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

AutoTrader

Edward Kent

Spark Summit Europe 2018

#SAISEnt9

AutoTrader   Buying new & used | Sell your car | Car reviews & more | Finance, insurance & more   **Sign in**

Home / Buying / Search Results

2,489 cars found

★ Save search   ⟳ Reset

| All | New | Used |

Sort by: Sponsored adverts first ▼

Read our latest reviews

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 >

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

£349
per month (PCP)
Finance example

CHORLEY GROUP

Get an insurance quote | Check its history

📷 20

WWW.PRESTIGE-MOTORSDIRECT.CO.UK

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

EXPERT REVIEW
2 months ago
Audi Q8 SUV (2018
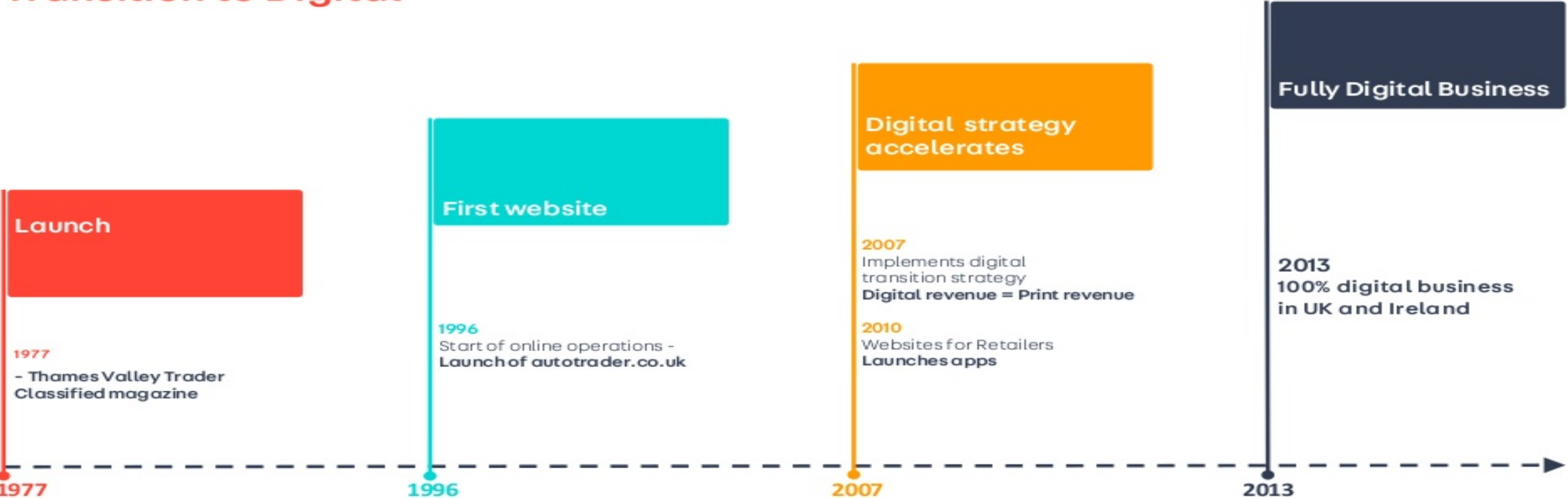-) review

