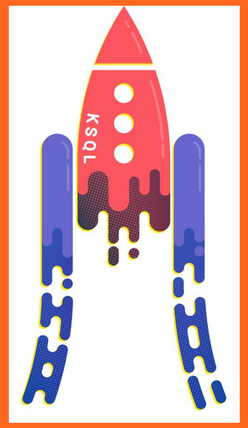
confluent

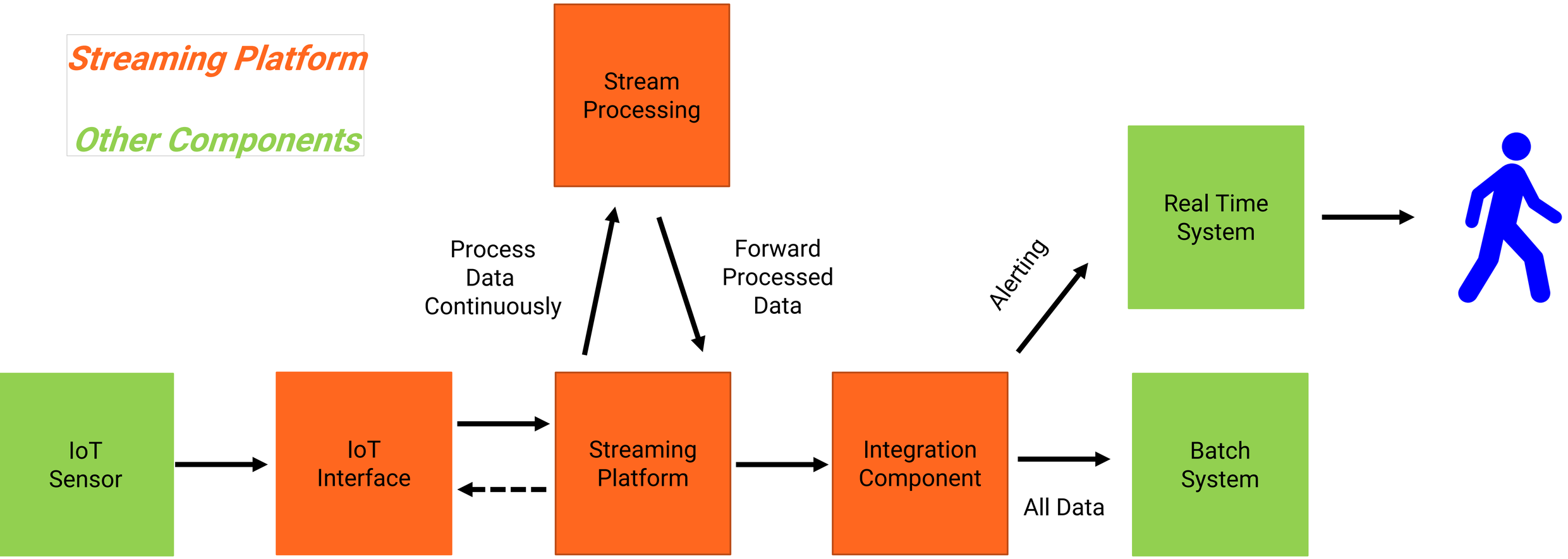
IoT Sensor Analytics with Apache Kafka, KSQL and TensorFlow

Kafka-Native End-to-End IoT Data Integration and Processing

Kai Waehner - Technology Evangelist
kontakt@kai-waehner.de - LinkedIn – Twitter : @KaiWaehner
www.confluent.io
www.kai-waehner.de



Internet of Things – End-to-End Data Processing



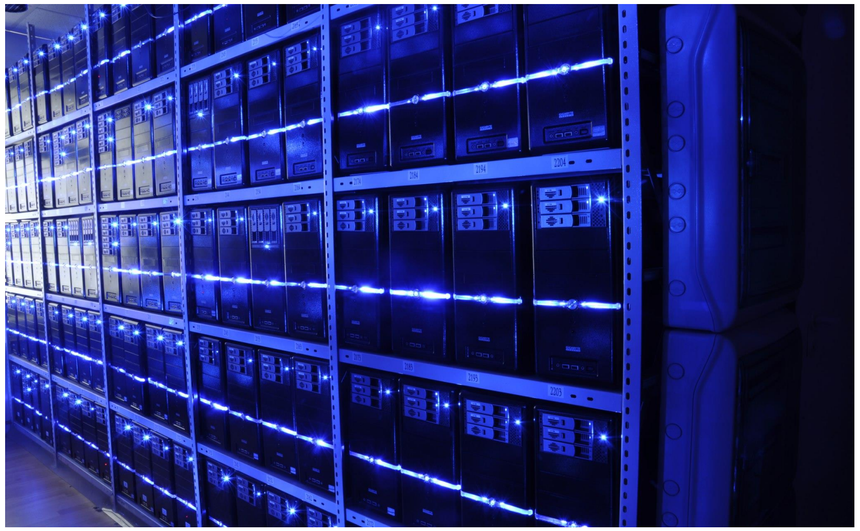
The first analytic model

python

TensorFlow

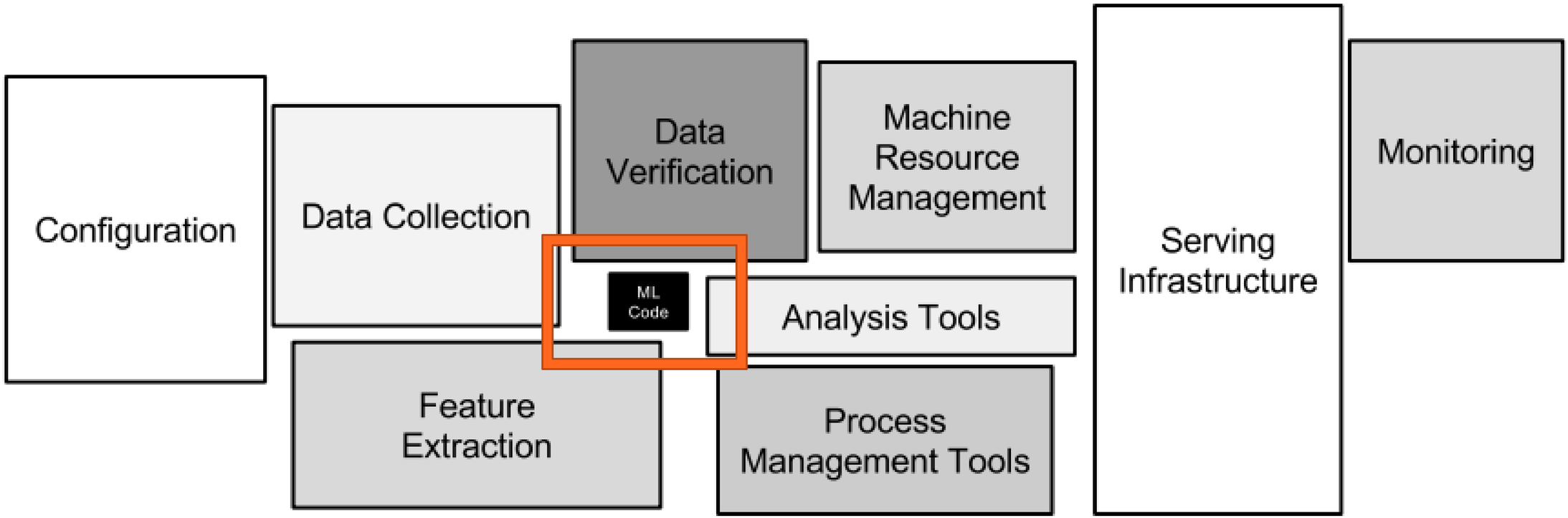
Keras

```
def queen(queens):
    for i in range(BOARD_SIZE):
        test_queens = queens + [i]
        try:
            validate(test_queens)
            if len(test_queens) == BOARD_SIZE:
                return test_queens
            else:
                return add_queen(test_queens)
        except BailOut:
```



