CONFLUENT

Resilient Real-time Data Streaming across the Edge and Hybrid Cloud

Use Cases, Architectures, and Examples for Data in Motion powered by Apache Kafka

Kai Waehner

Field CTO
kai.waehner@confluent.io
linkedin.com/in/kaiwaehner
confluent.io
kai-waehner.de
@KaiWaehner

KAI WAEHNER

Agenda

1) Resilient enterprise architectures
2) Real-time data streaming with the Apache Kafka ecosystem
3) Cloud-first and serverless Industrial IoT in automotive
4) Multi-region infrastructure for core banking
5) Hybrid cloud for customer experiences in retail
6) Disconnected edge for safety and security in the public sector

Agenda

1) **Resilient enterprise architectures**
2) Real-time data streaming with the Apache Kafka ecosystem
3) Cloud-first and serverless Industrial IoT in automotive
4) Multi-region infrastructure for core banking
5) Hybrid cloud for customer experiences in retail
6) Disconnected edge for safety and security in the public sector

AWS Cloud Outage hit Disney World Visitors…

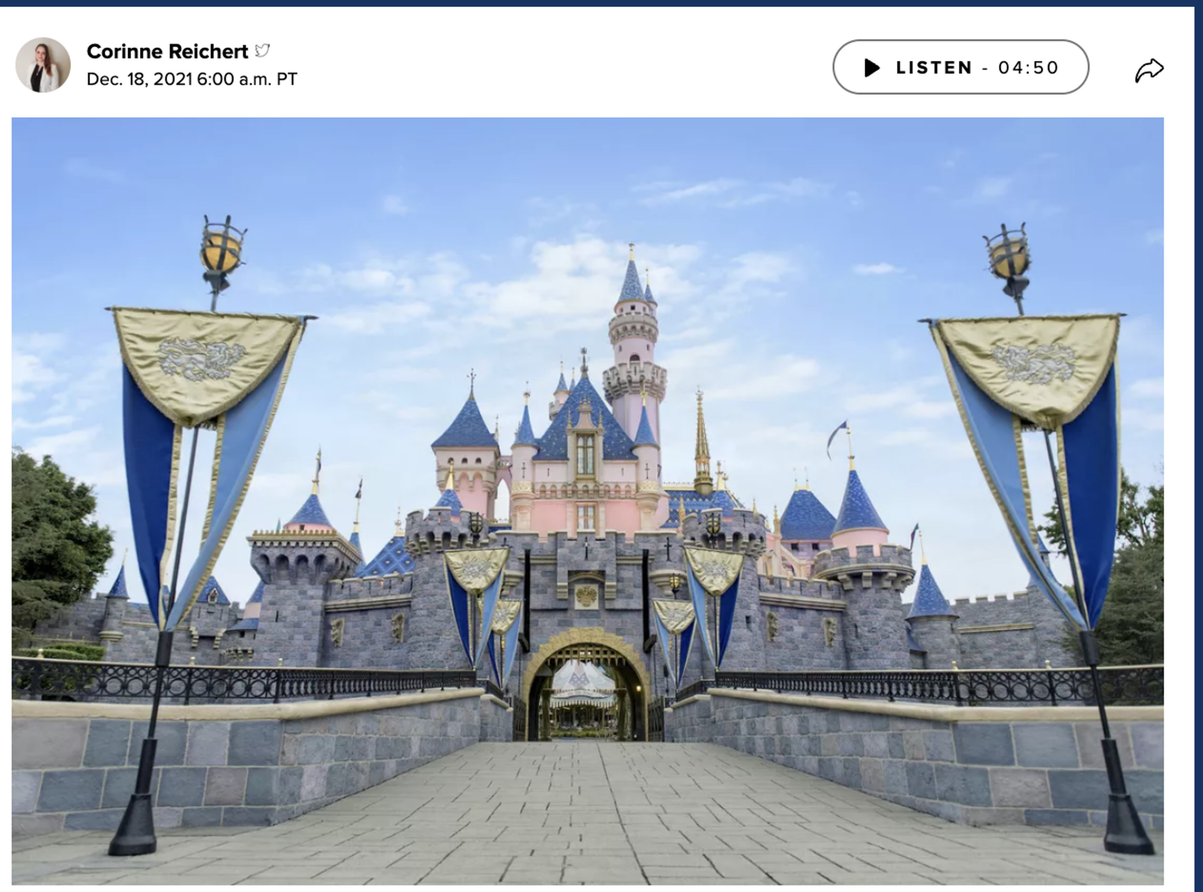
Disney parks were already facing heat from fans. Then an AWS outage came along

As Disney increasingly leans on apps for almost every facet of guest experience, tech problems have a wide-reaching impact on expensive days in the theme parks.

Not even Disney's vaunted magic could save its Disneyland park app from a widespread Amazon Web Services outage temporarily wrecking the day for its guests this week. But for fans of "the happiest place on Earth," this was just the latest in a string of problems.

Disney has been increasingly pushing its theme park guests to use their mobile devices to do everything from ordering food to accessing tickets and park reservations. It has also put a new paid version of its FastPass system, now re-branded Genie Plus, into the app. That means outages, including one that hit Walt Disney World last week, can bring enjoyment in the parks to a screeching halt.

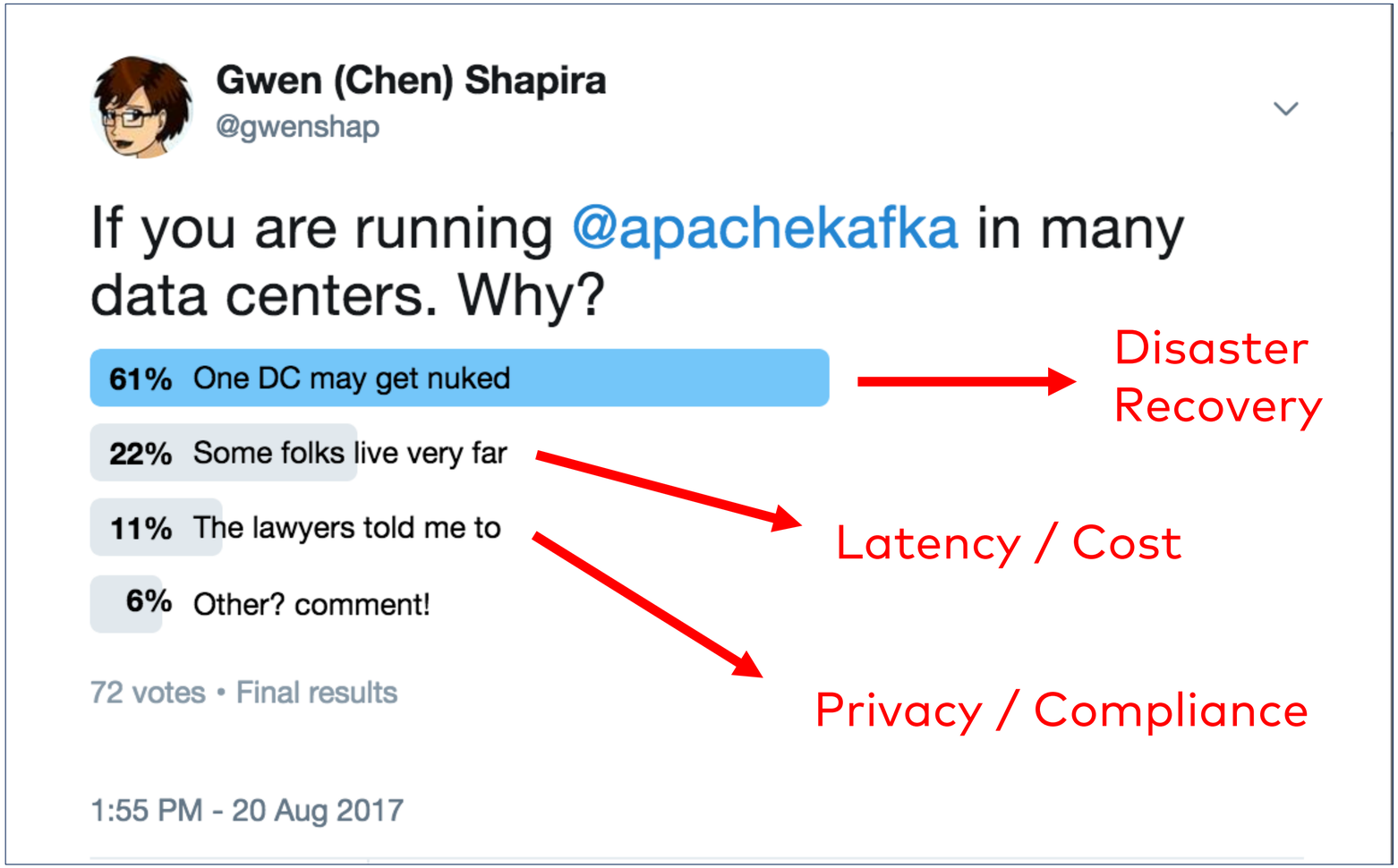
https://www.cnet.com/tech/services-and-software/disney-parks-were-already-facing-heat-from-fans-then-an-aws-outage-came-along/



