confluent

Unleashing Apache Kafka and TensorFlow in Hybrid Architectures

Kai Waehner

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

kafka

Google Cloud Platform

H_2O.ai

Microsoft Azure

TensorFlow

amazon web services

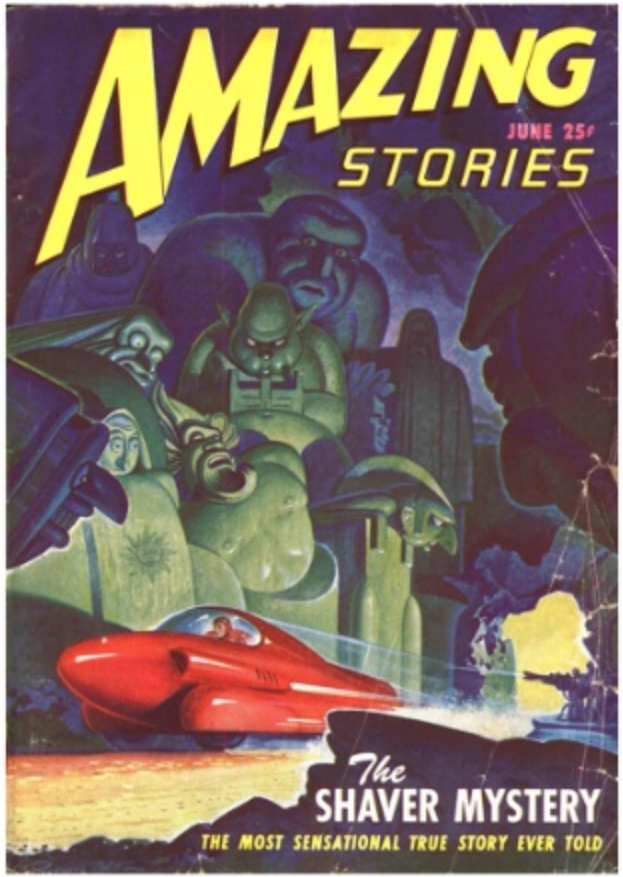
kubernetes

Poll

In which of these environments, if at all, are you using Apache Kafka today?

1. Self managed - On premise (Bare metal, VM)
2. Self managed - Public cloud (AWS EC2, etc.)
3. Self managed - Virtual private cloud (Kubernetes, Mesos, etc.)
4. Kafka as a Service – Public cloud
5. We do not currently use Apache Kafka

Disclaimer: This is a fictional story (but not far from reality)…



Global automotive company builds connected car infrastructure

Digital Transformation

- Improve customer experience
- Increase revenue
- Reduce risk

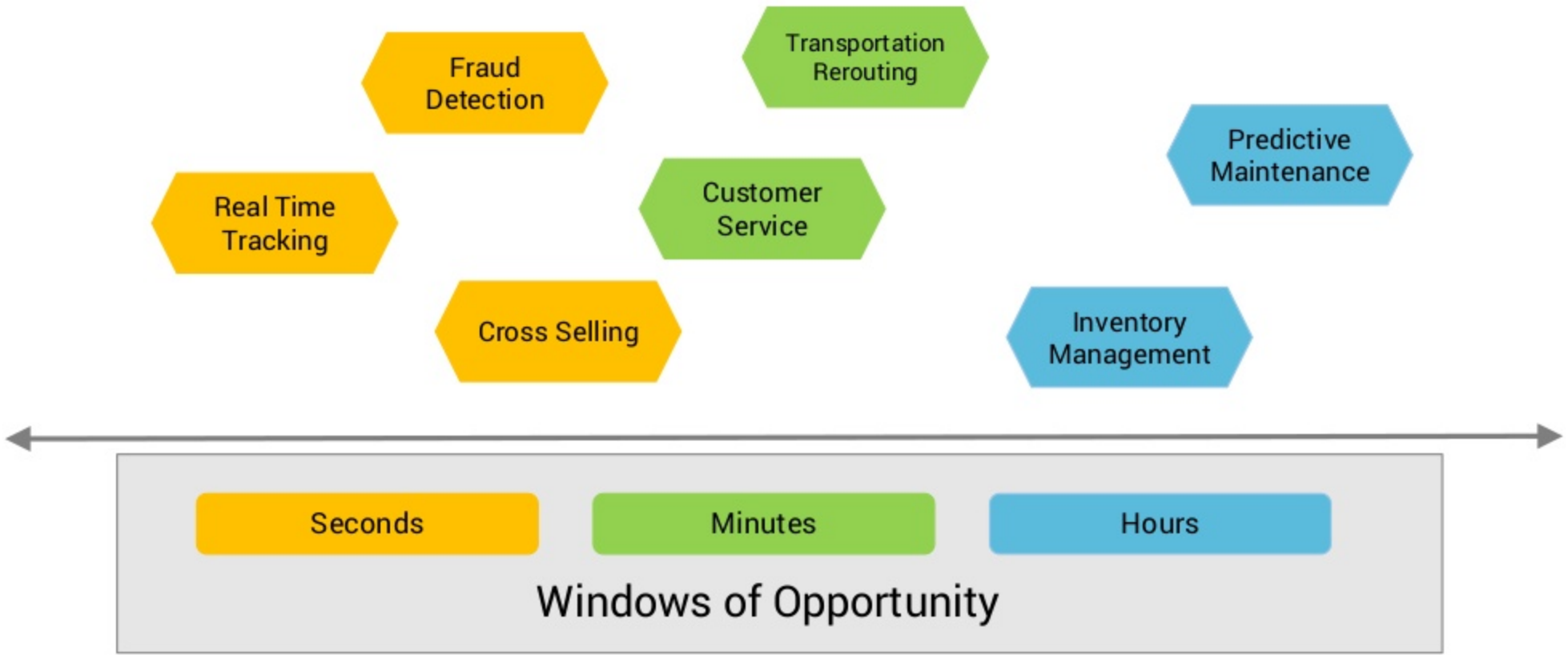
3 years ago — Project begins

Today — Connected car infrastructure in production for first use cases

2 years in the future — Improved processes leveraging machine learning

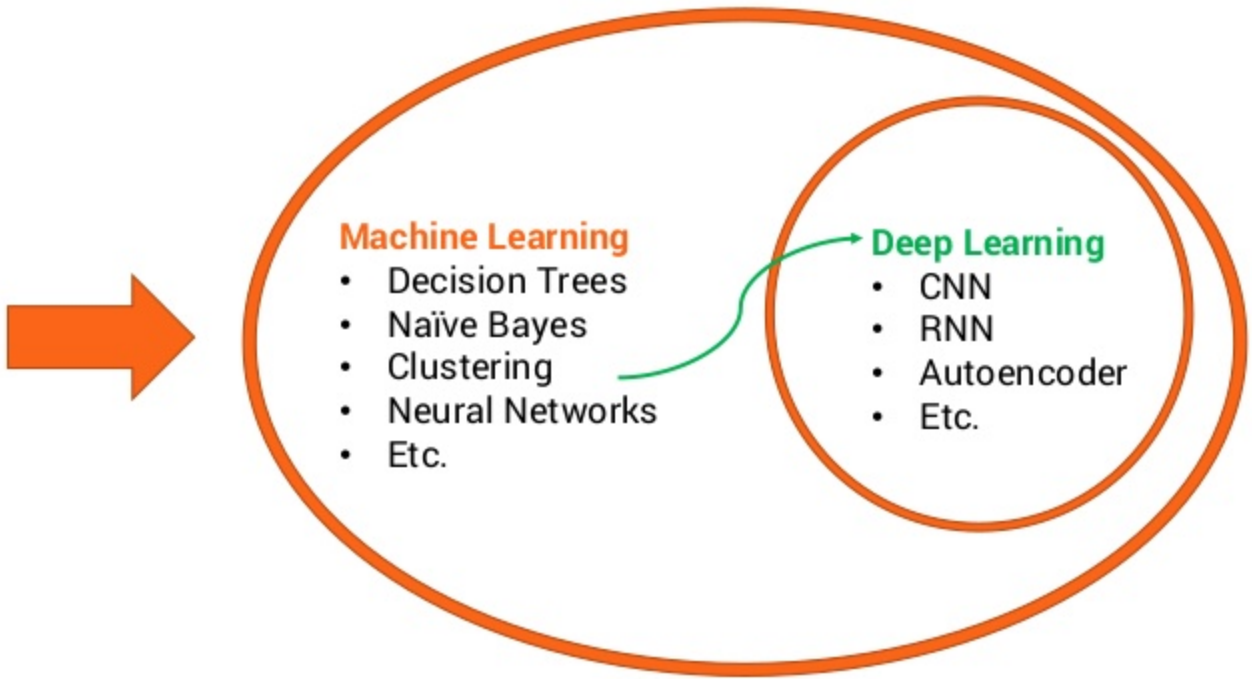
Time

Analyze and act on critical business moments



Machine Learning (ML)

...allows computers to find hidden insights without being explicitly programmed where to look.

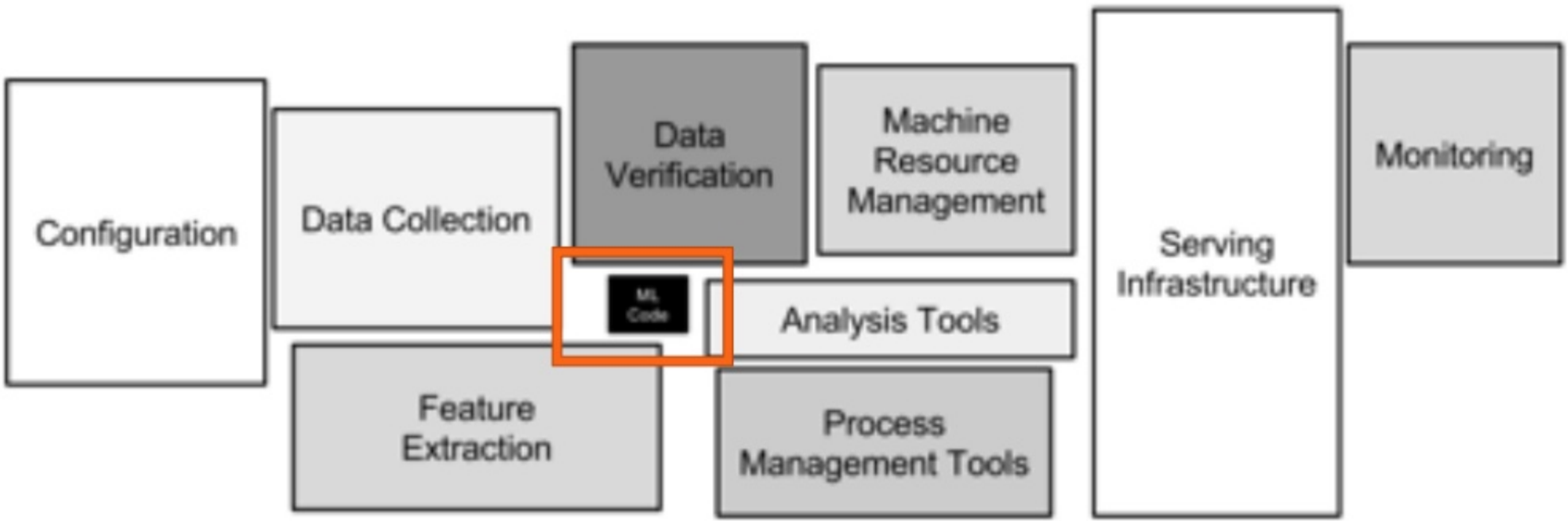


The First Analytic Models


```
def add_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:
```

