zalando

Operating Elasticsearch in Kubernetes

microxchg 2019

MIKKEL LARSEN

@mikkeloscar

OLIVER TROSIEN

@otrosien

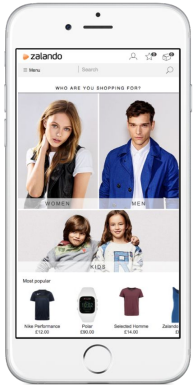
2019-04-02

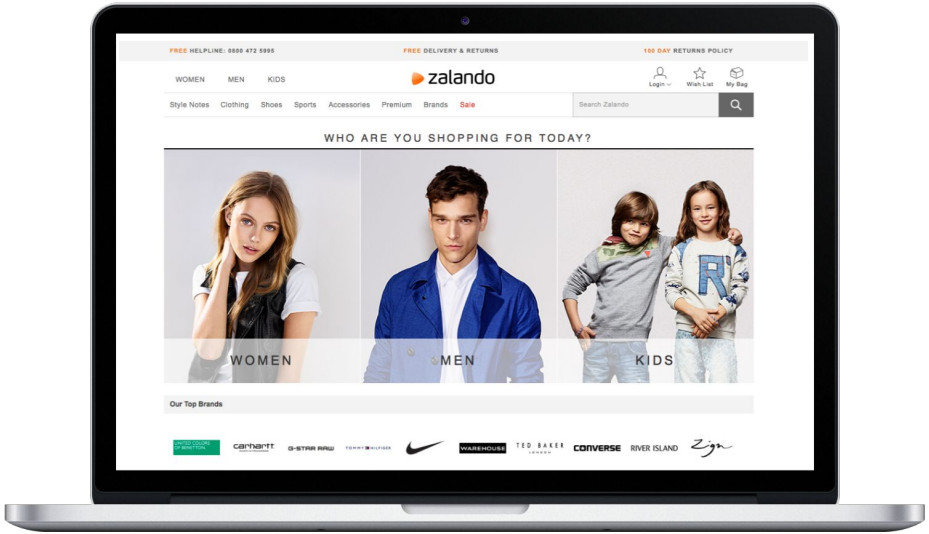
WHO ARE WE?

Oliver
Core Search Services
Elasticsearch

Mikkel
Cloud Infrastructure
Kubernetes

“EUROPE’S LEADING ONLINE FASHION PLATFORM”





