



# Best Practices for Streaming IoT Data with MQTT and Apache

## Kafka®

—  
Kai Waehner, Technology Evangelist, Confluent  
Dominik Obermaier, CTO, HiveMQ

# Speakers



**Kai Waehner**

**Technology Evangelist, Confluent**

[kai.waehner@confluent.io](mailto:kai.waehner@confluent.io)

[@KaiWaehner](https://twitter.com/KaiWaehner)



**Dominik Obermaier**

**CTO, HiveMQ**

[dominik.obermaier@hivemq.com](mailto:dominik.obermaier@hivemq.com)

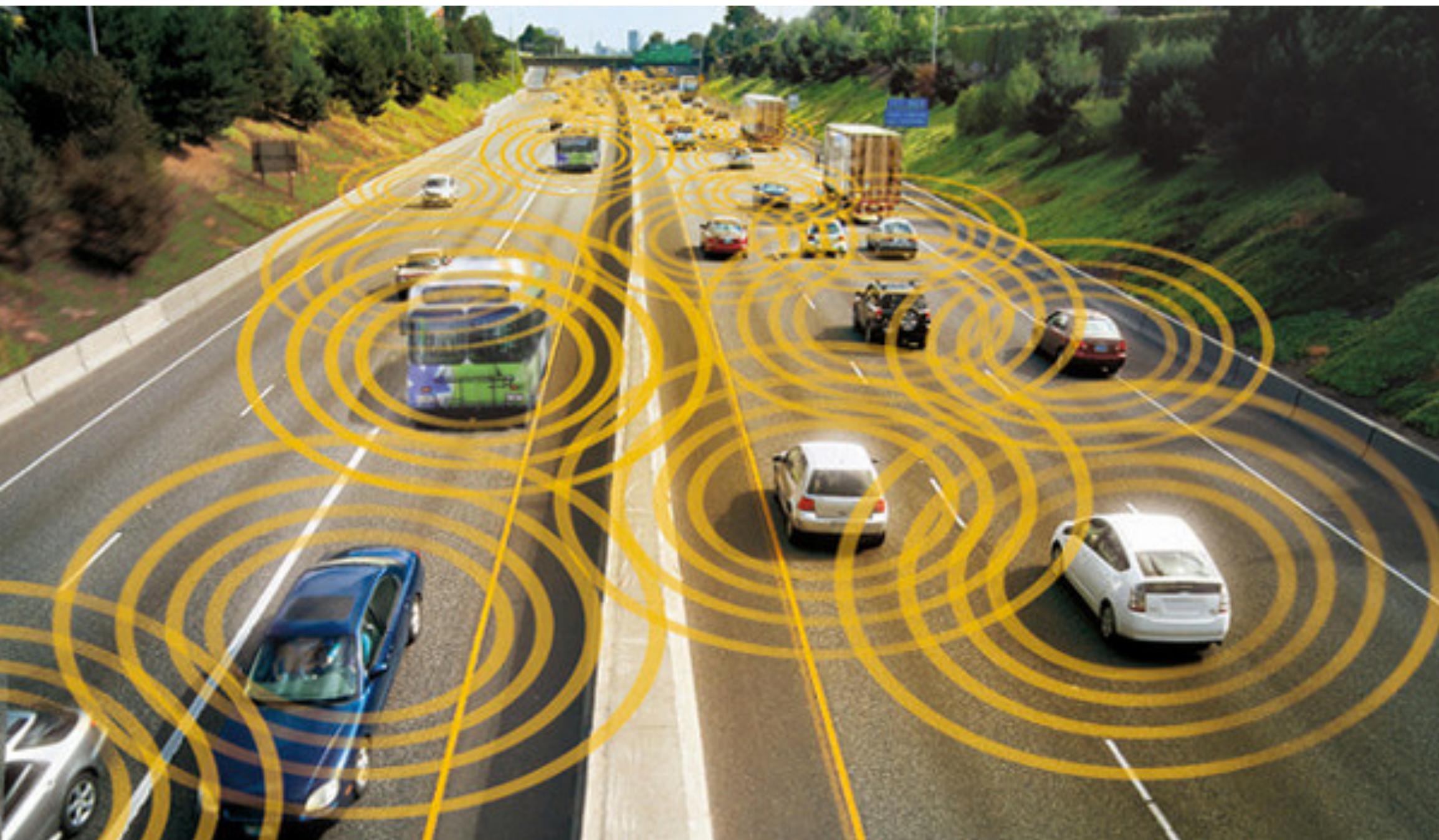
[@dobermai](https://twitter.com/dobermai)



# Agenda

---

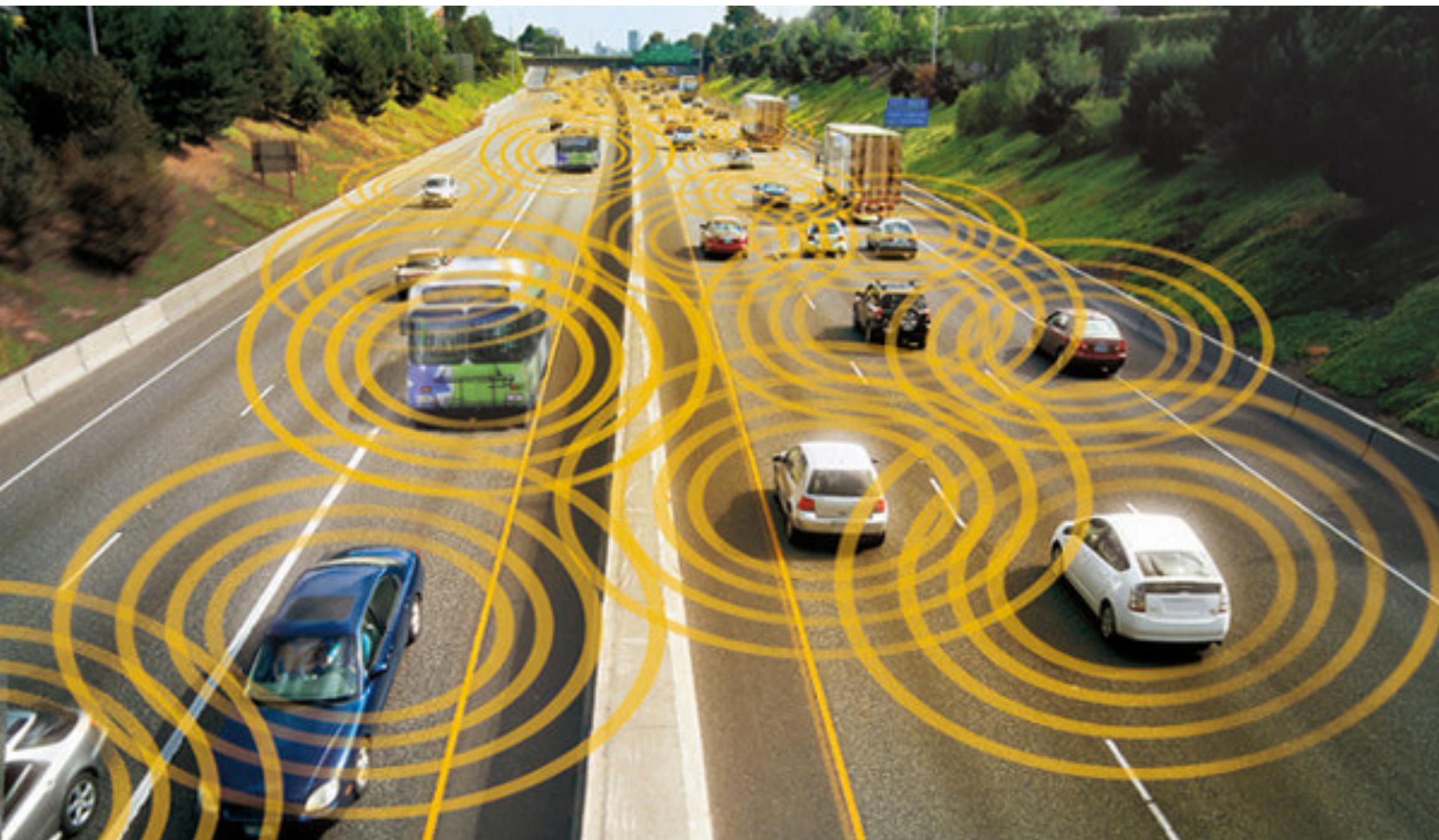
- Use Case
- Architecture
- Live Demo
- Best Practices
- Next steps



