Cloud Native Bern Meetup

Keda
Mastering Scale-to-Zero

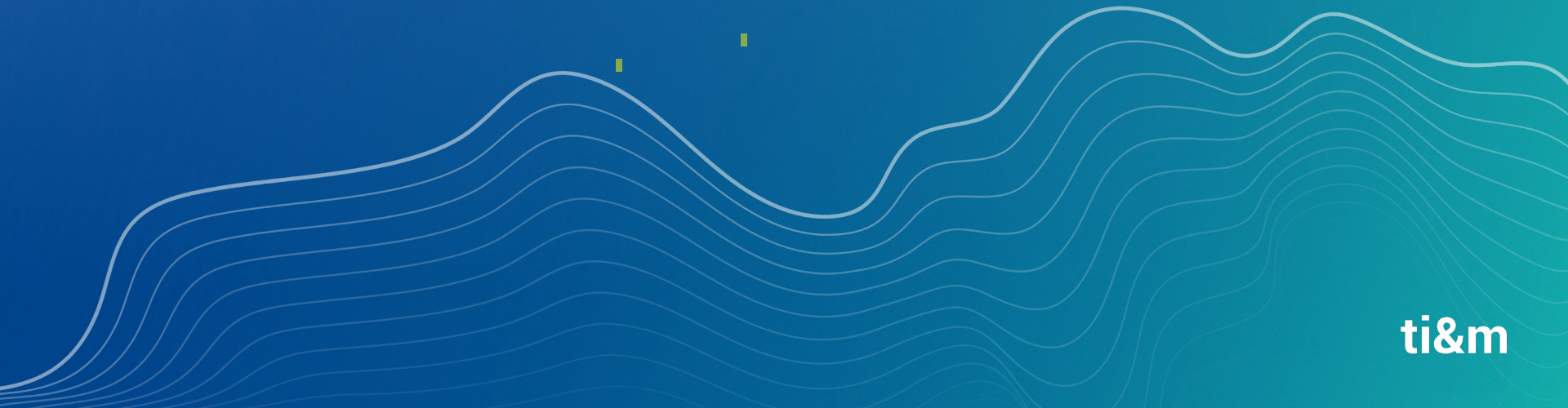
Daniel Hasselwander

Bern, 17.11.2022

ti&m

ti&m

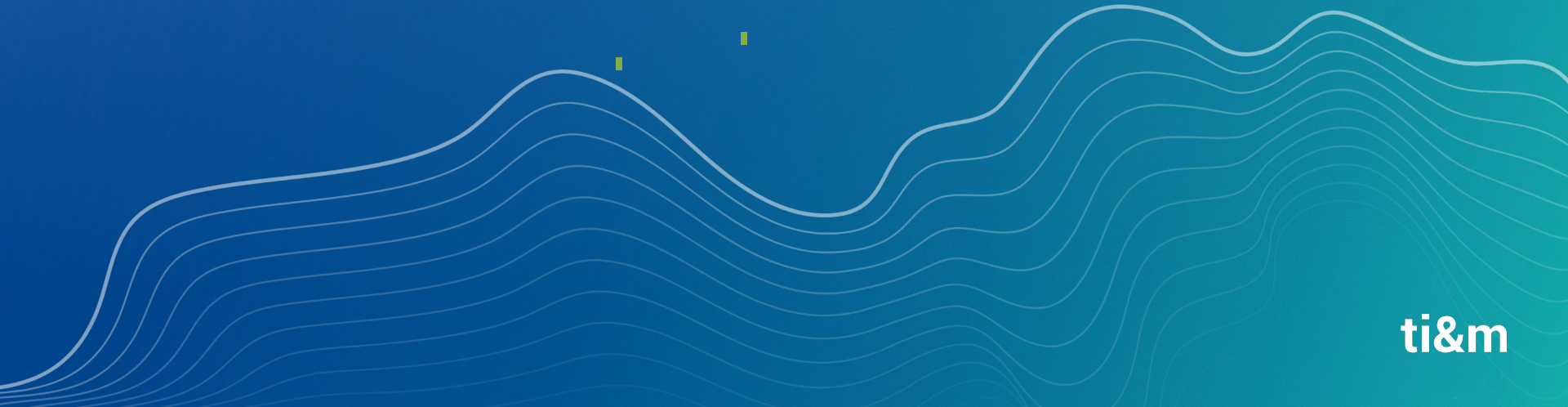
Each person leaves the house and goes to work



ti&m

ti&m

Same Picture – another Situation



ti&m

Cluster 1 Cluster 2 Cluster 3 Cluster 4

ti&m

It is Sunday, now

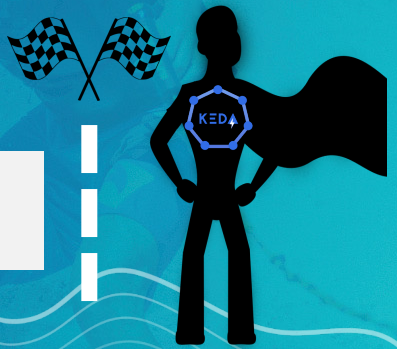
ti&m

Cluster 1 Cluster 2 Cluster 3 Cluster 4

ti&m

Our Path

What we want - What is Keda - How it works - Demo



Who am I ?

Azure

ti&m

big ideas. creative technology.

What we want ?

Scale-To-Zero

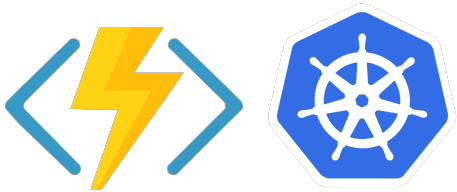
1 → 0

or

0 → 1

Scale Workload on external metric

Run Azure Functions in our cluster



What we have ?

Scale-To-1

Struggle with CPU und Ram Scale

What is Keda ?

– Kubernetes-based Event Driven Autoscaler

– Works alongside standard components

– Installable with helm

– 56 Built-in Scalers

How it works ? KEDA Parts

Agent – Scale to Zero

activate and deactivate Deployments

«keda-operator» container

Metrics Server

exposes event data for HPA to scale out

«keda-operator-metrics-apiserver» container

How it works ? Custom Ressources (CRD)

ScaledObjects

Events ➡ Deployment etc.

ScaledJobs

Events ➡ Kubernetes Job

TriggerAuthentication

Namespace useable credentials

ClusterTriggerAuthentication

Cluster useable credentials

Demo

– https://github.com/Hassel93/cn_meetup_keda

Cluster 1 Cluster 2 Cluster 3 Cluster 4

ti&m

KEDA – When to use it

– Should be your default scale mechanism

– Scale-To-Zero != Realtime

– Azure Function in your Cluster

KEDA – Good to know

– Don´t combine ScaleObject with HPA
– Can use multiple scaler
– HTTP Scaler are experimental
– Azure CLI implement a KEDA extension for Functions

Thank you for your attention !

Questions ?