zalando

WE BRING FASHION TO PEOPLE IN 17 COUNTRIES

17 markets

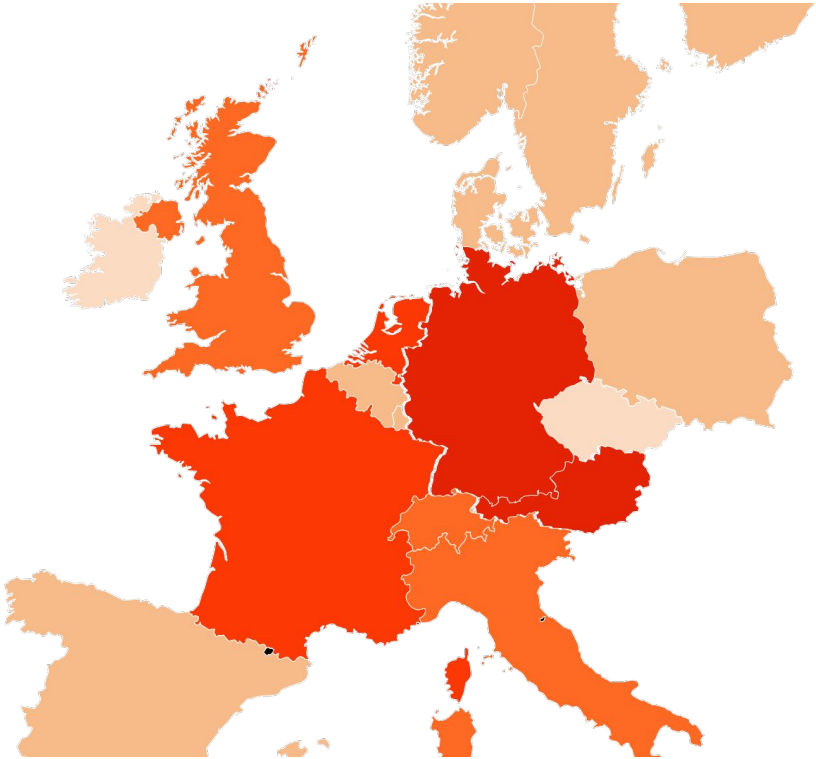
7 fulfillment centers

26 million active customers

5.4 billion € revenue 2018

250 million visits per month

15,000 employees in Europe



zalando

SEARCH @ ZALANDO

300k+
Products
per country

~2000
Brands

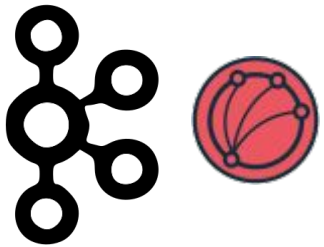
~700
Categories

45%
Mobile Traffic

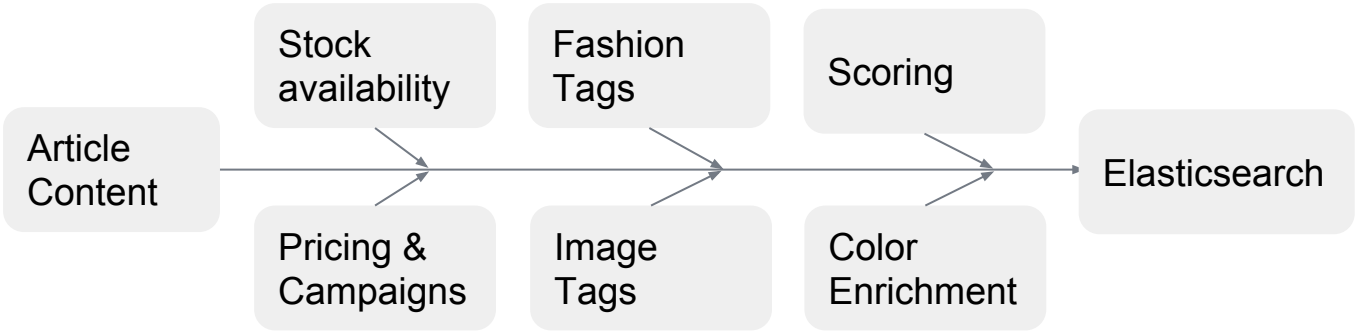
12K
QPS

8K
Updates/s

DATA INGESTION



Kafka / Nakadi *



* nakadi.io

zalando

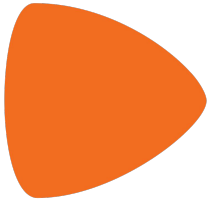
Search Overall
Zoom Out
Last 6 hours
Origami Max Search RPS
24988
4xx / 5xx
ES Cluster State
green
ES Active Shards
1391
ES Nodes
512
ES Indices
100
ORIGAMI
Clients to Origami RPS
CatalogAPI
Reco
Minerva
Clients to Origami Latency
minerva
reco
catalog-api
outfits
fsapi
Origami ELB Latency
latency
instances
p99.9 Max: 416 ms
p99 Max: 159 ms