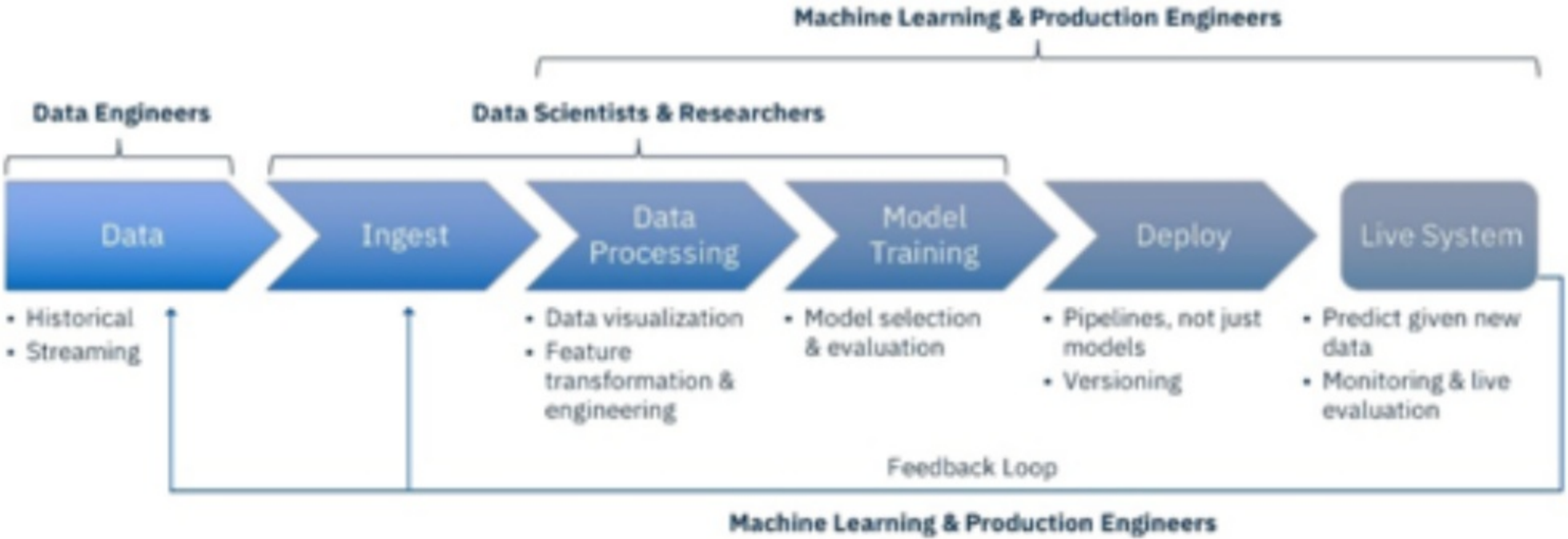
How to deploy the models in production?
...real-time processing?
...at scale?
...24/7 zero downtime?

Hidden Technical Debt in Machine Learning Systems



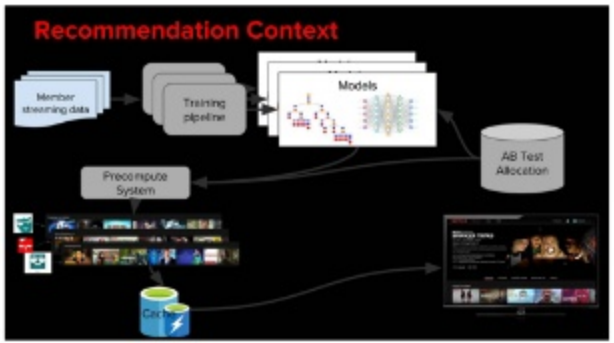
https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf

Impedance mismatch between model development and model deployment

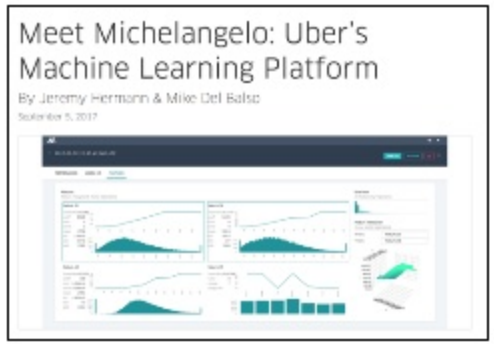


https://www.slideshare.net/NickPentreath/productionizing-spark-ml-pipelines-with-the-portable-format-for-analytics-100788521

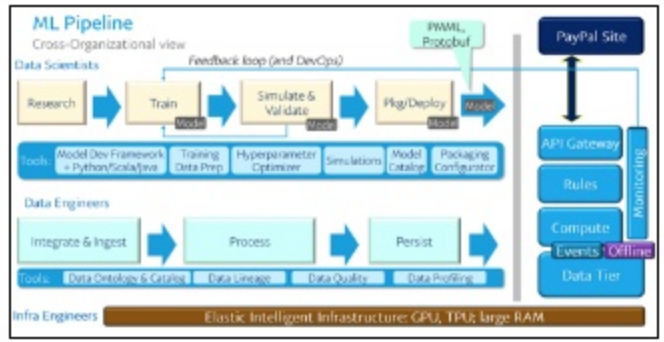
Scalable, Technology-Agnostic ML Infrastructures



NETFLIX



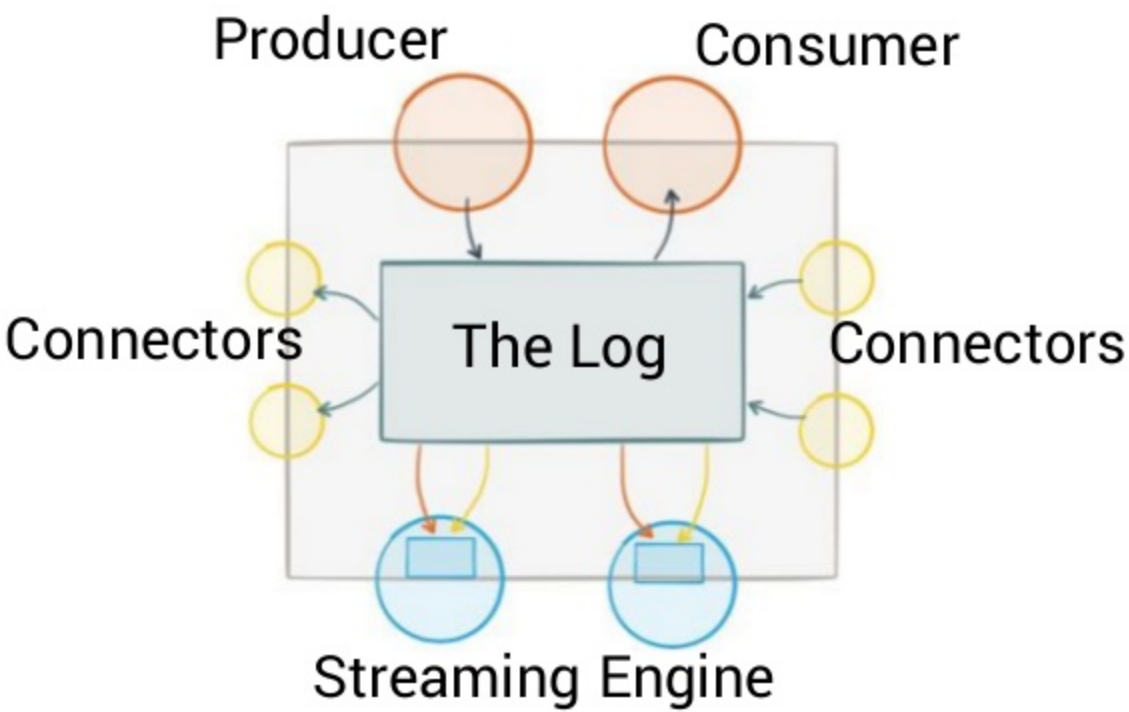
UBER



PayPal

https://www.infoq.com/presentations/netflix-ml-meson
https://eng.uber.com/michelangelo
https://www.infoq.com/presentations/paypal-data-service-fraud

Apache Kafka—The Rise of a Streaming Platform



kafka

1 PUB/SUB
2 STORE
3 PROCESS

Apache Kafka at Scale at Tech Giants

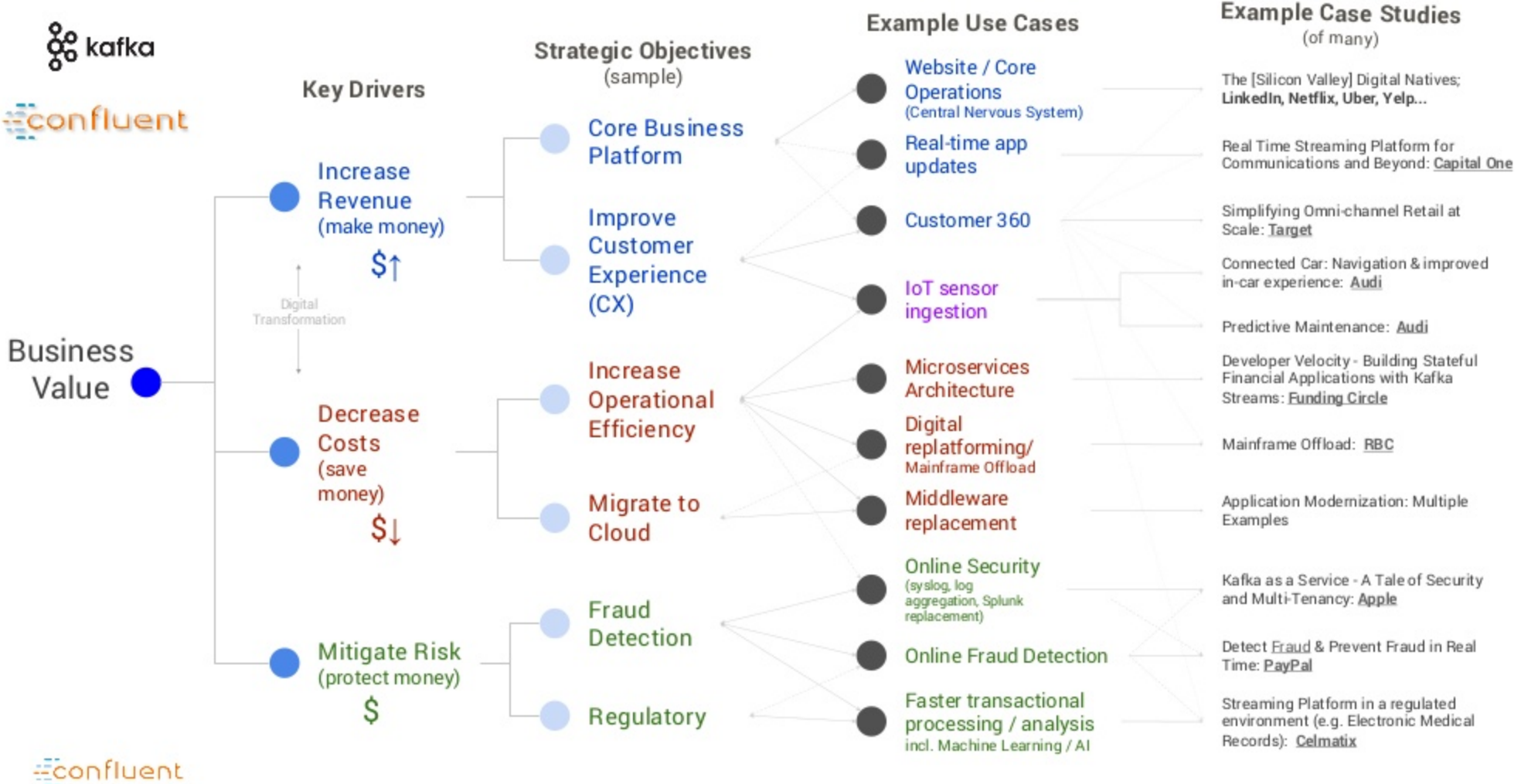
* Kafka Is not just used by tech giants
** Kafka is not just used for big data

