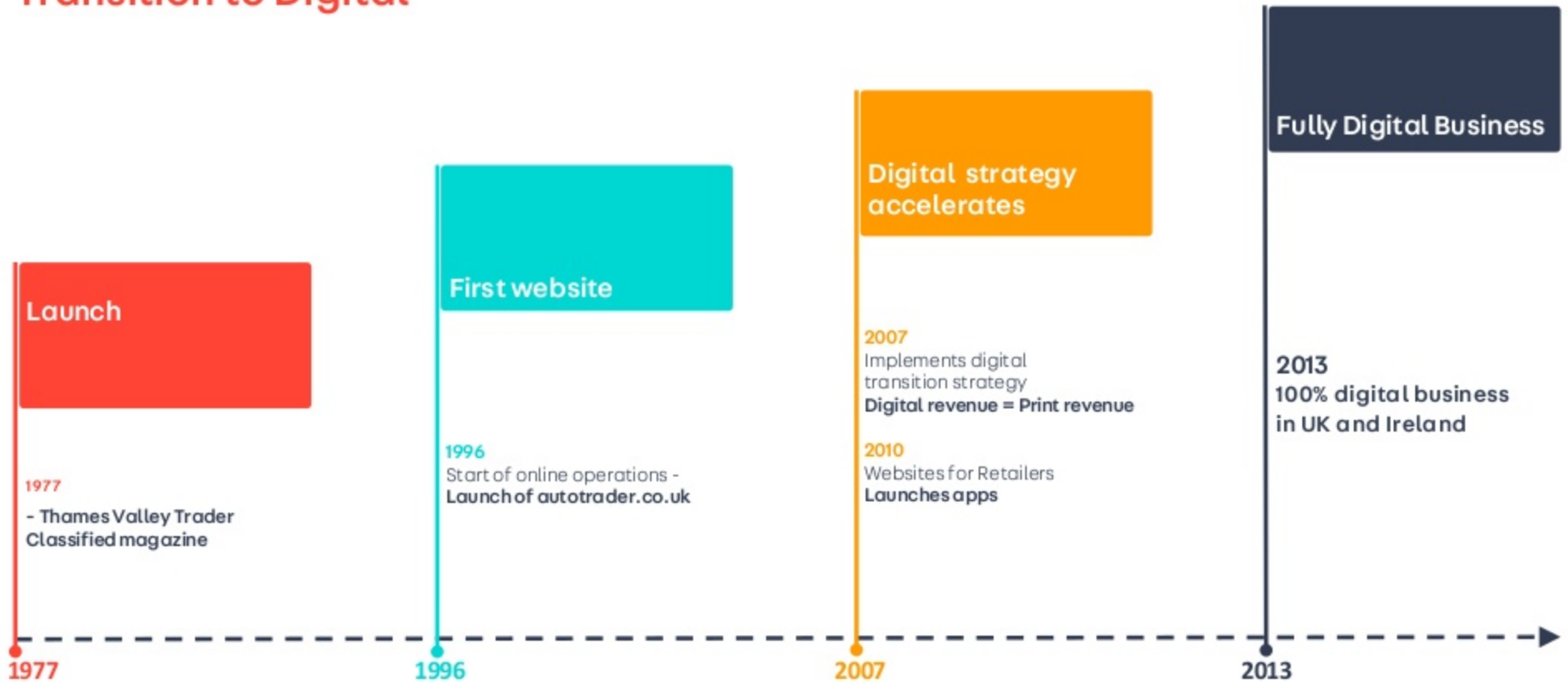
Prestige Motors

Read our latest reviews

EXPERT REVIEW
2 months ago
[Audi Q8 SUV \(2018 - \) review](#)EXPERT REVIEW
3 months ago
[Audi A6 saloon \(2018 - \) review](#)EXPERT REVIEW
7 months ago
[Audi A7 Hatchback \(2018 - \) review](#)EXPERT REVIEW
9 months ago
[Audi RS4 Avant estate \(2017 - \) review](#)

Introduction

Transition to Digital



12th
biggest UK
website

Page views, ComScore CY2017



450,000

Cars listed per day
(average)