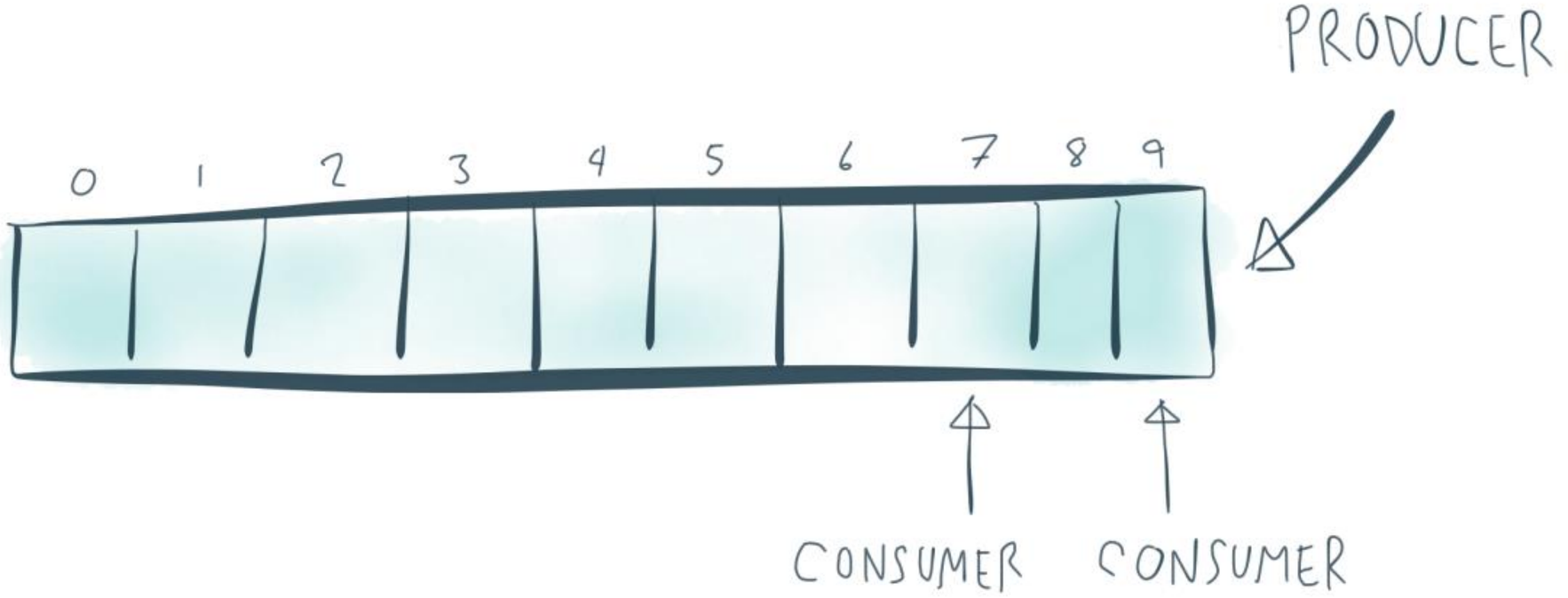
How to deploy the models
in production?
…real-time processing?
…at scale?
…24/7 zero downtime?

