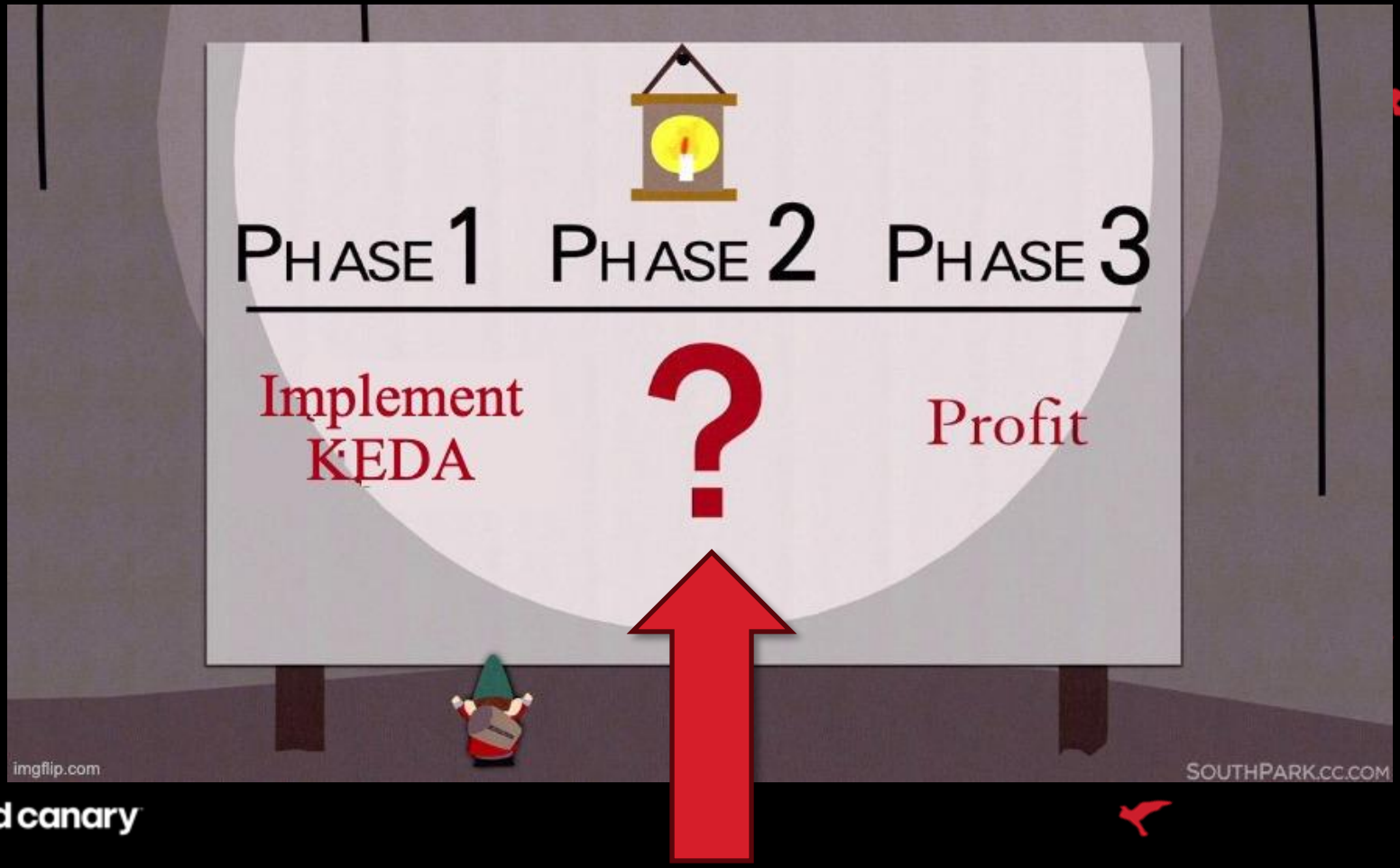
Is Everyone O-KEDA?