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

Transition to Digital

Launch

First website

Digital strategy accelerates

Fully Digital Business

1977

- Thames Valley Trader Classified magazine

1996
Start of online operations - Launch of autotrader.co.uk

2007
Implements digital transition strategy
Digital revenue = Print revenue

2010
Websites for Retailers
Launches apps

2013
100% digital business in UK and Ireland

1977   **1996**   **2007**   **2013**

AutoTrader

3

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

AutoTrader

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

AutoTrader

Services for retailers

 Classified advertising

 Valuations

 Finance solutions

 Creating a trusted marketplace

 Forecourt management tools

 Retailer education and insight

AutoTrader

Data at Auto Trader
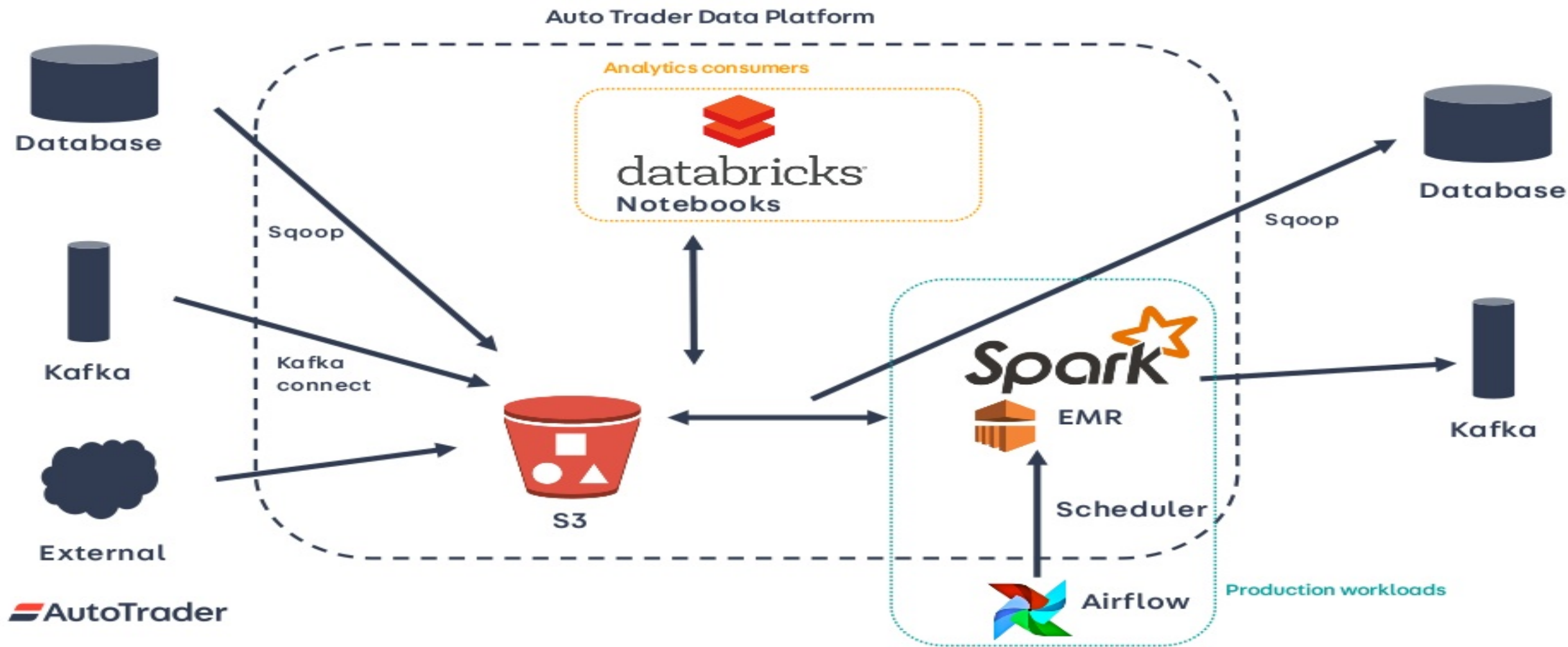
WORK IN PROGRESS

< 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

Production phase

Automated model training **Automated model deployment**

- Quick iterations
- Fast feedback
- Offline predictions

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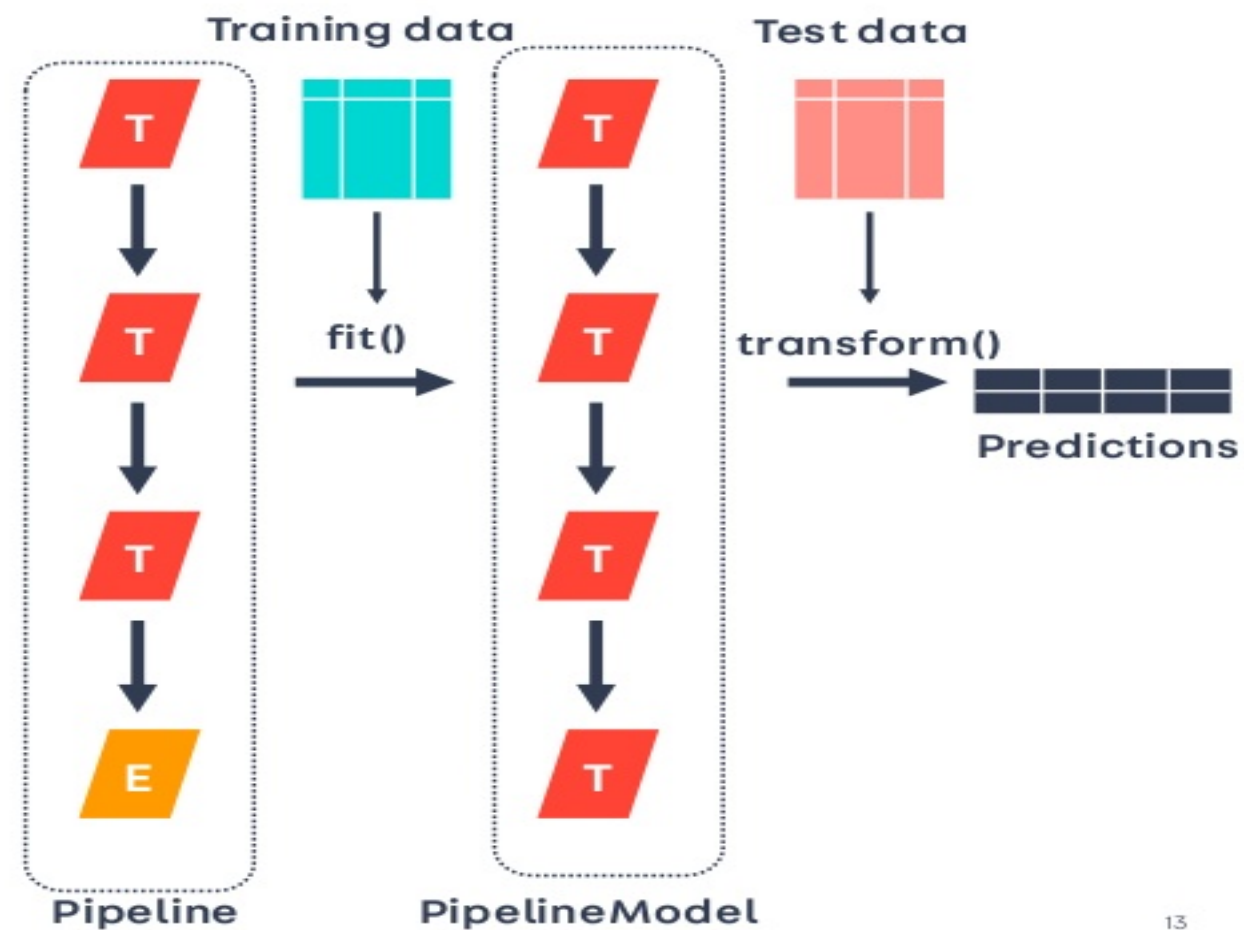
AutoTrader

