

# Executive Briefing: What You Need to Know about Fast Data

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



Based on  
this report

[lightbend.com/fast-data-platform](http://lightbend.com/fast-data-platform)  
2nd edition coming in October!

# Fast Data Architectures for Streaming Applications

Getting Answers Now from  
Data Sets that Never End

A black and white photograph of a rocky riverbed. The water is flowing rapidly, creating white foam and ripples around the large, smooth stones. The scene conveys a sense of constant motion and flow.

Dean Wampler



# What We'll Discuss

- 
- A night photograph of the London skyline, centered on the Tower Bridge. The bridge's towers are illuminated, and the city lights of London are visible in the background across the River Thames.
- Why streaming? Why now?
  - How to choose technologies
  - The impact streaming will have on your organization

What We'll Discuss



Why Streaming?

- New opportunities that require streaming
  - Media content is obviously one ;)
  - Upgrading batch applications for competitive advantage

Why Streaming?



## Similar IoT Architectures

# Fast Data Use Cases

### Predictive Analytics

Apply ML models to large volumes of device data to pre-empt failures / outages



**Hewlett Packard  
Enterprise**

### IoT

Real-time consumer and industrial Device and Supply Chain management at scale



### Real-time Personalization

Real-time marketing based on behavior, location, inventory levels, product promotions, etc.

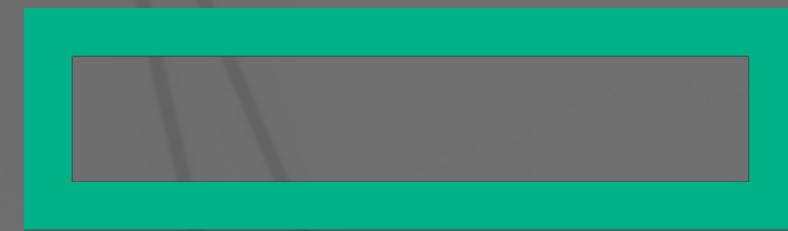


**RoyalCaribbean  
INTERNATIONAL**®

### Real-time Financial Processes

Drive better business outcomes through real-time risk, fraud detection, compliance, audit, governance, etc.





# Predictive Analytics

## Hewlett Packard Enterprise

- ML models applied to device telemetry to detect anomalies
- Preemptive maintenance prevents potential failures that would impact users

# Predictive Analytics - Core Idea

Handle anomaly: move activity off component, schedule maintenance window to replace it.



Corrective  
Actions

Anomaly  
Handler

Probable  
Anomalies

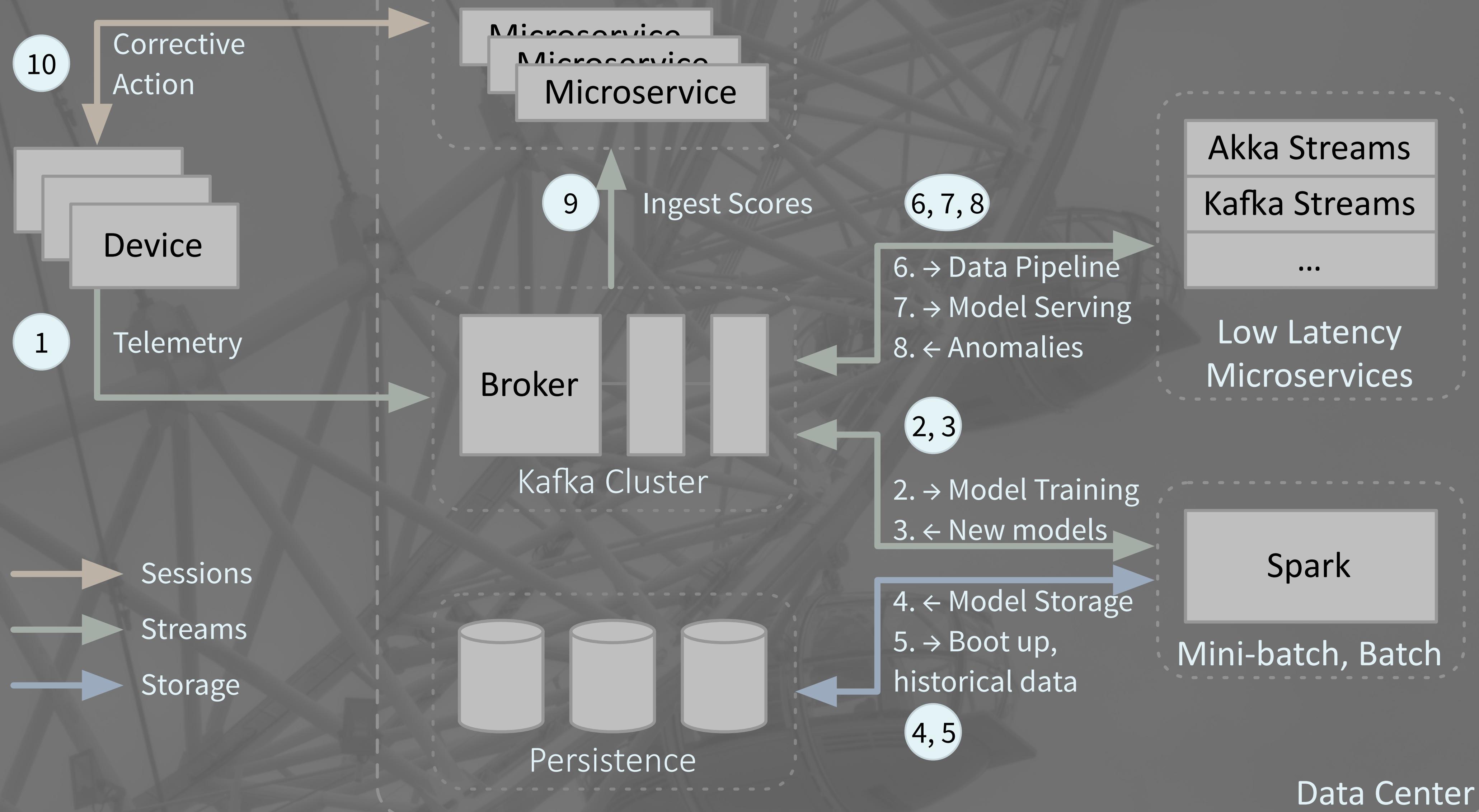
Anomaly  
Detection:  
Model

Telemetry  
Records

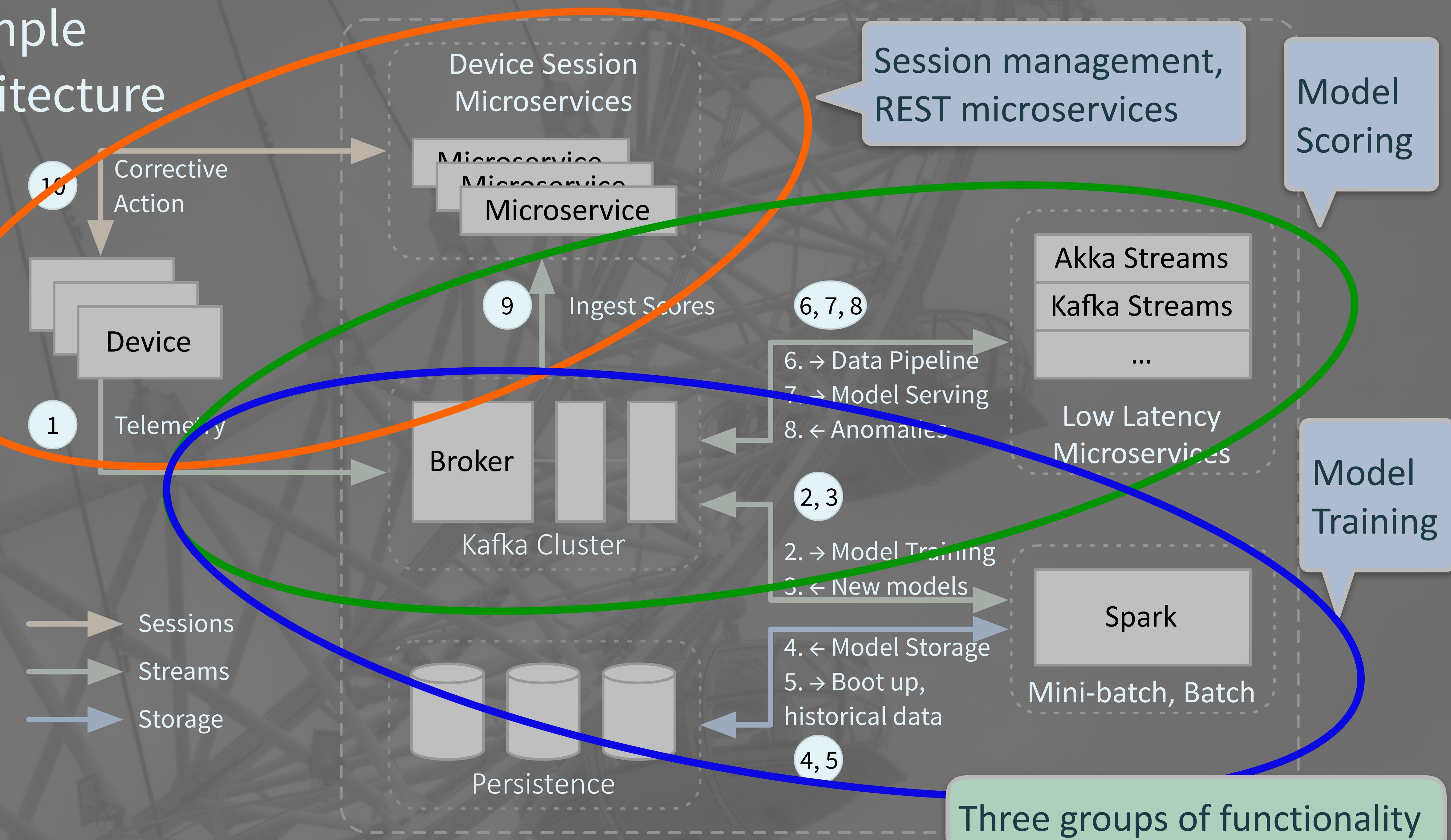
Ingest telemetry from  
edge devices.

Train models to look for  
anomalies... and score  
incoming telemetry.