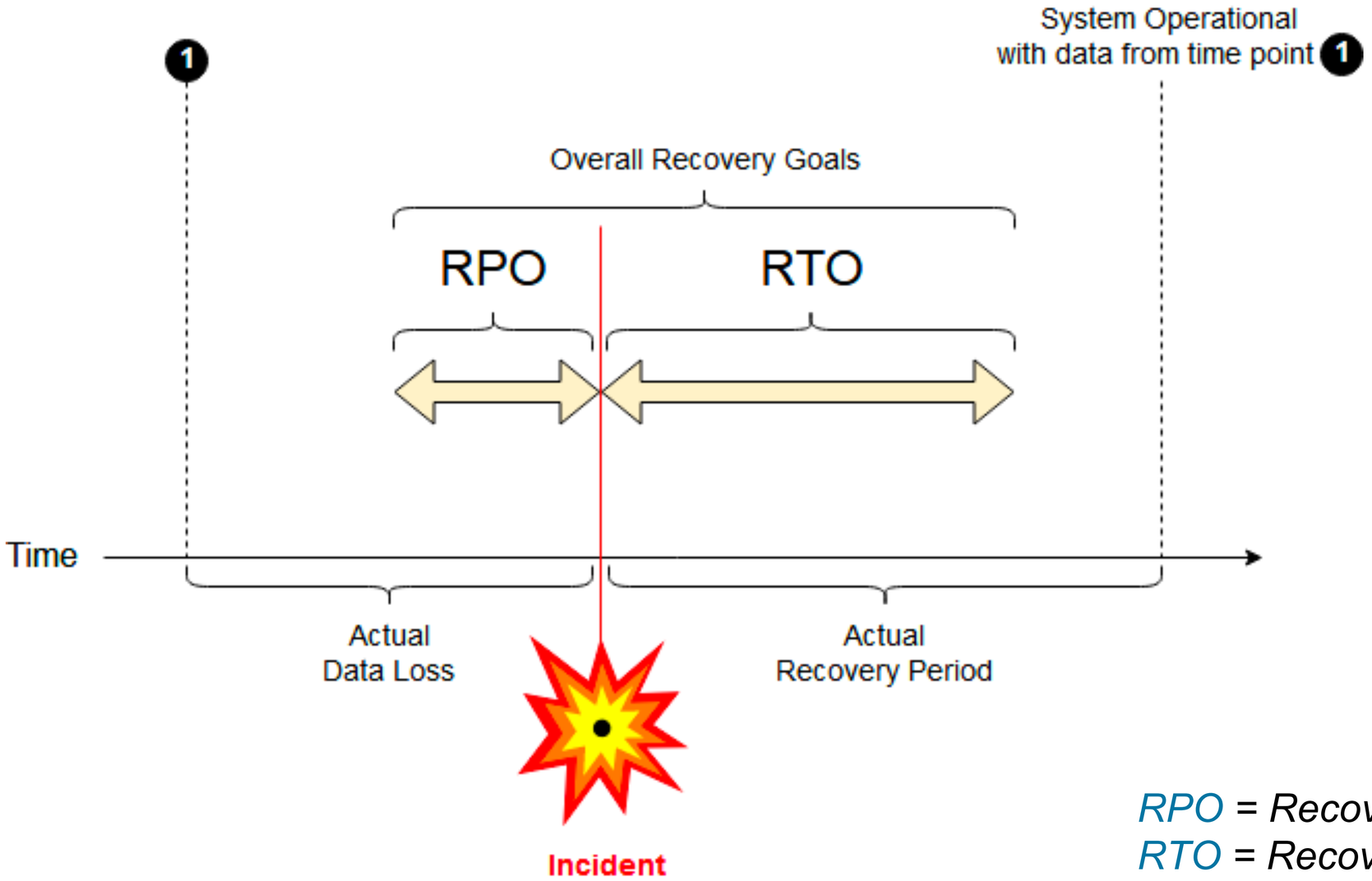
More and more Disneyland services require a mobile app.

Disney Parks

Why one data center or cloud region is not good enough?



Disaster Recovery – RPO and RTO



RPO = *Recovery Point Objective*
RTO = *Recovery Time Objective*

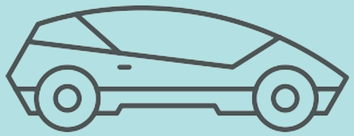
ZERO RPO requires synchronous replication

ZERO RTO requires seamless failover

Agenda

1) Resilient enterprise architectures
2) **Real-time data streaming with the Apache Kafka ecosystem**
3) Cloud-first and serverless Industrial IoT in automotive
4) Multi-region infrastructure for core banking
5) Hybrid cloud for customer experiences in retail
6) Disconnected edge for safety and security in the public sector

Real-time Data in Motion beats Slow Data.



Transportation

Predictive maintenance

Driver-rider match

ETA updates

Banking

Instant payments

Fraud detection

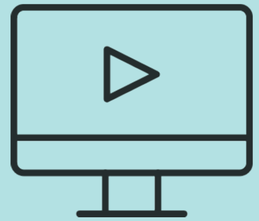
Mobile applications / customer experience