"Exciting" lessons learned in our journey
to use KEDA Pod Autoscaling

Brian Davis
Principal Software Engineer, Red Canary

@brianthedavis

red canary

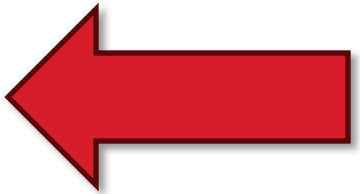

PHASE 1
PHASE 2
PHASE 3
Implement KEDA
?
Profit
imgflip.com
SOUTHPARK.CC.COM

red canary

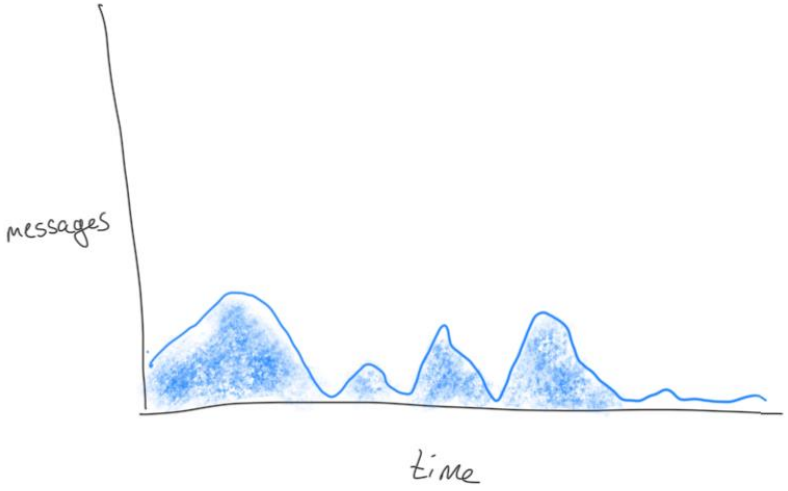
Five Phases of adopting KEDA

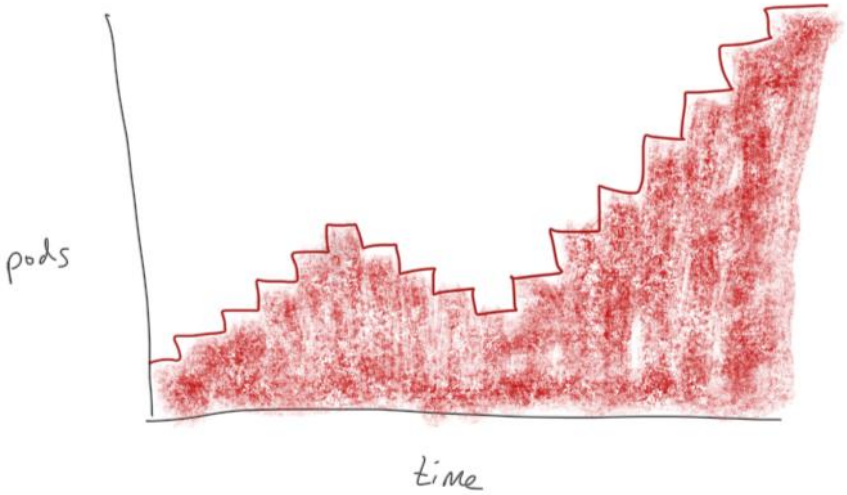
red canary

1. Accepting your current auto scaling is no good
2. Figuring out how to pronounce it
3. Implement it
4. Trough of despair
5. Profit