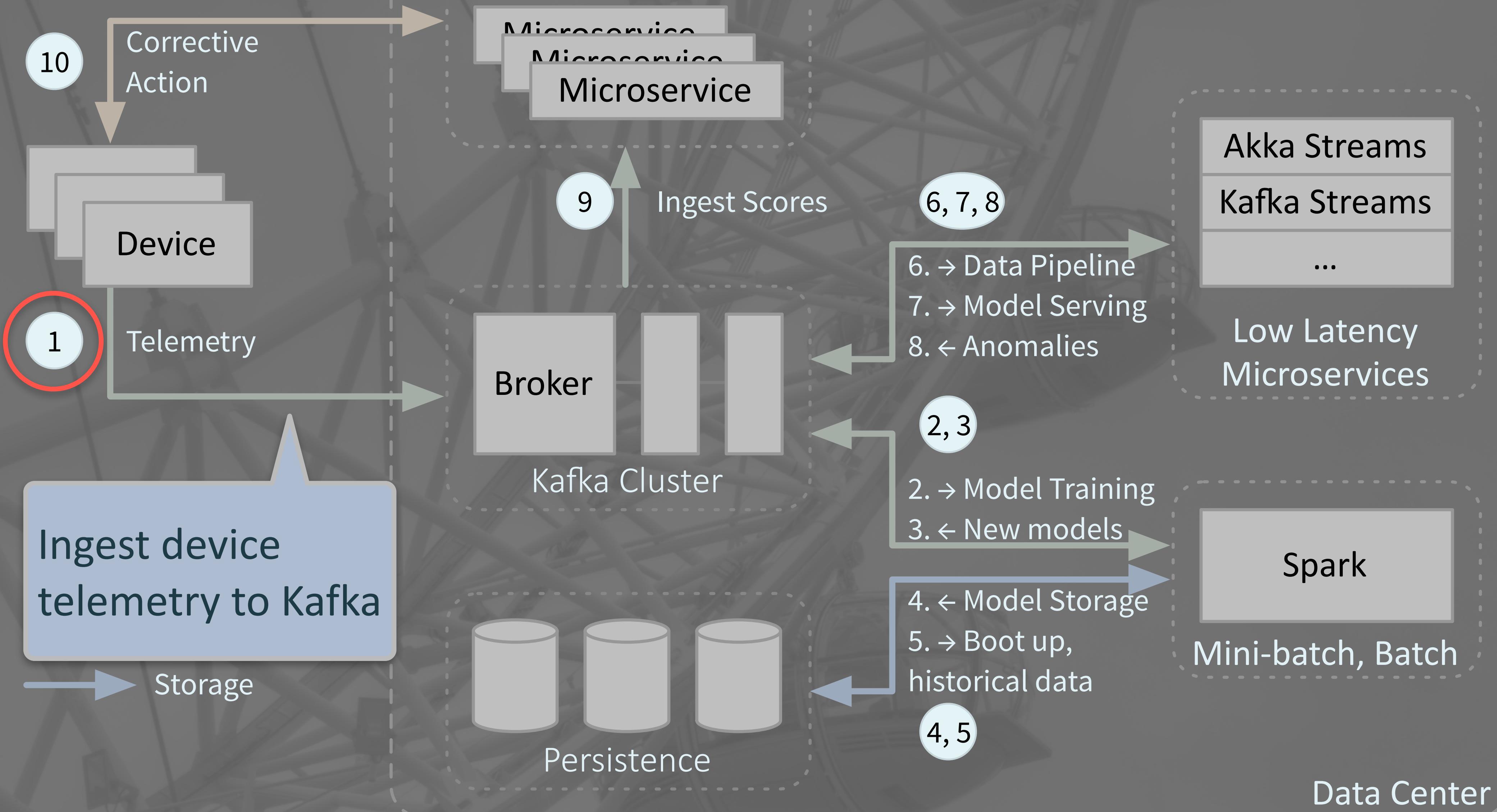
# Example Architecture



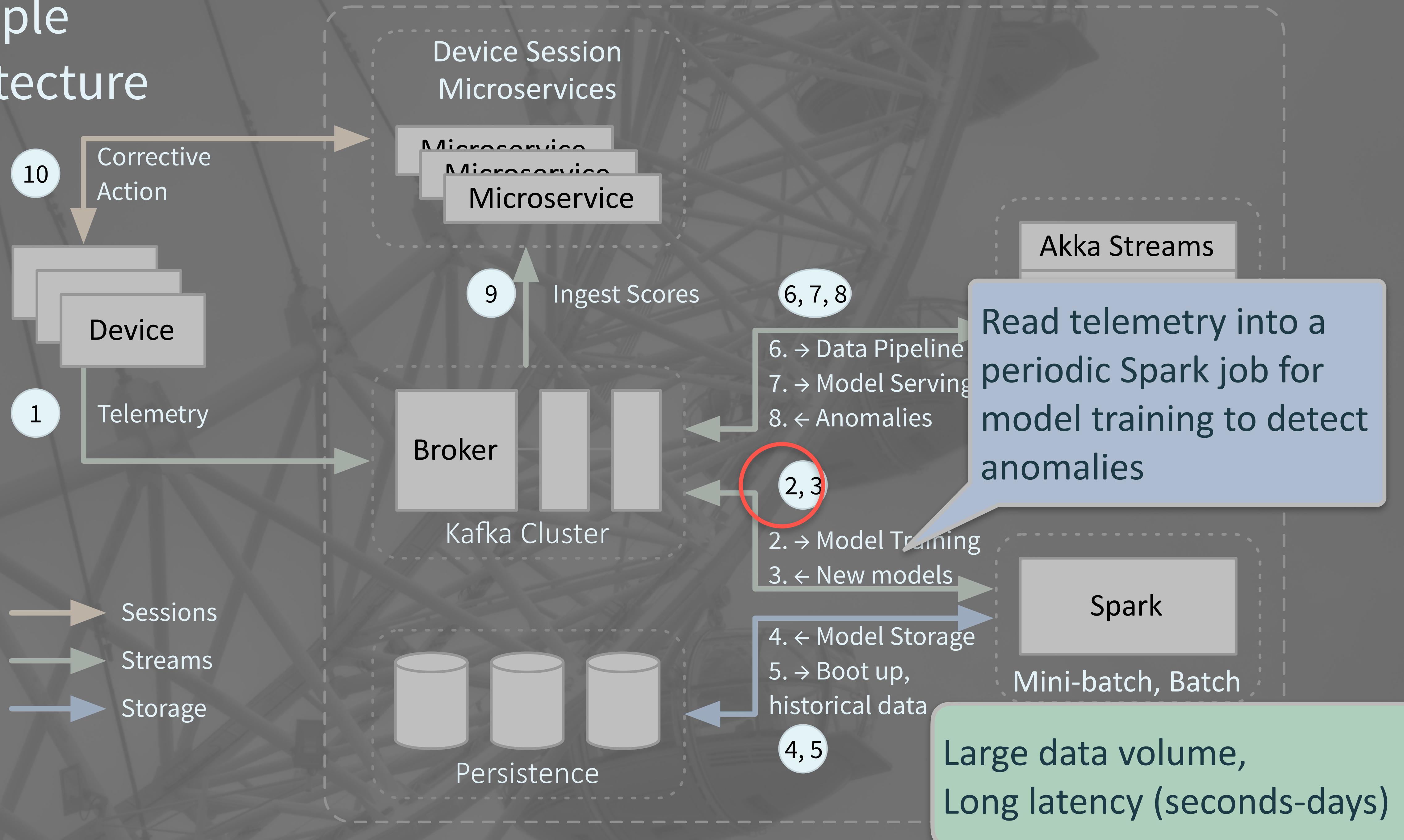
# Example Architecture



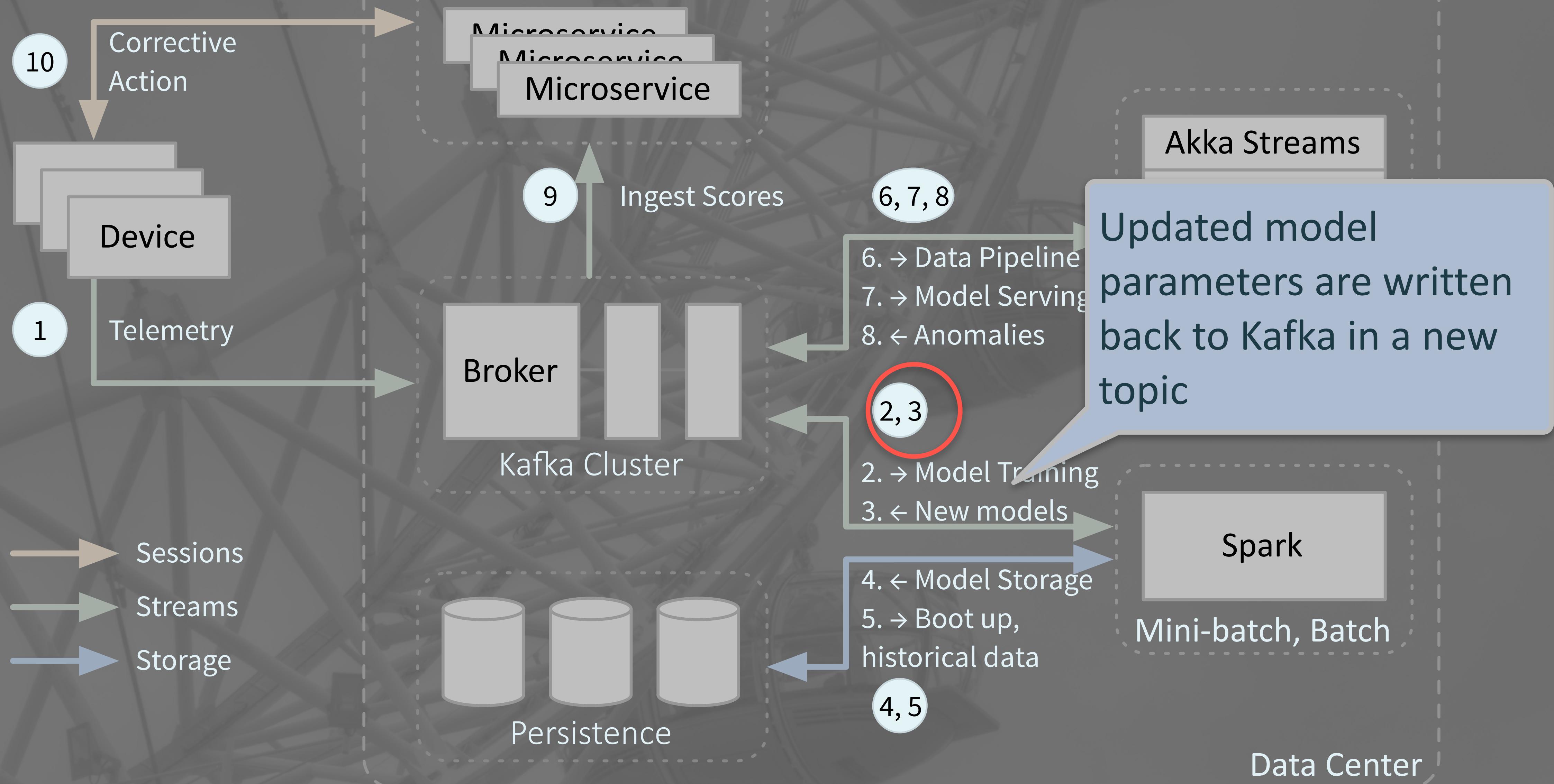
# Example Architecture



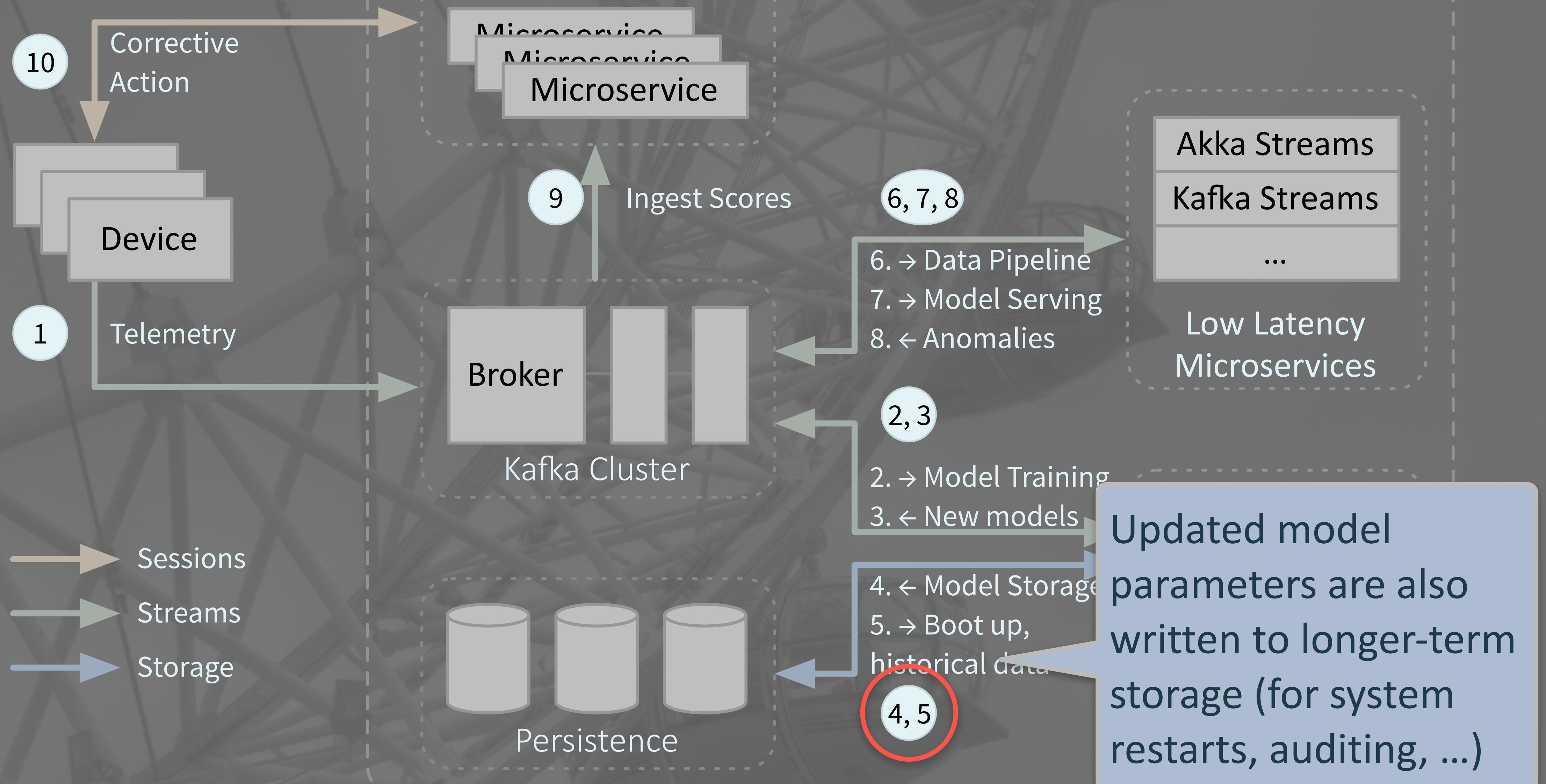
# Example Architecture



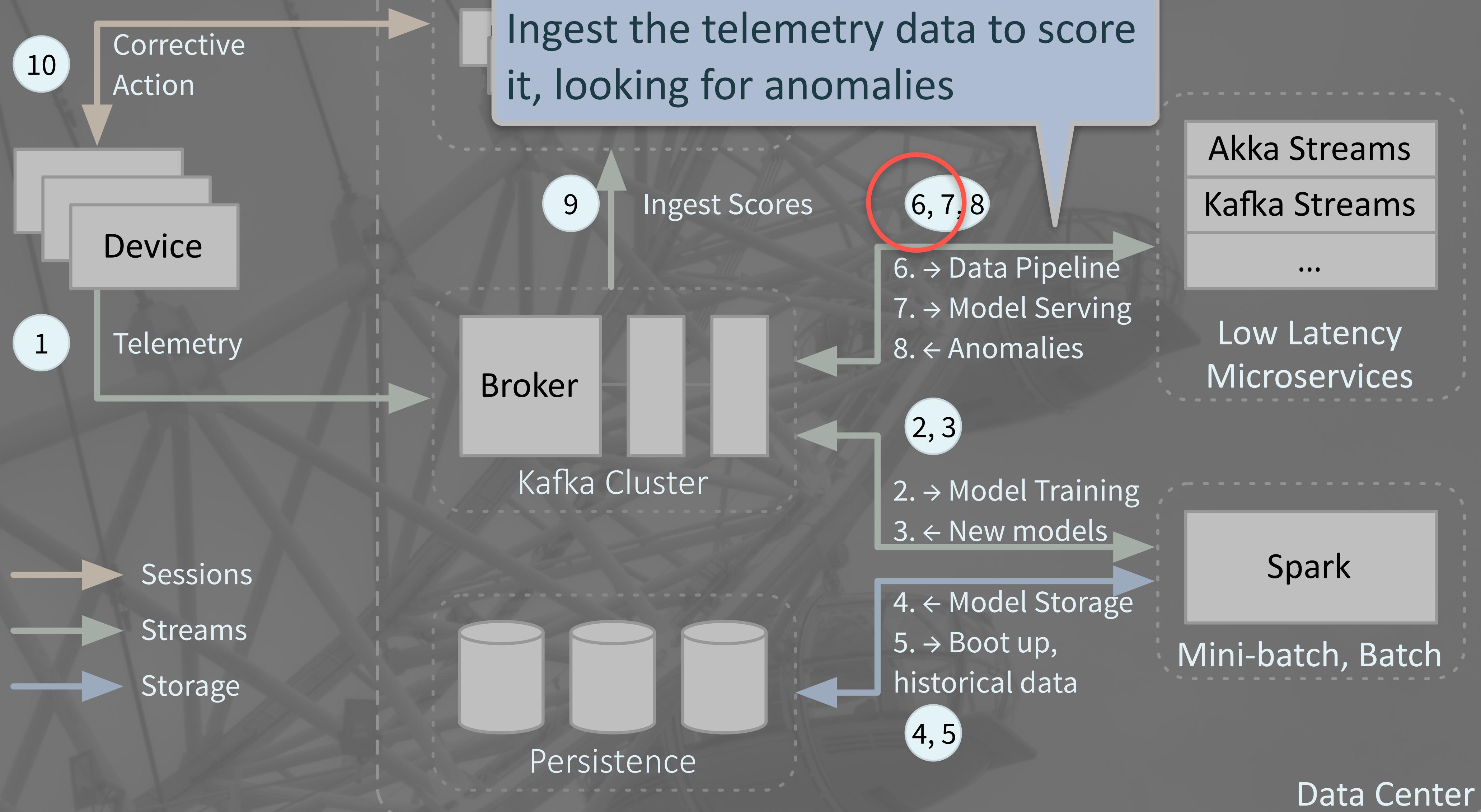
# Example Architecture



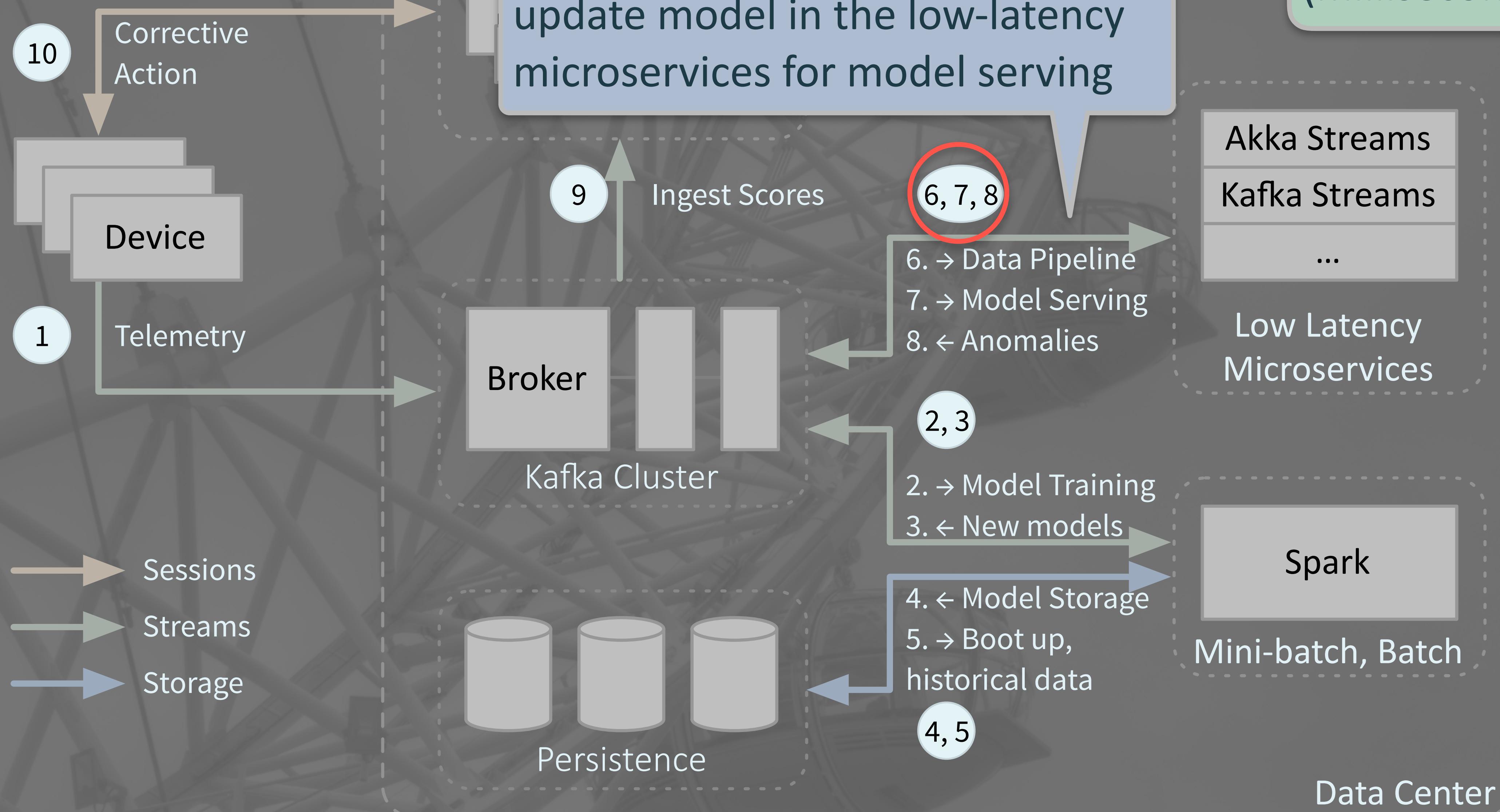
# Example Architecture



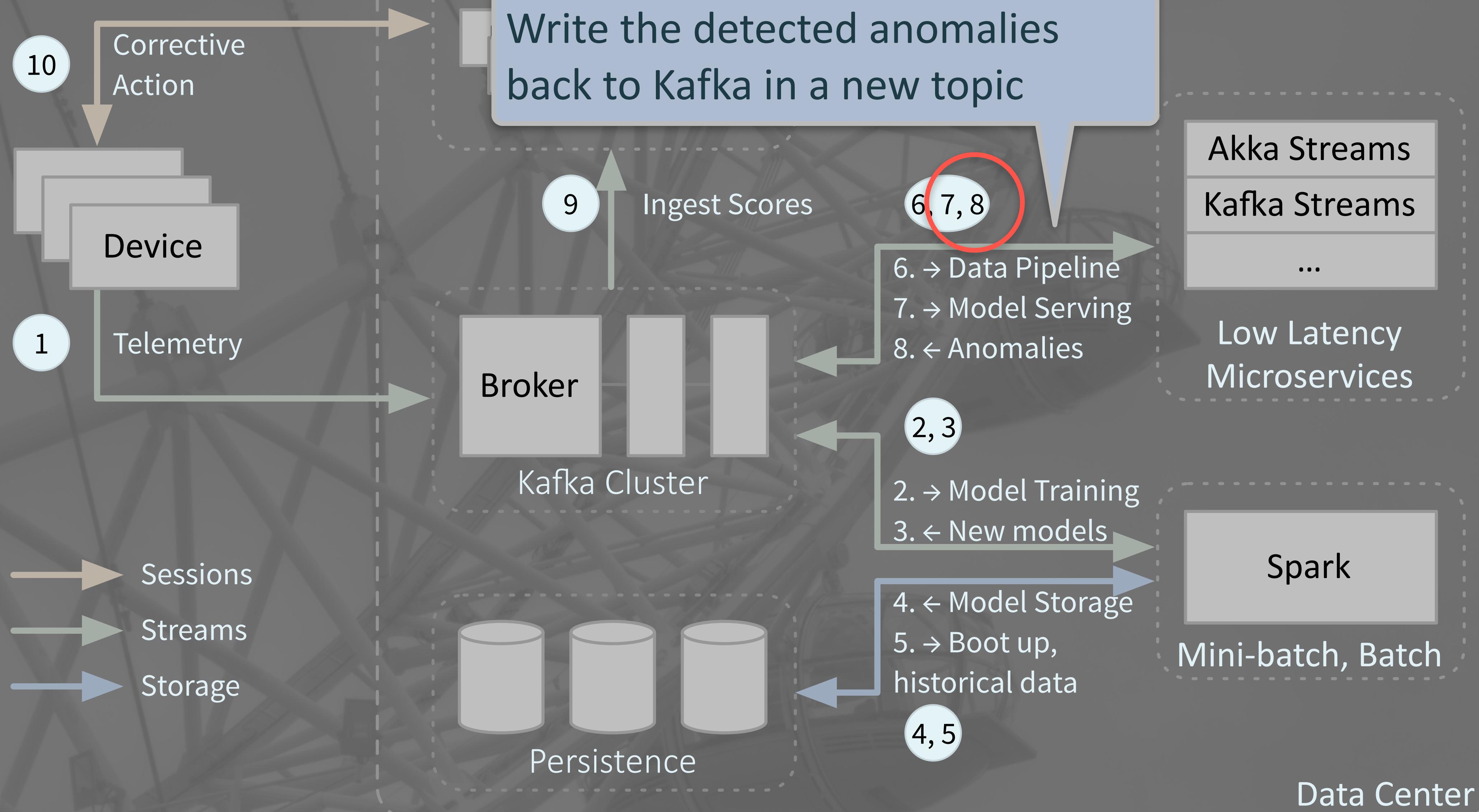
# Example Architecture



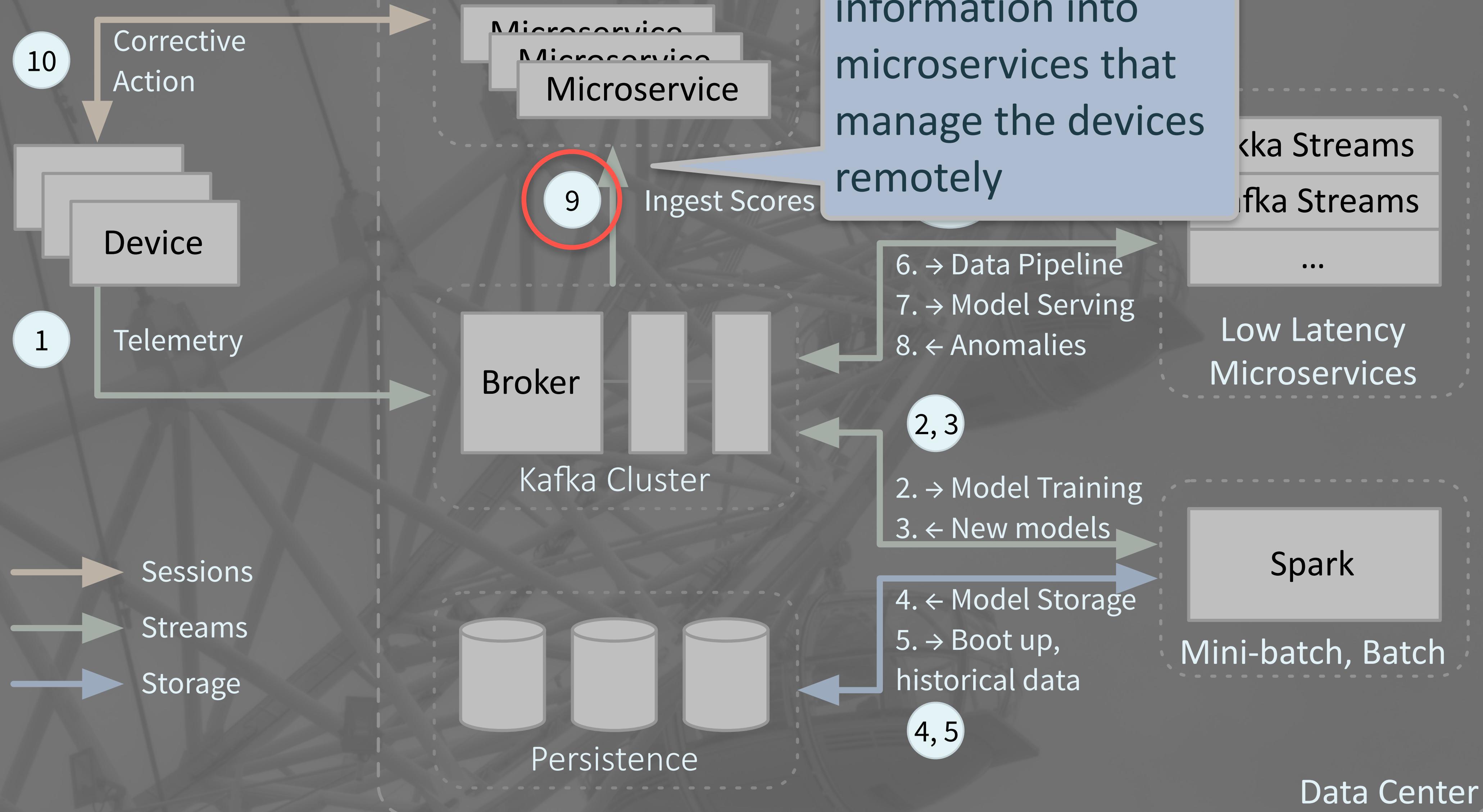
# Example Architecture



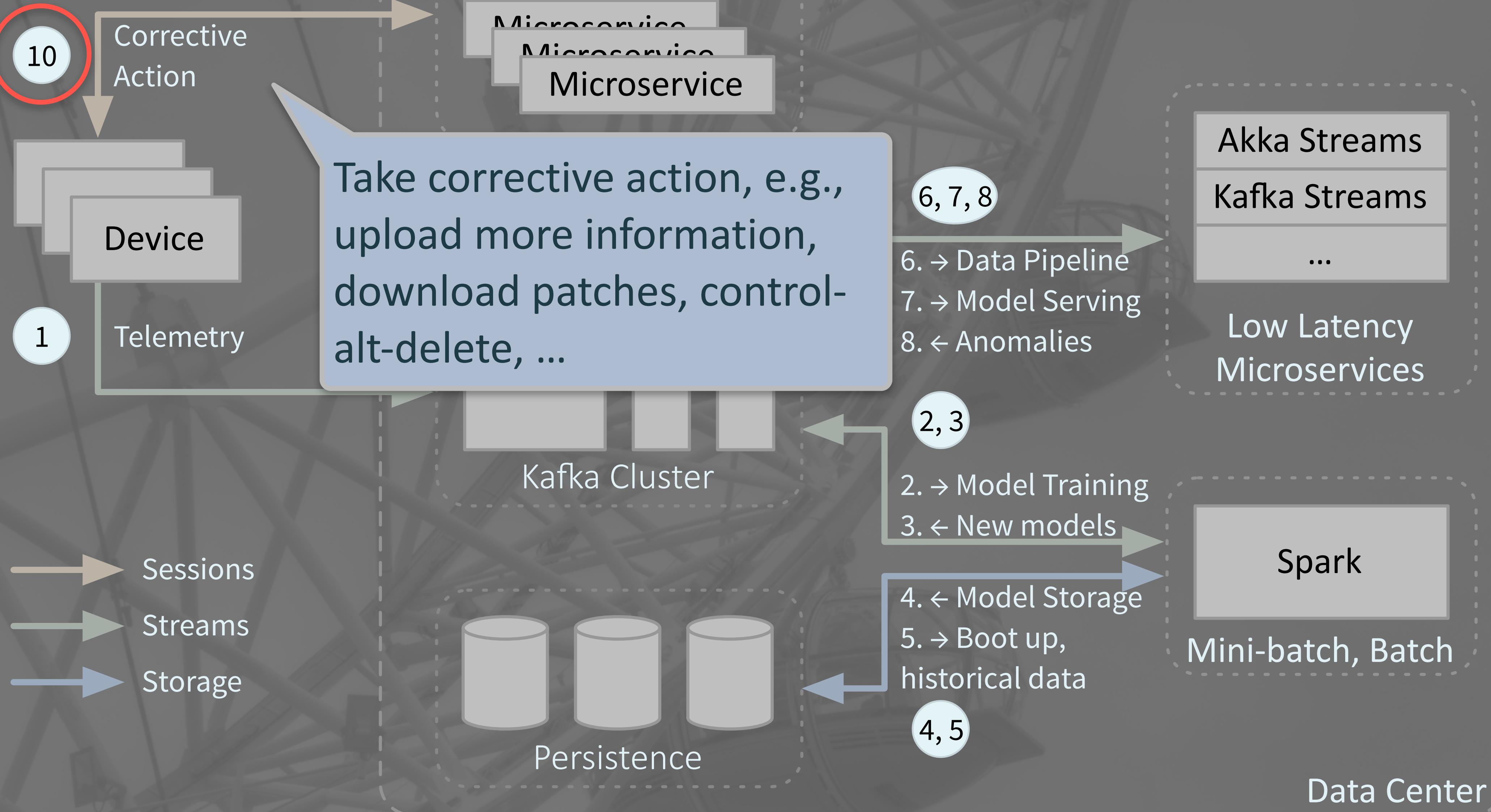
# Example Architecture



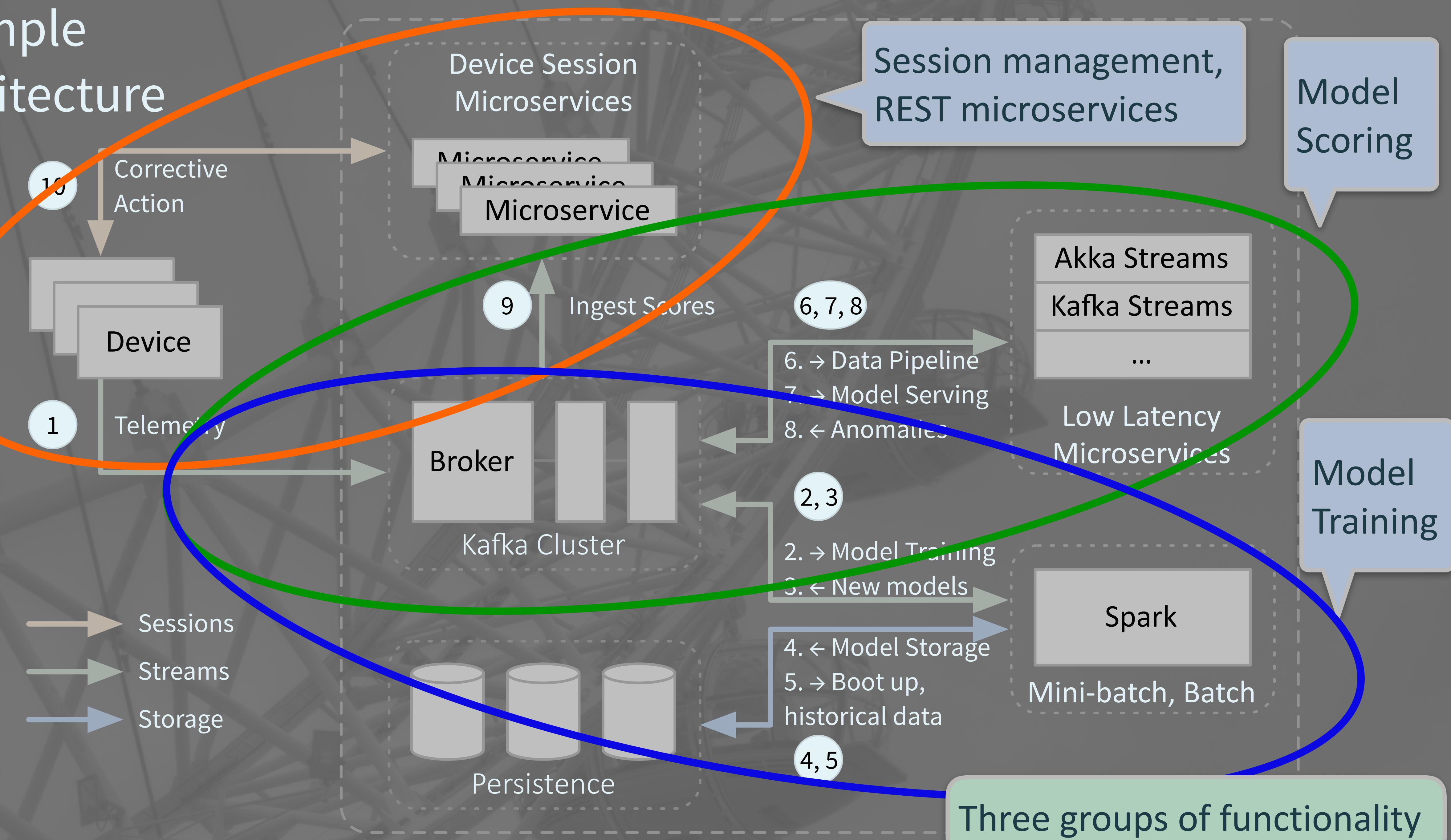
# Example Architecture



# Example Architecture

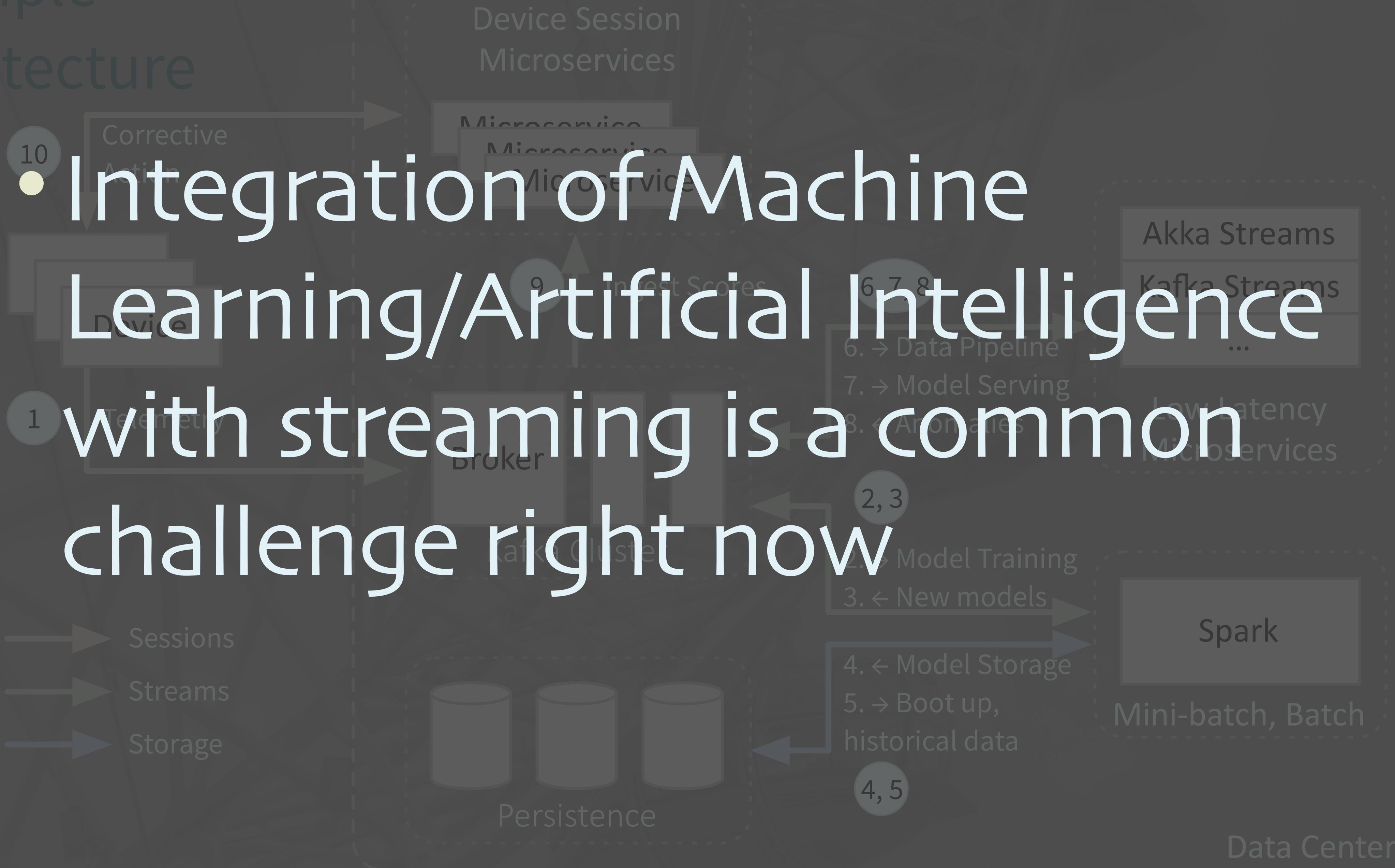