Retail

Real-time inventory

Real-time POS reporting

Personalization



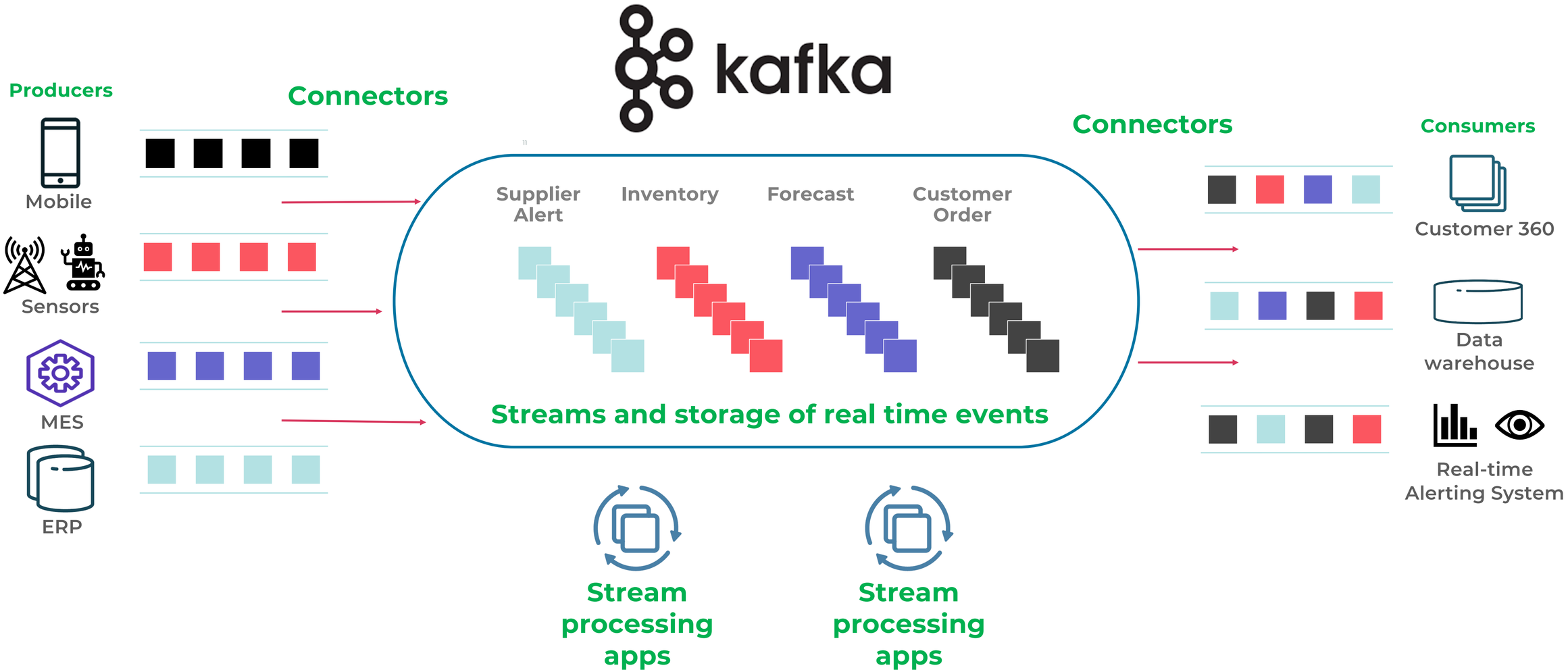
Entertainment

Real-time recommendations

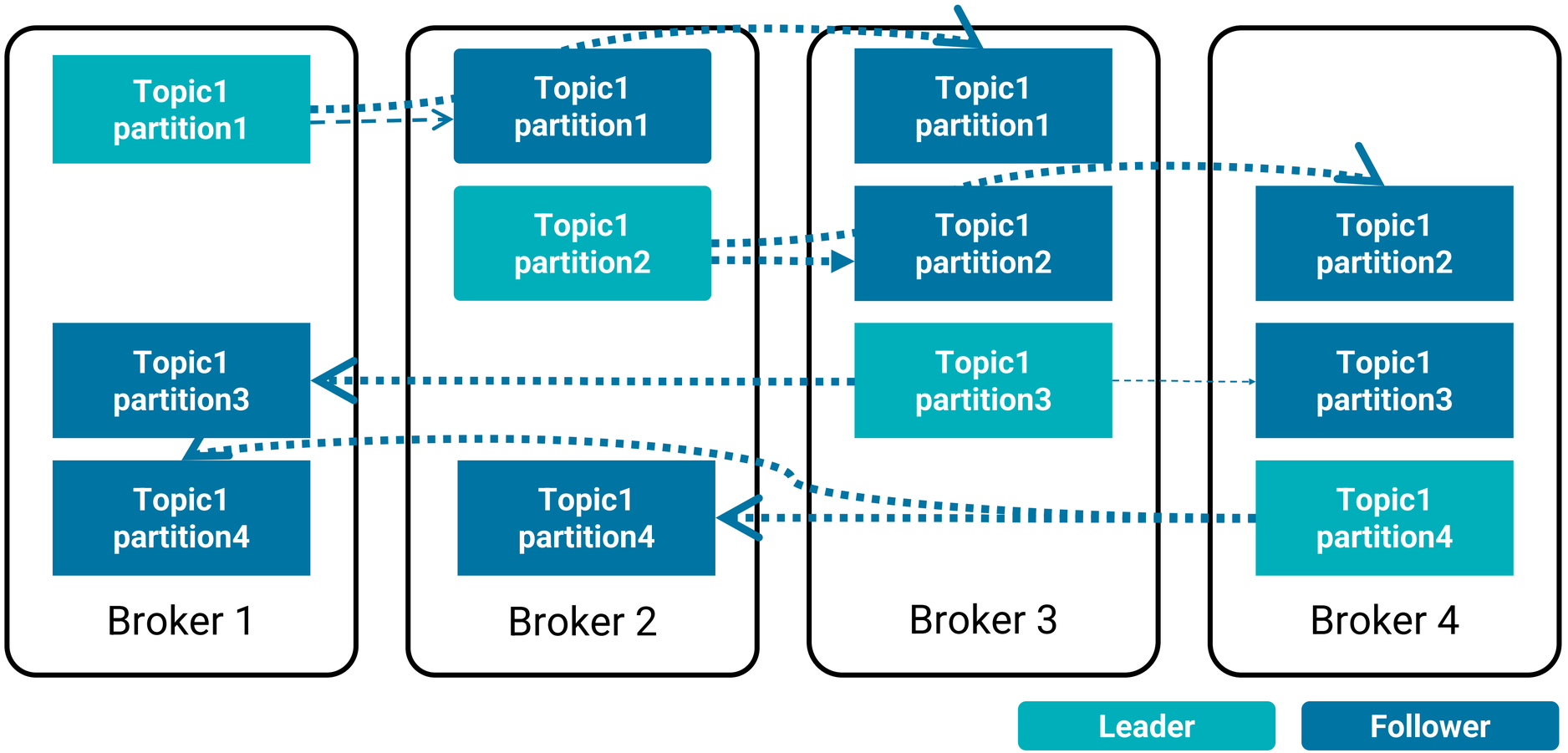
Personalized news feed

In-car purchases

Apache Kafka is the Platform for Data in Motion



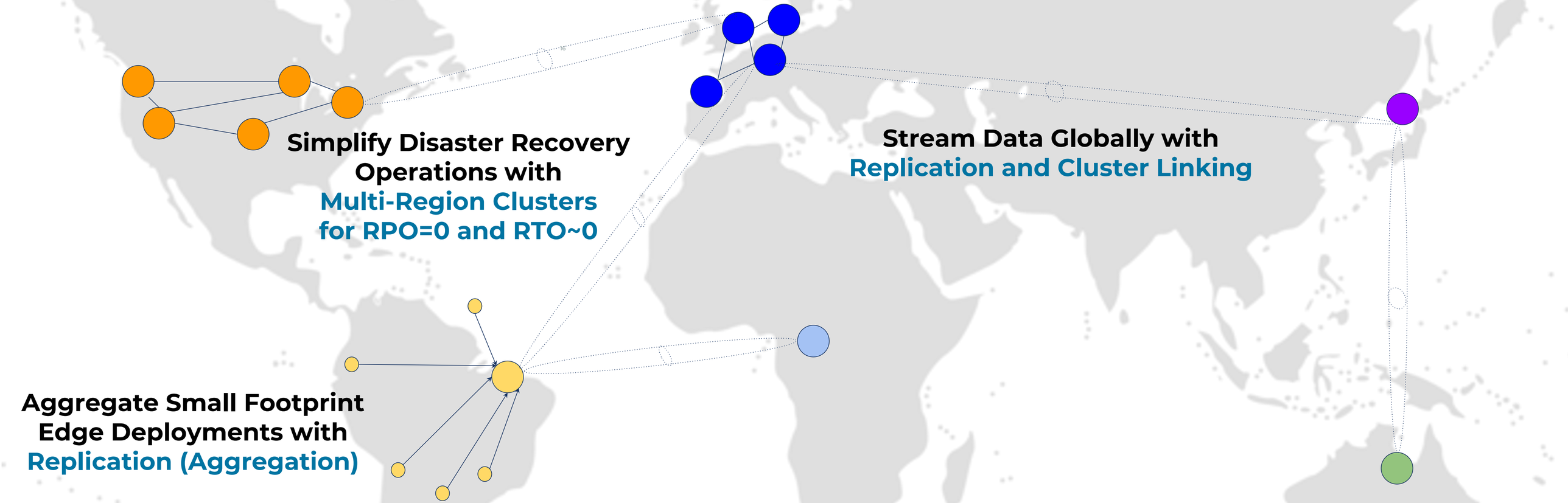
Apache Kafka =
A Resilient, Distributed System



Resilient Data Streaming across Edge and Hybrid Cloud

Streaming Replication between Kafka Clusters

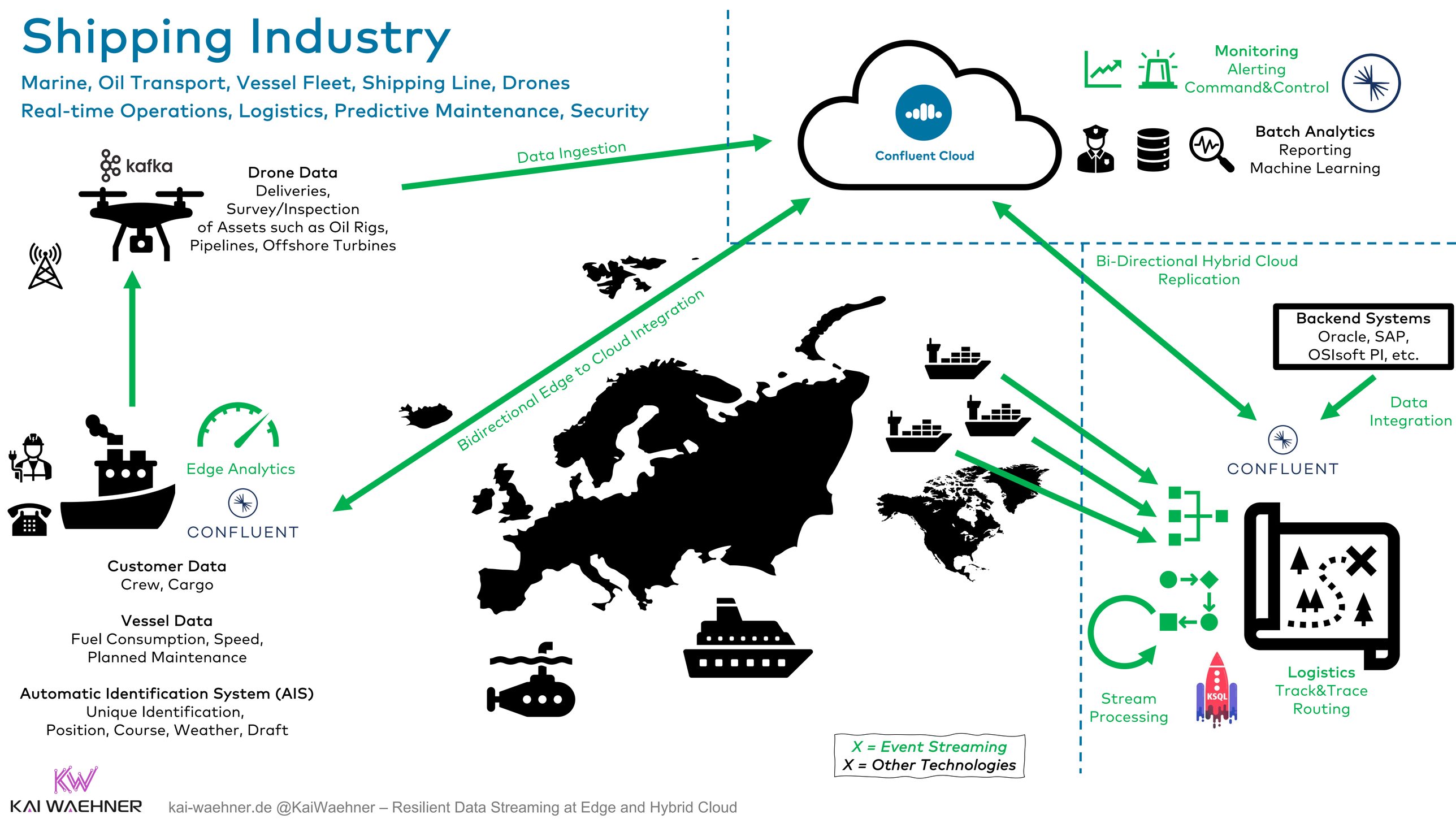
Bridge to Databases, Data Lakes, Apps, APIs, SaaS



Shipping Industry

Marine, Oil Transport, Vessel Fleet, Shipping Line, Drones

Real-time Operations, Logistics, Predictive Maintenance, Security



Agenda

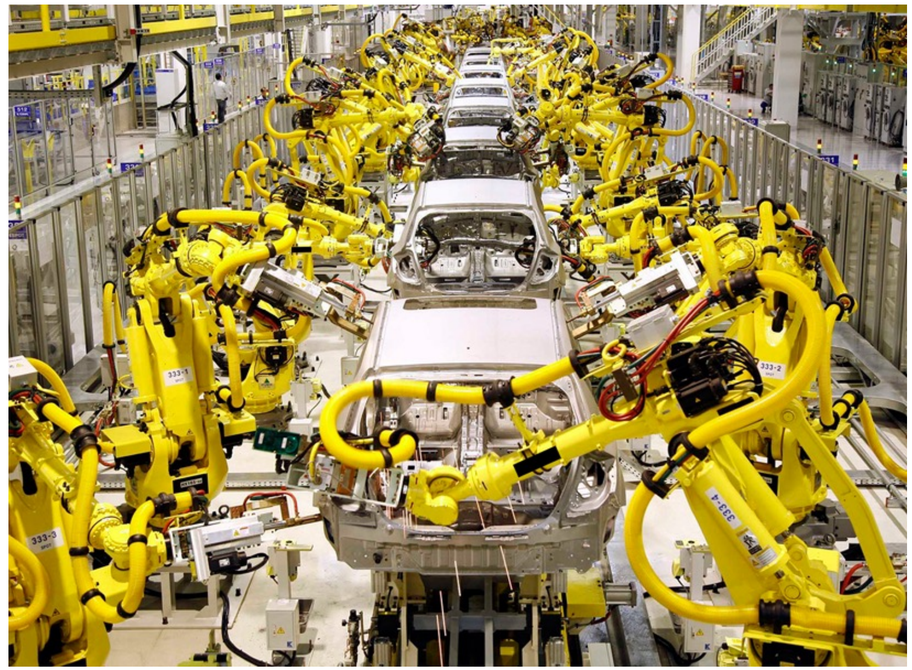
1) Resilient enterprise architectures
2) Real-time data streaming with the Apache Kafka ecosystem
3) **Cloud-first and serverless Industrial IoT in automotive**
4) Multi-region infrastructure for core banking
5) Hybrid cloud for customer experiences in retail
6) Disconnected edge for safety and security in the public sector

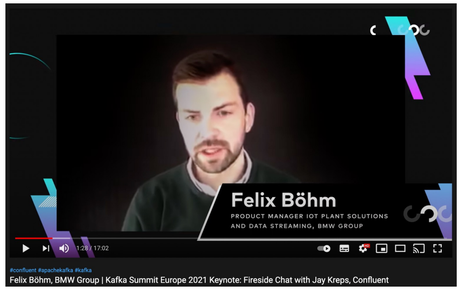
BMW Group

Mission-critical workloads across the edge and cloud

- Why Kafka? **Decoupling. Transparency. Innovation.**
- Why Confluent? **Stability is key in manufacturing**
- Decoupling between logistics and production systems
- **Cloud-first event streaming on Azure Cloud with serverless Confluent Cloud**