Our old scaling



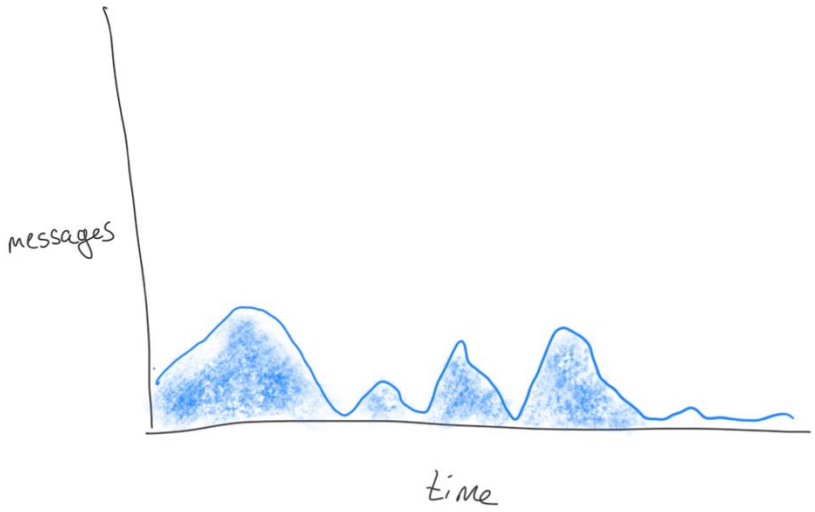


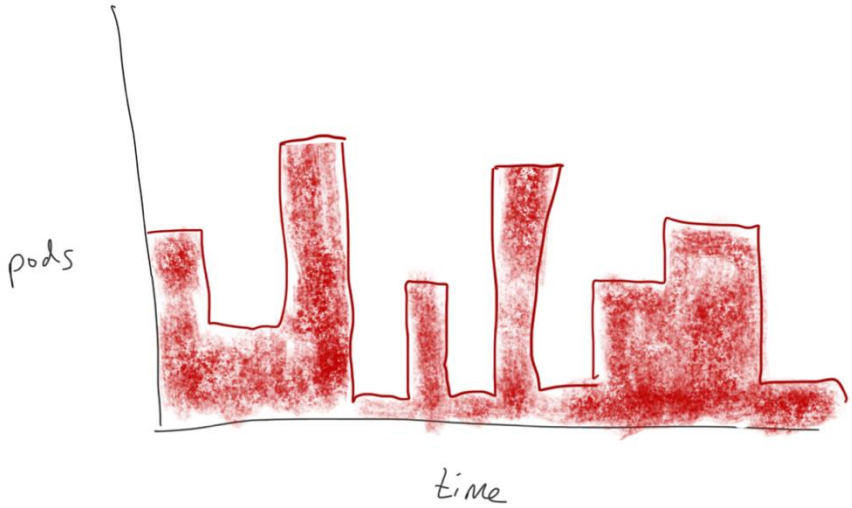
Zero to N Scaling:
One of the coolest features of KEDA is its ability to scale your application back to zero when there are no events to process. This can lead to significant cost savings.

Saul Fernandez - https://dev.to/sarony11/hpa-vs-keda-in-kubernetes-the-autoscaling-guide-to-know-when-and-where-to-use-them-m96