# Example Architecture



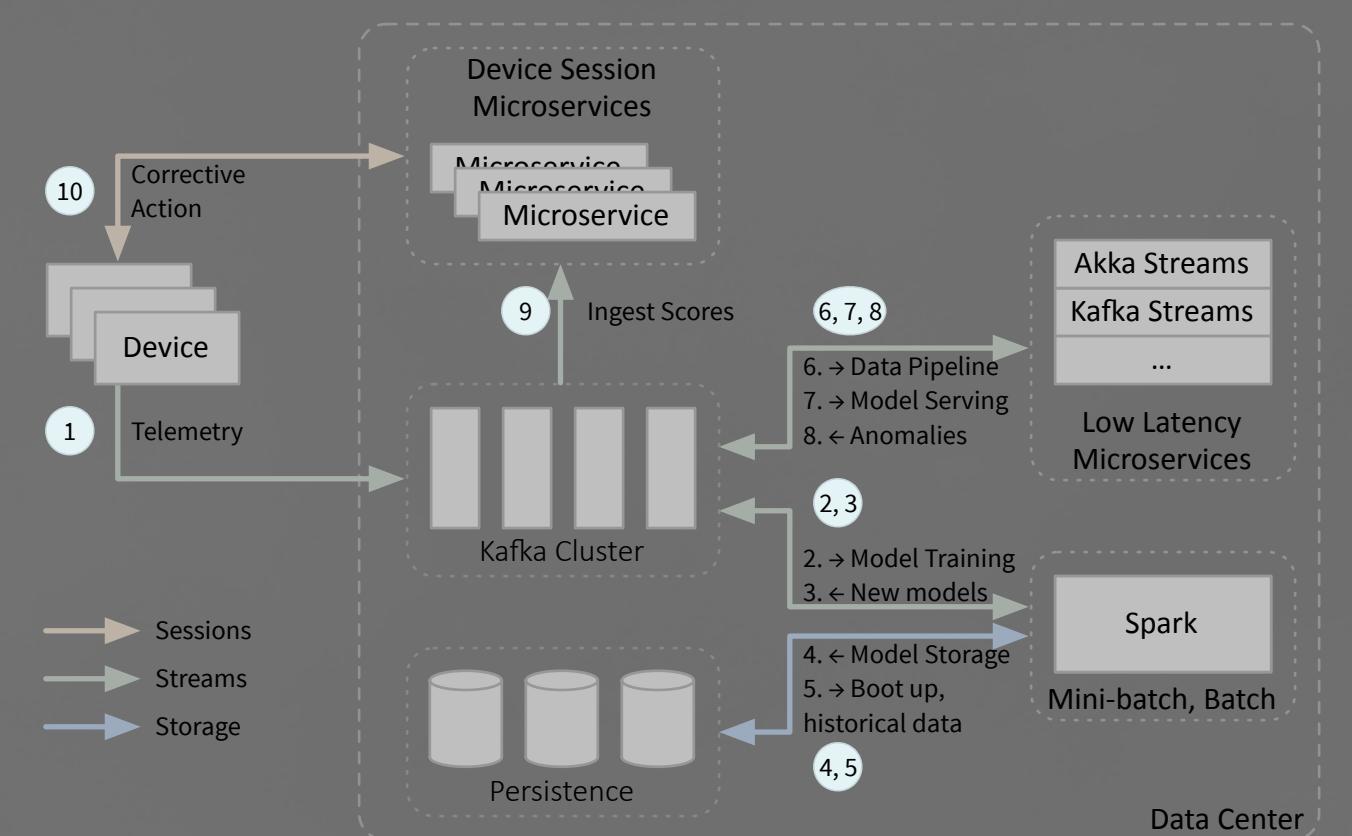
# Example Architecture

• Integration of Machine Learning/Artificial Intelligence with streaming is a common challenge right now



# Challenges

- Network overhead for telemetry ingestion too high?
- Model serving latency too long?
- Datacenter unavailable?
- Idea: Serve models on the device!

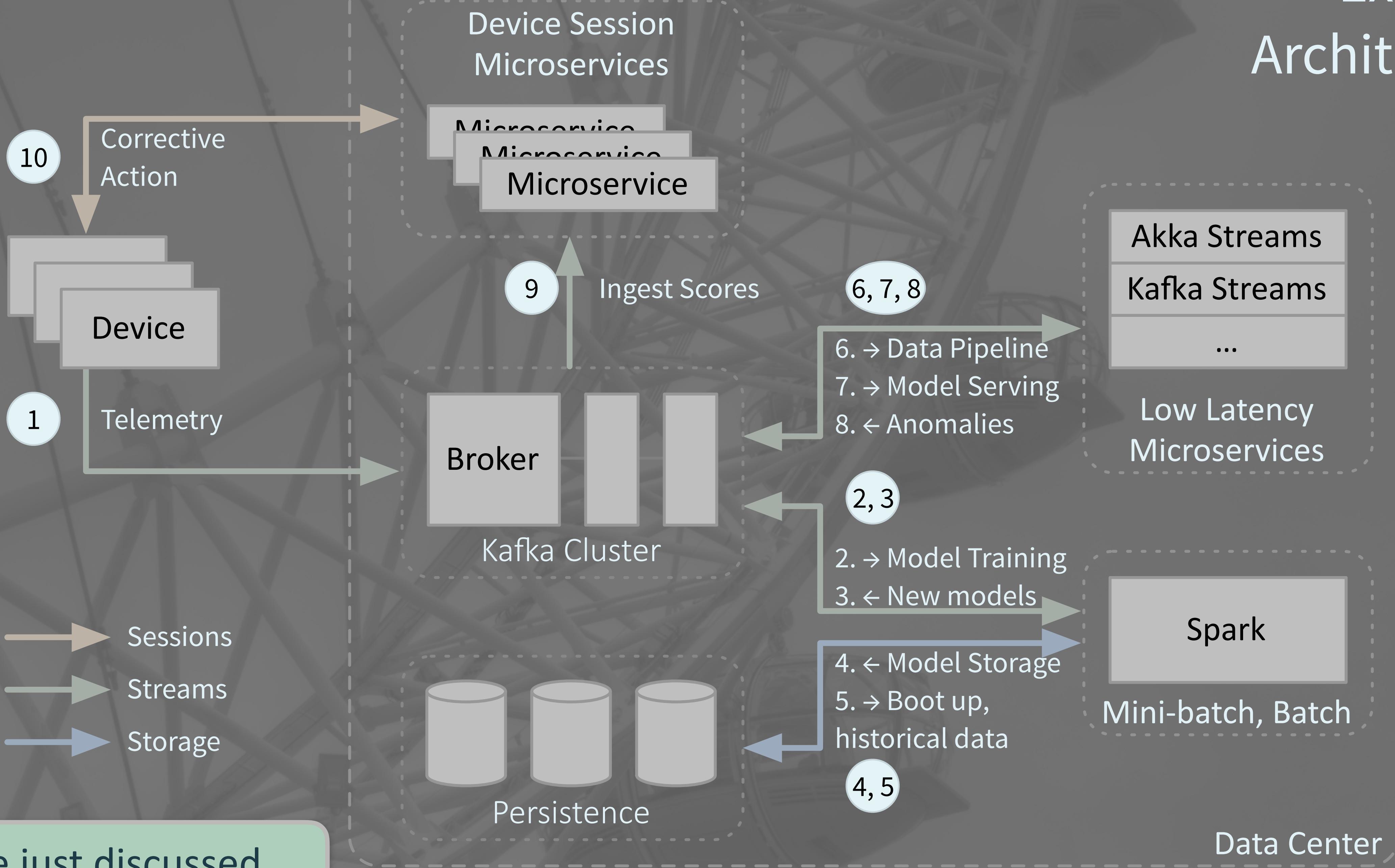


# Internet of Things

- Real-time consumer and industrial device and supply chain management at scale

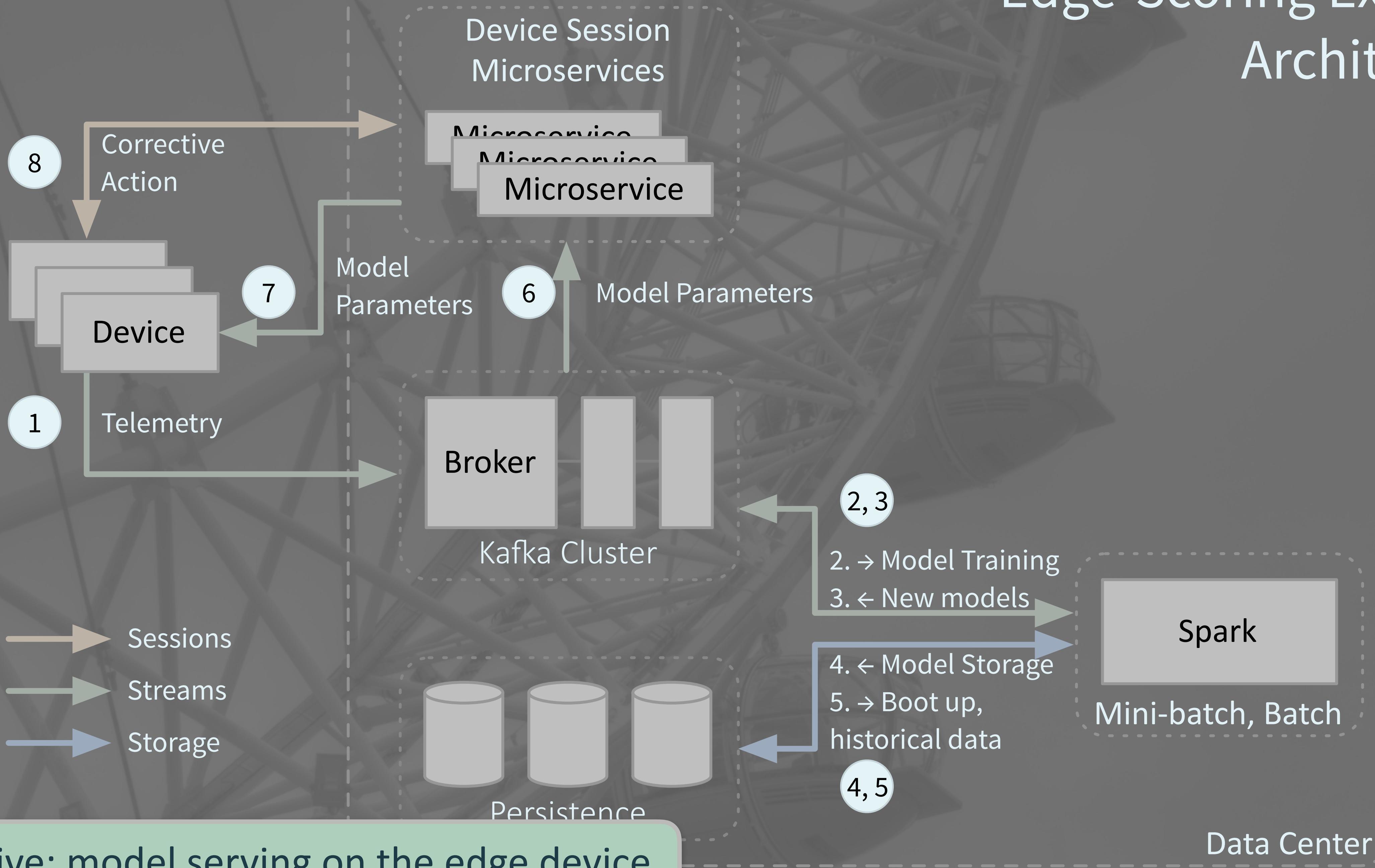


# Example Architecture

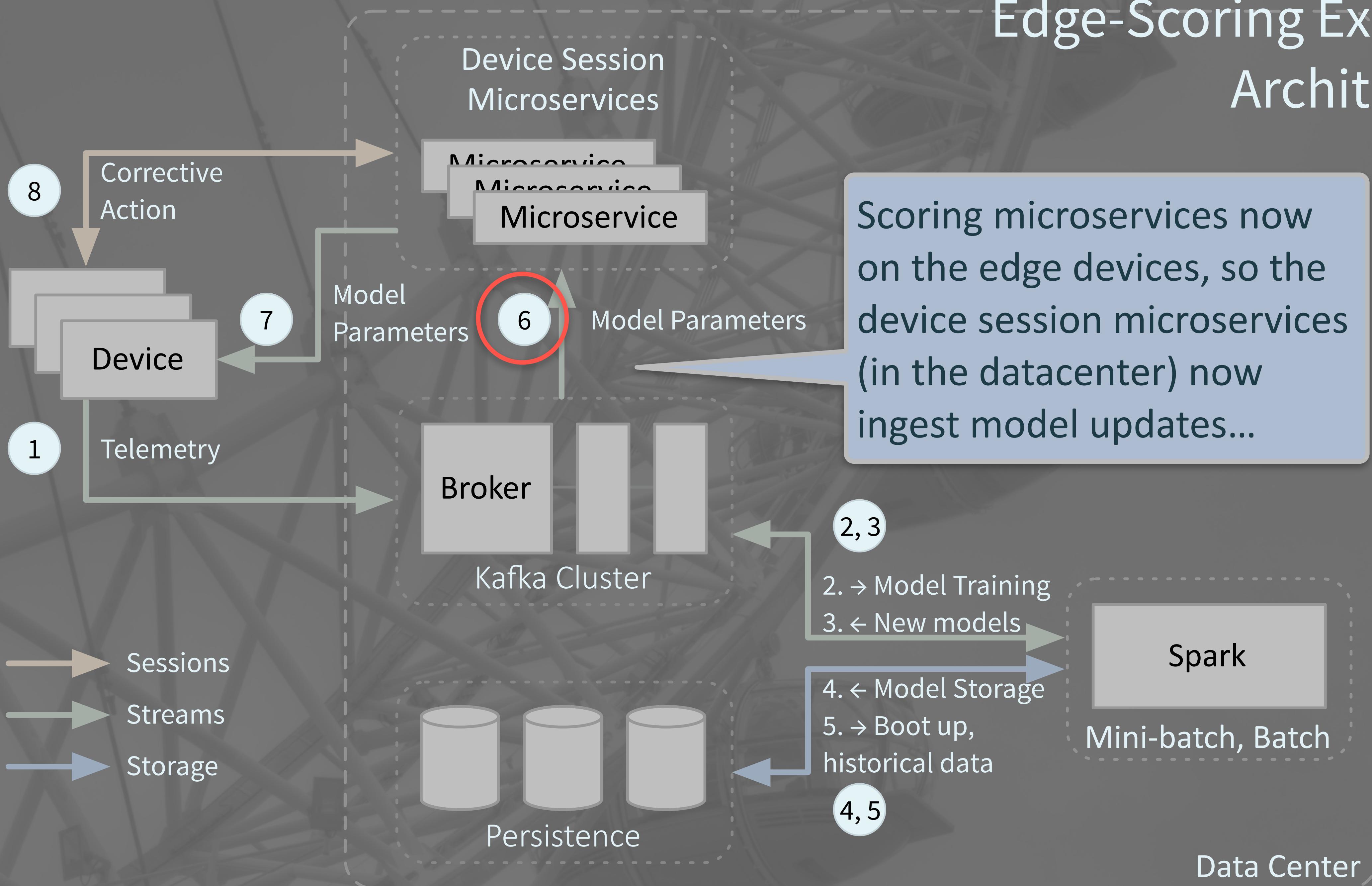


What we just discussed...

