



# STRIMZI: HOW APACHE KAFKA HAS FALLEN IN LOVE WITH KUBERNETES

Running Apache Kafka on Kubernetes

Paolo Patierno  
Principal Software Engineer, Messaging & IoT team  
3/4/2019

# MY CUSTOM RESOURCE

```
kubectl apply -f paolopatierno.yaml
```

**apiVersion:** redhat/v1

**kind:** PrincipalSoftwareEngineer

**metadata:**

**name:** Paolo Patierno

**namespace:** Red Hat, Messaging & IoT team

**annotations:**

eclipse/committer: Vert.x, Hono & Paho

microsoft/mvp: Azure & IoT

**labels:**

family: dad of two, husband of one

sports: running, swimming, motogp, vr46, ssc napoli

**spec:**

**replicas:** 1

**containers:**

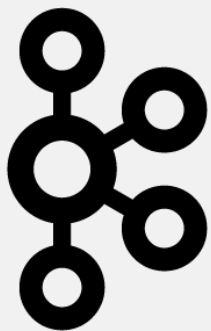
- **image:** patiernohub.io/paolo:latest



@ppatierno

# WHAT AM I GOING TO TELL YOU?

... a true love story



**kafka**



**kubernetes**

# APACHE KAFKA

What is that?



“ ... a publish/subscribe  
messaging system ...”

“ ... a streaming  
data platform ...”

“ ... a distributed, horizontally-scalable,  
fault-tolerant, commit log ...”

# APACHE KAFKA

What is that?



- Developed at LinkedIn back in 2010, open sourced in 2011
- Designed to be fast, scalable, durable and available
- Distributed by nature
- Data partitioning (sharding)
- High throughput / low latency
- Ability to handle huge number of consumers

# APACHE KAFKA

What is that?



- Messages are sent to and received from a topic
  - Topics are split into one or more partitions (aka shards)
  - All actual work is done on partition level, topic is just a virtual object
- Each message is written only into a one selected partition
  - Partitioning is usually done based on the message key
  - Message ordering within the partition is fixed
- Retention
  - Based on size / message age
  - Compacted based on message key
- Replication
  - Each partition can exist in one or more Backup copies to achieve high availability in case of node failures

# KUBERNETES

What is that?



# kubernetes

“ A system for ...”

“ ... automating deployment ...”

“ ... scaling ...”

“ ... management ...”

“ ... of containerized applications ...”

“ It’s like a Linux kernel ... but for distributed systems”

# KUBERNETES

What is that?



# kubernetes

- Comes from Google experience with project “Borg”
- Abstract the underlying hardware in terms of “nodes”
- On the nodes a set of different “resources” can be deployed and handled
- Containerized applications are deployed, using and sharing “resources”



# KUBERNETES

What is that?



# kubernetes

Container  
scheduling

Self healing

Secret &  
configuration  
management

Horizontal  
scaling

Load  
balancing

Service  
discovery

Automated  
rollout/rollback

Batch  
execution

Storage  
orchestration



“ ... every love story has many challenges ...”

# THE CHALLENGES

... of running Apache Kafka on Kubernetes

- A Kafka cluster requires ...
  - ... a stable broker identity and stable network address
  - ... a way for brokers to discover each other and communicate
  - ... durable state on brokers and storage recovery
  - ... to have brokers accessible from clients, directly
- ... and if it's not enough, it runs alongside a Zookeeper ensemble which requires ...
  - ... each node has the configuration of the others
  - ... to have nodes able to communicate each others
- Accessing Kafka isn't so simple

# THE CHALLENGES

... how Kubernetes can help on that?

- Kubernetes provides ...
  - ... StatefulSets for stable identity and network ...
  - ... together with Headless services for internal discovery
  - ... Services for accessing the cluster
  - ... Secrets and ConfigMap for handling configurations
  - ... PersistentVolume and PersistentVolumeClaim for durable storage

“ ... but it’s not so easy ... ”

“ ... you would need an operator handling all these Kubernetes resources for you ...”