> 4.5 trillion messages / day

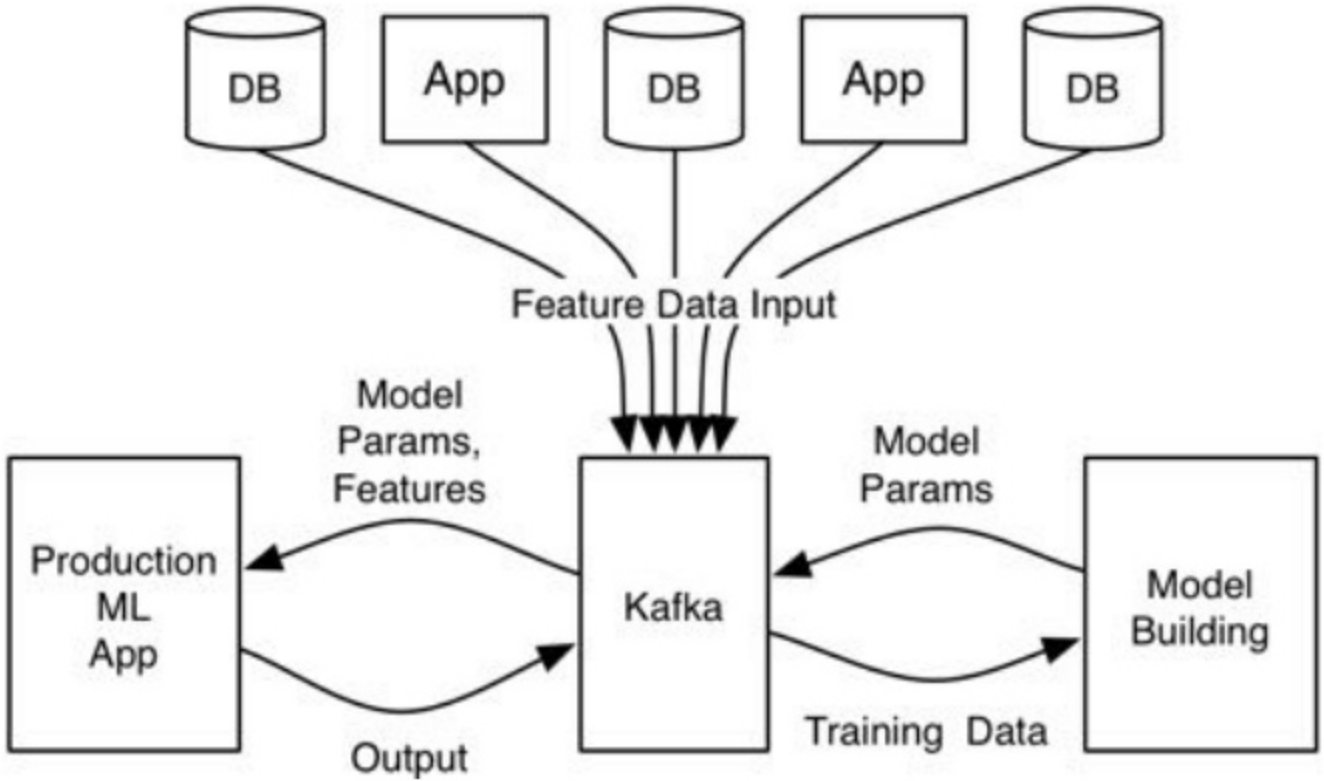
> 6 Petabytes / day

"You name it"

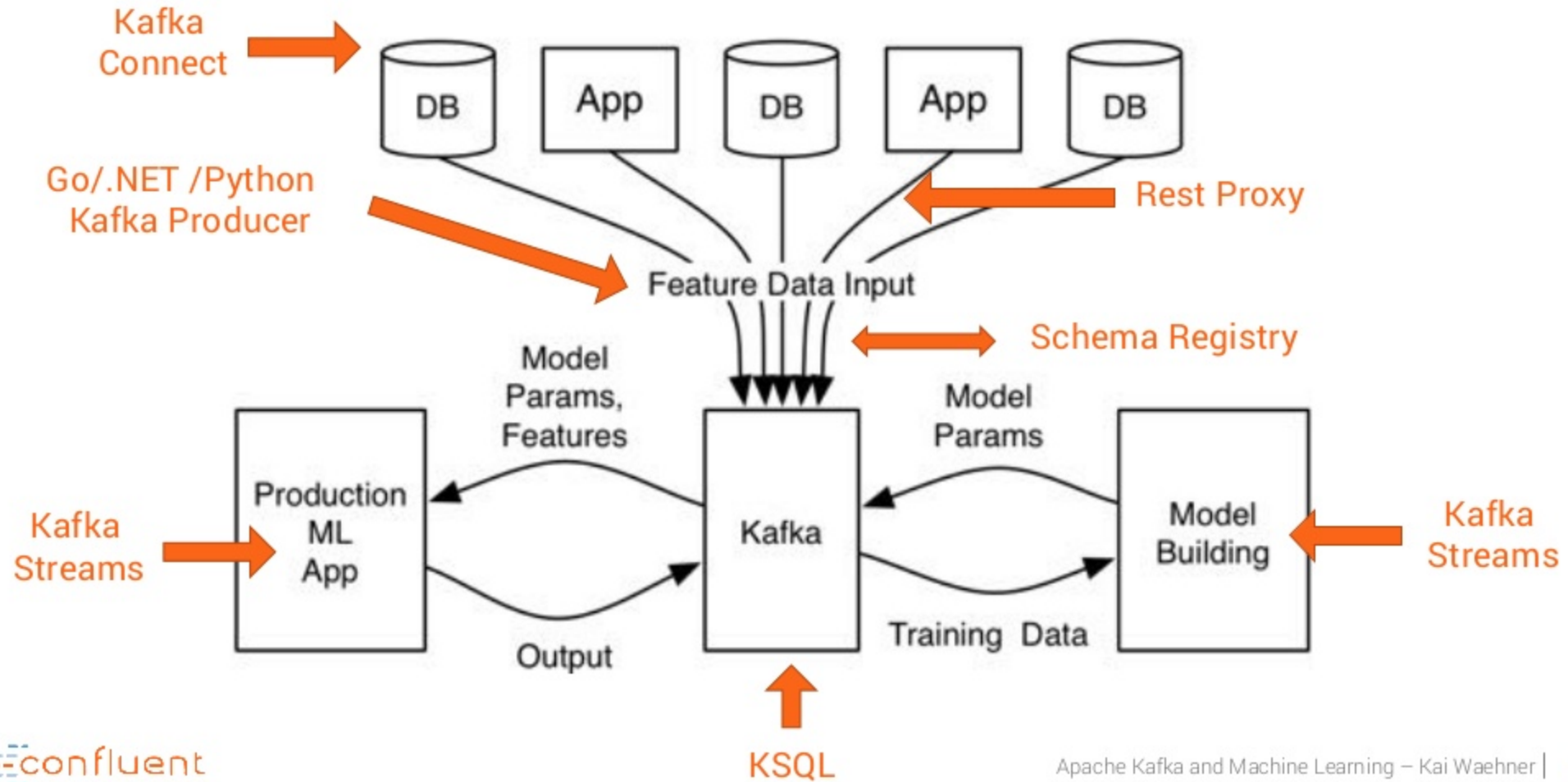
Confluents - Business Value per Use Case



Apache Kafka's Open Source Ecosystem as Infrastructure for ML

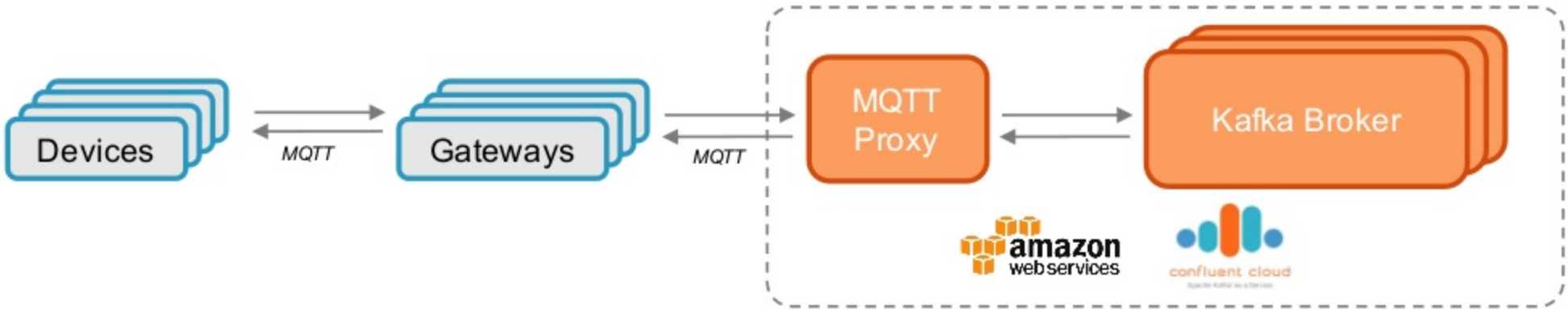


Apache Kafka's **Open Ecosystem** as Infrastructure for ML



Getting Started

Connected Car Infrastructure in Production on AWS



Real time tracking of the cars
to enable new, innovative digital services

Why did they choose Confluent Cloud?

