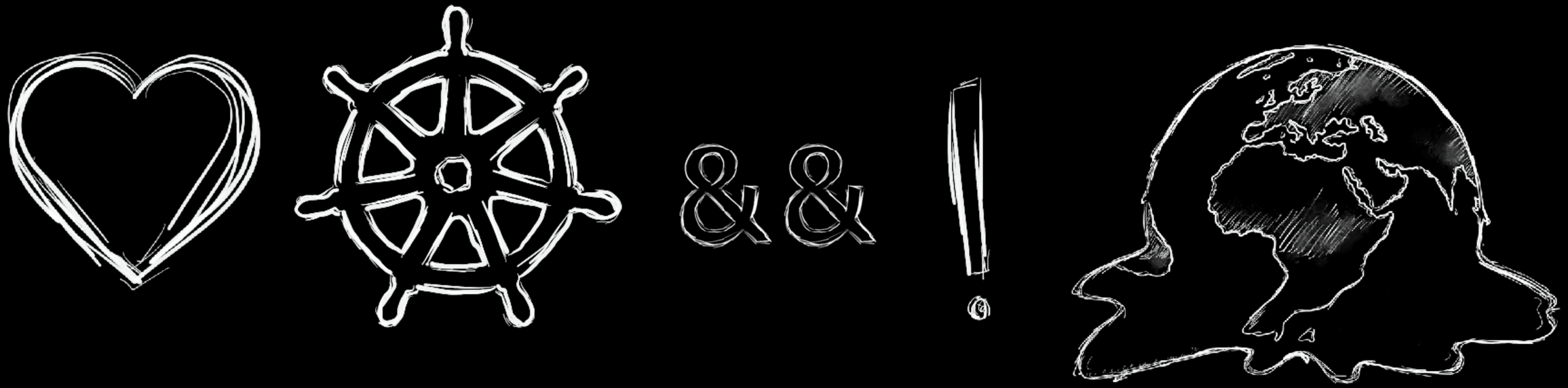
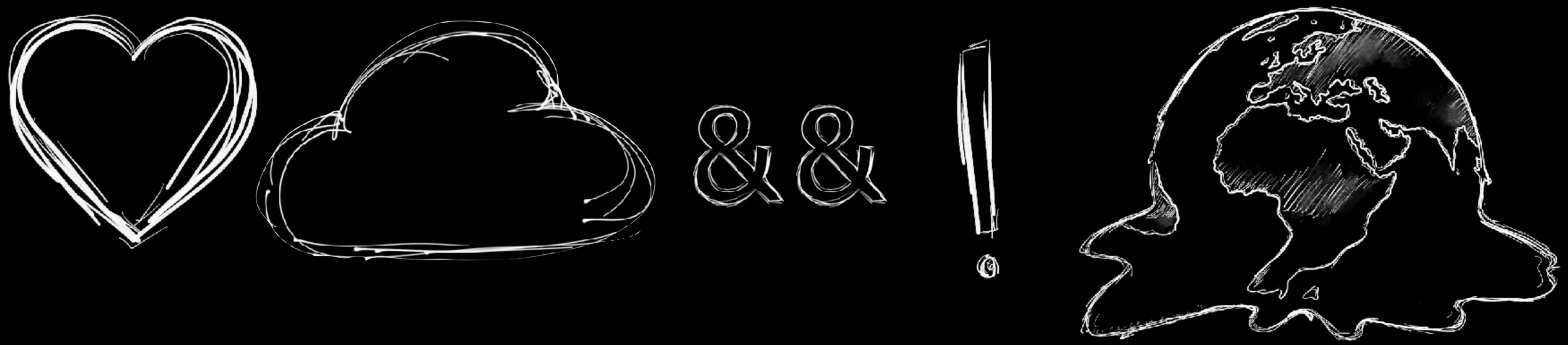
How to Love K8s and Not Wreck the Planet



Holly Cummins
IBM
@holly_cummins

How to Love Cloud Native and Not Wreck the Planet



Holly Cummins
IBM
@holly_cummins


THE LEAN SERIES
ERIC RIES, SERIES EDITOR
Jez Humble, Joanne Molesky & Barry O'Reilly
LEAN
ENTERPRISE
How High Performance
Organizations
Innovate at Scale
O'REILLY

Continuous integration is the practice of working in small batches and using automated tests to detect and reject changes that introduce a regression. It is, in our opinion, the most important technical practice in the agile canon, and it forms the foundation of continuous delivery, for which we require in addition that each change keeps the code on trunk releasable. However, that can be hard to adopt for teams that are not used to it.

In our experience, people tend to fall into two camps: those who can't understand how it is possible (particularly at scale) and those who can't believe people would work in any other way. We assure you that it is possible, both at small scale and large scale, whatever your domain.

Let's first address the scale problem with two examples. First, the HP FutureSmart case study demonstrates continuous integration being effective with a distributed team of 400 people working on an embedded system. Second, we'll note that almost all of Google's 10,000+ developers distributed over 40 offices work off a single code tree. Everyone working off this tree develops and releases from trunk, and all builds are created from source. 20 to 60 code changes are submitted every minute, and 50% of the codebase changes every month.[6] Google engineers have built a powerful continuous integration system