p95 Max: 83 ms
p90 Max: 59 ms
mean Max: 25 ms
RPS-per-CPU (prio0)
de-de Max: 2.02 ops Avg: 1.22 ops Current: 0.89 ops
CATALOG API
RPS-per-CPU (prio1)
max current
fr-fr 2.26 ops 0.88 ops
nl-nl 2.16 ops 0.94 ops
it-it 1.47 ops 0.97 ops
pl-pl 1.33 ops 0.73 ops
RPS-per-CPU (prio2)
max avg current
se-sv 2.89 ops 1.79 ops 1.46 ops
ch-de 1.69 ops 1.02 ops 0.75 ops
es-es 1.58 ops 0.95 ops 1.09 ops
RPS-per-CPU (prio3)
max avg current
be-nl 4 ops 2 ops 1 ops
be-fr 3 ops 1 ops 1 ops
no-no 2 ops 2 ops 2 ops
dk-da 2 ops 1 ops 1 ops
RPS-per-CPU (prio4)
max avg current
de-en 1.04 ops 0.69 ops 0.72 ops
cz-cs 0.42 ops 0.20 ops 0.09 ops
ch-it 0.41 ops 0.24 ops 0.19 ops
ie-en 0.14 ops 0.08 ops 0.10 ops
ELASTICSEARCH
CPU (prio0)
max avg
prio0b 21% 10%
prio0a 21% 8%
CPU (prio1)
max avg
prio1b 40% 16%
prio1a 37% 13%
CPU (prio2)
max avg
prio2b 41% 18%
prio2a 32% 14%
CPU (prio3)
max
94%
prio3a 15%