“ ... you could ask him to apply the changes you want for you or recovery from problems automatically ...”

“ ... that’s a great idea! Why not?”

# THE OPERATOR

“Someone” who works for you

- It has the knowledge of the application to control
- It lets you to “describe” your application ...
  - ... and deploy it for you
- It watches the “desired” state and the “actual” state ...
  - ... taking appropriate actions
- It can handle the entire lifecycle of an application
  - Upgrades, security, ...



# OPERATORS FTW!

Welcome to their new home

- OperatorHub.io
  - A new home for the Kubernetes community to share operators
- Operators available today:
  - AWS service
  - Etcd
  - MongoDB
  - Prometheus
  - Redis
  - ... and many more





# WELCOME TO STRIMZI


The open source Apache Kafka operator



OperatorHub.io

Search OperatorHub...

Contribute ▾



## Strimzi Kafka

Run an Apache Kafka cluster, including Kafka Connect, ZooKeeper and more.

[Home](#) > [Strimzi Kafka](#)

## Strimzi Kafka

Strimzi provides a way to run an [Apache Kafka](#) cluster on [Kubernetes](#) or [OpenShift](#) in various deployment configurations. See our [website](#) for more details about the project.

### Supported Features

- **Manages the Kafka Cluster** - Deploys and manages all of the components of this complex application, including dependencies like ZooKeeper that are traditionally hard to administer.
- **Includes Kafka Connect** - Allows for configuration of common data sources and sinks to move data into and out of the Kafka cluster.
- **Topic Management** - Creates and manages Kafka Topics within the cluster.
- **User Management** - Creates and manages Kafka Users within the cluster.

Install

OPERATOR VERSION

0.11.1 (latest) ▾

CAPABILITY LEVEL ⓘ

- ✓ Basic Install
- ✓ Seamless Upgrades

# WELCOME TO STRIMZI

The open source Apache Kafka operator



- Open source project licensed under Apache License 2.0
- Focuses on running Apache Kafka on Kubernetes and OpenShift:
  - Container images for Apache Kafka and Apache Zookeeper
  - Operators for managing and configuring Kafka clusters, topics or users
- Provides Kubernetes-native experience for running Kafka on Kubernetes and OpenShift
  - Kafka cluster, topic and user as Kubernetes custom resources



Web site: <http://strimzi.io/>



GitHub: <https://github.com/strimzi>



Twitter: [@strimziio](https://twitter.com/strimziio)

