



Dean Wampler, Ph.D.  
[dean@lightbend.com](mailto:dean@lightbend.com)  
[polyglotprogramming.com/talks](http://polyglotprogramming.com/talks)

# Trends Affecting System Architectures: Mobile Engagement, Streaming Data, and Machine Learning

# Data Streaming Architectures

[lbnd.io/fast-data-ebook](https://lbnd.io/fast-data-ebook)

O'REILLY®

Compliments of  
**Lightbend**

## Fast Data Architectures for Streaming Applications

Getting Answers Now from Data Sets That Never End

2nd Edition



Dean Wampler, PhD

## Streaming Engines



## Microservices



## Machine Learning



**Intelligent Management & Monitoring and Security**



## Data Backplane



Storage Options

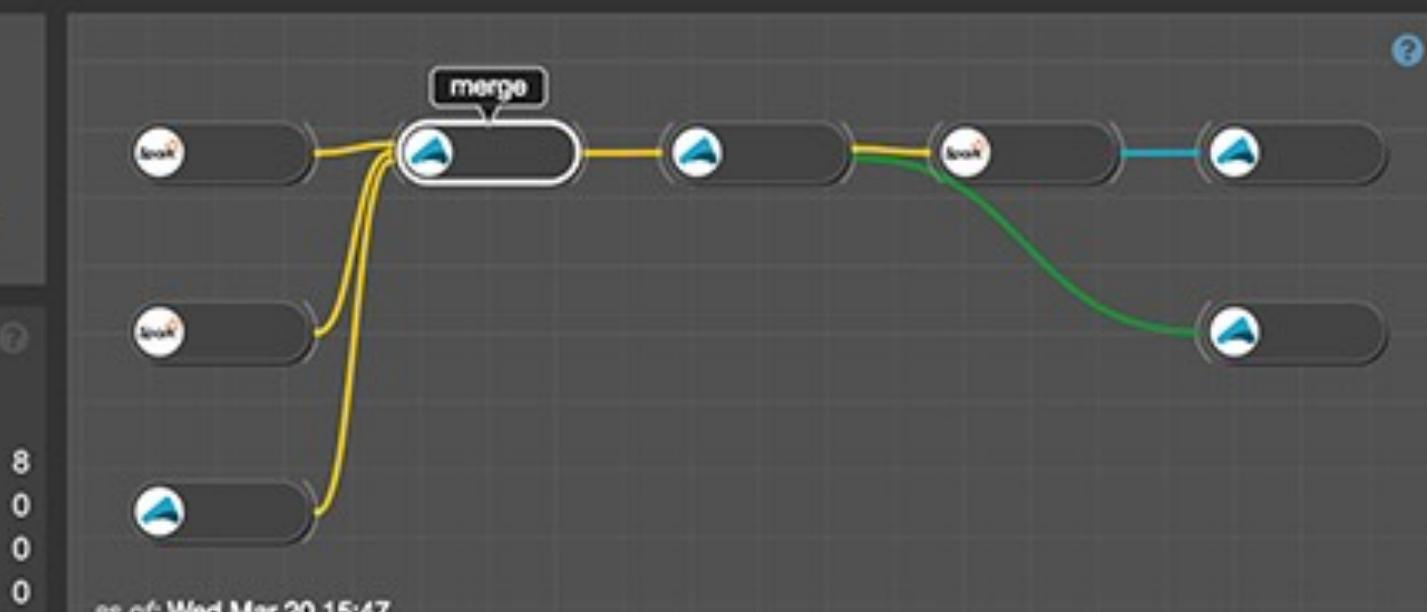
HDFS

SQL, NoSQL

Cloud Storage (S3 etc)

Search





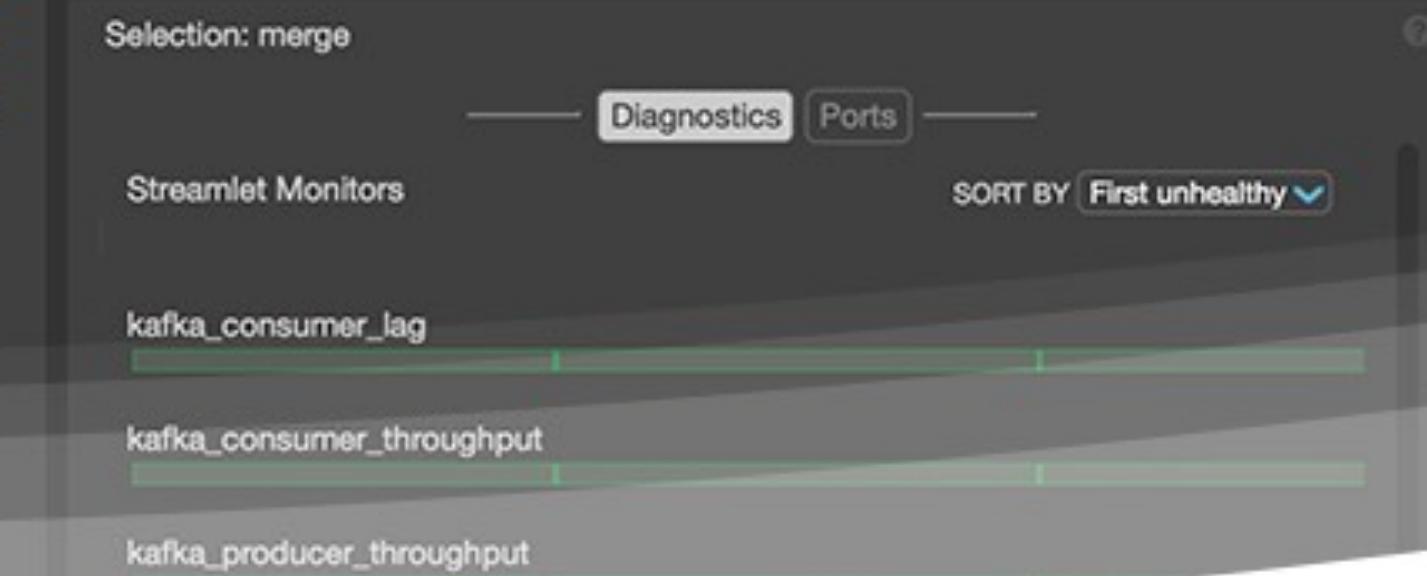
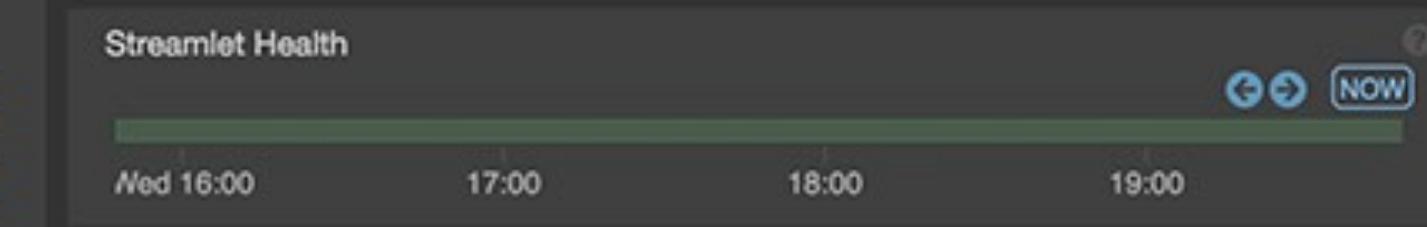
## Application Details

## Streamlet Current Health

Healthy	8
Warning	0
Critical	0
Unknown	0

## Streamlet Health Events

cdr-validator	---
cdr-aggregator	---
merge	---
console-egress	---
error-egress	---
cdr-generator1	---
cdr-generator2	---
cdr-ingress	---

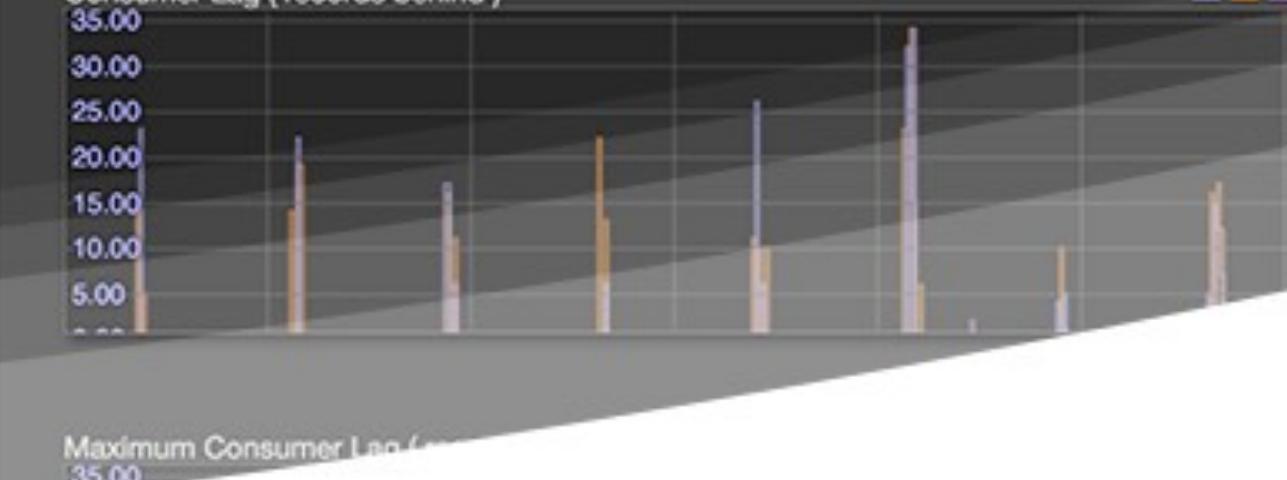


## Metrics

## Throughput (records / second)



## Consumer Lag ( records behind )



What we're up to at Lightbend...  
[lightbend.com/lightbend-pipelines-demo](http://lightbend.com/lightbend-pipelines-demo)





# What We'll Discuss

- Forces driving us towards stream processing
- Streaming architectures
  - Requirements
  - An example
- Mixing data science and data engineering
- Other production concerns

What We'll Discuss

# Forces driving us towards stream processing

A black and white photograph of a vast wind farm in a desert landscape under a dramatic, cloudy sky. The foreground is filled with the silhouettes of numerous wind turbines standing in long, curved rows. In the middle ground, rolling hills or mountains are visible against a sky filled with heavy, layered clouds. The lighting creates strong highlights and shadows on the clouds, suggesting a bright sun obscured by the weather.