KUBERNETES @ ZALANDO

Default Deployment Target

1400~ nodes

Since **Oct 2016**

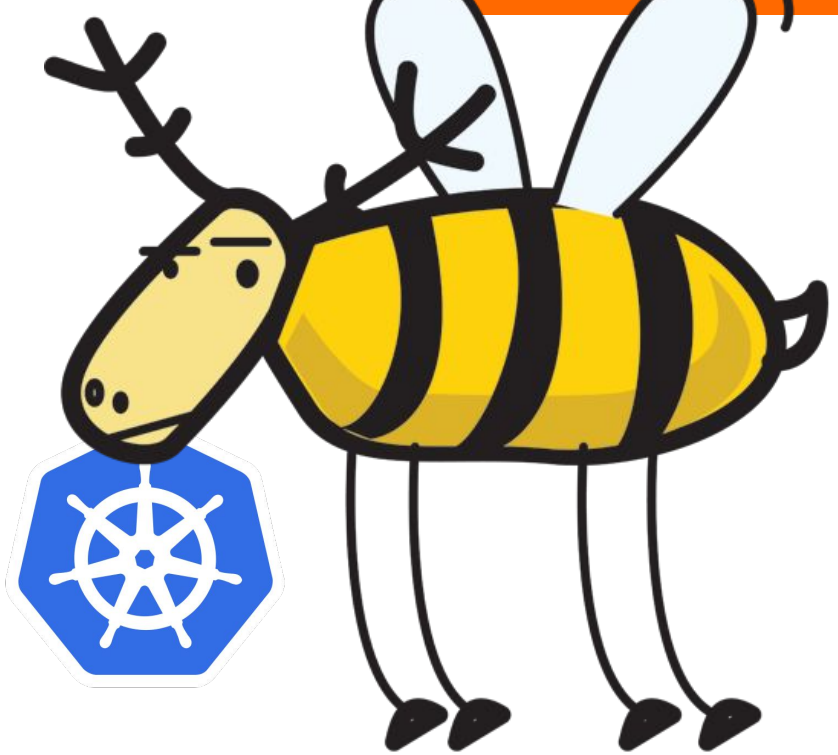


zalando

114 clusters

Node **Autoscaling**

From **v1.4** to **v1.12**



Elasticsearch
2.500 vCPUs
1 TB RAM

Elasticsearch in Kubernetes

RUNNING ELASTICSEARCH IN KUBERNETES

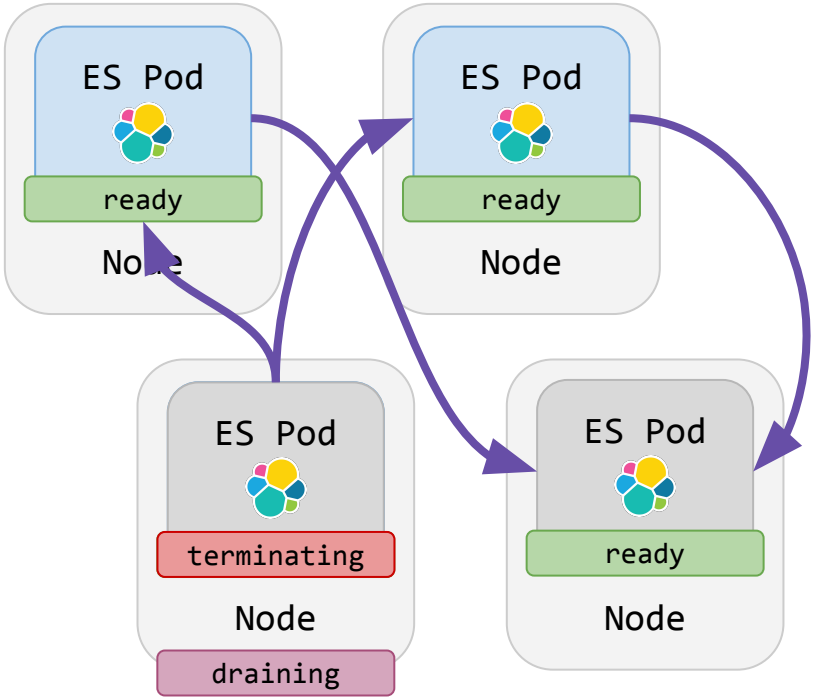
1. **Safe automatic updates**
   (Including Kubernetes cluster updates)

2. **Advanced auto-scaling for cost efficiency**

UPDATING ELASTICSEARCH (STATEFULSET)

1) PreStop Hook (bash script)
- Exclude node in ES
- Wait for node to drain (up to 1h)
- Data is moved to existing nodes

2) PostStart Hook (bash script)
- Remove all excludes
- Let ES rebalance from existing nodes