Discovery phase

AutoTrader

The Model

Spark ML Pipeline

Transformers

Estimators

Training data

Test data

fit()

transform()

Predictions

Pipeline

PipelineModel

AutoTrader

13

The Model

Days to Sell ML Pipeline

Historic advert data →

Vehicle taxonomy →

Valuations →

Geographic data →

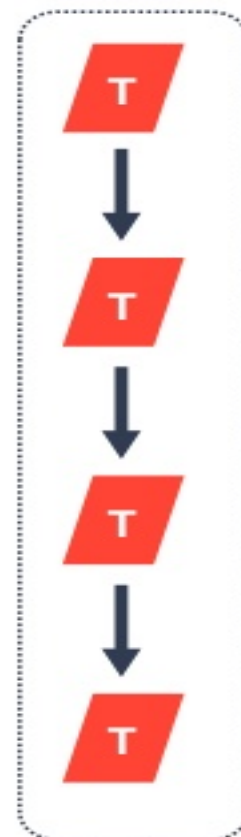**Training data**
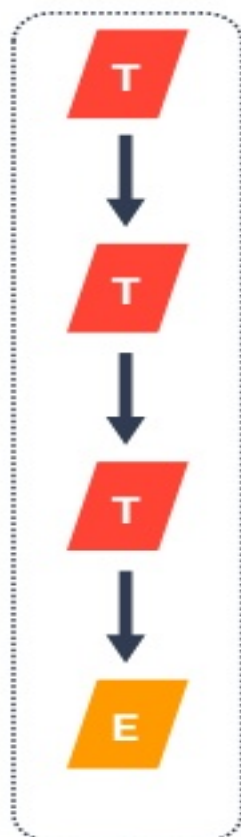
Pipeline

fit()

PipelineModel

Make/model/...
Age
Valuation
Location

transform()