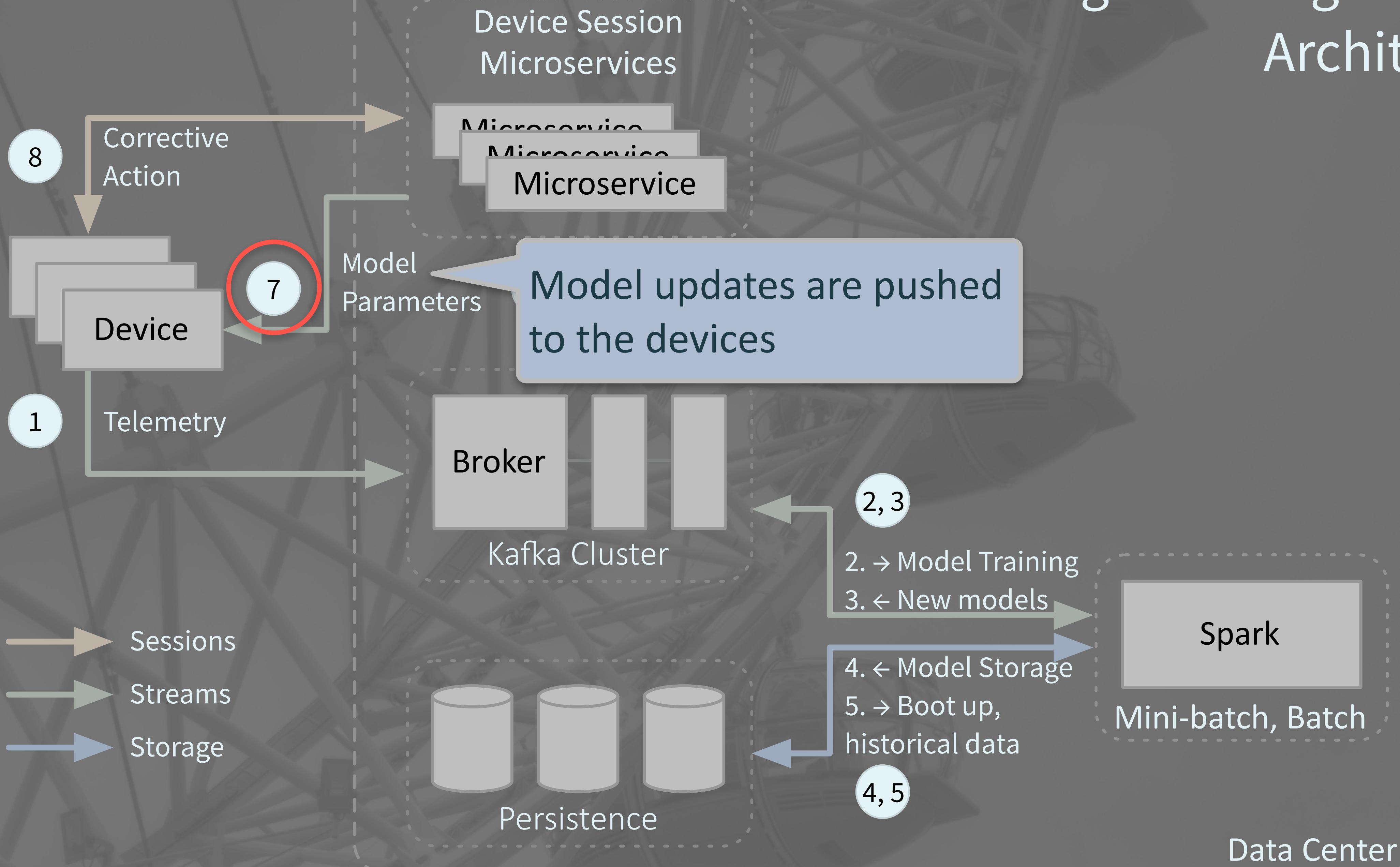
# Edge-Scoring Example Architecture



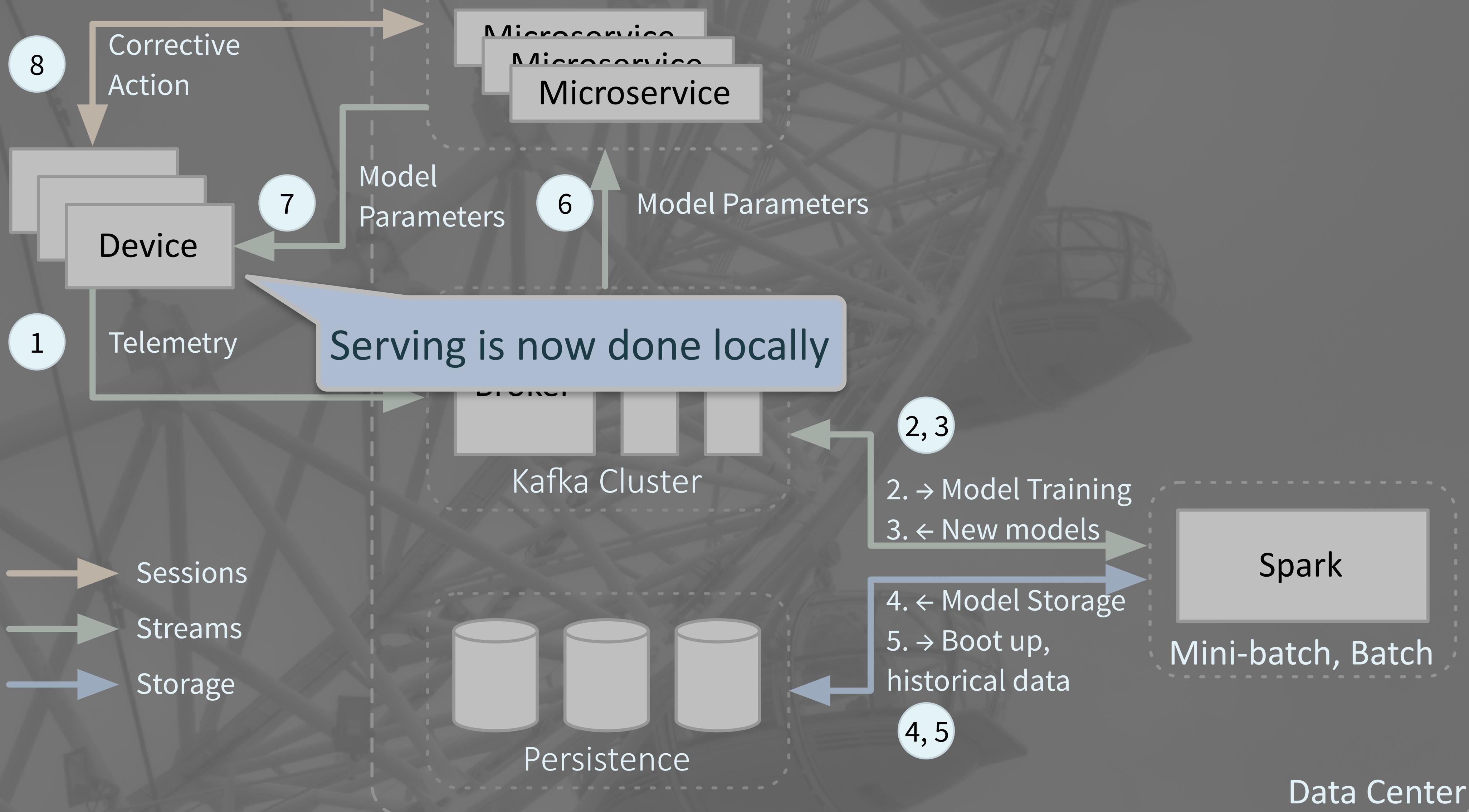
# Edge-Scoring Example Architecture



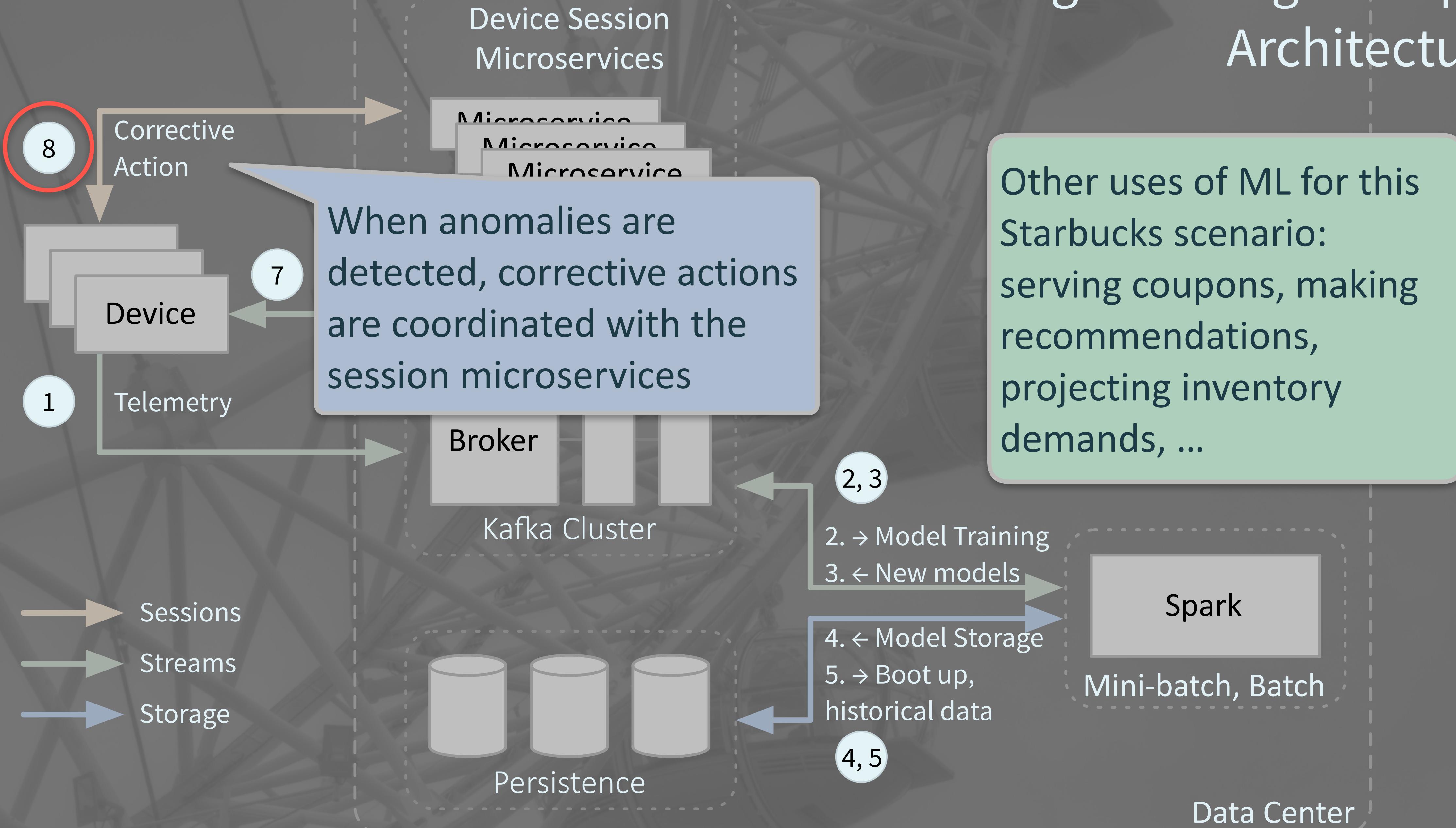
# Edge-Scoring Example Architecture



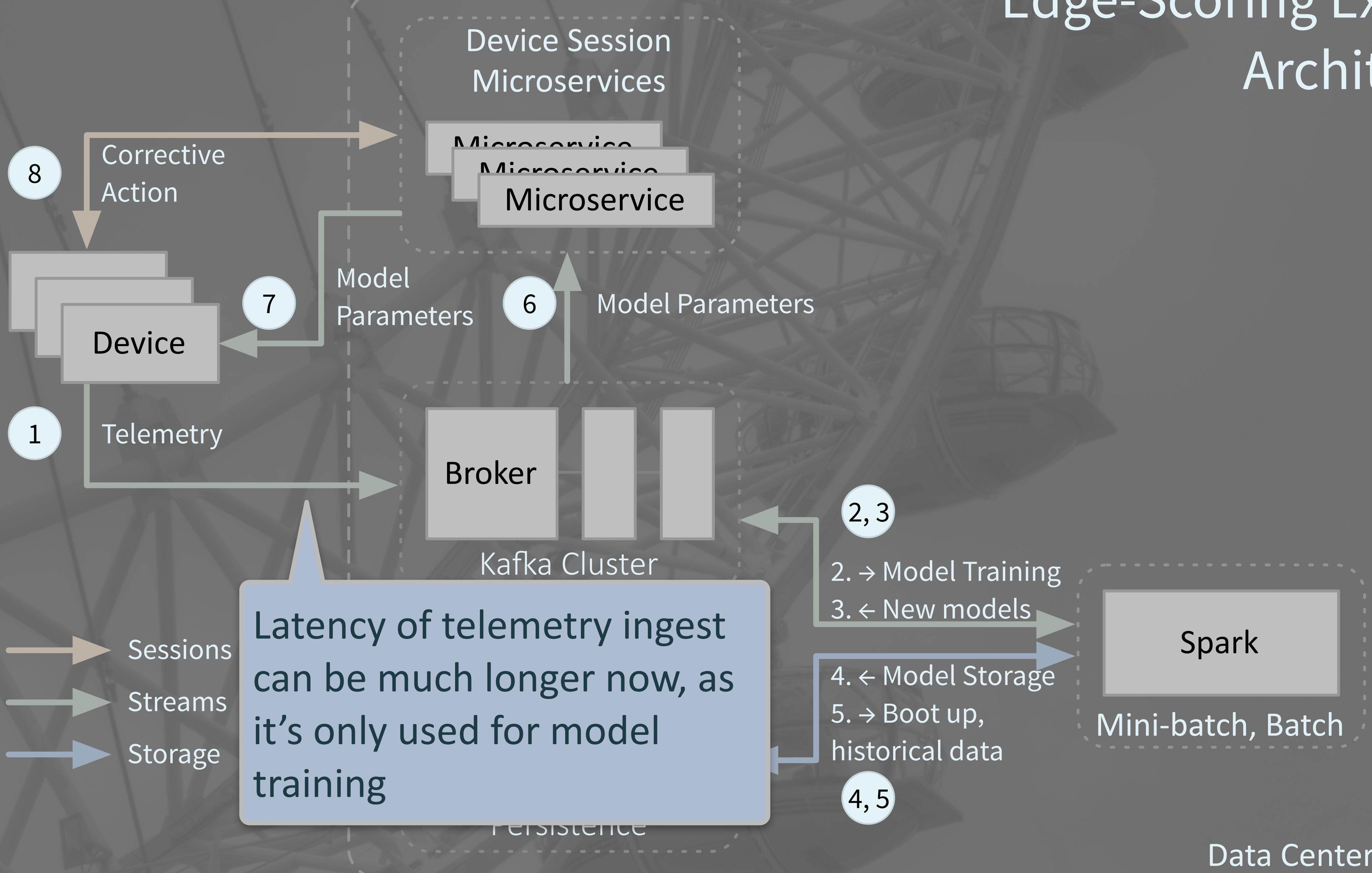
# Edge-Scoring Example Architecture



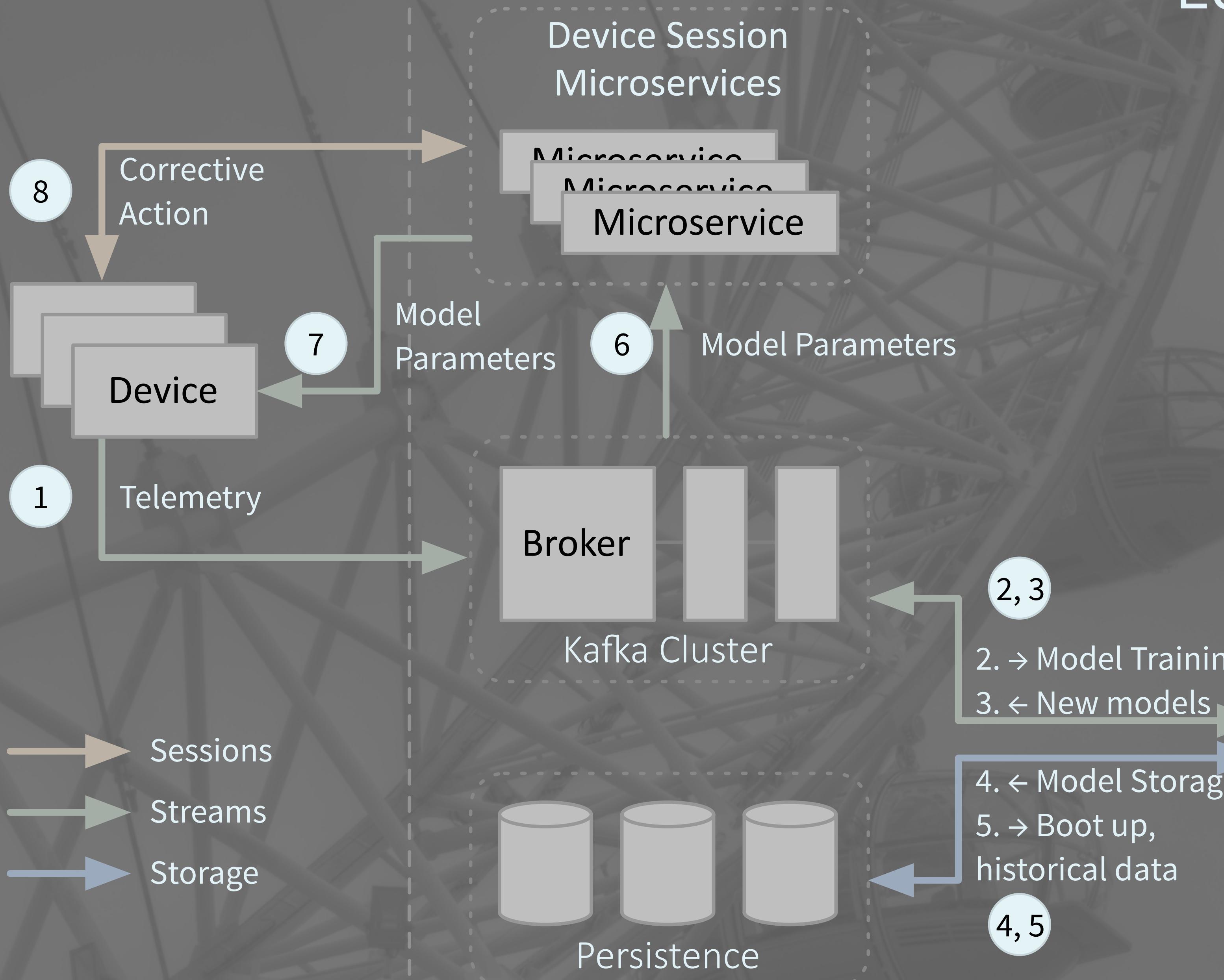
# Edge-Scoring Example Architecture



# Edge-Scoring Example Architecture



# Edge-Scoring Example Architecture



Recap: Edge Serving

# Fas

Batch changed to streaming  
for competitive advantage

## Predictive Analytics

Apply ML models to large volumes of device data to pre-empt failures / outages



**Hewlett Packard  
Enterprise**

## IoT

Real-time consumer and industrial Device and Supply Chain management at scale



## Real-time Personalization

Real-time marketing based on behavior, location, inventory levels, product promotions, etc.



RoyalCaribbean  
INTERNATIONAL®

## Real-time Financial Processes

Drive better business outcomes through real-time risk, fraud detection, compliance, audit, governance, etc.



# Technology Choices



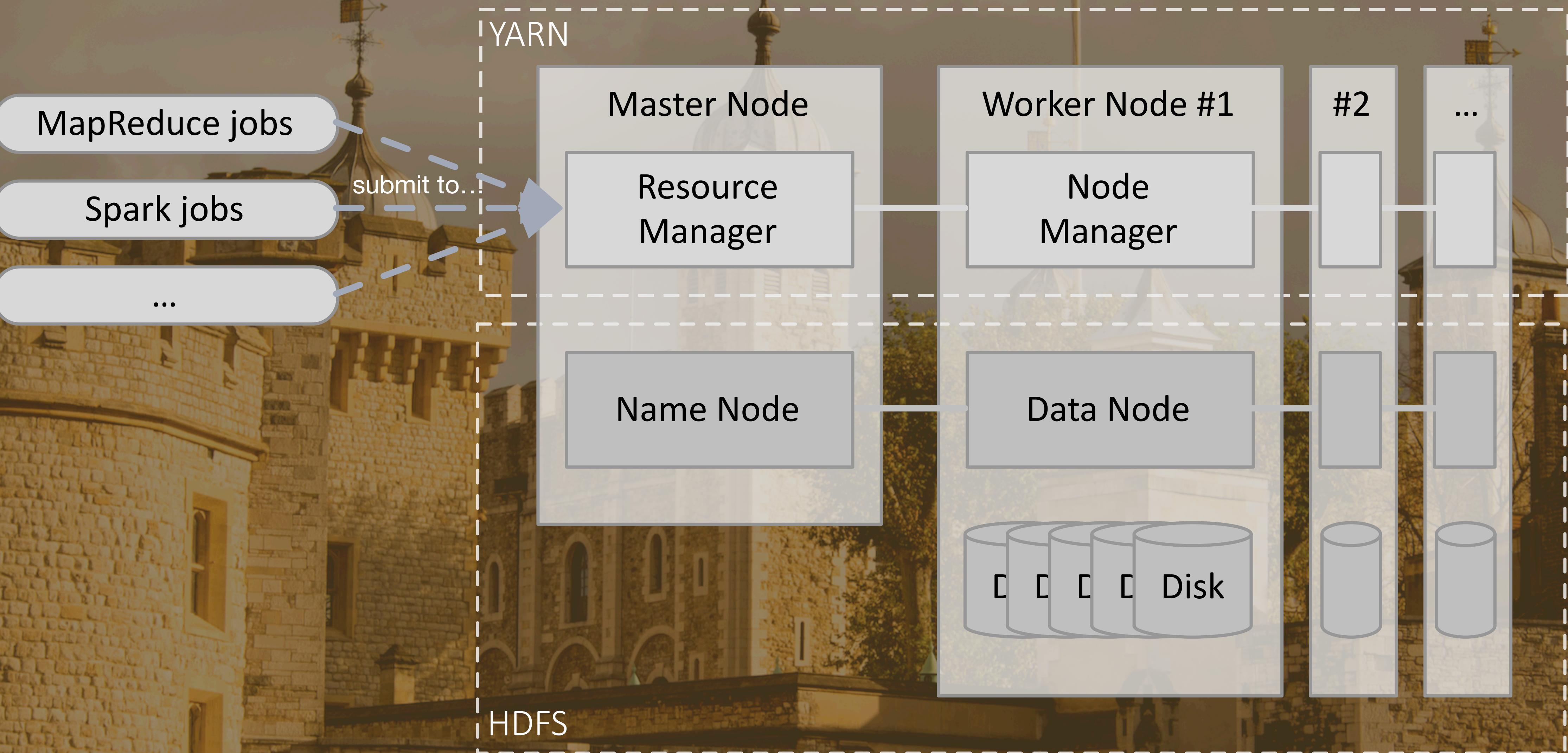
- 
- More than “faster” Hadoop...
  - New architectures that merge data processing with microservices

# Technology Choices

# Recall Hadoop...



- Data warehouse replacement
- Historical analysis
- Interactive exploration
- Offline training of machine learning models
- ...



# Resource Management

Compute

MapReduce jobs

Spark jobs

...

submit o...

YARN

Master Node

Resource Manager

Worker Node #1

Node Manager

#2

...

Name Node

Data Node



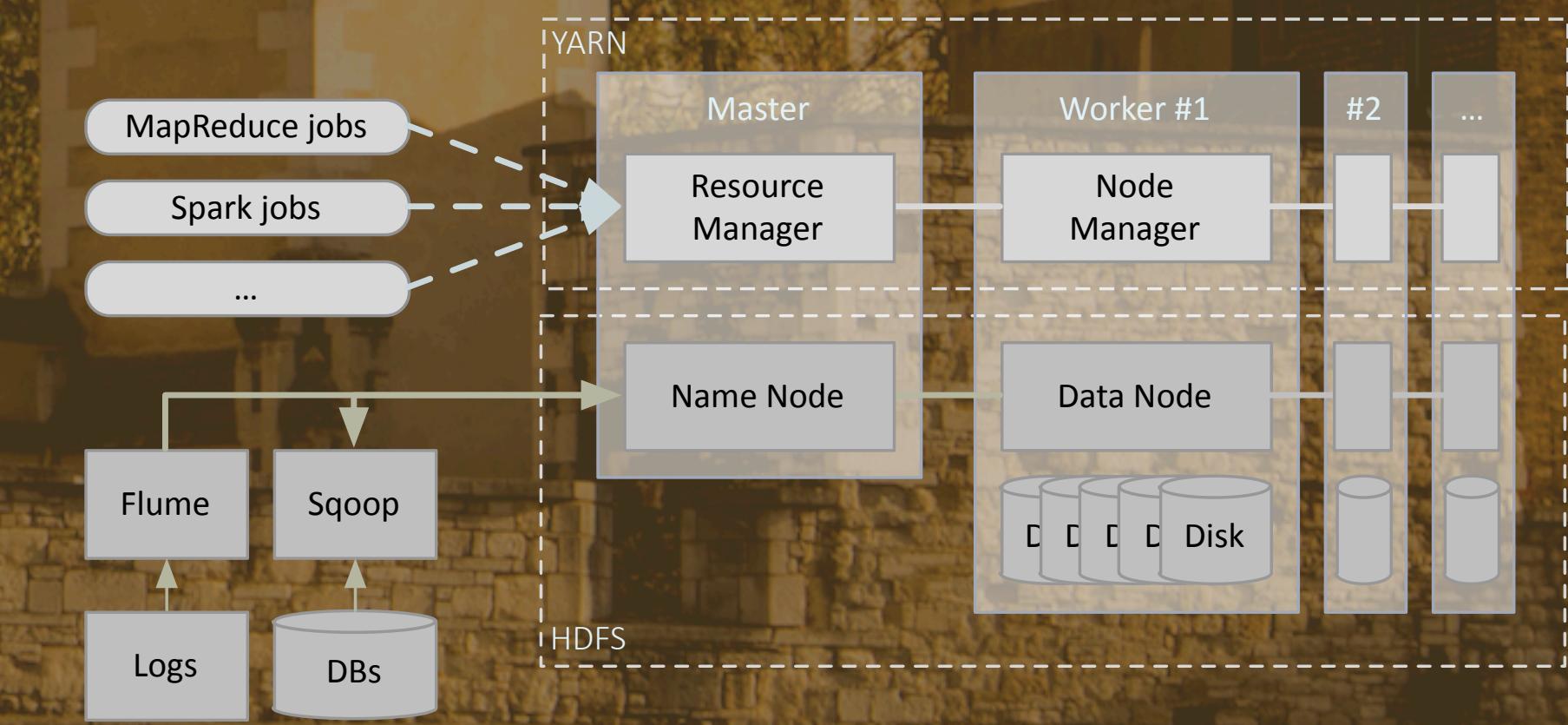
Disk

HDFS

Storage

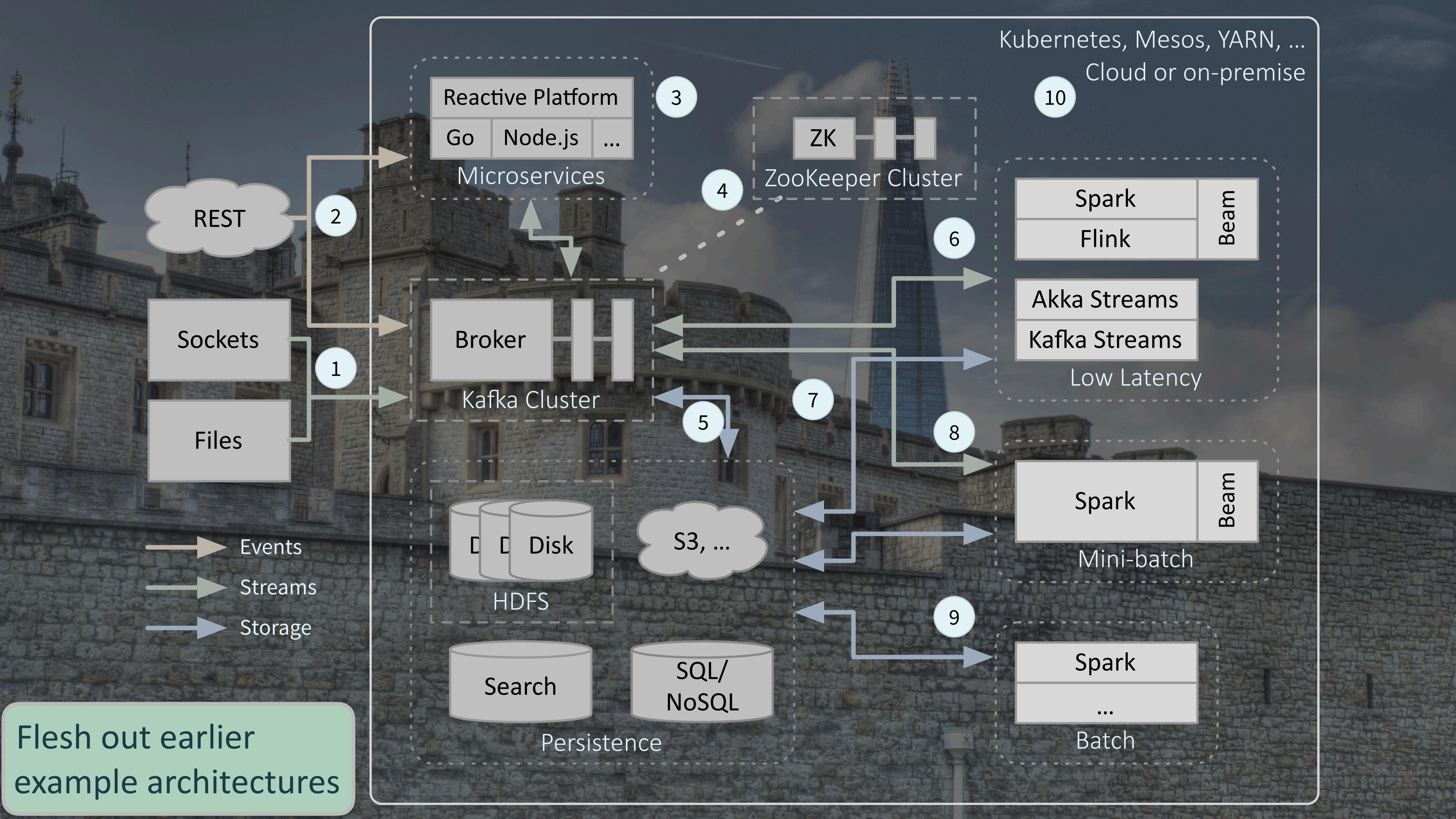
Optimized for storing lots of data *at rest*, with subsequent processing, but not optimized for data *in motion*.