Why Cloud?

- Extreme Scale
- Dynamic Instances
- Special Hardware (GPUs, TPUs,)

Why Confluent Cloud?

- No operations burden
- 99.95 Enterprise SLA, guaranteed high throughput, low ms latency end-to-end
- Confluent Ecosystem, Multi-Cloud + on premise Deployments,
- End-to-End monitoring with Confluent Control Center

Replication of IoT Data from AWS to GCP

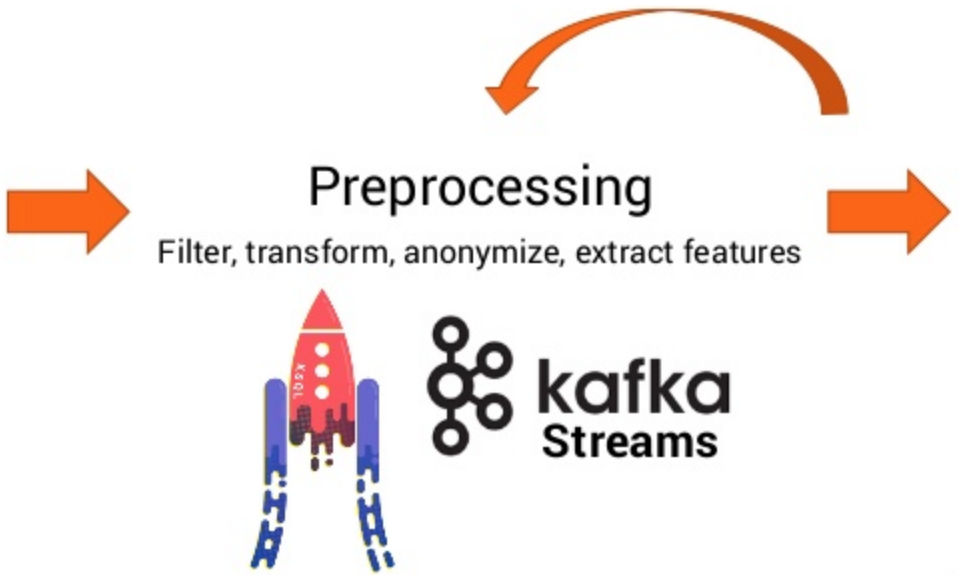
We should also use Kafka, but—oh no…GCP is the strategic cloud for the analytics team!

Poll

Which of the following use cases would you expect for hybrid Kafka deployments and replication of data?

1. Backup / Separation of Concerns
2. Disaster recovery
3. Active / active deployments
4. Cloud migration
5. We don't anticipate using a hybrid deployment

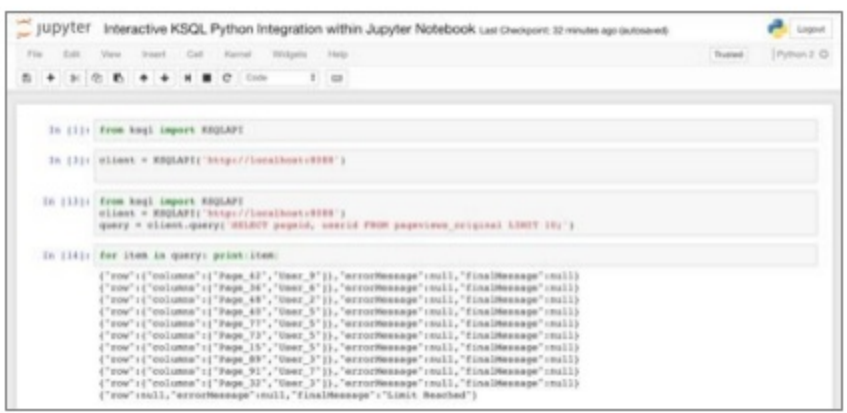
Data Preprocessing



- Use KSQL to preprocess data at scale without coding
- Use SQL statements for interactive analysis
  + deployment to production at scale
- Leverage e.g. Python with KSQL REST interface

Preprocessing with KSQL

```
SELECT car_id, event_id, car_model_id, sensor_input
FROM car_sensor c
LEFT JOIN car_models m ON c.car_model_id =
     m.car_model_id
WHERE m.car_model_type ='Audi_A8';
```

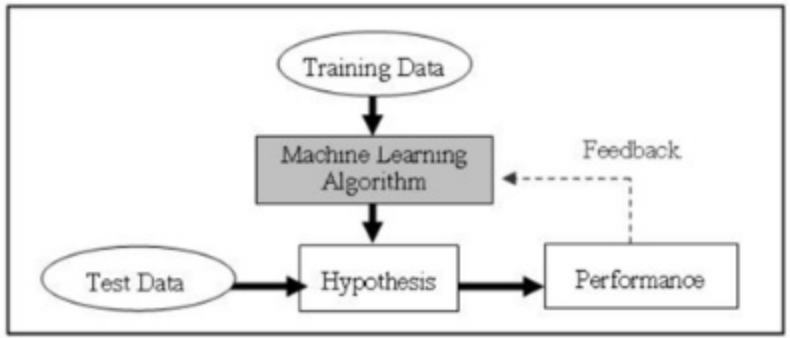


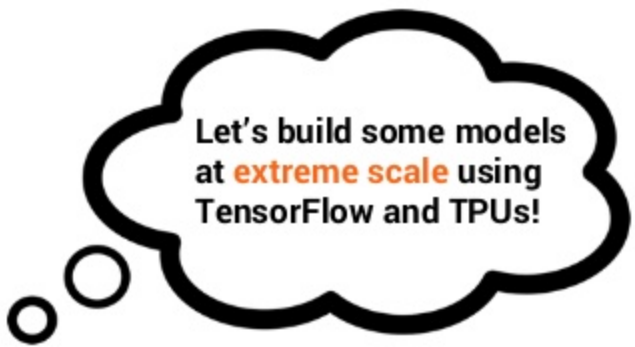
confluent

Data Ingestion into a Data Store

- "Kafka Benefits Under the Hood"
- Out-of-the-box connectivity
- Data format conversion
- Single message transformation (including error-handling)

Model Training using a Data Store





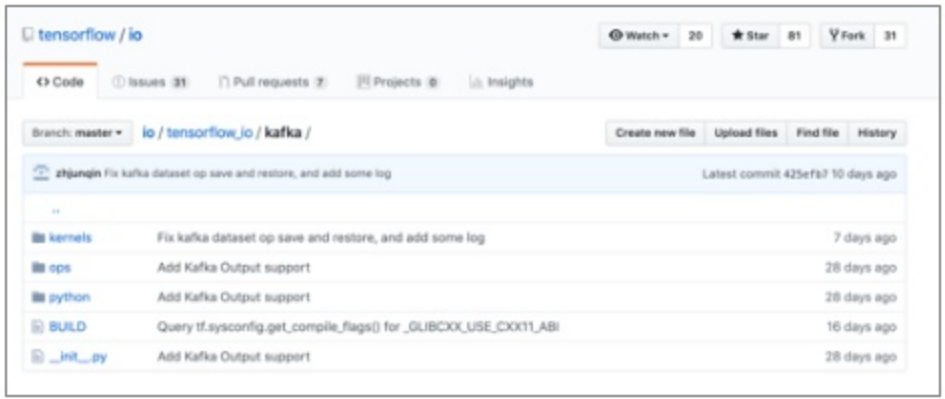
Google Cloud Storage

Cloud Machine Learning Engine

TensorFlow

Analytic Model

Model Training without additional Data Store



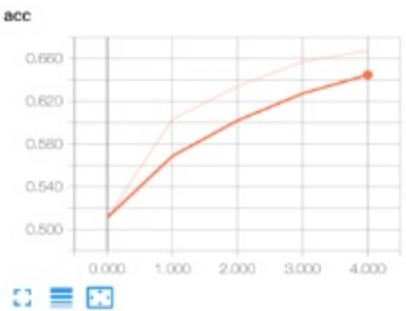
https://github.com/tensorflow/io/tree/master/tensorflow_io/kafka

kafka

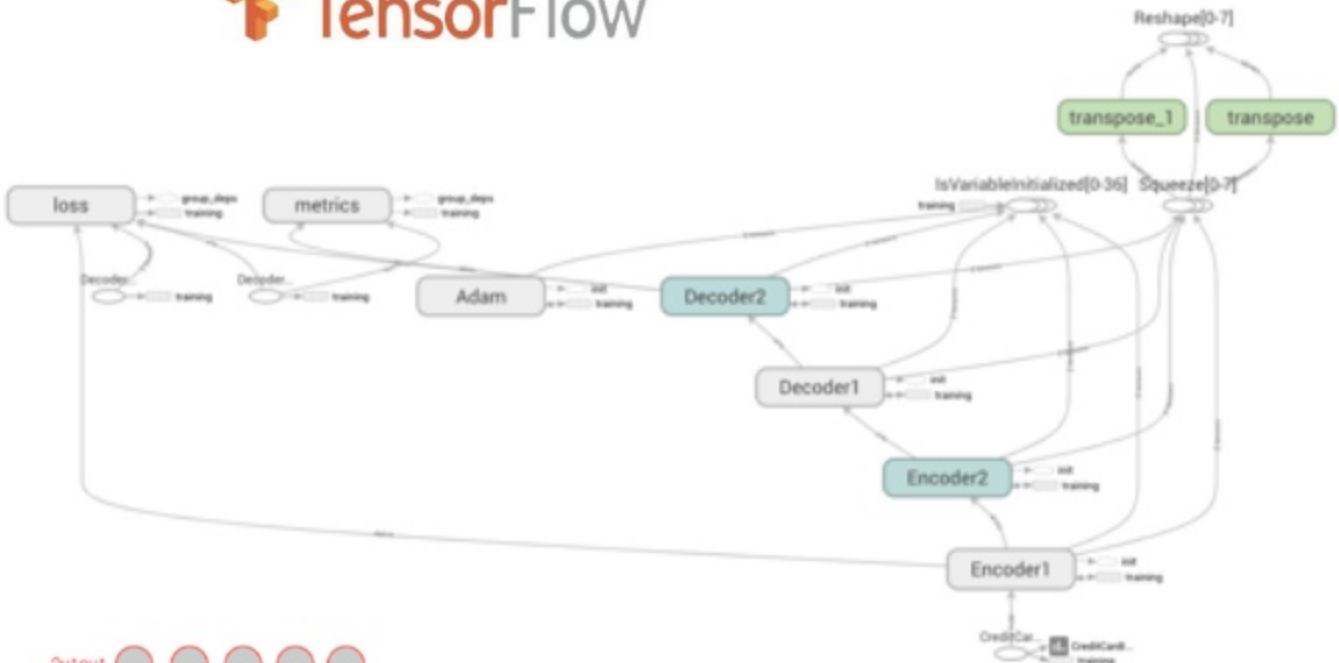
TensorFlow

- Native integration between Kafka and TensorFlow
- KafkaDataSet and KafkaOutputSequence for TensorFlow
- Written in C++ (linked with librdkafka
- Part of the graph in TensorFlow
- Direct training and inference from streaming data
- No data storage like S3 or HDFS needed