Telecom



Finance

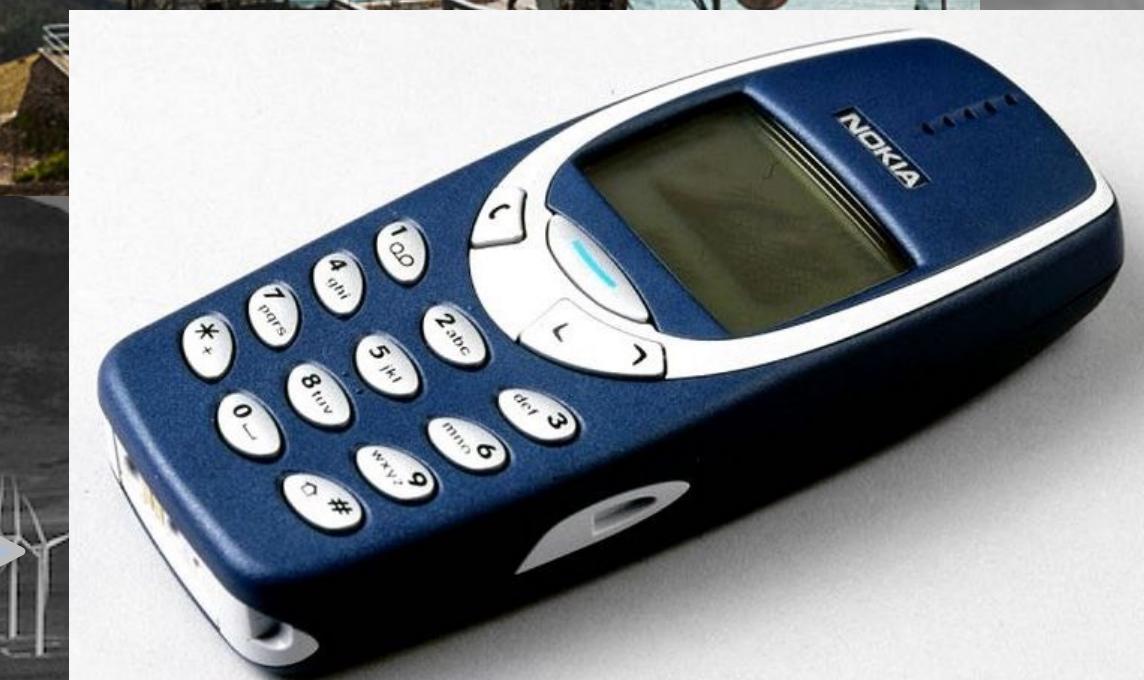


Energy



... and IoT

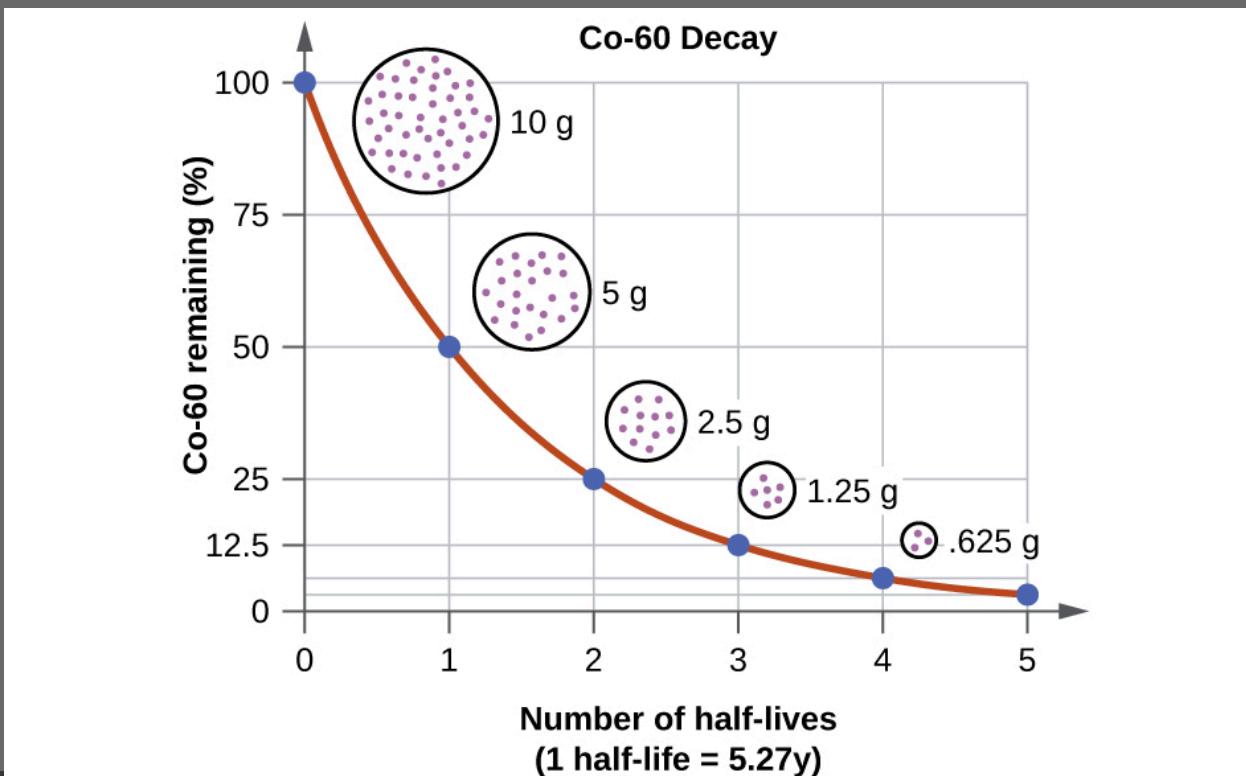
State of the art phone!

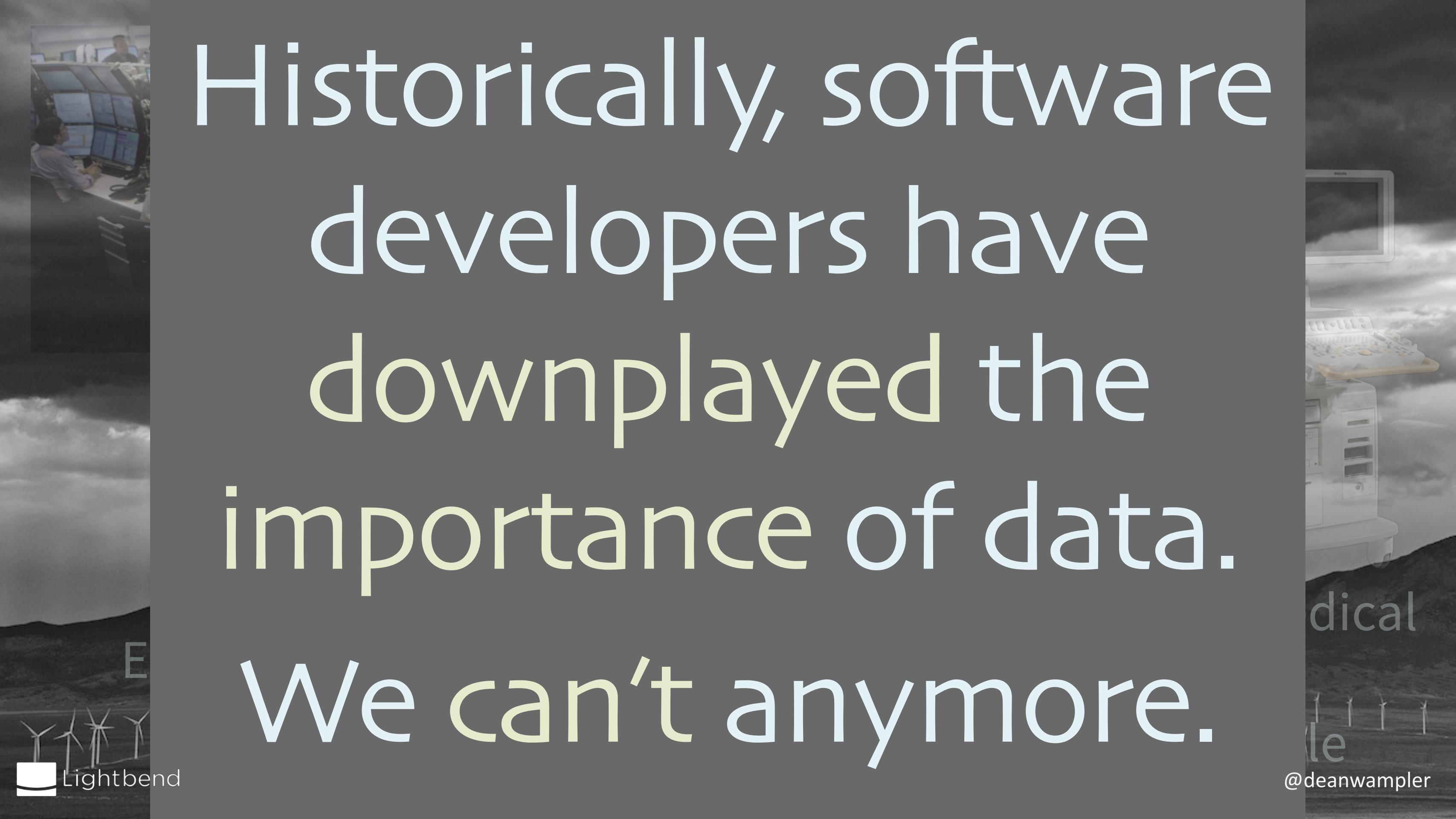


Medical  
Mobile



# Information value has a half life; it decays with time





Historically, software  
developers have  
downplayed the  
importance of data.  
We can't anymore.



Lightbend

@deanwampler

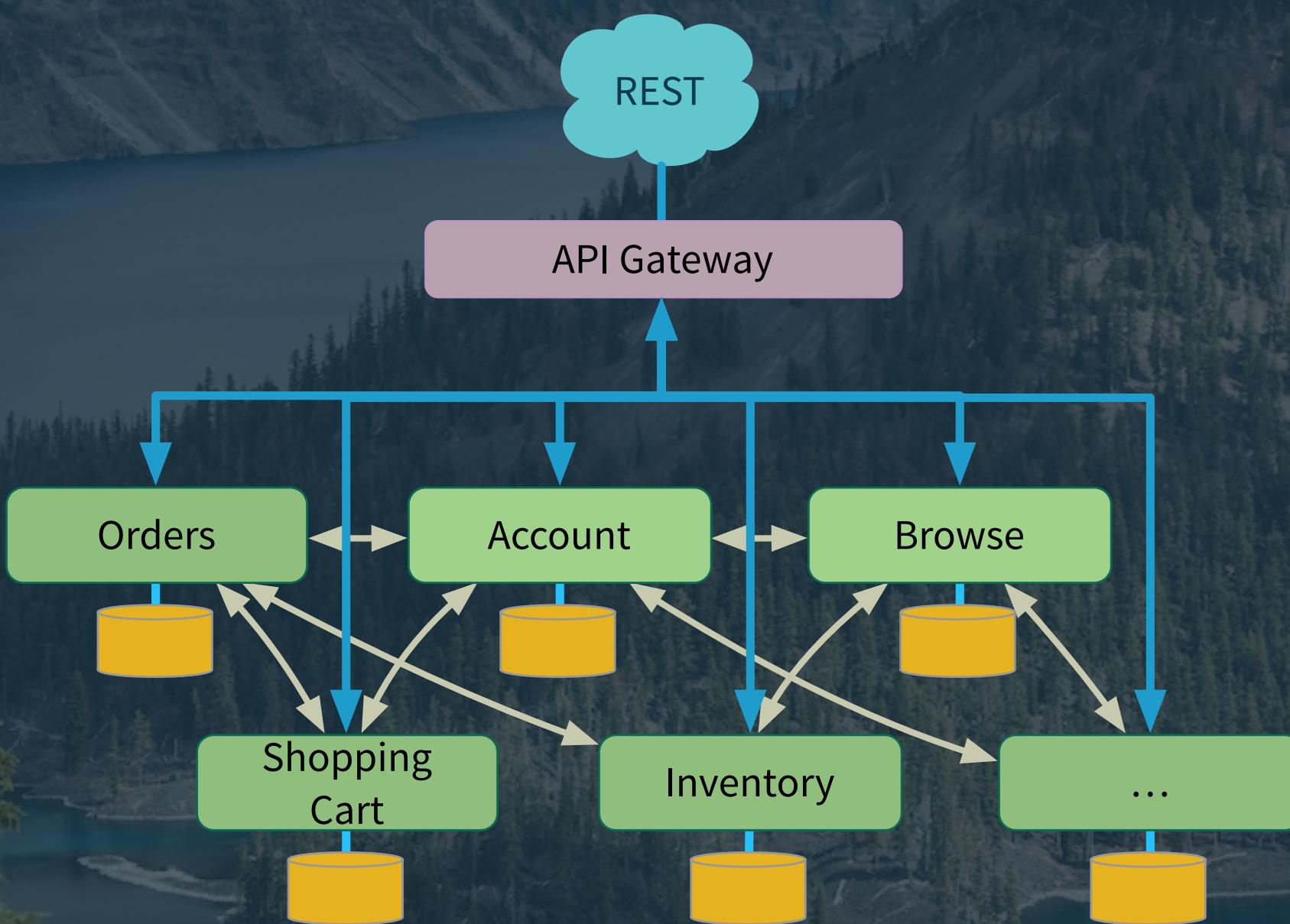
A wide-angle landscape photograph of a volcanic crater lake, likely Crater Lake in Oregon. The lake is a vibrant blue, contrasting with the dark, rocky slopes of the surrounding mountains. The mountains are covered in dense forests of coniferous trees, with some exposed rock and scree on the steeper slopes. The sky is filled with soft, white clouds.

# Streaming Architectures: Requirements

- Reliability - fault and “surprise” tolerant
- Availability - “always on”
- Low latency - for some definition of “low”
- Scalability - up and down
- Adaptability - ideally without restarts