- **Use case**
  - **Logistics and supply chain in global plants**
  - Right stock in place (physically and in ERP systems like SAP)
  - Just in time, just in sequence
  - Lot of critical applications



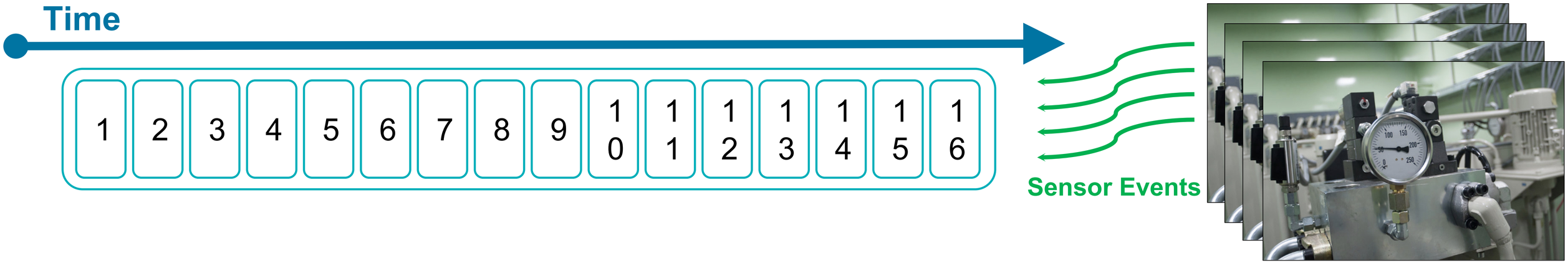


Jay Kreps, Confluent CEO
Felix Böhm, BMW Plant Digitalization and Cloud Transformation

Keynote at Kafka Summit Eurpoe 2021:
https://www.youtube.com/watch?v=3cG2ud7TRs4

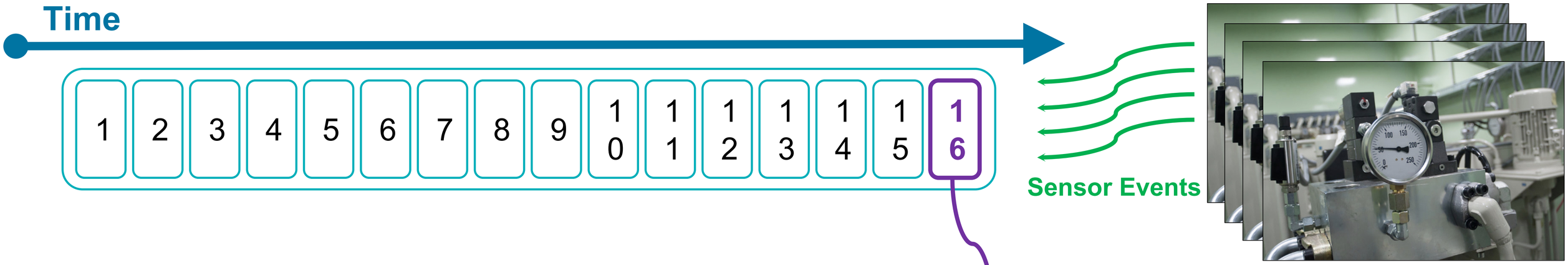
Condition Monitoring and Predictive Maintenance

Stateless and stateful stream processing for real-time data correlation with Kafka-native tools (Kafka Streams / ksqlDB)



Condition Monitoring and Predictive Maintenance

Stateless and stateful stream processing for real-time data correlation with Kafka-native tools (Kafka Streams / ksqlDB)




```
builder
  .stream(”temperature-sensor")
  .filter((key, sensor-data) ->
        sensor-data.temperature > 100)
  .to(”temperature-spikes");
```

kafka Streams

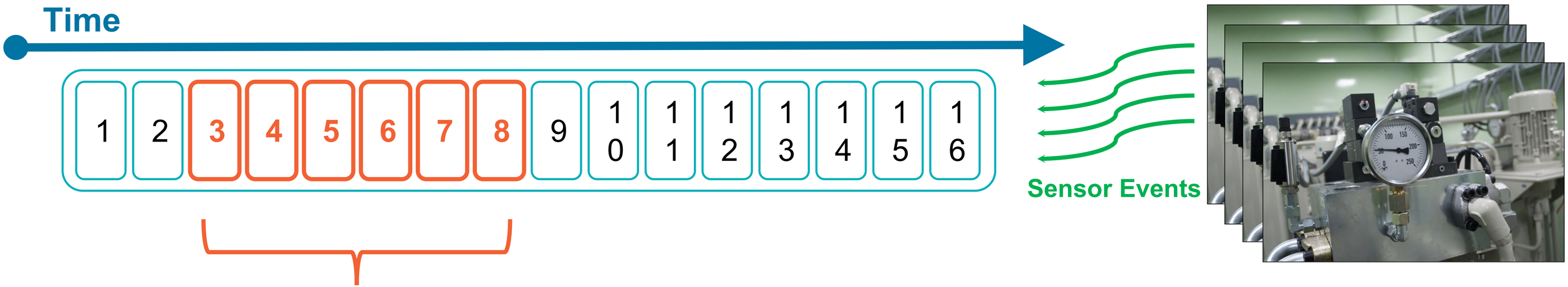
Stateless Filter Above-Threshold Events

Condition Monitoring

(Temperature Spikes)

Condition Monitoring and Predictive Maintenance

Stateless and stateful stream processing for real-time data correlation with Kafka-native tools (Kafka Streams / ksqlDB)



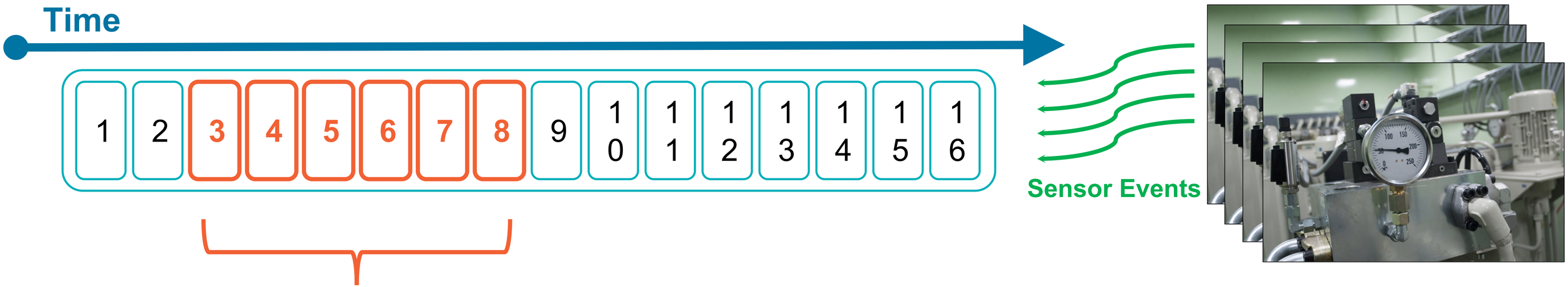
Predictive Maintenance

(Continuous Anomaly Detection)

```
CREATE TABLE anomaly_detection AS
  SELECT temperature_spike_id, COUNT(*) AS total_spikes,
       AVG(temperature) AS avg_temperature
  FROM sensor-data
  WINDOW TUMBLING (SIZE 1 HOUR)
  GROUP BY temperature_spike_id
  EMIT CHANGES;
```

Condition Monitoring and Predictive Maintenance

Stateless and stateful stream processing for real-time data correlation with Kafka-native tools (Kafka Streams / ksqlDB)



KSQL

TensorFlow

Real-time Machine Learning

Predictive Maintenance

(Continuous Anomaly Detection)

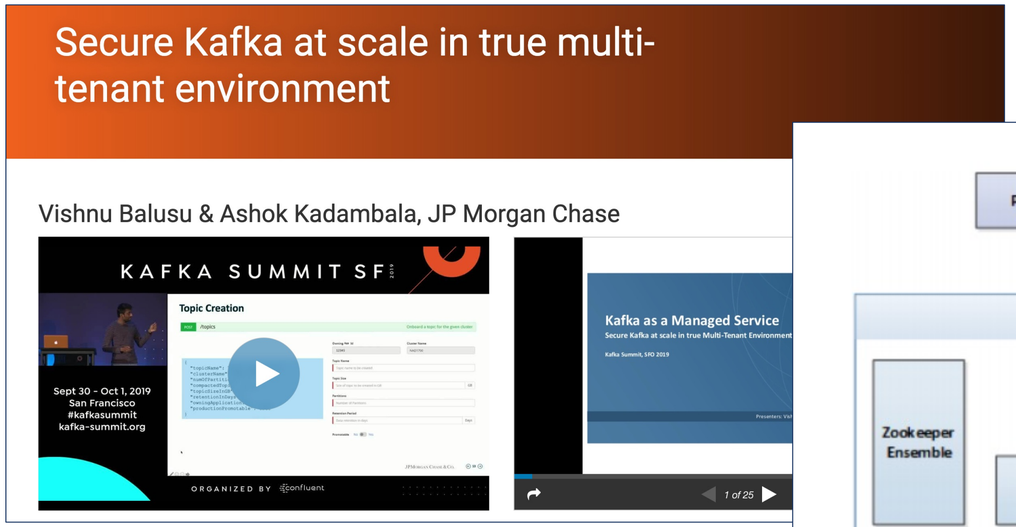
```
CREATE STREAM anomaly_detection AS
SELECT sensor_id, detect_anomaly(sensor_values)
FROM machine;
```