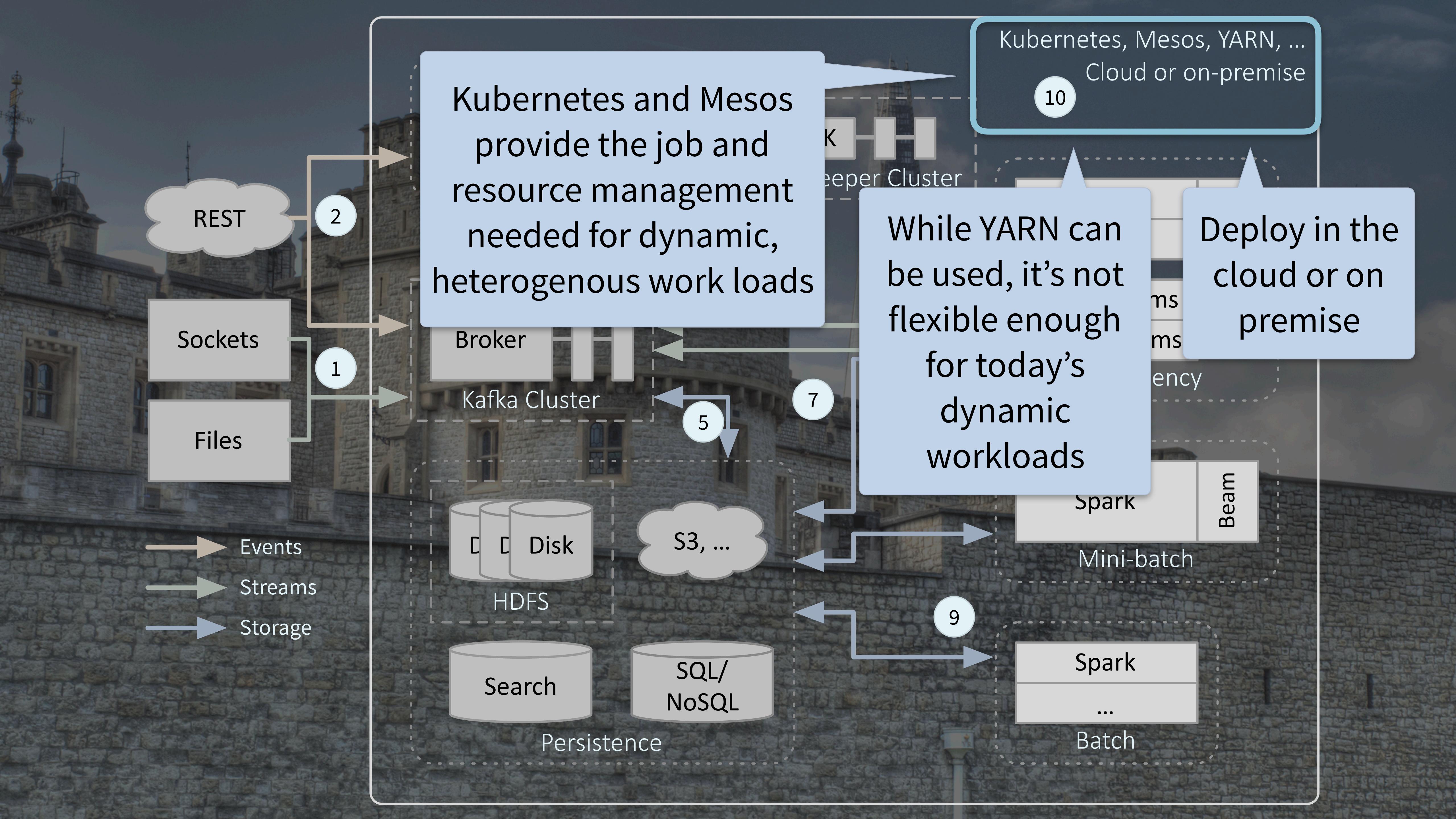
- Hadoop is ideal for batch and interactive apps
- ... but also constrained by that model

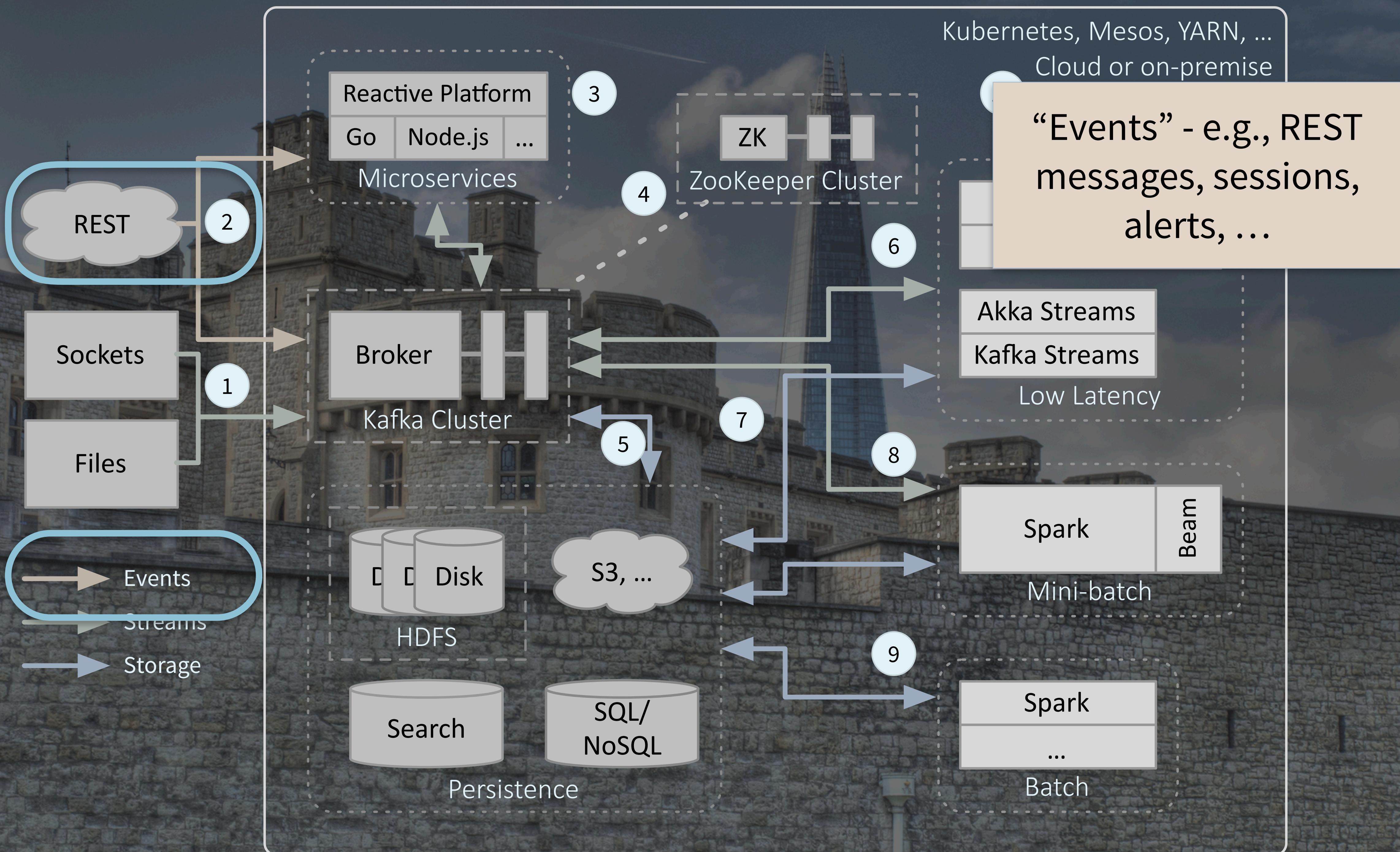


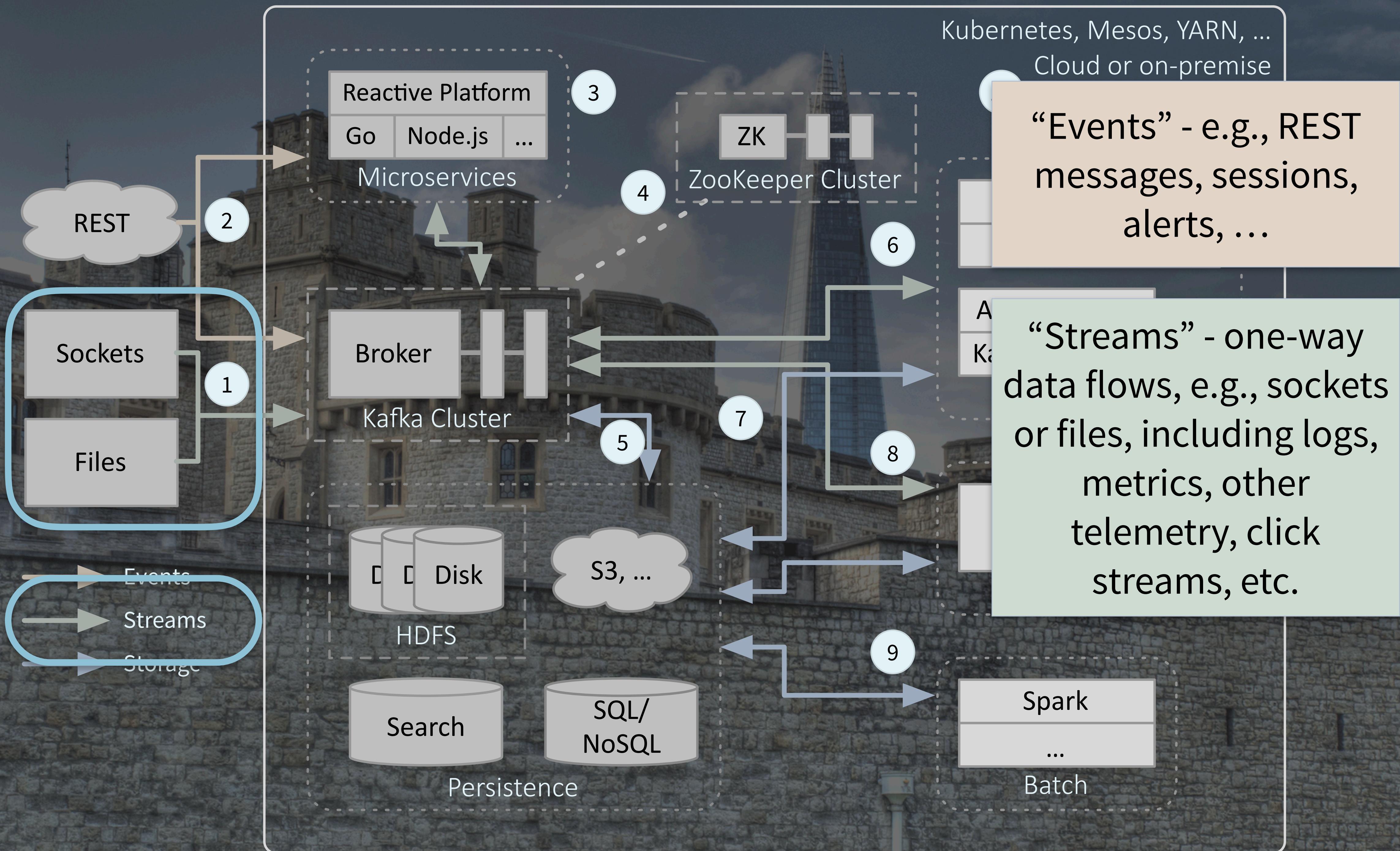
# New Fast Data Architecture

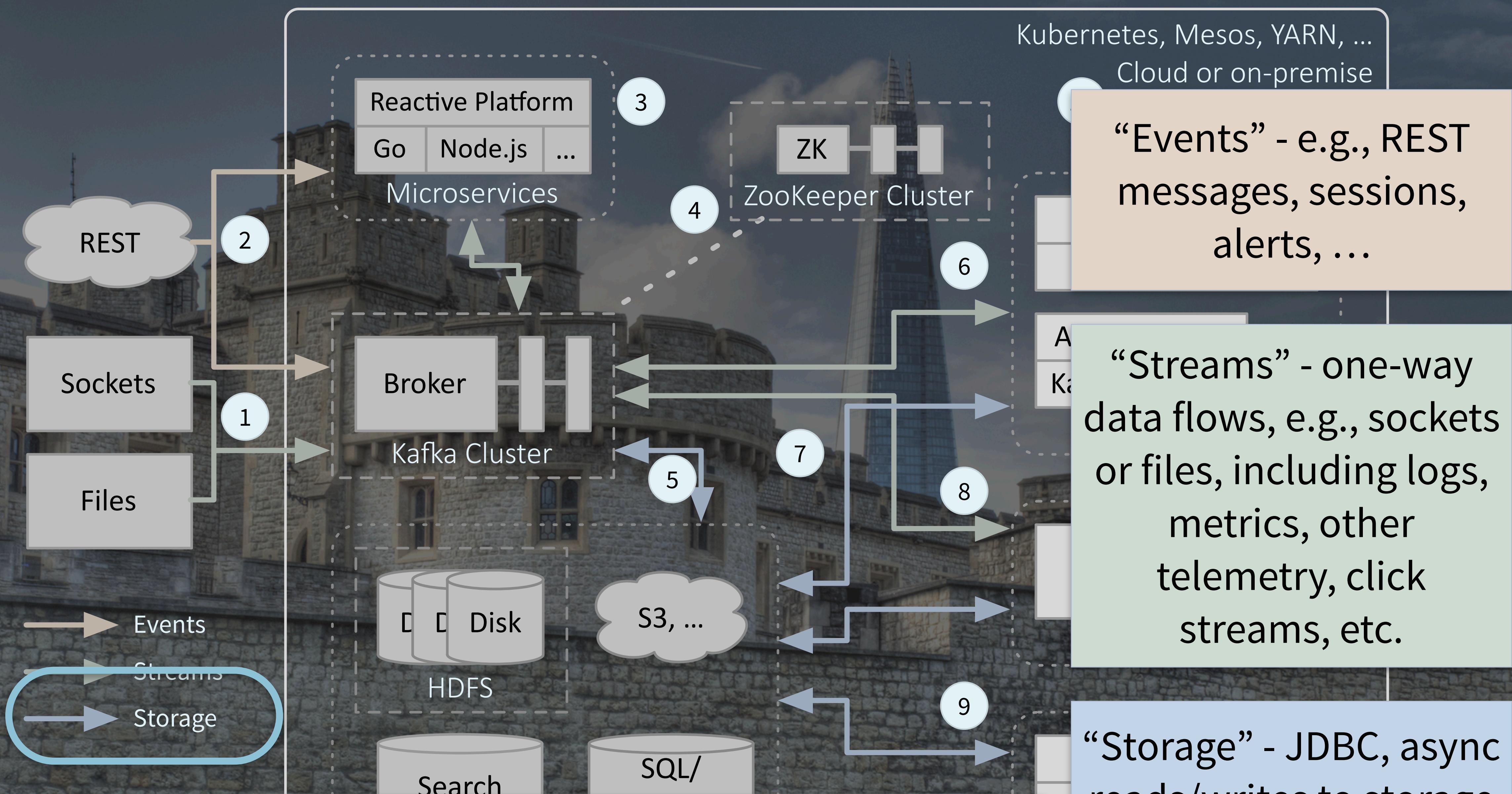












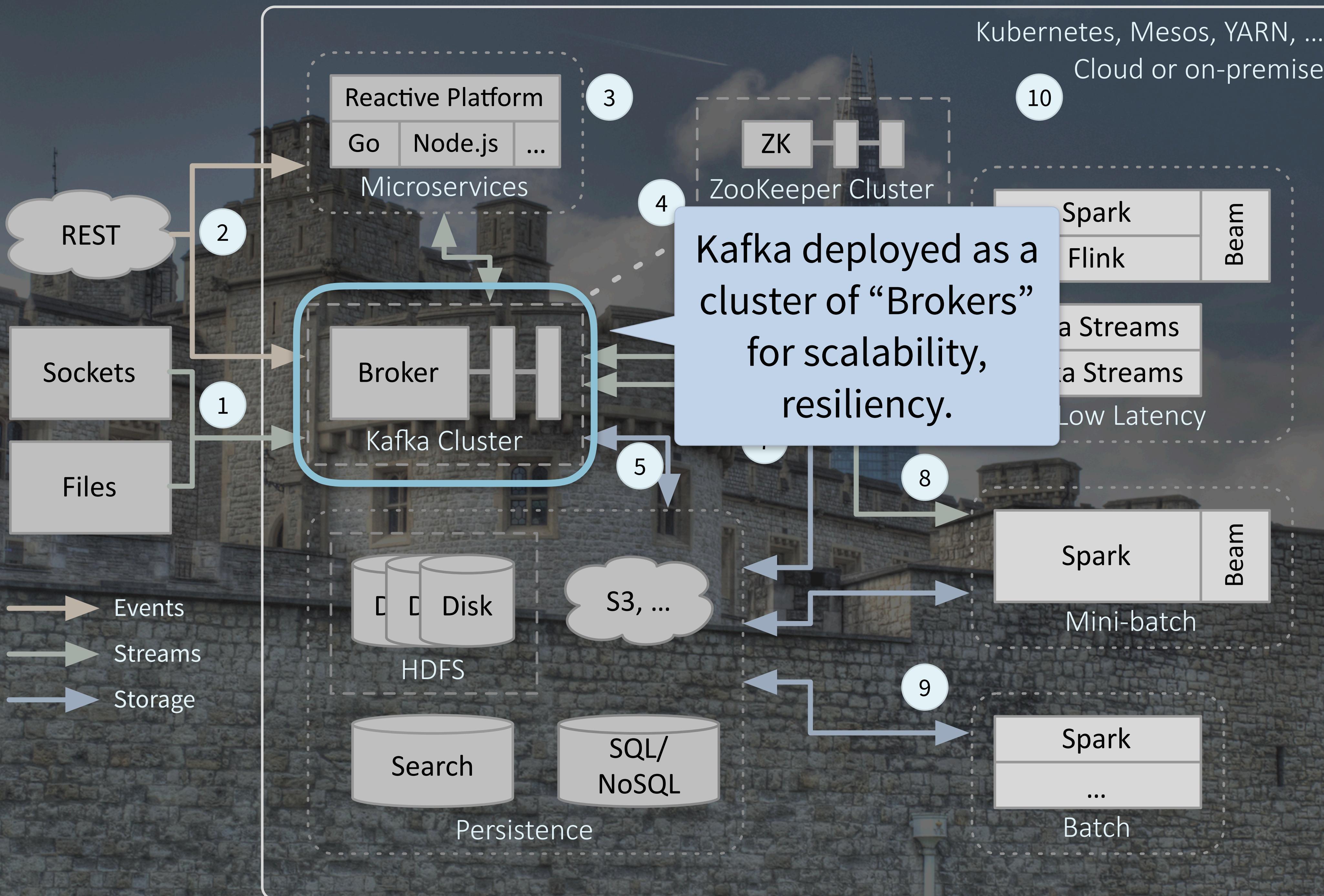
Kubernetes, Mesos, YARN, ...  
Cloud or on-premise

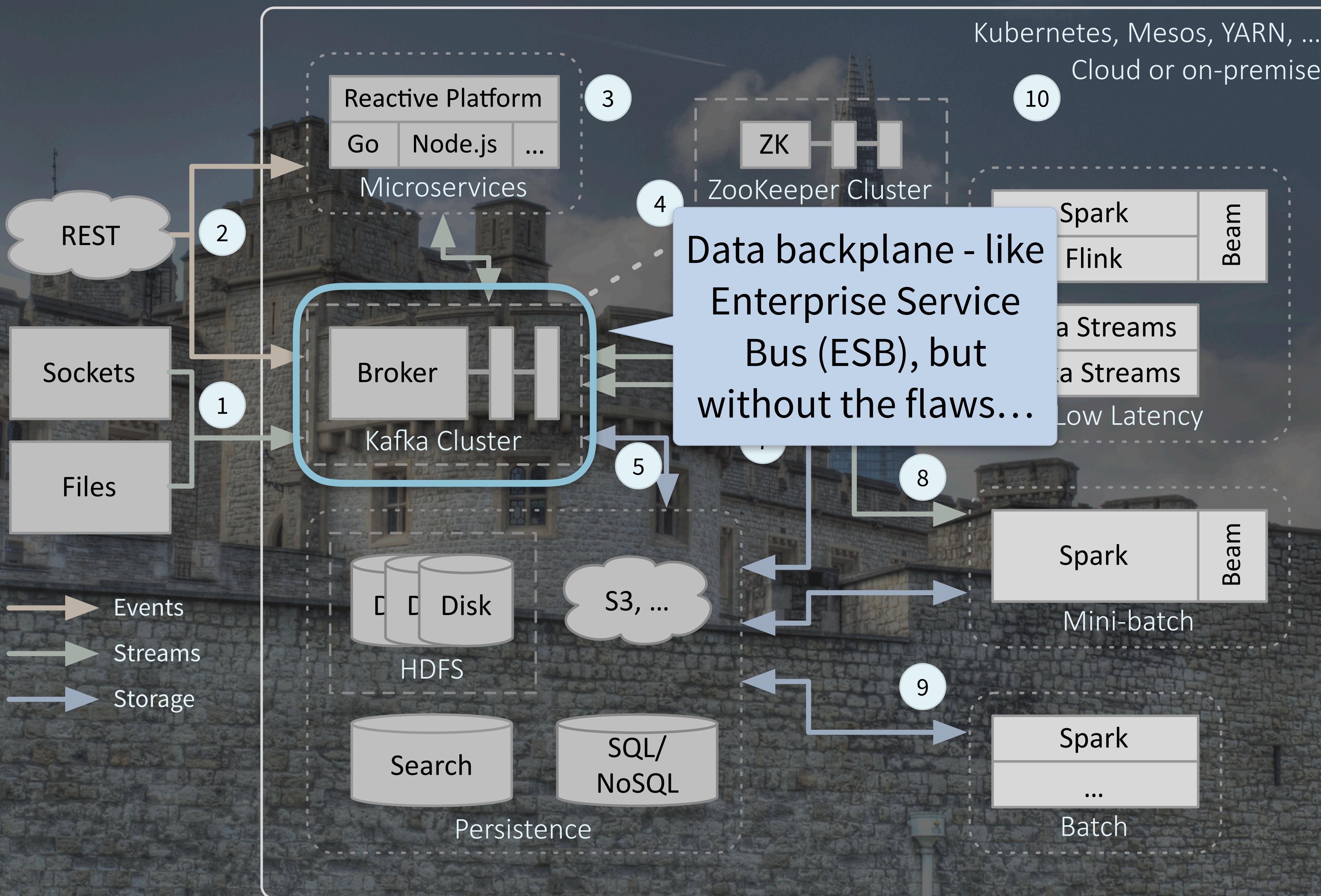
“Events” - e.g., REST messages, sessions, alerts, ...

“Streams” - one-way data flows, e.g., sockets or files, including logs, metrics, other telemetry, click streams, etc.

“Storage” - JDBC, async reads/writes to storage

Each has different volumes, velocities, latency characteristics, protocols, etc.