[6] http://bit.ly/1v701cY

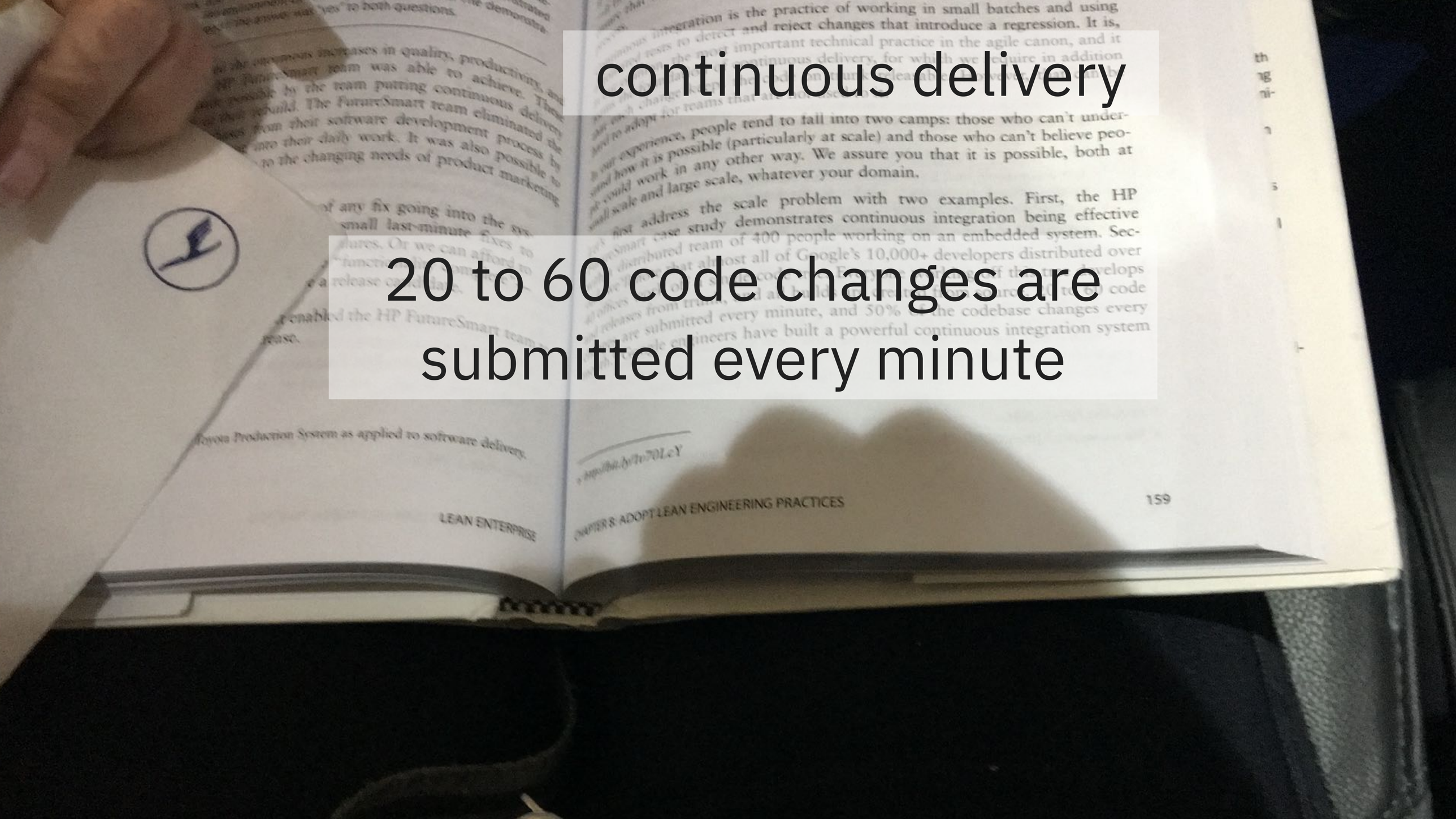
continuous delivery

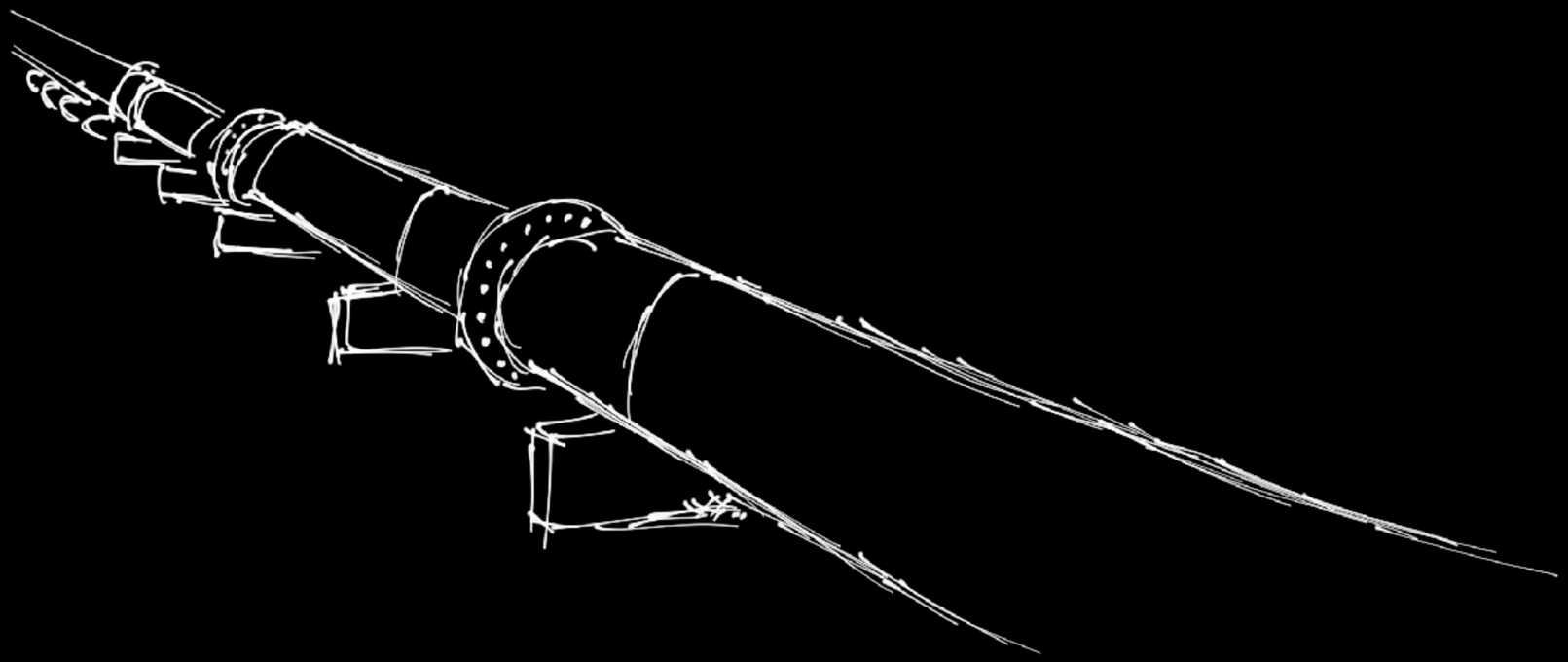
integration is the practice of working in small batches and using
and reject changes that introduce a regression. It is,
important technical practice in the agile canon, and it
continuous delivery, for which we require in addition
for teams that are not using

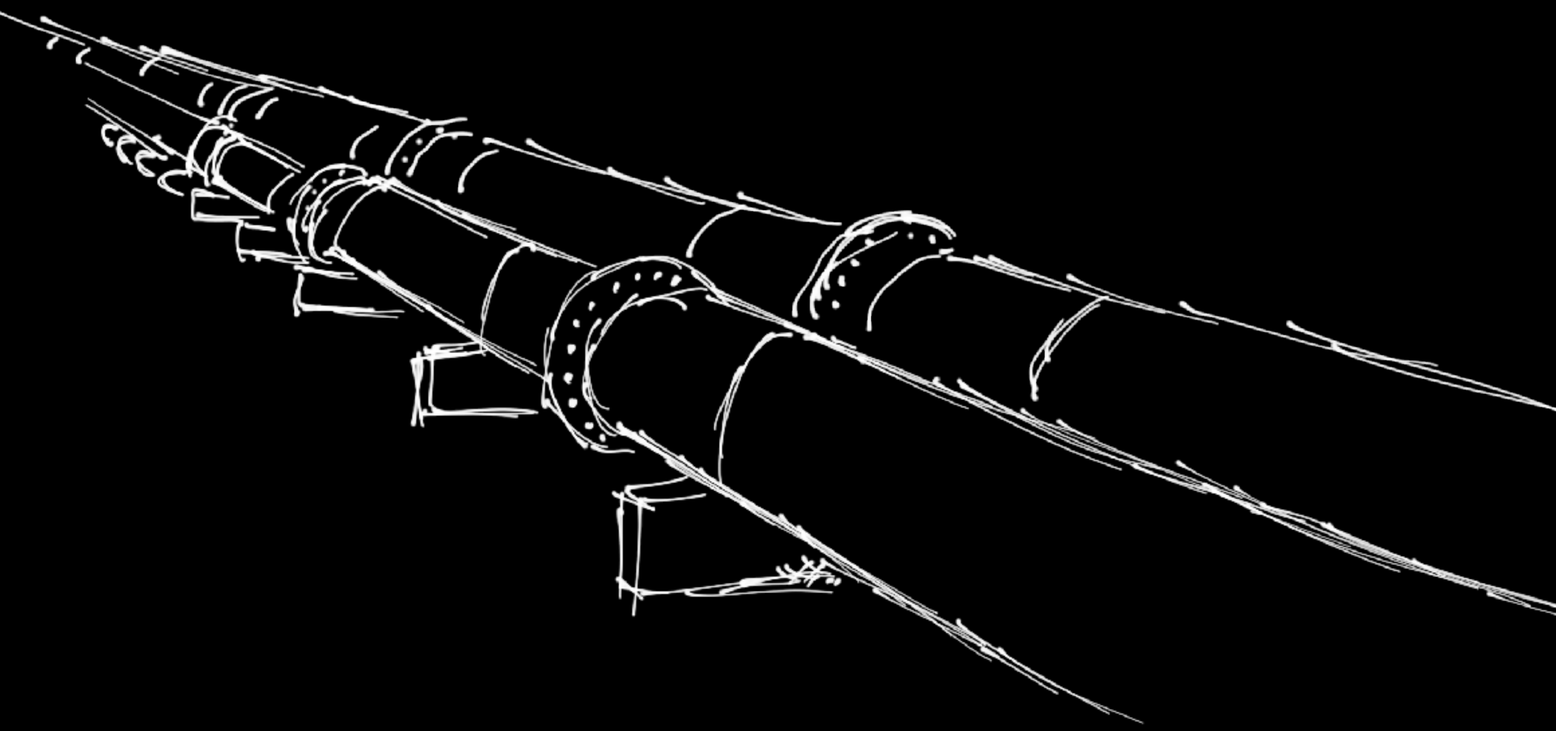
adopt. In our experience, people tend to fall into two camps: those who can't understand how it is possible (particularly at scale) and those who can't believe people could work in any other way. We assure you that it is possible, both at small scale and large scale, whatever your domain.