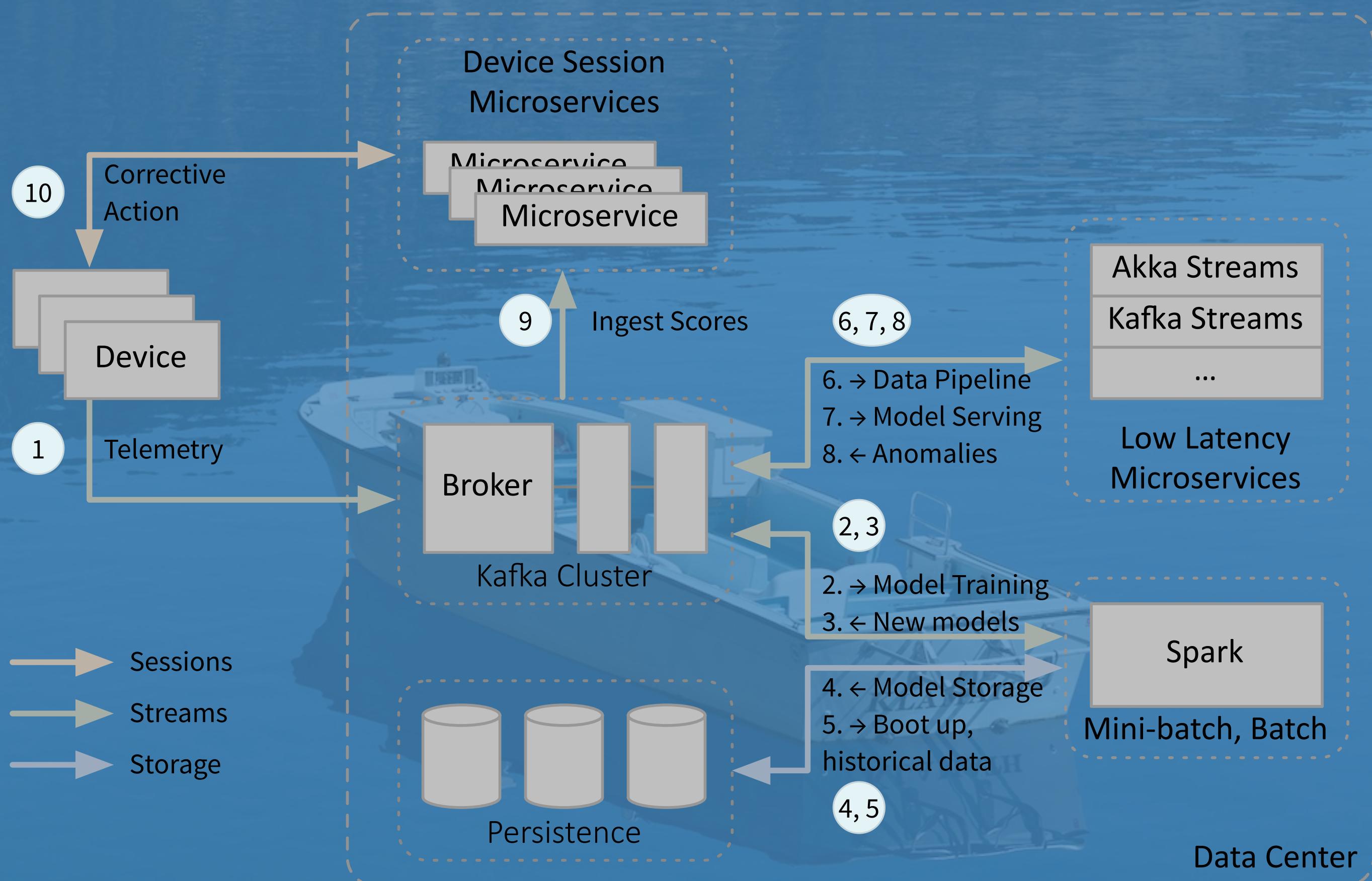


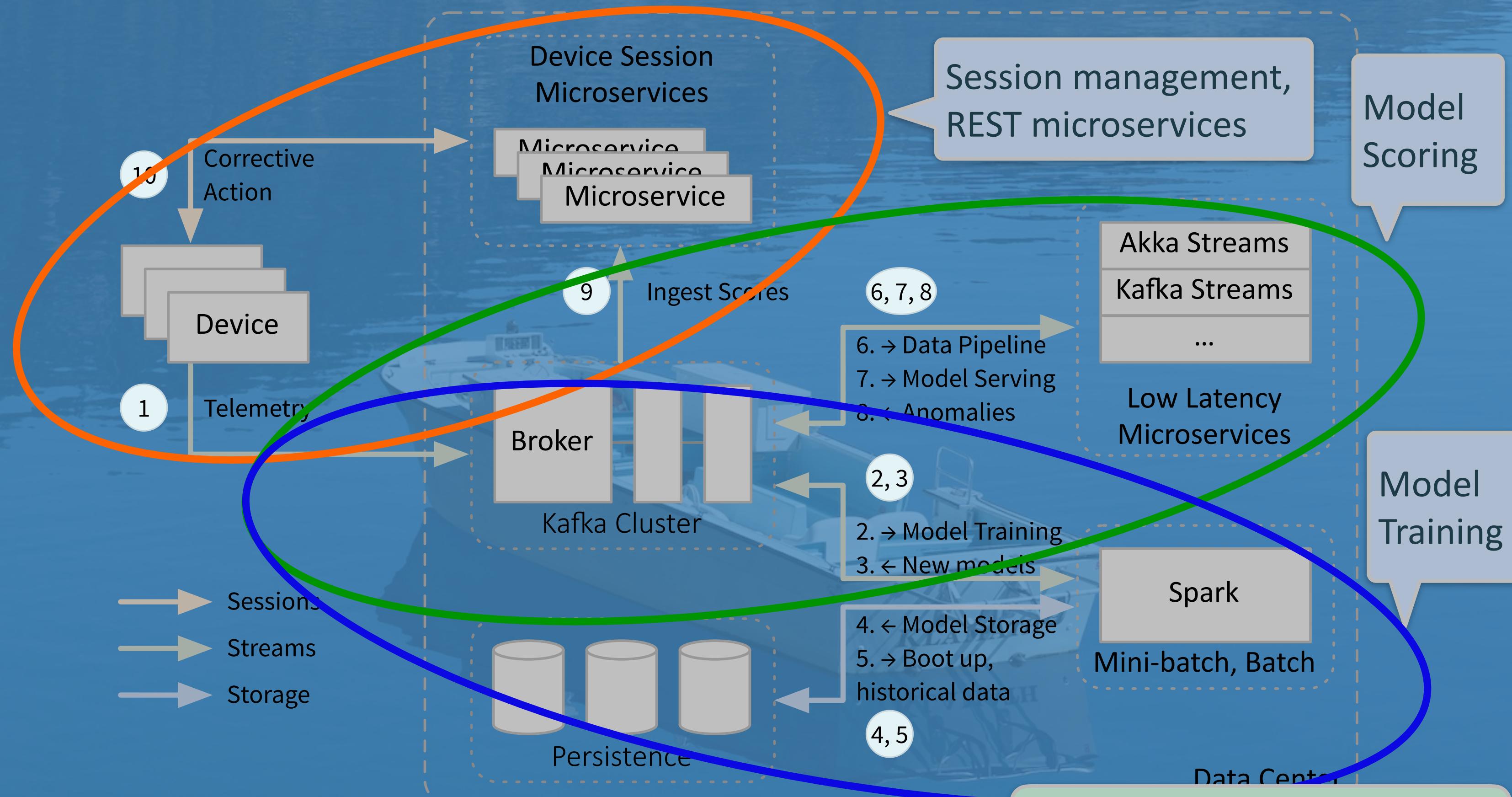
# Reminds me of microservices

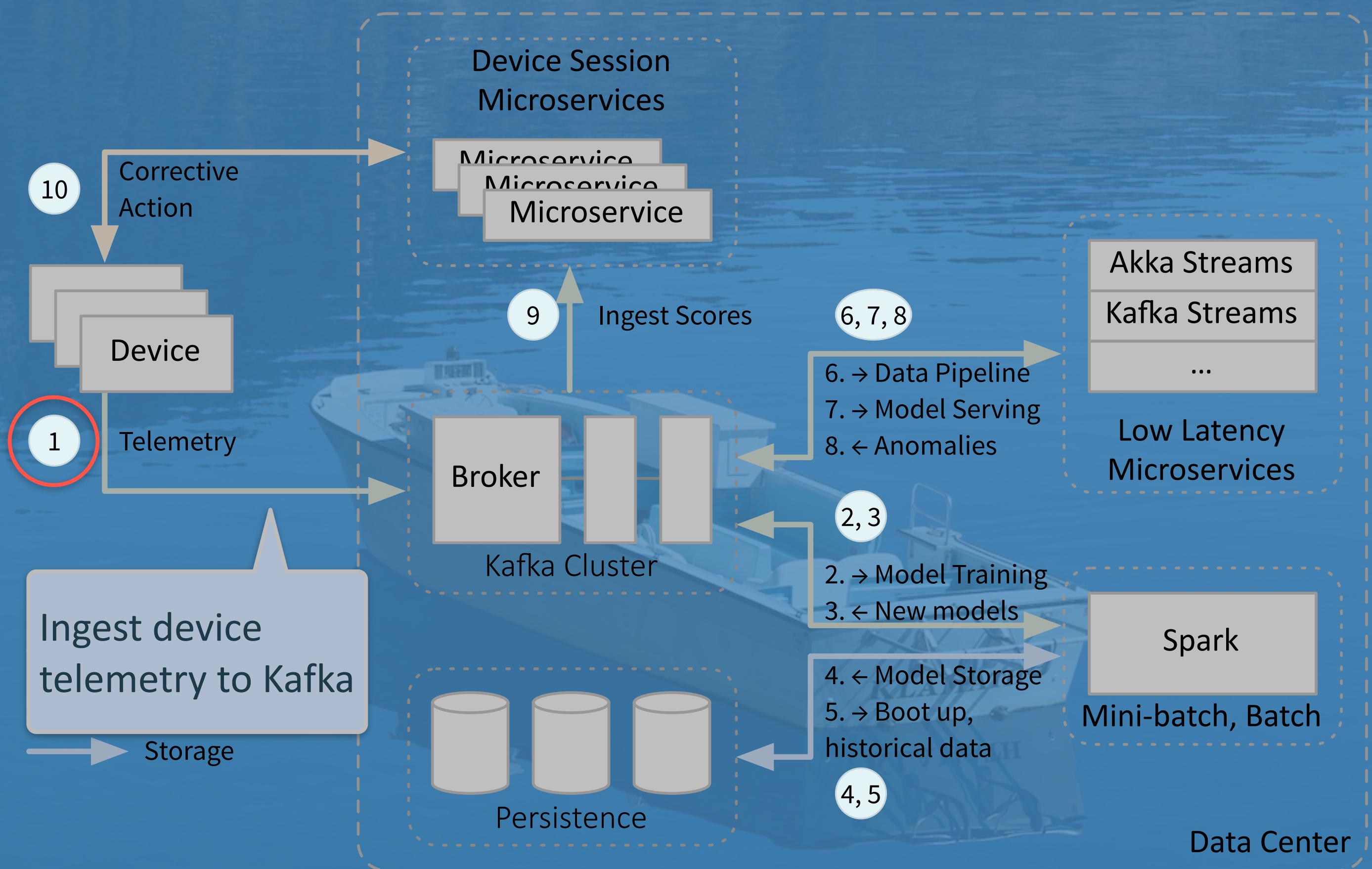


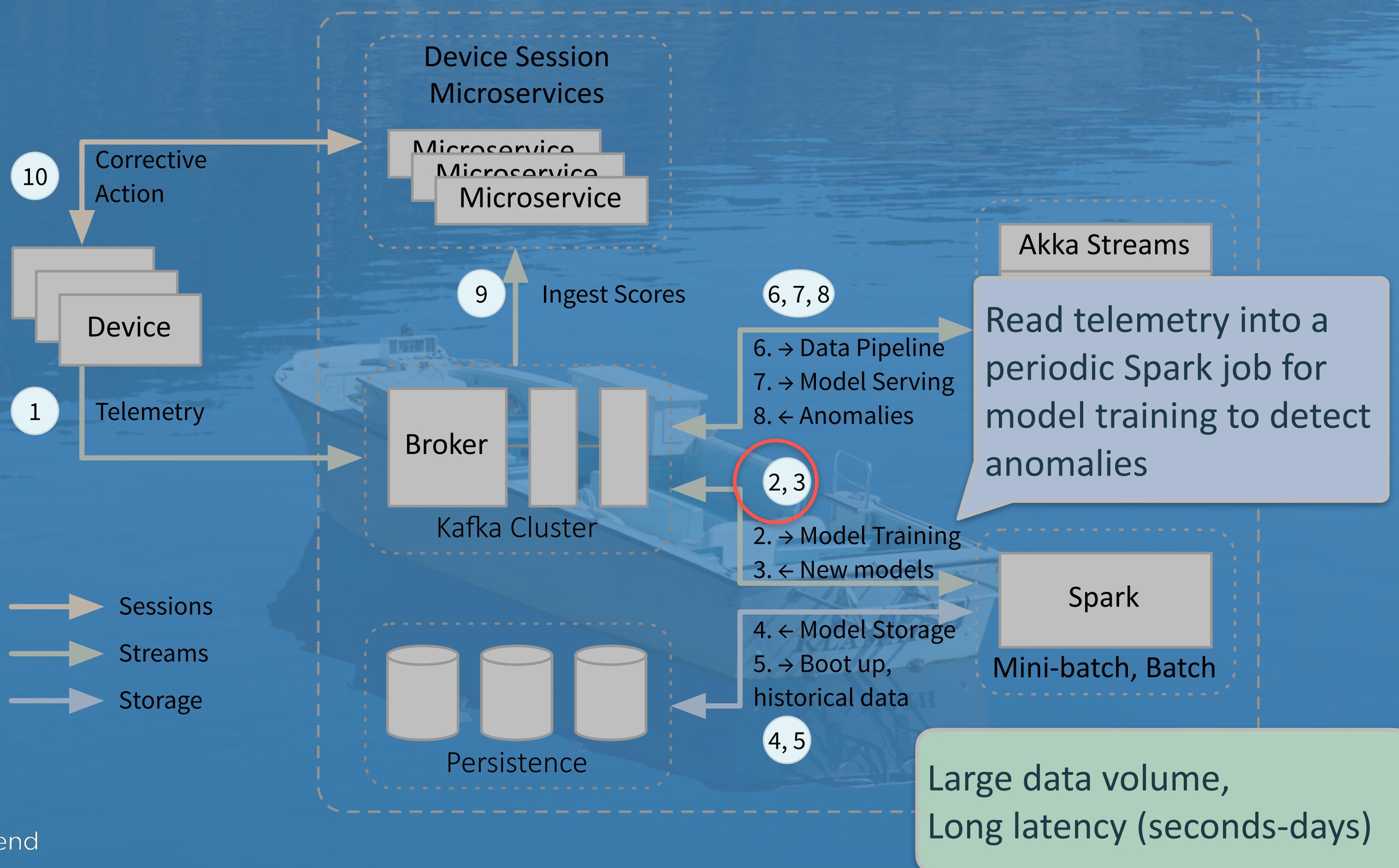