# Red Hat AMQ Streams

Apache Kafka for the Enterprise



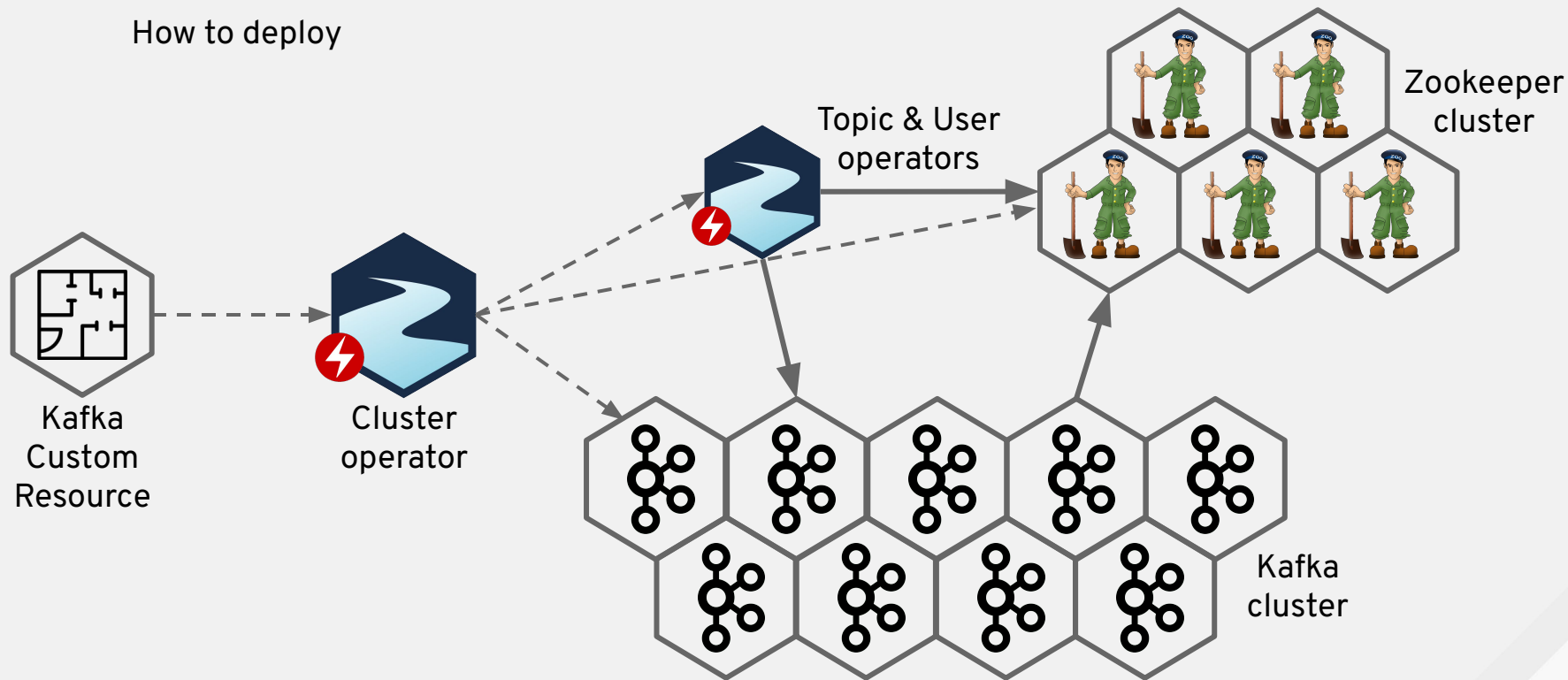
- Part of the Red Hat AMQ suite
- AMQ Streams on OCP
  - Running Apache Kafka on OpenShift Container Platform
  - Based on the Strimzi project
- AMQ Streams on RHEL
  - Running Apache Kafka on “bare metal”