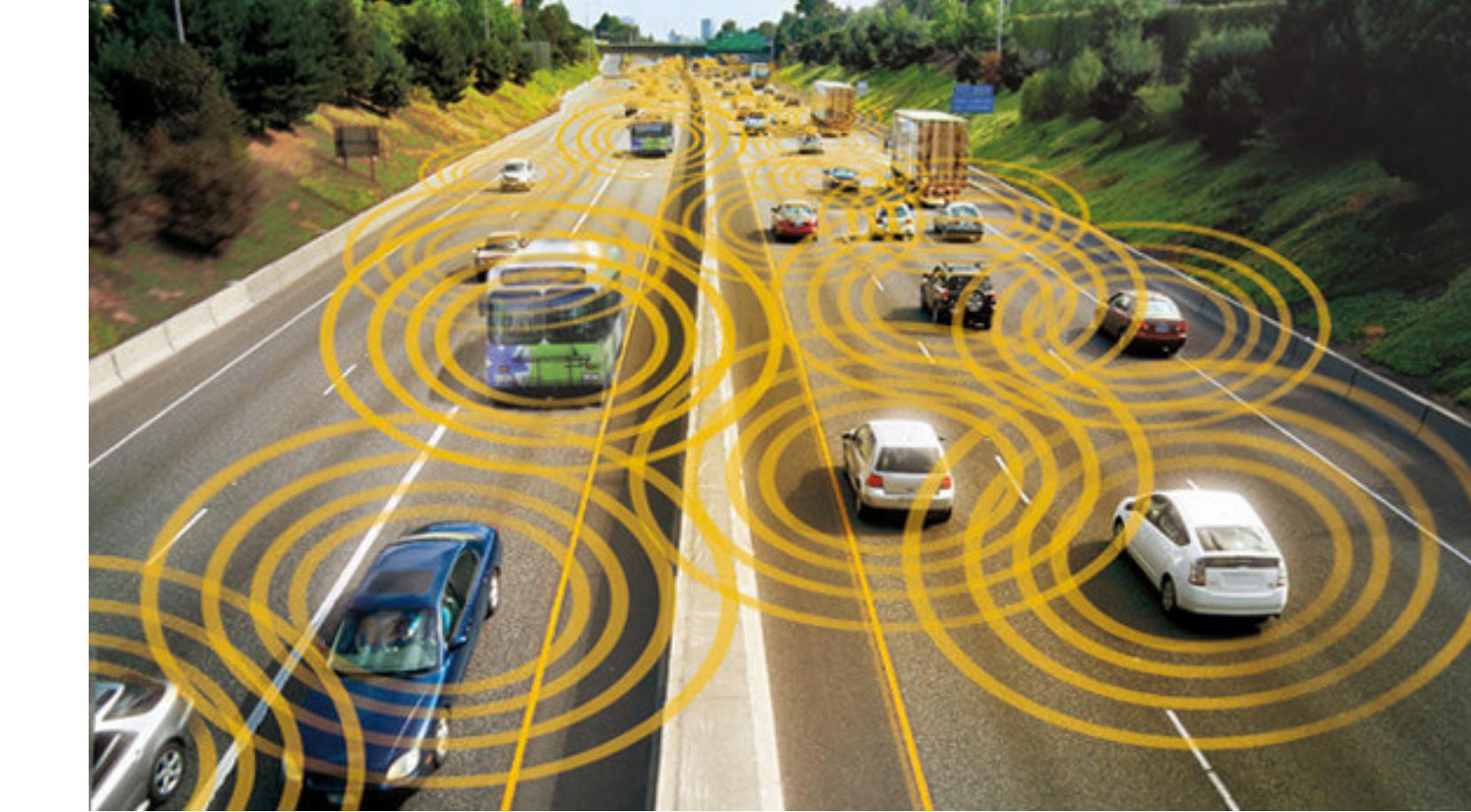
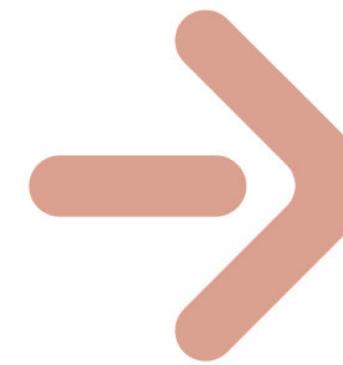
# Agenda

---

- Use Case
- Architecture
- Live Demo
- Best Practices
- Next steps



# Global Automotive Company Builds Connected Car Infrastructure



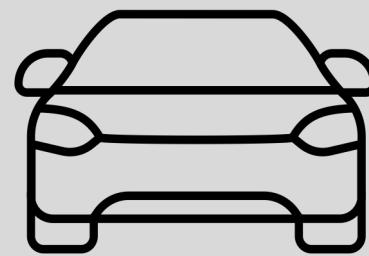
## Use Cases:

- **Connected Car Infrastructure (Cars, Traffic Lights, Cloud Services, etc.)**
- **Real Time Analytics (Predictive Maintenance, etc.)**
- **Continuous Services / Sales**
- **Partner Integration (Car workshop, gas station, food market, etc.)**
- ...

# Kafka Ecosystem

## Other Components

Car Sensor



HiveMQ  
MQTT Broker



(8a)  
Alert Car

MQTT  
Connector

Kafka Connect  
or  
Confluent Proxy  
or  
HiveMQ Plugin

(1)  
Ingest Data

KSQL



Tensor  
Flow

(04)  
Train Model

(5) Deploy  
Model  
----->

Real Time  
Kafka Streams  
Application  
(Java / Scala)

Tensor Flow

Real Time  
Edge  
Computing  
(C / librdkafka)

Tensor Flow  
Lite

TensorFlow  
Serving

gRPC

Real Time  
Kafka App

(6a) Consume  
Car Data  
(7) Potential  
Defect

Elastic  
Search



(6b)  
All Data

Grafana

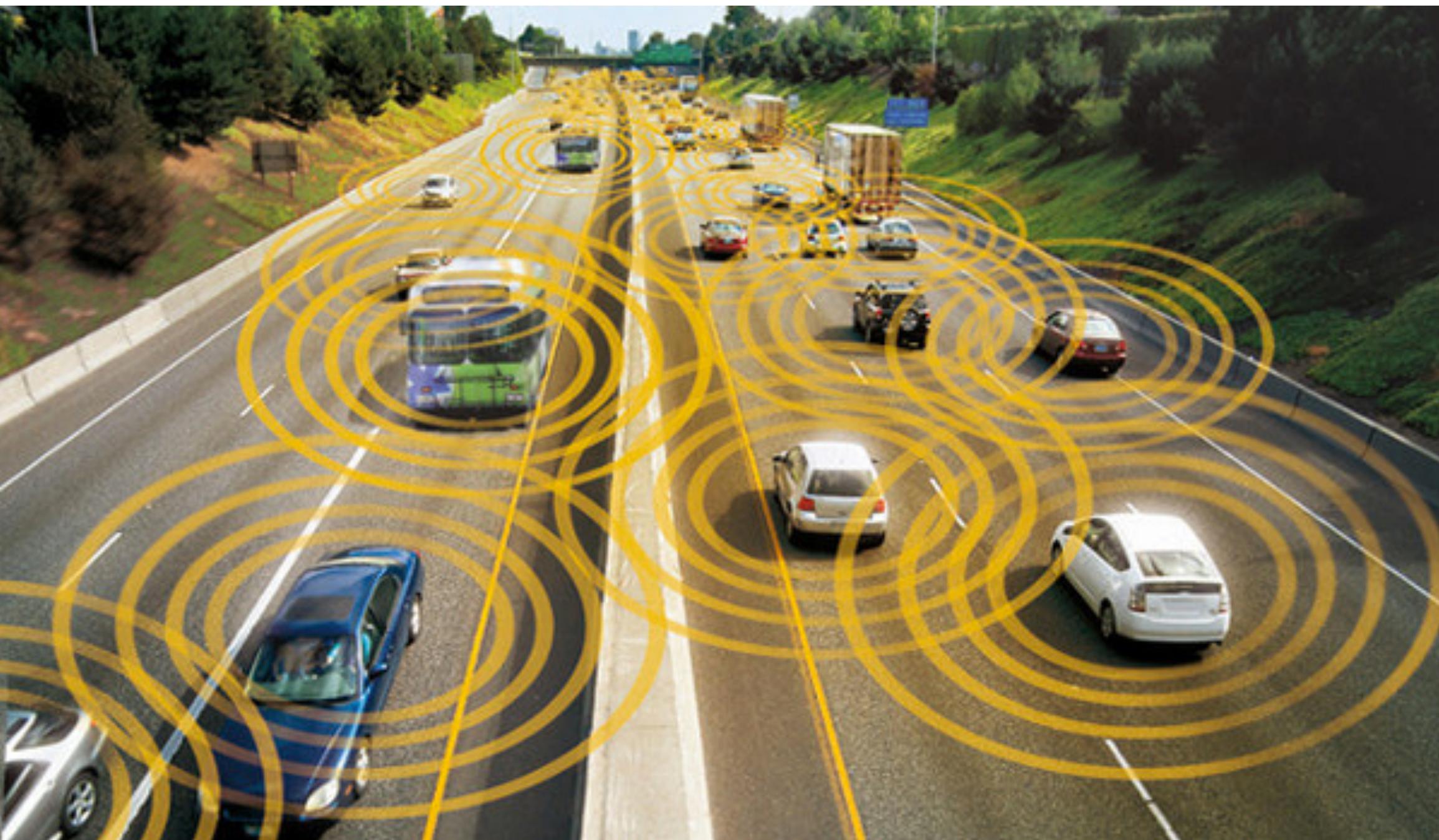


(8b)  
Alert Driver  
(e.g. mobile App)