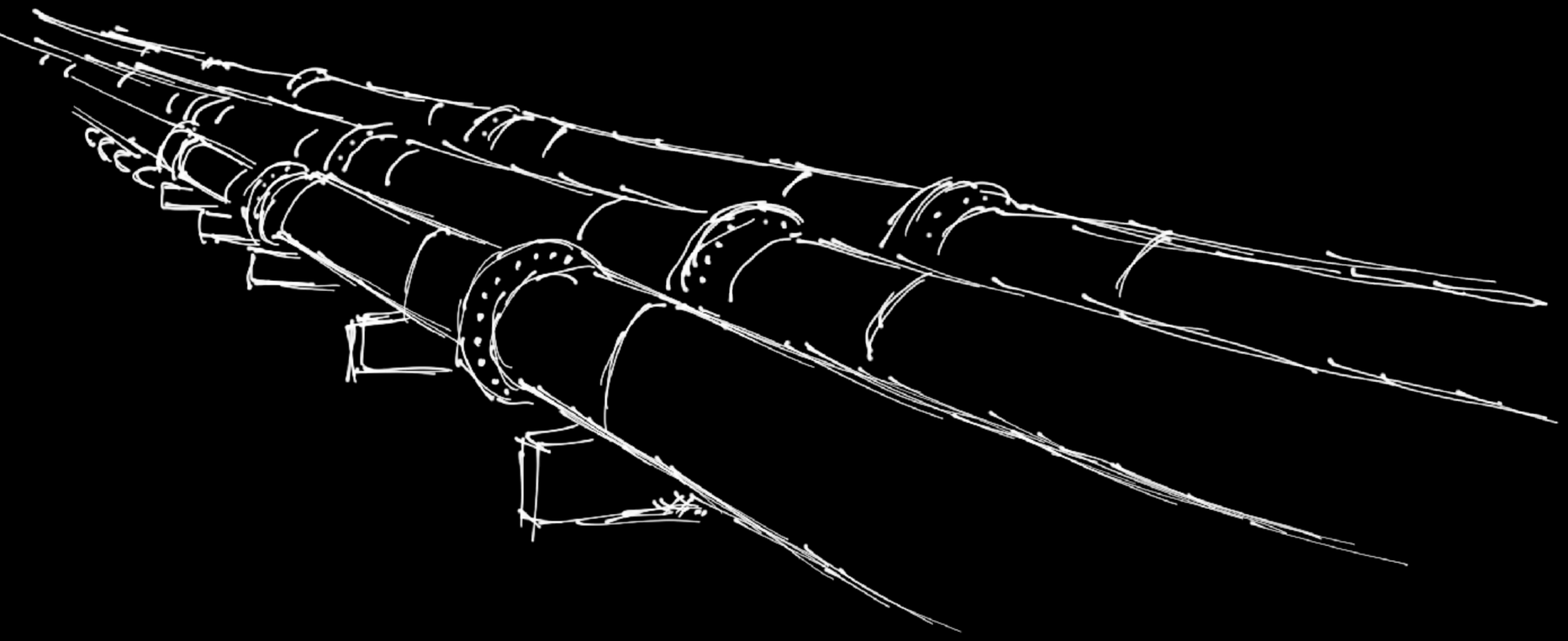
Let's first address the scale problem with two examples. First, the HP FutureSmart case study demonstrates continuous integration being effective with a distributed team of 400 people working on an embedded system. Second, we'll note that almost all of Google's 10,000+ developers distributed over 40 offices work off a single code tree. Everyone working off this tree develops and releases from trunk, and all builds are created from source. 20 to 60 code changes are submitted every minute, and 50% of the codebase changes every month.[9] Google engineers have built a powerful continuous integration system