Web site: <https://access.redhat.com/products/red-hat-amq#streams>

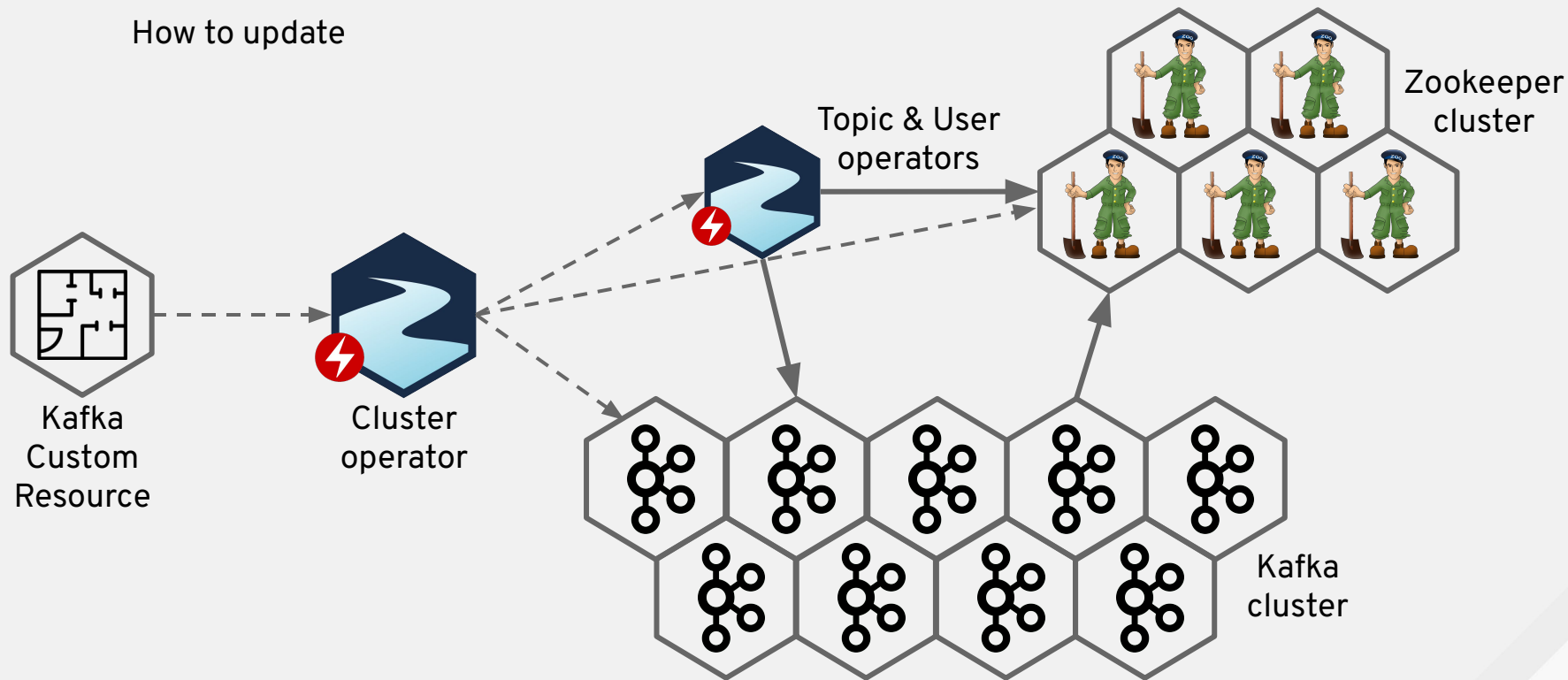
# KAFKA CLUSTER

How to deploy



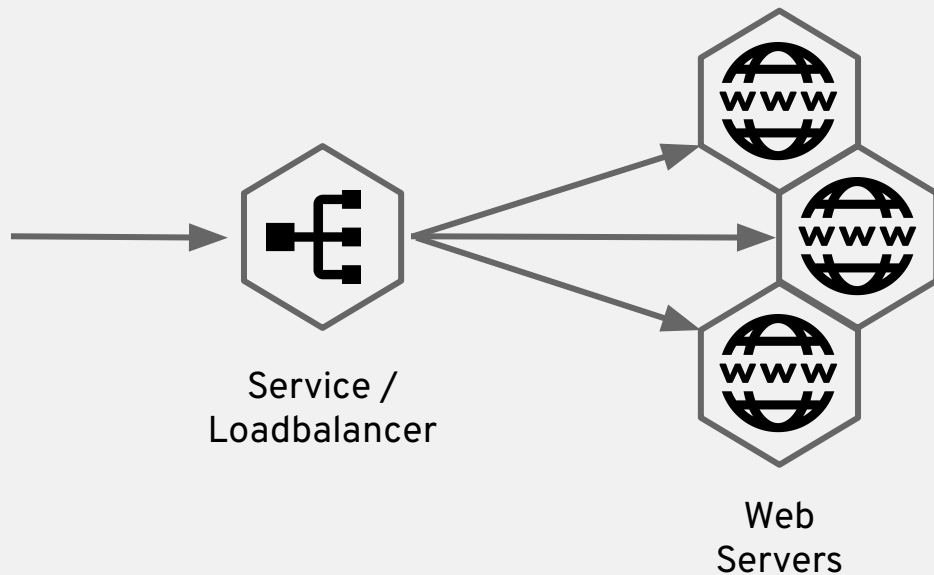