Days to sell

Prediction

AutoTrader

14

Production phase
Automated model training

Automated model training

Converting to a Spark job

databricks
Notebook

Spark
Self-contained JAR

Scheduler

Airflow

- Automated testing
- Integrated release automation
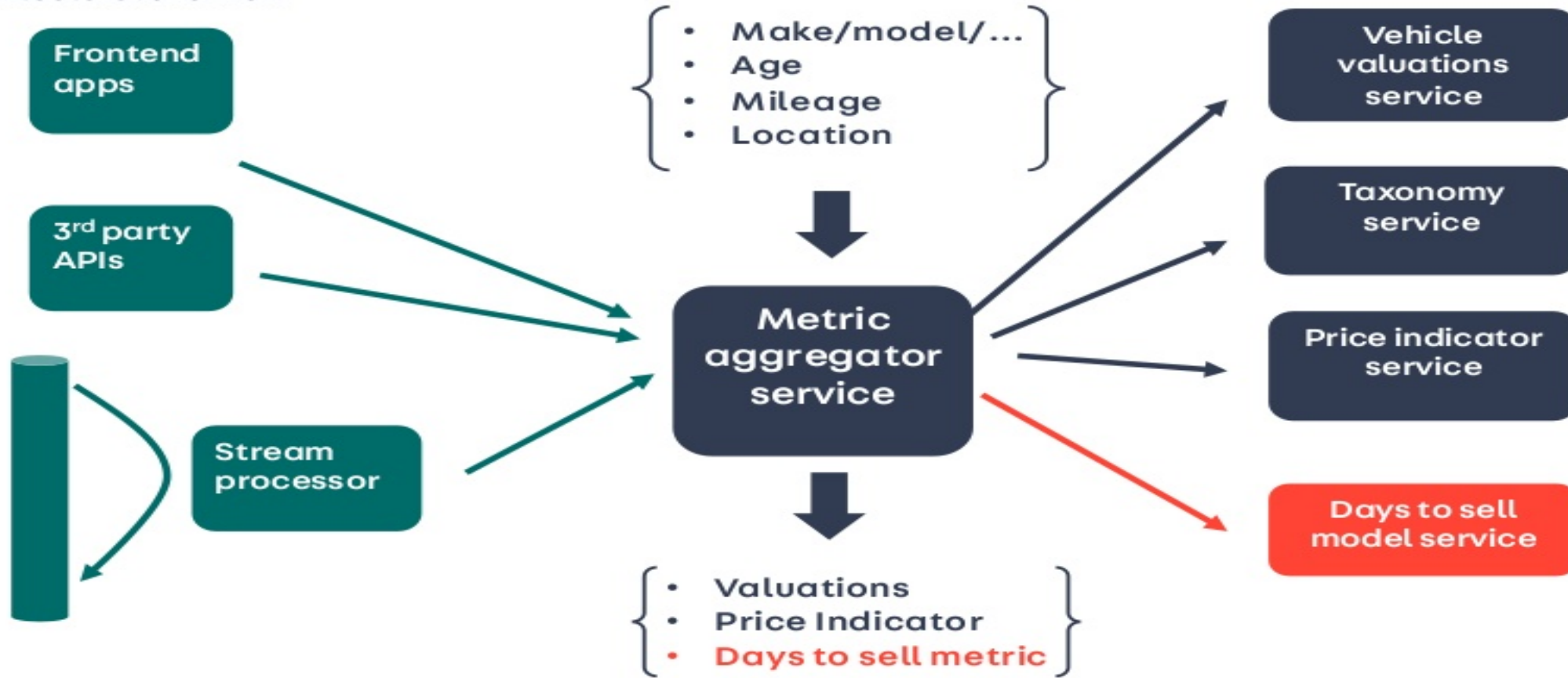- Airflow integration
- Extra stage of code review
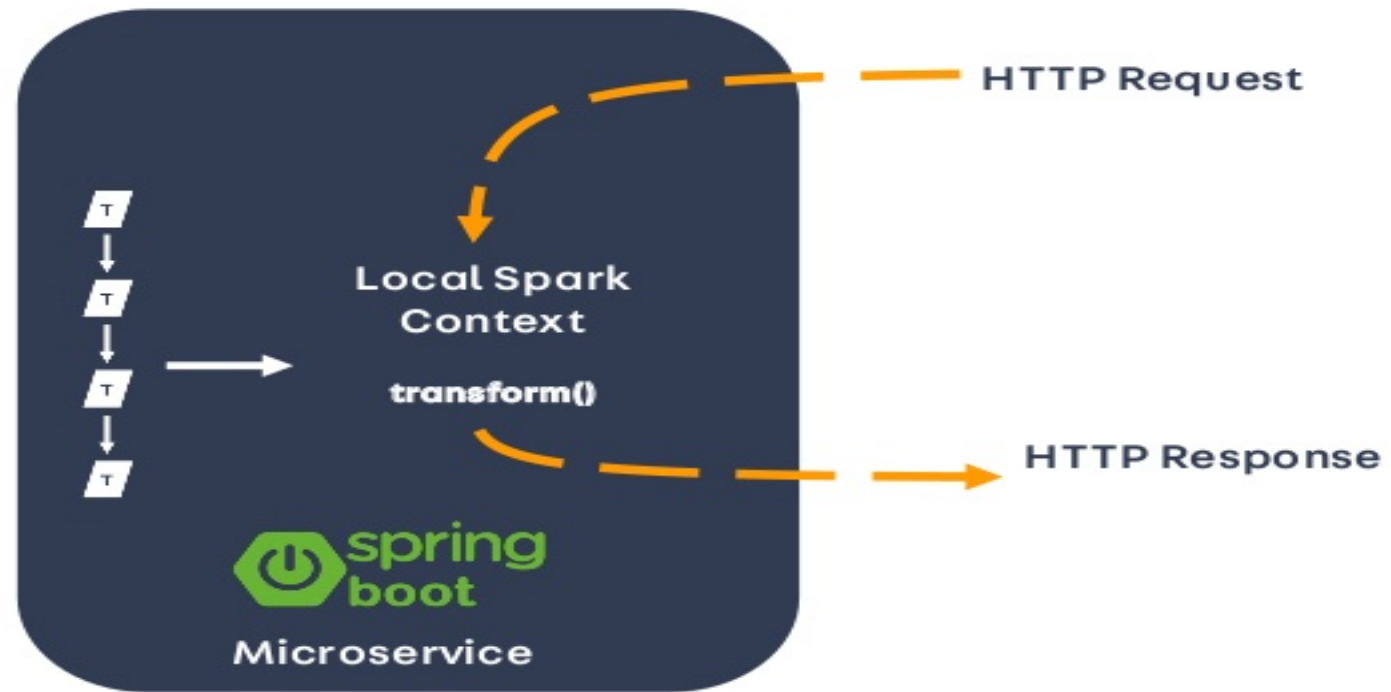
AutoTrader
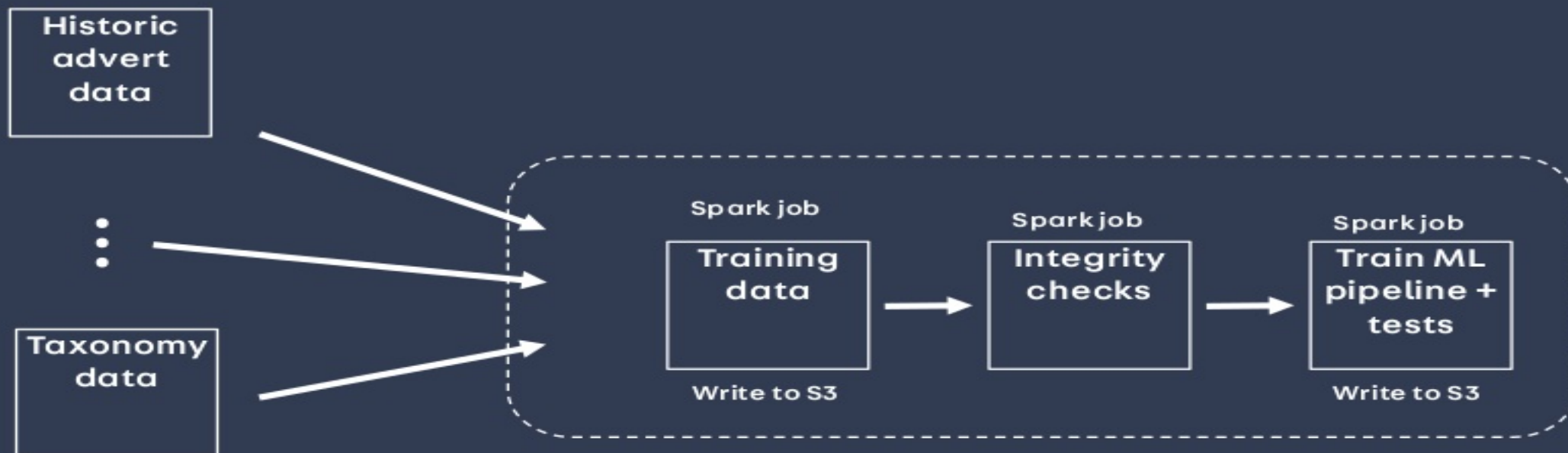
Production phase
Automated model deployment
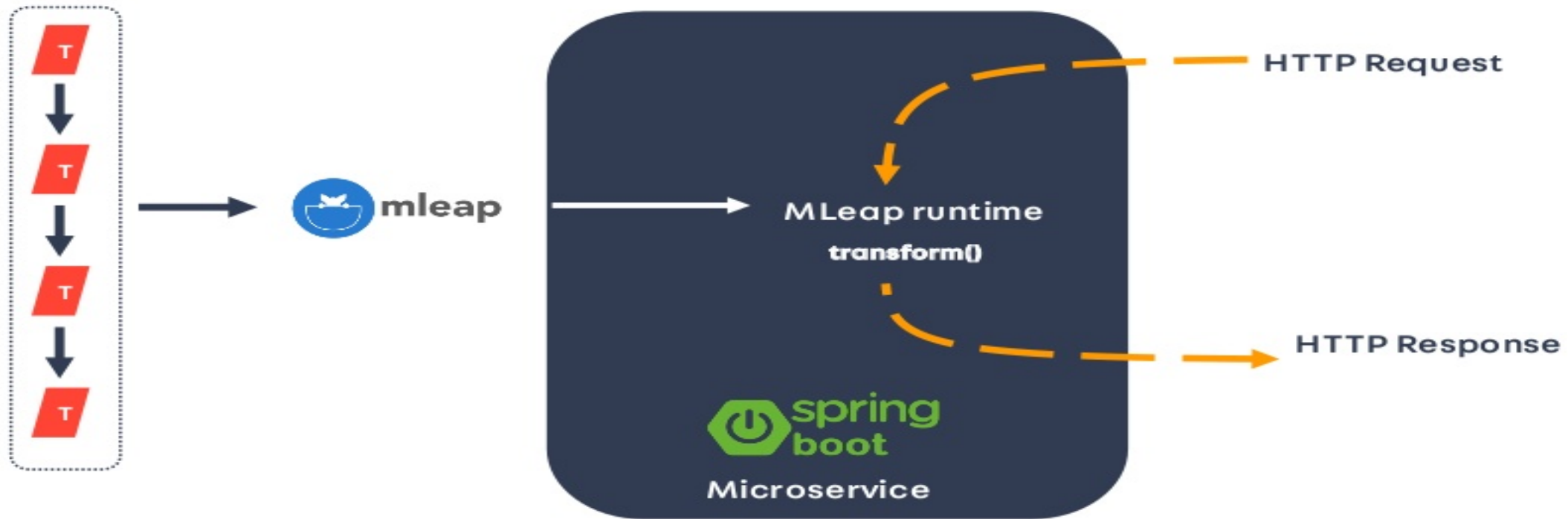
Architectural overview