Hidden Technical Debt in Machine Learning Systems

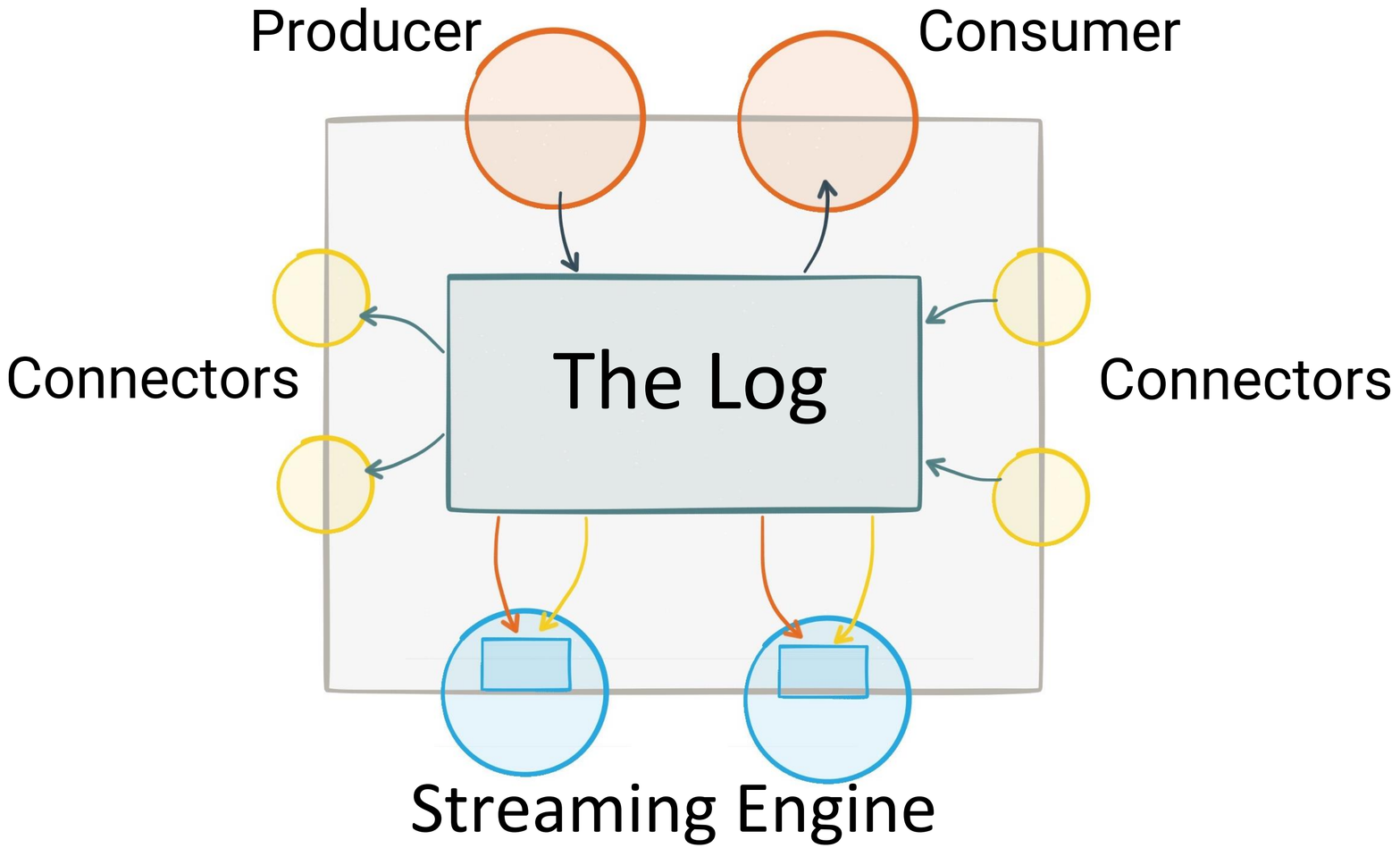


https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf

kafka



Apache Kafka – The Rise of an Event Streaming Platform



kafka

1 PUB/SUB

2 STORE

3 PROCESS

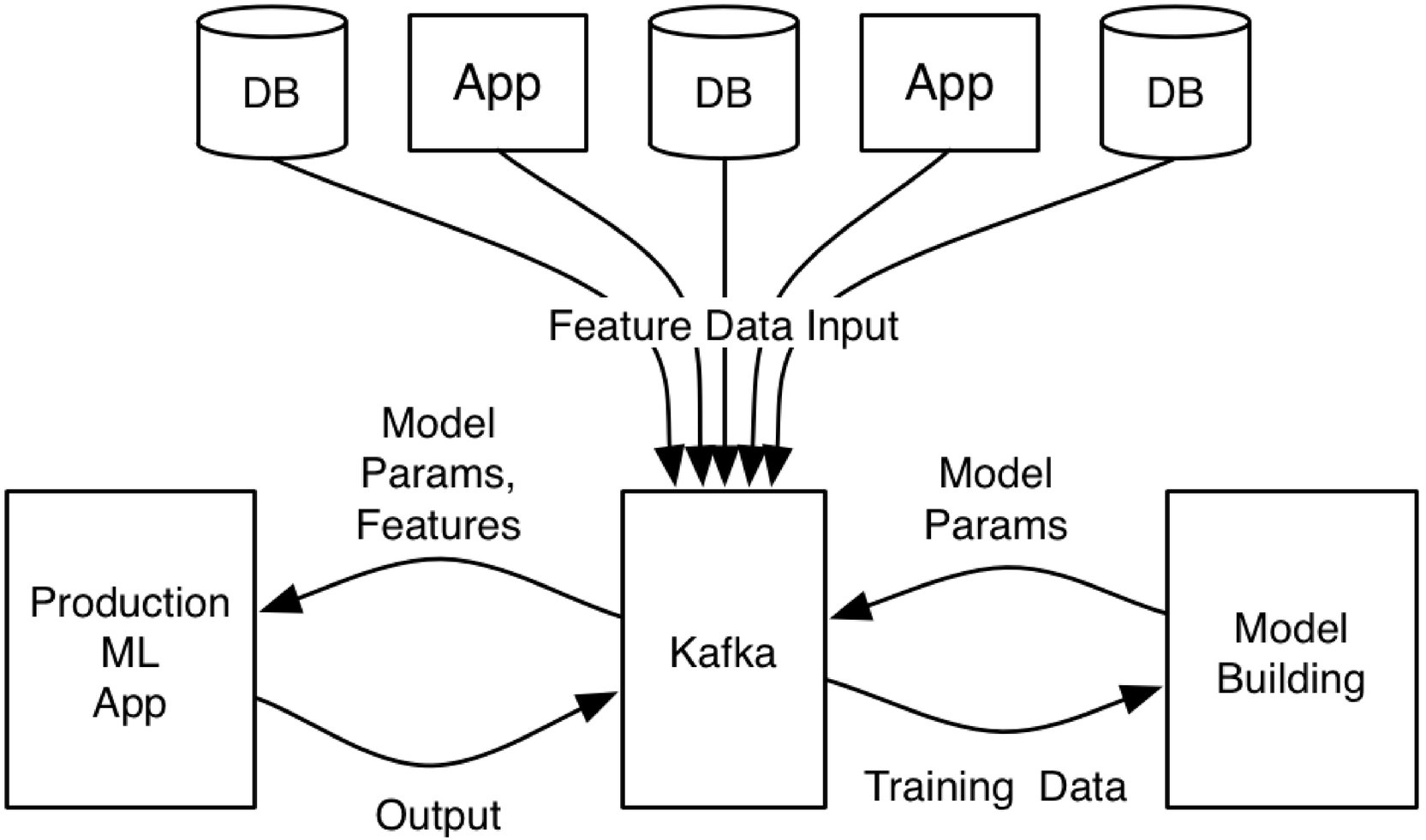
Apache Kafka at Scale at Tech Giants

> 4.5 trillion messages / day

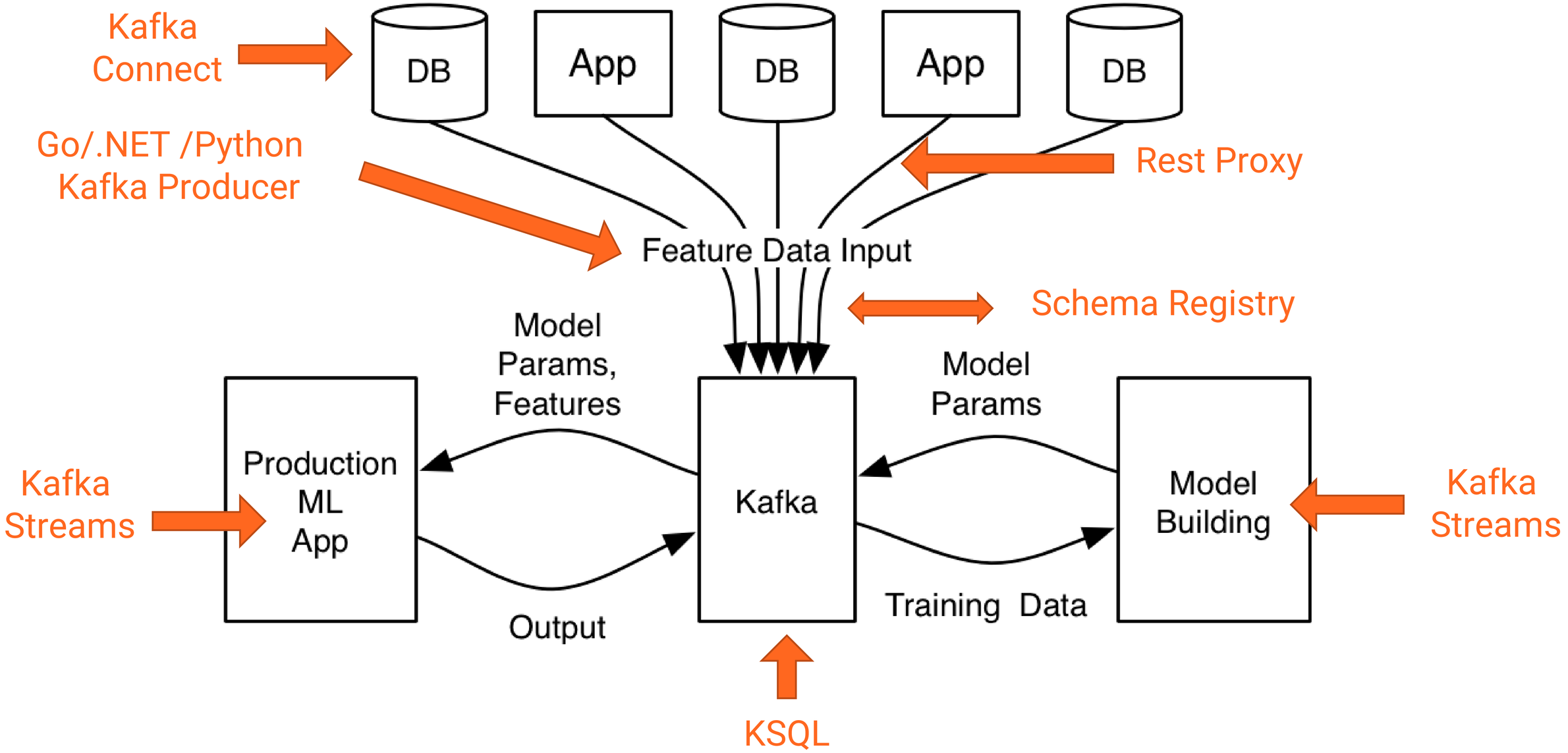
> 6 Petabytes / day

“You name it”