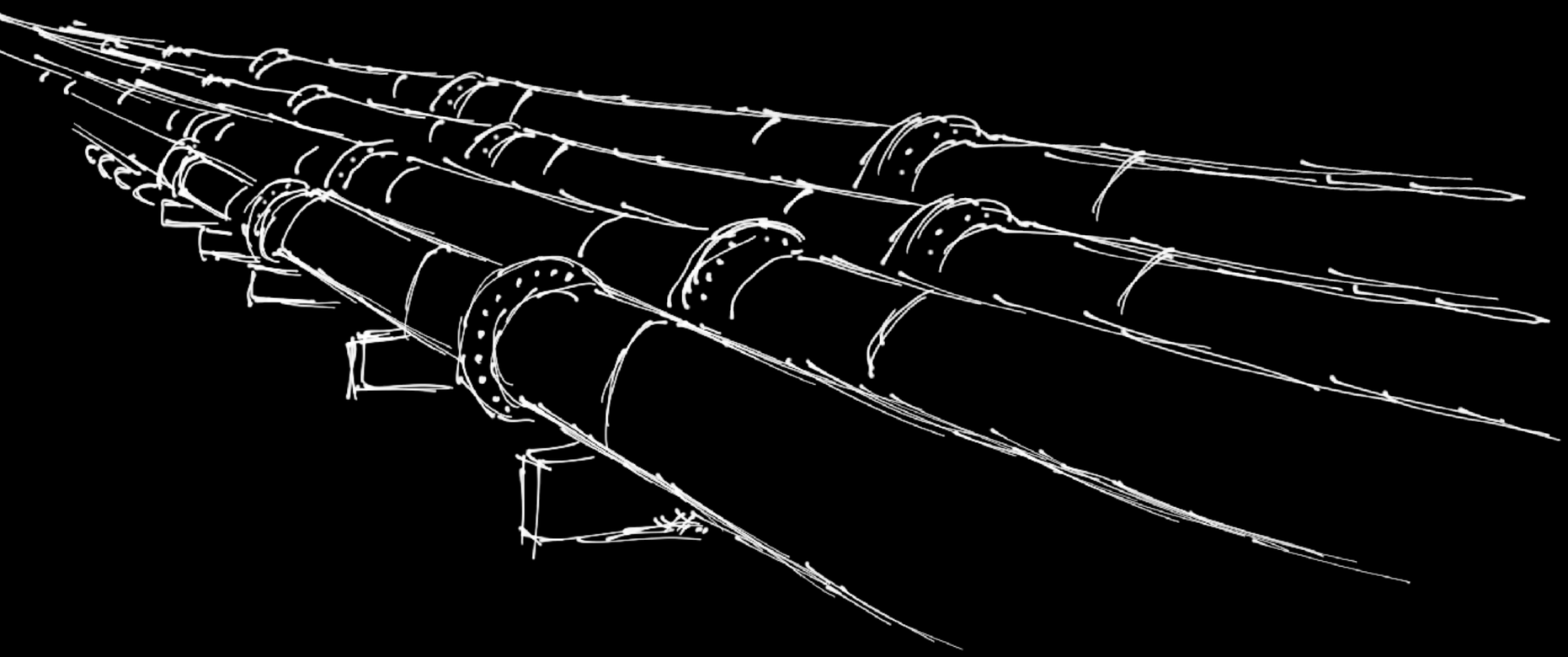
[9] http://bit.ly/1v701cY


continuous delivery
20 to 60 code changes are submitted every minute

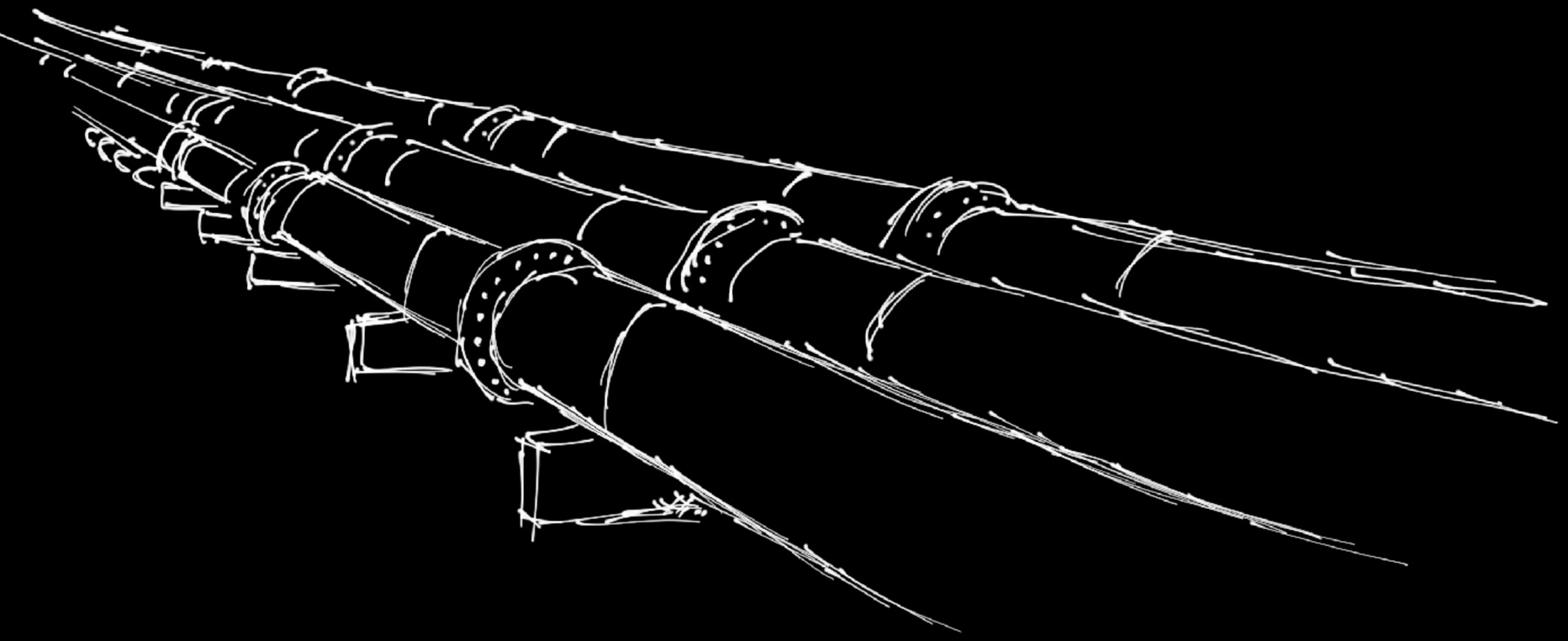




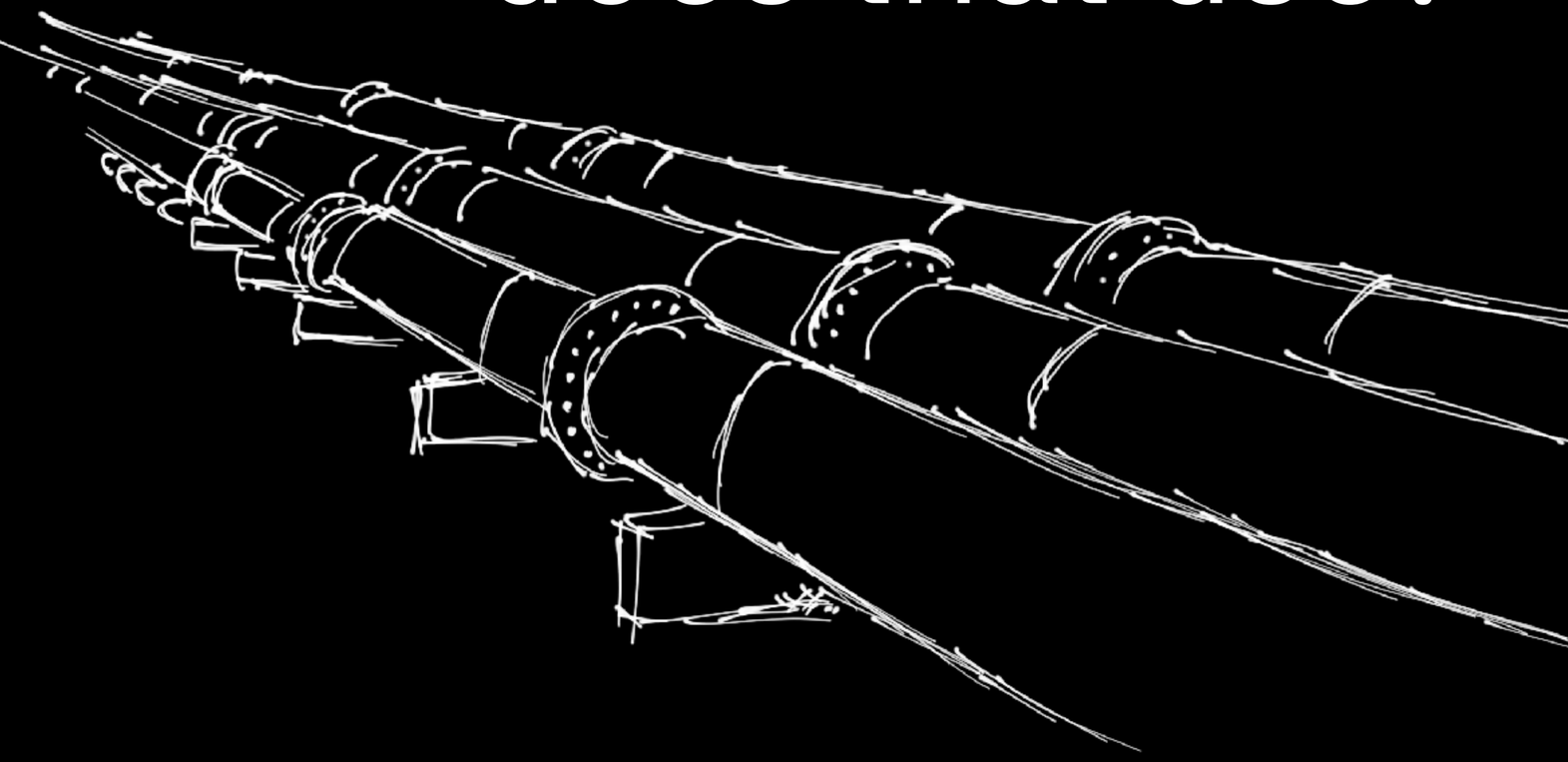




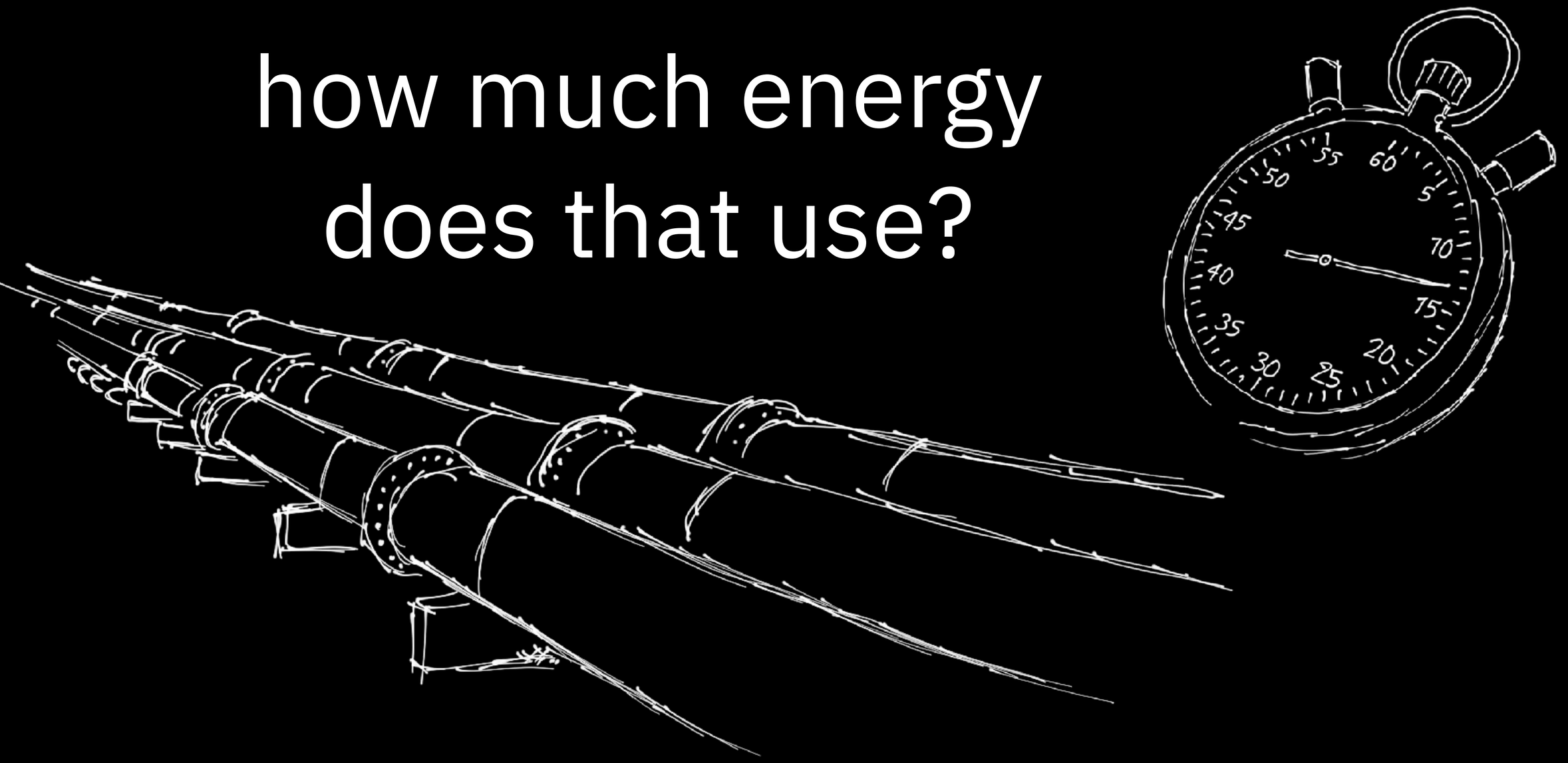
how many pipelines?