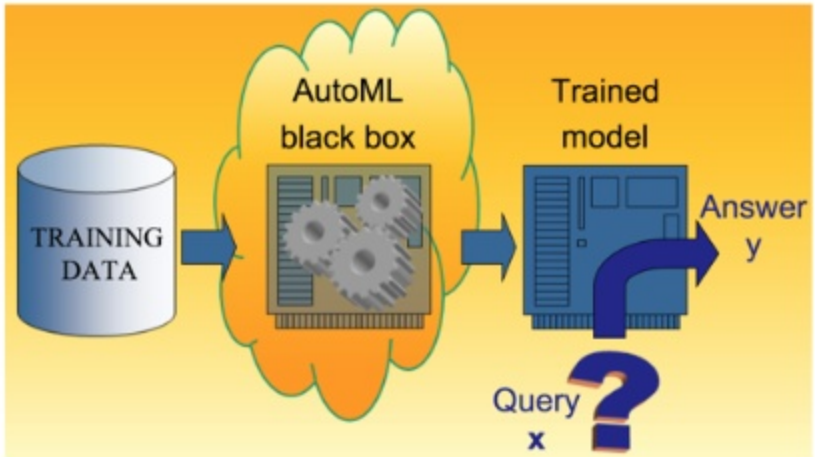
Analytic Model (Autoencoder for Anomaly Detection)



TensorFlow



AutoML



http://slideplayer.com/slide/10575150/

Cloud AutoML

H2O.ai

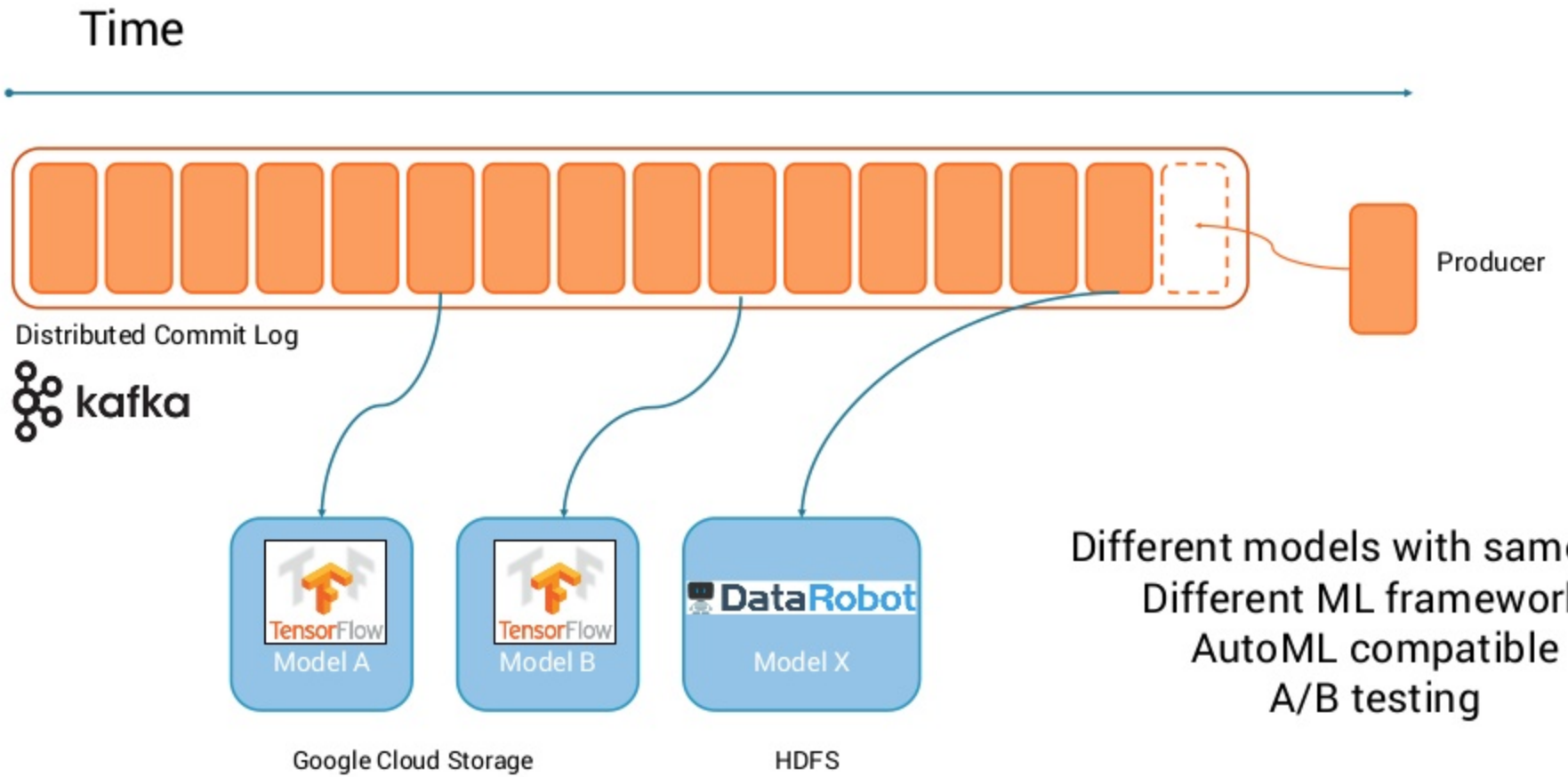
Driverless AI: Your Expert System for AI

DataRobot

tazi.ai

"One-click Data-in Model-out Simplicity"

Replayability – a log never forgets!



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

The Need for Local Data Processing

Confluent Replicator

Analytic Model

CLOUD

PII data

Local Processing

ON PREMISE

We are ready to use our models for predictions, BUT all the PII data needs to be processed in our local data center!

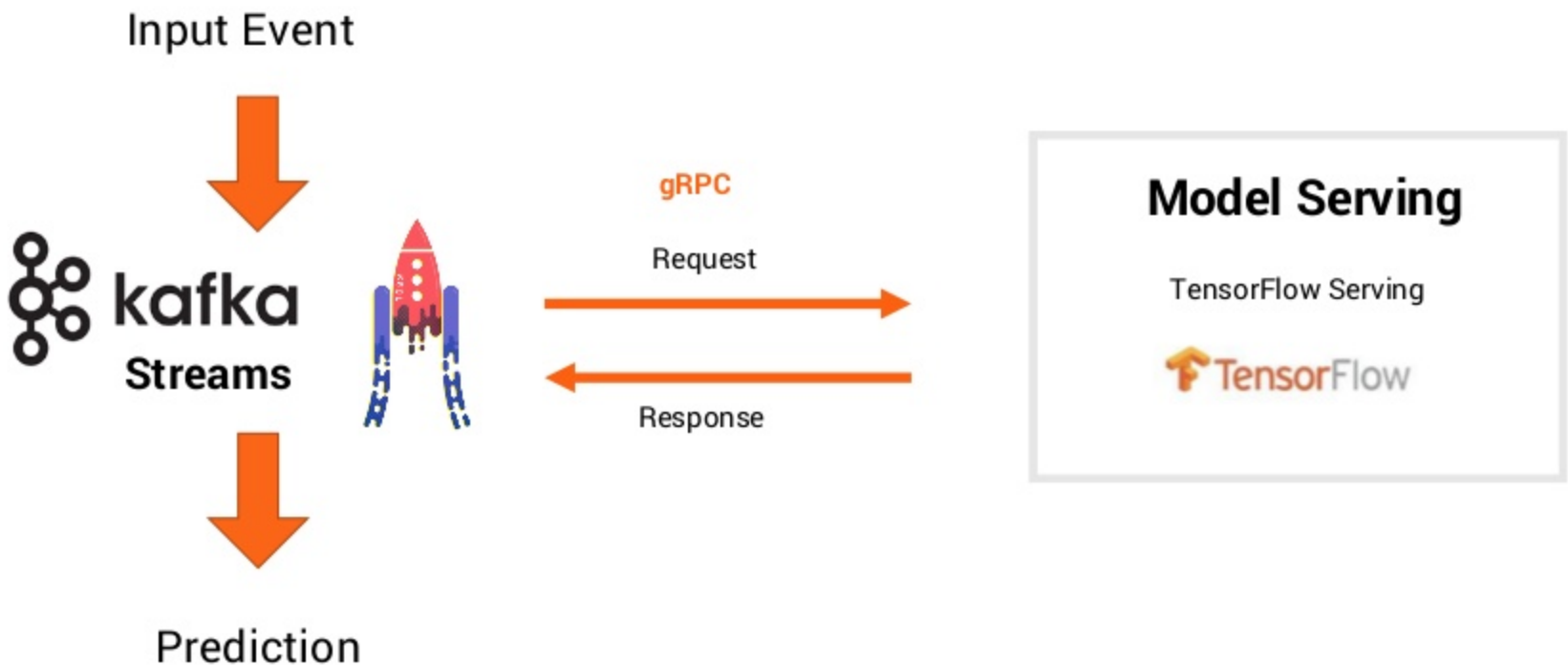
Self managed on premise deployment for model deployment and monitoring

Confluent Operator takes over the challenge of operating Kafka on Kubernetes!