# Why Kafka?

Organized into topics

Topics are partitioned,  
replicated, and  
distributed

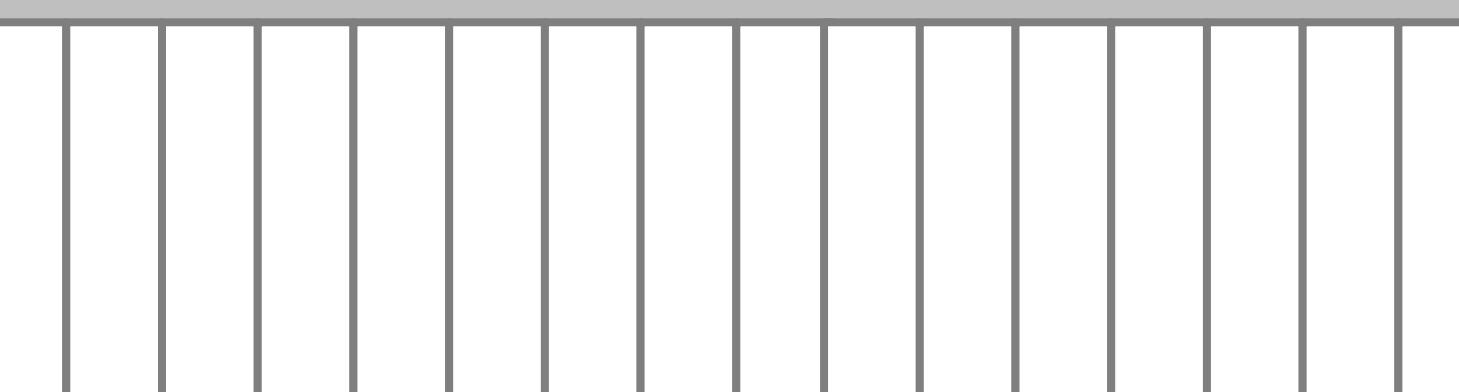
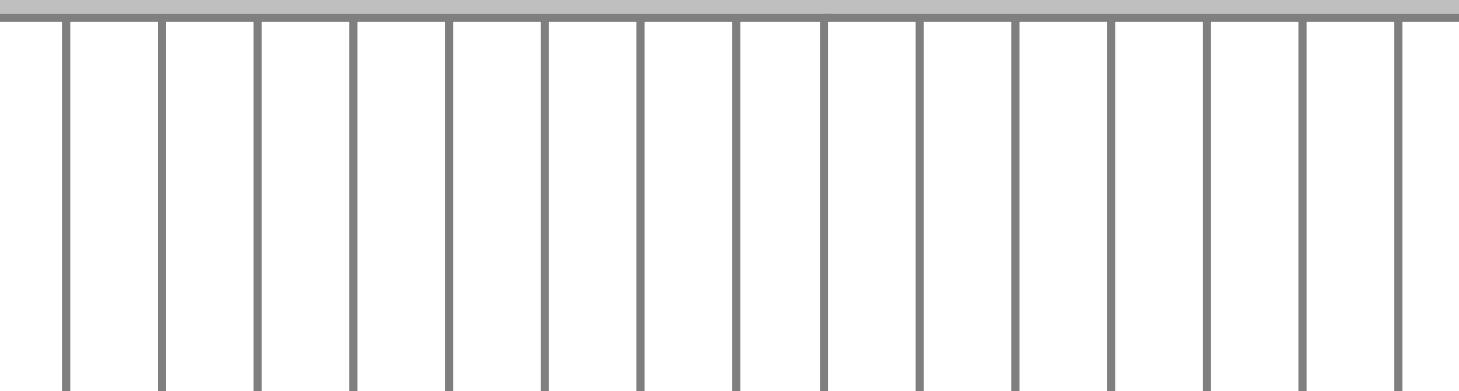
Kafka

Partition 1

Topic A

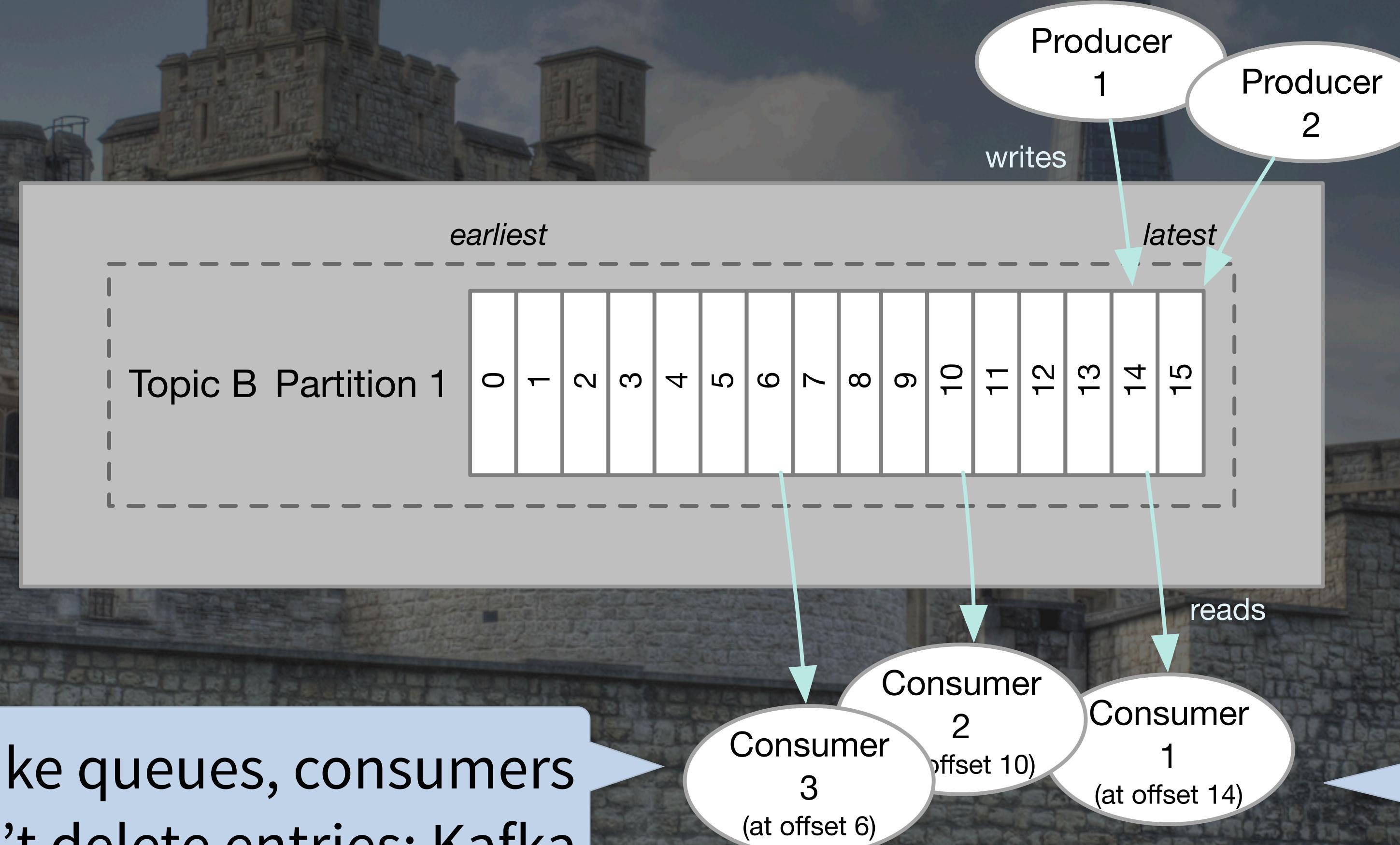
Partition 2

Topic B Partition 1



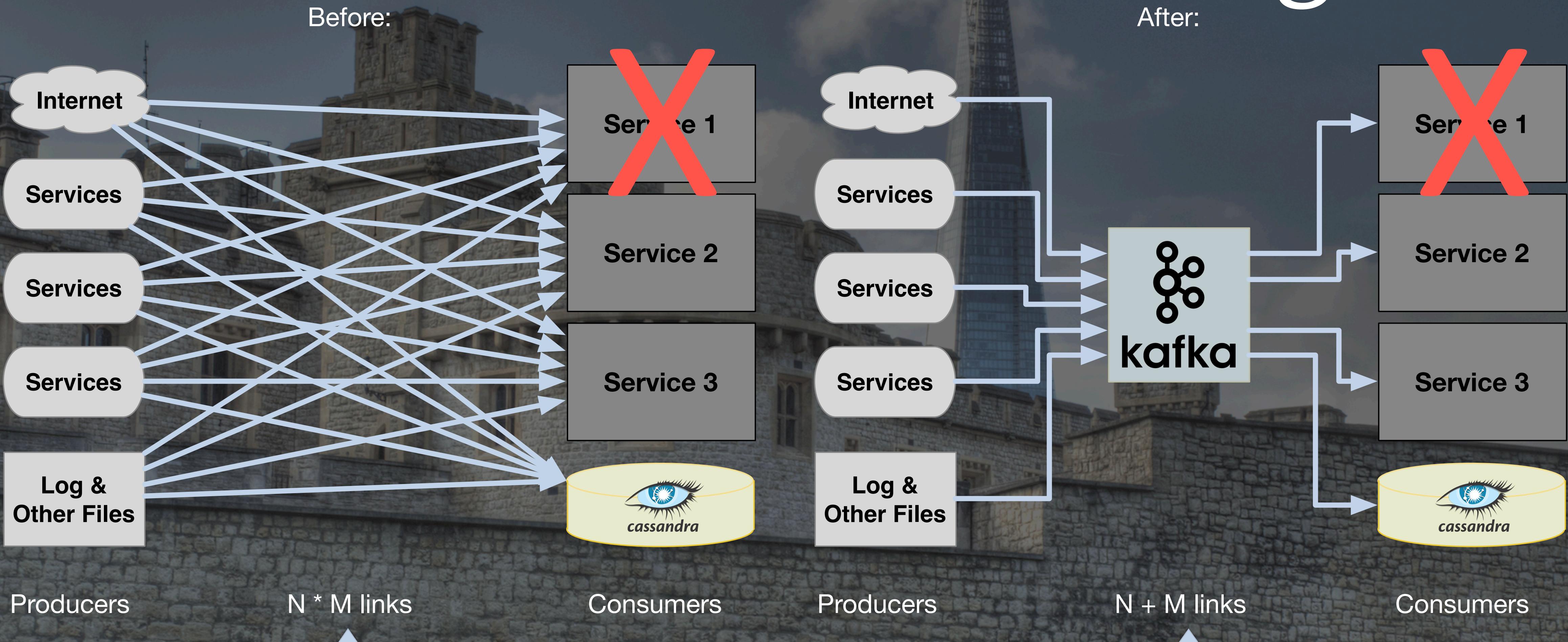
# Why Kafka?

Logs, not queues!



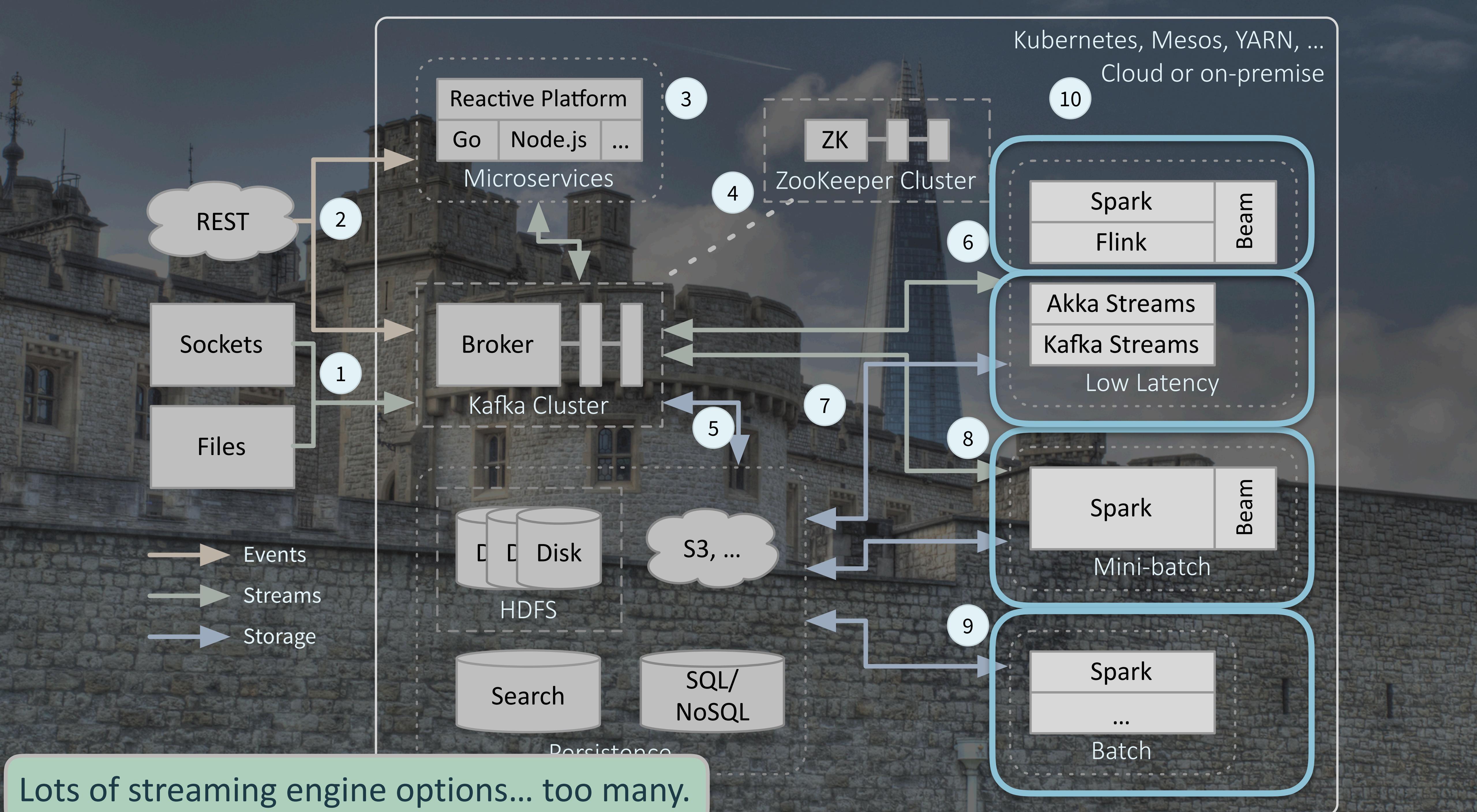
Unlike queues, consumers don't delete entries; Kafka manages their lifecycles

# Using Kafka



Messy and fragile;  
what if “Service 1”  
goes down?

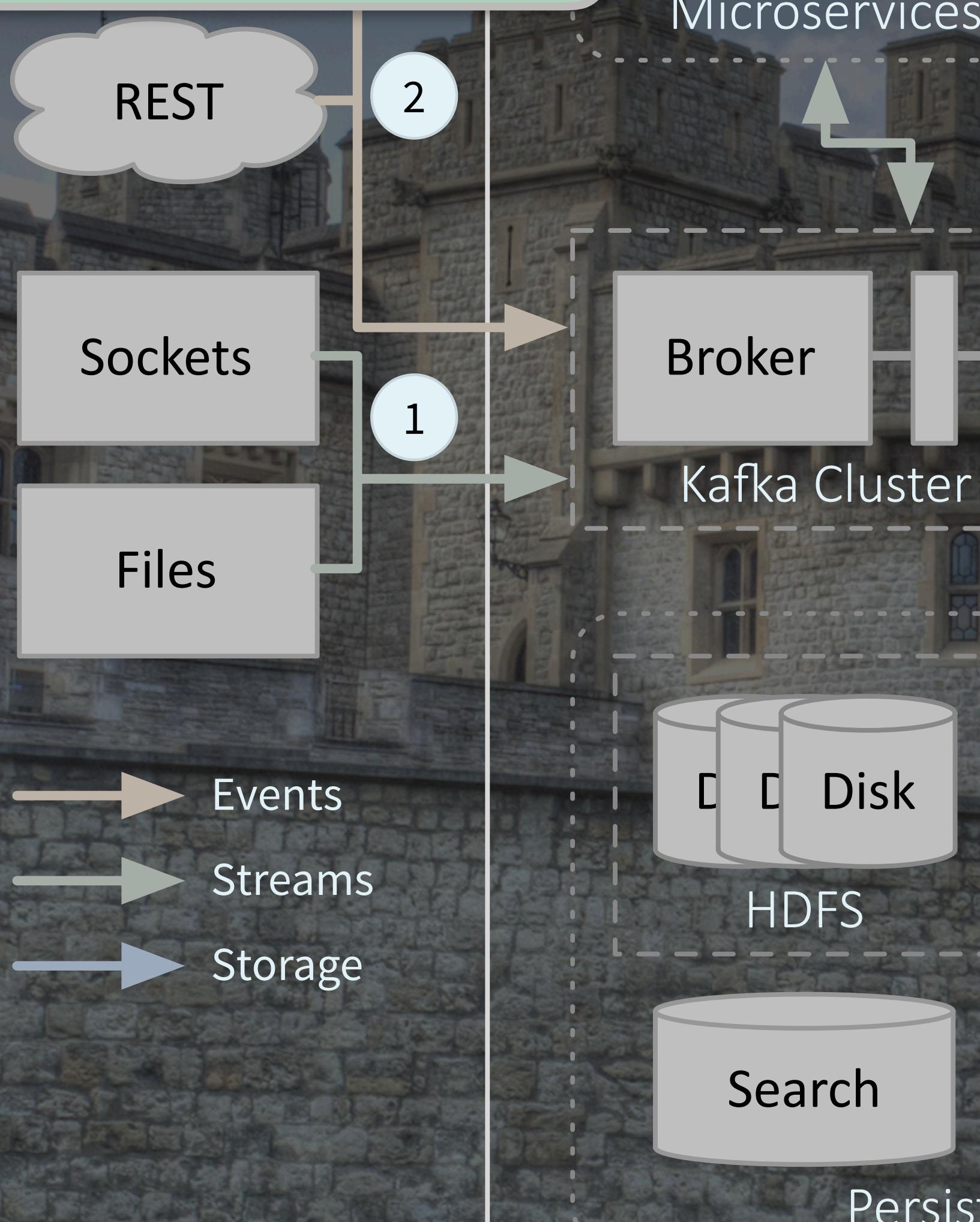
Simpler and more  
robust! Loss of Service  
1 means no data loss.



# How do you choose?

- Latency: how low?
- Volume per unit time: how high?
- Data processing: which kinds?
- Build, deploy, and manage services: what are your preferences?

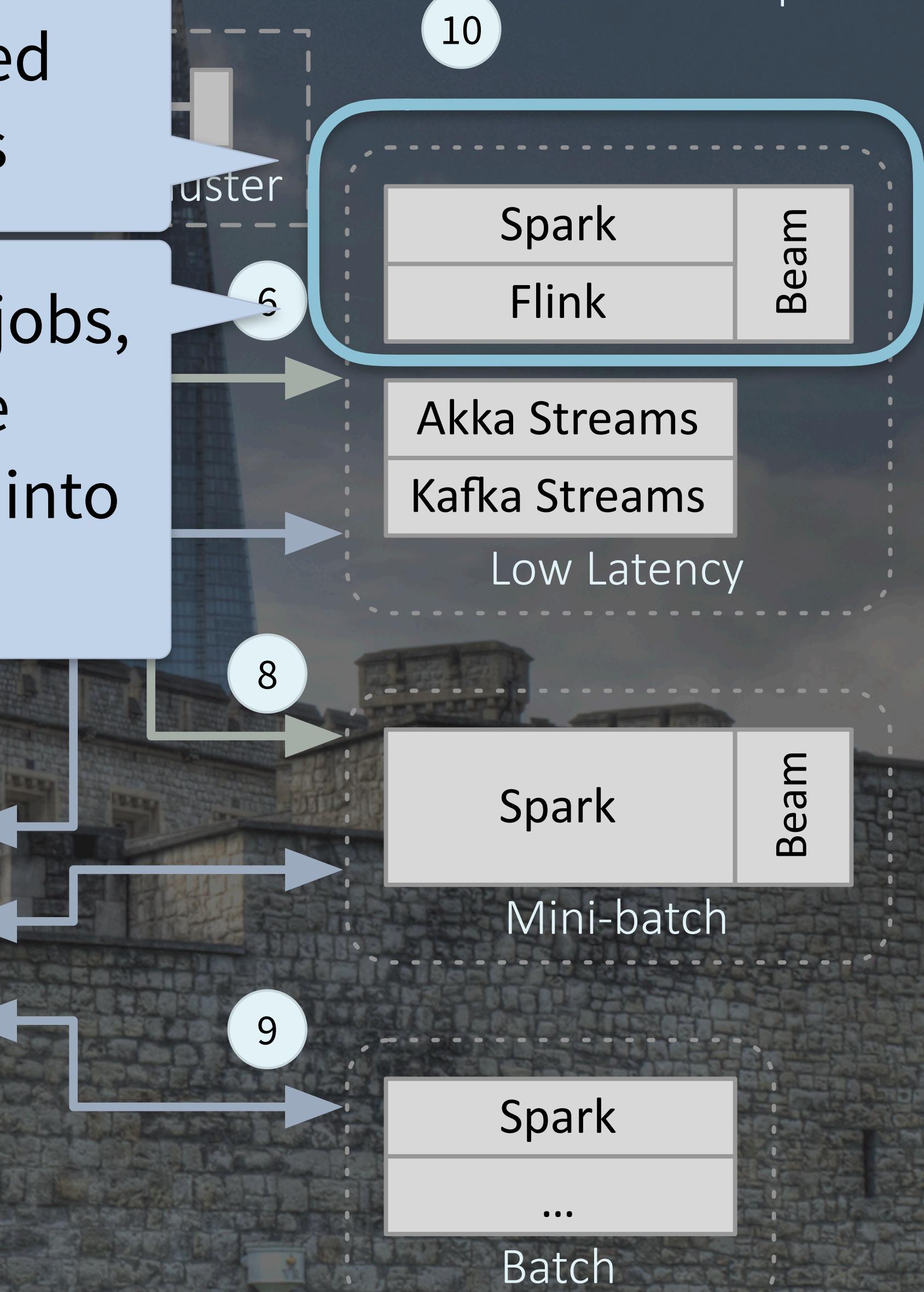
The streaming engines form two groups:



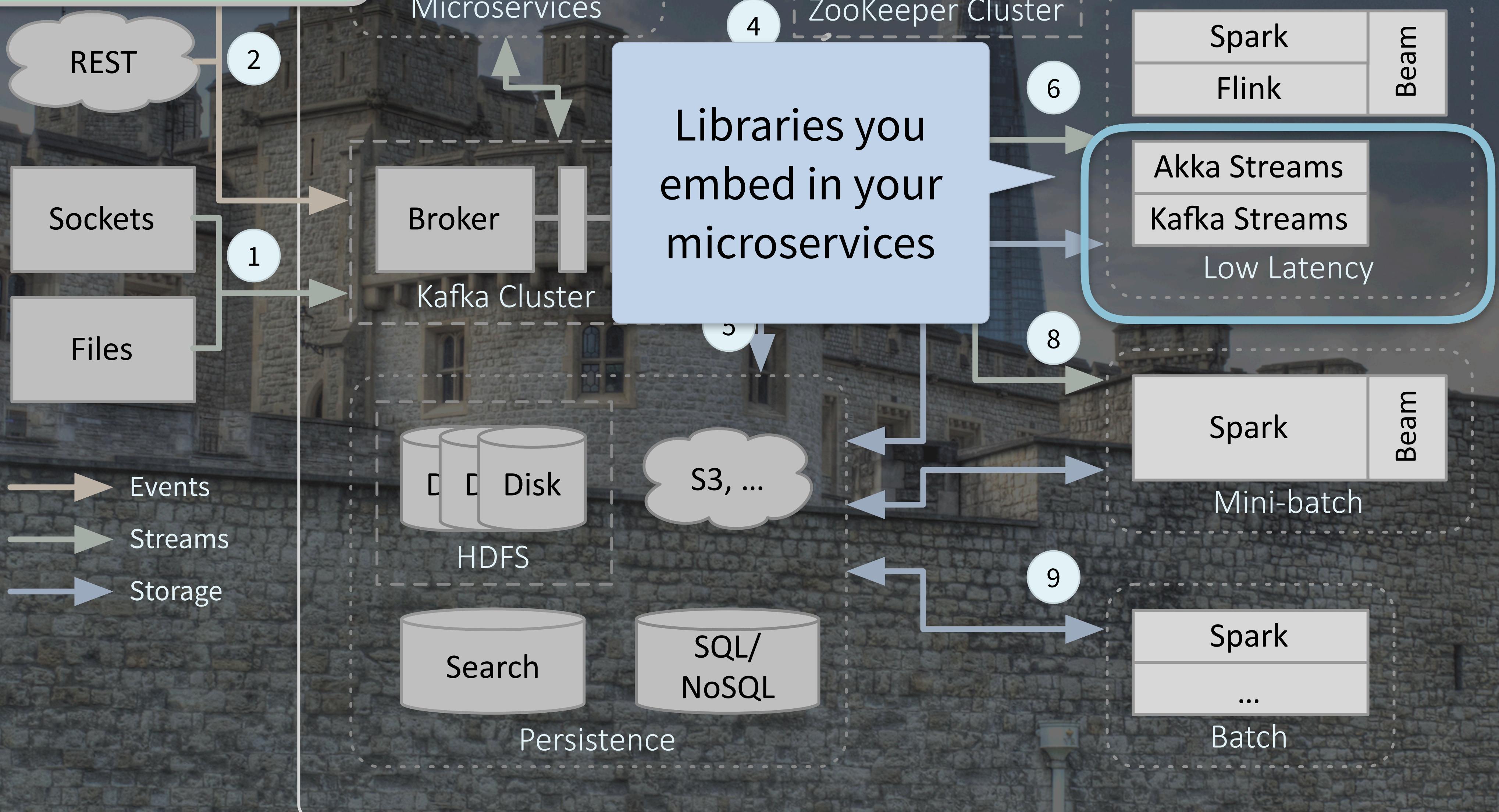
Run as distributed services

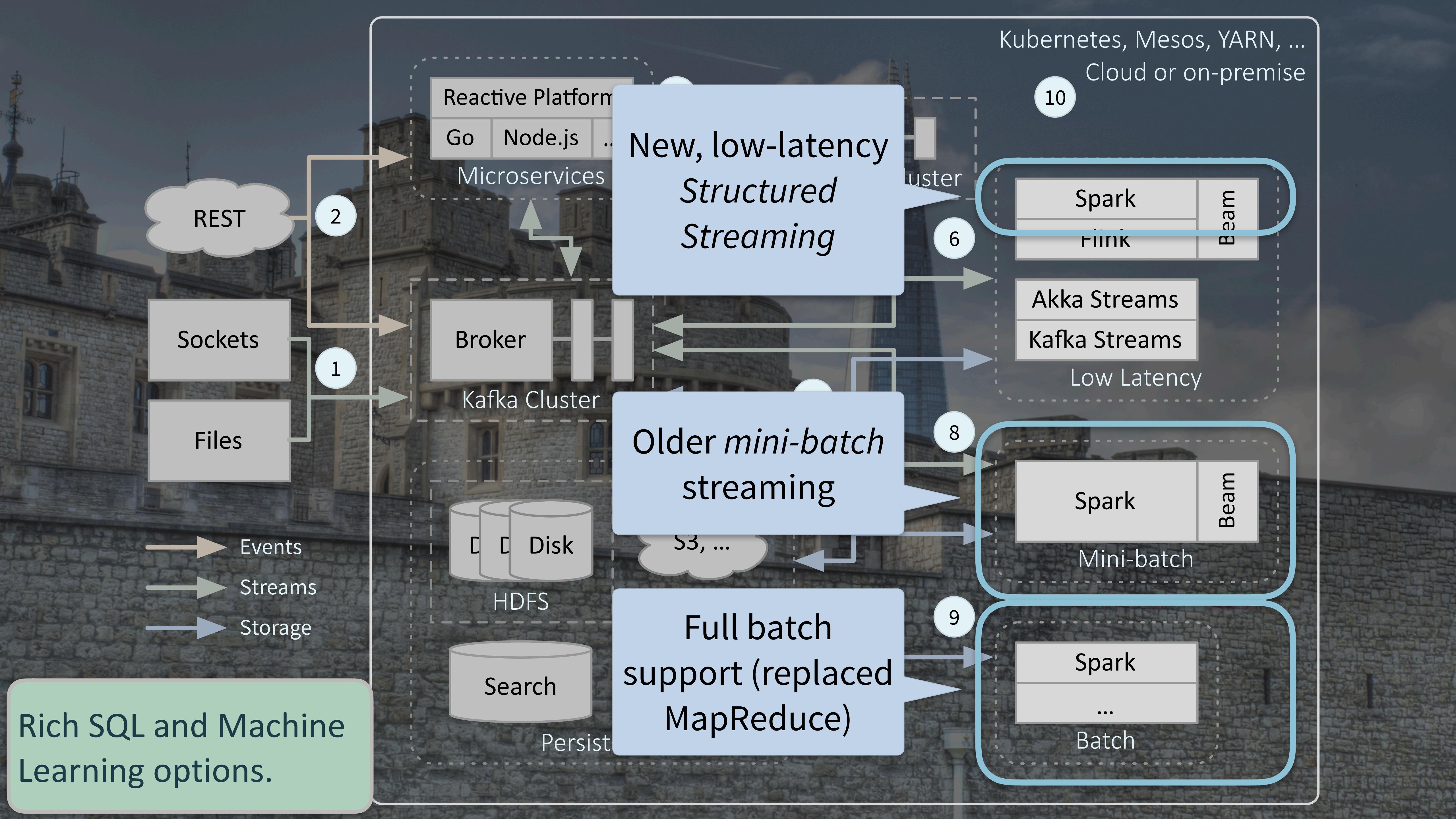
You submit jobs, they are partitioned into tasks