# KAFKA CLUSTER

How to update



# KUBERNETES APPLICATIONS

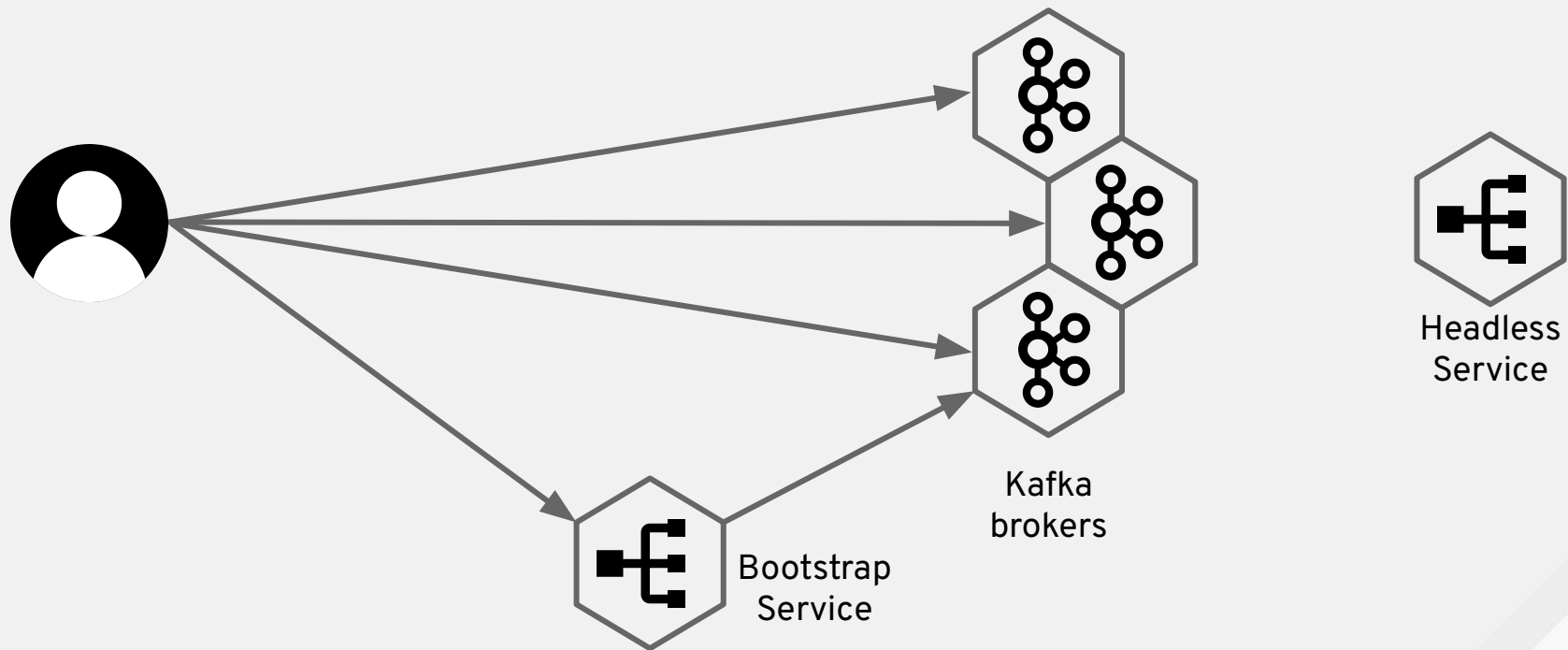
How to access



“ ... but load balancing cannot work with  
Kafka!”

