Strimzi: running your **Apache Kafka** cluster on **Kubernetes**

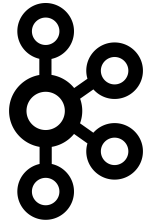
DevDay 2020

Paolo Patierno, Principal Software Engineer @Red Hat
@ppatierno

```
apiVersion: redhat/v1
kind: PrincipalSoftwareEngineer
metadata:
  name: Paolo Patierno
  namespace: Red Hat, Messaging & IoT team
  annotations:
    cncf/maintainer: Strimzi
    eclipse/committer: Vert.x, Hono & Paho
    microsoft/mvp: Azure
  labels:
    family: dad of two, husband of one
    sports: running, swimming, motogp, vr46, formula1, ssc napoli
    community: cncf napoli, devday
spec:
  replicas: 1
  containers:
      - image: patiernohub.io/paolo:latest
```

Apache Kafka & its ecosystem

STRIMZI

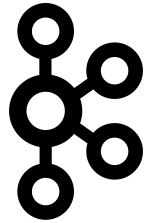


Apache Kafka

" ... a publish/subscribe messaging system ..."

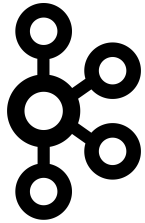
" ... a streaming data platform ..."

" ... a distributed, horizontally-scalable, fault-tolerant, commit log ..."



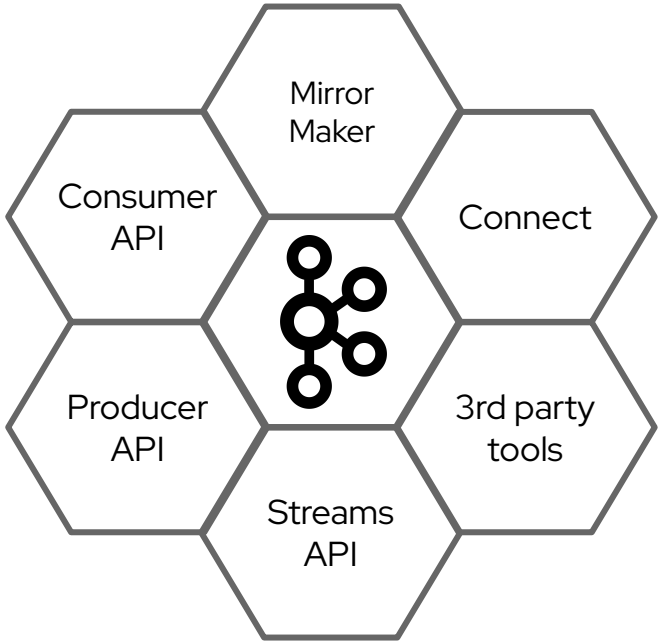
Apache Kafka

- Open Source project originally created by LinkedIn
- 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
- Broader ecosystem more than just the broker



Apache Kafka

- Broader ecosystem more than just the broker
  - Producer/Consumer clients
  - Streams API for real time processing
  - Syncing systems
  - Mirroring clusters



Kubernetes: the Linux kernel for the Cloud

STRIMZI

Kubernetes

" A system for ..."

" ... automating deployment ..."

" ... scaling ..."

" ... management ..."

" ... of containerized applications ..."

" It's like a Linux kernel ... but for distributed systems"

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

Container scheduling

Self healing

Secret & configuration management

Service discovery

Horizontal scaling

Load balancing

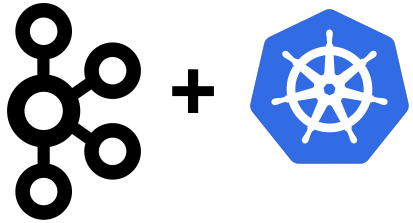
Storage orchestration

Automated rollout/rollback

Batch execution

Why running Kafka on Kubernetes?

STRIMZI

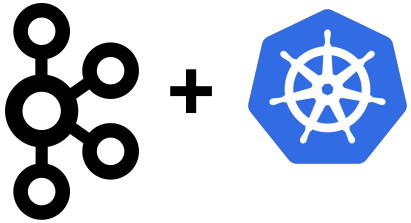


Kafka on Kubernetes

- Apache Kafka
  - distributed by nature
  - workloads using Apache Kafka are also distributed and scalable
- Kubernetes
  - great abstraction for running software everywhere
  - enables cloud-native development
- Why not using the Kubernetes knowledge to run Apache Kafka?