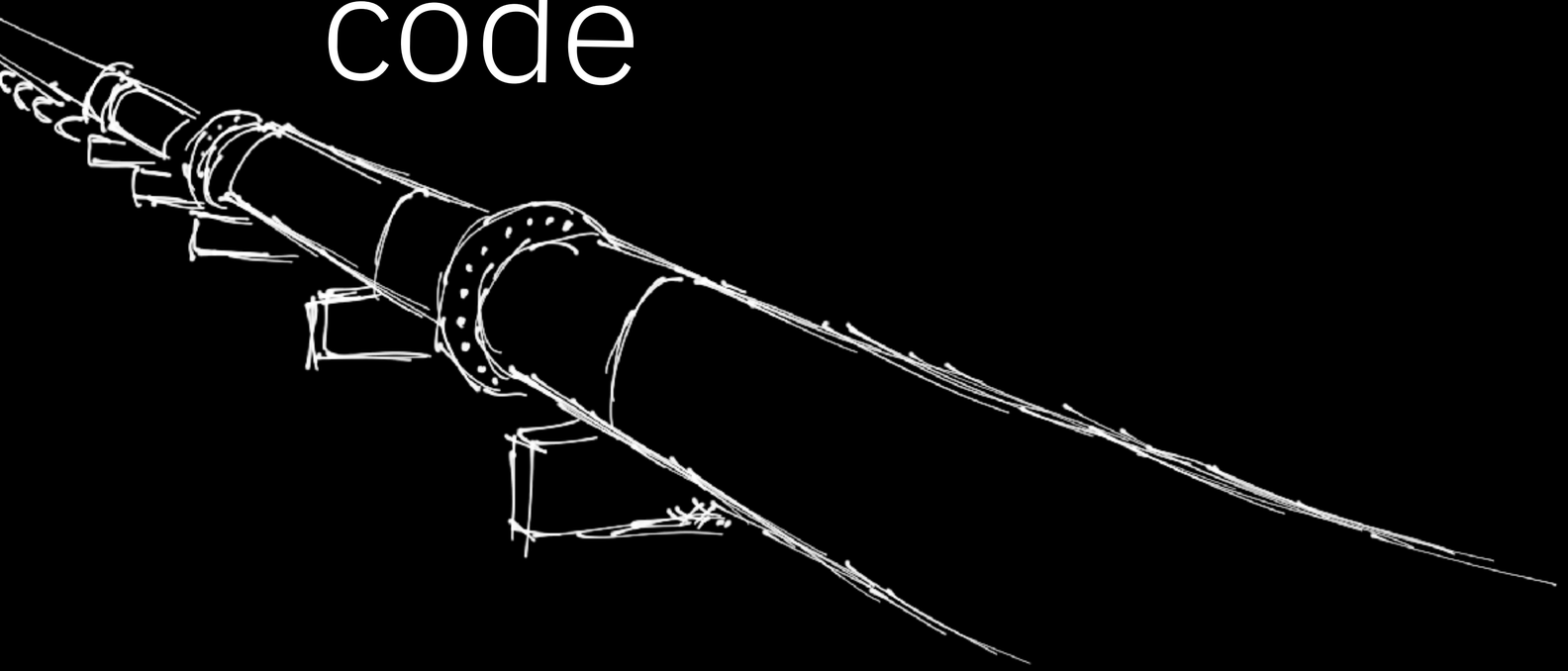
how much energy does that use?



how much energy does that use?



code




code
oil

code

~=

oil

80% of energy is fossil fuels

80% of (US) energy is fossil fuels

why does this matter?

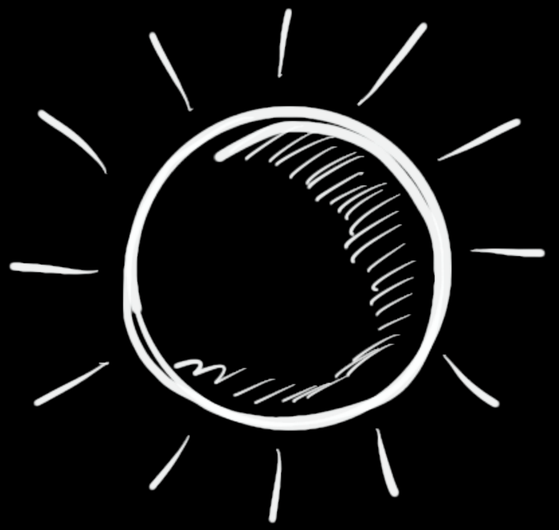
why does
this matter?

oh.