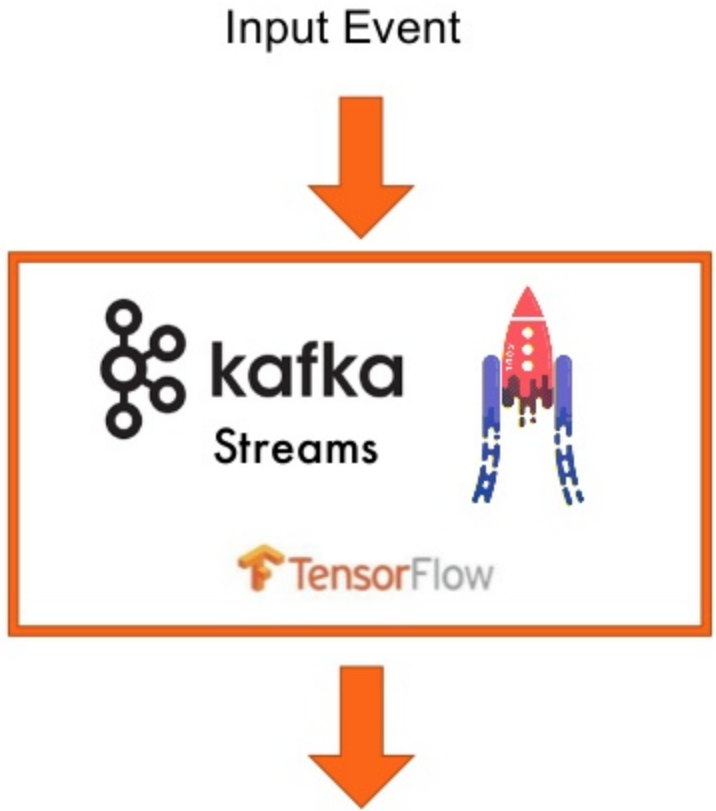
(Automated provisioning, scaling, fail-over, partition rebalancing, rolling updates, monitoring, …)

Stream Processing vs. Request-Response for Model Serving

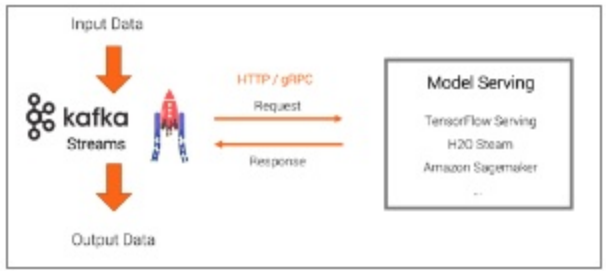
Model Deployment - Option 1:
RPC communication to do model inference



Model Deployment - Option 2: Model interference natively integrated into the App



Stream Processing vs. Request-Response for Model Serving



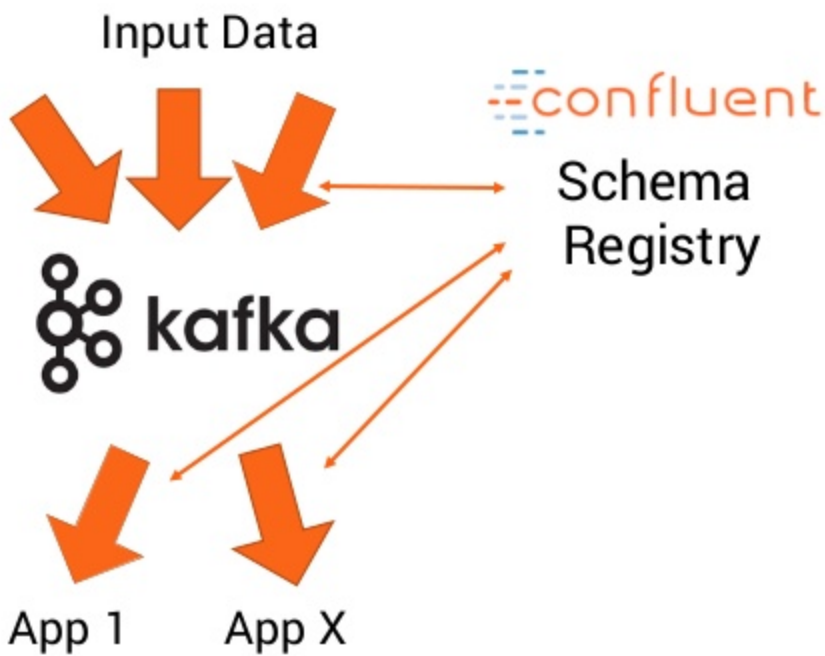
Pros of a Model Server:

- **Simple integration** with existing technologies and organizational processes
- **Easier to understand** if you come from non-streaming world
- **Later migration** to real streaming is also **possible**
- **Model management built-in** for different models, versioning and A/B testing

Cons (== Pros of Deployment in the Streaming App):

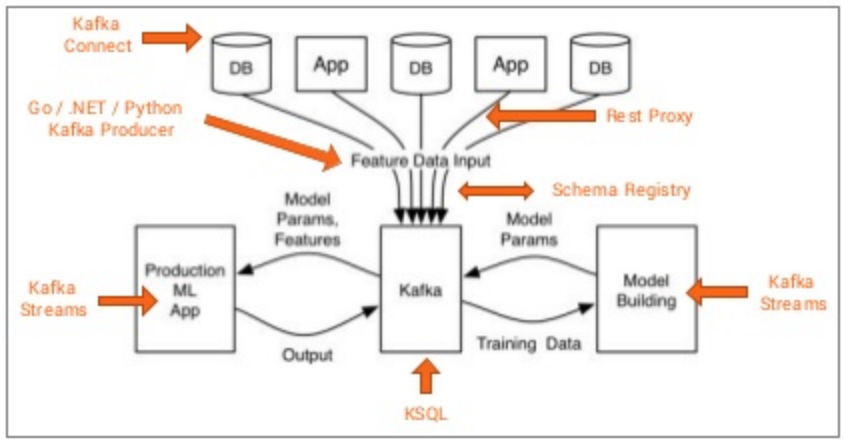
- **Worse latency** as remote call instead of local inference
- **No offline inference** (devices, edge processing, etc.)
- **Coupling** the availability, scalability, and latency/throughput of your Kafka Streams application with the **SLAs of the RPC interface**
- **Side-effects** (e.g., in case of failure) **not covered by Kafka processing** (e.g., exactly once)

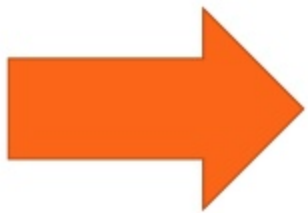
Confluent Schema Registry for Message Validation



- "Kafka Benefits Under the Hood"
- Schema definition + evolution
- Forward and backward compatibility
- Multi data center deployment