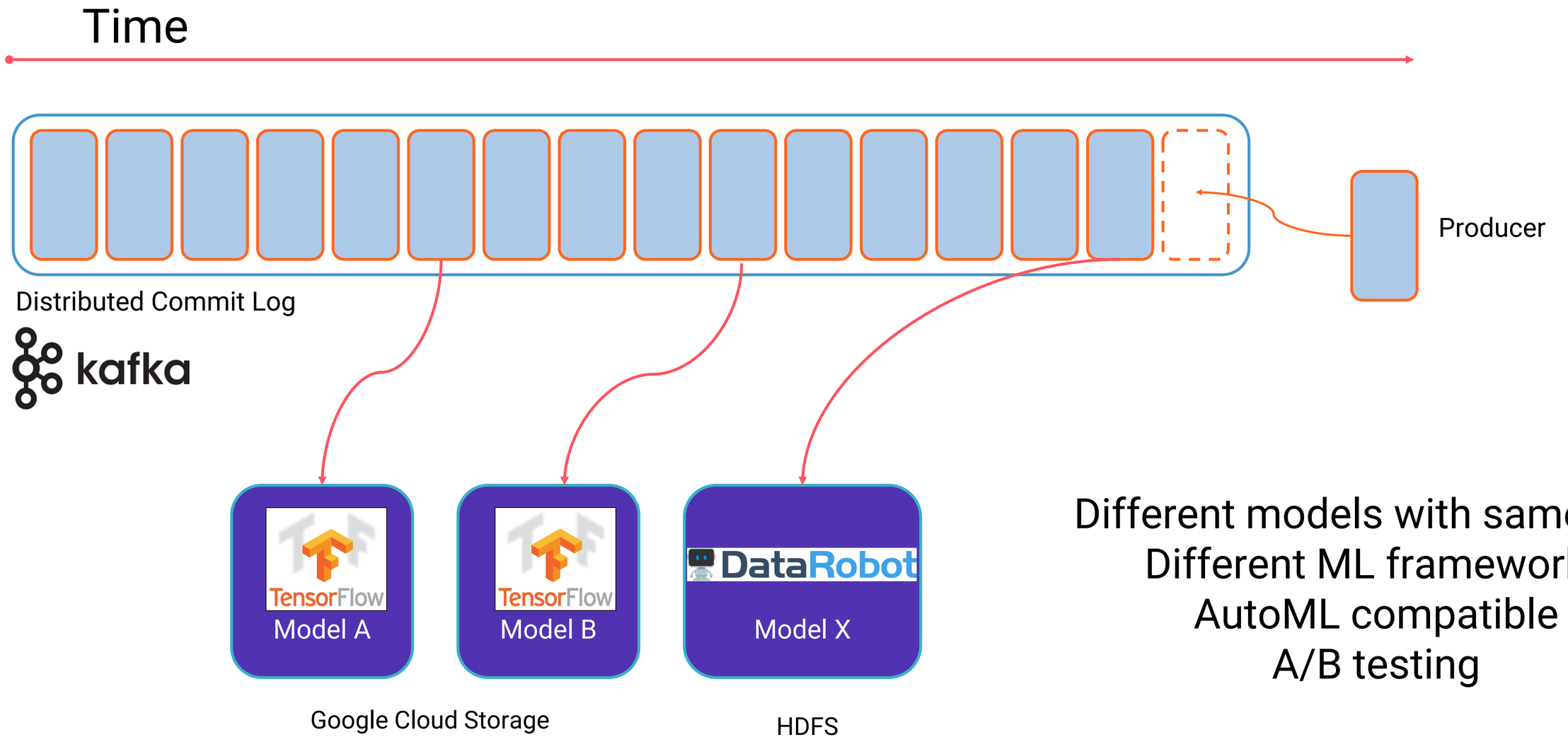
Apache Kafka as Infrastructure for ML



Apache Kafka's Open Ecosystem as Infrastructure for ML

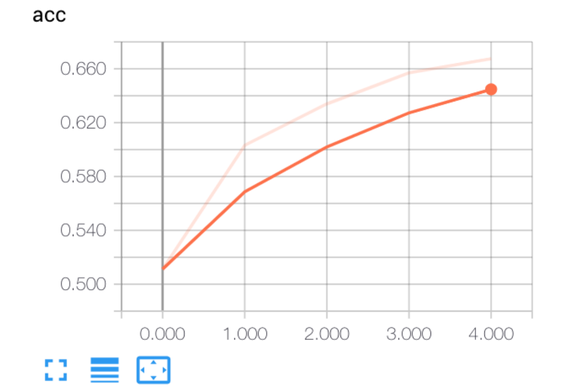


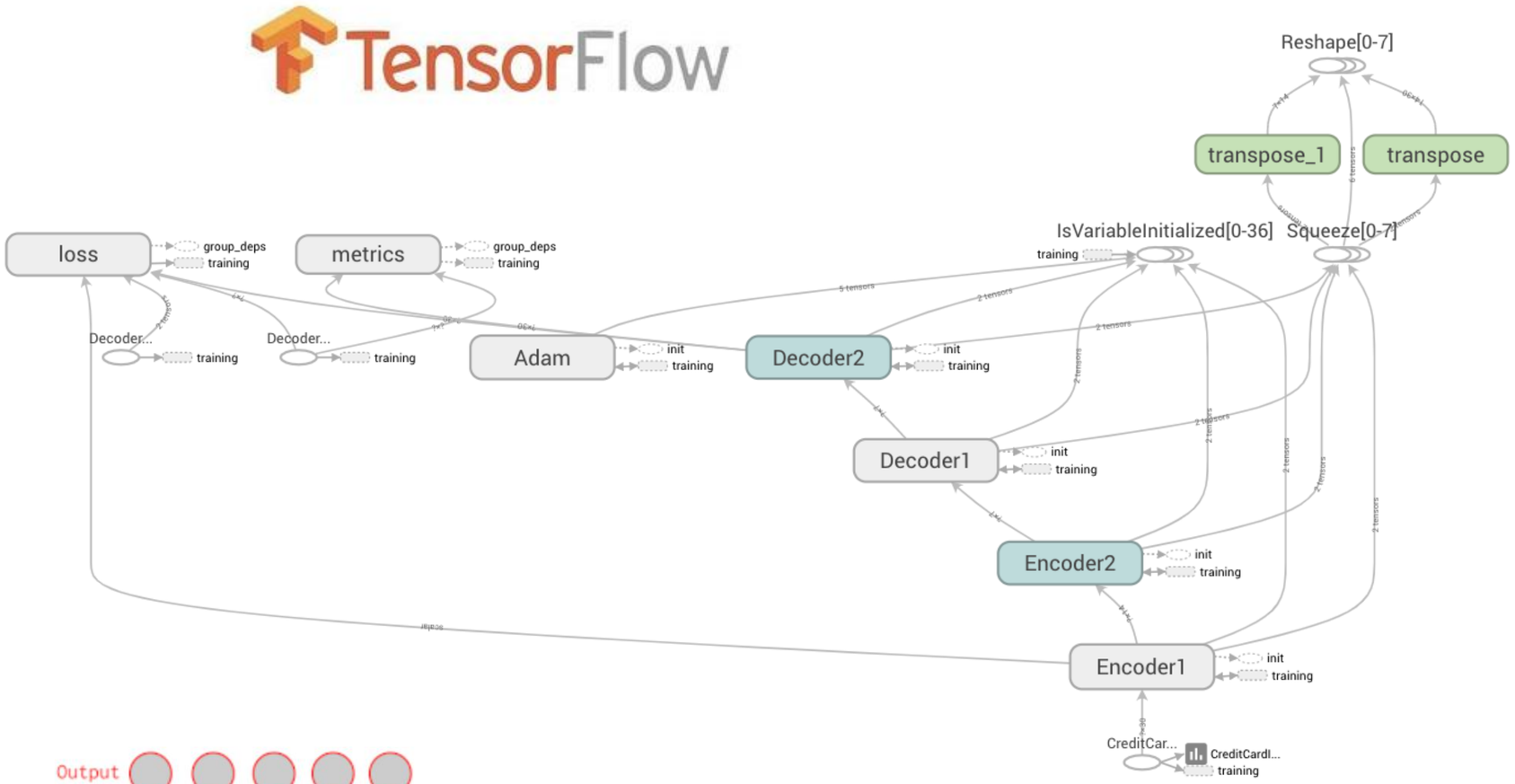
Replayability — a log never forgets!