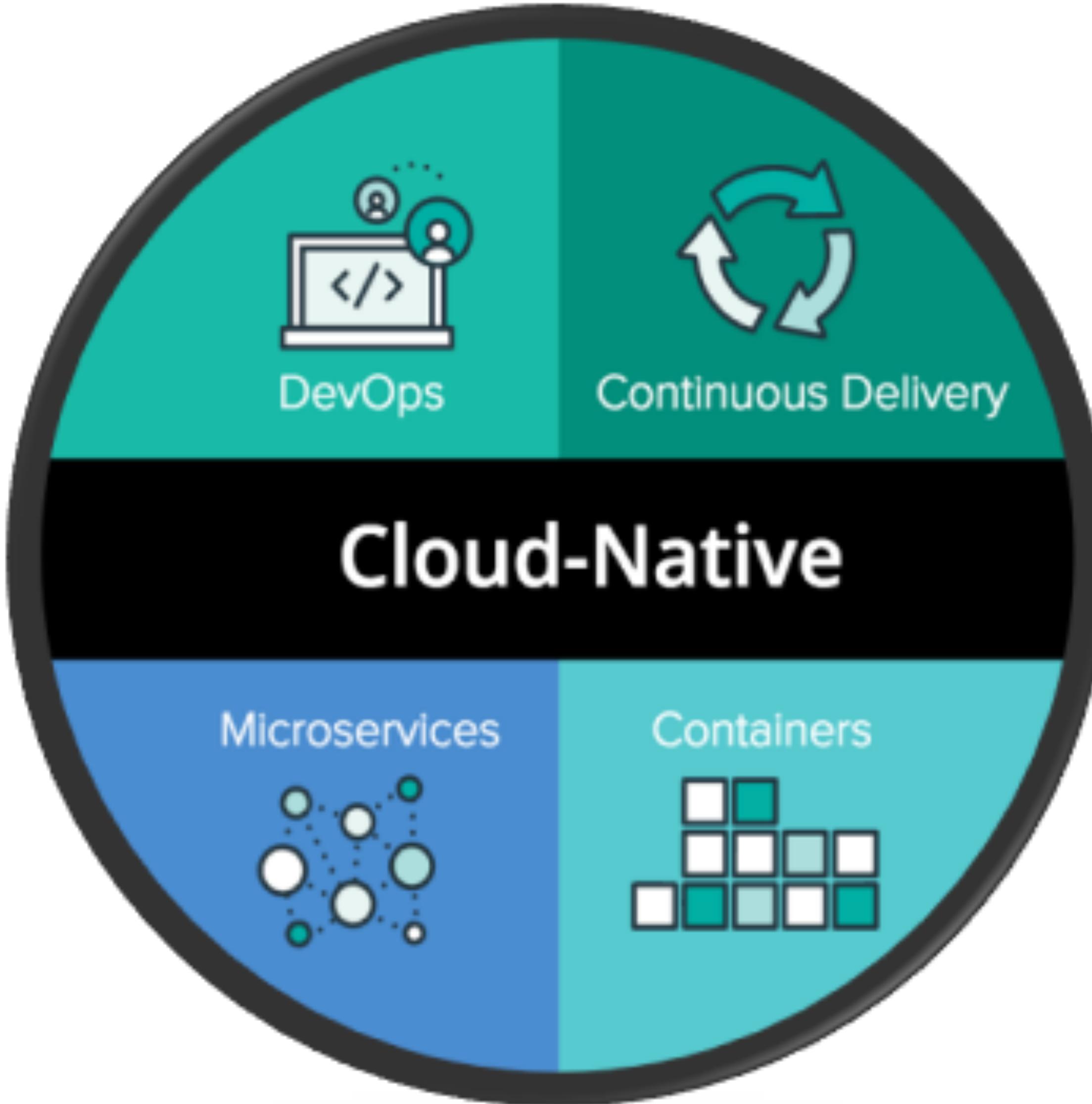
# Agenda

---

- Use Case
- **Architecture**
- Live Demo
- Best Practices
- Next steps



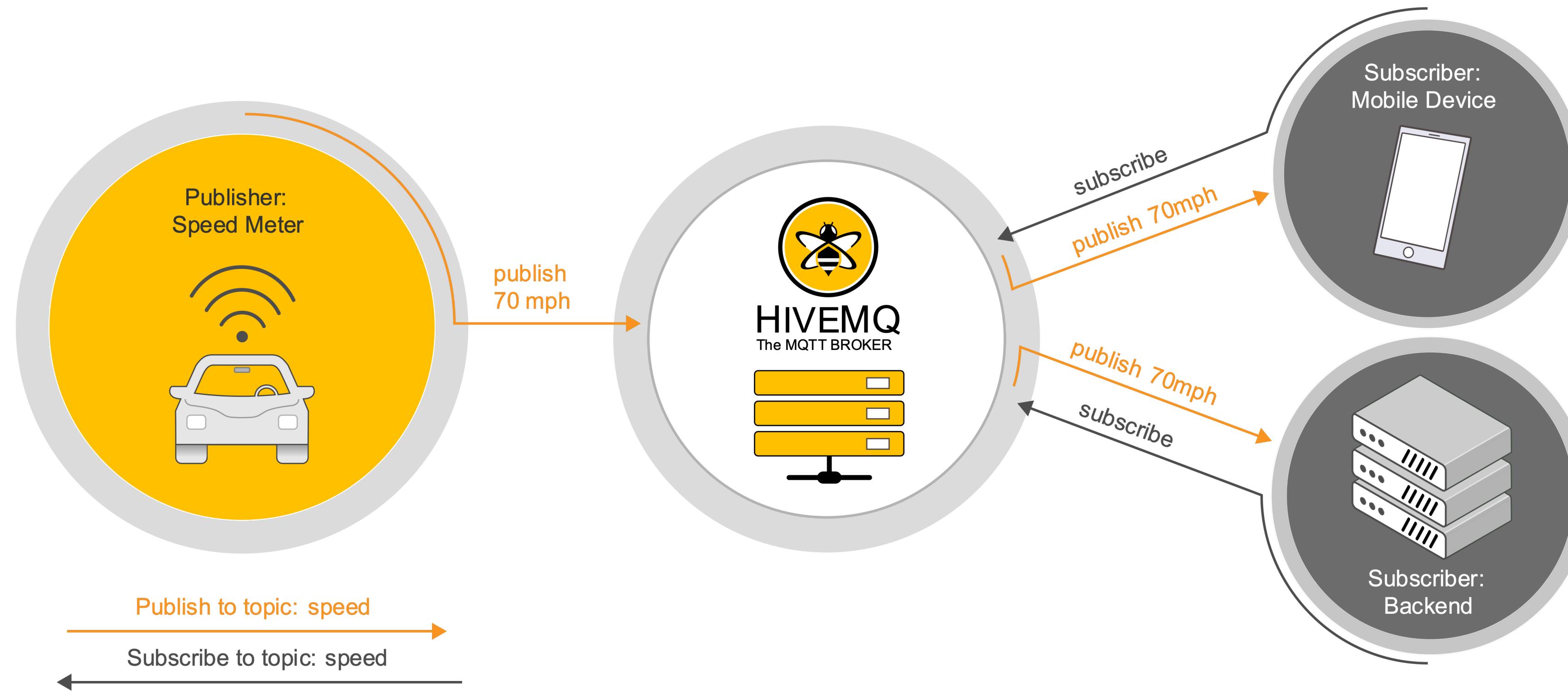
# Cloud Native Infrastructure



## Benefits

- Scalable
- Flexible
- Agile
- Elastic
- Automated
- Etc.

# MQTT - Publish / subscribe messaging protocol



- Built on top of TCP/IP for constrained devices and unreliable networks
- Many (open source) broker implementations
- Many (open source) client libraries
- IoT-specific features for bad network / connectivity
- Widely used (mostly IoT, but also web and mobile apps via MQTT over WebSocket)

# MQTT Trade-Offs

## Pros

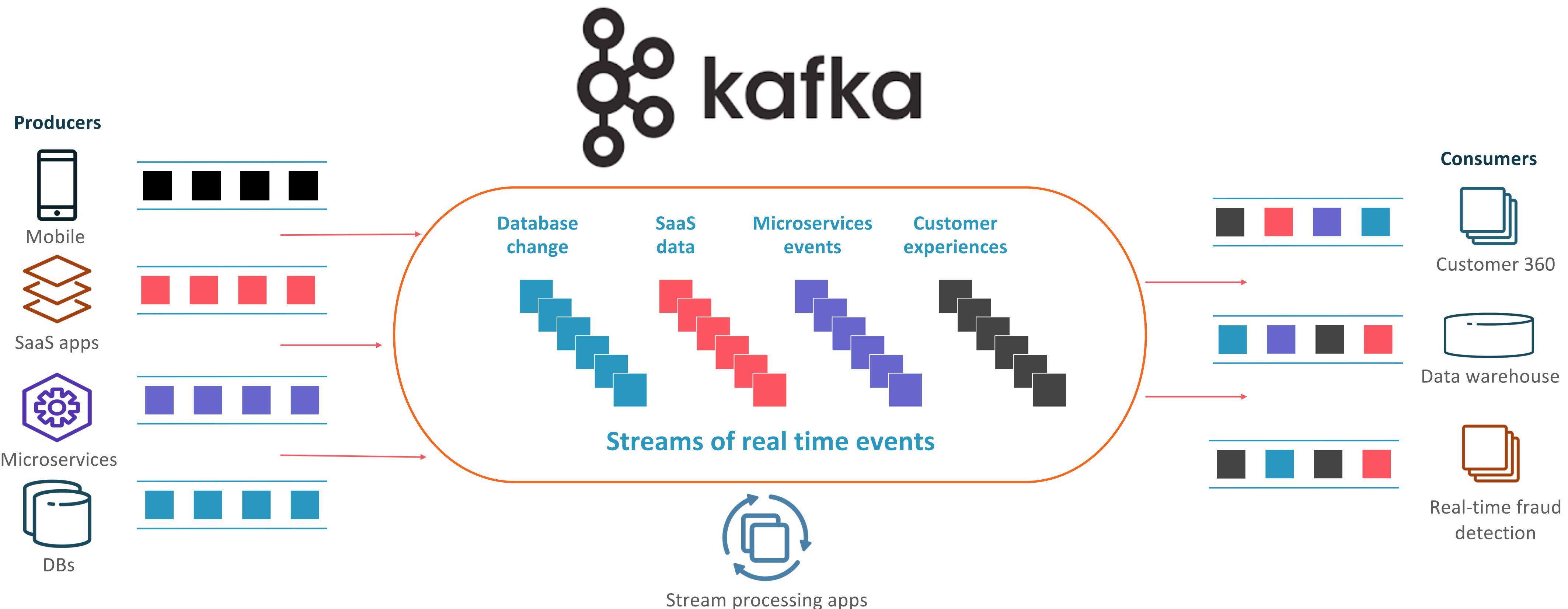
- Lightweight
- All programming languages supported
- Built for poor connectivity / high latency scenarios (e.g. mobile networks!)
- High scalability and availability \*
- ISO Standard
- Most popular IoT protocol

## Cons

- Only pub/sub, not stream processing
- Asynchronous processing (clients can be offline for long time)
- No reprocessing of events