Monitoring the infrastructure for ML





Grafana

Prometheus

confluent Control Center

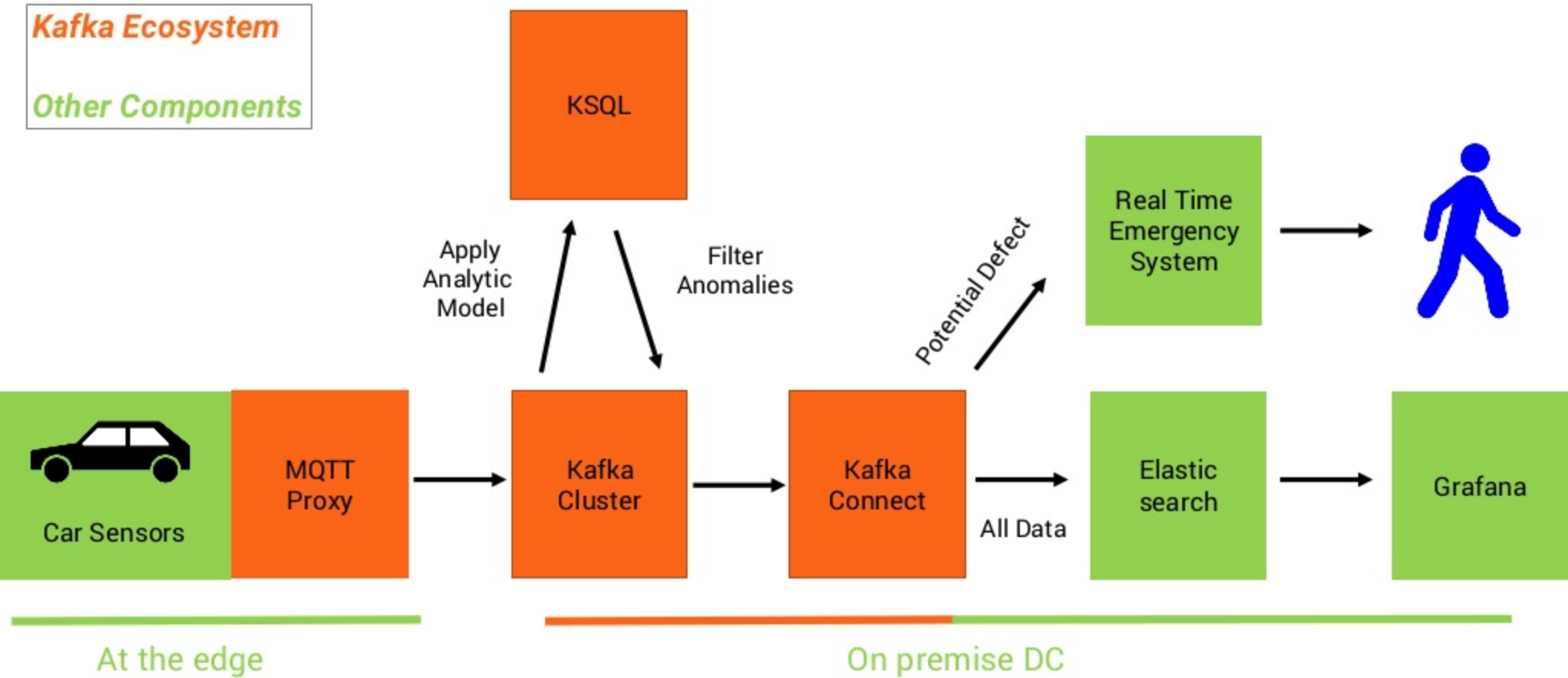
Java

Build vs. Buy

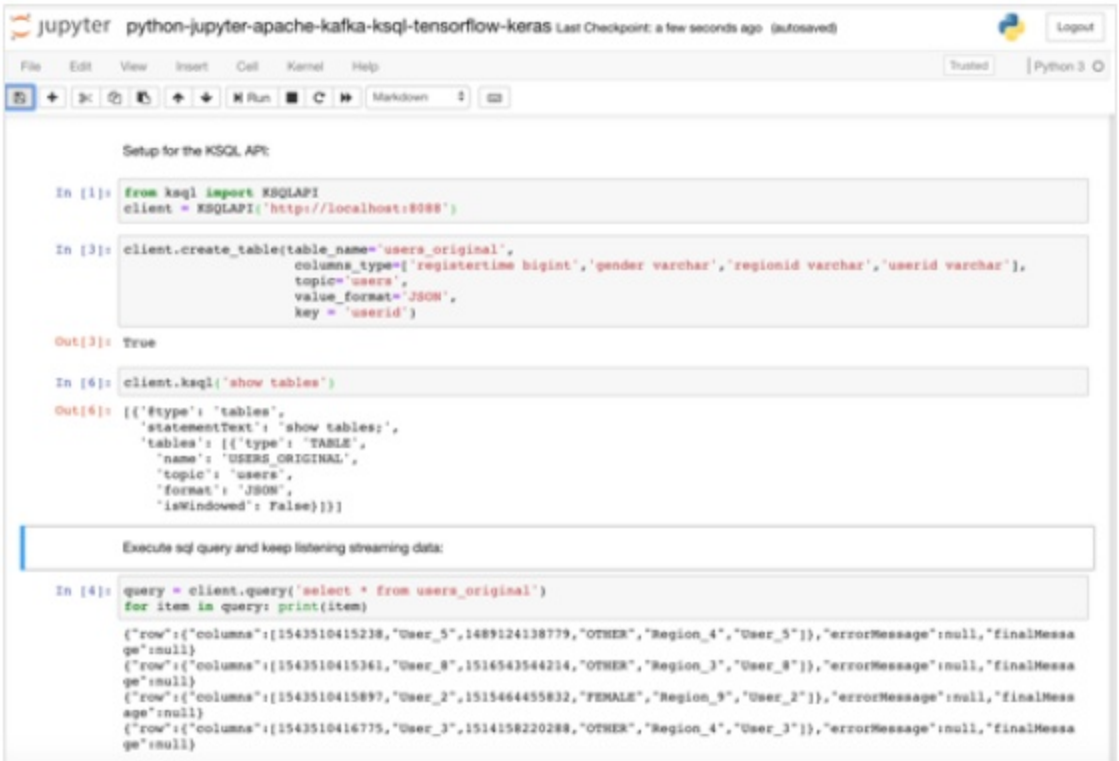
Hosted vs. Managed

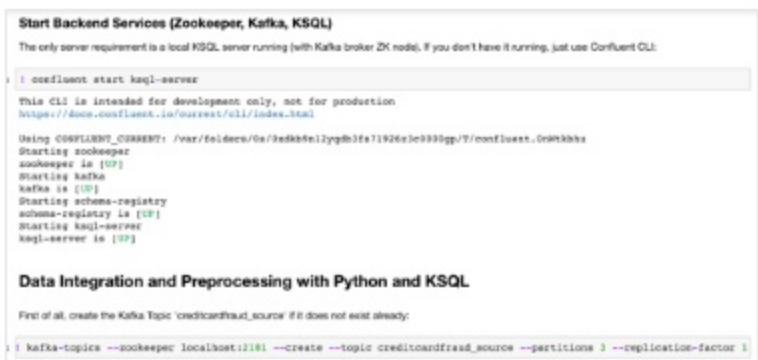
Basic vs. Advanced

KSQL and Deep Learning (Auto Encoder) for Anomaly Detection



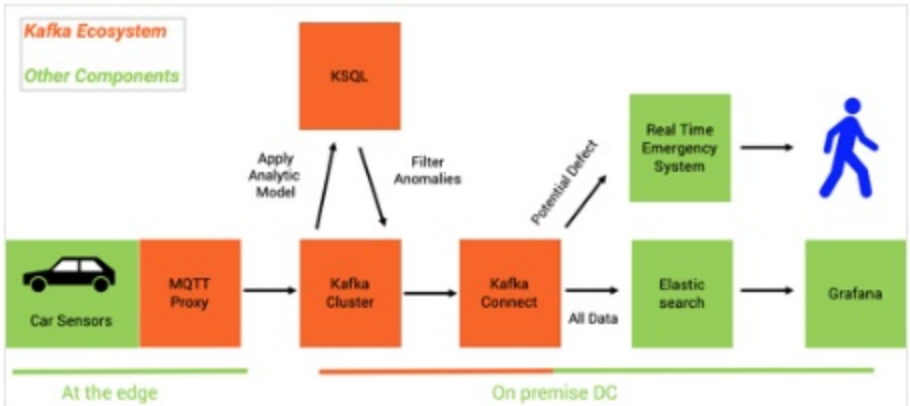
Model Training with Python, KSQL, TensorFlow, Keras and Jupyter





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

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)

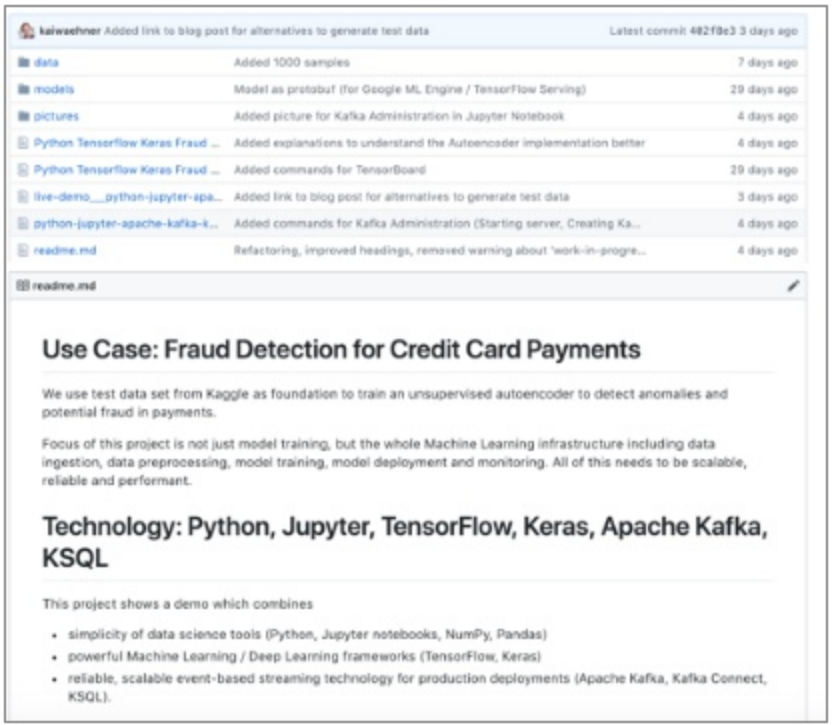
TensorFlow

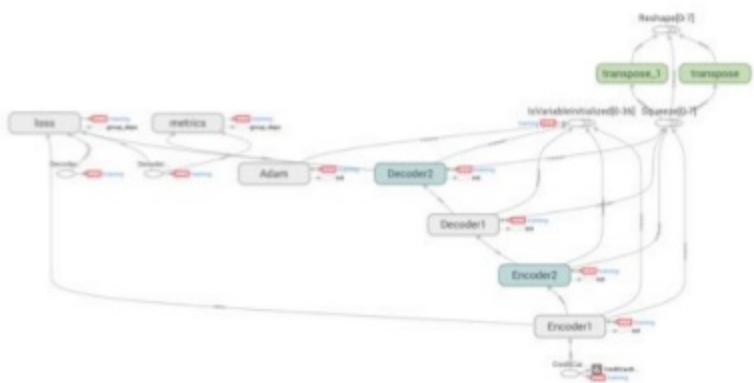
Live Demo

End-to-End Sensor Analytics…

Python, Jupyter Notebook, TensorFlow, Keras, Apache Kafka, KSQL and MQTT

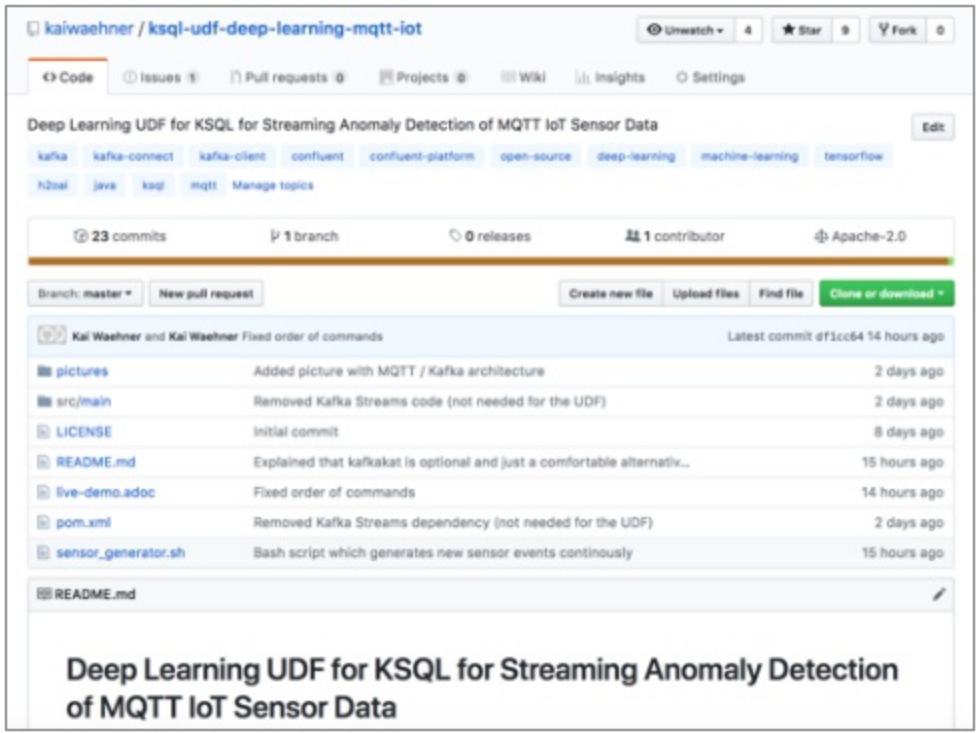
Model Training with Python, KSQL, TensorFlow, Keras and Jupyter





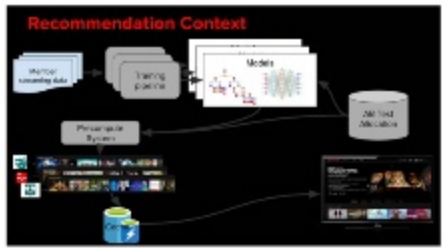
https://github.com/kaiwaehner/python-jupyter-apache-kafka-ksql-tensorflow-keras

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



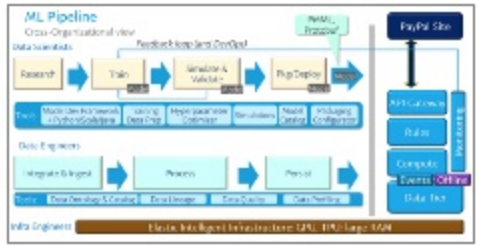
https://github.com/kaiwaehner/ksql-udf-deep-learning-mqtt-iot