# Streaming Architectures: An Example

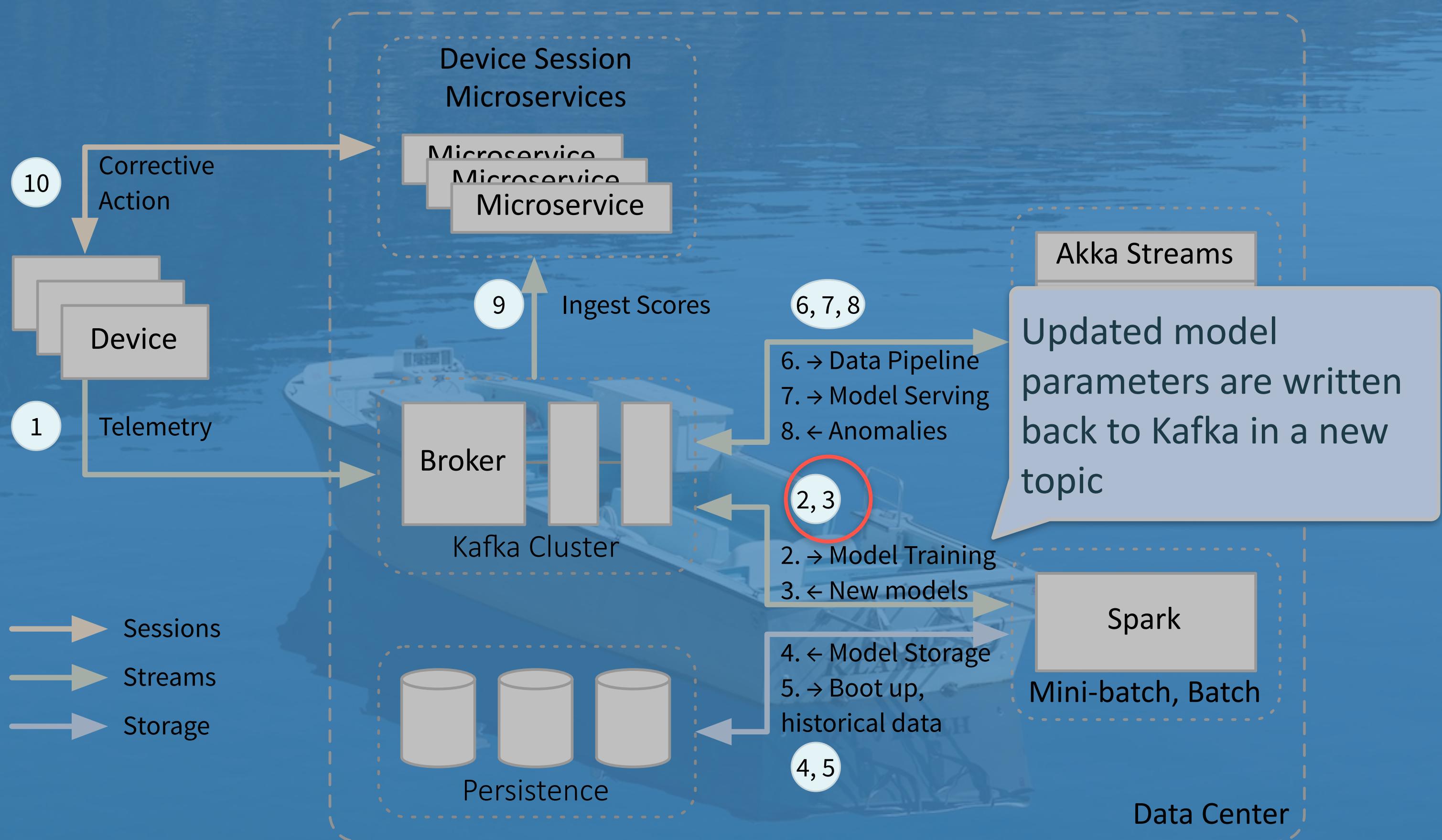


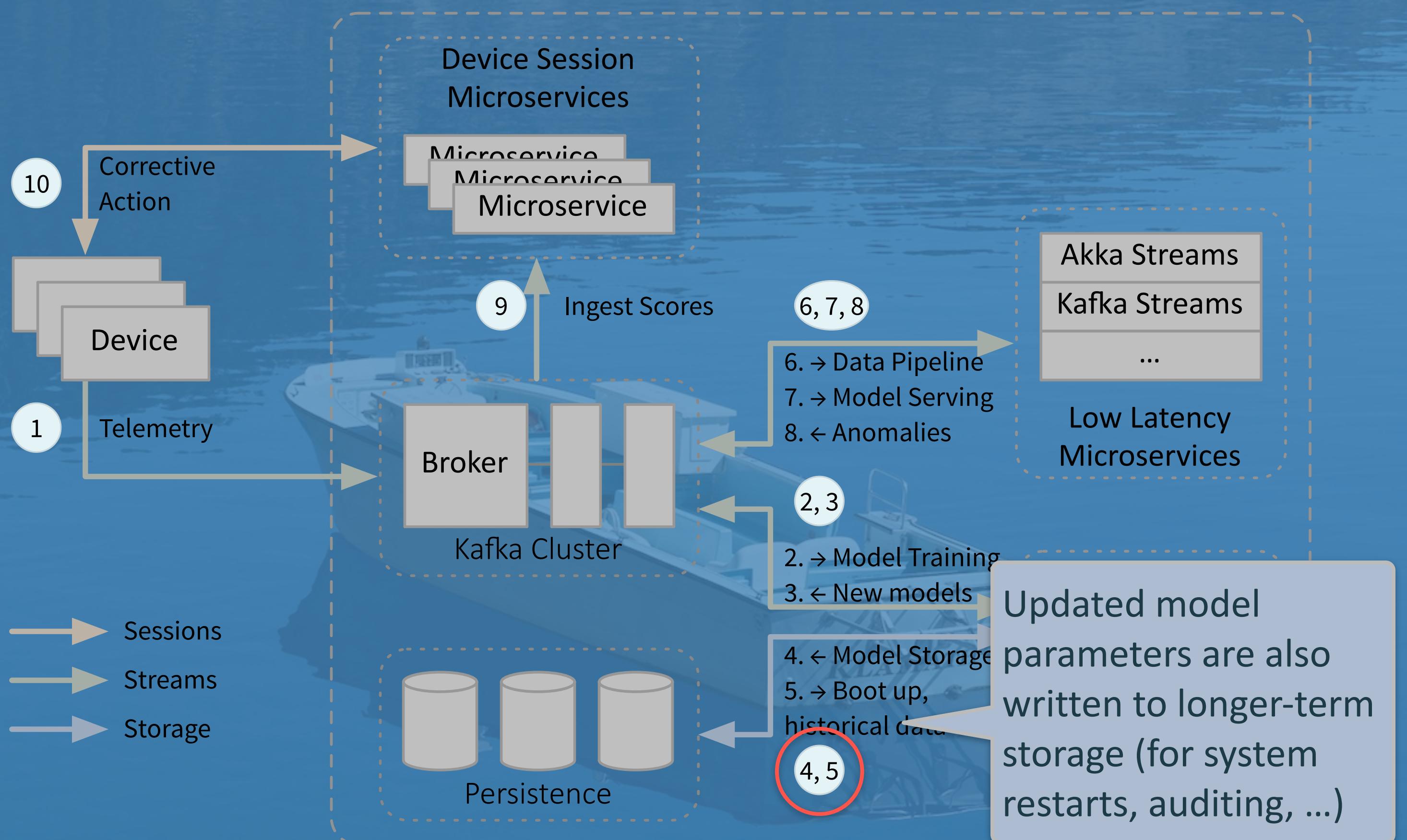


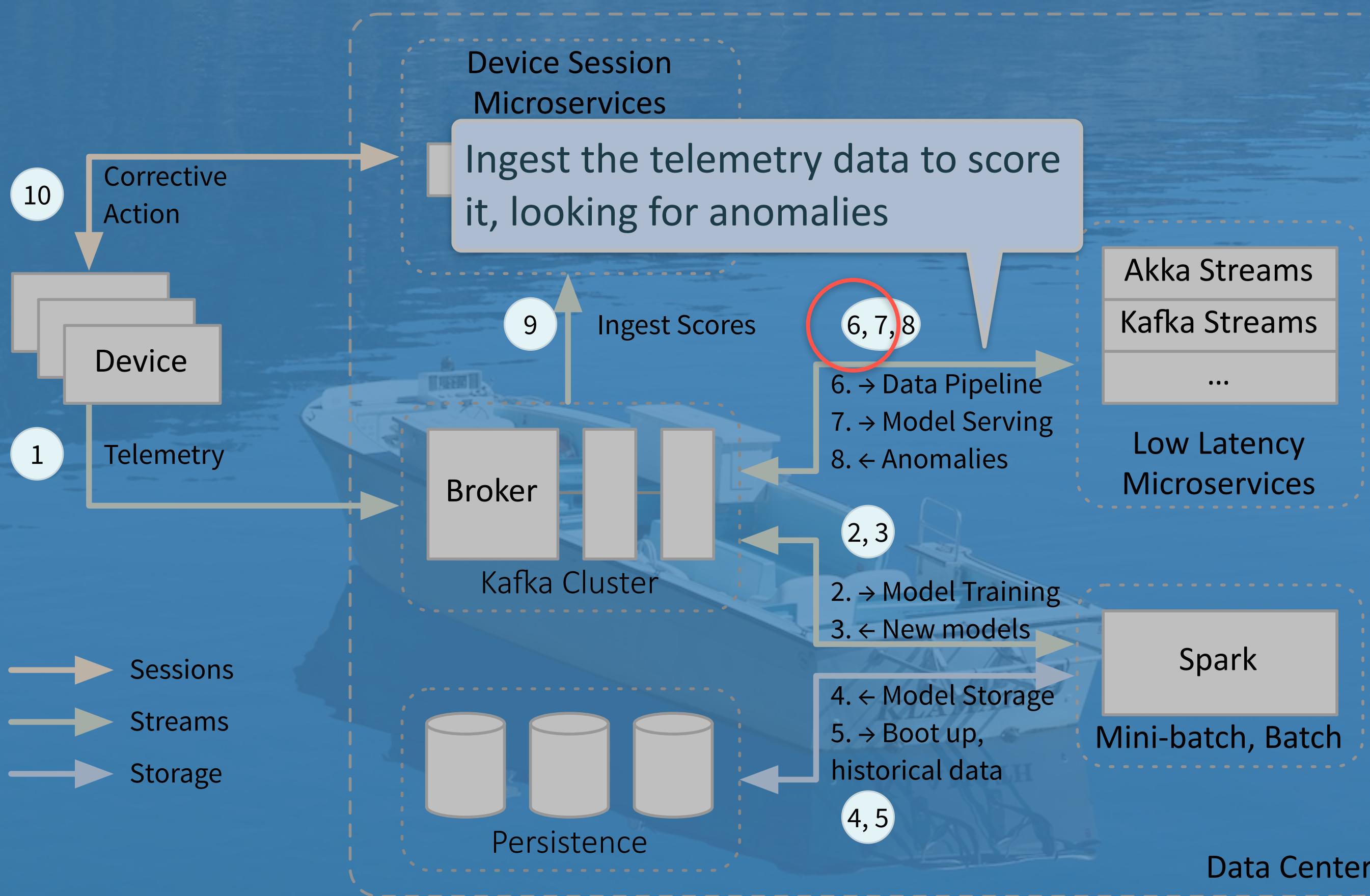


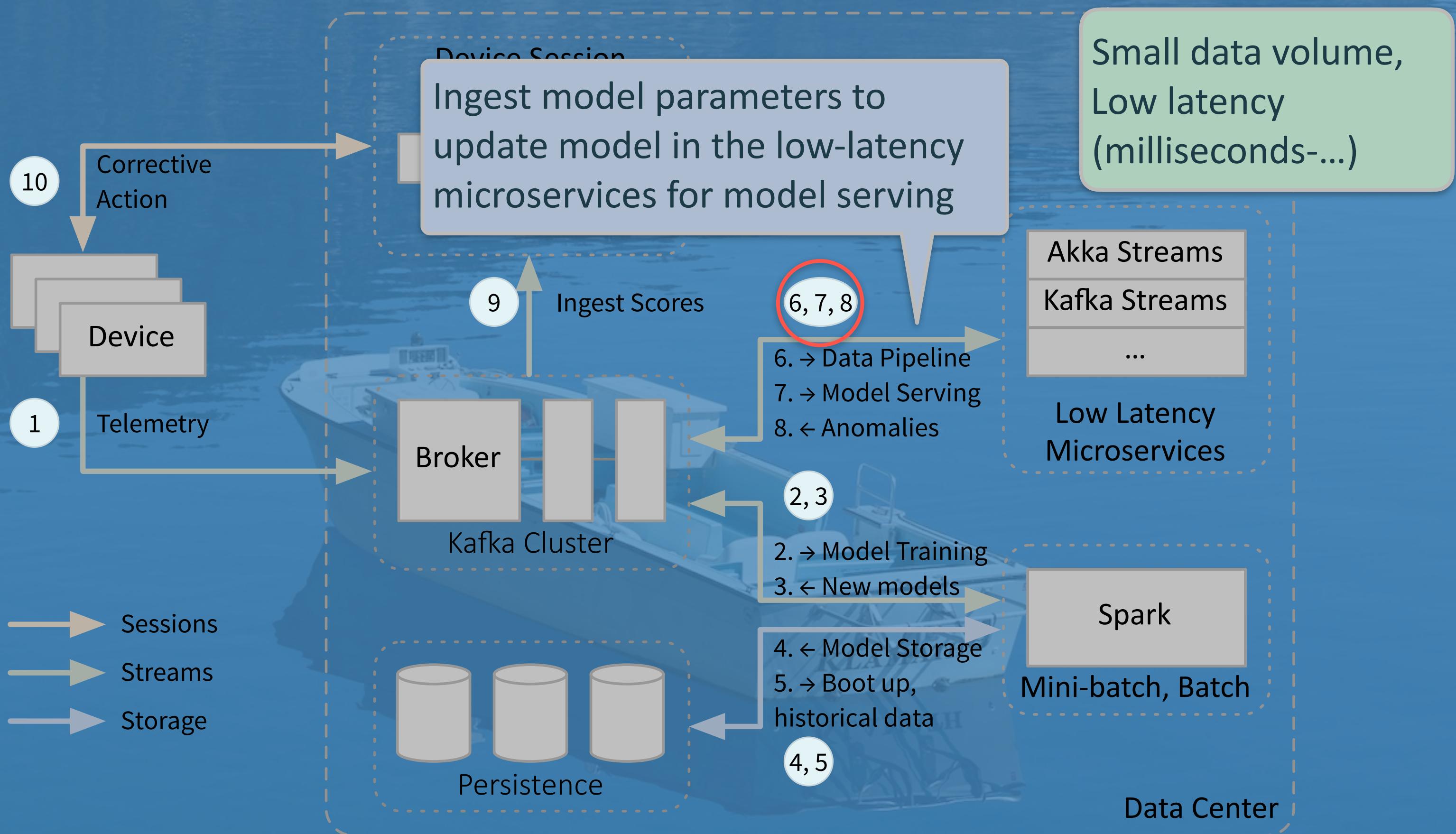