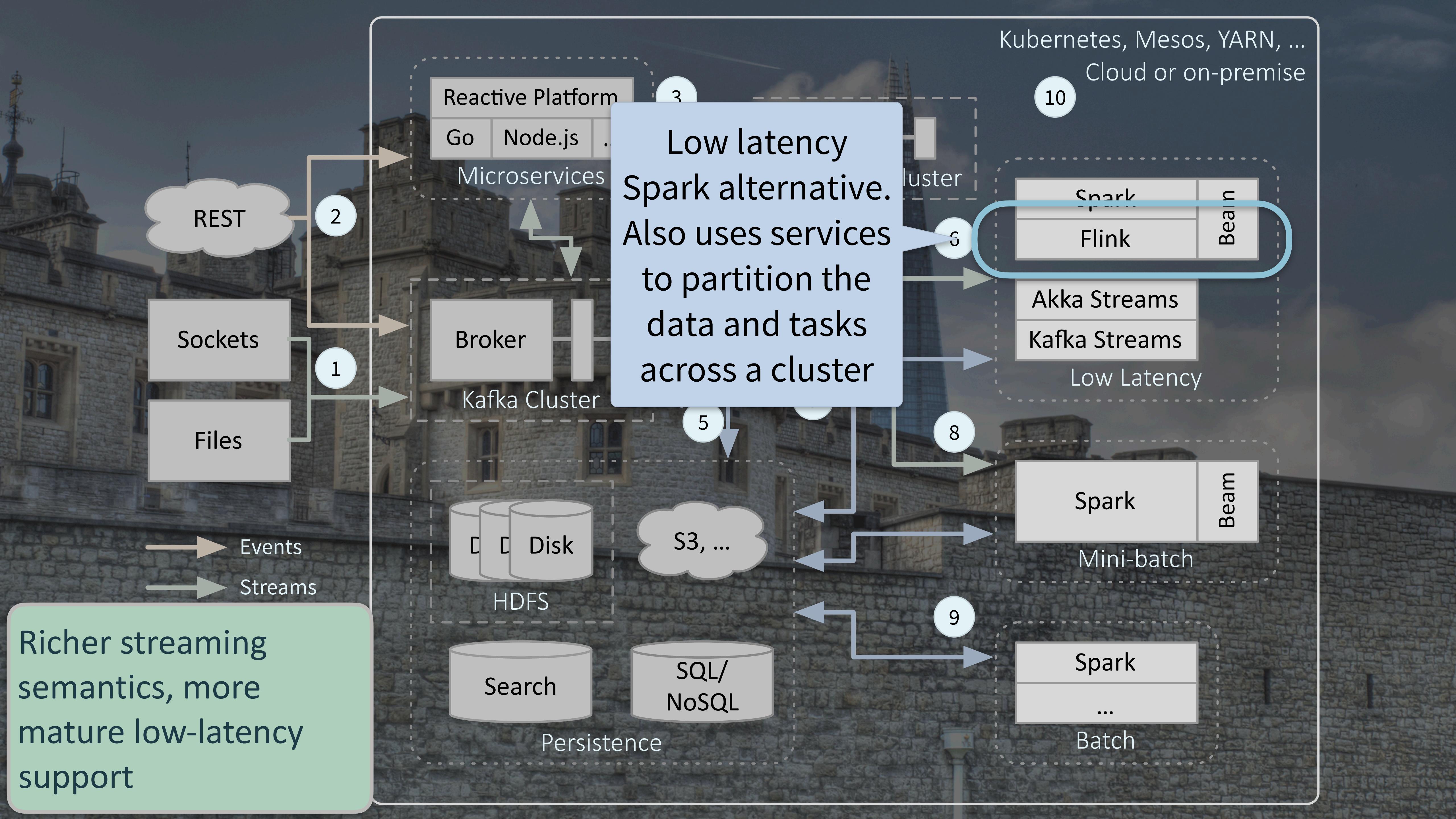
Kubernetes, Mesos, YARN, ...  
Cloud or on-premise

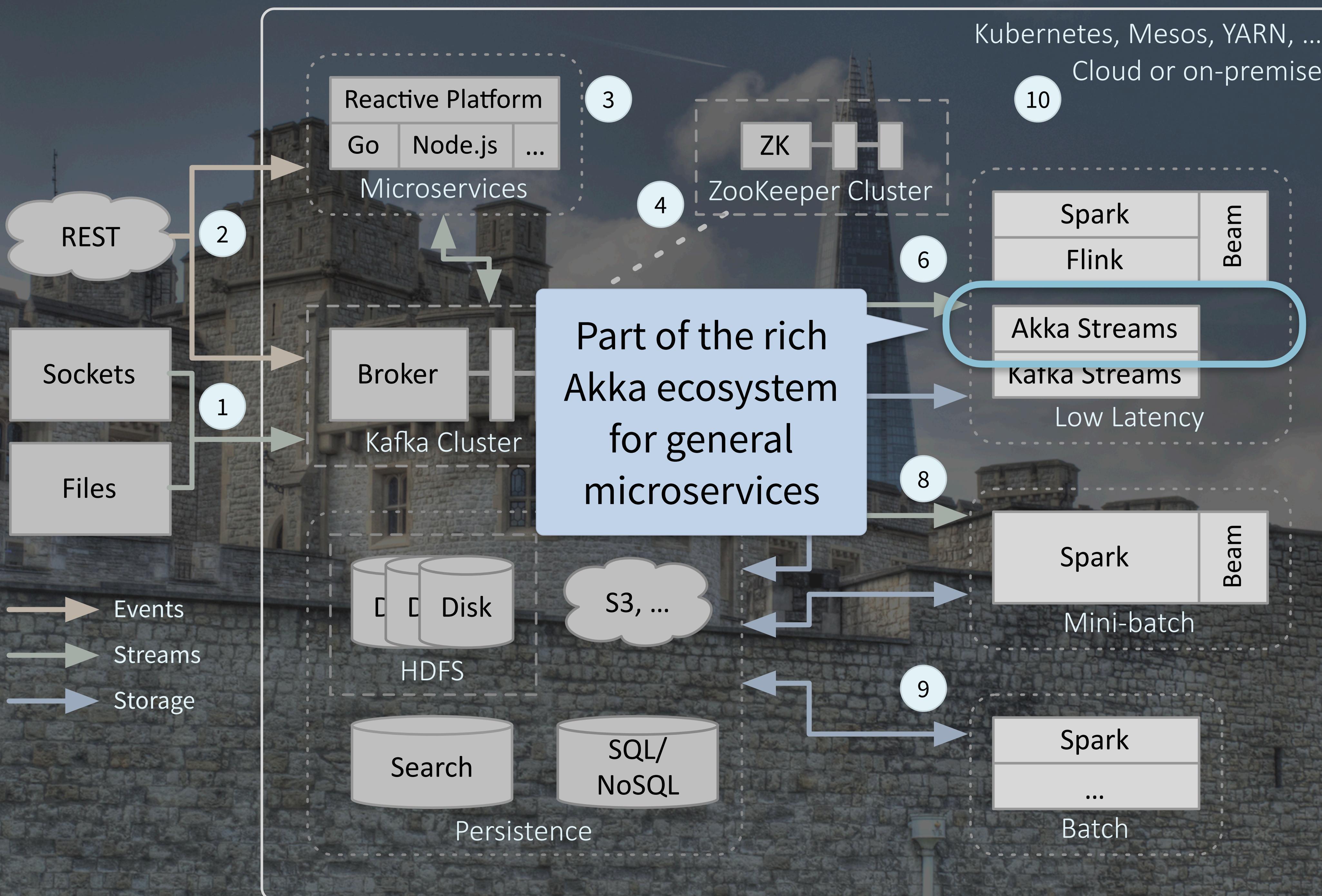


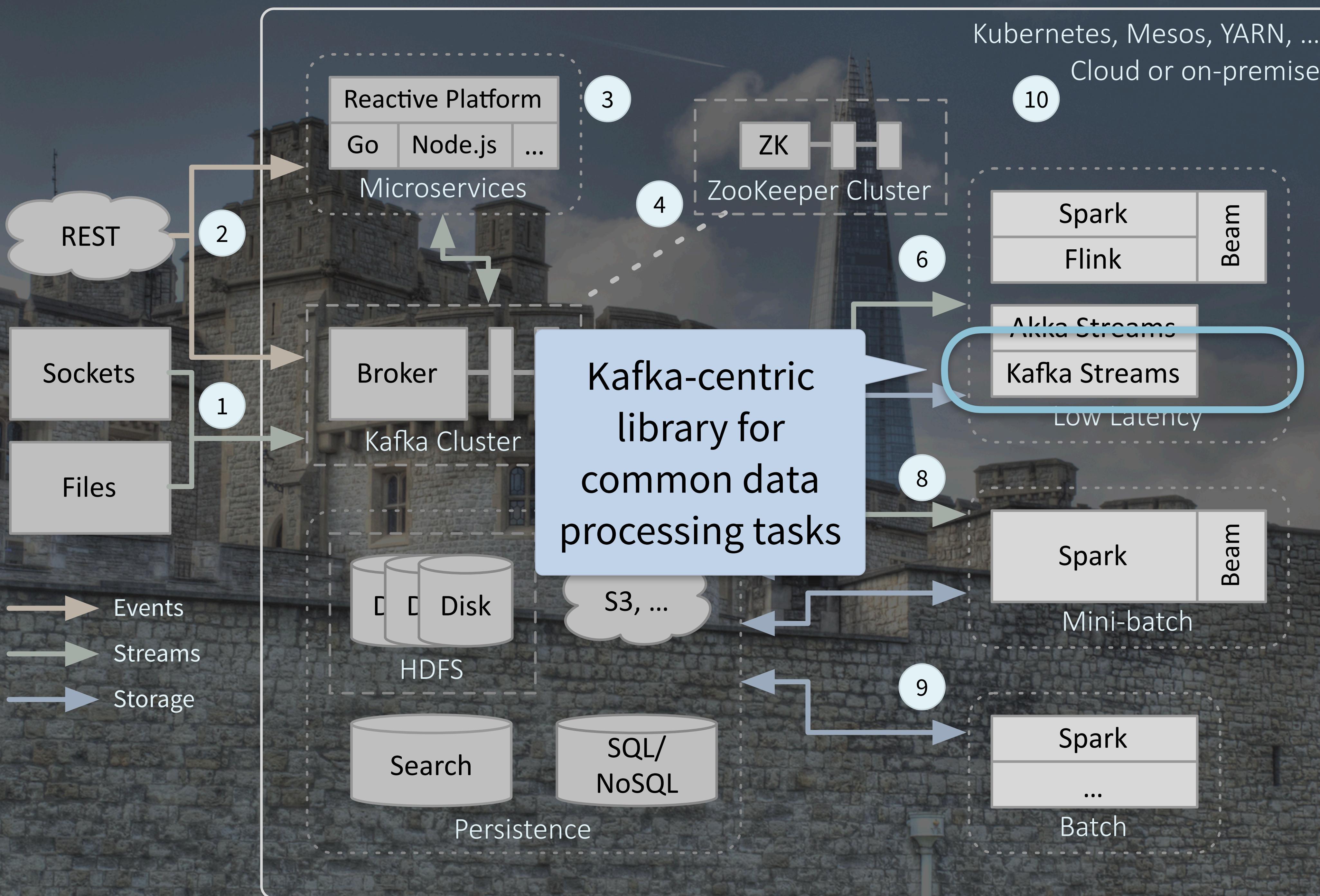
The streaming engines form two groups:

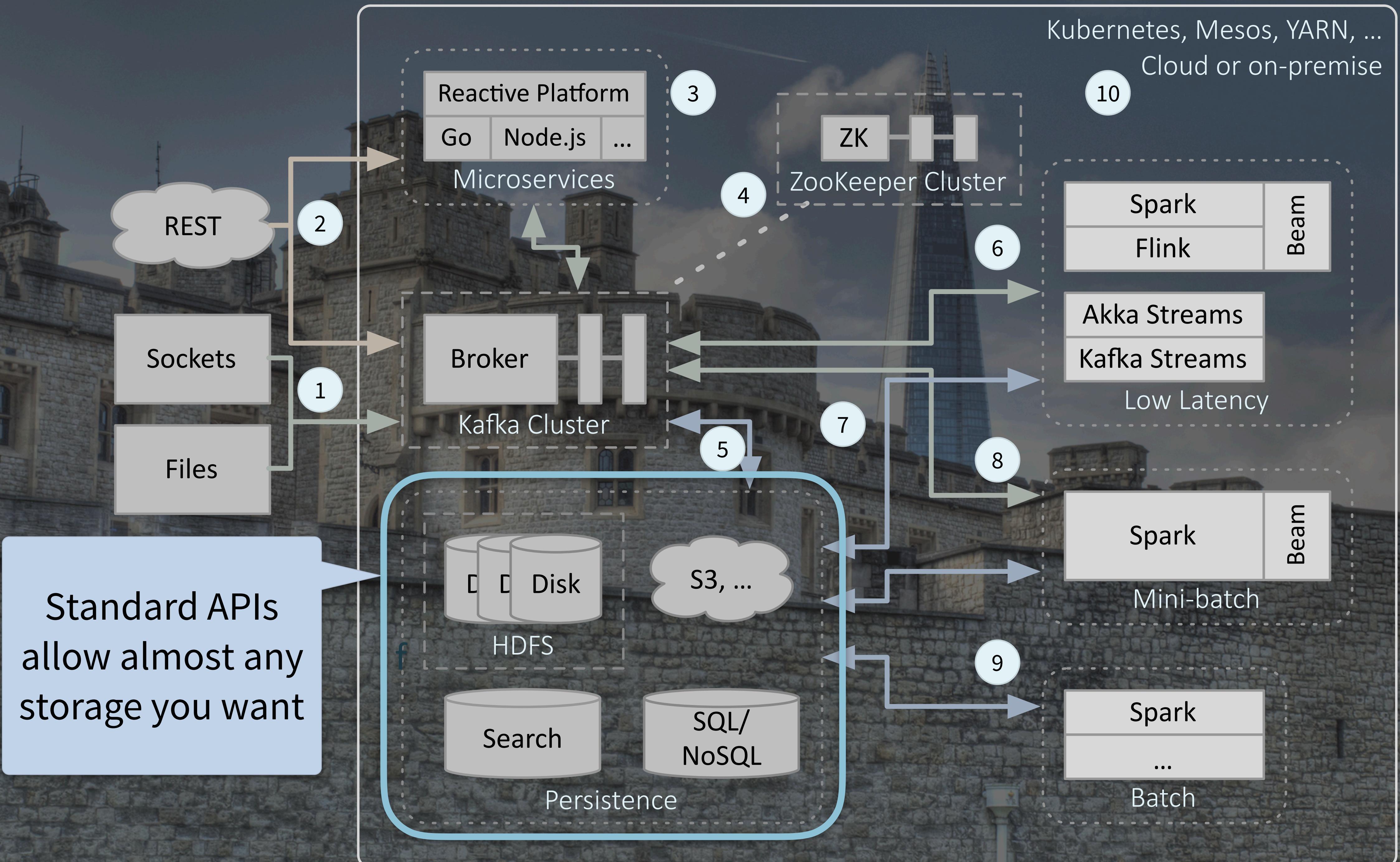


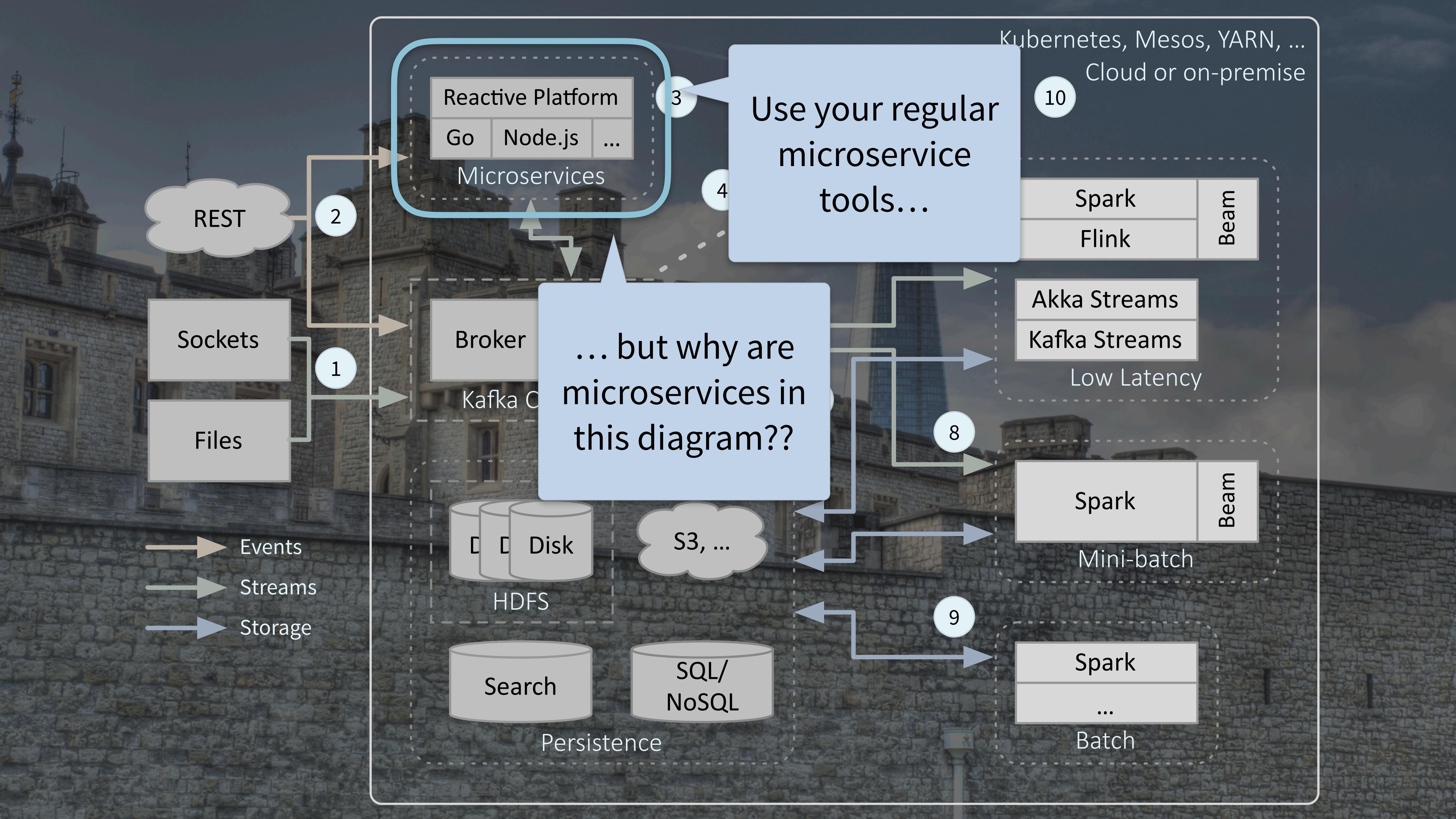












# Why Microservices in Fast Data?

1. The trend is to run everything in big clusters using Kubernetes or Mesos
  - In the cloud or on-premise

# Why Microservices in Fast Data?

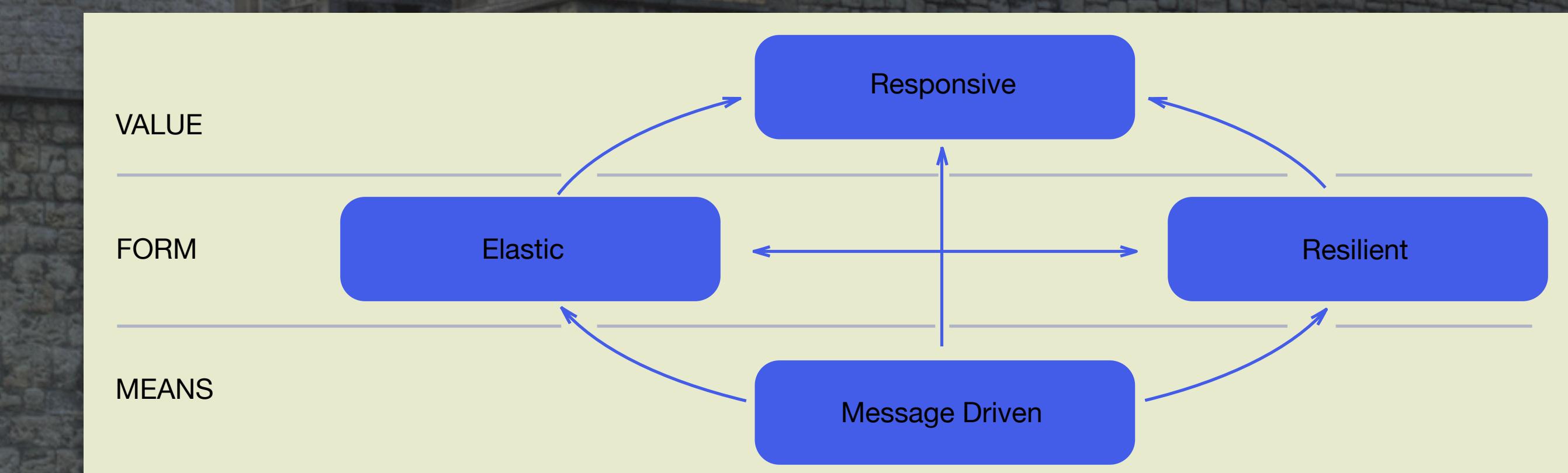
2. If streaming gives you information faster...

- ... you'll want quick access to it in your other services!