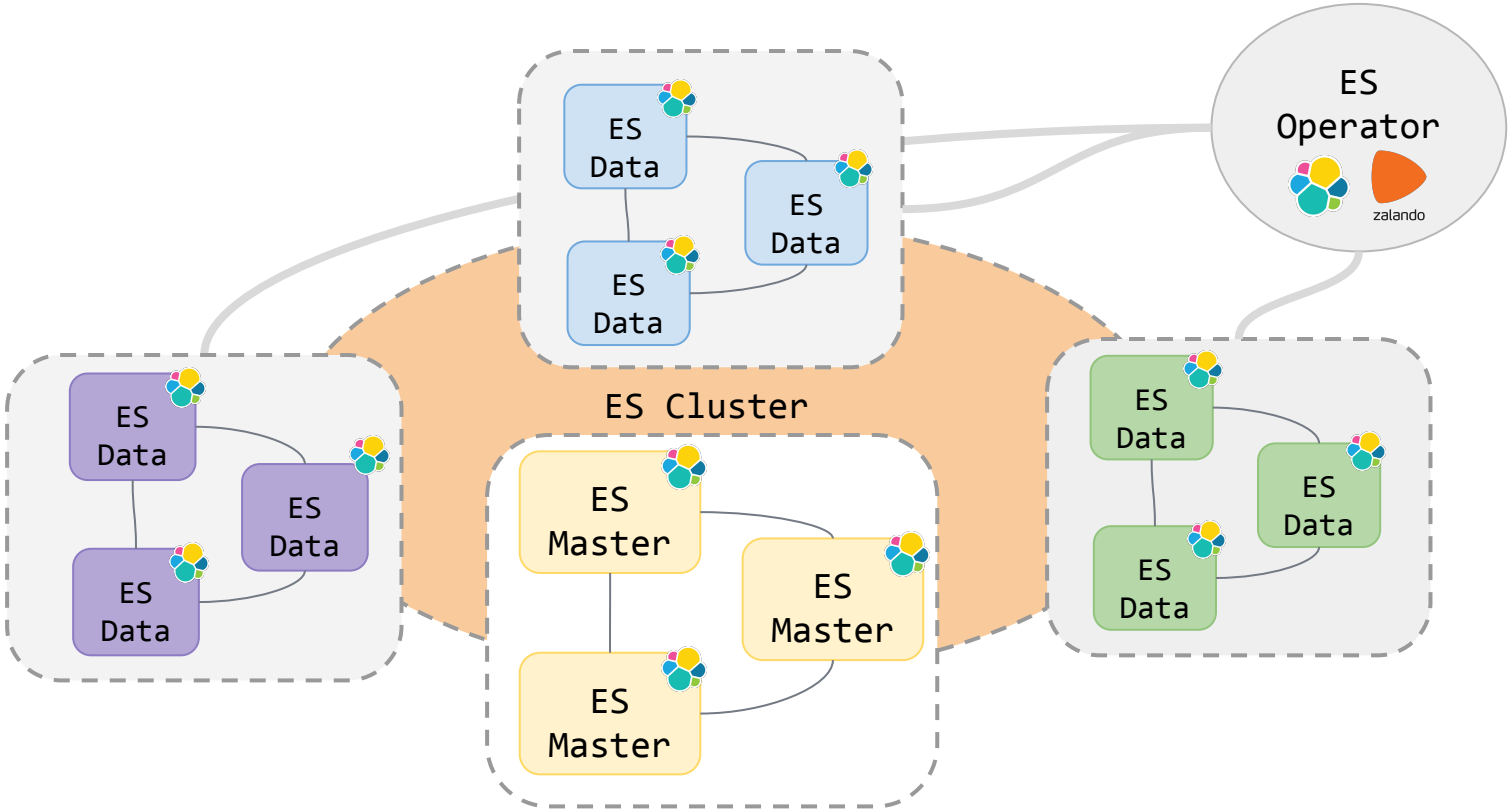
zalando

ELASTICSEARCH DATA SETS

```yaml
apiVersion: zalando.org/v1
kind: ElasticsearchDataSet
metadata:
  name: test-cluster
spec:
  scaling:
    {...}
  replicas: 3
  template: # PodTemplate
    {...}
  volumeClaimTemplates:
    {...}
```