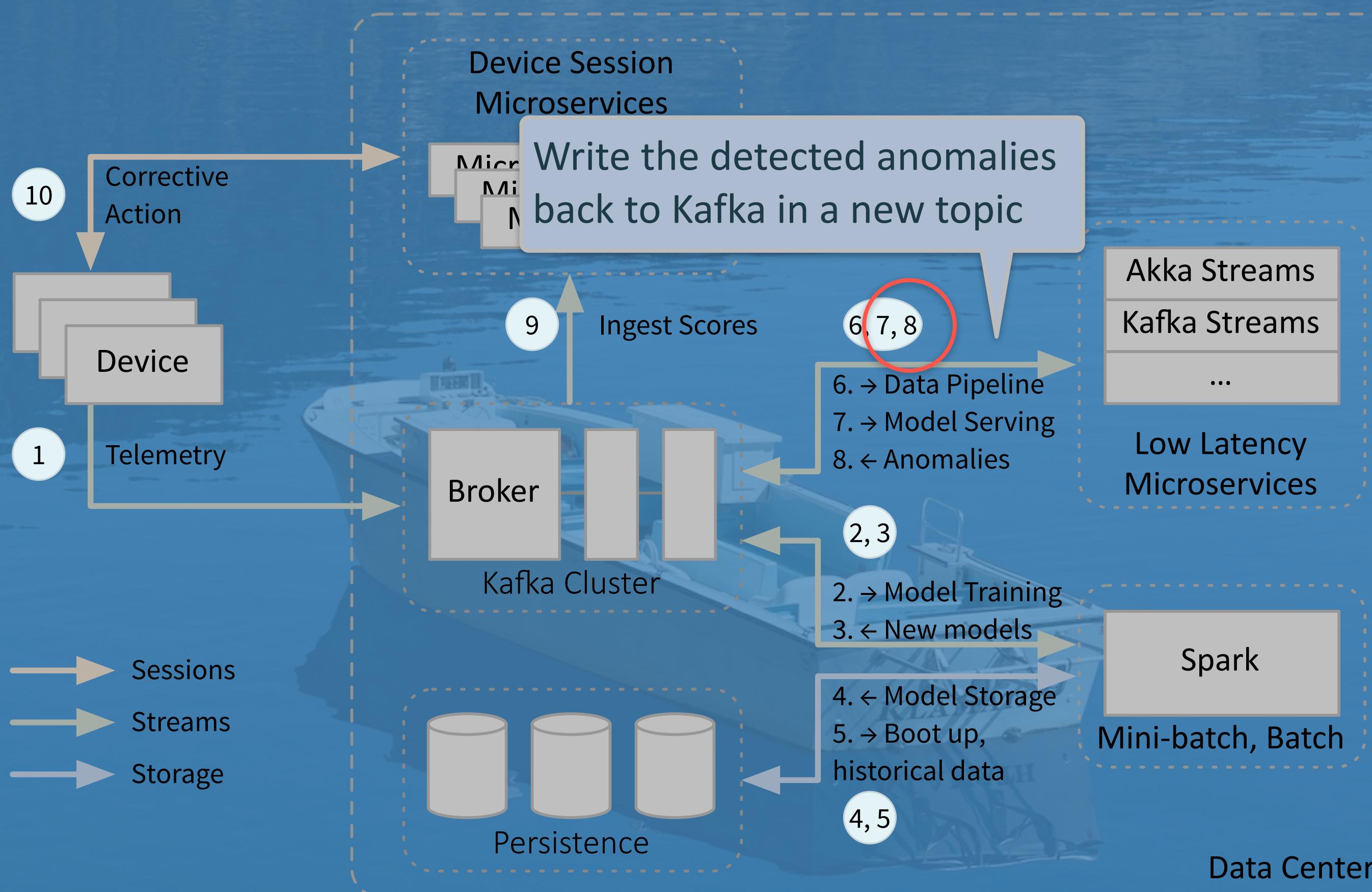


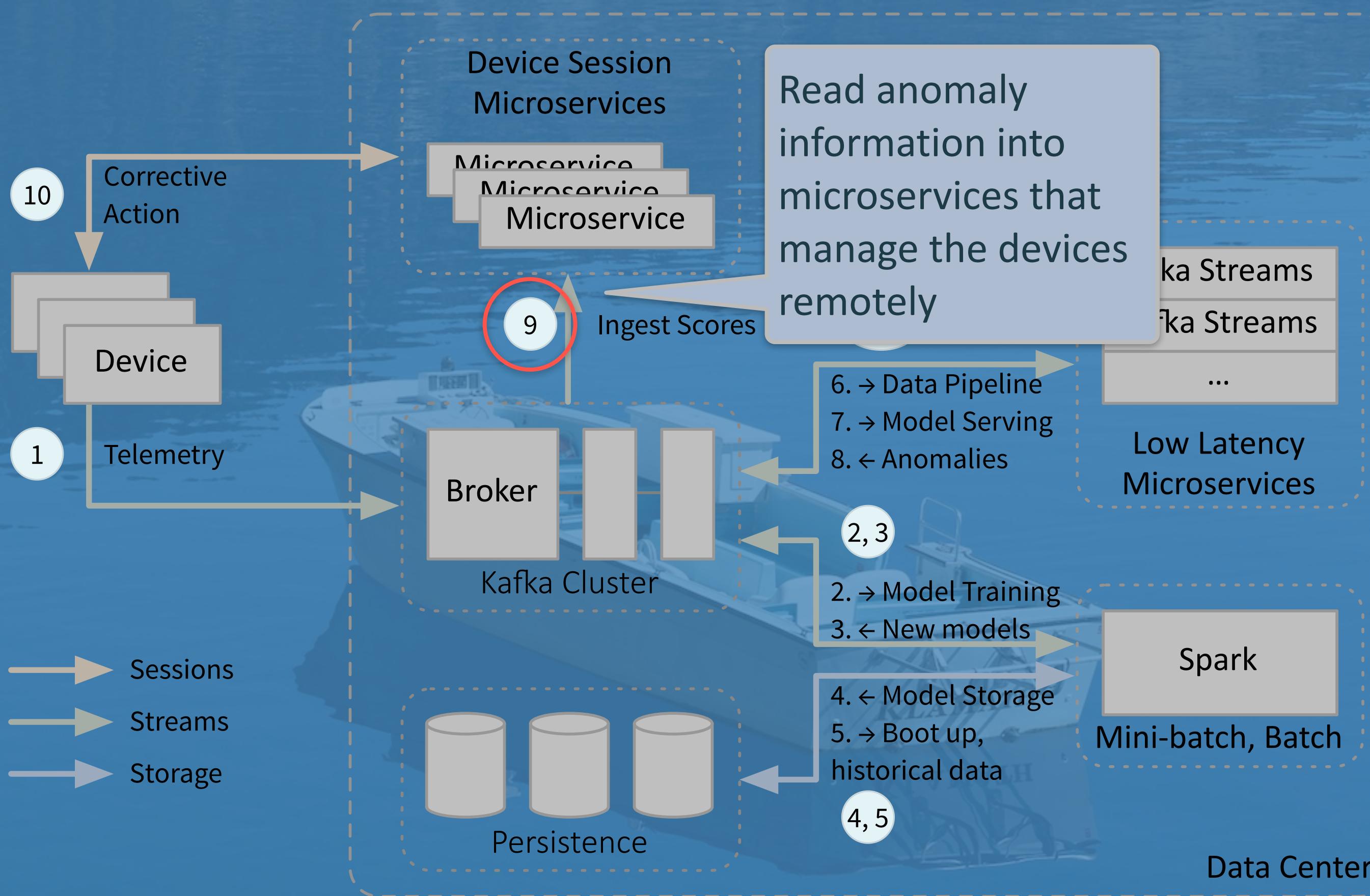


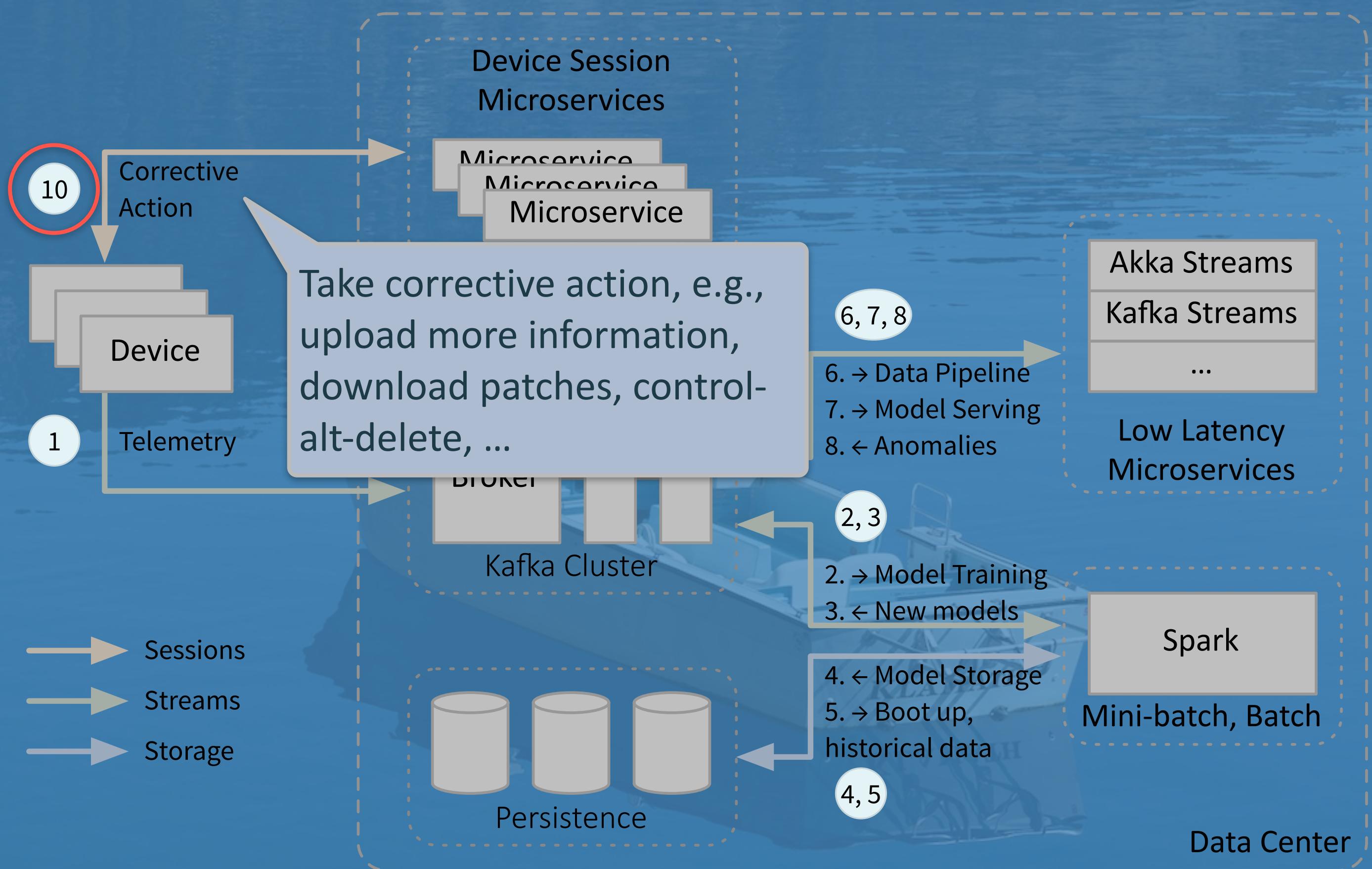


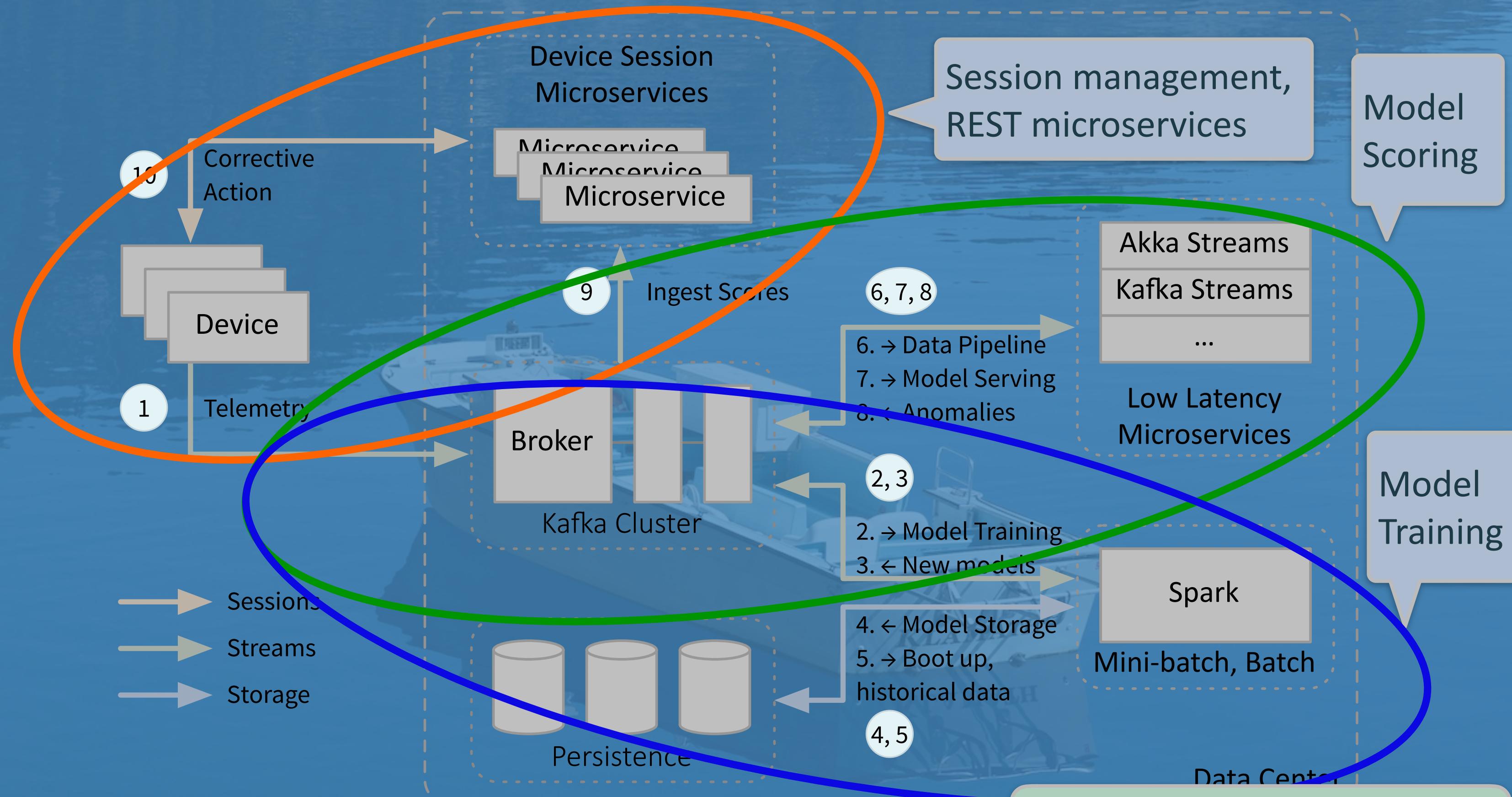