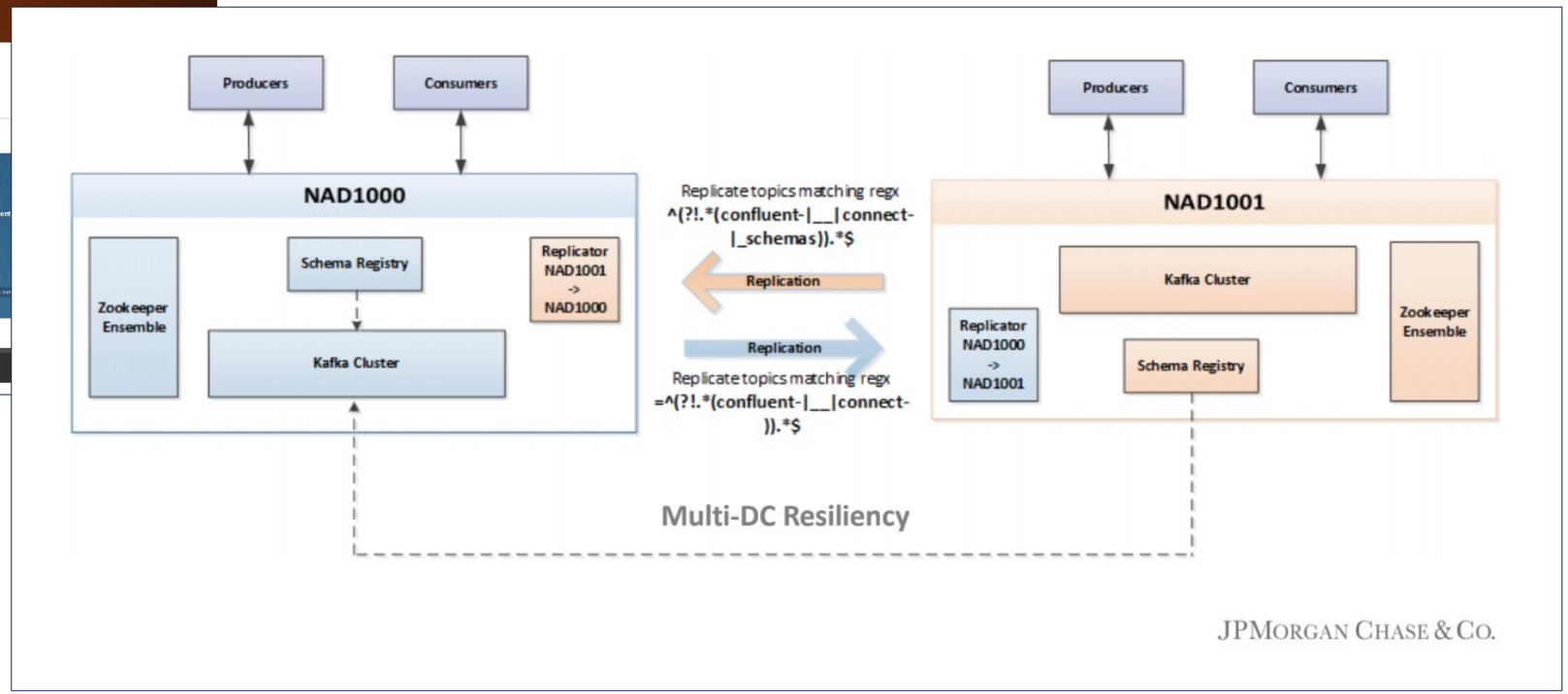
TensorFlow model embedded in User Defined Function (UDF)

Agenda

1) Resilient enterprise architectures
2) Real-time data streaming with the Apache Kafka ecosystem
3) Cloud-first and serverless Industrial IoT in automotive
4) **Multi-region infrastructure for core banking**
5) Hybrid cloud for customer experiences in retail
6) Disconnected edge for safety and security in the public sector

Disaster Recovery @ JPMorgan



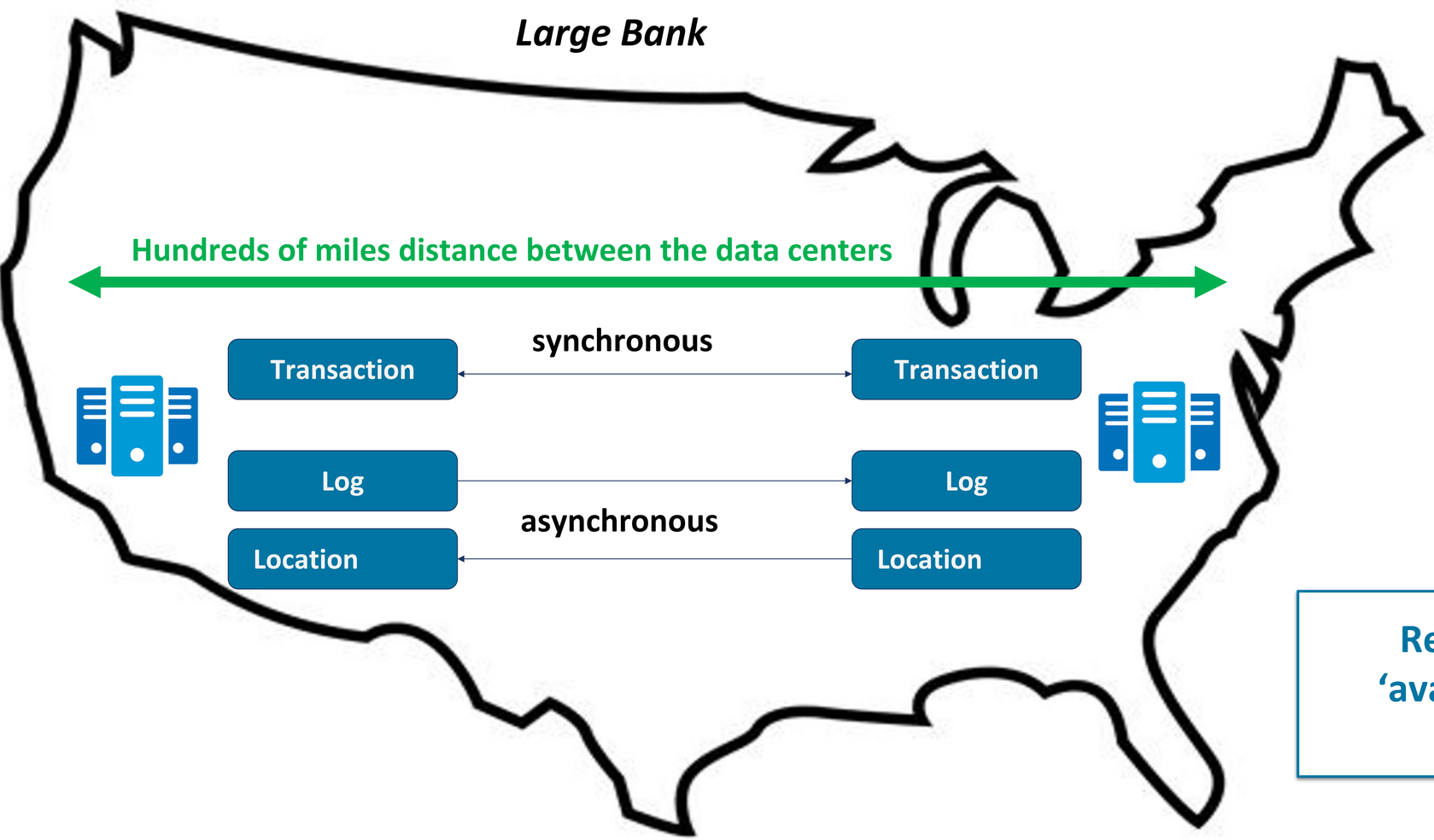


JPMorganChase

https://www.confluent.io/kafka-summit-san-francisco-2019/secure-kafka-at-scale-in-true-multi-tenant-environment

Multi-Region Kafka Cluster in Financial Services

Zero downtime + zero data loss (RPO=0 and RTO~0) + automated disaster recovery

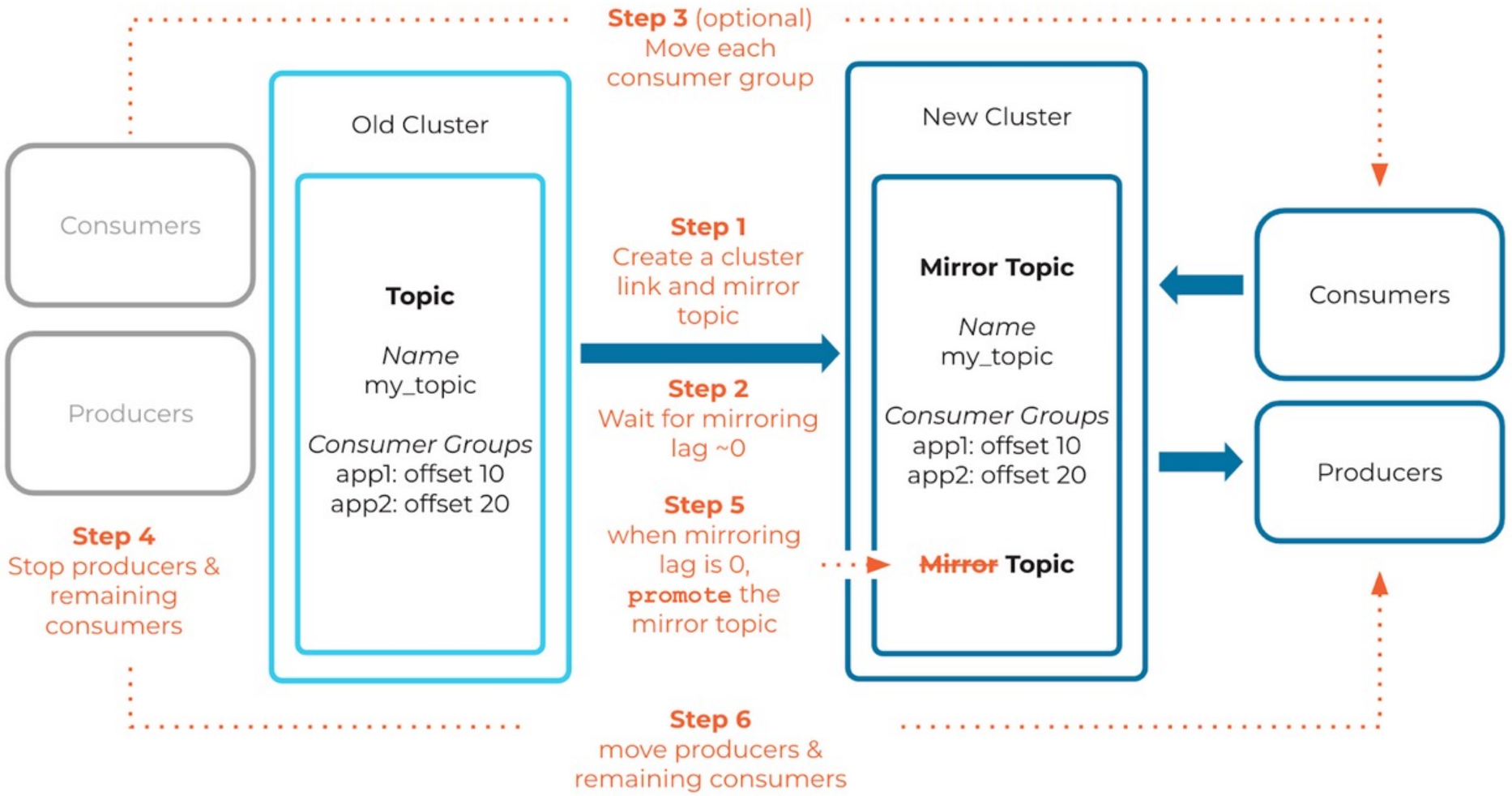


- 'Payment' transactions enter from us-east and us-west with fully **synchronous replication**
- 'Log' and 'Location' information in the same cluster use async - **optimized for latency**
- **Automated disaster recovery** (zero downtime, zero data loss)