2x
more
influential
for new car
buyers
than nearest
competitor

Annual Car Buying Report 2016

55m
monthly
cross-platform
visits

ComScore – Nov 17

4x
more
Auto
searches
than
Google

80% UK Auto
Retailers
advertise on Auto Trader

9m
monthly
unique users

ComScore – Nov 17

10x
More minutes on
site than all OEMS
combined

Introduction

Services for consumers



New and used car search listings



Searching by monthly budget



Dealer reviews and ratings



Vehicle check



Valuations



Price indicator



Private sales



Car reviews

Introduction

Services for retailers



Classified advertising



Valuations



Finance solutions



Creating a trusted marketplace

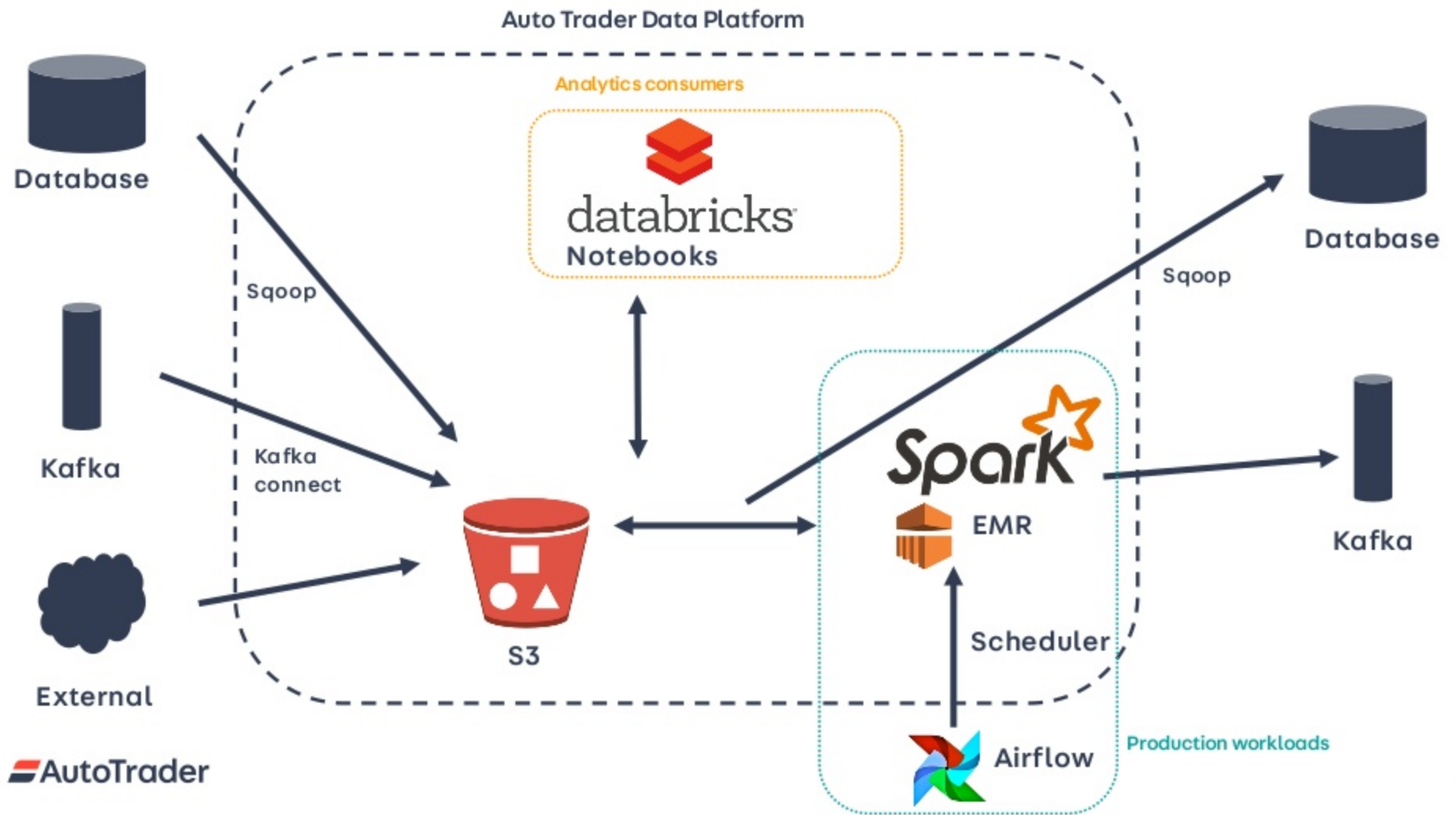


Forecourt management tools



Retailer education and insight

Data at Auto Trader



Days to Sell

[Stock](#)[Performance](#)[Sourcing](#)[i-Control](#)[My website](#)[News & Explore](#)[Help](#)[Account](#)[Notifications](#)**WORK IN PROGRESS**[< Forecourt](#)

Ford Fiesta 1.0 EcoBoost Titanium 5dr (start/stop)

EO14UHX 2014 (14) 50,000 miles Petrol Manual Blue 2 owners

[Add to stock](#)[Customer view](#)[Save appraisal](#)

Get an adjusted valuation

[Add spec](#)

Get a more accurate valuation by adding the full spec of the car.

Car information

Once a car is added to your forecourt we'll start to display more information here, such as the:

- Appraisal history of the car
- Quality of your advert
- Performance of your advert

Auto Trader insight

Retail market average

Retail adjusted

£8,384 Part ex £6,530

No added spec

AT Retail Rating

[More on our rating](#)

74 out of 100



Typical Ford range

Days to sell

38 days

Live market condition

[View detail](#)

Higher demand than normal

NATIONAL MARKET SUMMARY

Matching cars live on Auto Trader 15

Lowest priced car £10,000

Highest priced car £12,000

Discovery to production

Discovery phase

- Quick iterations
- Fast feedback
- Offline predictions



Production phase

Automated model
training

Automated model
deployment

- Slow feedback cycle
- Fully automated
 - Unit tests
 - Data integrity checks
 - Deployment
- Resilient
- Online, low-latency predictions