# A Streaming Platform is the Underpinning of an Event-driven Architecture



## Ubiquitous connectivity

Globally scalable platform for all event producers and consumers

## Immediate data access

Data accessible to all consumers in real time

## Single system of record

Persistent storage to enable reprocessing of past events

## Continuous queries

Stream processing capabilities for in-line data transformation

# Kafka Trade-Offs (from IoT perspective)

---

## Pros

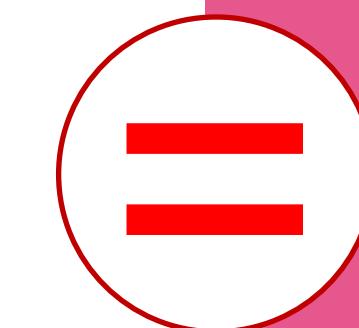
- **Stream processing**, not just pub/sub
- High throughput
- Large scale
- High availability
- Long term storage and buffering
- Reprocessing of events
- **Good integration to rest of the enterprise**

## Cons

- Not built for tens of thousands connections
- Requires stable network and good infrastructure
- No IoT-specific features like keep alive, last will or testament



# (De facto) Standards for Processing IoT Data



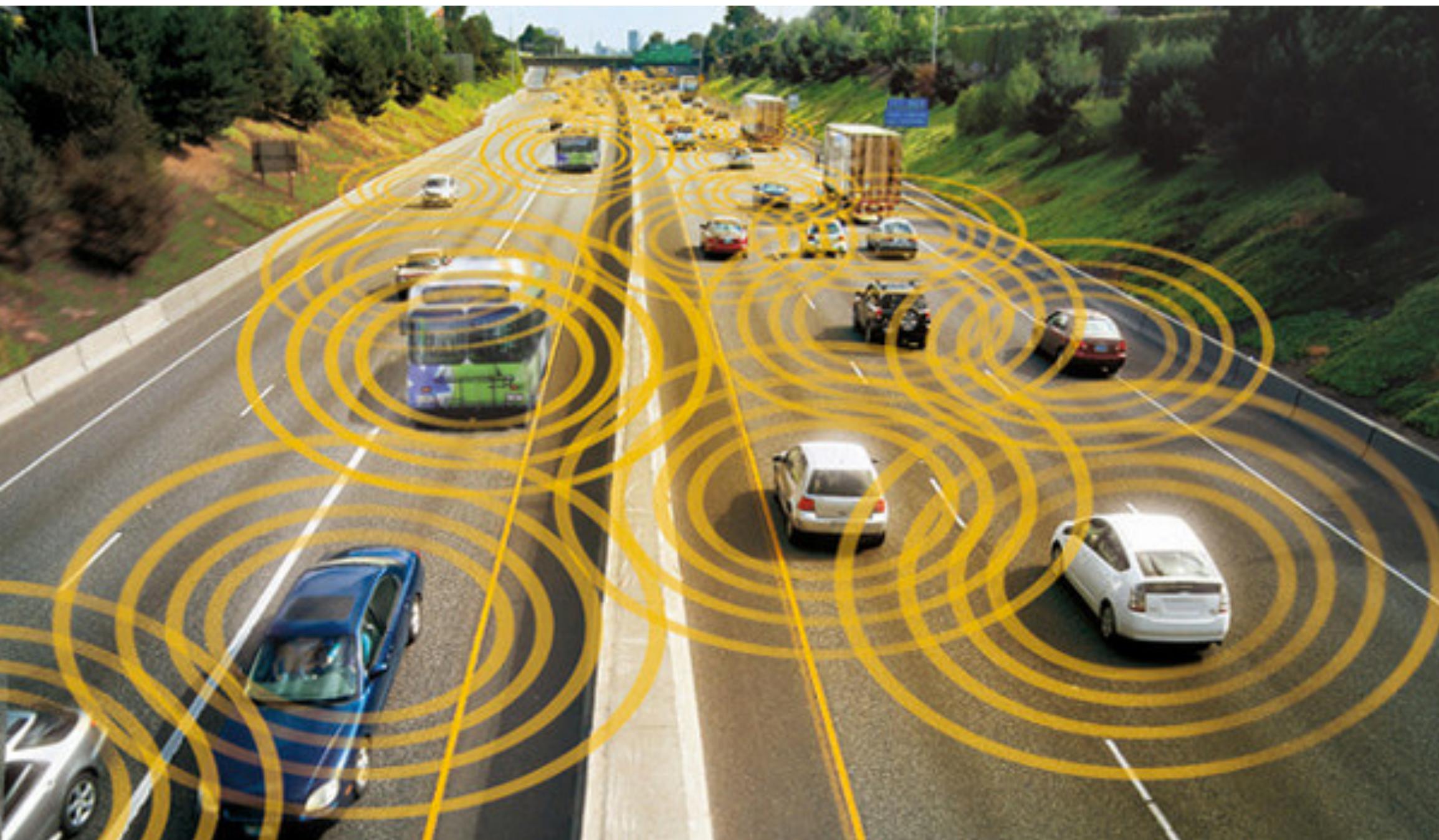
A Match Made In Heaven

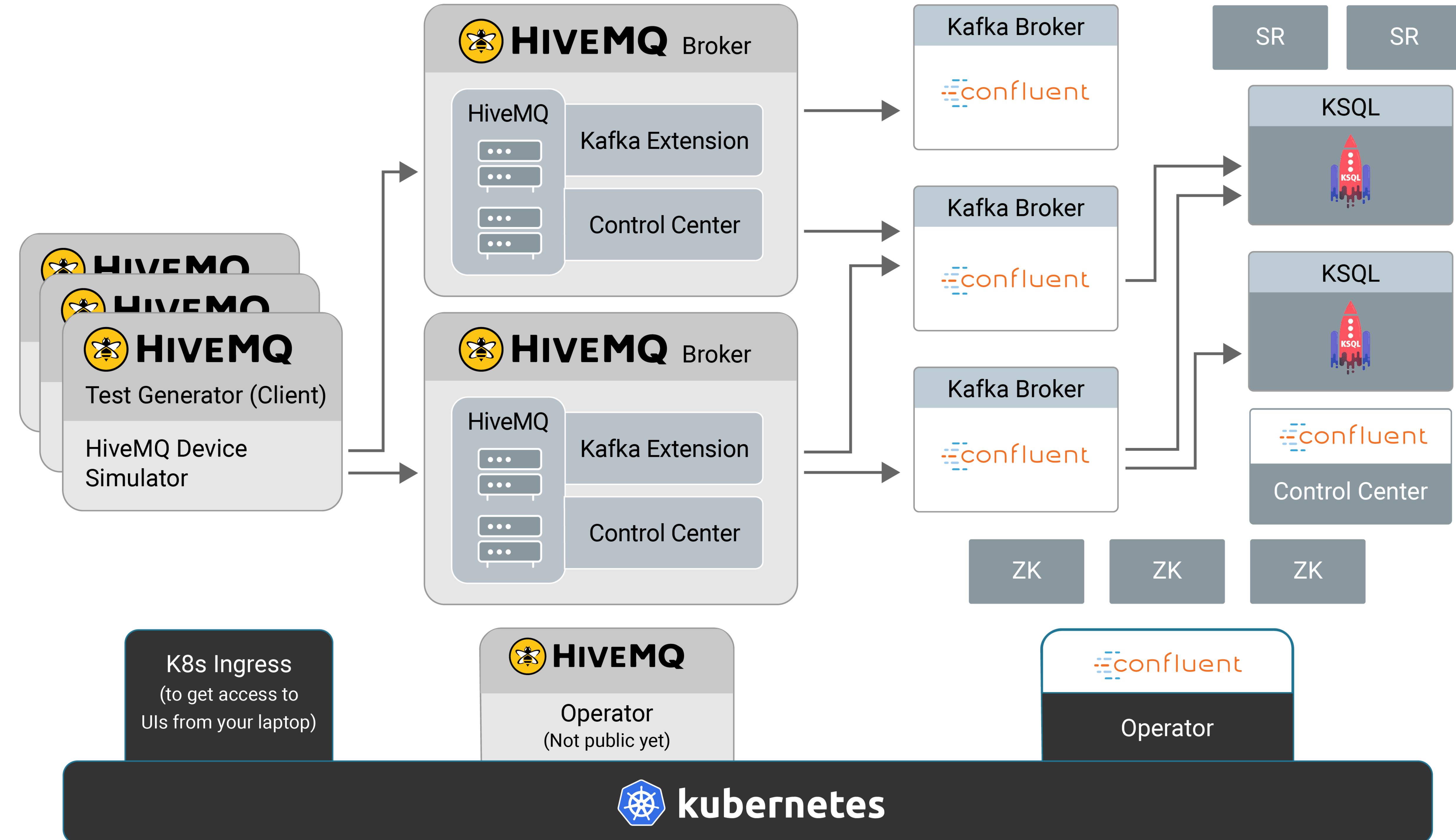


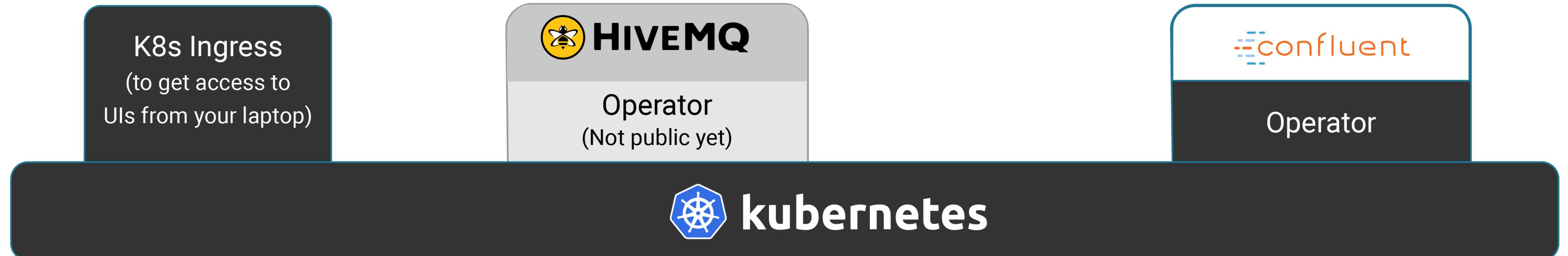
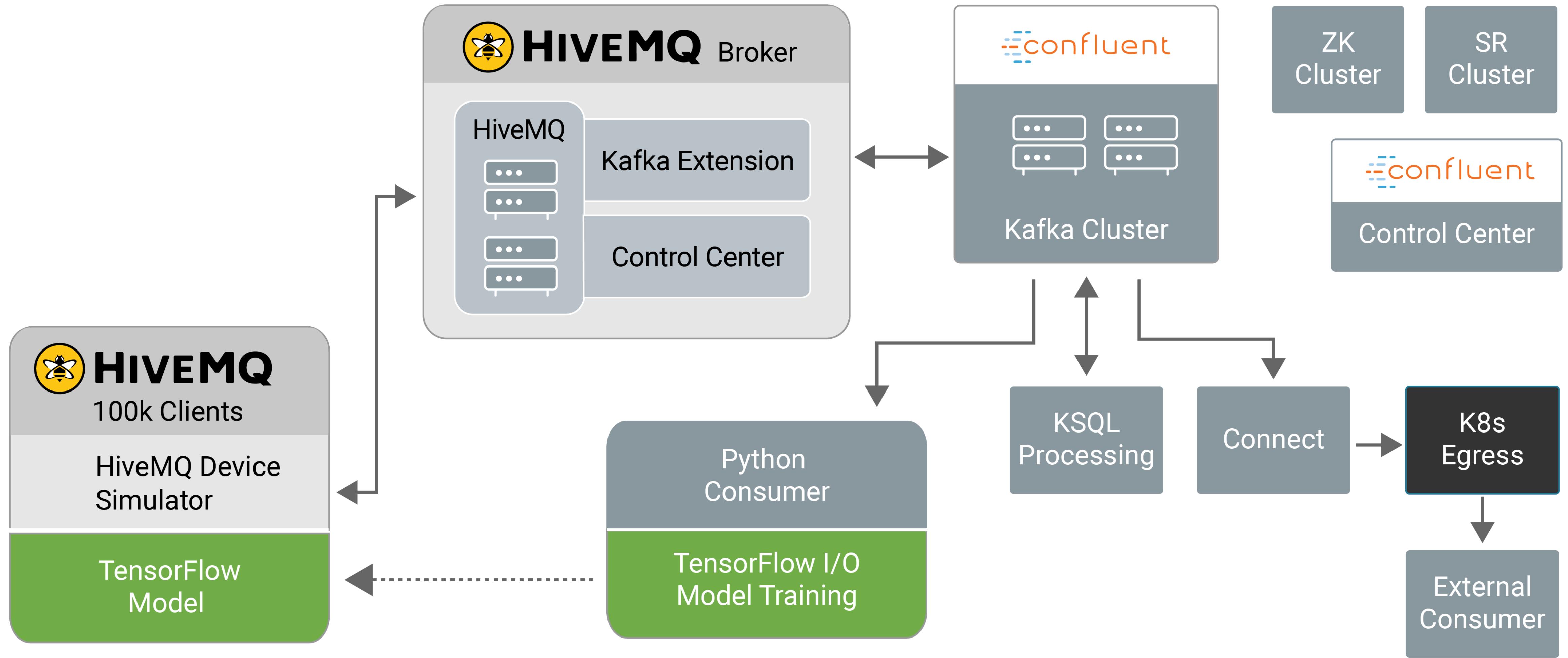
# Agenda

---

- Use Case
- Architecture
- Live Demo
- Best Practices
- Next steps







## Live Demo

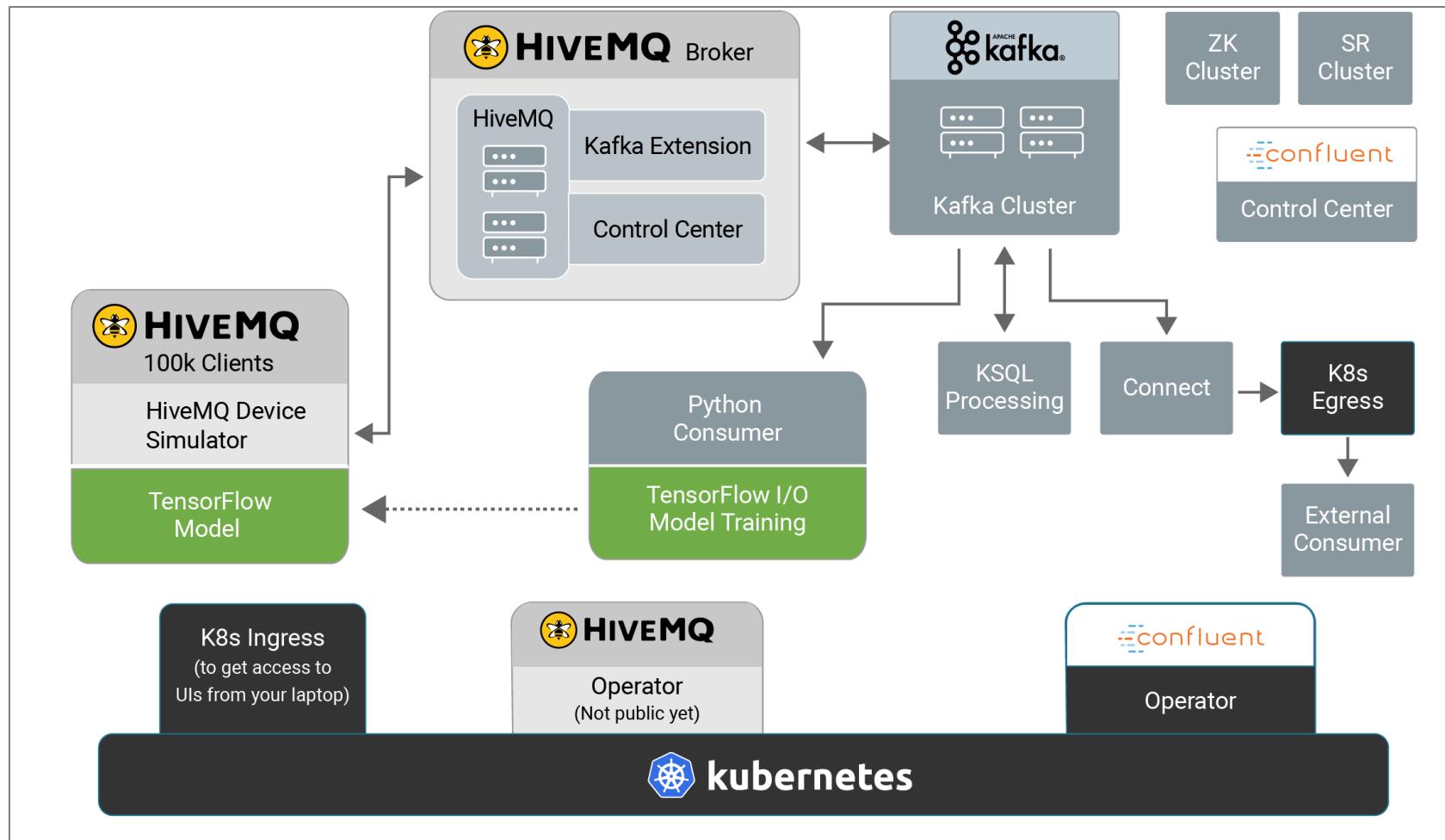


**End-to-End Integration and Data Processing for  
100000 Connected Cars**