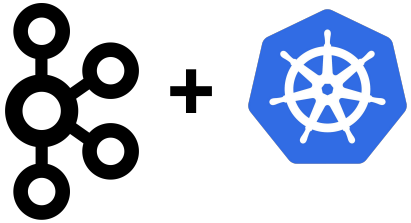
… but how many challenges!?

STRIMZI



Challenges

- 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

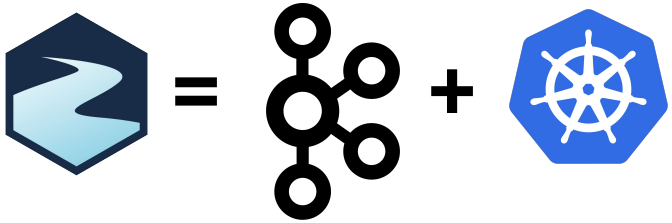


Challenges

- 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
  - ... Ingress for enabling access from outside
  - ... NetworkPolicy for defining access rules
  - ... and many more ...

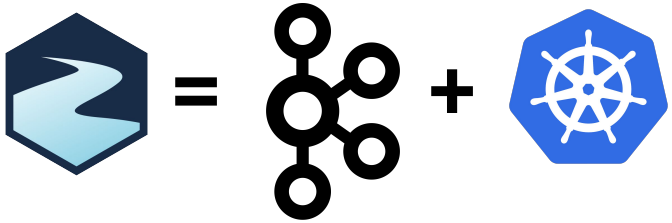
Let’s make it easy!

STRIMZI



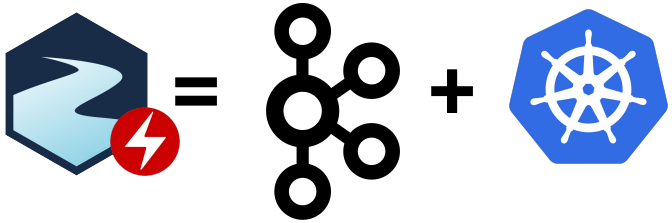
Strimzi

- Open source project licensed under Apache License 2.0
- Focuses on running Apache Kafka on Kubernetes
  - Container images for Apache Kafka, Apache ZooKeeper and other components
  - Operators for deploying, managing and configuring Kafka clusters
- Provides a Kubernetes-native experience
  - Not only Kafka clusters, but also users, topics and the rest of Kafka ecosystem
- CNCF sandbox project since September 2019



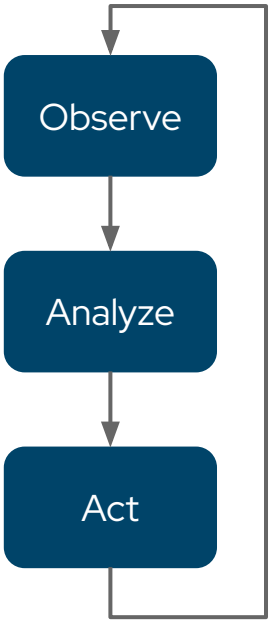
Kafka Resources

- Kafka is a Kubernetes native resource
- … as well as the other components
  - KafkaUser and KafkaTopic for handling users and topics
  - KafkaConnect and KafkaConnector for handling a Kafka Connect deployment
  - KafkaBridge for enabling HTTP access to the cluster
  - KafkaMirrorMaker and KafkaMirrorMaker2 for mirroring data across clusters
  - KafkaRebalance for rebalancing the cluster through Cruise Control