Result: Clearing time from 'deposit' to 'available' goes from 5 days to 5 seconds (including security checks)

(Only available in Confluent Platform)

Migration with Cluster Linking

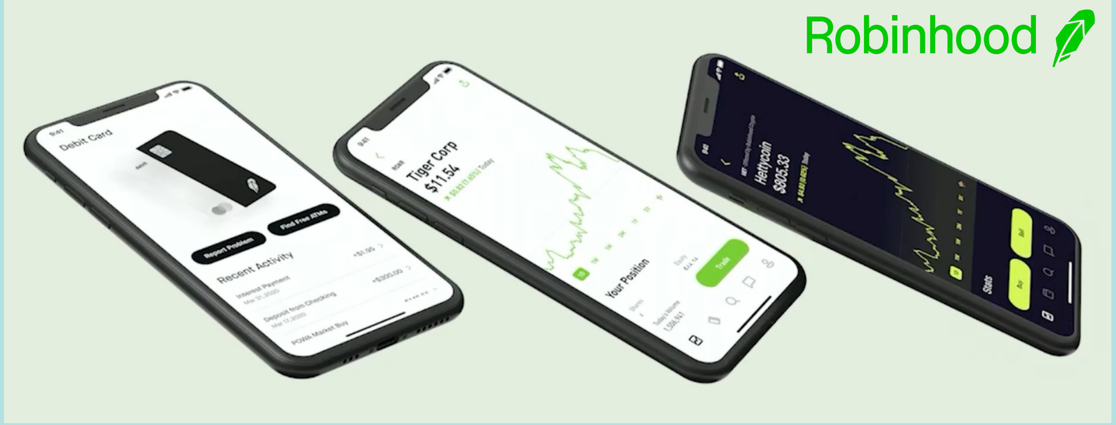


Robinhood

Mission: "Democratize finance for all"

Kafka for mission-critical and analytics use cases

Microservices using various technologies

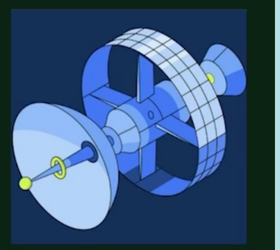


Kafka use cases at Robinhood

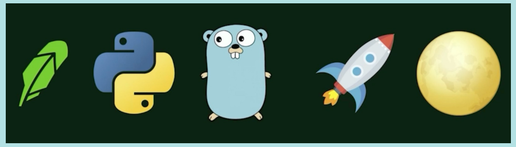
Stock trading

Clearing

Crypto trading



Messaging and push notifications



kafkahood

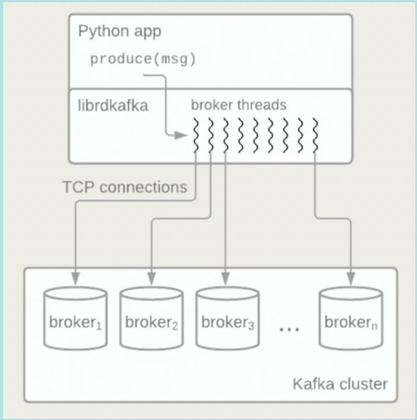
Taming non-conforming and potentially dangerous application patterns via Robinhood's **internal wrapper library** for Python and Golang Kafka clients

kafkaproxy

Taming a massive fleet of Python Kafka clients via a sidecar proxy written in Rust

FAUST

Clusters	Brokers	Topics/partitions
6	160	1400 topics 90k partitions
Messages/second (all clusters)	**API requests/sec (largest cluster)**	**Inbound client connections (largest cluster)**
2.2M	500k	> 100k

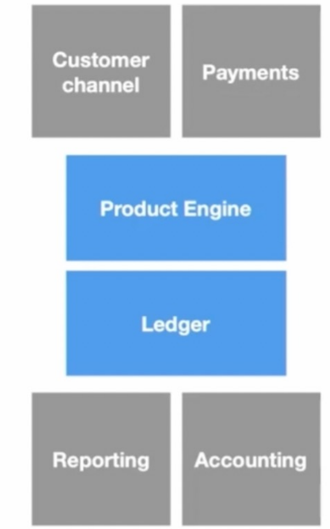


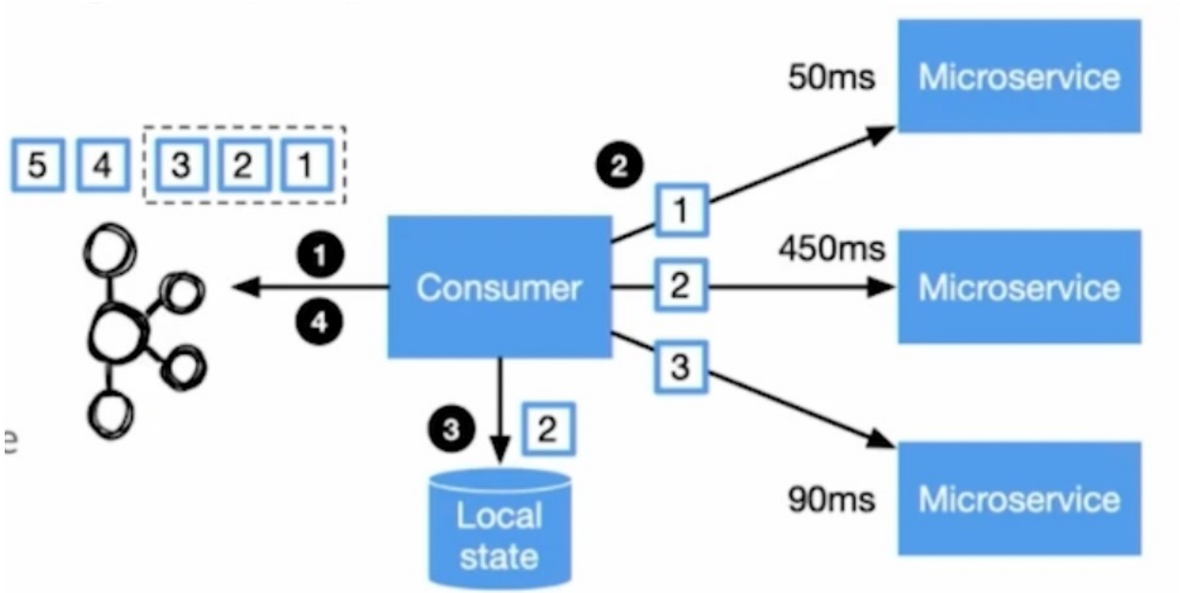
https://www.confluent.io/events/kafka-summit-americas-2021/taming-a-massive-fleet-of-python-based-kafka-apps-at-robinhood/

Thought Machine – Core Banking

Thought Machine

- Cloud-native core banking software
- Transactional workloads (24/7, zero data loss)
- Flexible product engine powered by smart contracts (not blockchain)





Al Rajhi Bank · ARVEST · Atom bank · CURVE · IMPACT CREDIT SOLUTIONS · ING
INTESA SANPAOLO · JPMorgan Chase · LLOYDS BANK · LUNAR · Mascoma Bank · MAGYAR BANKHOLDING
monese · SEB · standard chartered · transferGo

https://www.confluent.io/events/kafka-summit-apac-2021/scaling-a-core-banking-engine-using-apache-kafka/