# Demo 100.000 Connected Cars

(Kafka + MQTT + TensorFlow)



Kai Waehner's GitHub profile page. He is a Technology Evangelist at Confluent. His bio states: "Kai Währner works as Technology Evangelist at Confluent. He is regular speaker at international IT conferences. References: www.kai-waehner.de". He is associated with Confluent and located in Munich, Germany. His email is megachucky@gmail.com. His GitHub stats show 62 repositories, 0 projects, 0 packages, 60 stars, and 22 followers. His pinned projects include:

- kafka-streams-machine-learning-examples**: This project contains examples which demonstrate how to deploy analytic models to mission-critical, scalable production environments leveraging Apache Kafka and its Streams API. Models are built with...
- tensorflow-serving**: Kafka Streams + Java + gRPC combined with RPC / Request...
- ksql-udf-deep-learning-mqtt-iot**: Deep Learning UDF for KSQL for Streaming Anomaly Detection of MQTT IoT Sensor Data.
- kafka-connect-iot**: Internet of Things Integration + MQTT Connector + Sensors...
- python-jupyter-apache-kafka-ksql-tensorflow-keras**: Making Machine Learning Simple and Scalable with Python, Jupyter Notebook, TensorFlow, Keras, Apache Kafka and KSQL.
- hivemq-mqtt-tensorflow-learning-training-inference**: Real Time Big Data / IoT Machine Learning with HiveMQ (MQTT), TensorFlow, Keras, and a data store like S3, HDFS or...