Comparing our current project status to others



NETFLIX

UBER



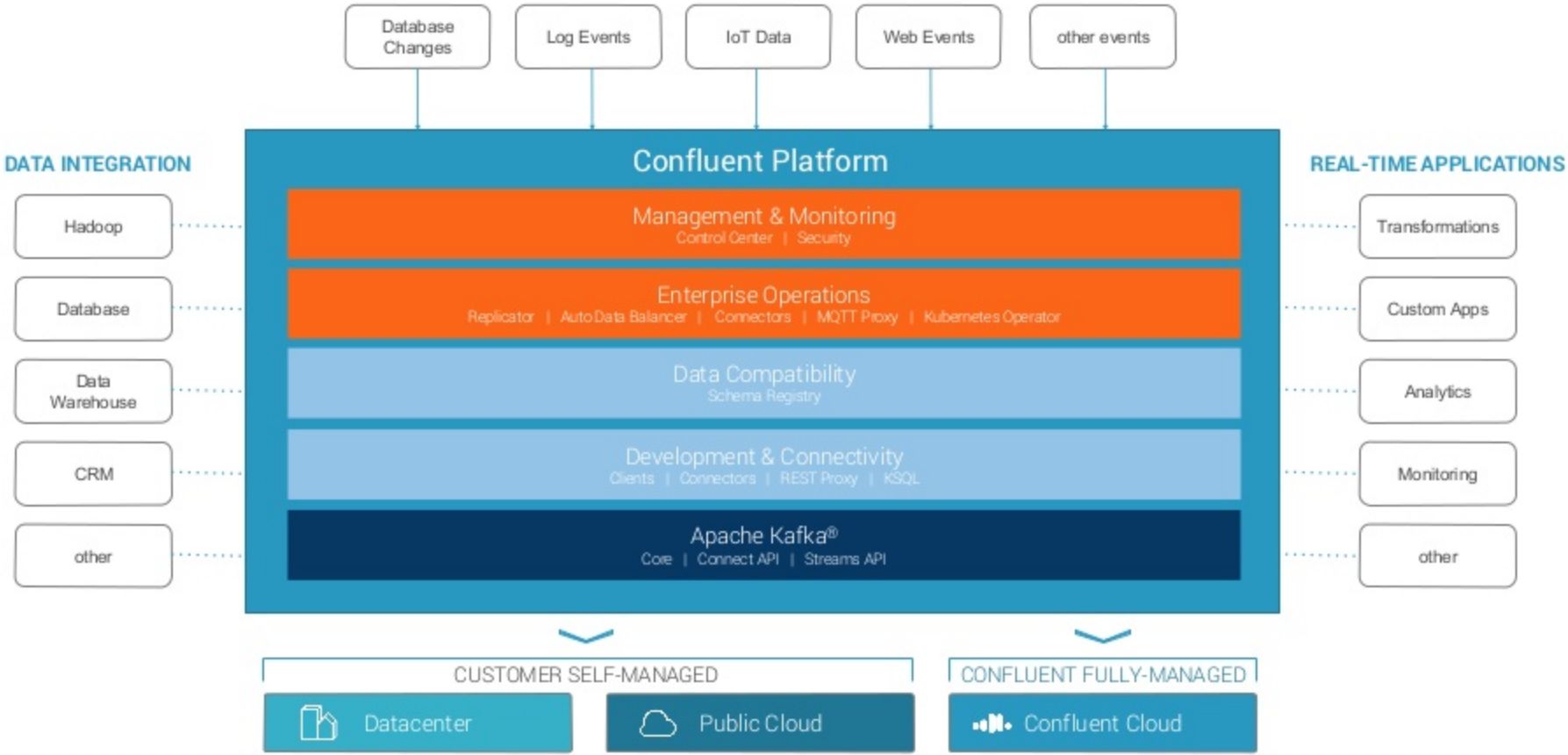
PayPal

Poll

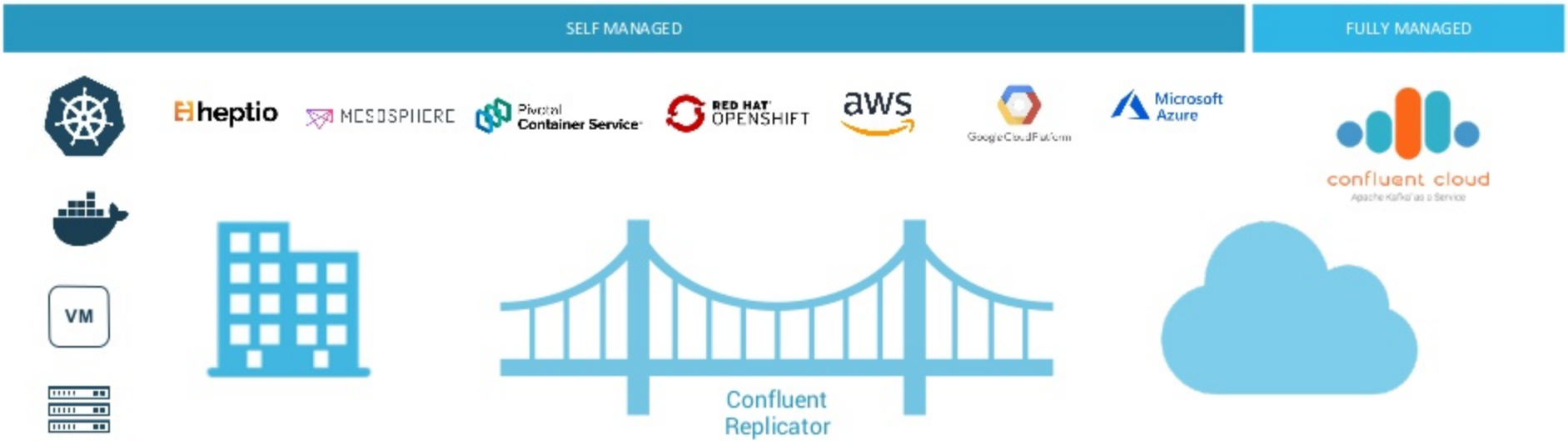
Which of the following use cases are you most likely to to utilize Kafka for over the next year?

1. Data ingestion (processing e.g. batch in Spark)
2. Data pipeline (processing with Kafka)
3. Stream processing (e.g. Kafka Streams, KSQL)
4. Stream processing with machine learning
5. Other (like microservices, event sourcing, storage)

Confluent Delivers a Mission-Critical Event Streaming Platform



Best-of-breed Platforms, Partners and Services for Multi-cloud Streams

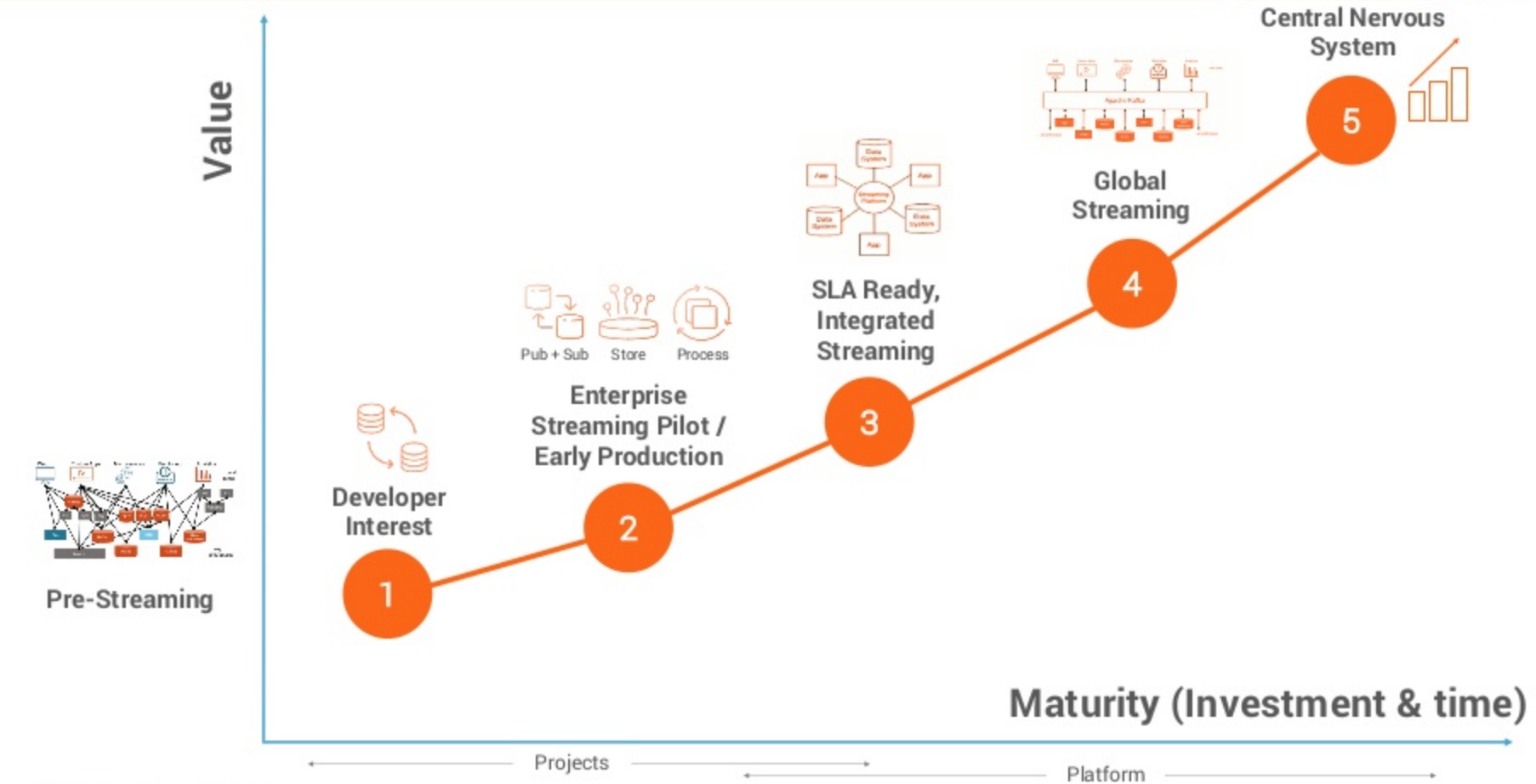


Private Cloud
Deploy on bare-metal, VMs, containers or Kubernetes in your datacenter with **Confluent Platform** and **Confluent Operator**

Hybrid Cloud
Build a persistent bridge between datacenter and cloud with **Confluent Replicator**

Public Cloud
Implement self-managed in the public cloud or adopt a fully managed service with **Confluent Cloud**

Confluent's Streaming Maturity Model - where are you?



Resources and Next Steps

https://confluent.io

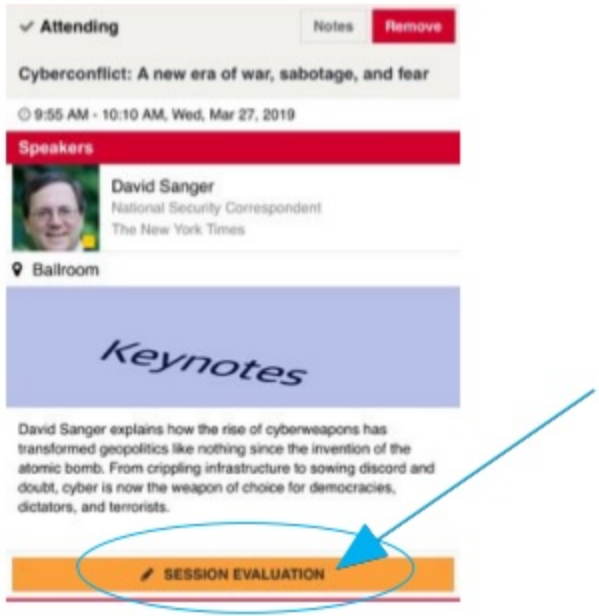
https://confluent.io/cloud

https://slackpass.io/confluentcommunity
#confluent-cloud

@confluentinc

Rate today's session → Please rate my session!

Session page on conference website



O'Reilly Events App

confluent

kafka TensorFlow H2O.ai

Google Cloud Platform amazon web services Microsoft Azure kubernetes

Questions? Feedback? Let's connect!

Kai Waehner

Technology Evangelist

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