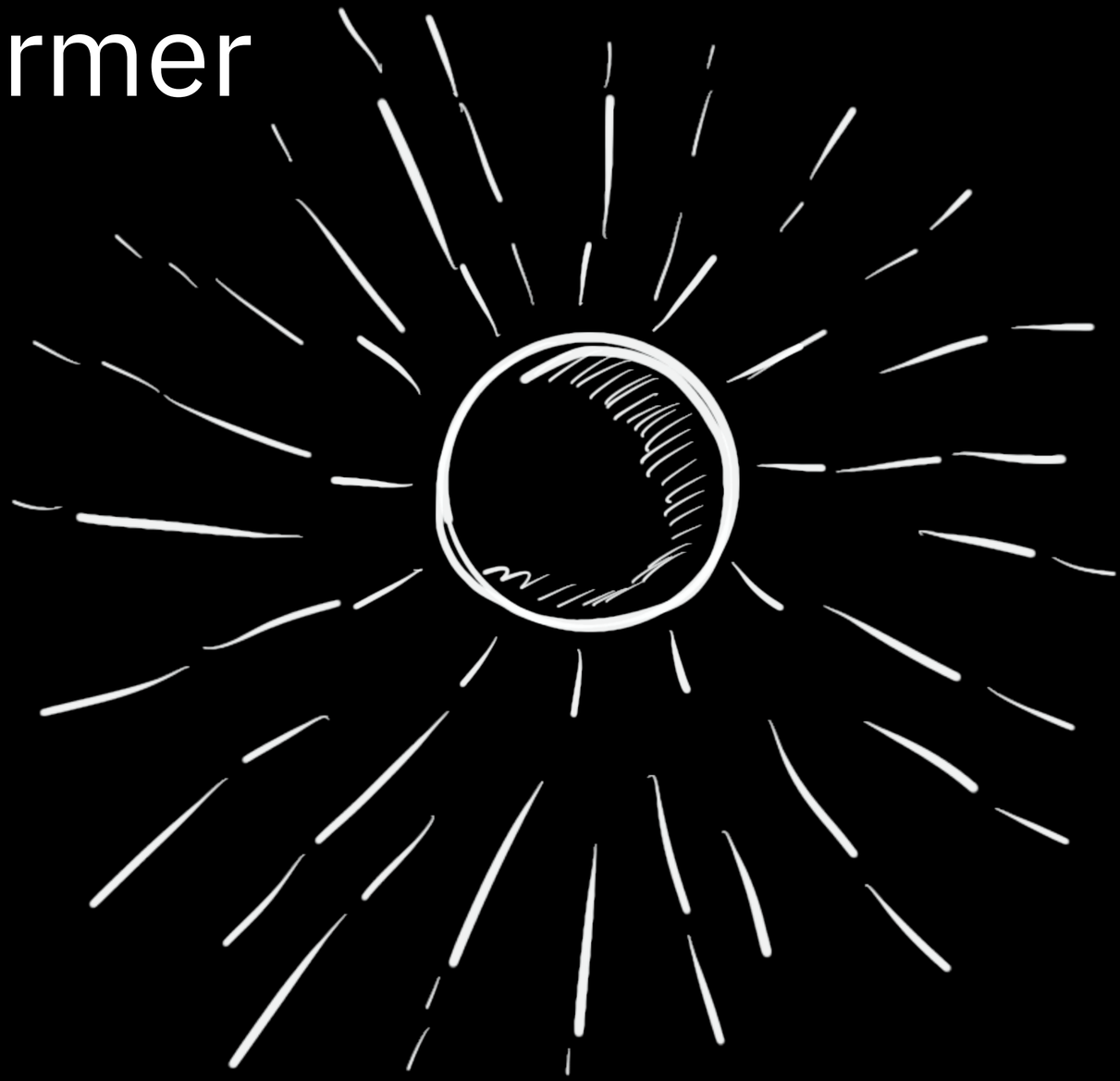
the earth is
getting **warmer**

the earth is
getting **warmer**

warmer

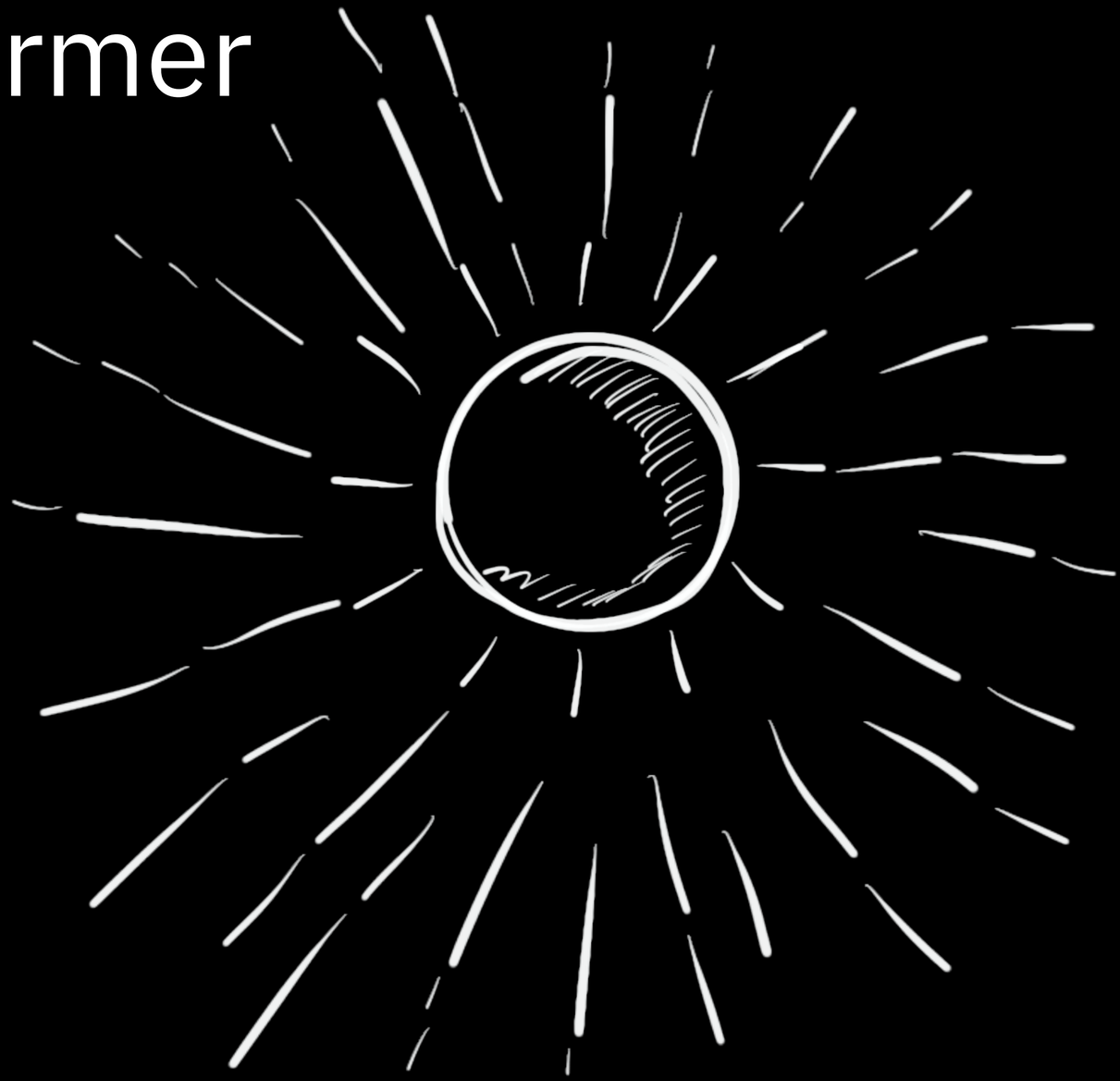


uncomfortably warmer

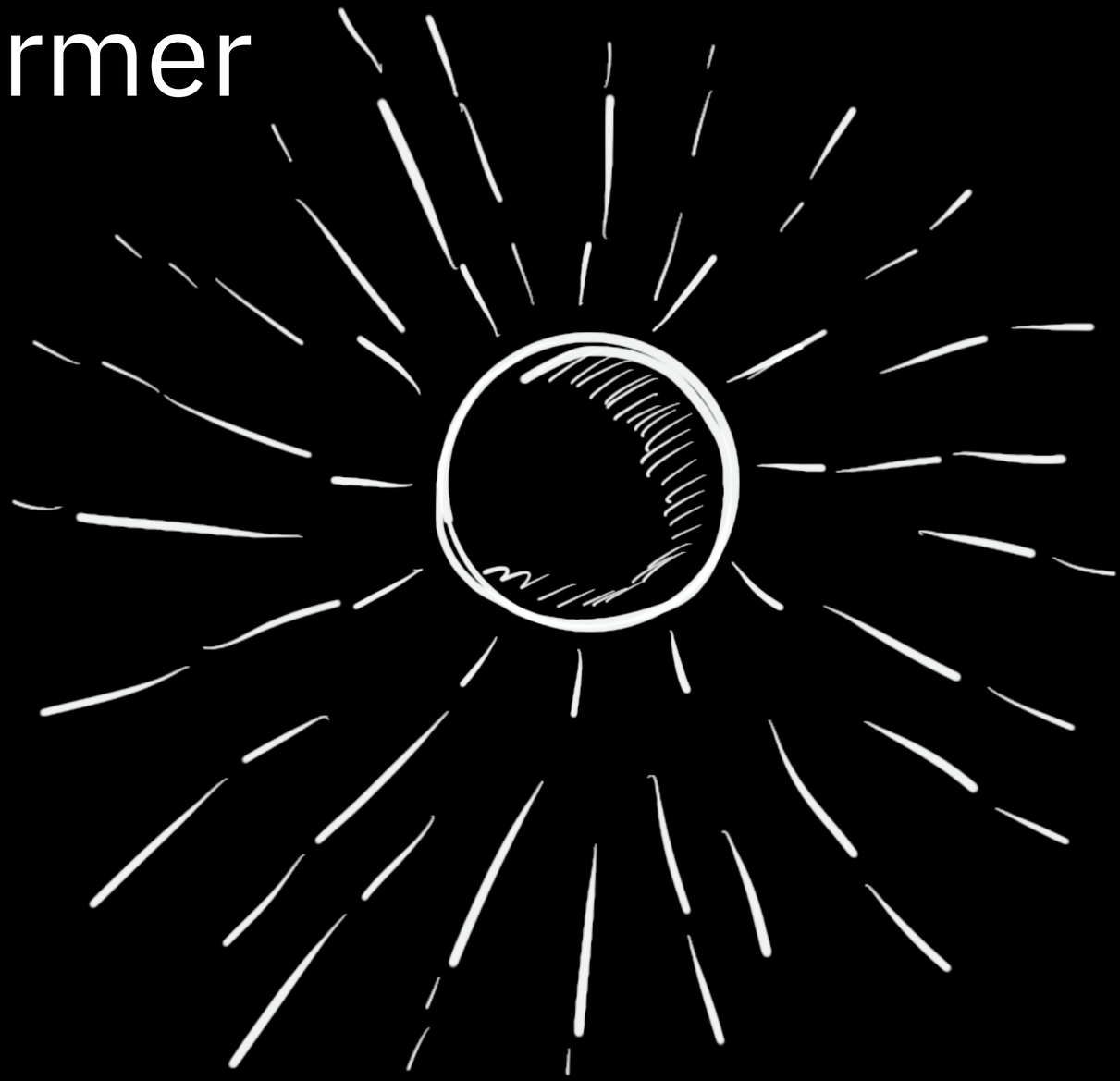