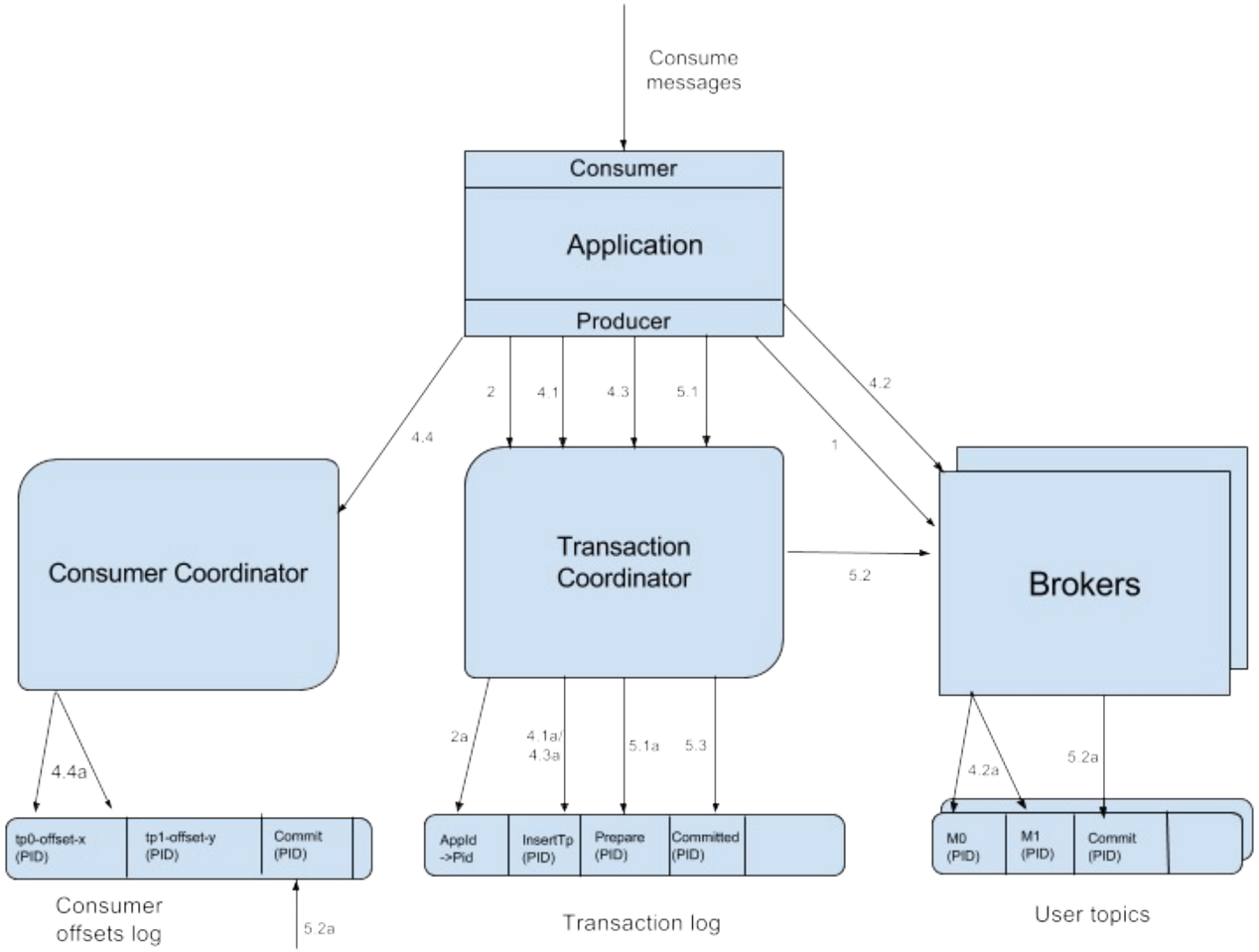
“Transactions” in Apache Kafka

Exactly-Once Semantics (EOS)

available since Kafka 0.11 (June 2017):

```
producer.initTransactions();
try {
  producer.beginTransaction();
  producer.send(record1);
  producer.send(record2);
  producer.commitTransaction();
} catch(ProducerFencedException e) {
  producer.close();
} catch(KafkaException e) {
  producer.abortTransaction();
}
```

kafka

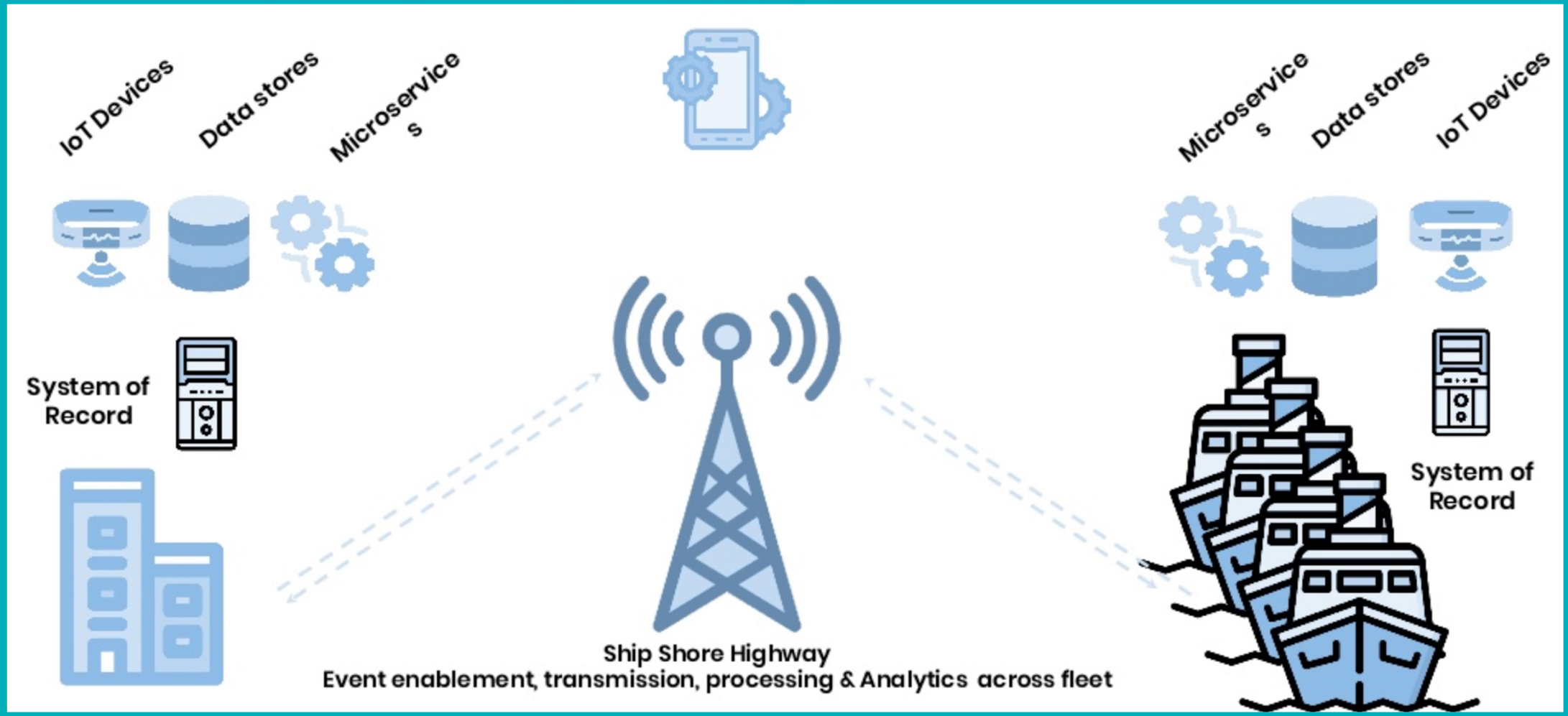


https://cwiki.apache.org/confluence/display/KAFKA/KIP-98+-+Exactly+Once+Delivery+and+Transactional+Messaging
https://www.confluent.io/kafka-summit-london18/dont-repeat-yourself-introducing-exactly-once-semantics-in-apache-kafka/

Agenda

1) Resilient enterprise architectures
2) Real-time data streaming with the Apache Kafka ecosystem
3) Cloud-first and serverless Industrial IoT in automotive
4) Multi-region infrastructure for core banking
5) Hybrid cloud for customer experiences in retail
6) Disconnected edge for safety and security in the public sector

Royal Caribbean - Offline Edge for Swimming Retail Stores



https://www.confluent.io/kafka-summit-lon19/seamless-guest-experience-with-kafka-streams/

Hybrid Retail Architecture

Confluent Cloud

Manager

API

Get report

Context-specific real-time upsell

Payment processing and fraud detection as a service

3rd party payment provider

Customer data

Payment data

Train schedule

Loyalty information

Streams of real time events

CRM

Customer data

CONFLUENT

Customer data

Payment data

Train schedule

Loyalty information

Streams of real time events

Customer

Customer data

Payment data

Train schedule

Loyalty information

Streams of real time events

Customer

CONFLUENT

Event Streaming at the Edge in the Smart Retail Store

Item Availability

Local Inventory Management

Global Inventory Management

Confluent Cloud

Customer data

Payment data

Train schedule

Loyalty information

Streams of real time events

Point of Sale (POS)

Payment

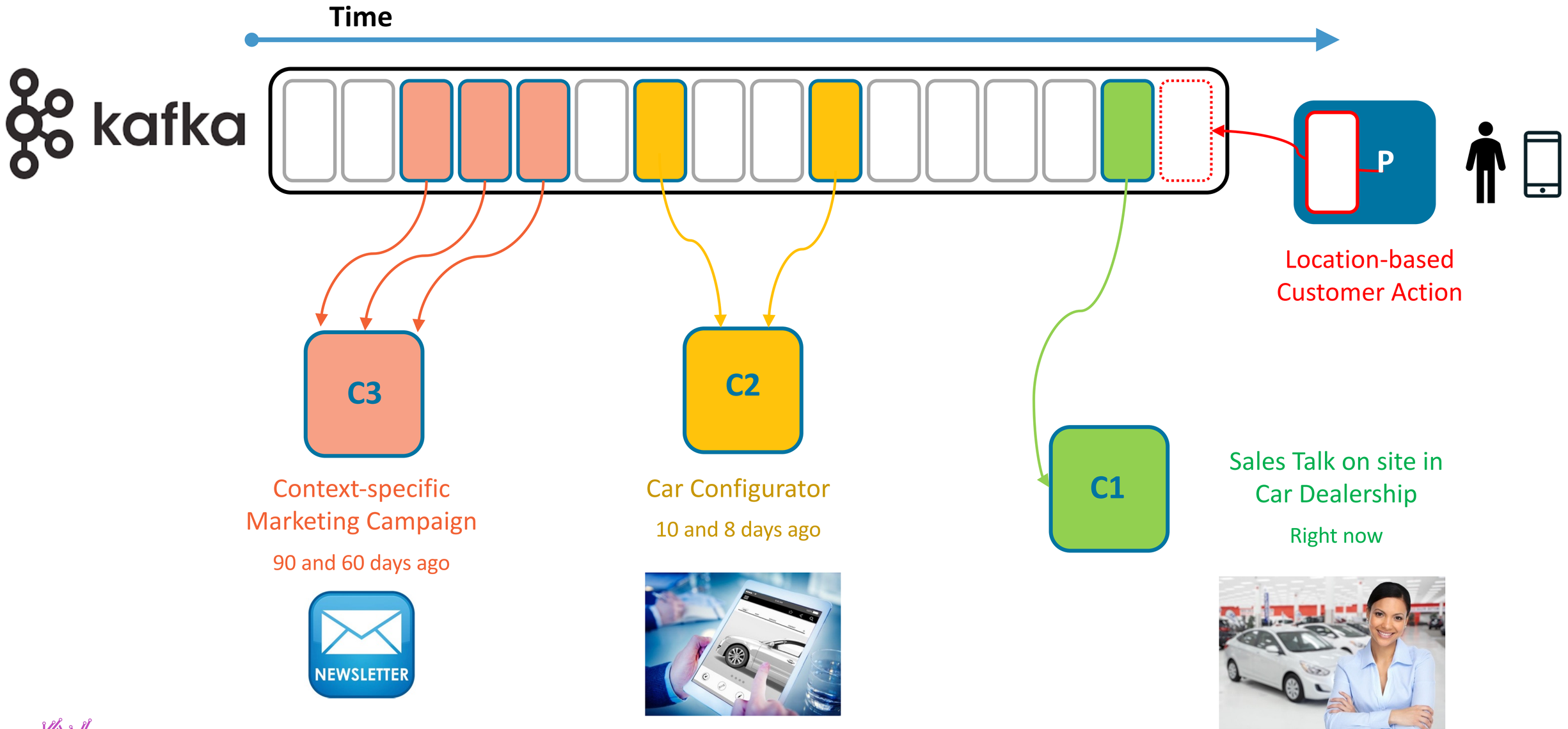
Discount

Loyalty System

CONFLUENT

Omnichannel Retail

Customer 360
(Website, Mobile App, On Site in Store, In-Car)