- Frontend apps
- 3rd party APIs
- Stream processor

Make/model/...
Age
Mileage
Location

Metric aggregator service

- Vehicle valuations service
- Taxonomy service
- Price indicator service
- Days to sell model service

- Valuations
- Price Indicator
- Days to sell metric

AutoTrader

19
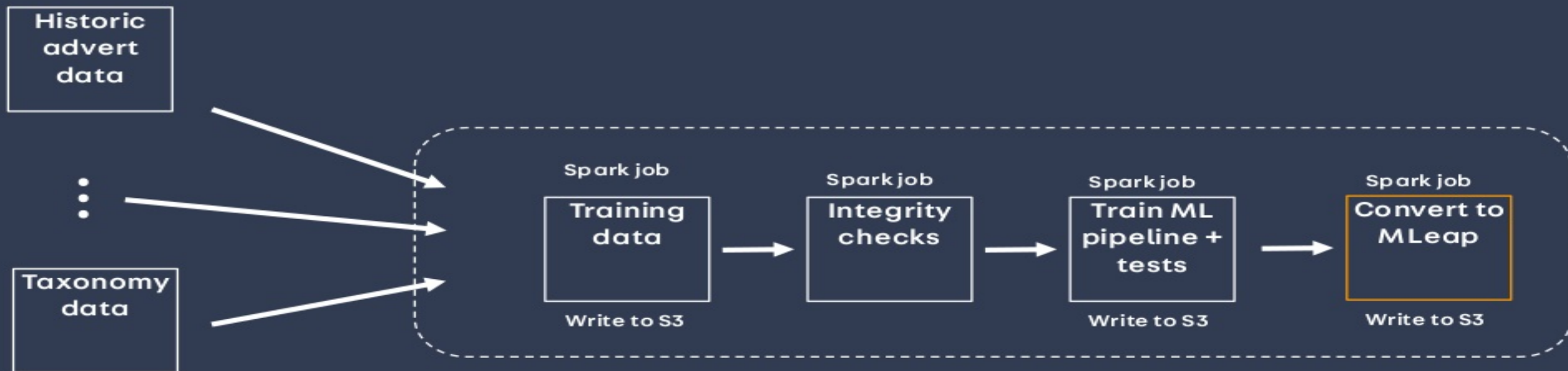
Automated model deployment

Online serving attempt 2
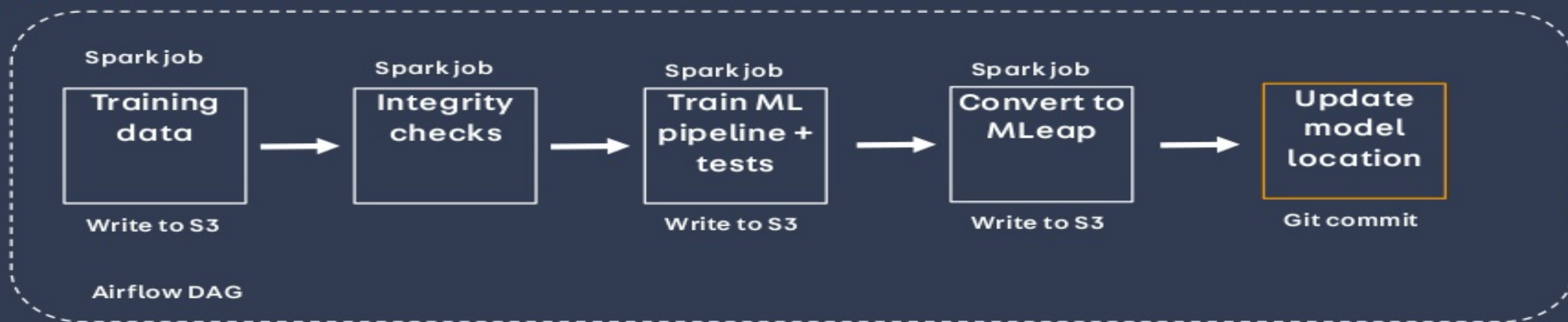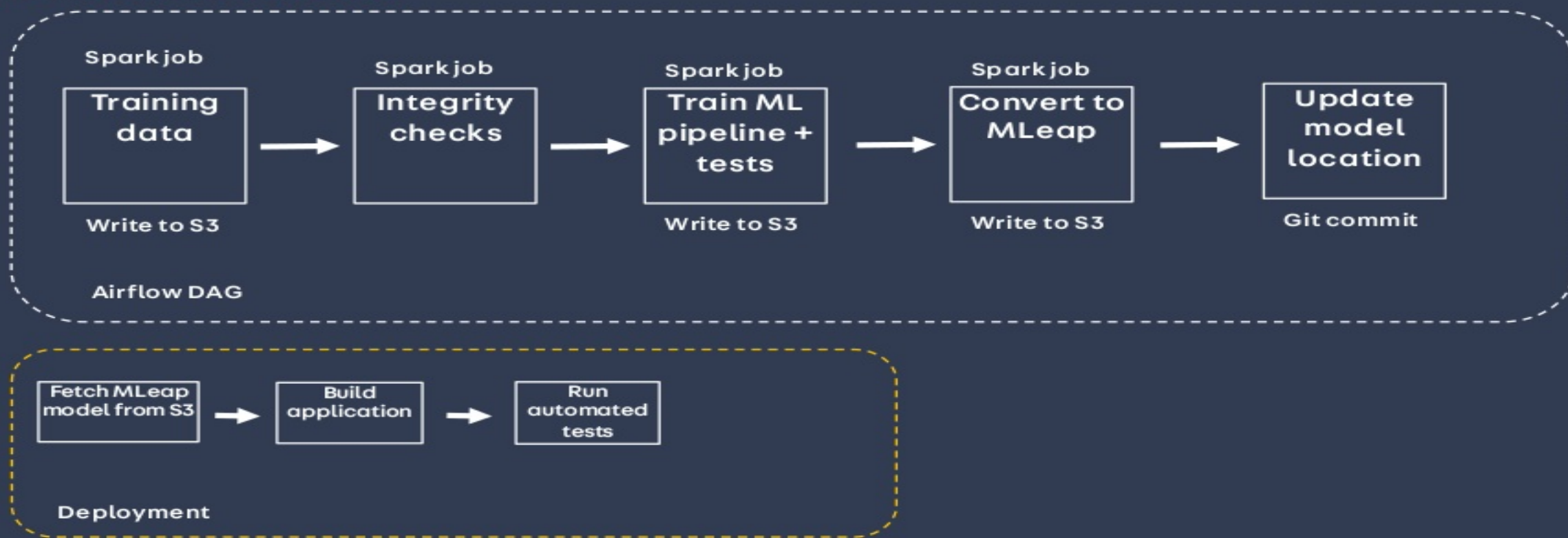
Automated model deployment

Testing Strategy

- # Unit tests ✓

- # Integration tests?