Discovery phase

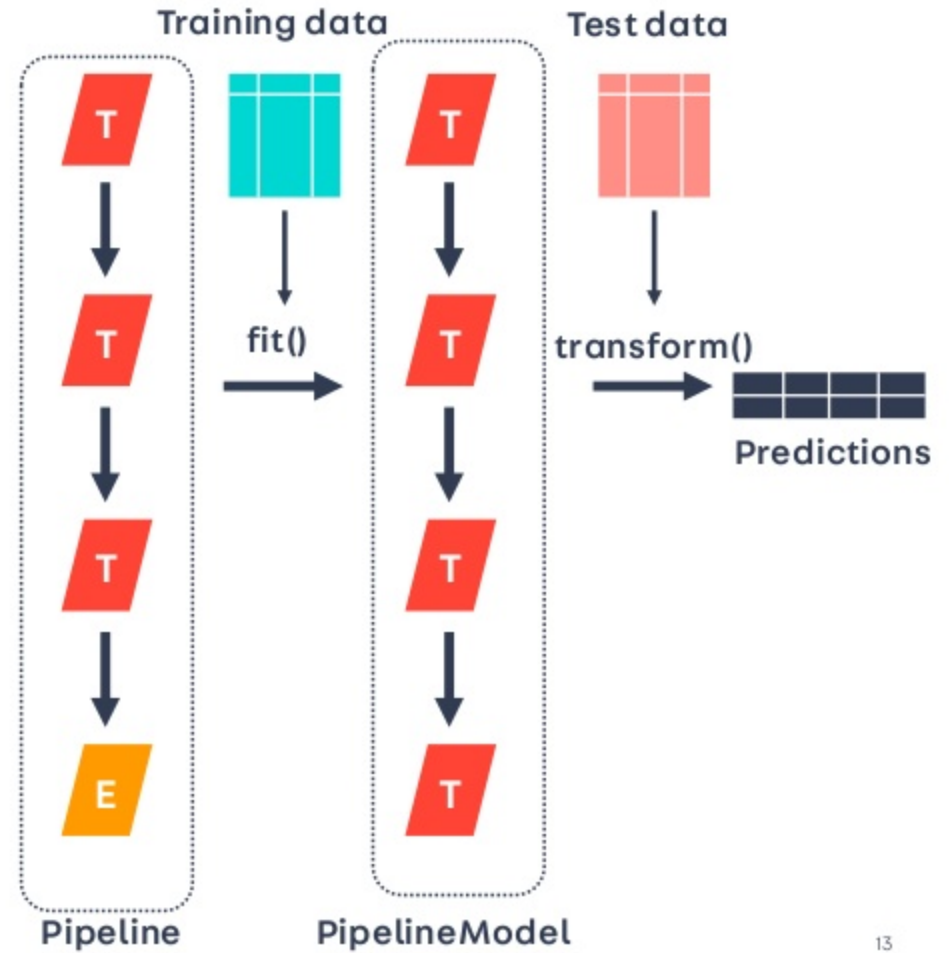
The Model

Spark ML Pipeline

Transformers



Estimators

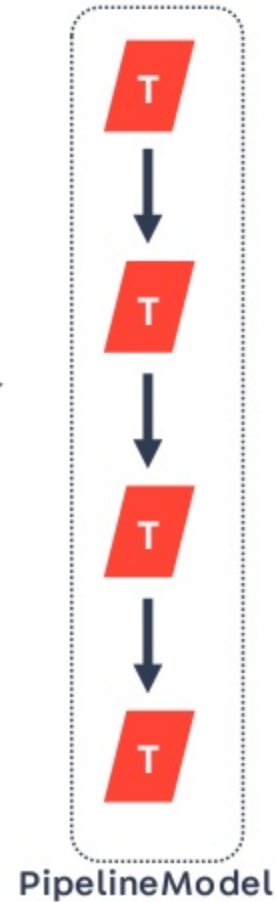


The Model

Days to Sell ML Pipeline



fit()

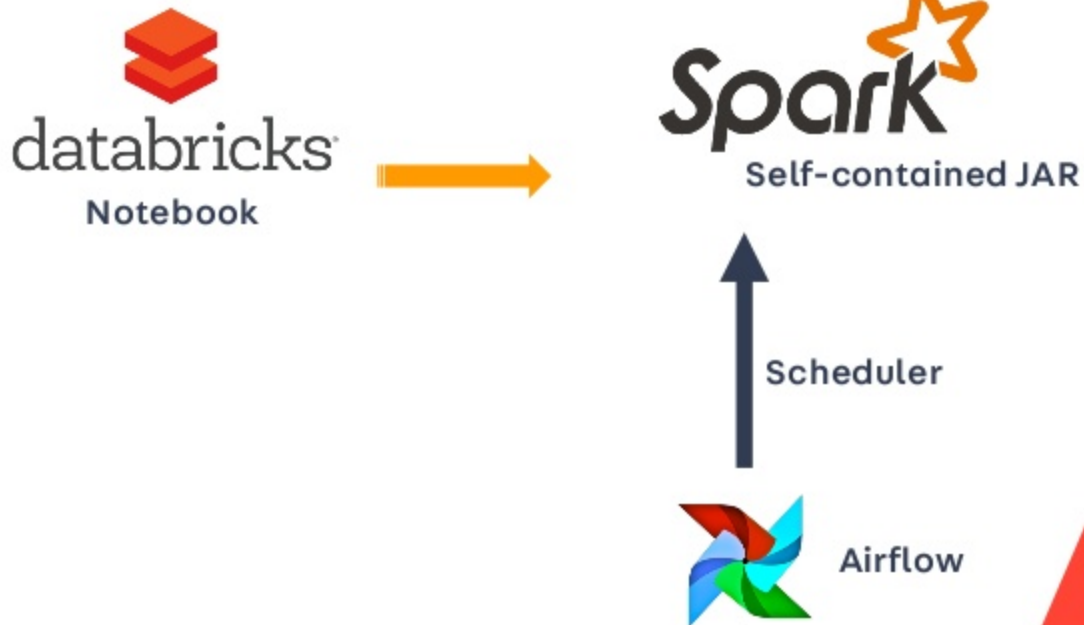


Production phase

Automated model training

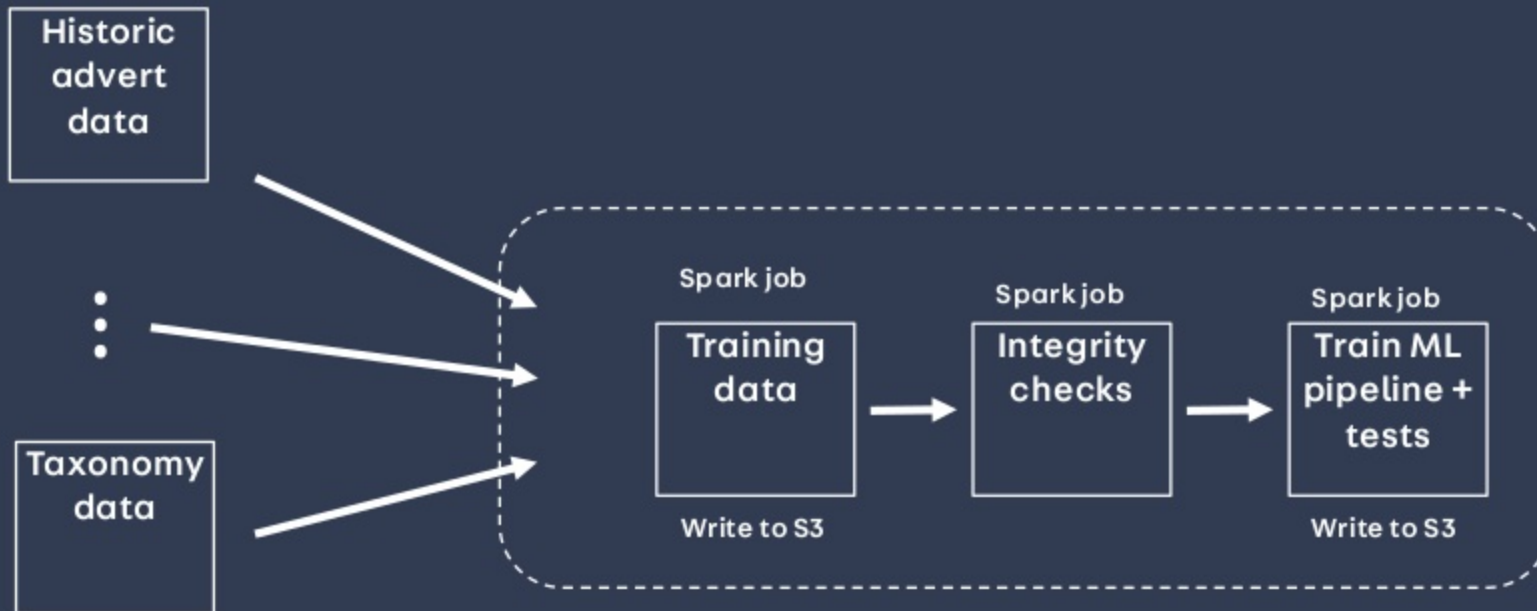
Automated model training

Converting to a Spark job



- Automated testing
- Integrated release automation
- Airflow integration
- Extra stage of code review