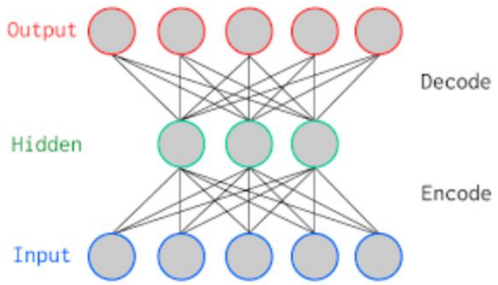


Different models with same data
Different ML frameworks
AutoML compatible
A/B testing

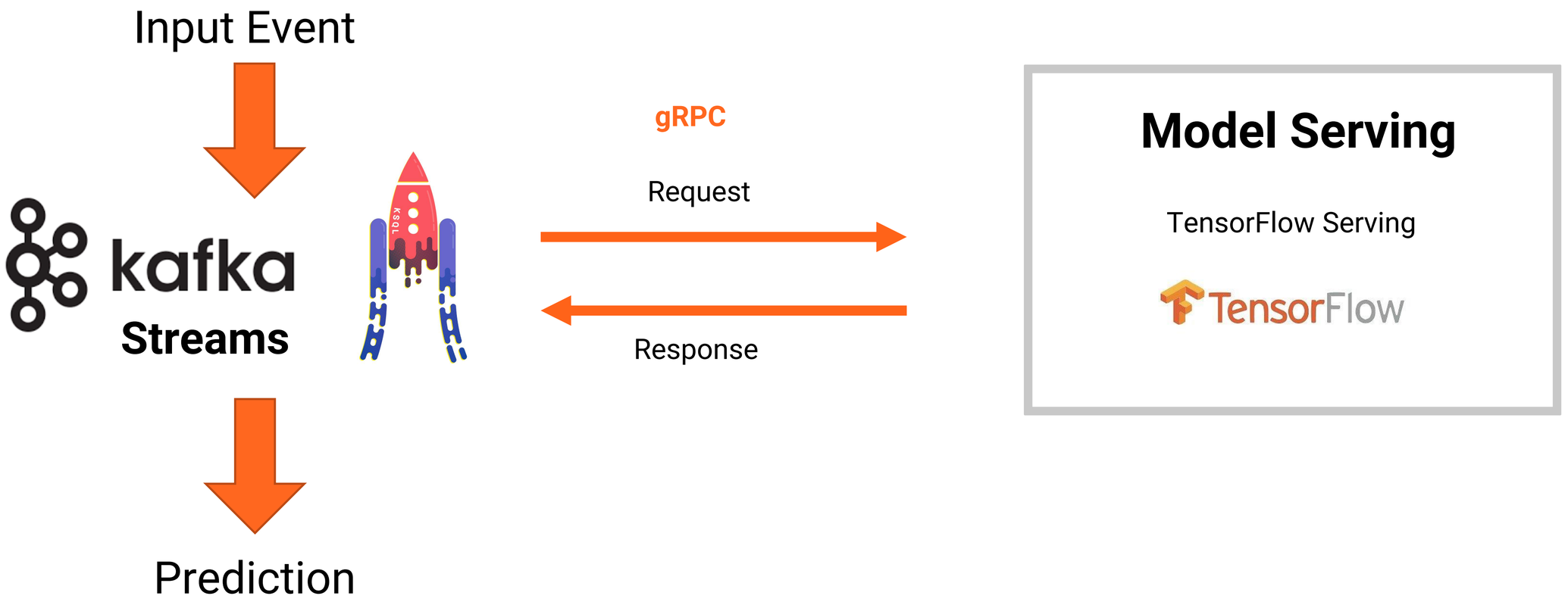
Analytic Model (Autoencoder for Anomaly Detection)