The new hotness

red canary

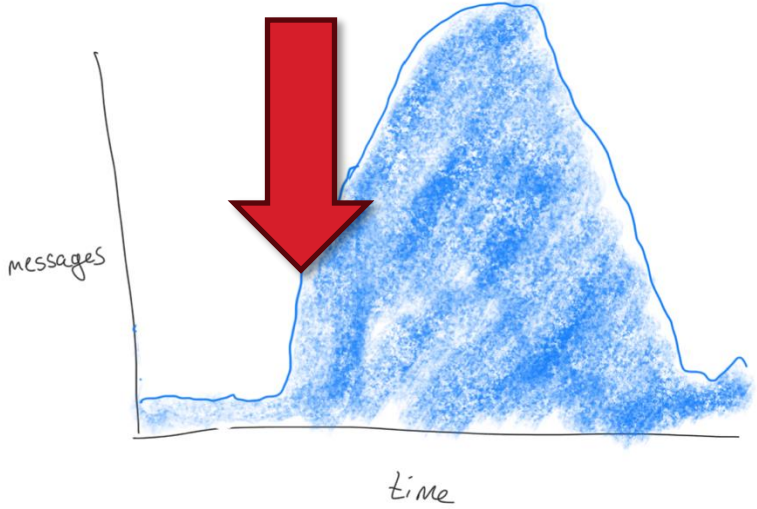


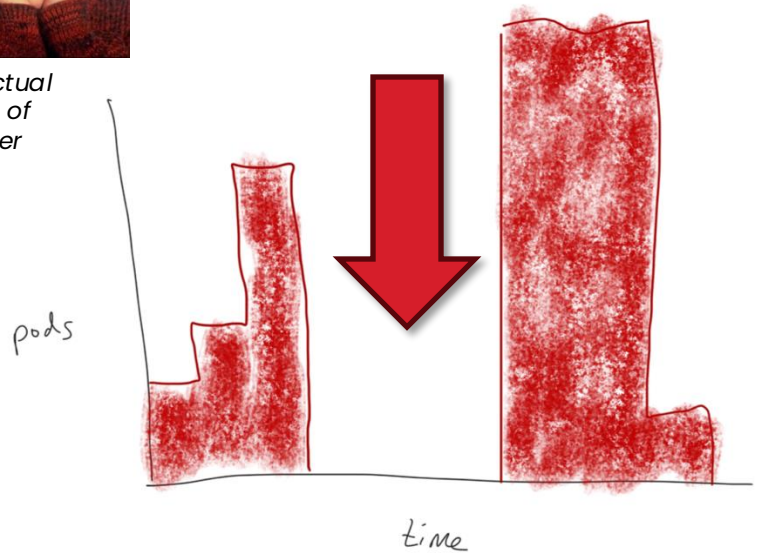


...until

*Not actual picture of engineer

Engineer gets paged at 3am





Zero to N Scaling:
One of the coolest features of KEDA is its ability to scale your application back to zero when there are no events to process **OR THE K8S METRICS SERVER GETS EVICTED**.

"I'll take scalability for 1000, Alex"

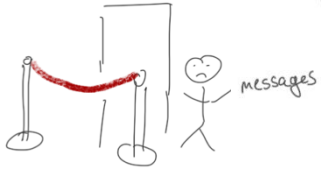
- AWS STS Rate Limiting
- AWS S3 Rate Limiting

Old scaling patter
(pyramids)

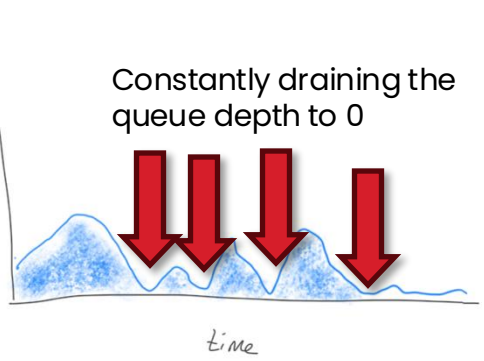
New scaling pattern
(skyscrapers)

Tuning KEDA

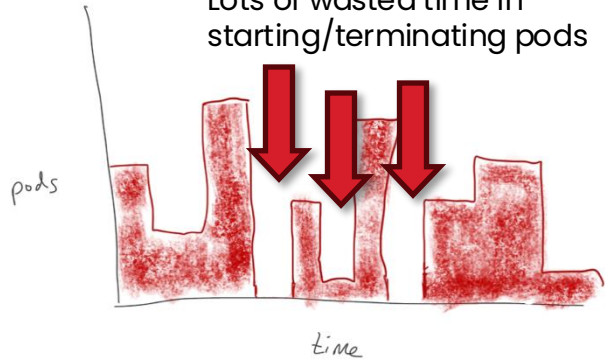
Make messages wait around for *a little bit* to smooth out the peaks and valleys and waste less resources



Constantly draining the queue depth to 0

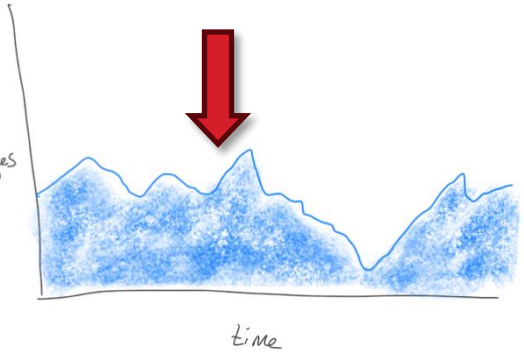


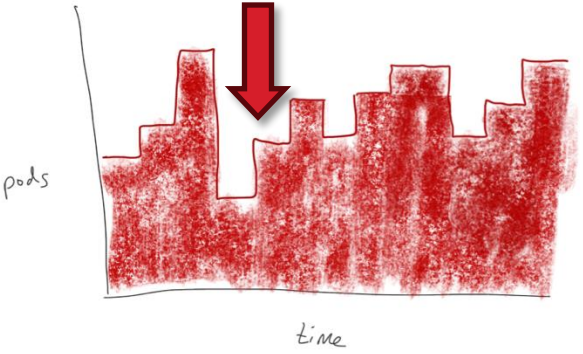
Lots of wasted time in starting/terminating pods



The night club theory
You want a line out the nightclub – otherwise people won't see how cool it really is....

More steady state queue depth and pod count





Are you O-KEDA?

KEDA is AWESOME!

...but

be prepared for unexpected issues
when you implement a new auto scaler

@brianthedavis

red canary