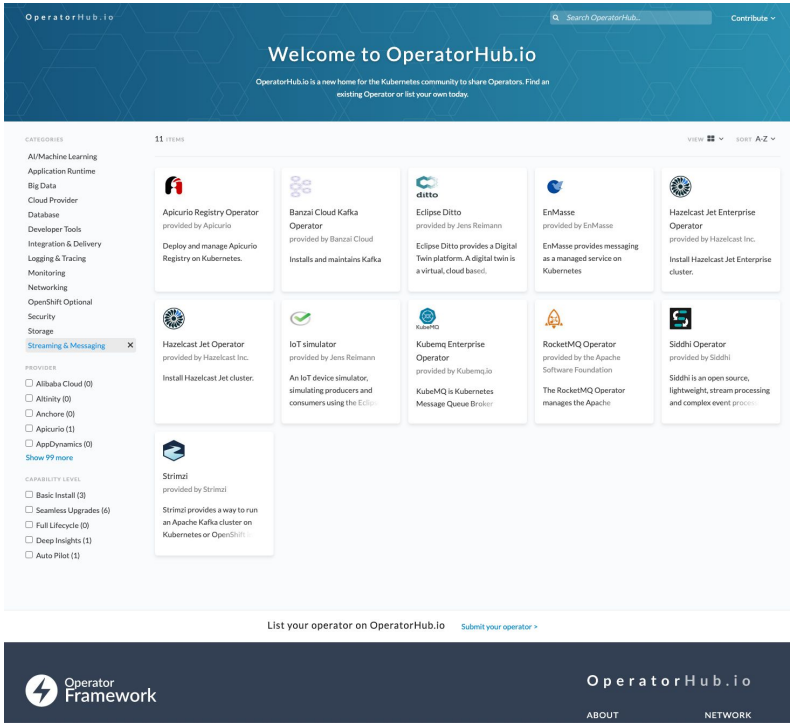
Operator

- 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, …

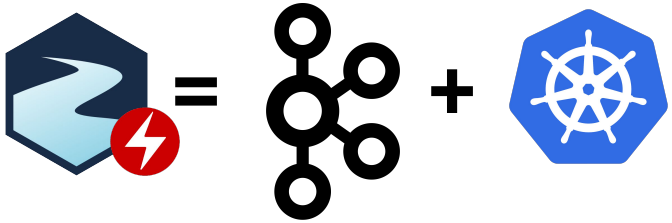


OperatorHub.io

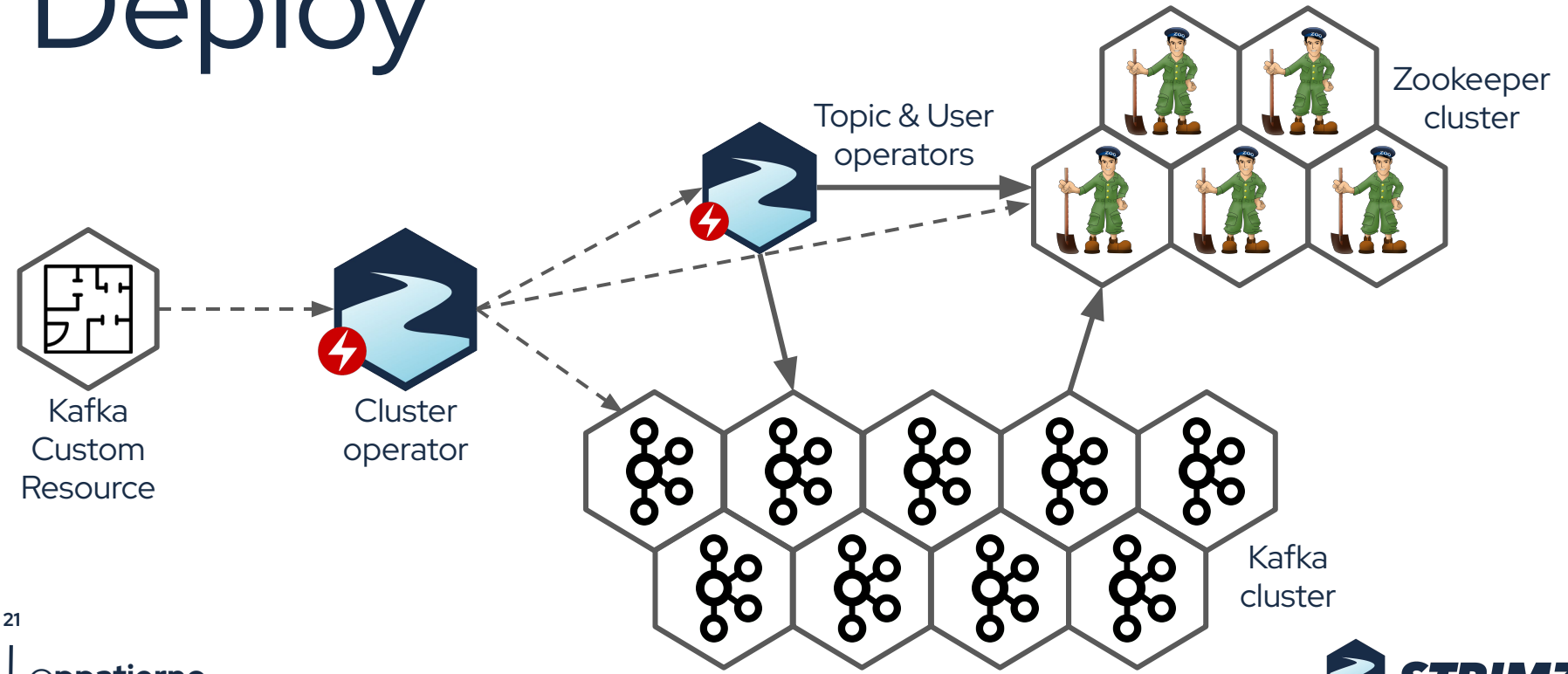
- Home for Kubernetes operators
  - A lot of categories (Database, Streaming & messaging, Logging & Tracing, …)
  - Installation via Helm Charts or YAML files
  - You can develop your own and provide to the community
  - https://operatorhub.io/