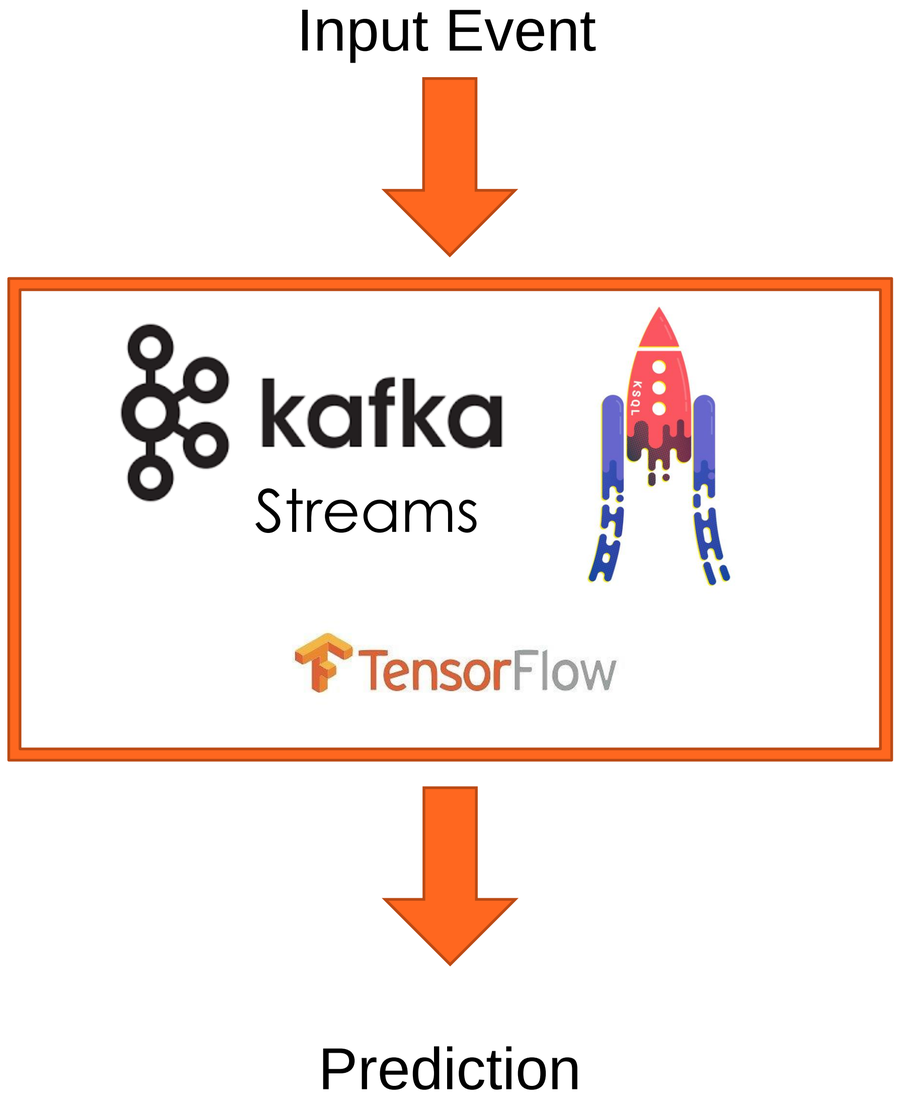




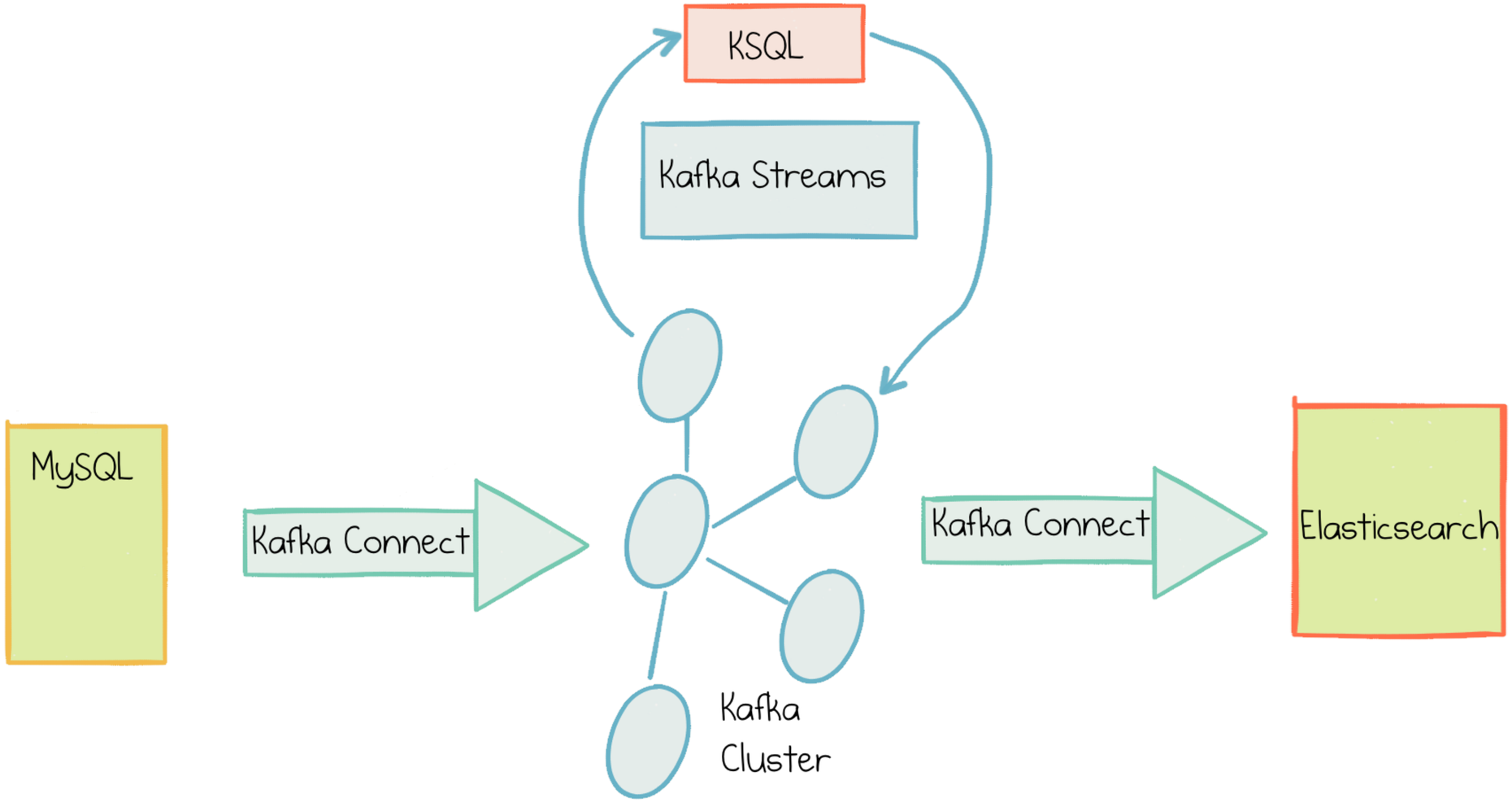
Model Deployment #1: RPC Communication to do Model Inference



Model deployment #2: Model interference natively in the App



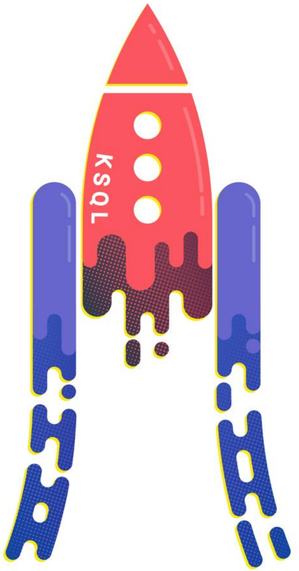
KSQL – The Streaming SQL Engine for Apache Kafka