github.com/zalando-incubator/es-operator

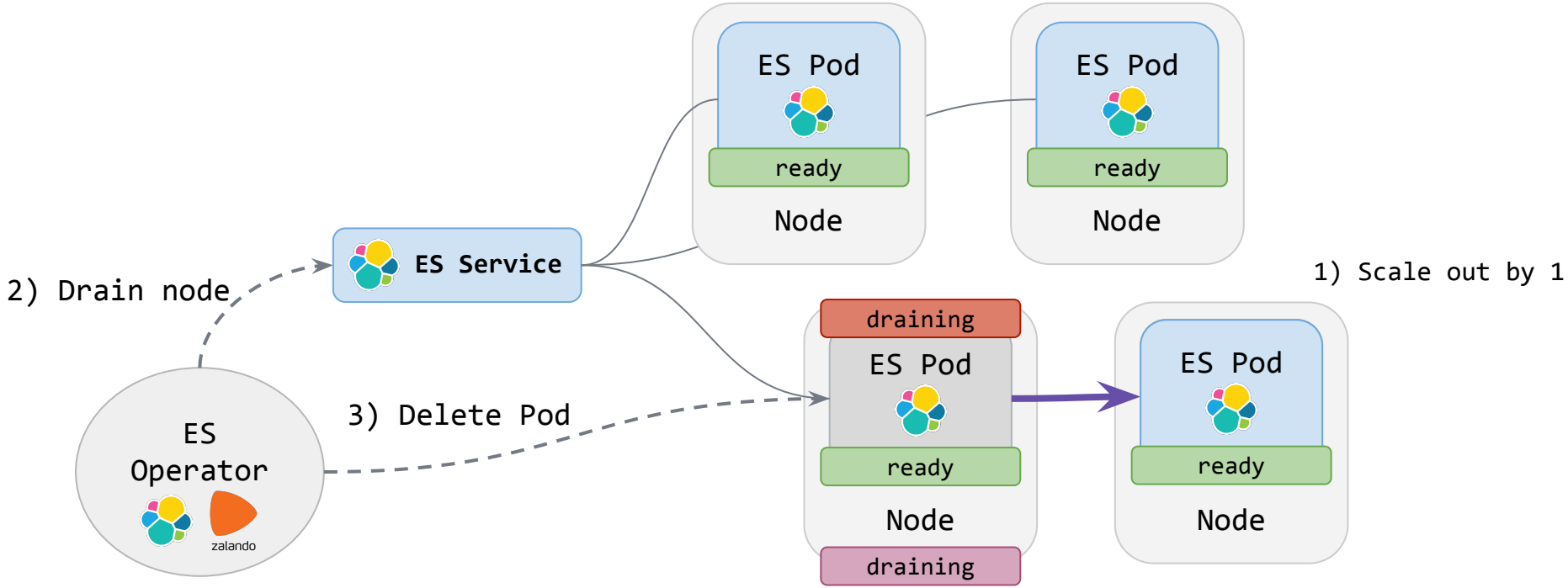
ELASTICSEARCH DATA SETS



github.com/zalando-incubator/es-operator

zalando

UPDATING ELASTICSEARCH (OPERATOR)

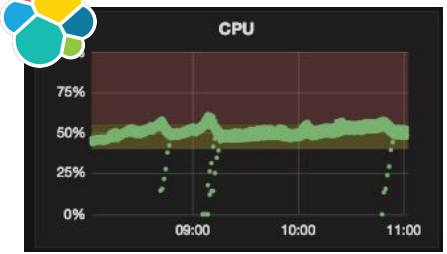


github.com/zalando-incubator/es-operator

zalando

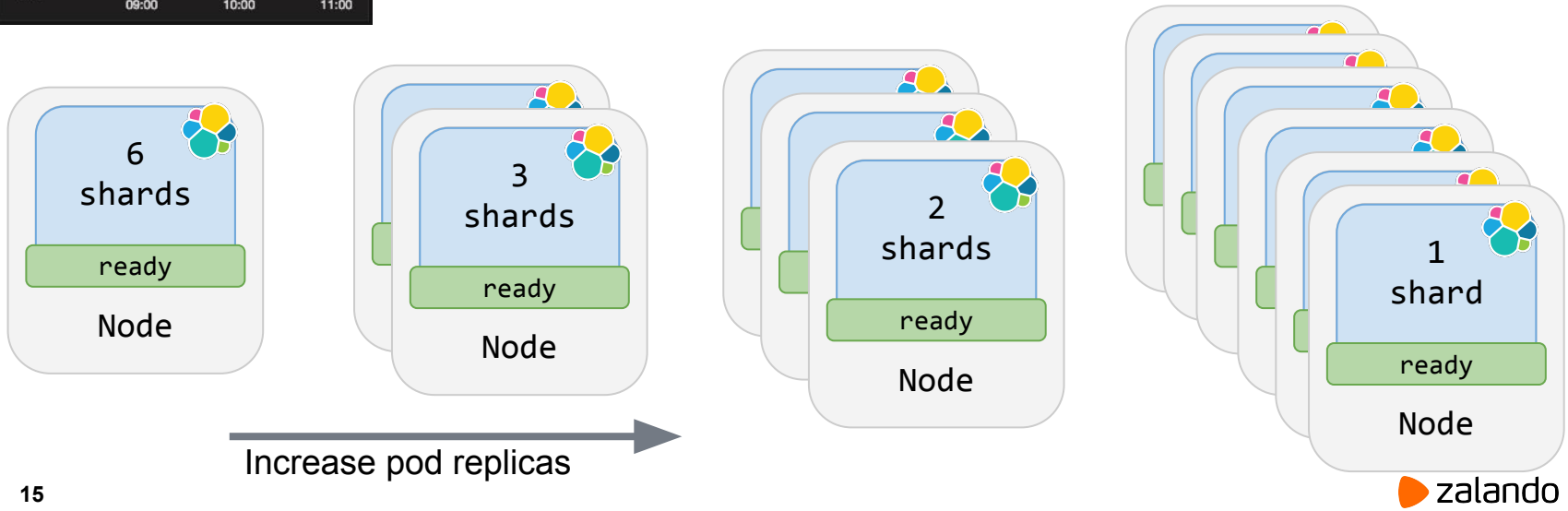
SCALING UP ELASTICSEARCH (1)