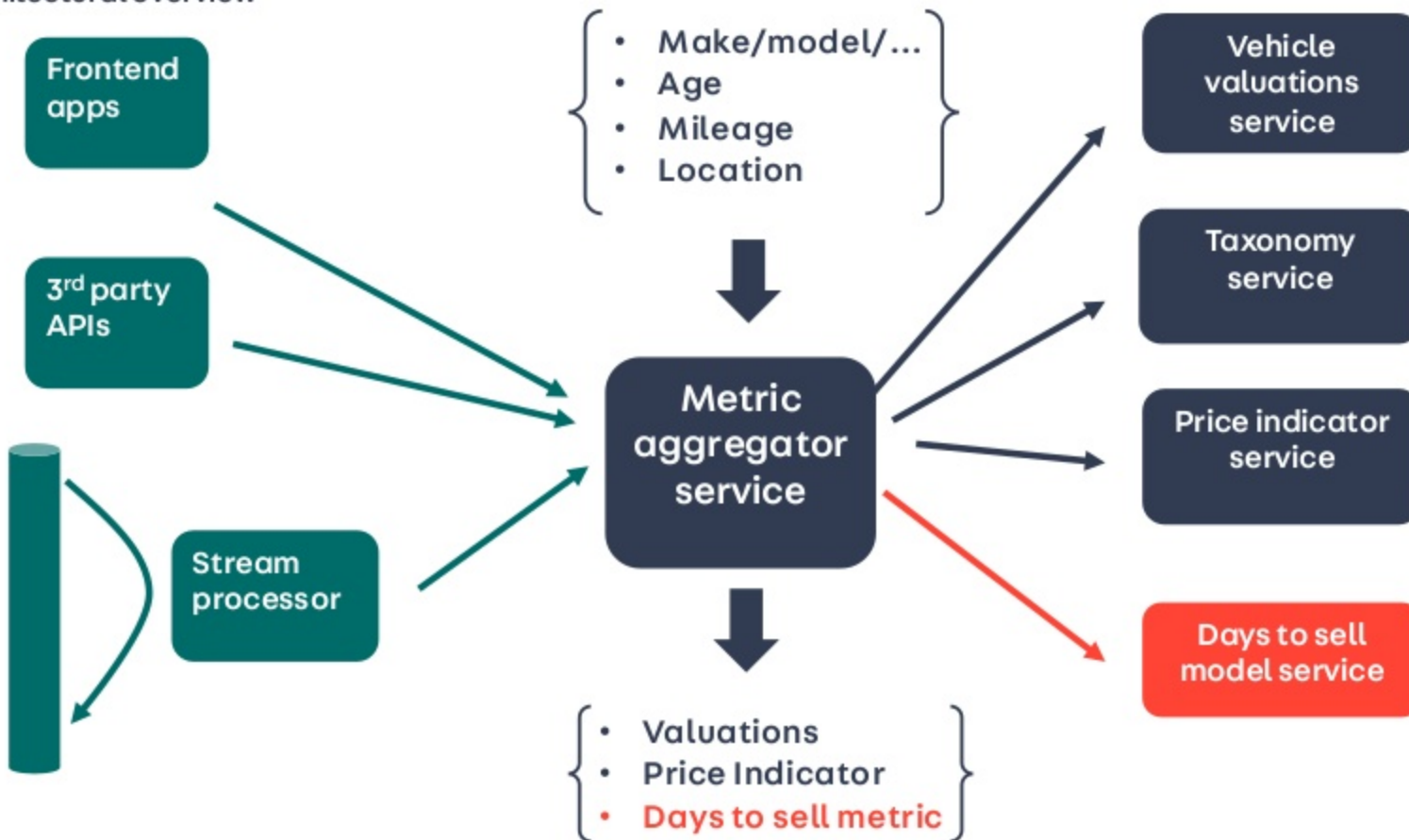
DAG Scheduling



Production phase

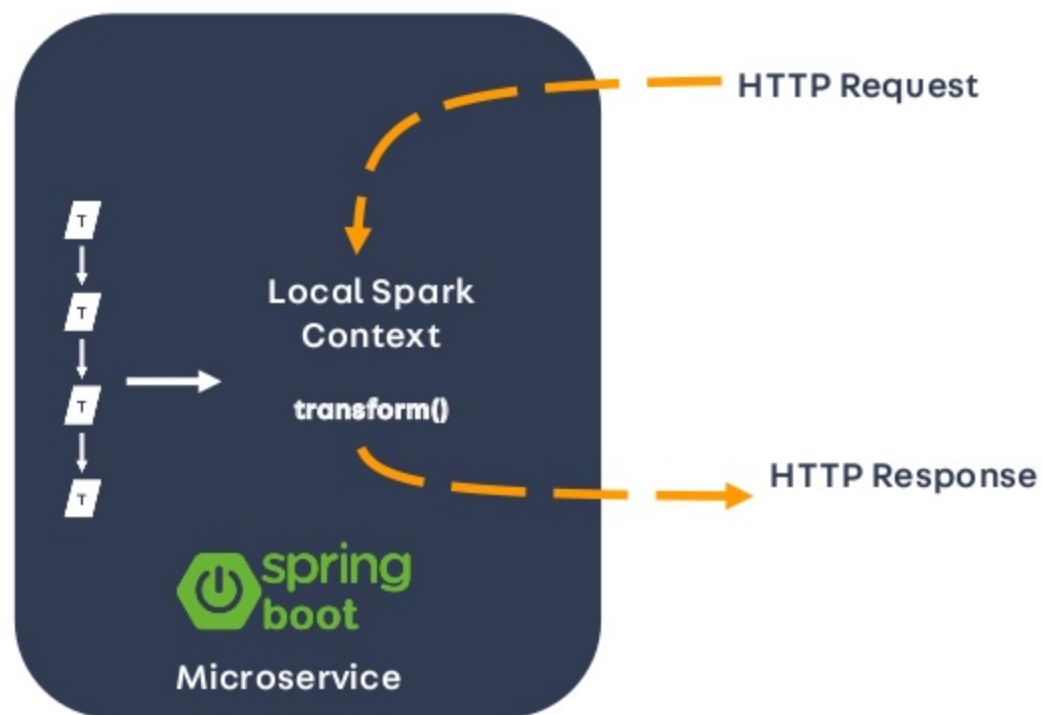
Automated model deployment

Architectural overview



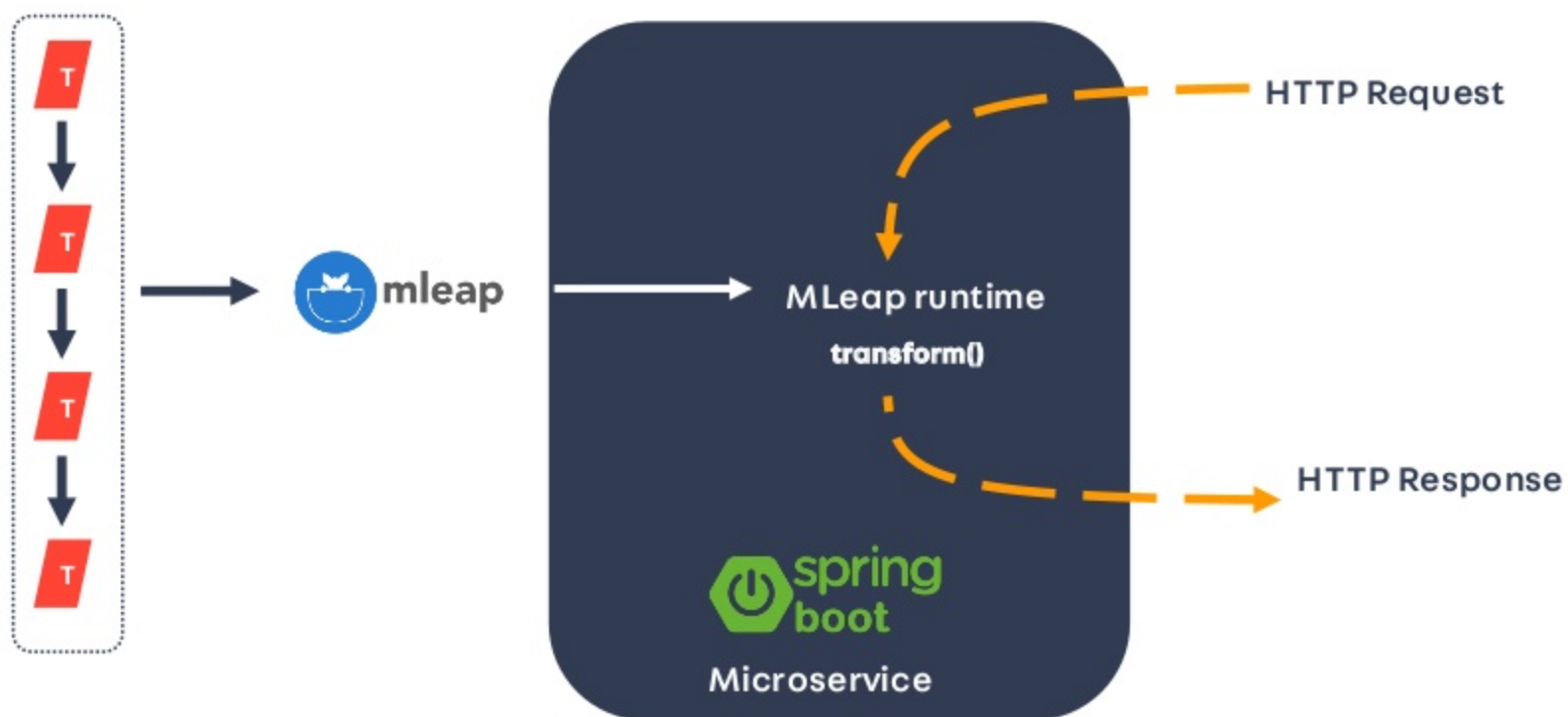
Automated model deployment