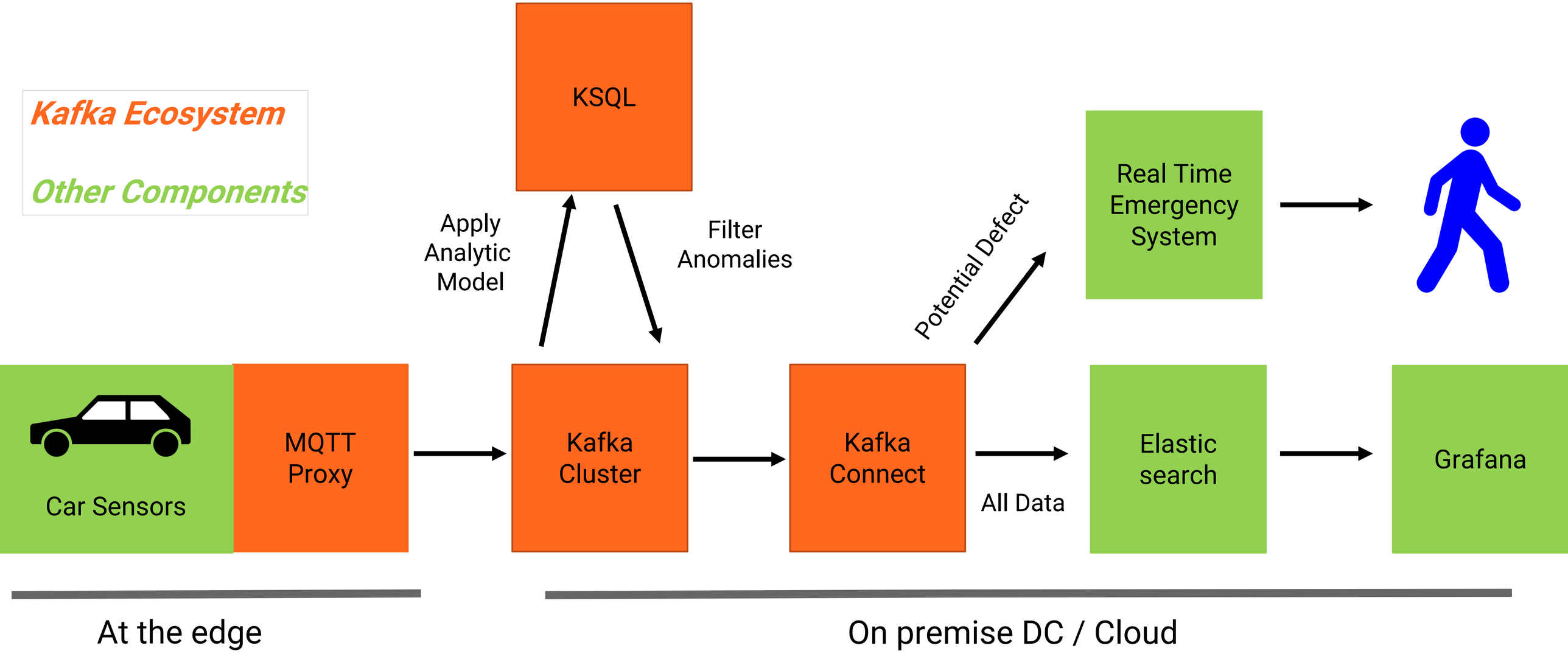
KSQL – Continuous Queries for Streaming ETL / Anomaly Detection

```
CREATE STREAM car_sensor_XYZ AS
SELECT car_ID, car_model, owner_id value FROM car c
LEFT JOIN users u ON c.owner_id = u.user_id
WHERE c.model = ‘Luxury Car ABC';
```

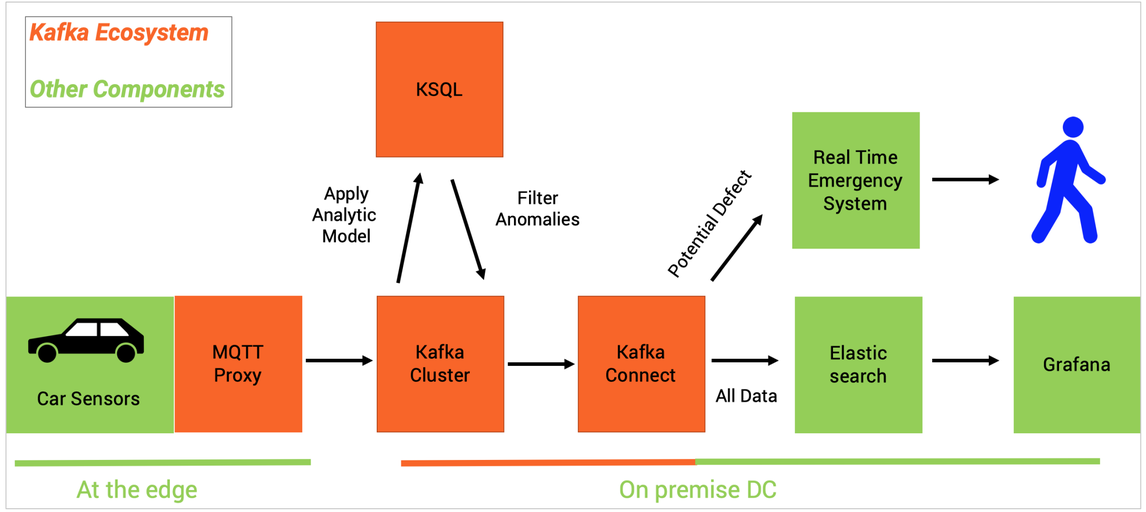
```
CREATE TABLE possible_detect AS
SELECT sensor_value, count(*)
FROM car_sensor
WINDOW TUMBLING (SIZE 120 MINUTES)
GROUP BY car_id
HAVING count(*) > 5;
```