uncomfortably warmer

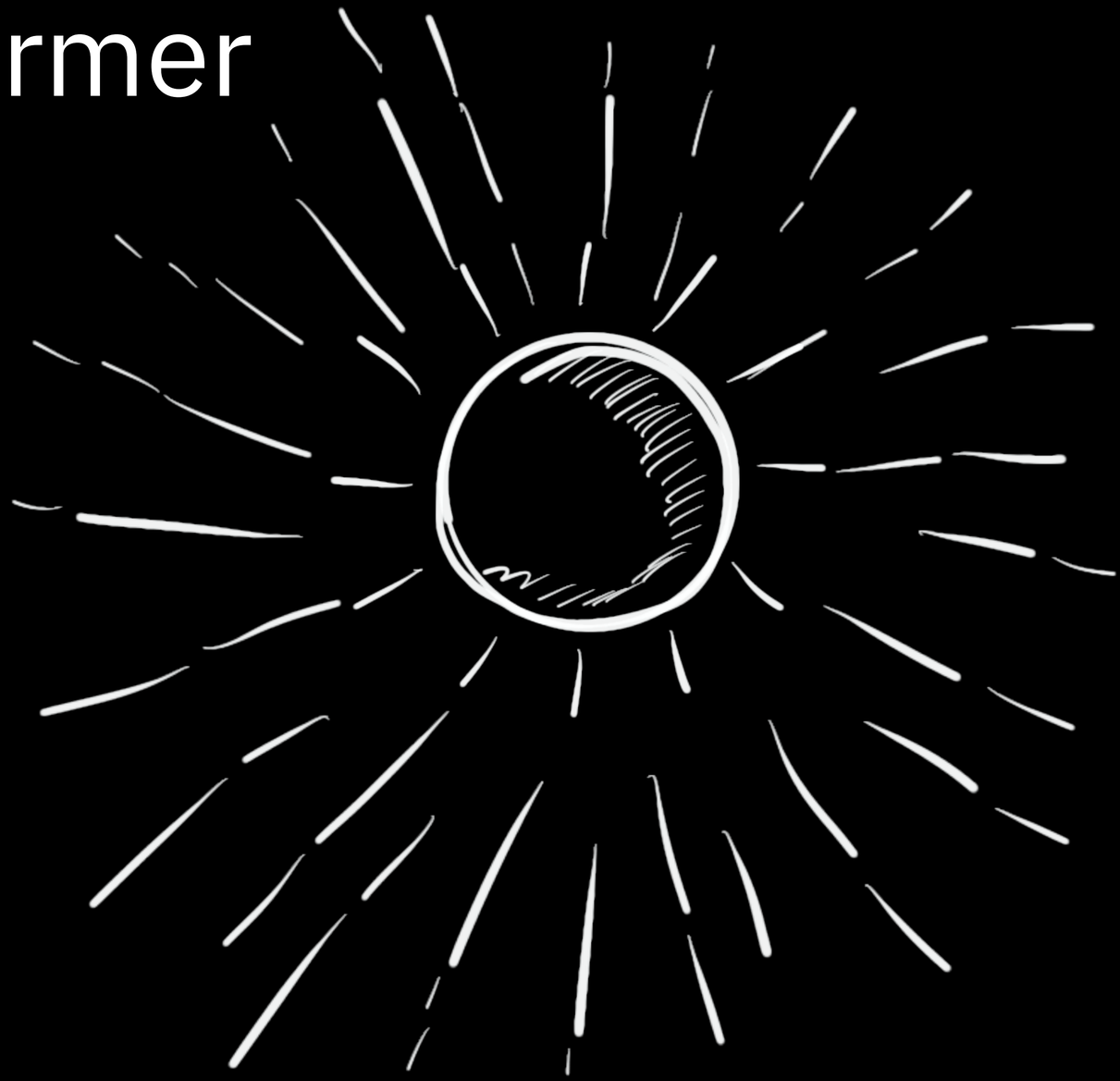
drought



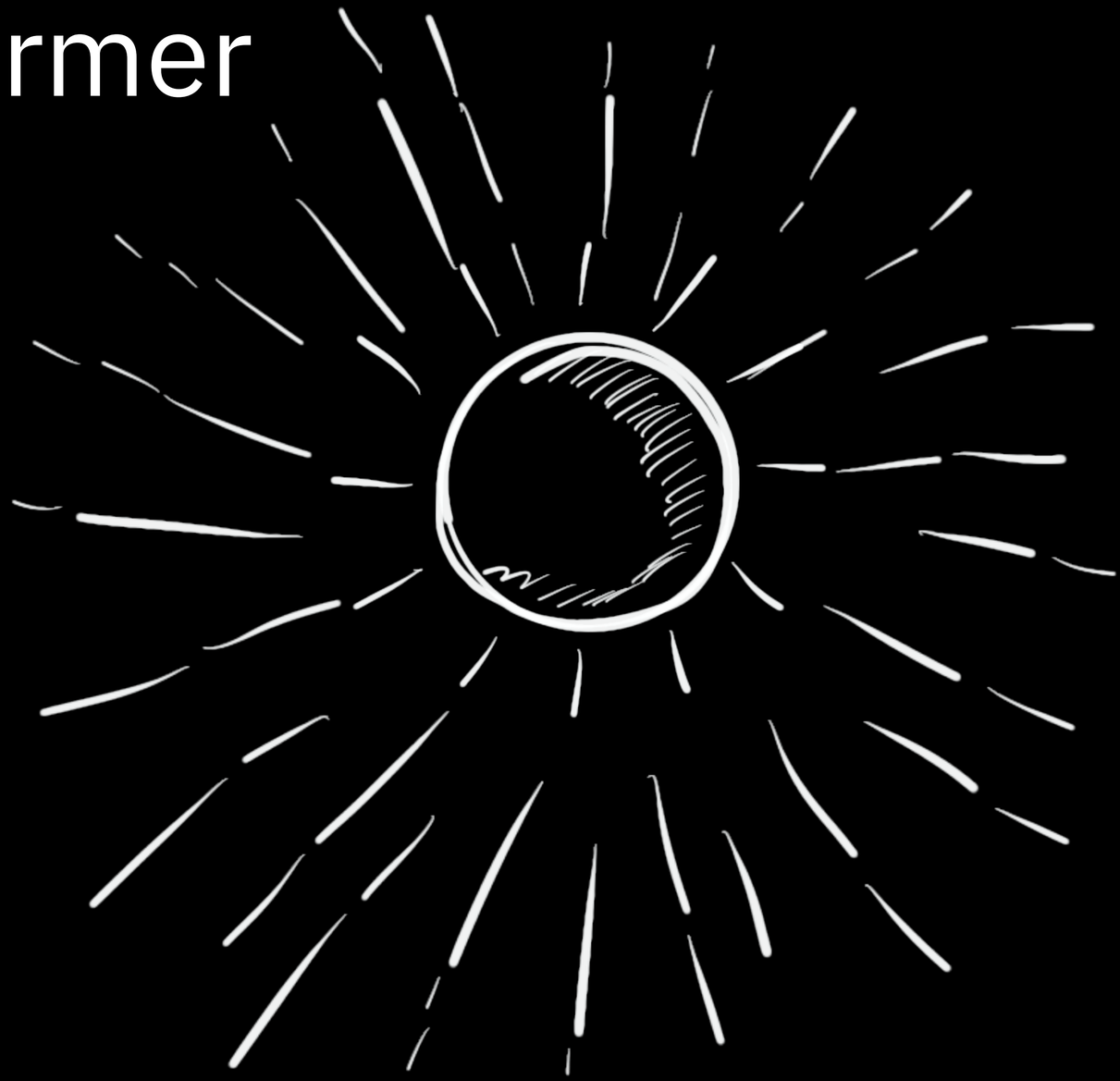
uncomfortably warmer
drought
floods



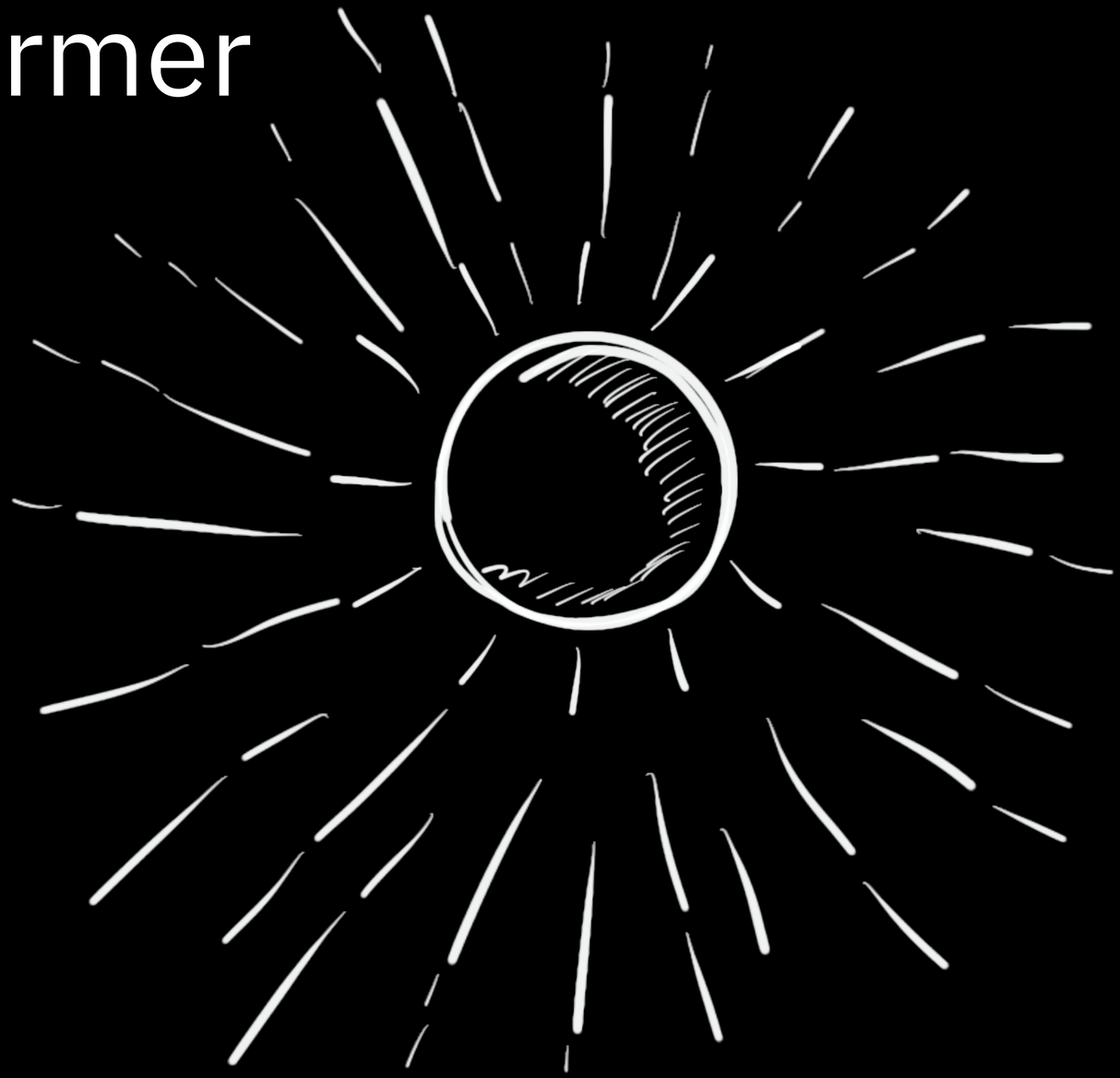
uncomfortably warmer
drought
floods
submersion