# KAFKA CLUSTER

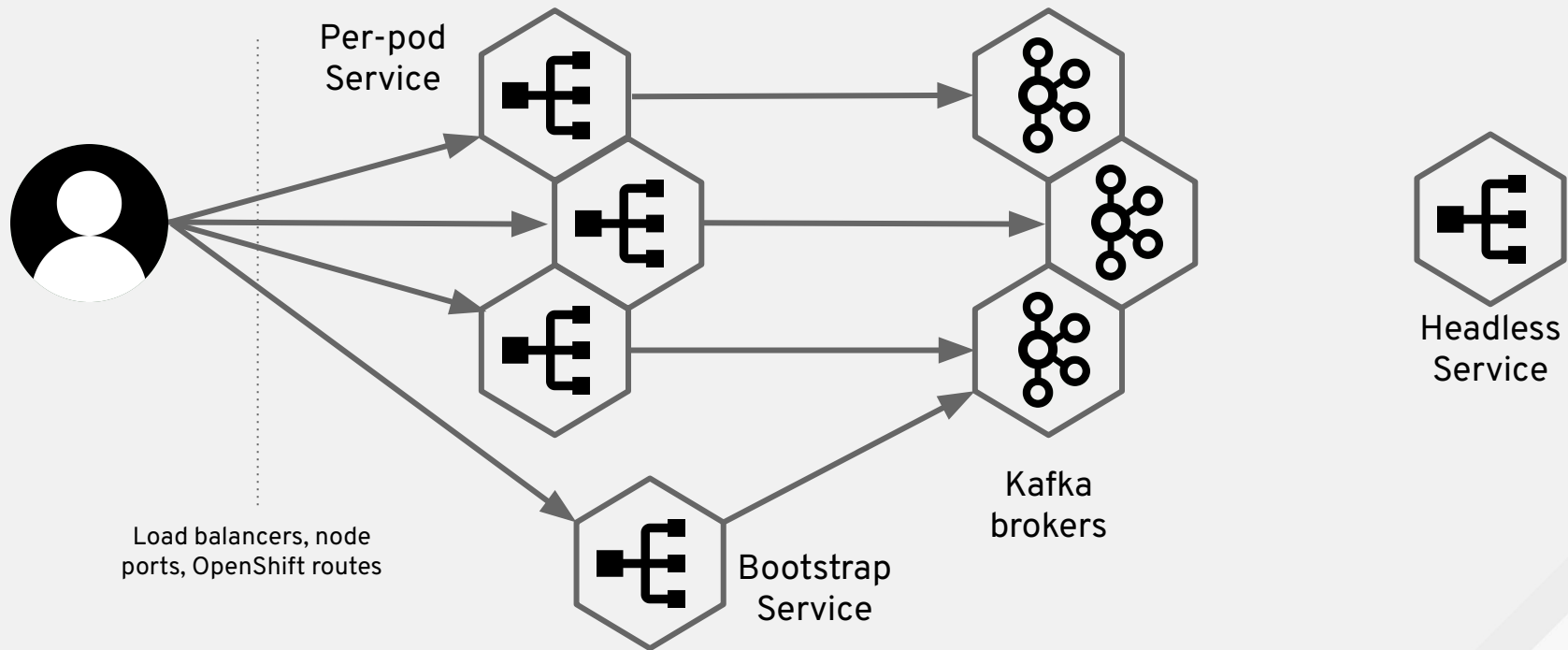
How to access: inside Kubernetes





# KAFKA CLUSTER

How to access: from outside Kubernetes



# FEATURES

It's really more

Storage

Annotations

Mirroring

Labels

Prometheus

Scale Up

Off cluster access

Tolerations

Metrics

Affinity

Encryption

CPU and  
RAM

Logging

Healthchecks

Users

Upgrades

Secrets

Authentication

Authorization

Configuration

Kafka  
Connect

Pod Disruption  
Budgets

Scale  
Down

Topic

Grafana

Source2Image

ACLs

HA

Network  
Policies

ImagePullSecrets

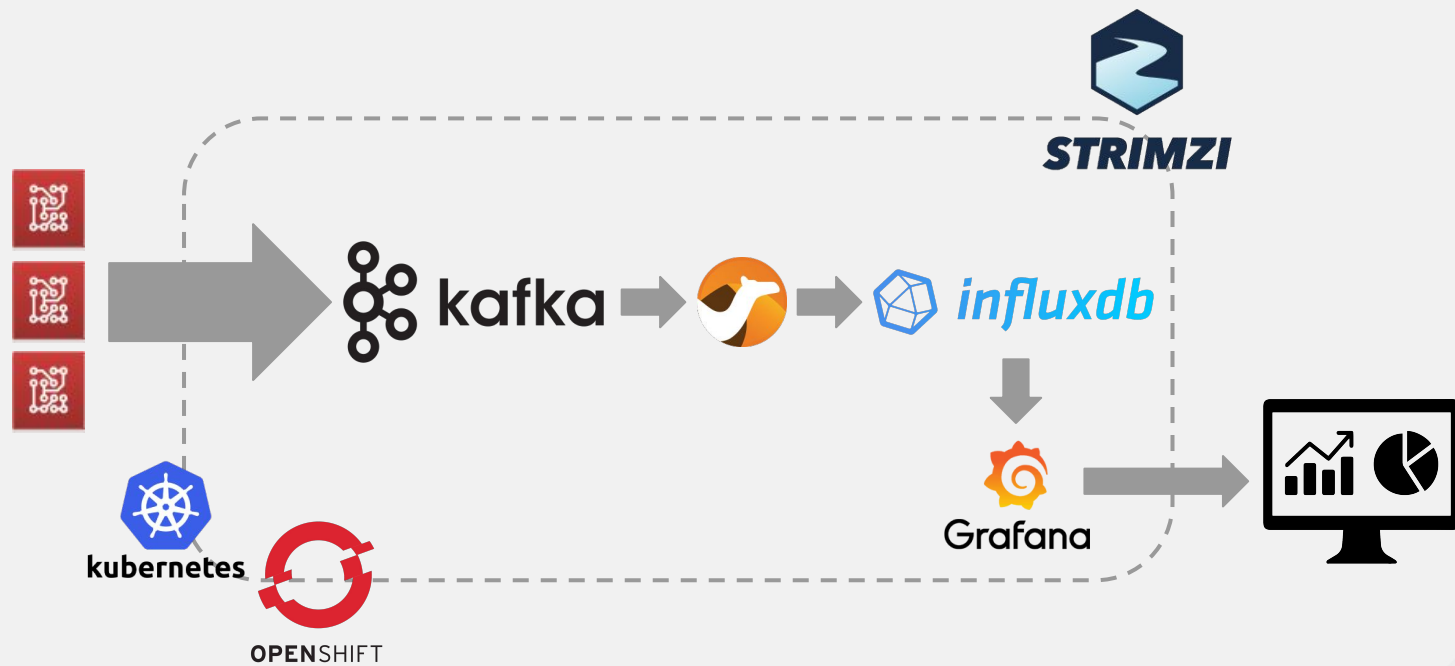
Zookeeper

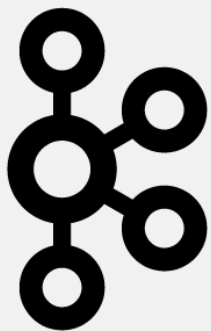
JVM  
Configuration

“Do you want to see this really working?  
Demo time!”

# DEMO

IoT devices monitoring





kafka



kubernetes

***STRIMZI***

# RESOURCES

Where I can find more?

- Strimzi : <https://strimzi.io/>
- OperatorHub.io : <https://www.operatorhub.io/>
- Apache Kafka : <https://kafka.apache.org/>
- Kubernetes : <https://kubernetes.io/>
- OpenShift : <https://www.openshift.com/>
- Operator framework : <https://github.com/operator-framework>
- Demo: <https://github.com/ppatierno/modern-integration-and-application-development-day-2019>

# THANK YOU!!



@ppatierno

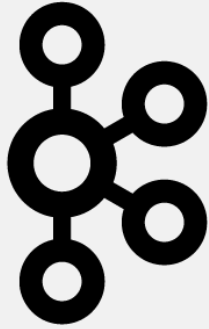


***STRIMZI***



**kubernetes**





**kafka**



**kubernetes**