Agenda

1) Resilient enterprise architectures
2) Real-time data streaming with the Apache Kafka ecosystem
3) Cloud-first and serverless Industrial IoT in automotive
4) Multi-region infrastructure for core banking
5) Hybrid cloud for customer experiences in retail
6) **Disconnected edge for safety and security in the public sector**

Data Processing at the Edge

Always on (even “offline”)
Replayability
Cost-efficiency
Low latency

kafka

Time

P

C1

Connect to the gaming server for kids

Play games, earn rewards, communicate with other kids in the train, …

C2

Estimated time of arrival

C3

Know-your-customer

Loyalty app, predictive behavior, …

Devon Energy

Oil & Gas Industry

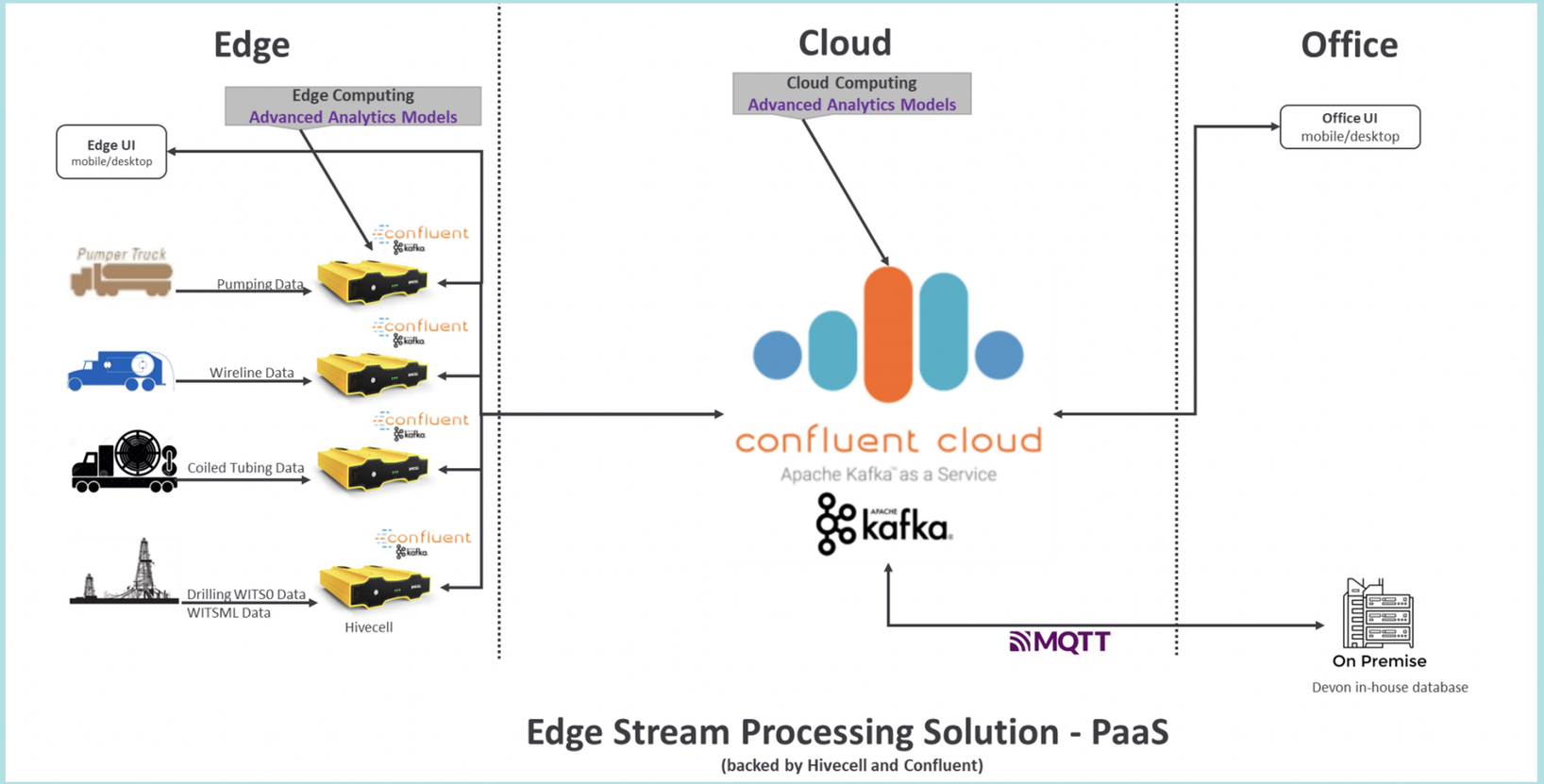
Improve drilling and well completion operations

Edge stream processing/analytics + closed-loop control ready

Vendor agnostic (pumping, wireline, coil, offset wells, drilling operations, producing wells)

Replication to the cloud in real-time at scale

Cloud agnostic (AWS, GCP, Azure)



Source: Energy in Data - Powered by AAPG, SEG & SPE: energyindata.org

Smart Soldiers at the Edge

Confluent Cloud

Command Post running Confluent Platform aggregating information from soliders and other sensor data

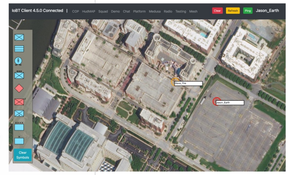
Weather

Personnel

Logistics

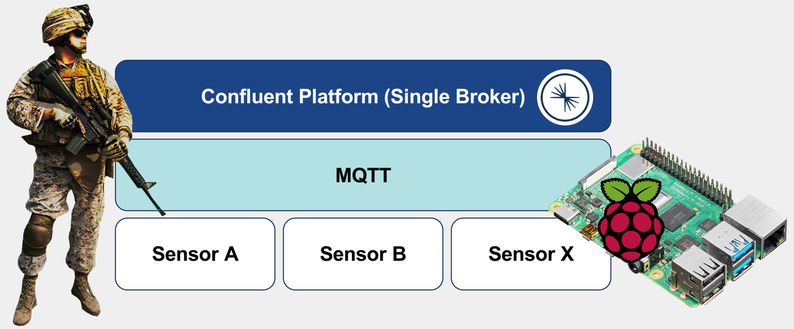
Targets

CONFLUENT



Enhanced Situational Awareness

Sensor data published to Command Post when connected to network

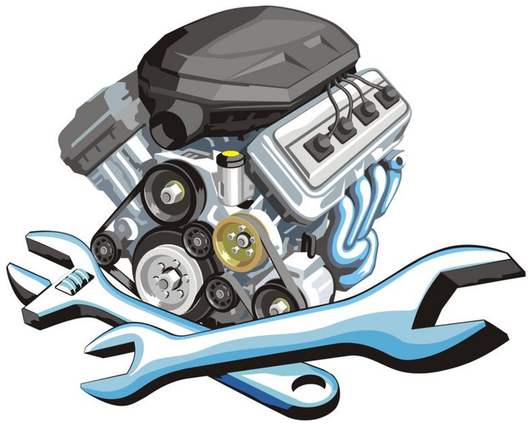


Single Kafka broker deployed on a small computer and leveraging Cluster Linking to publish sensor data to the Command Post.

Why people choose Confluent for building resilient architectures?

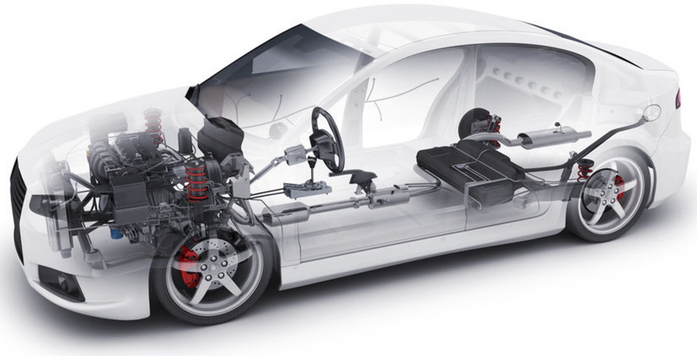
Confluent completes Apache Kafka. Cloud-native. Everywhere.

Car Engine



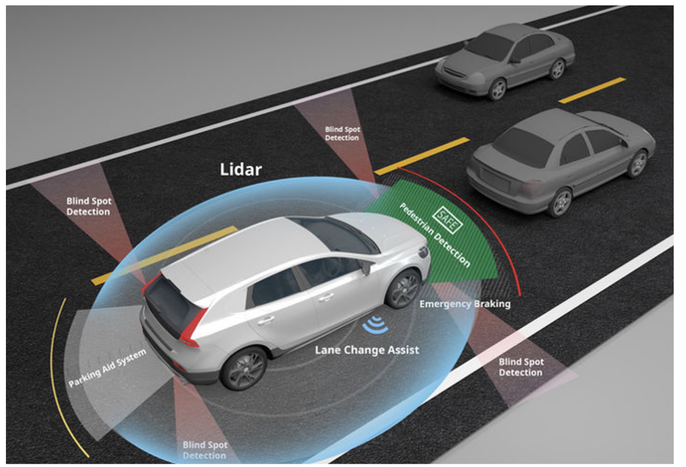
APACHE kafka

Car



CONFLUENT

Self-driving Car



Confluent Cloud

Questions? Feedback?
Let’s connect!

Kai Waehner

Field CTO
Confluent

KAI WAEHNER

kai.waehner@confluent.io
@KaiWaehner
confluent.io
kai-waehner.de
linkedin.com/in/kaiwaehner