Online serving attempt 1

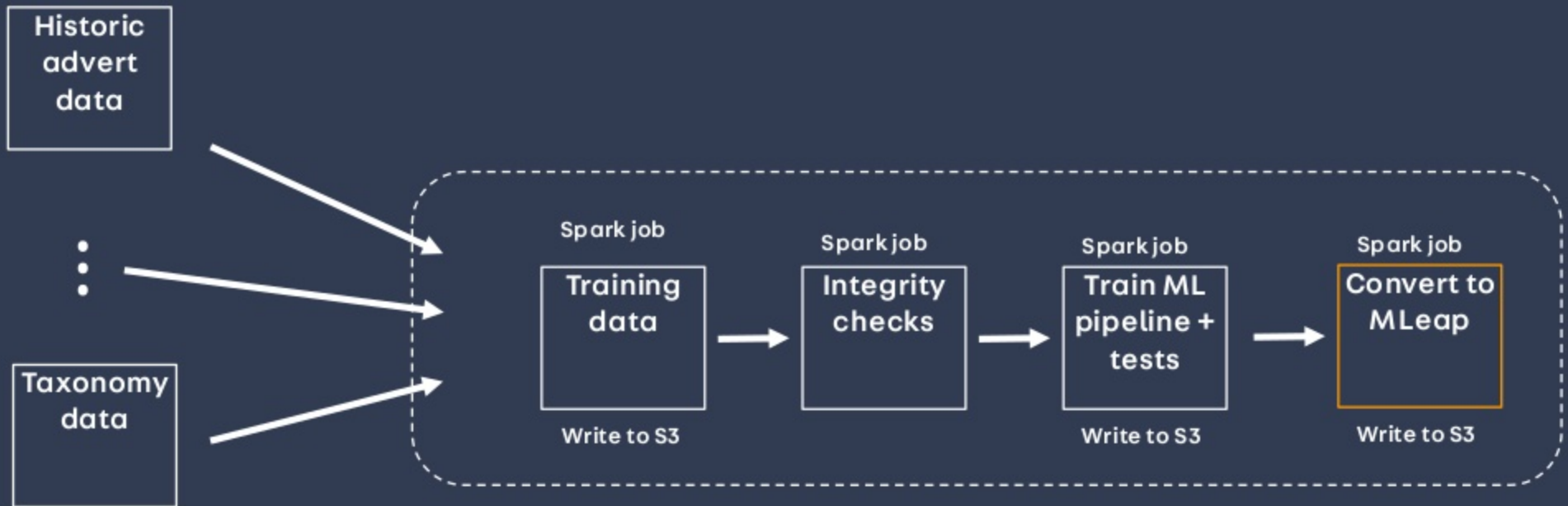


Automated model deployment

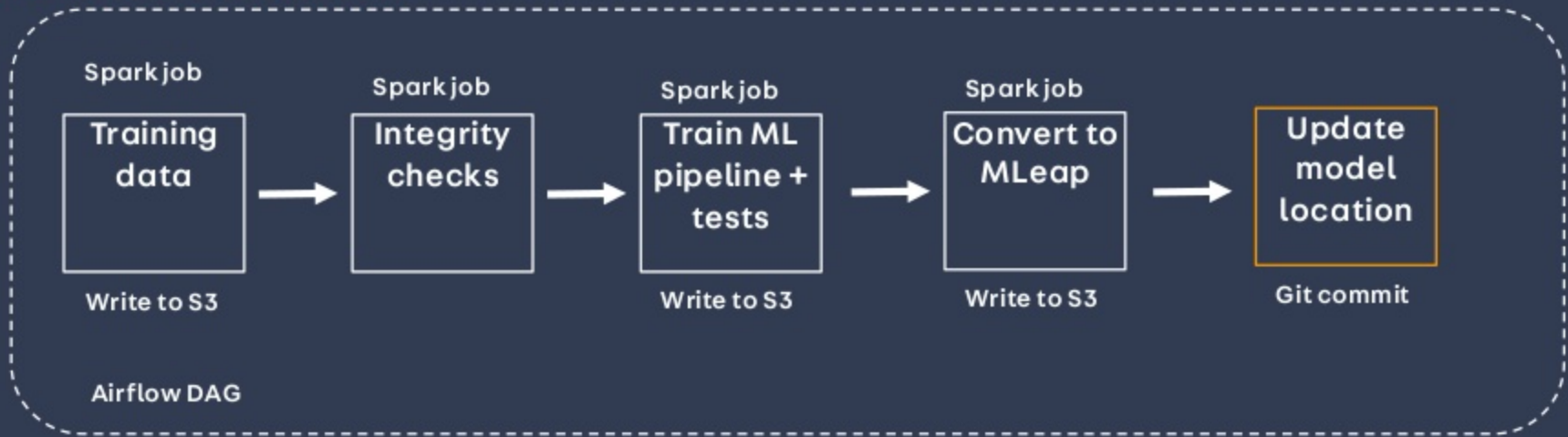
Online serving attempt 2



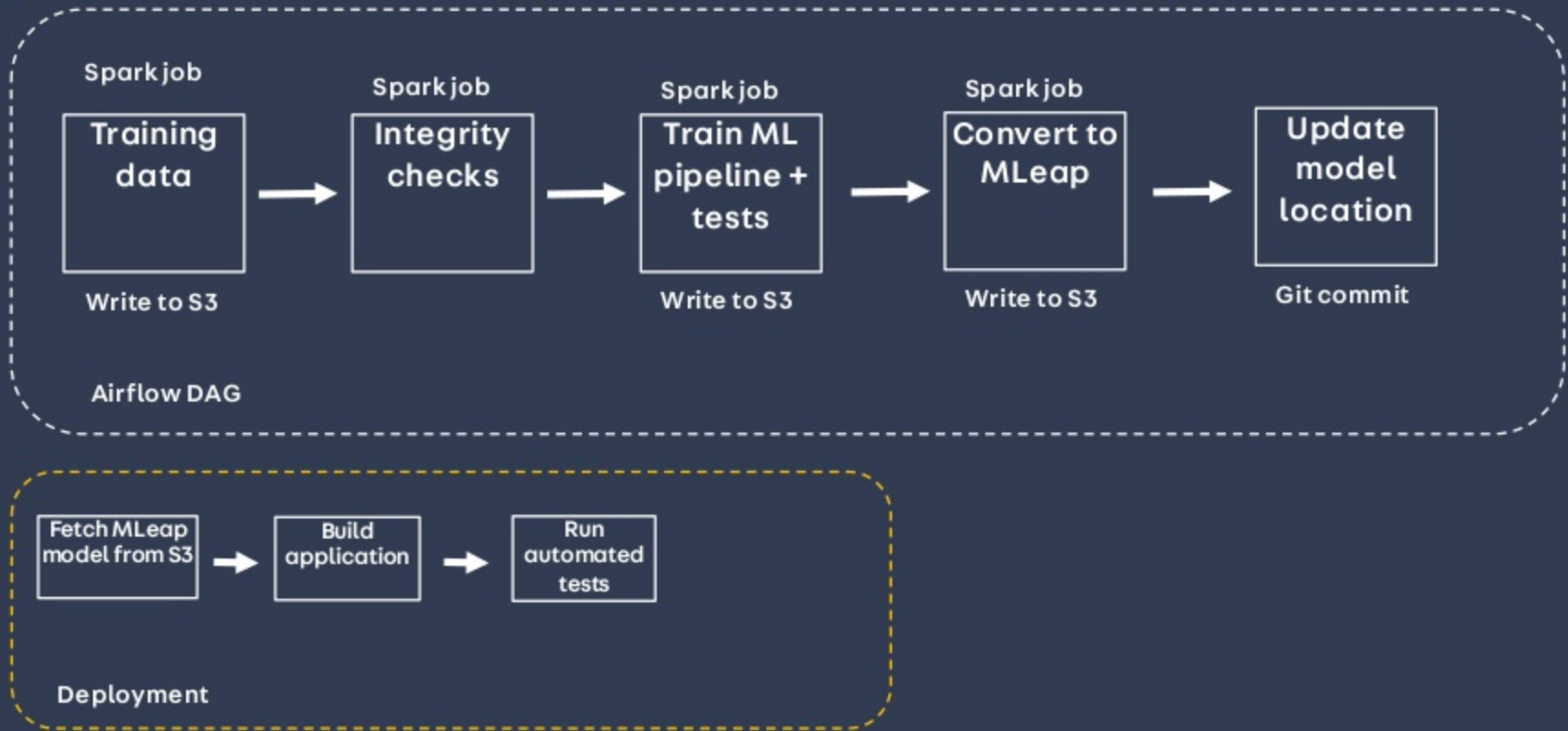
DAG Scheduling