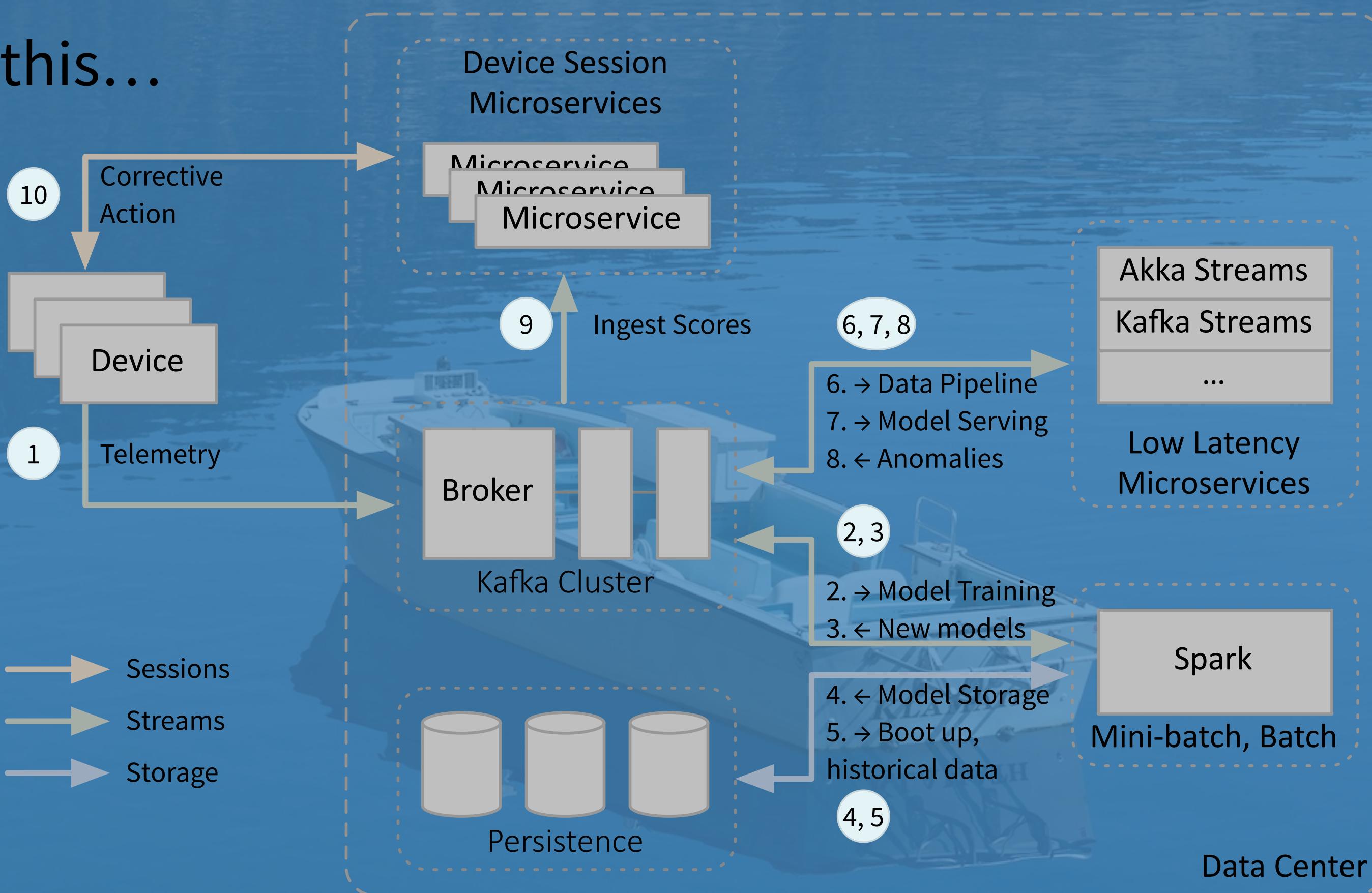
Three groups of functionality

# Streaming Architectures: An Example

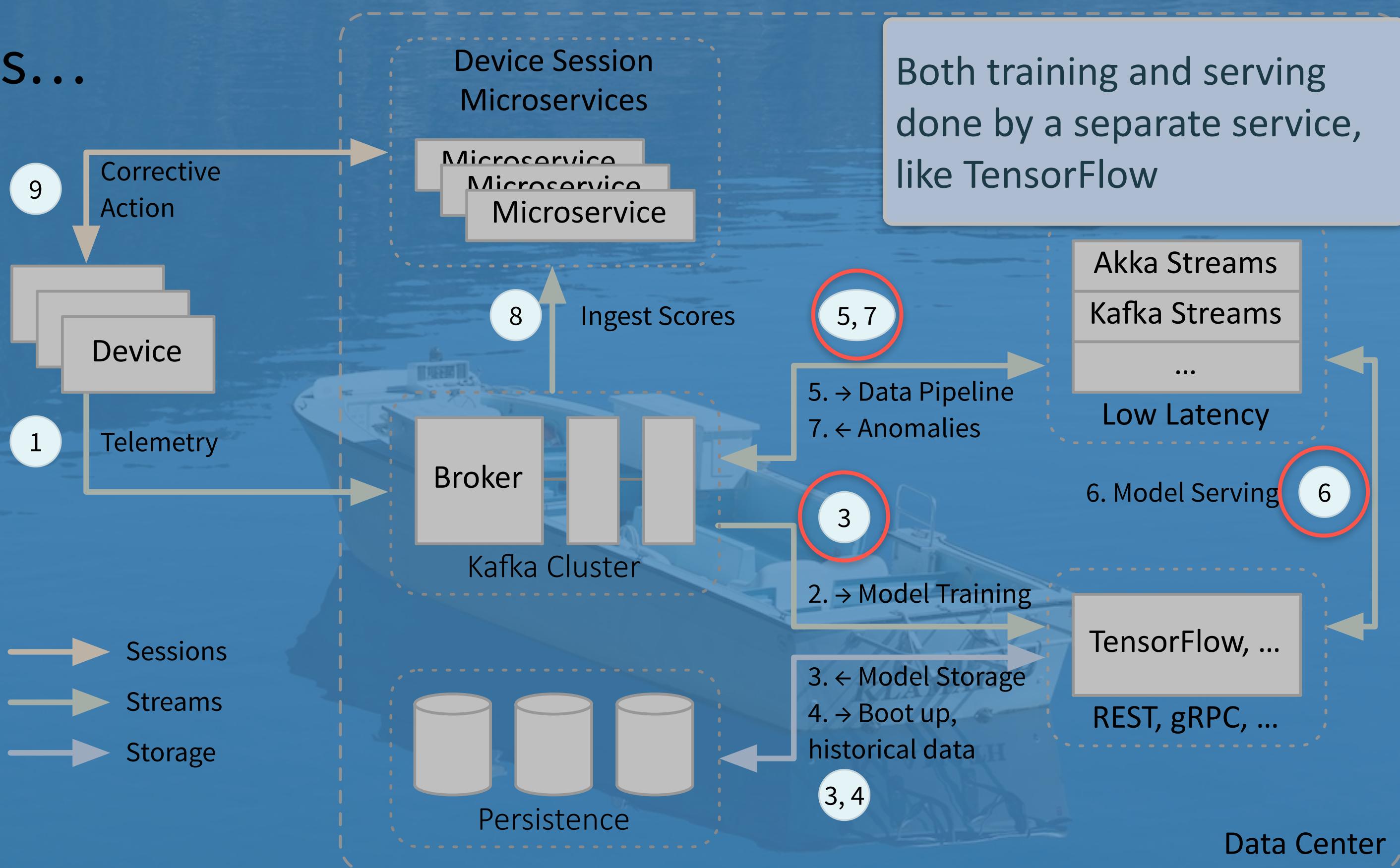
## Variation: Model Serving as a Service