<https://github.com/kaiwaehner/hivemq-mqtt-tensorflow-kafka-realtime-iot-machine-learning-training-inference>

or

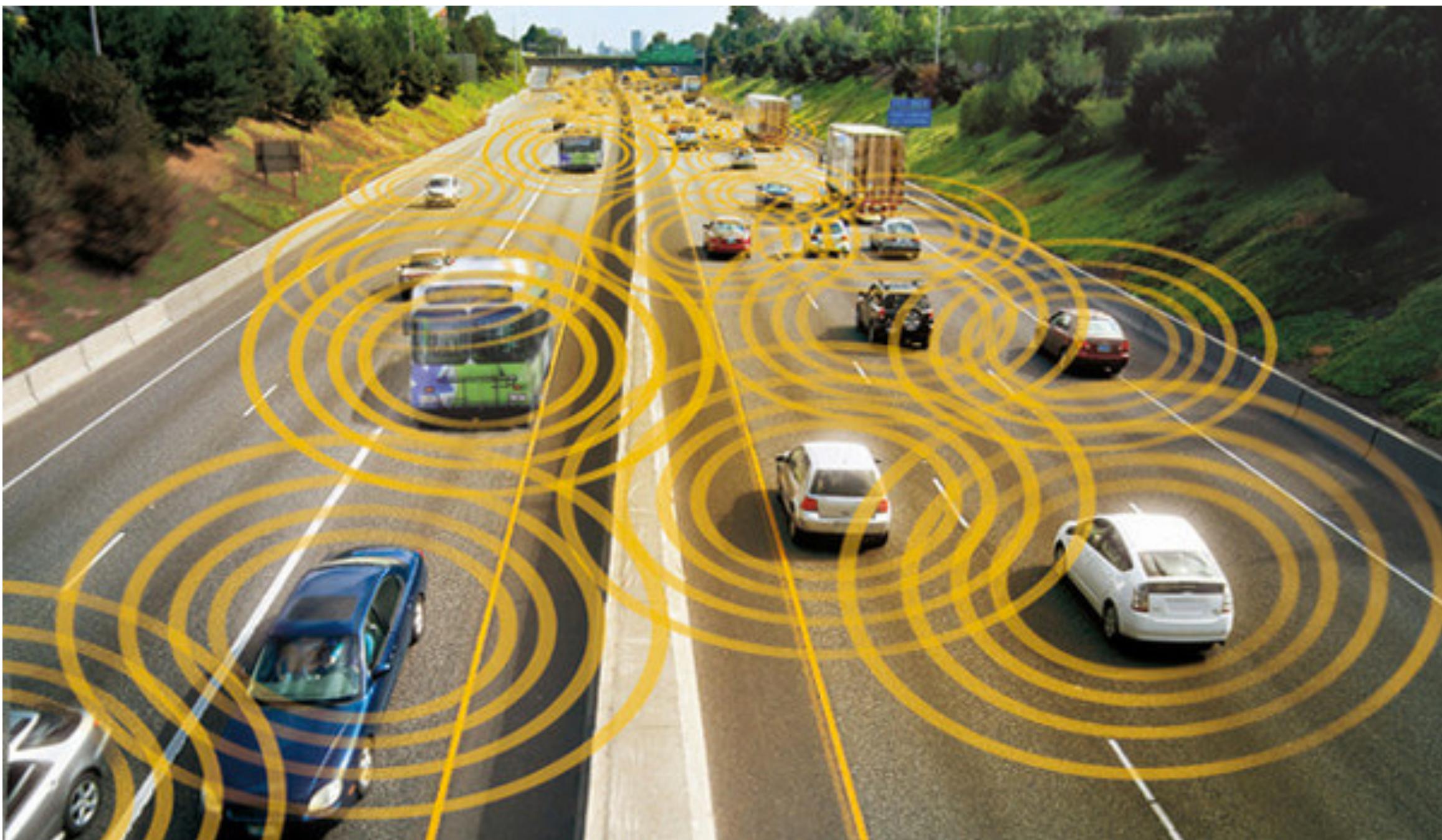
<http://bit.ly/kafka-mqtt-ml-demo>

=> Try it out in 30 minutes!

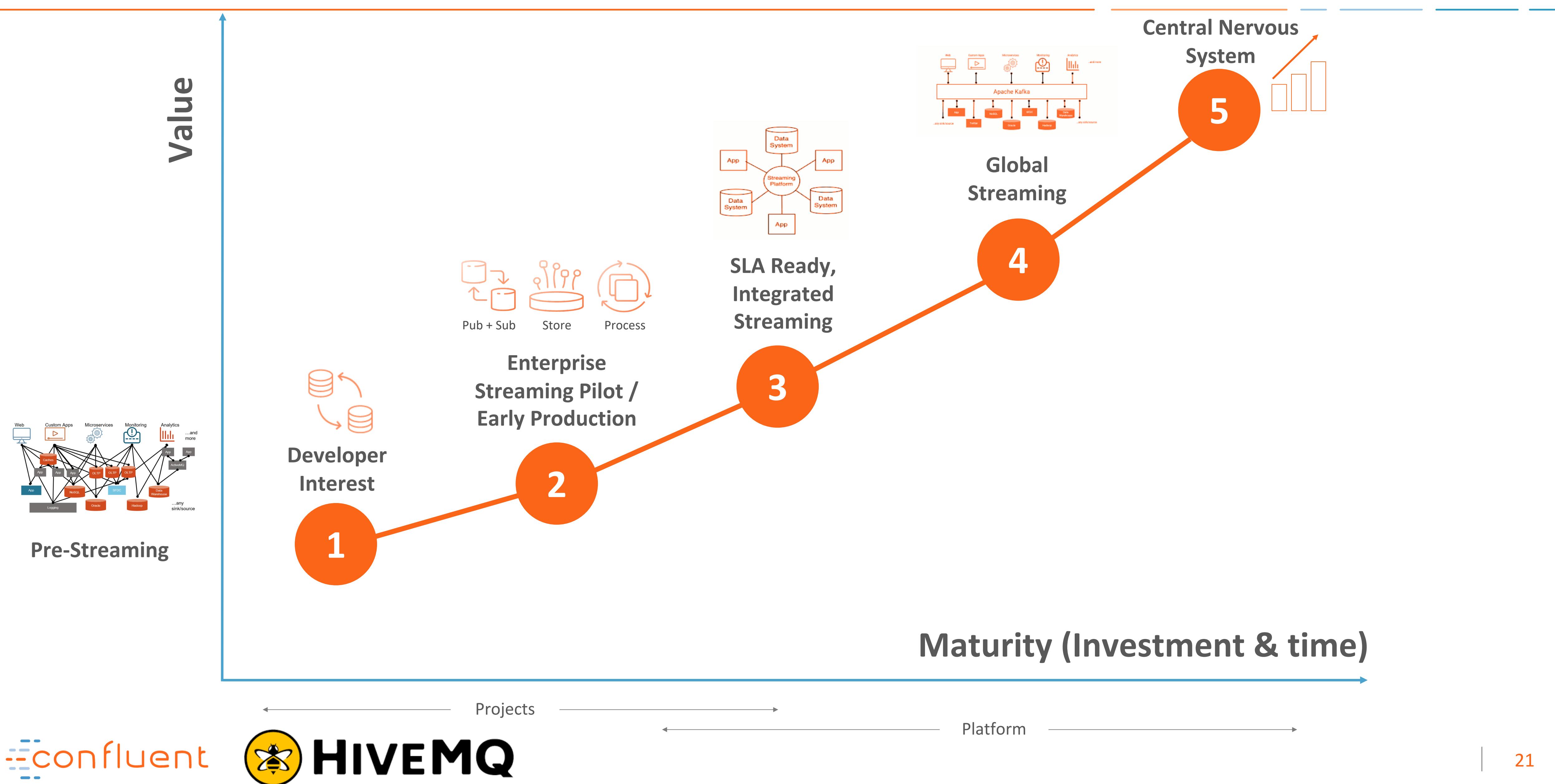
# Agenda

---

- Use Case
- Architecture
- Live Demo
- **Best Practices**
- Next steps



# Typical Journey



# Start Small, but prepare for Scalability from Beginning

---

## 1. Use cloud native and scalable components

- Confluent Platform is cloud native and built for scale
- HiveMQ is cloud native and built for scale

## 2. Don't deep dive too much in the beginning – but understand options

- HiveMQ Kafka Extension?
- Confluent MQTT connectors?
- Customer Integration?