Automated model deployment



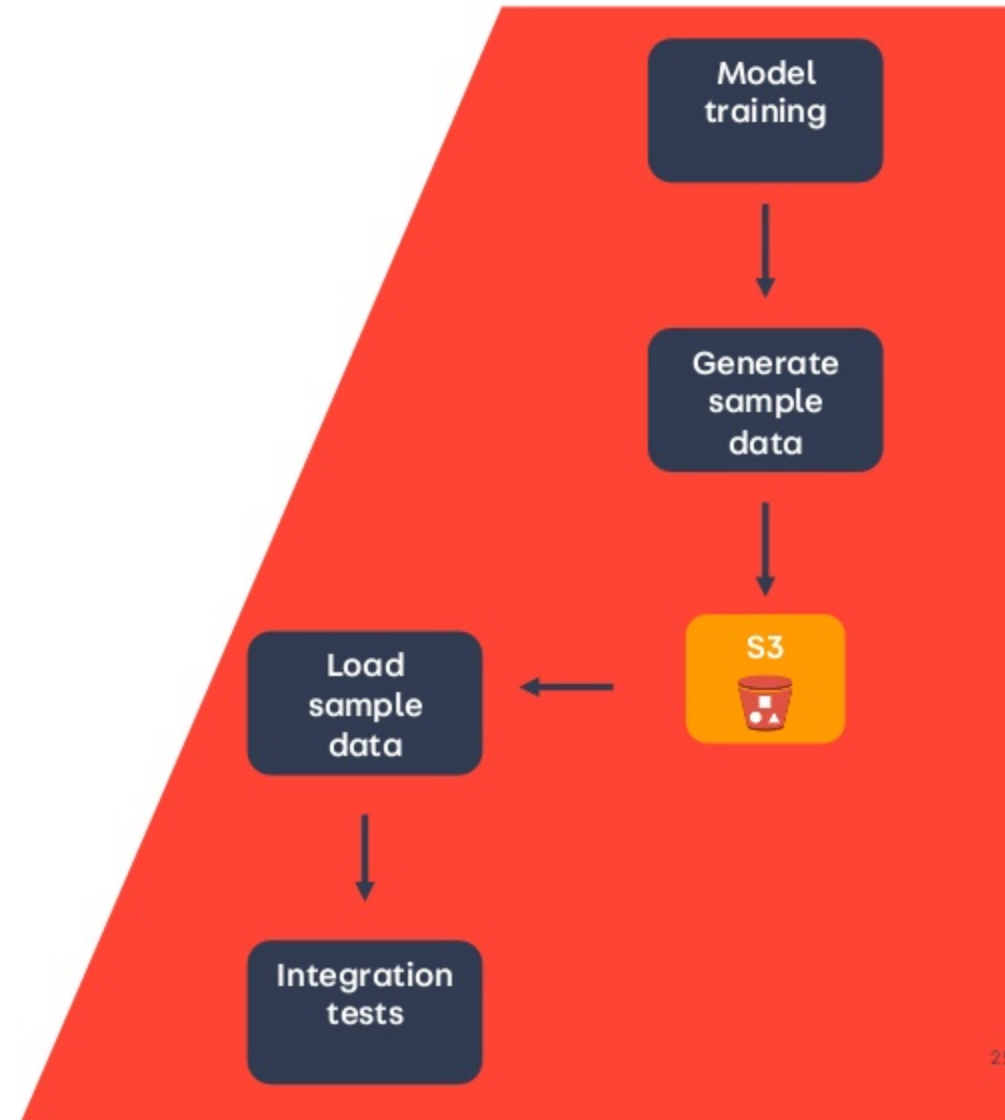
Automated model deployment



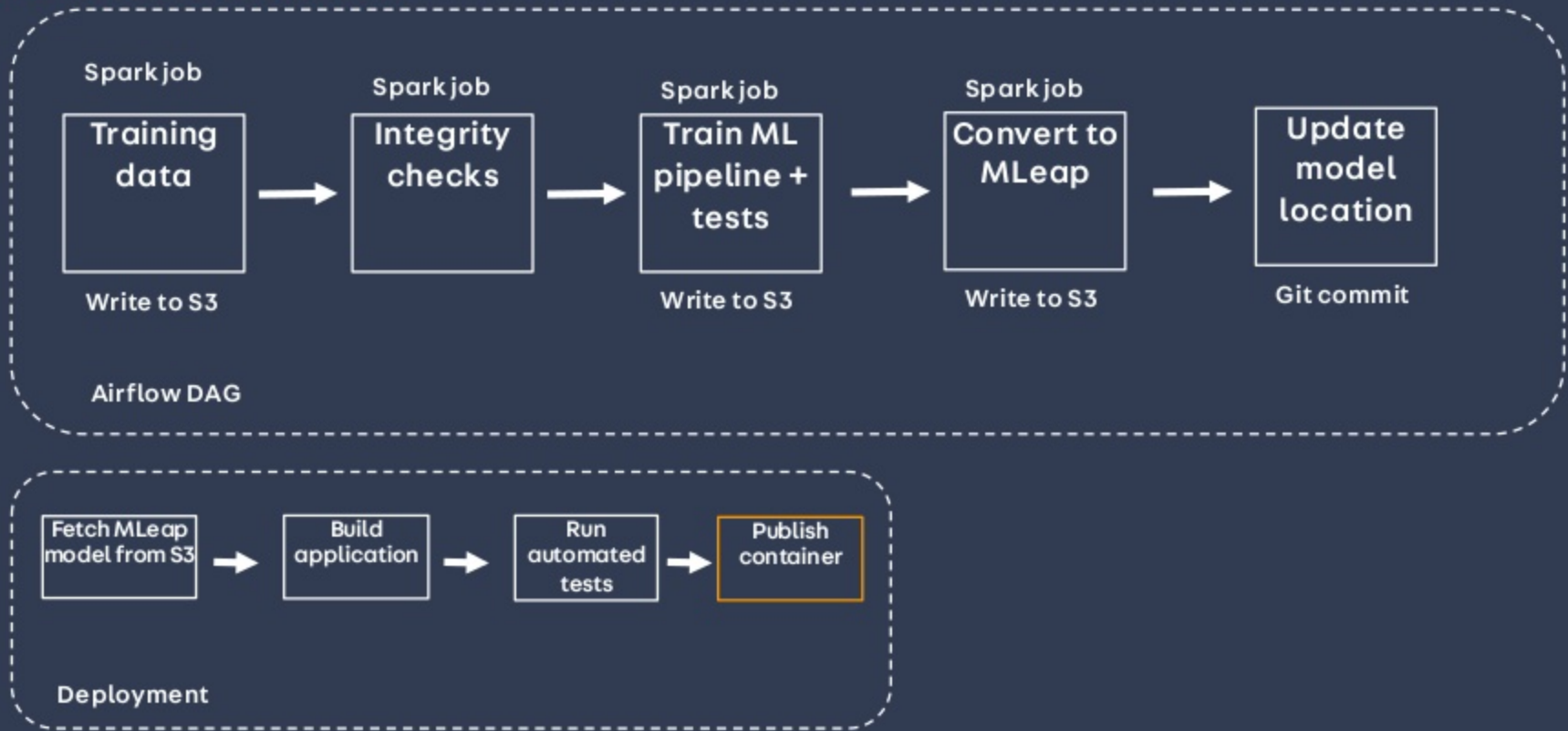
Automated model deployment

Testing Strategy

- **Unit tests ✓**
- **Integration tests?**