AutoTrader

Automated model deployment

Airflow DAG

Spark job	Spark job	Spark job	Spark job	
Training data →	**Integrity checks** →	**Train ML pipeline + tests** →	**Convert to MLeap** →	**Update model location**
Write to S3		Write to S3	Write to S3	Git commit

Deployment

Fetch MLeap model from S3 → Build application → Run automated tests → Publish container

AutoTrader
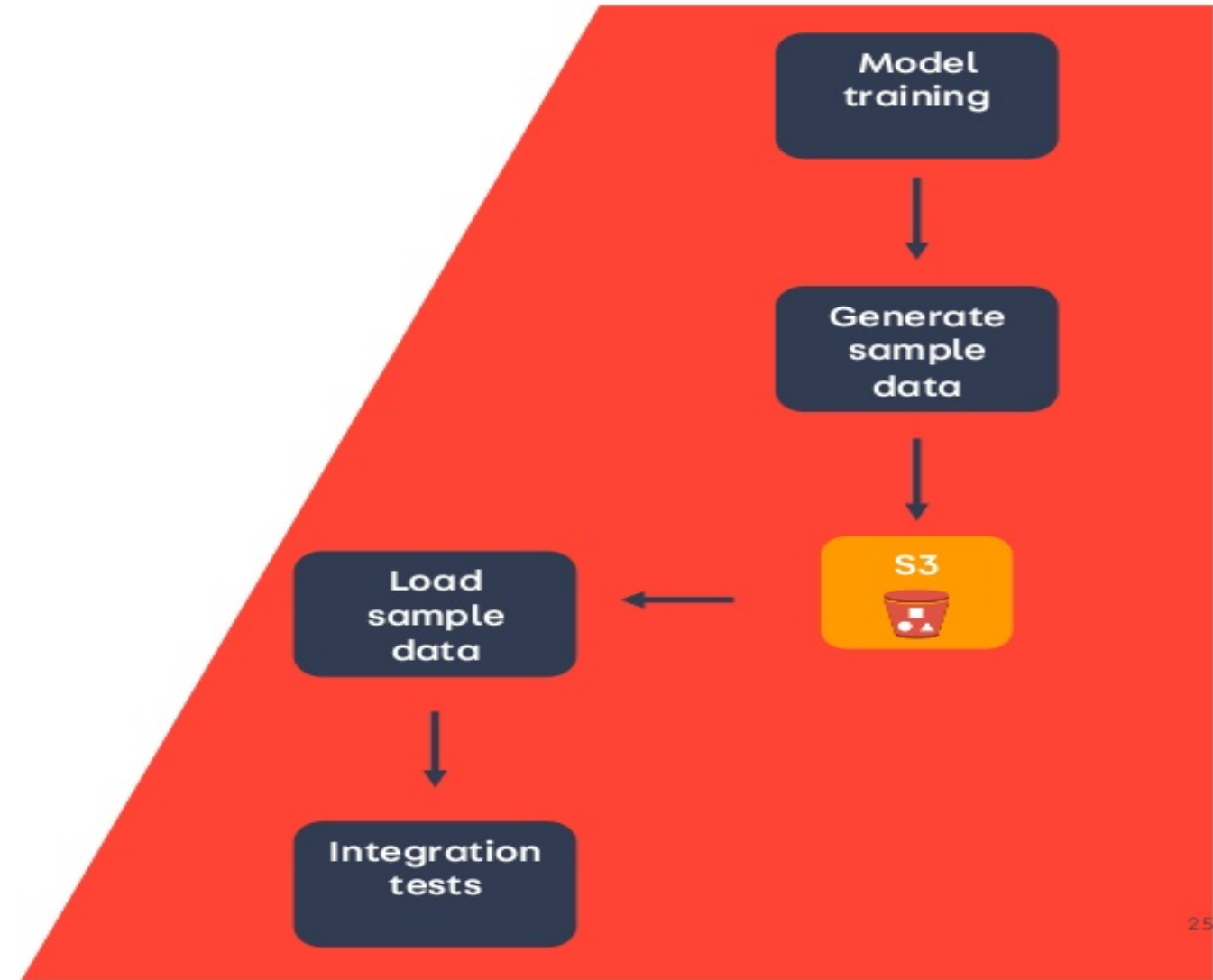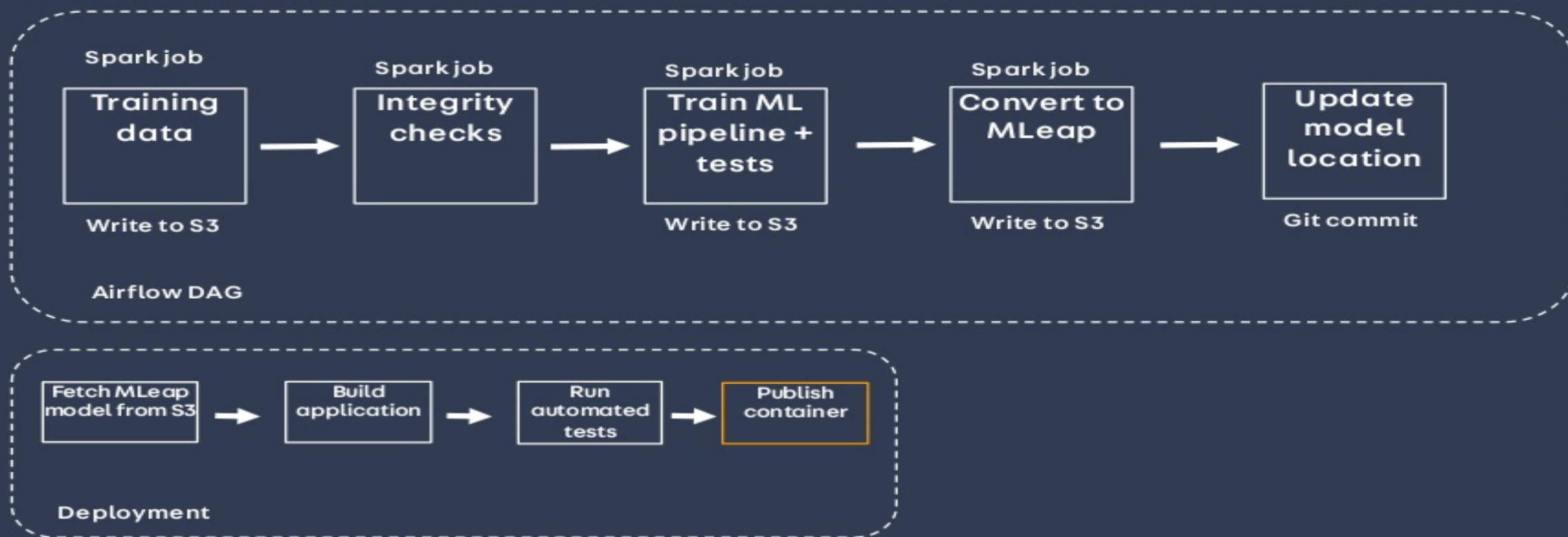
Brief introduction to Kubernetes