# From this...



# To this...

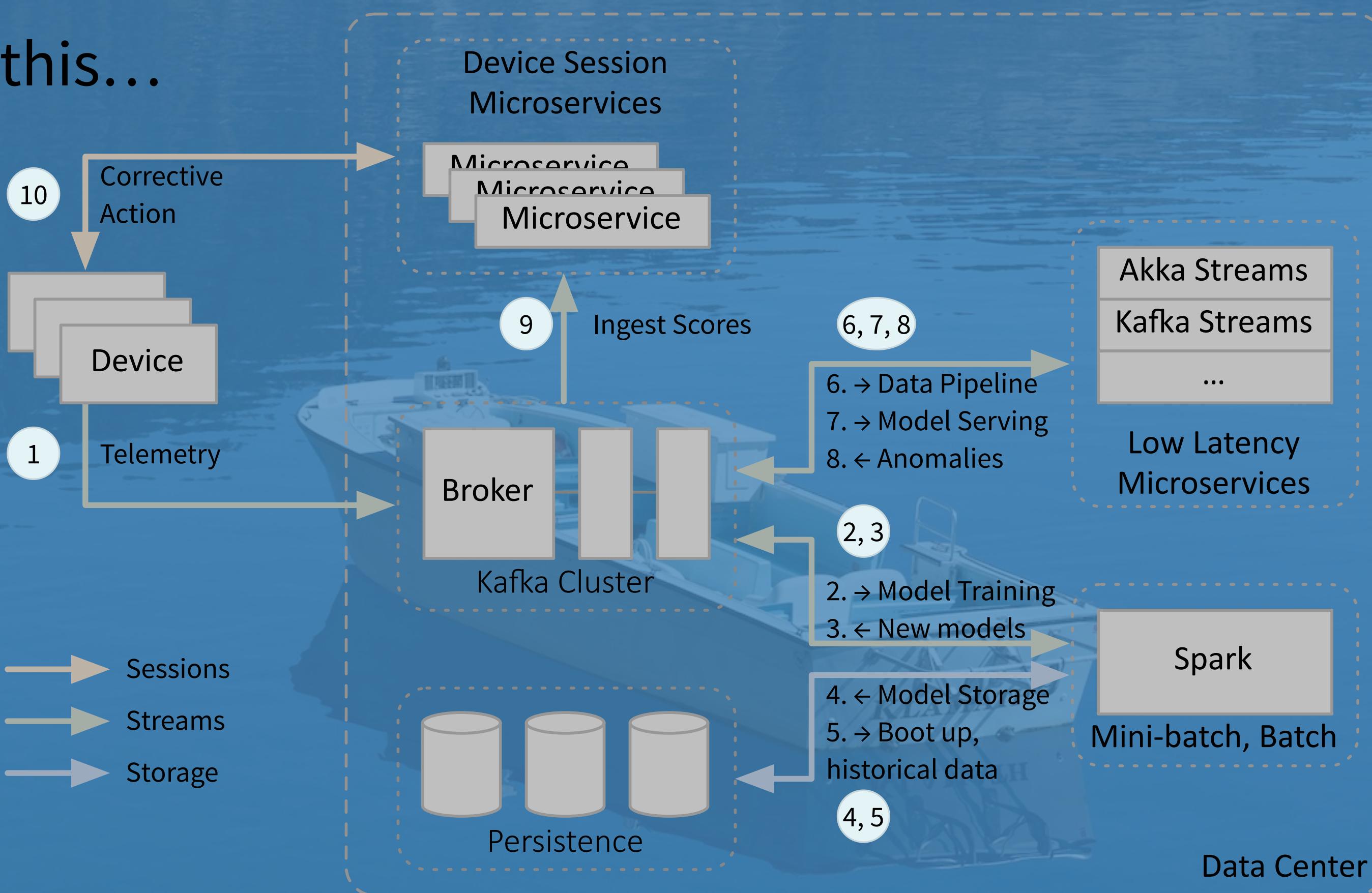


# Streaming Architectures: An Example

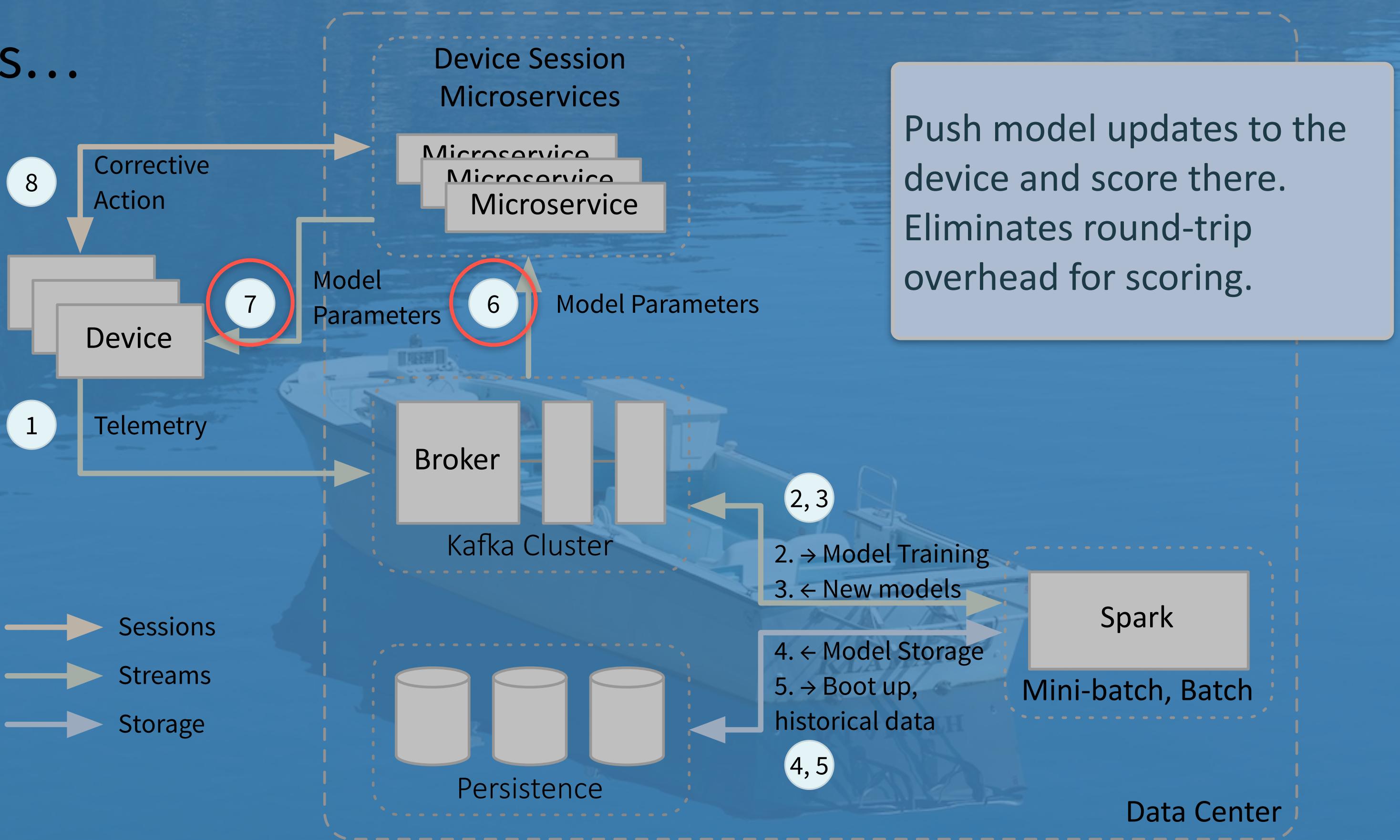
## Variation: Model Serving on the Edge



# From this...



# To this...



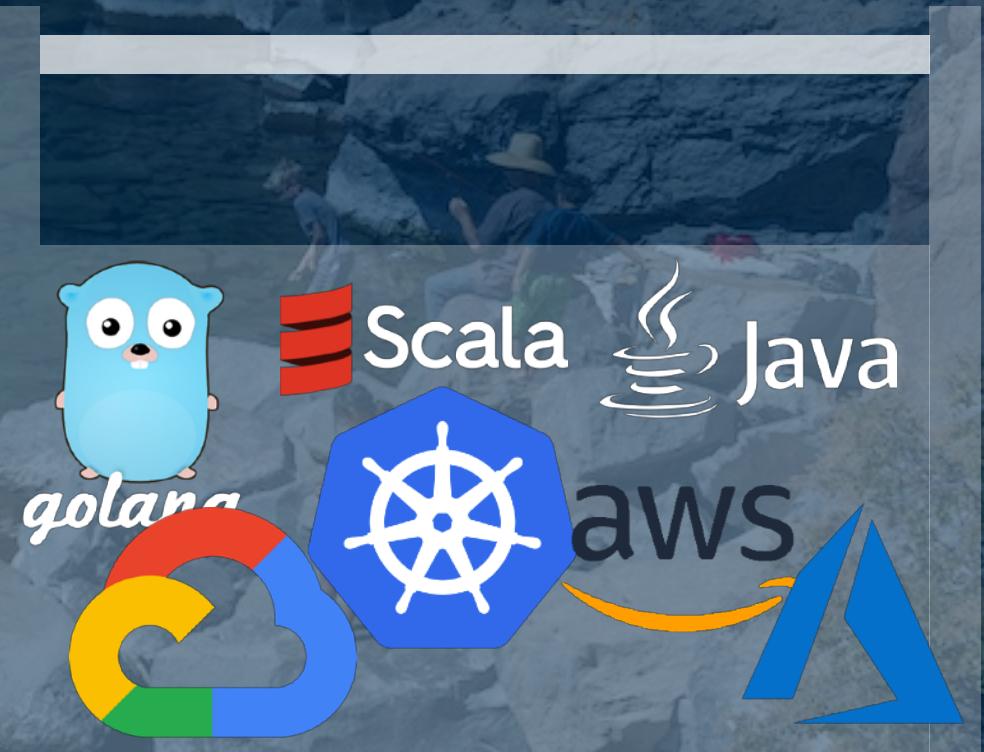
# Mixing data science and data engineering



# Tools



Data Science toolbox



Software  
Engineering toolbox

# Process

- Less process oriented
  - Iterative, experimental
  - .... but still within the Scientific Method
- Process oriented
  - Agile Manifesto
  - .... which does not mention data!

# Philosophy

- Comfortable with uncertainty
  - Probabilities and Statistics
- Uncomfortable with uncertainty
  - Prefer determinism
  - ... while admitting that distributed systems lack guarantees

# Other Production Concerns

Auditing

Models  
Are  
Data

# Auditing

- Kind of Model
- Parameters and hyperparameters
- When trained
- Data used for training
- When deployed, undeployed, etc.
- ...

# Auditing

- Quality metrics
- Serving metrics (how many records, scoring times...)
- Provenance of decision to retrain
  - The metrics gathered above that were used to decide when to retrain