Automated model deployment



Automated deployment

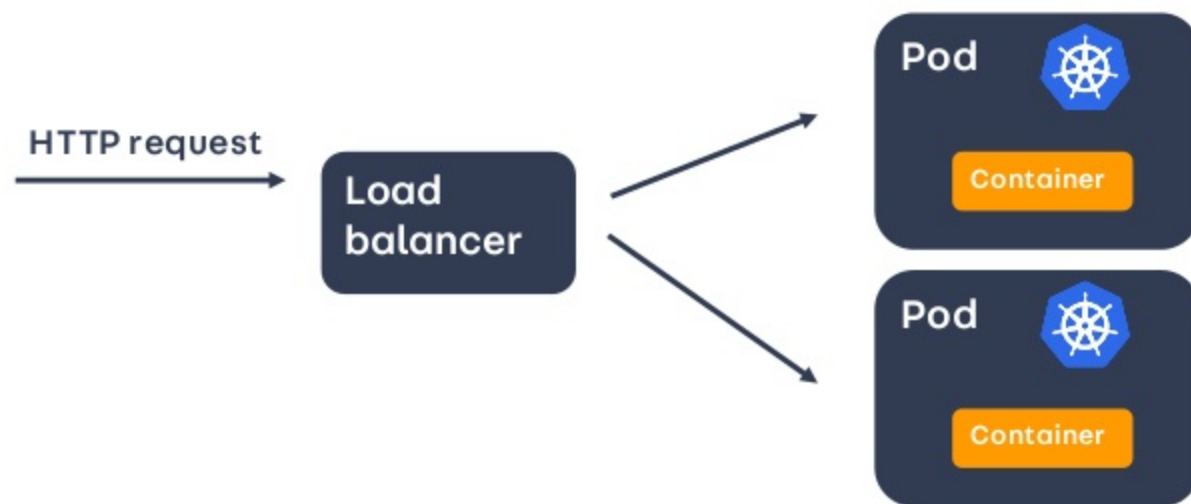
Brief introduction to Kubernetes

“Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications.”

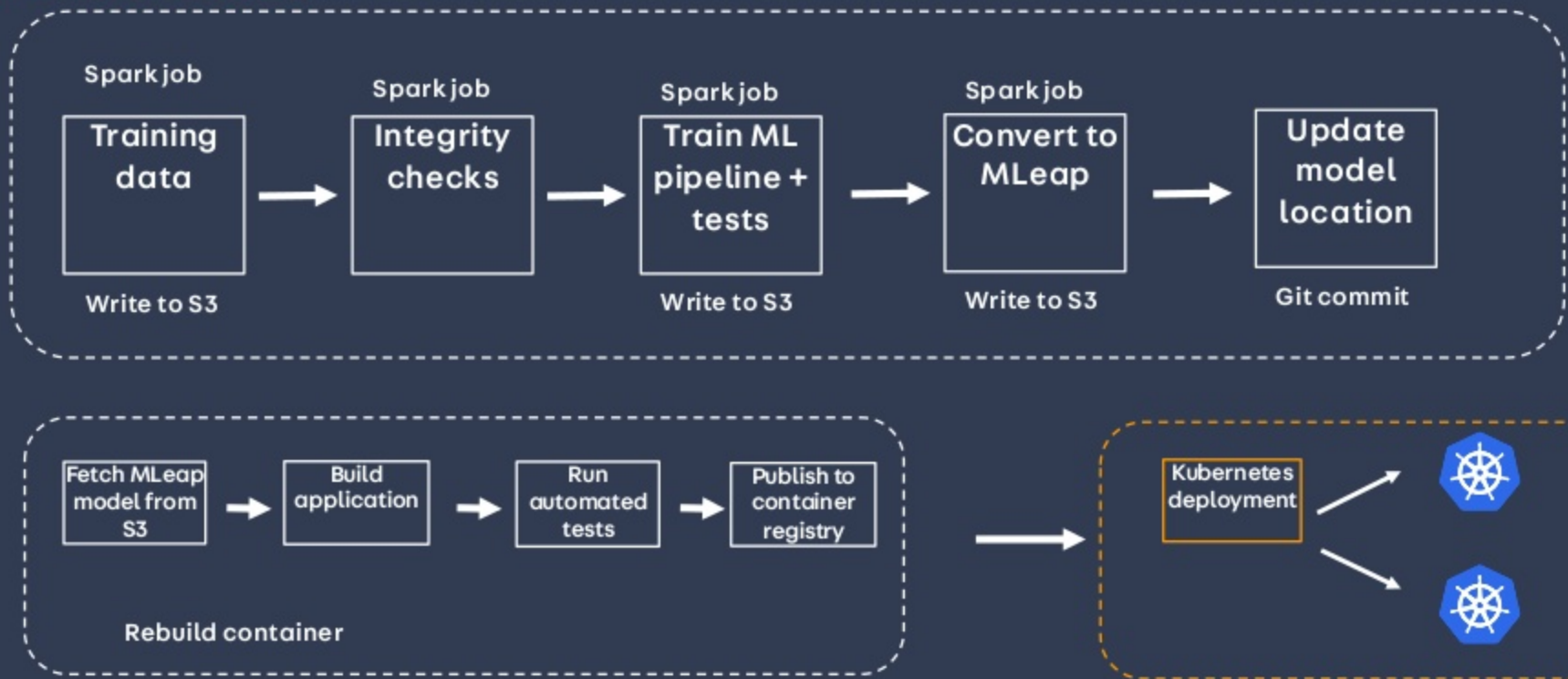


Automated deployment

Brief introduction to Kubernetes



Automated deployment



Summary

Discovery



Production



Summary

Discovery



Production



 @ed_kent

 **AutoTrader**

engineering.autotrader.co.uk

careers.autotrader.co.uk