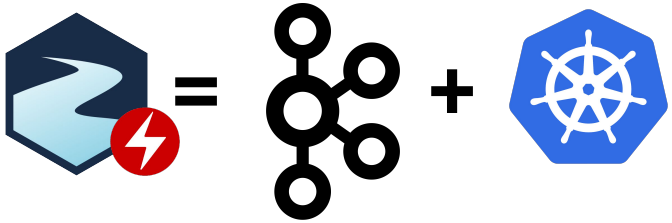
STRIMZI



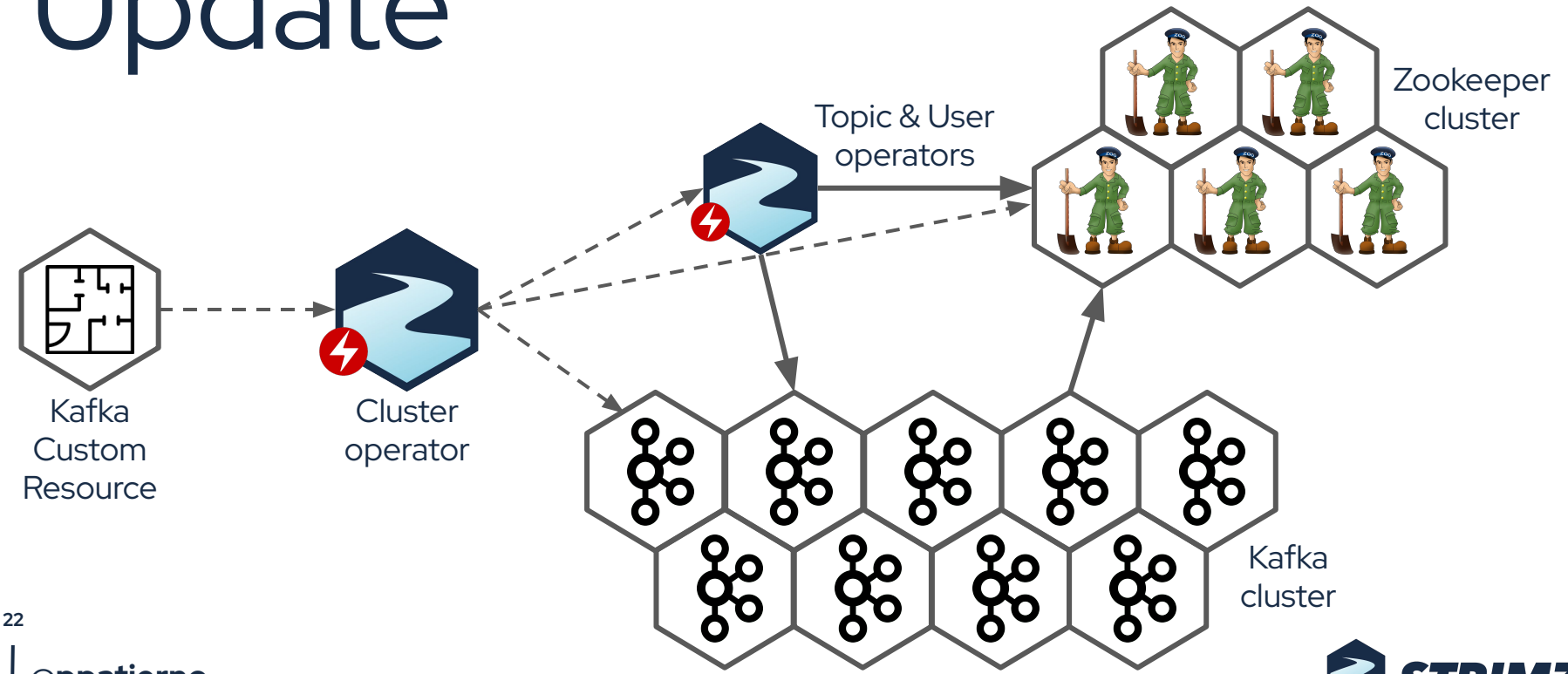
Deploy



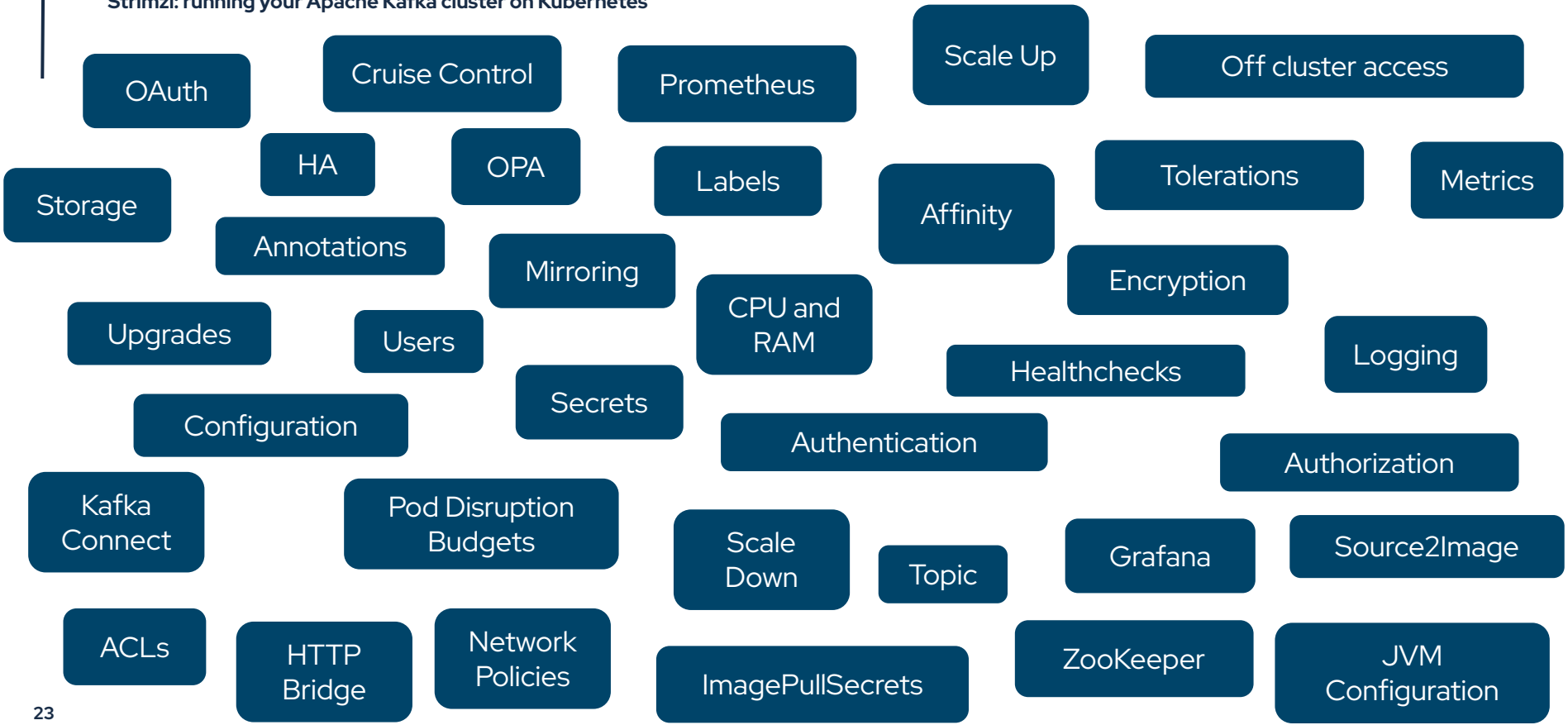
STRIMZI



Update



STRIMZI


OAuth
Cruise Control
Prometheus
Scale Up
Off cluster access
HA
OPA
Labels
Tolerations
Metrics
Storage
Affinity
Annotations
Mirroring
Encryption
CPU and RAM
Upgrades
Users
Logging
Healthchecks
Secrets
Configuration
Authentication
Authorization
Kafka Connect
Pod Disruption Budgets
Scale Down
Topic
Grafana
Source2Image
ACLs
HTTP Bridge
Network Policies
ImagePullSecrets
ZooKeeper
JVM Configuration