METRICS



Thresholds
- CPU
- Duration
- Cooldown

Boundaries
- Max # Pod replicas

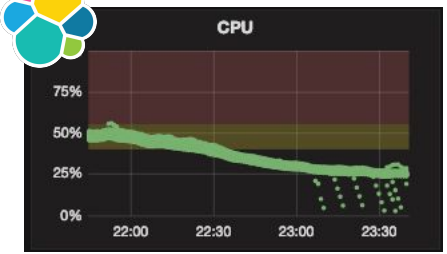


zalando

SCALING DOWN ELASTICSEARCH

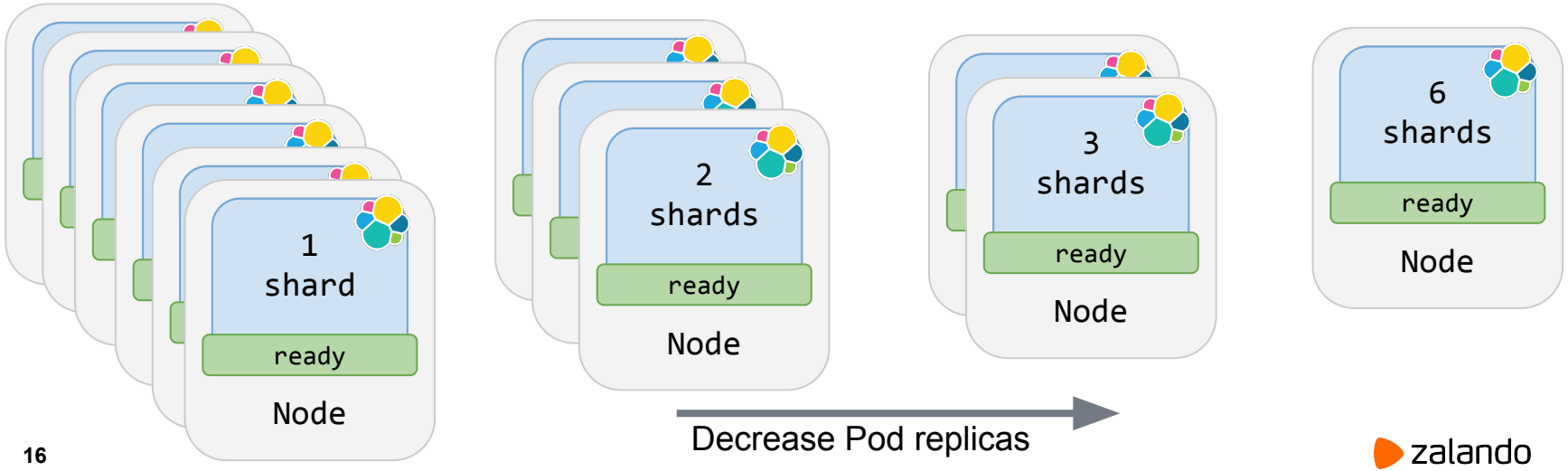
DON'T OPERATE WHEN CLUSTER IS NOT GREEN!