# Why Microservices in Fast Data?

## 3. Streaming raises the bar on data services

- Compared to batch services, long-running streaming services must be more:
- Scalable
- Resilient
- Flexible



# Why Microservices in Fast Data?

4. This leads to our last major point...



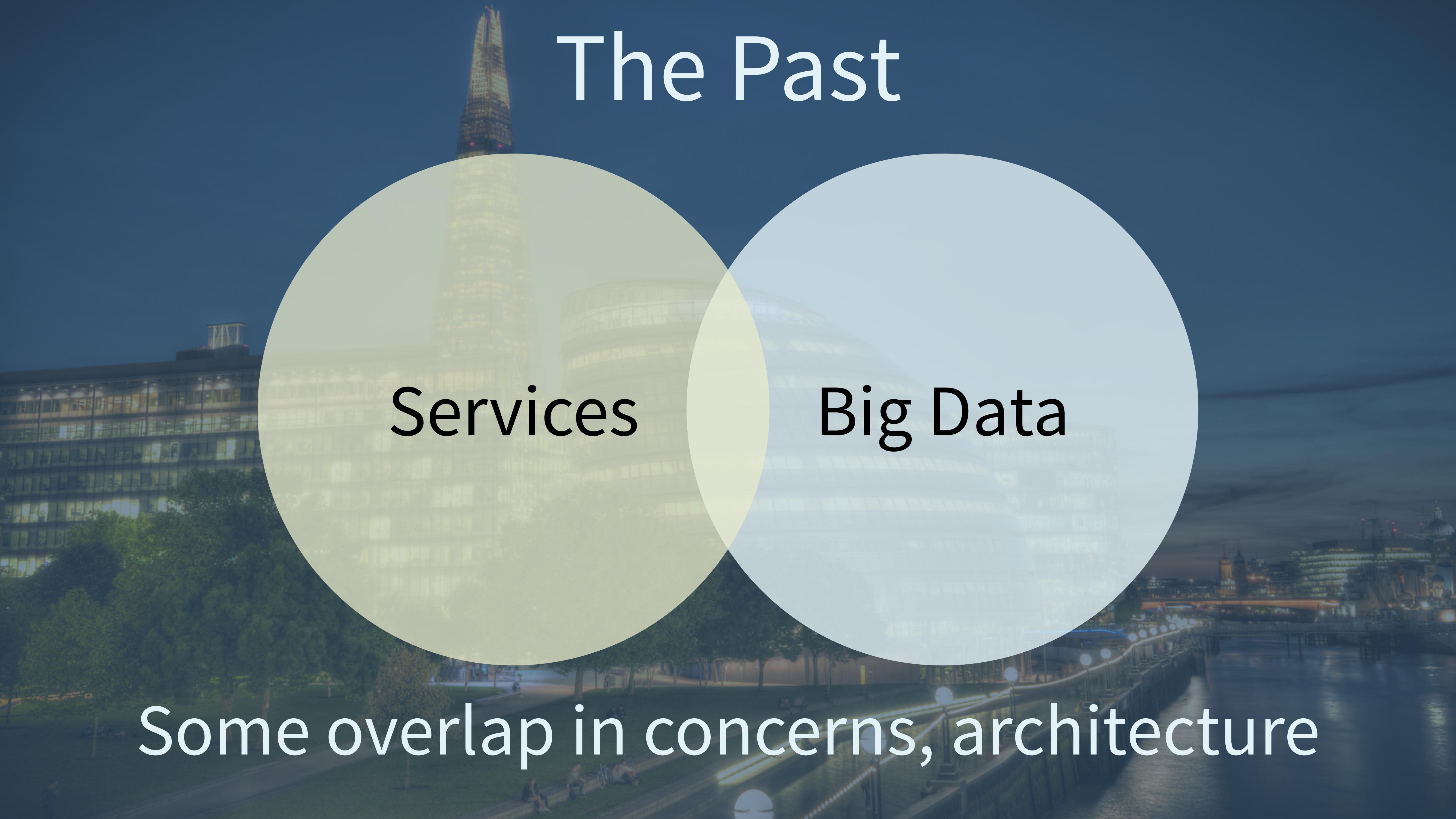
# Organizational Impact



# Organizational Impact

- Data scientists have to understand production issues
- Data engineers have to become good at highly-available microservices
- Microservice engineers have to become good at data

# The Past

A Venn diagram with two overlapping circles is overlaid on a photograph of a city skyline at dusk. The left circle is yellow and contains the word 'Services'. The right circle is light blue and contains the words 'Big Data'. The background shows the Shard skyscraper, the London Eye, and other buildings along the River Thames.

Services

Big Data

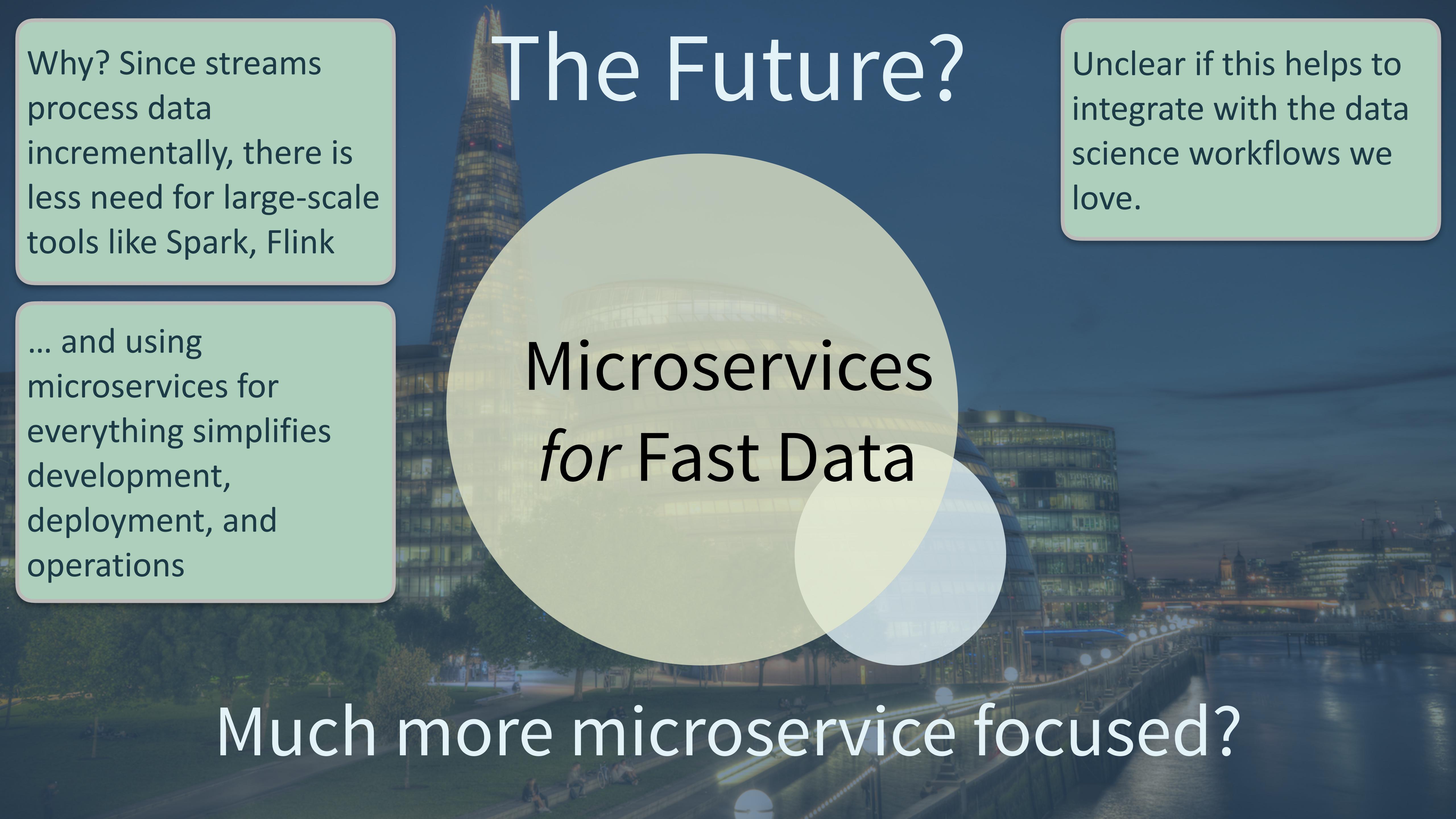
Some overlap in concerns, architecture

# The Present



Microservices  
& Fast Data

Much more overlap



Why? Since streams process data incrementally, there is less need for large-scale tools like Spark, Flink

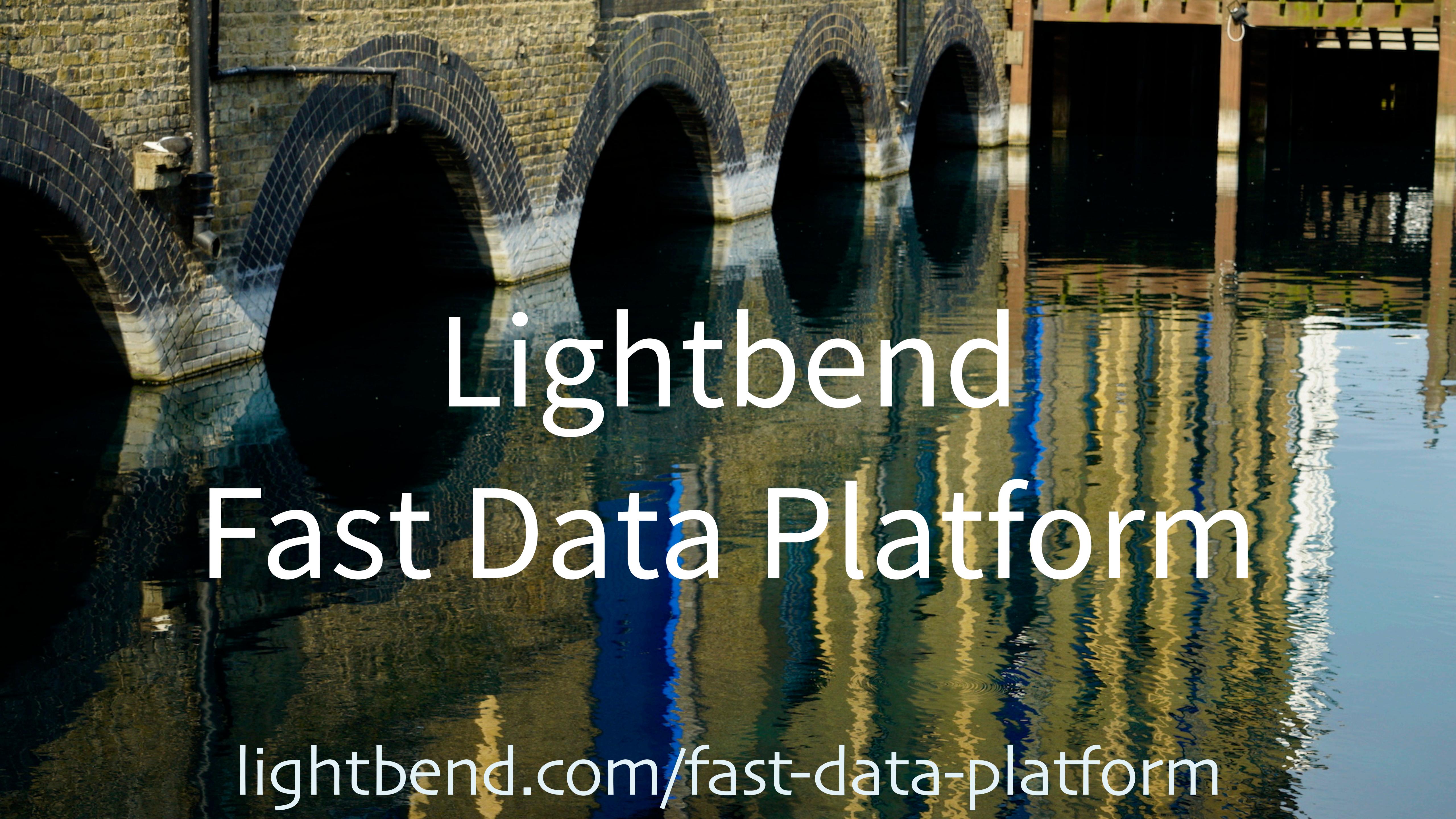
... and using microservices for everything simplifies development, deployment, and operations

# The Future?

## Microservices for Fast Data

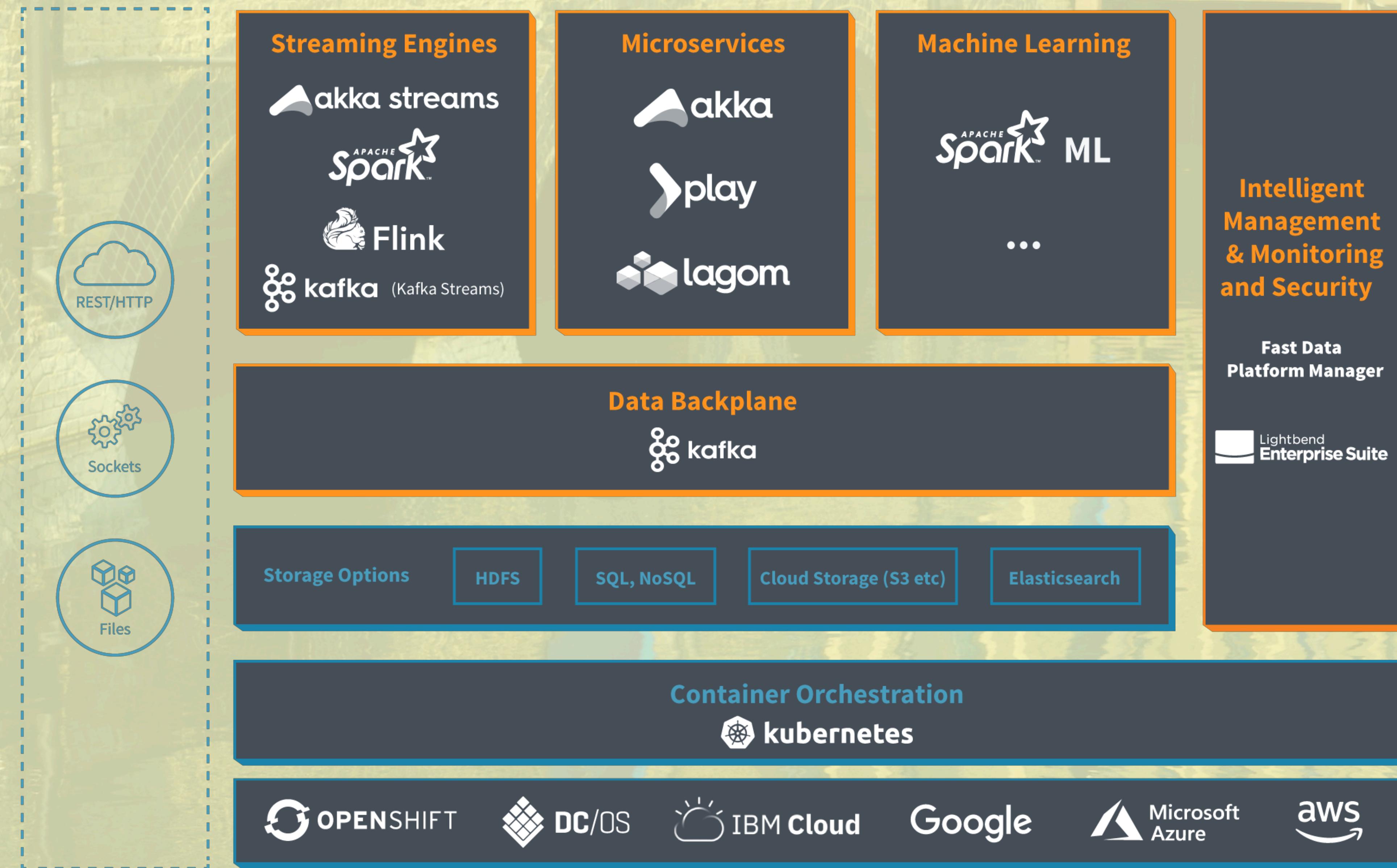
Much more microservice focused?

Unclear if this helps to integrate with the data science workflows we love.

A photograph of a multi-arched brick bridge reflected perfectly in the still water below. The bridge's structure is composed of several large, semi-circular arches made of dark-colored bricks. The reflection in the water is sharp and clear, mirroring the bridge's arches and the surrounding environment. The overall scene is peaceful and symmetrical.

# Lightbend Fast Data Platform

[lightbend.com/fast-data-platform](http://lightbend.com/fast-data-platform)



[lightbend.com/fast-data-platform](http://lightbend.com/fast-data-platform)

What we  
discussed

### Streaming Engines

akka streams



Flink

kafka (Kafka Streams)



### Microservices

akka

play

lagom

### Machine Learning

APACHE ML

...

Intelligent  
Management  
& Monitoring  
and Security

Fast Data  
Platform Manager

Lightbend  
Enterprise Suite

### Data Backplane

kafka

### Storage Options

HDFS

SQL, NoSQL

Cloud Storage (S3 etc)

Elasticsearch

### Container Orchestration

kubernetes

OPENSHIFT

DC/OS

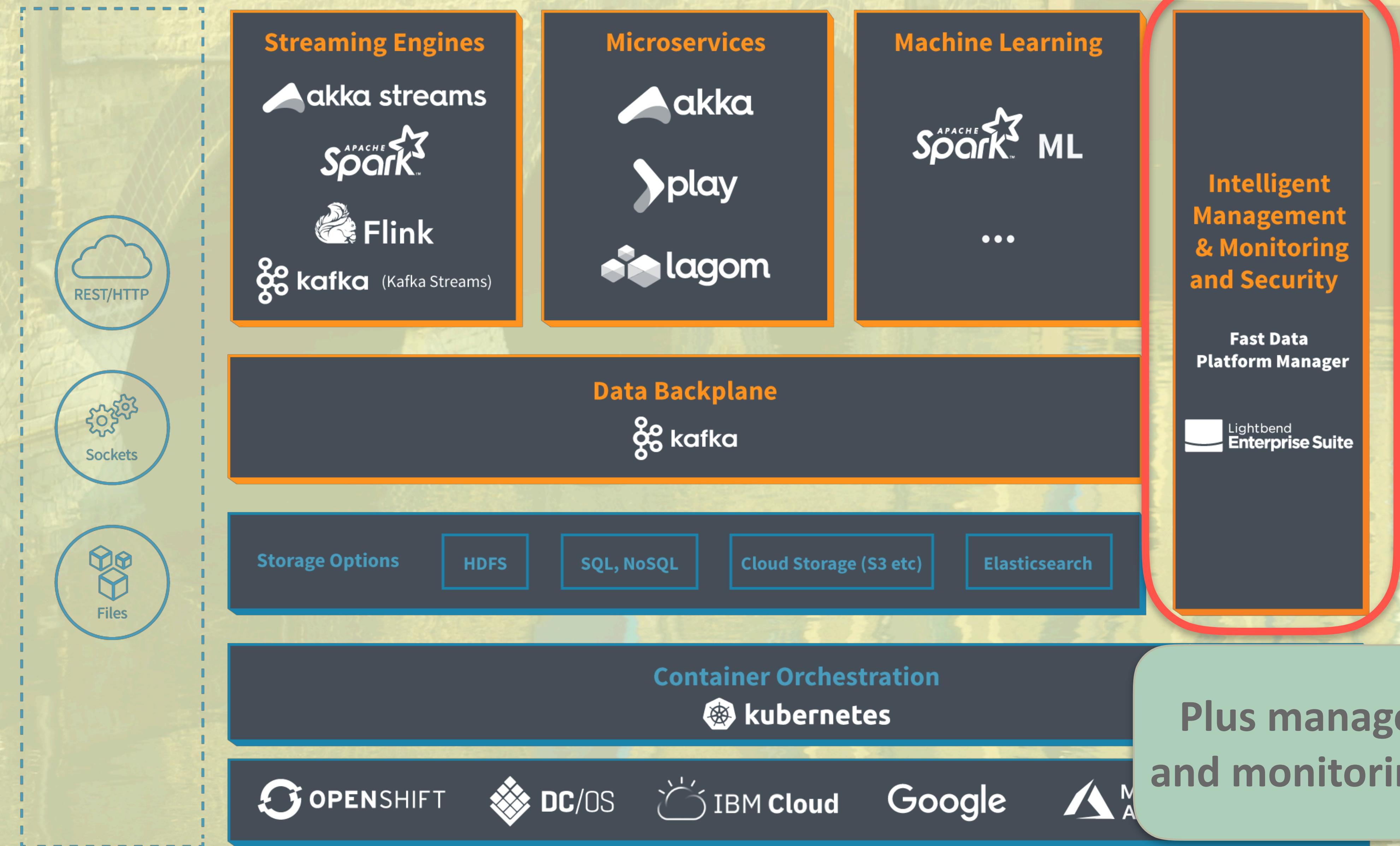
IBM Cloud

Google

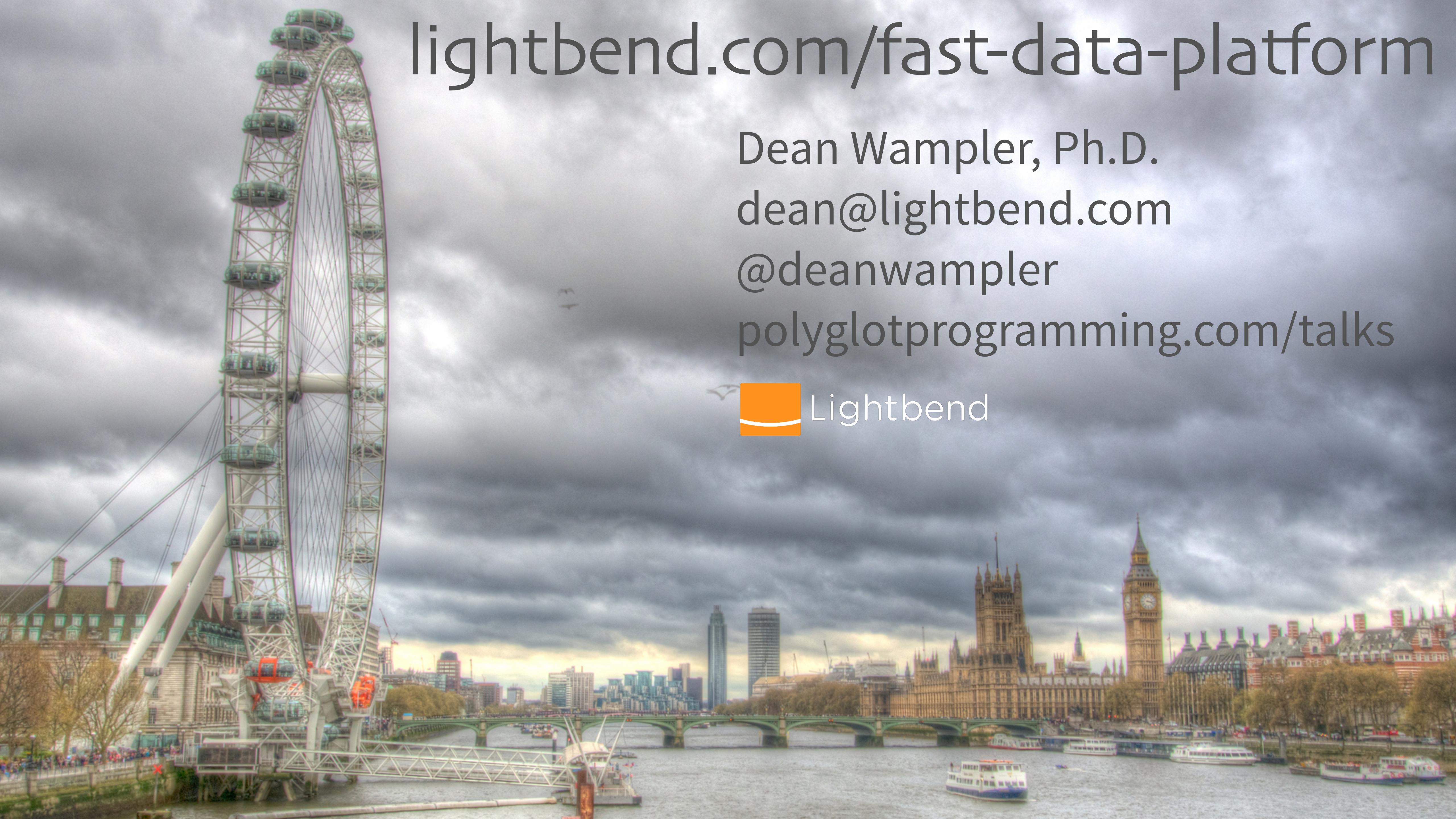
Microsoft  
Azure

aws

[lightbend.com/fast-data-platform](http://lightbend.com/fast-data-platform)



[lightbend.com/fast-data-platform](http://lightbend.com/fast-data-platform)

A photograph of the London skyline featuring the London Eye (Millennium Wheel) on the left and the Palace of Westminster with Big Ben on the right. The sky is overcast with dramatic clouds. In the foreground, the River Thames is visible with several boats, and a green bridge spans the river.

# [lightbend.com/fast-data-platform](http://lightbend.com/fast-data-platform)

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



Lightbend