KSQL and Deep Learning (Auto Encoder) for Anomaly Detection



Model Deployment with Apache Kafka, KSQL and TensorFlow

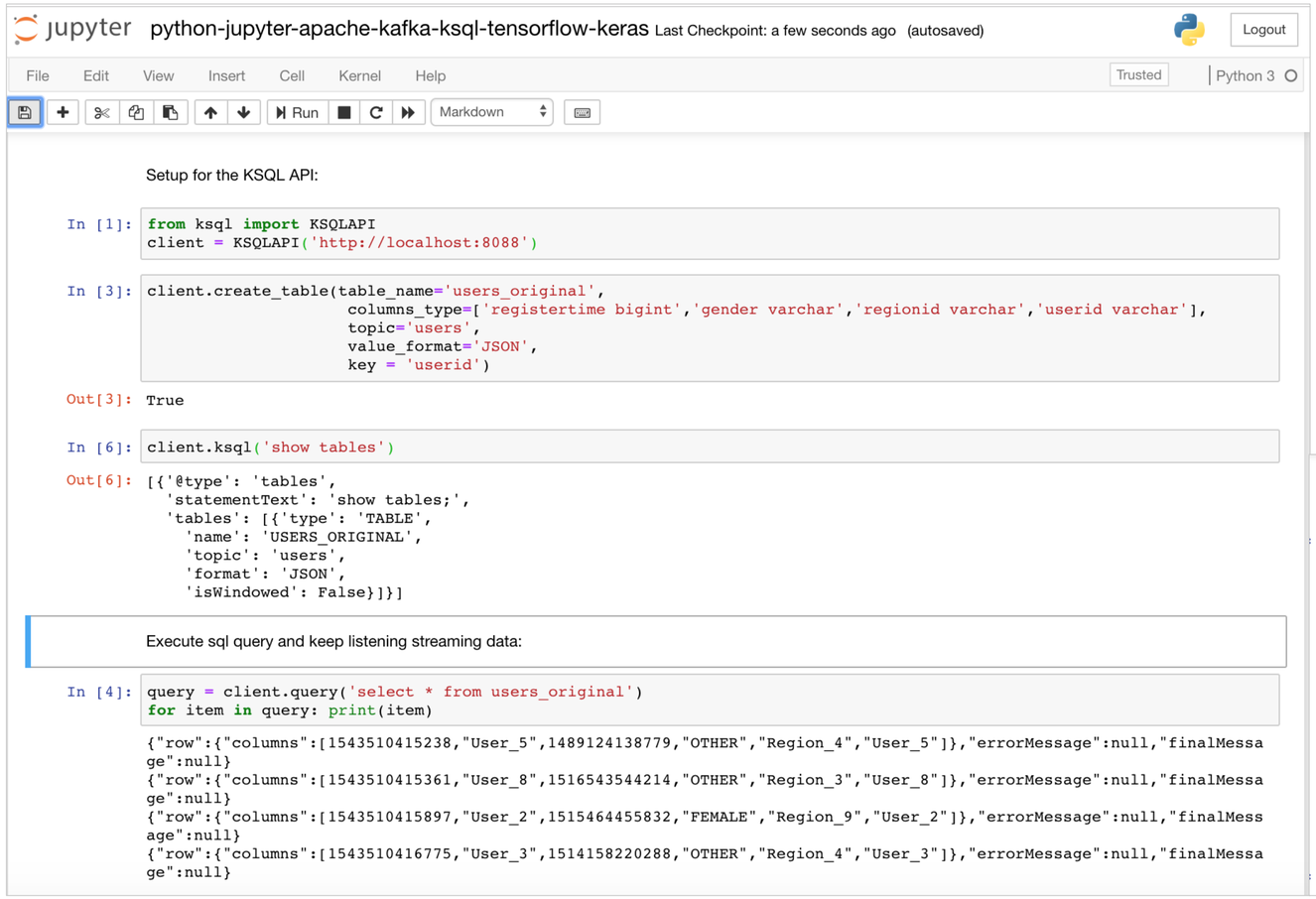


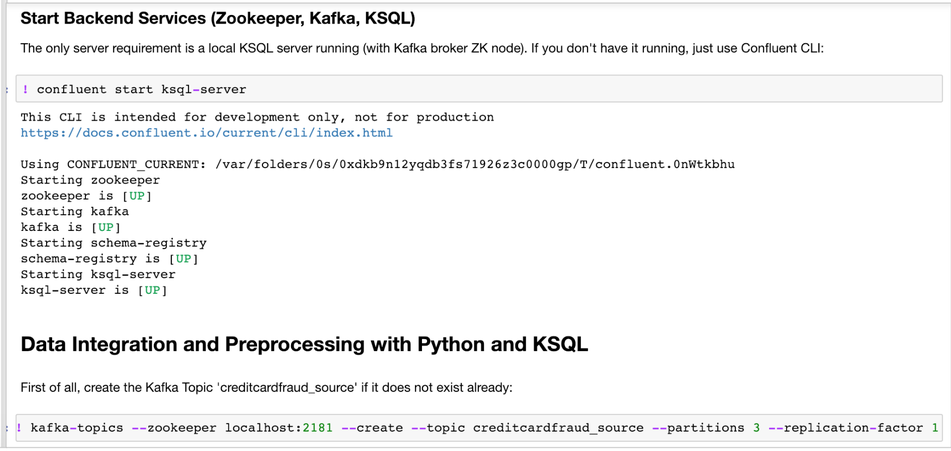
“CREATE STREAM AnomalyDetection AS
SELECT sensor_id, detectAnomaly(sensor_values)
FROM car_engine;“

User Defined Function (UDF)

TensorFlow

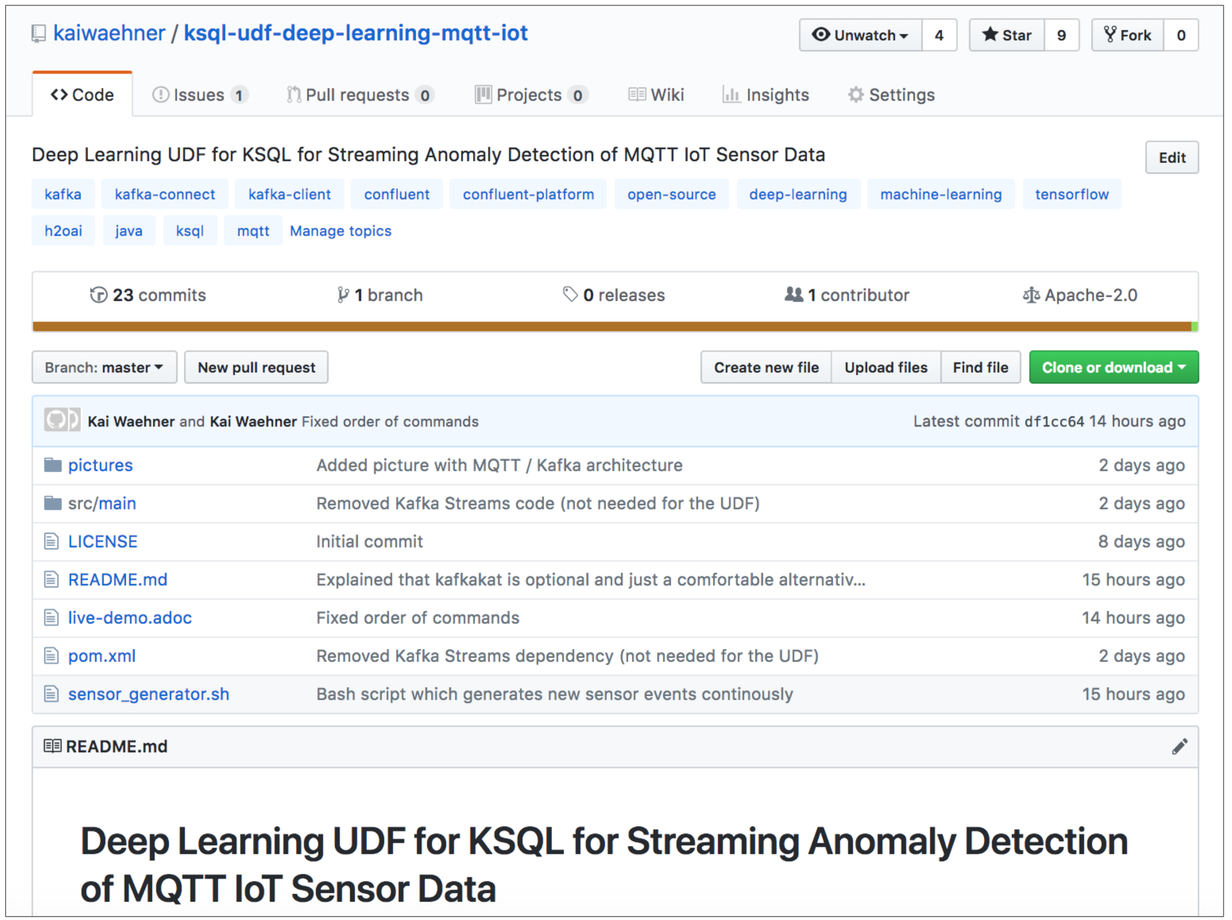
Model Training with Python, KSQL, TensorFlow, Keras and Jupyter





https://github.com/kaiwaehner/python-jupyter-apache-kafka-ksql-tensorflow-keras

Deep Learning UDF for KSQL for Streaming Anomaly Detection of MQTT IoT Sensor Data



https://github.com/kaiwaehner/ksql-udf-deep-learning-mqtt-iot

This is just the beginning of new era! Confluent Vision for Kafka:

Global

Automated **disaster recovery**

Global applications with **geo-awareness**

Infinite

Efficient and infinite data with **tiered storage**

Unlimited horizontal scalability for single clusters

Faster elastic scaling for brokers and partition

Elastic

Easy Kubernetes- based orchestration and management with Confluent operator

Faster **elastic scaling** when adding brokers and partitions

Cloud-Native Apache Kafka

Kai Waehner

Technology Evangelist

kontakt@kai-waehner.de
@KaiWaehner
www.kai-waehner.de
www.confluent.io
LinkedIn

Questions? Feedback? Please contact me!

Come and find out more about Kafka & Confluent on **Booth A3**


confluent