## 3. Plan for Enterprise-readiness

- Security
- Monitoring
- Operations tooling
- Bi-directional communication



# Choose the right tool stack and infrastructure

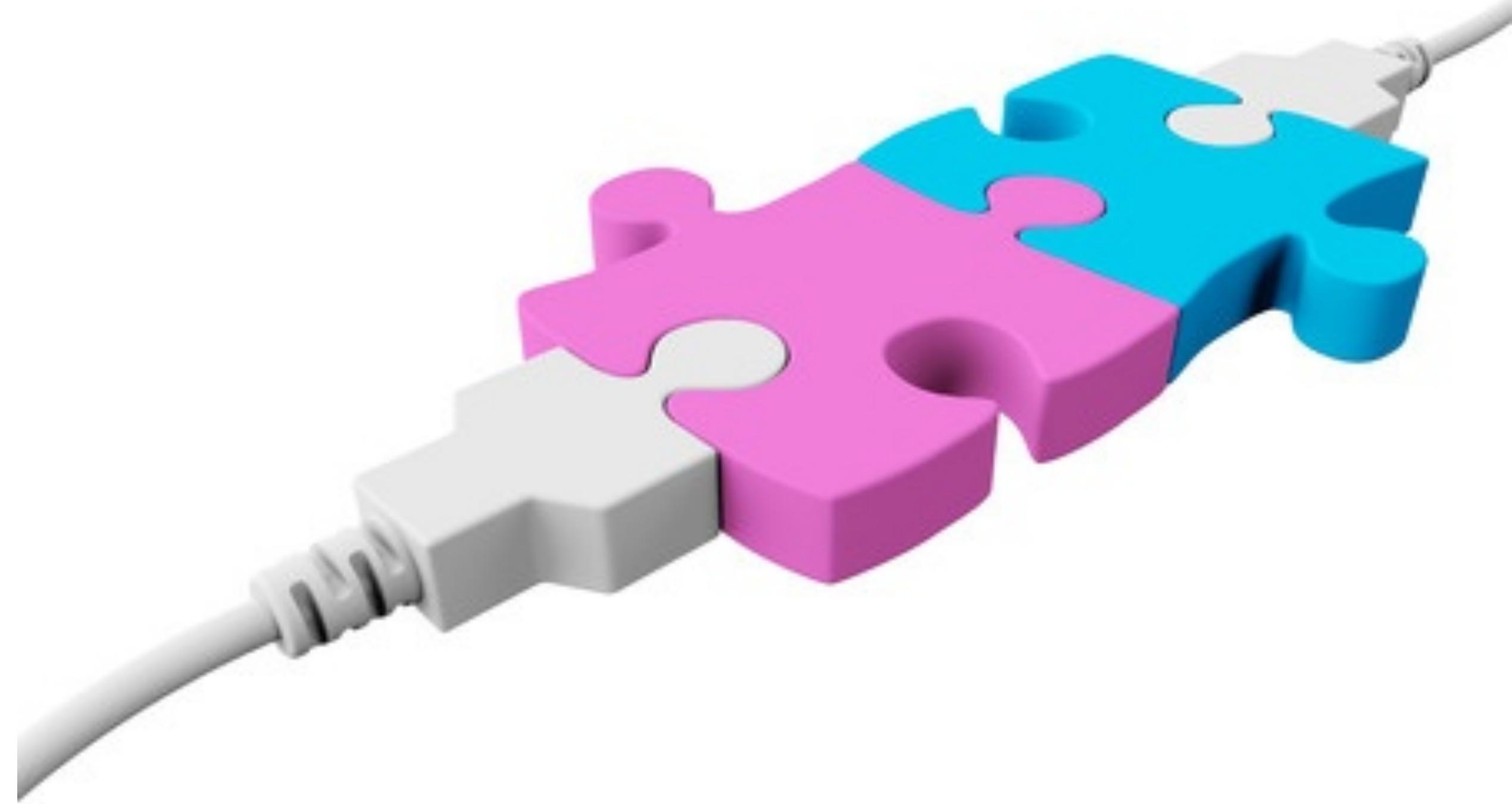
---

## Understand Trade-Offs and choose the right options for deployments

- Edge
- On Premise
- Cloud

## Use the best tools for the job

- Confluent Platform for Event Streaming
- HiveMQ for MQTT messaging and connectivity



# Separation of concerns

1. Devices
2. Gateway
3. Integration
4. Data Streaming
5. Consumer Apps

## Decouple tasks

- Source integration
- Data processing
- Business logic
- Sink integration
- Analytics
- ...



# Different data for different use cases

---

- Database, Data Lake
- Search
- Real time, Near Real Time, Batch
- Streaming, Request-Response
- CQRS, Event Sourcing
- Machine Learning

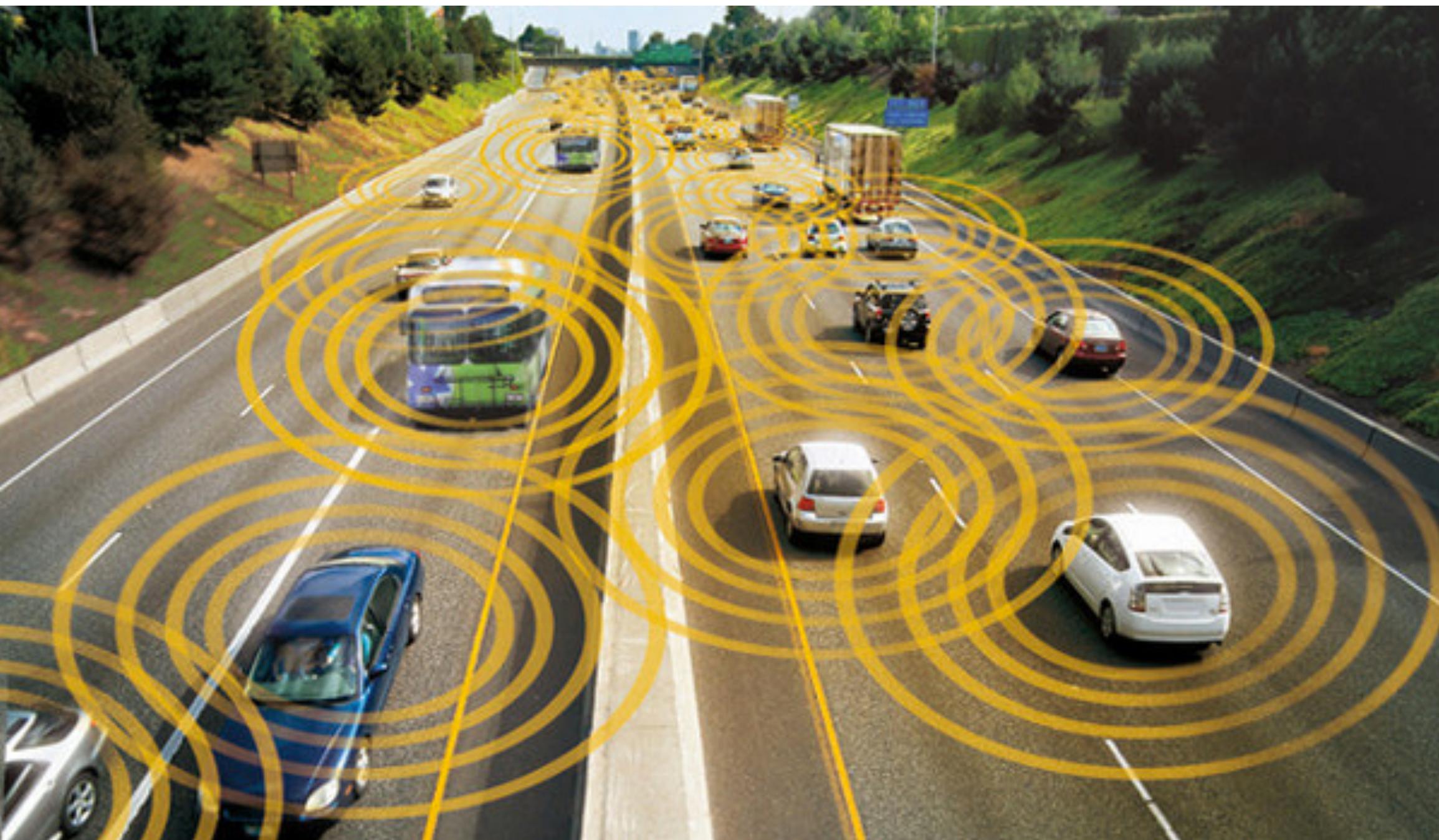
There is no single  
**MASTER DATA EVENT...**