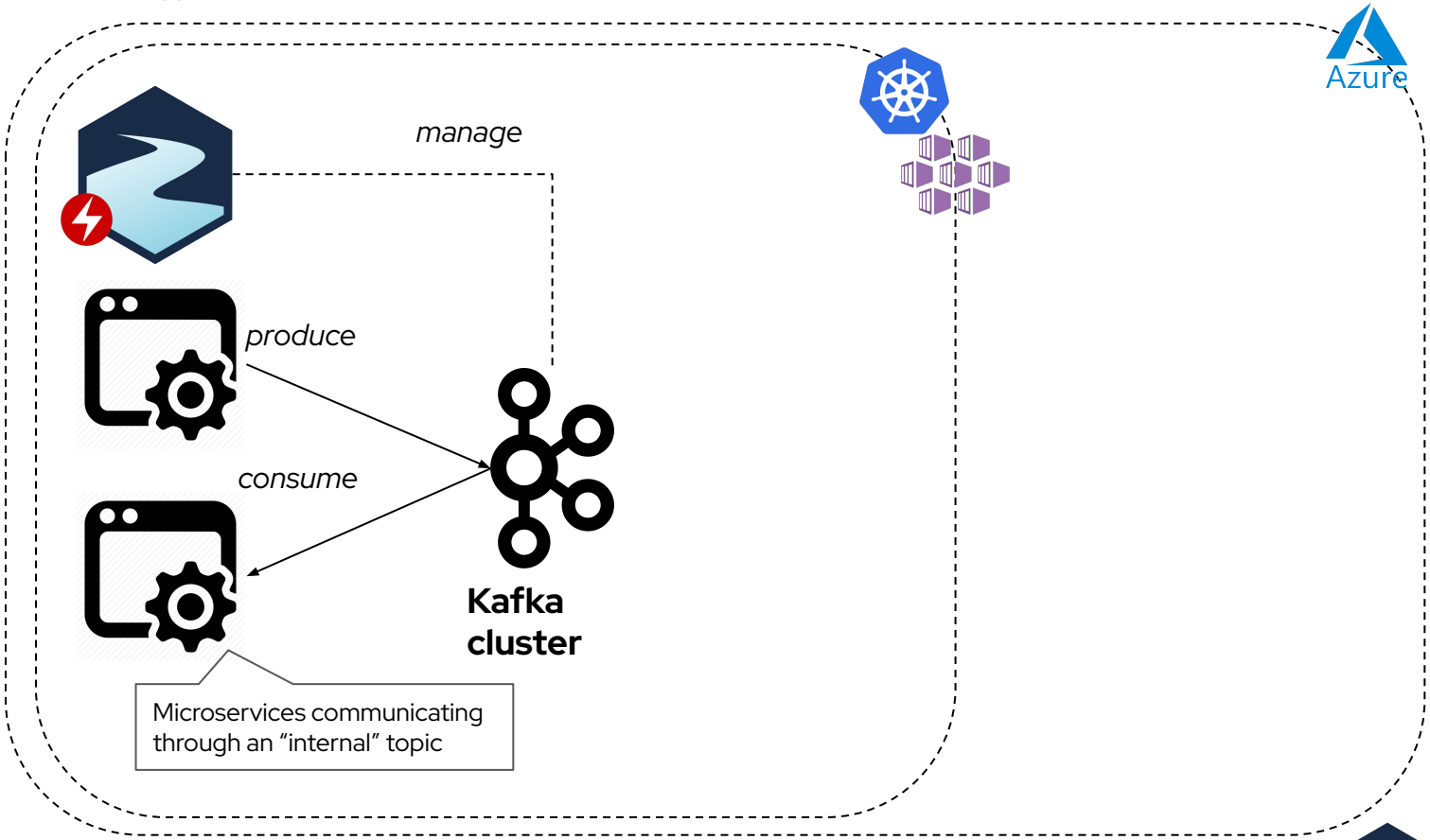
STRIMZI

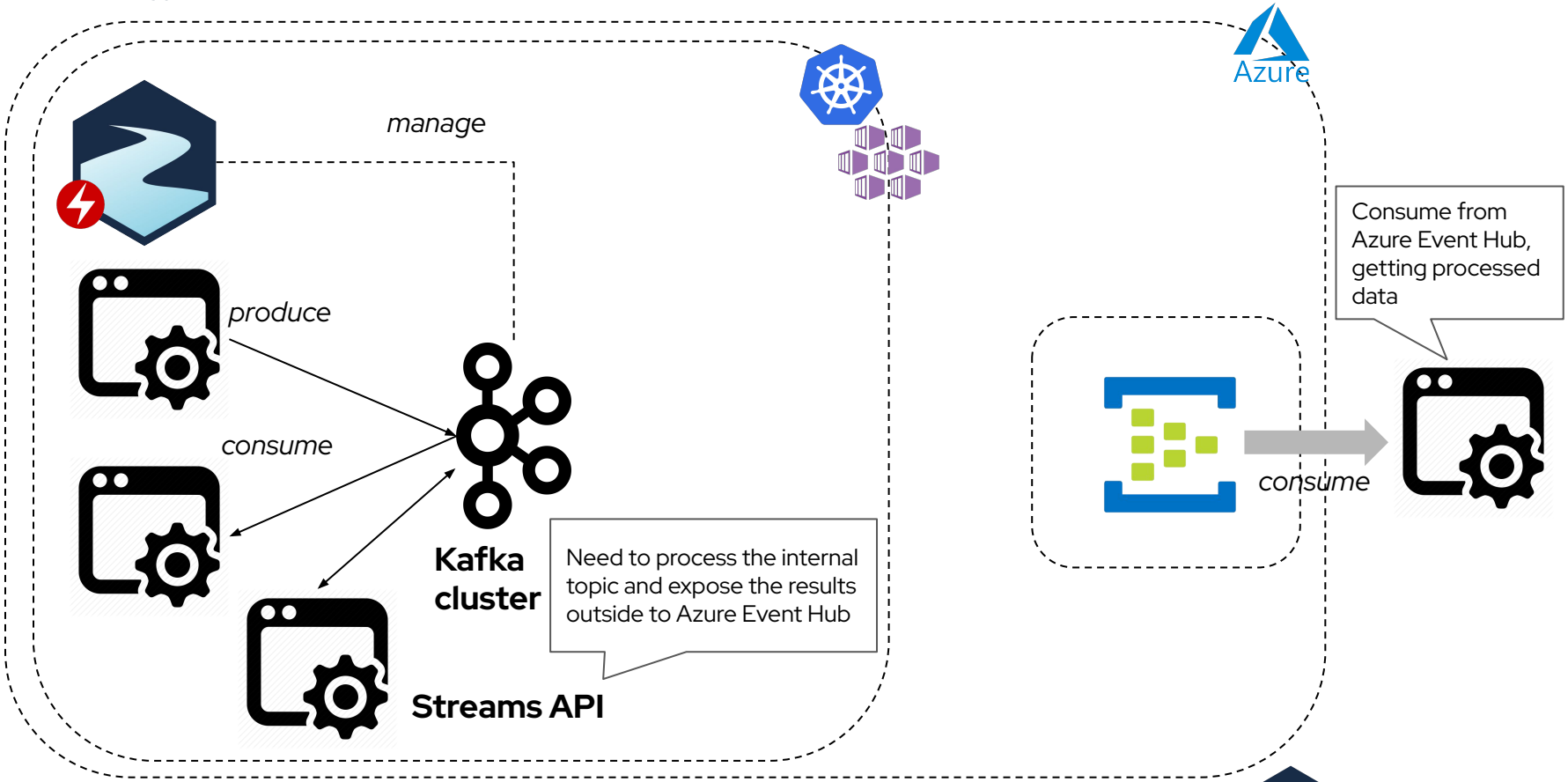
Demo

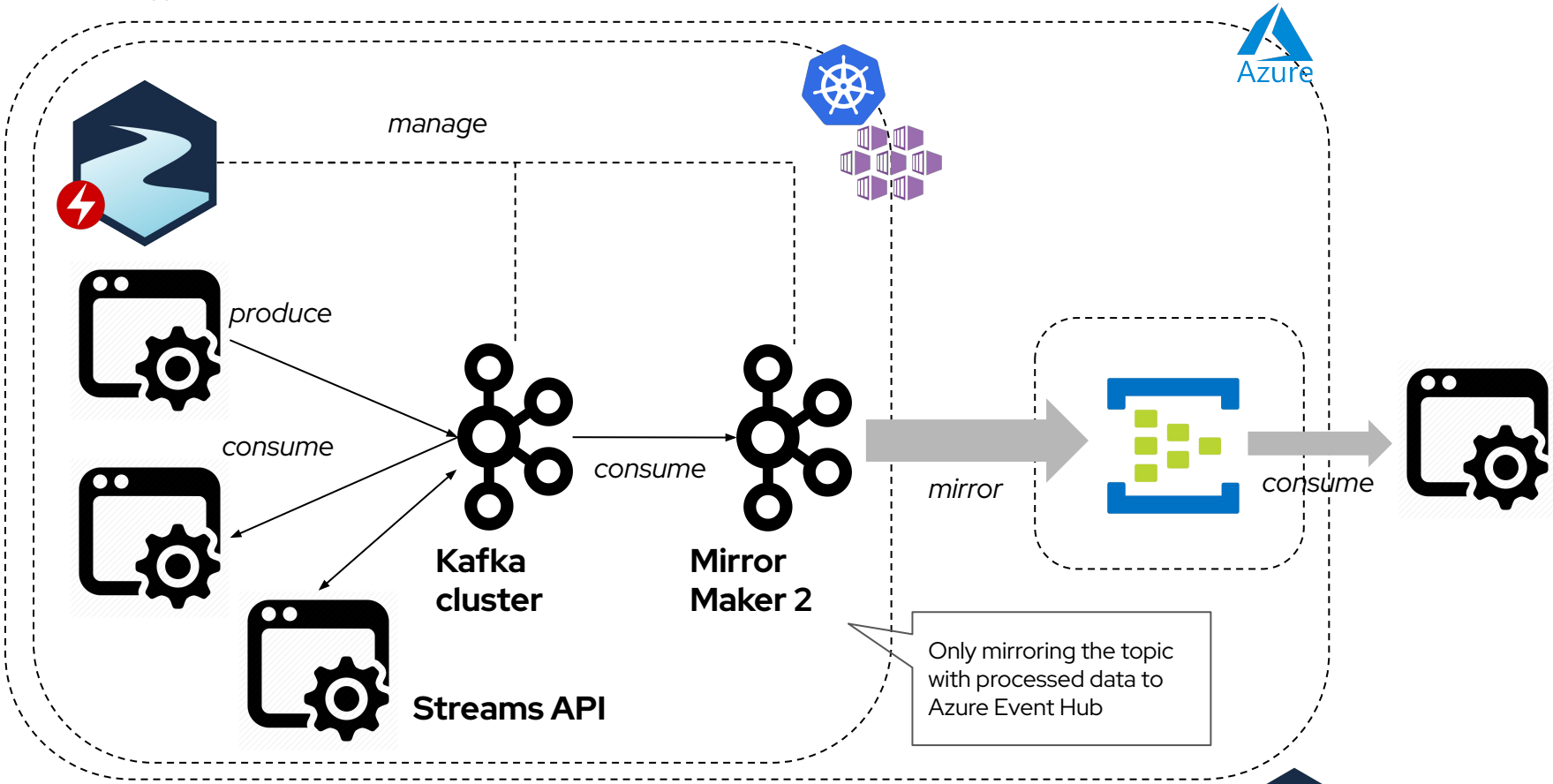
https://github.com/ppatierno/devday-2020-strimzi-aks-eventhub

- Kafka cluster deploy
- Kafka topics
- Mirroring to Azure Event Hub
- Kafka producer, Event Hub consumer

STRIMZI


Azure
manage
produce
consume
Kafka
cluster
Microservices communicating
through an “internal” topic


Azure
manage
produce
consume
Kafka cluster
Need to process the internal topic and expose the results outside to Azure Event Hub
Streams API
Consume from Azure Event Hub, getting processed data
consume


Azure
manage
produce
consume
consume
mirror
consume
Kafka cluster
Mirror Maker 2
Streams API
Only mirroring the topic with processed data to Azure Event Hub

Strimzi's friends ...

STRIMZI

kubernetes

JAEGER

Prometheus

KEDA

OPENTRACING

fluentd

Open Policy Agent

STRIMZI

Reach the community


https://strimzi.io


https://github.com/strimzi


@strimziio


#strimzi on the https://slack.cncf.io


cncf-strimzi-users@lists.cncf.io

Thank you

STRIMZI