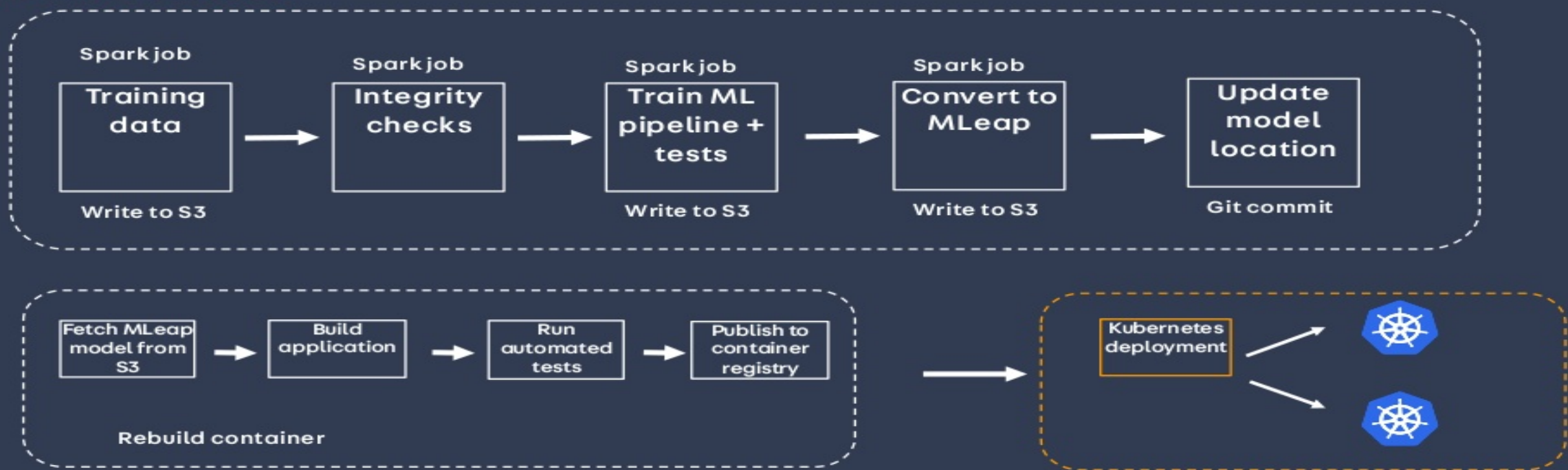"Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications."

kubernetes

Automated deployment

Spark job
Training data → Spark job **Integrity checks** → Spark job **Train ML pipeline + tests** → Spark job **Convert to MLeap** → **Update model location**

Write to S3 — Write to S3 — Write to S3 — Git commit

Fetch MLeap model from S3 → Build application → Run automated tests → Publish to container registry

Rebuild container

Kubernetes deployment

AutoTrader

Summary

Discovery

Production

Airflow

Model training

Deployment

kubernetes

spring boot

mleap

docker

databricks Notebooks

Spark

mleap

AutoTrader

30

@ed_kent

AutoTrader

engineering.autotrader.co.uk

careers.autotrader.co.uk