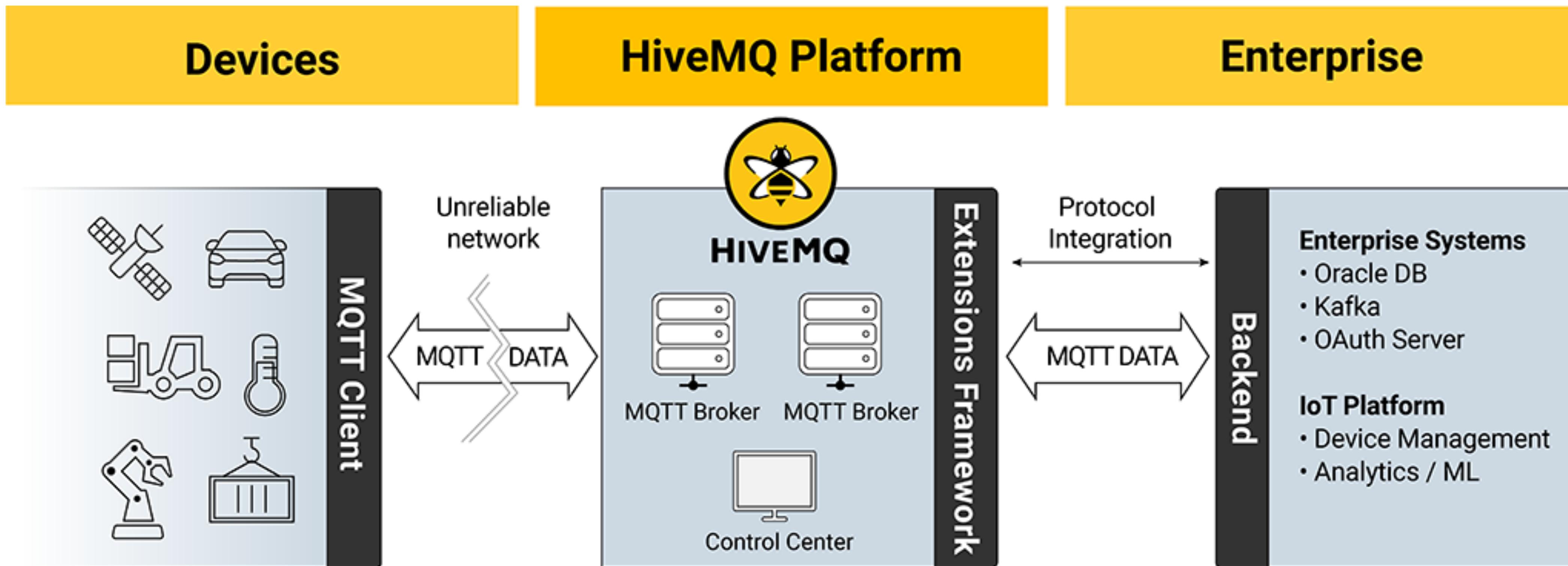
# Agenda

---

- Use Case
- Architecture
- Live Demo
- Best Practices
- **Next steps**



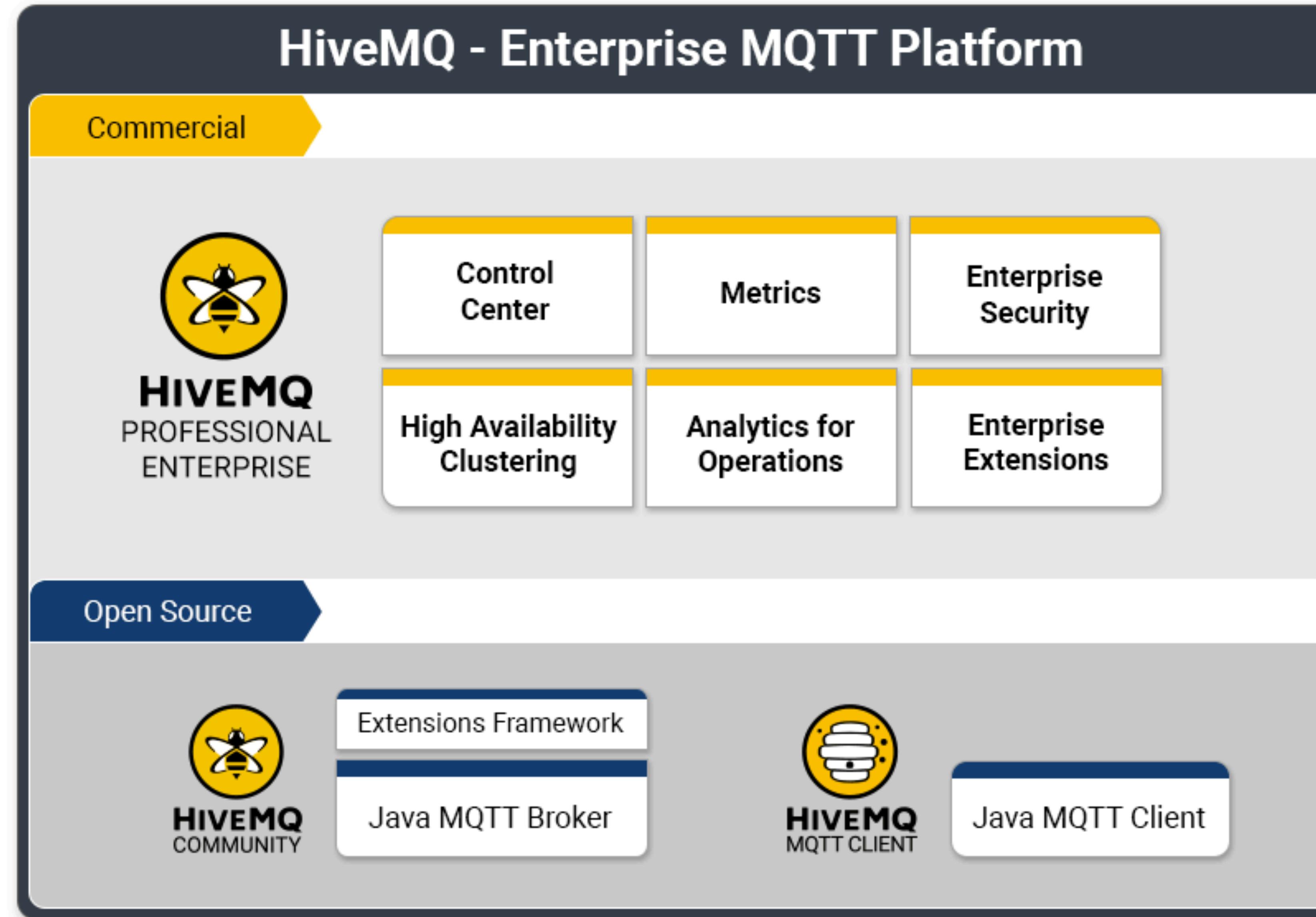
# The HiveMQ Platform

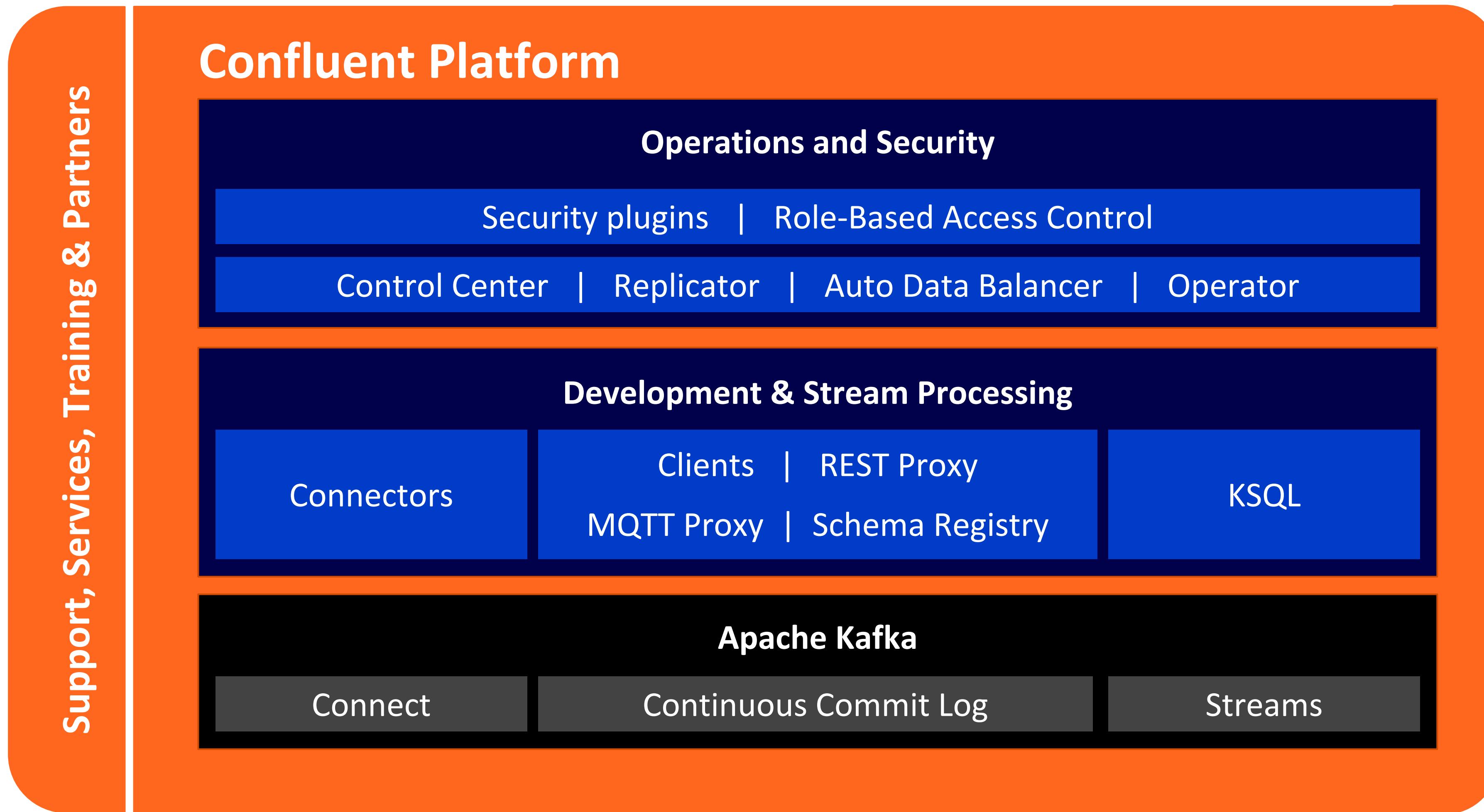


Kubernetes, Docker, OpenShift

Public or private cloud (AWS, MS Azure...) or on-premise

# The HiveMQ Platform – Open Source and Enterprise-grade





**Mission-critical  
Reliability**

**Complete Event  
Streaming Platform**

Self-Managed Software

Datacenter

Public Cloud

Fully-Managed Service

Confluent Cloud

**Freedom of Choice**

Spend your time on your applications!

---



## Confluent Cloud

Cloud-Native Confluent Platform **Fully-Managed Service**

Available on the leading public clouds with **mission-critical SLAs** and **consumption-based pricing**.



Google Cloud Platform



**Serverless Kafka** characteristics:

Pay-as-you-go, elastic auto-scaling, abstracting infrastructure (topics not brokers)

## Next steps...

---

Try out the demo in 30 minutes:

<https://github.com/kaiwaehner/hivemq-mqtt-tensorflow-kafka-realtime-iot-machine-learning-training-inference>

<http://bit.ly/kafka-mqtt-ml-demo>

Check out the documentation and blog posts

- HiveMQ and Apache Kafka - Streaming IoT Data and MQTT Messages:  
<https://www.hivemq.com/blog/streaming-iot-data-and-mqtt-messages-to-apache-kafka/>
- Internet of Things (IoT) and Event Streaming at Scale with Apache Kafka and MQTT:  
<https://www.confluent.io/blog/iot-with-kafka-connect-mqtt-and-rest-proxy>

Contact us for questions or any other feedback:

- Website, Email, Slack, Phone, ...
- Dominik: [dominik@hivemq.com](mailto:dominik@hivemq.com) , Kai: [kai.waehner@confluent.io](mailto:kai.waehner@confluent.io)



# Questions? Feedback?

Please contact us!



**Kai Waehner**  
Technology Evangelist

[kai.waehner@confluent.io](mailto:kai.waehner@confluent.io)  
[LinkedIn](#)  
[@KaiWaehner](#)  
[www.confluent.io](http://www.confluent.io)  
[www.kai-waehner.de](http://www.kai-waehner.de)



**Dominik Obermaier**  
CTO HiveMQ

[dominik.obermaier@hivemq.com](mailto:dominik.obermaier@hivemq.com)  
[www.linkedin.com/in/dobermai](http://www.linkedin.com/in/dobermai)  
[www.hivemq.com](http://www.hivemq.com)  
[www.twitter.com/dobermai](http://www.twitter.com/dobermai)