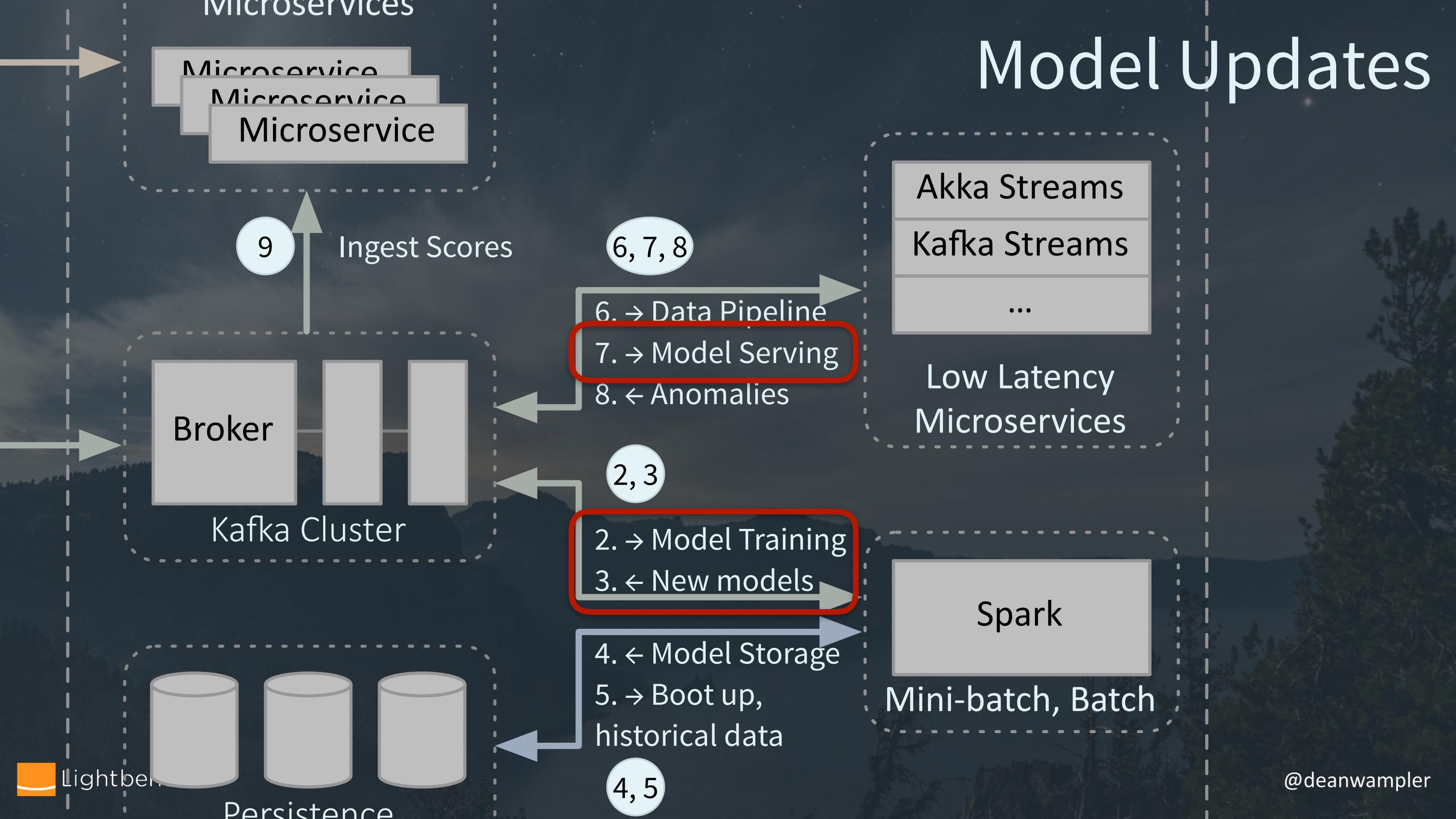
# Model Updates

## Concept Drift

Models have a half life, too

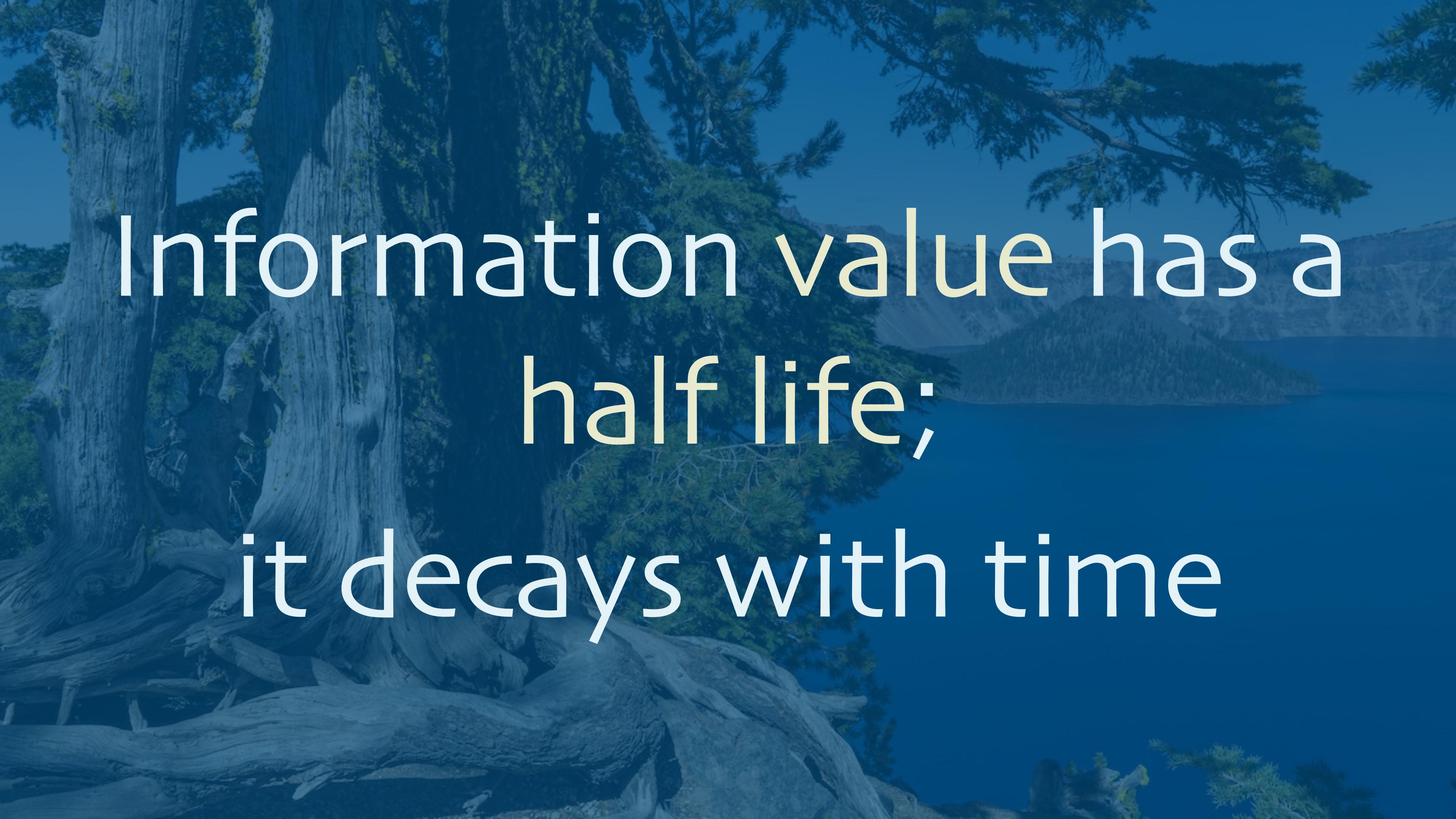
# Microservices

# Model Updates



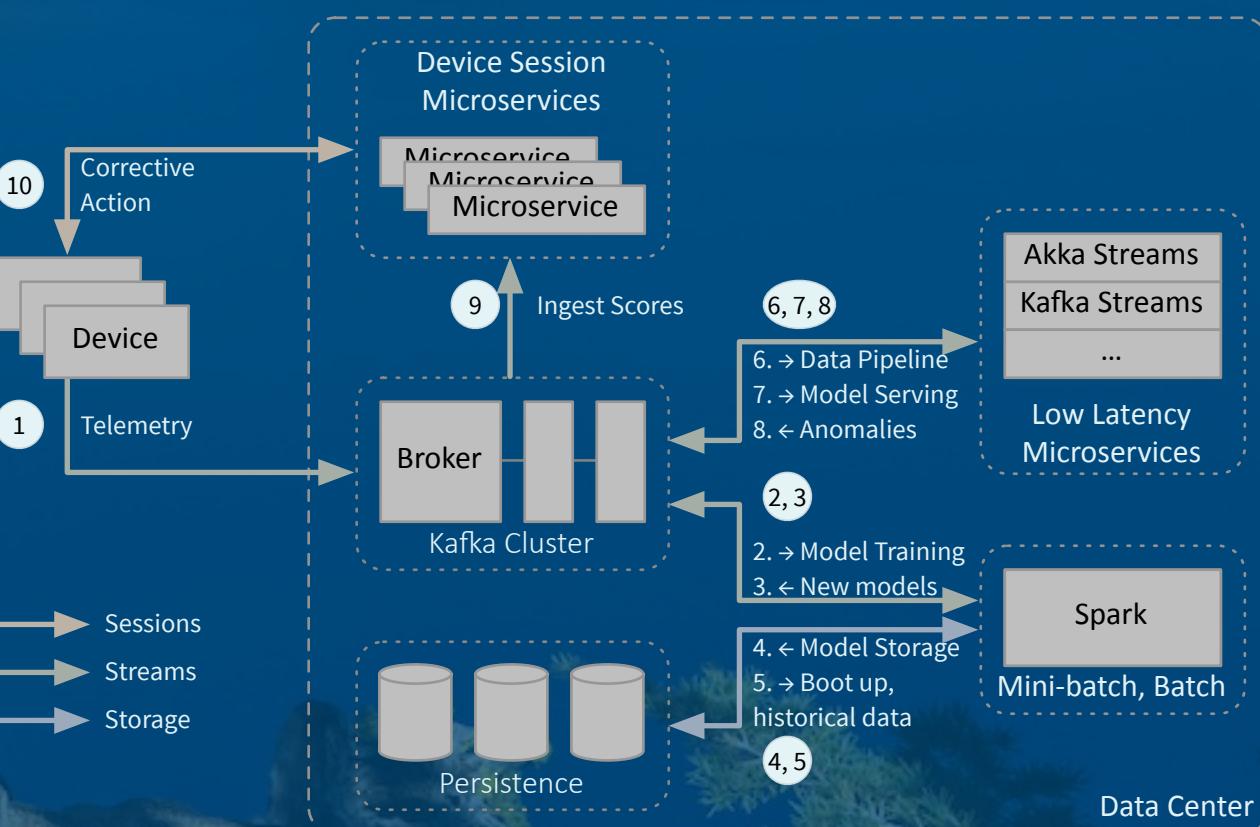
A scenic view of Crater Lake National Park. In the foreground, several ancient, gnarled tree trunks stand on a rocky outcrop. One tree on the left has bright yellow lichen growing on its bark. Behind the trees, a deep blue lake stretches towards a rugged, layered mountain range. The sky is clear and blue.

Recap

A large, ancient tree trunk with a textured bark surface, set against a background of green foliage and a body of water.

Information value has a  
half life;  
it decays with time

# Streaming systems have similar requirements as microservices



A photograph of a massive tree trunk, likely a sequoia, with intricate, layered bark. The trunk is positioned on the left side of the frame, receding towards a dense forest and a range of mountains under a clear blue sky.

Integrating  
data science and  
data engineering  
is hard



Dusty Milky Way

Mars last Summer



Lightroom

@deanwampler

# References

- General Information about Stream Processing
- [My O'Reilly Report on Architectures](#)
- [Streaming Systems Book](#)
- [Stream Processing with Apache Spark](#)
- [Designing Data-Intensive APPS book](#)

# References

- Other Talks
  - [Strata Talk on ML in a Streaming Context](#)
  - [Stream All the Things! \(video\)](#)
  - [Streaming Microservices with Akka Streams and Kafka Streams \(video\)](#)

# References

- Tutorials
  - [Model serving in streams](#)
  - [Stream processing with Kafka and microservices](#)

Controls

GRAFANA WORKLOADS

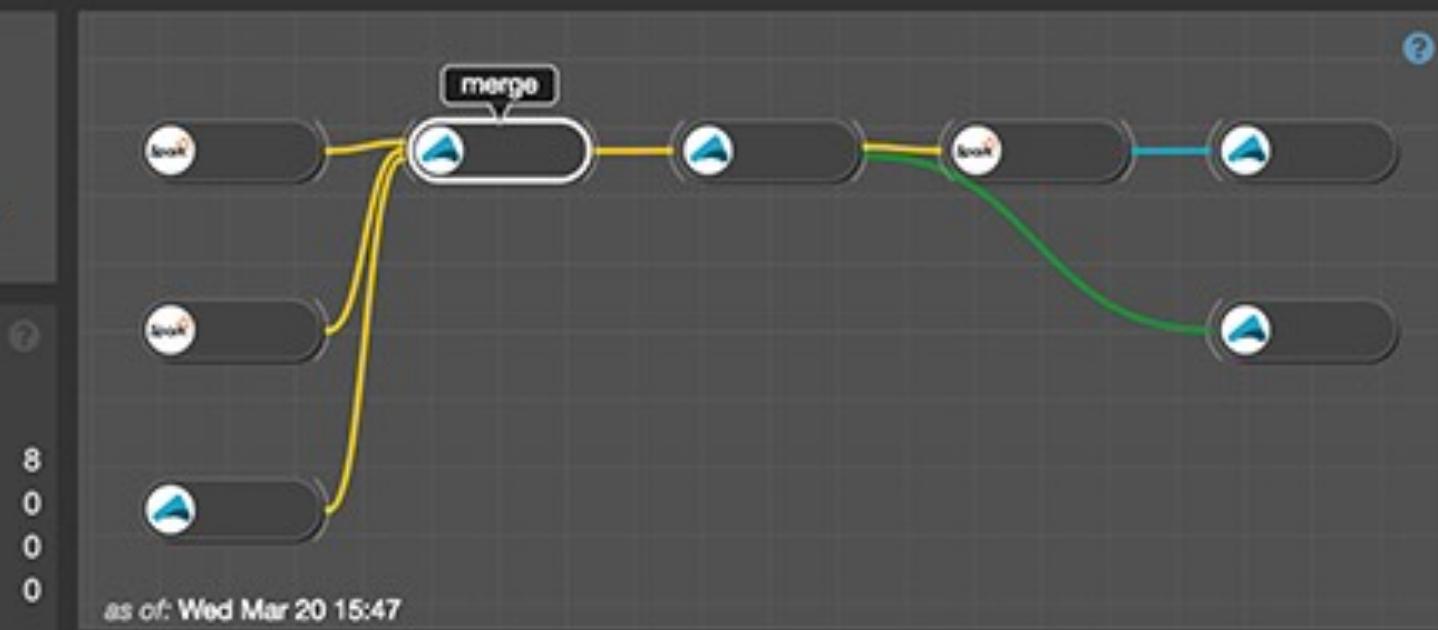
## Application Details

## Streamlet Current Health

Healthy	8
Warning	0
Critical	0
Unknown	0

## Streamlet Health Events

cdr-validator	—
cdr-aggregator	—
merge	—
console-egress	—
error-egress	—
cdr-generator1	—
cdr-generator2	—
cdr-ingress	—



## Streamlet Health

Wed 16:00

Buy my stuff!!

Diagnostics Ports

SORT BY First unhealthy

## Streamlet Monitors

kafka\_consumer\_lag

kafka\_consumer\_throughput

kafka\_producer\_throughput

## Consumer Lag (records behind)

35.00  
30.00  
25.00  
20.00  
15.00  
10.00  
5.00  
0.00

Maximum Consumer Lag (records behind)

35.00



What we're up to at Lightbend...  
[lightbend.com/lightbend-pipelines-demo](http://lightbend.com/lightbend-pipelines-demo)

# Questions?

Dean Wampler, Ph.D.  
[dean@lightbend.com](mailto:dean@lightbend.com)  
[@deanwampler](https://twitter.com/deanwampler)  
[lightbend.com/lightbend-platform](http://lightbend.com/lightbend-platform)  
[polyglotprogramming.com/talks](http://polyglotprogramming.com/talks)