METRICS



Thresholds
- CPU
- Duration
- Cooldown

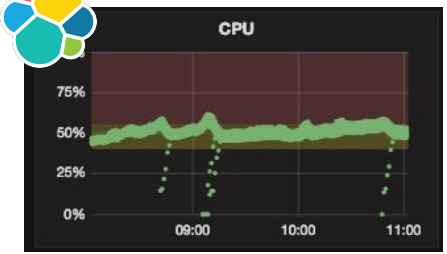
Boundaries
- Min # Replica
- Max # Shards per node
- Max disk usage (%)



zalando

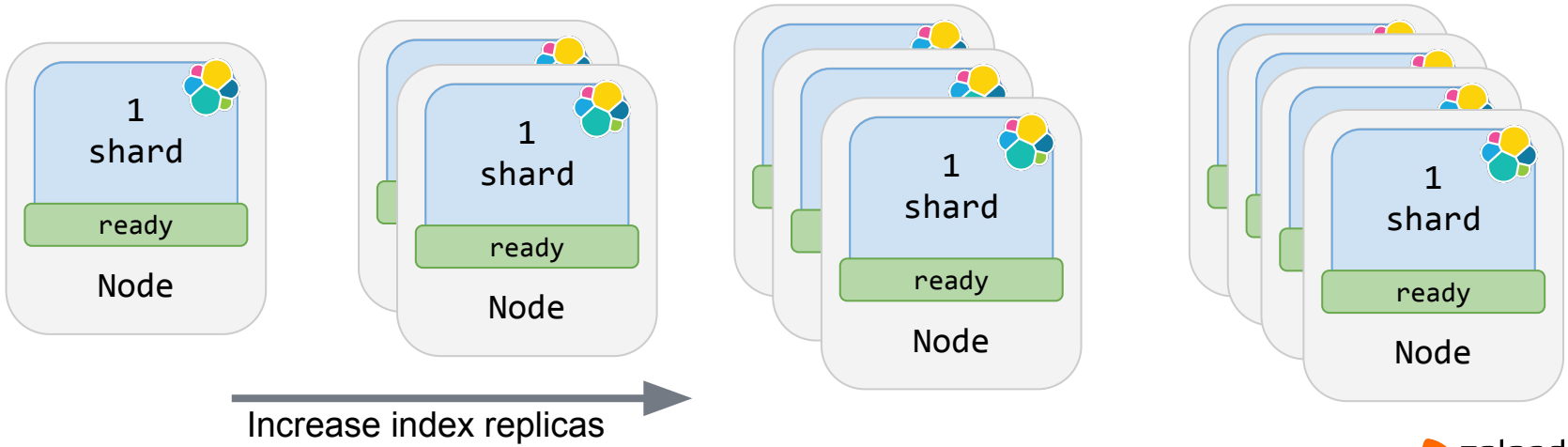
SCALING UP ELASTICSEARCH (2)

METRICS



Thresholds
- CPU
- Duration
- Cooldown

Boundaries
- Min # Shards per node
- Max # Pod replicas



zalando

DEMO

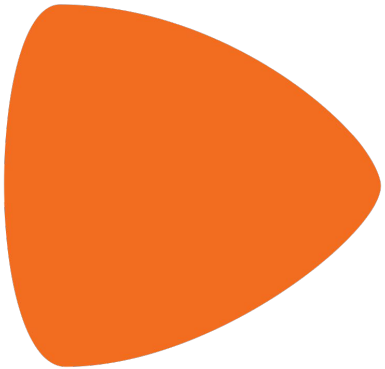
OPEN SOURCE

ES Operator
github.com/zalando-incubator/es-operator

microXchg demo
github.com/otrosien/microxchg19-demo

Nakadi
nakadi.io

Kubernetes on AWS
github.com/zalando-incubator/kubernetes-on-aws



zalando

MIKKEL LARSEN

mikkel.larsen@zalando.de

@mikkeloscar

OLIVER TROSIEN

oliver.trosien@zalando.de

@otrosien

2019-04-02