uncomfortably warmer
drought
floods
submersion
hurricanes



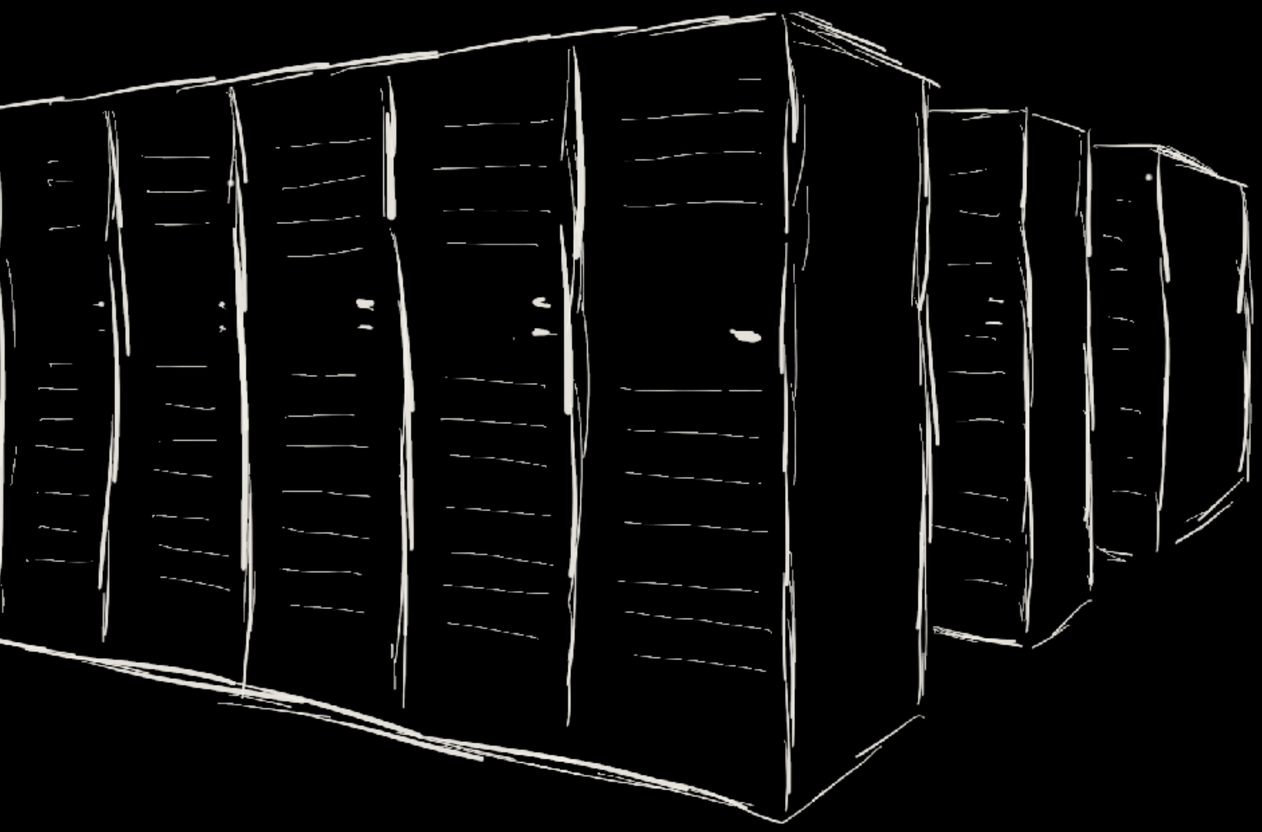
uncomfortably warmer
drought
floods
submersion
hurricanes
fires



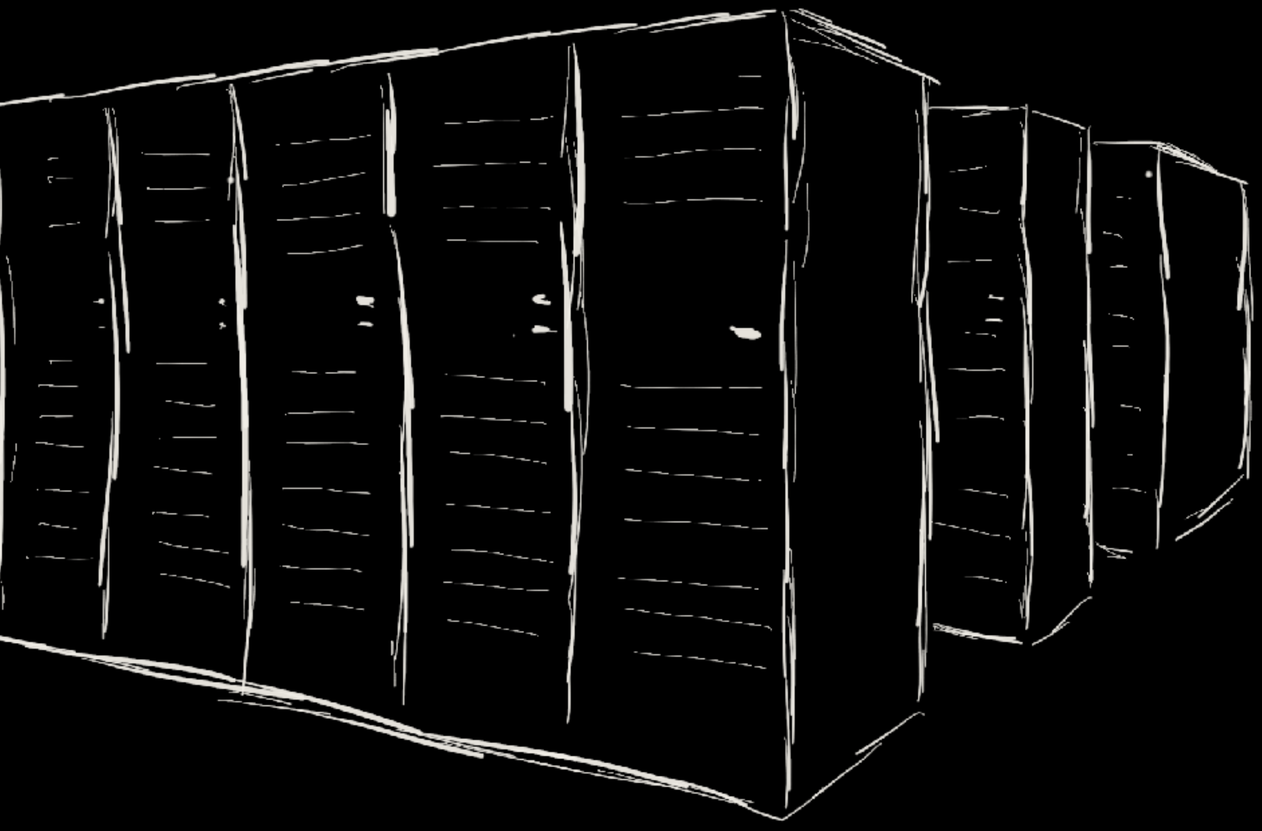
our industry
contributes to
climate change

we
contribute to
climate change

data centres

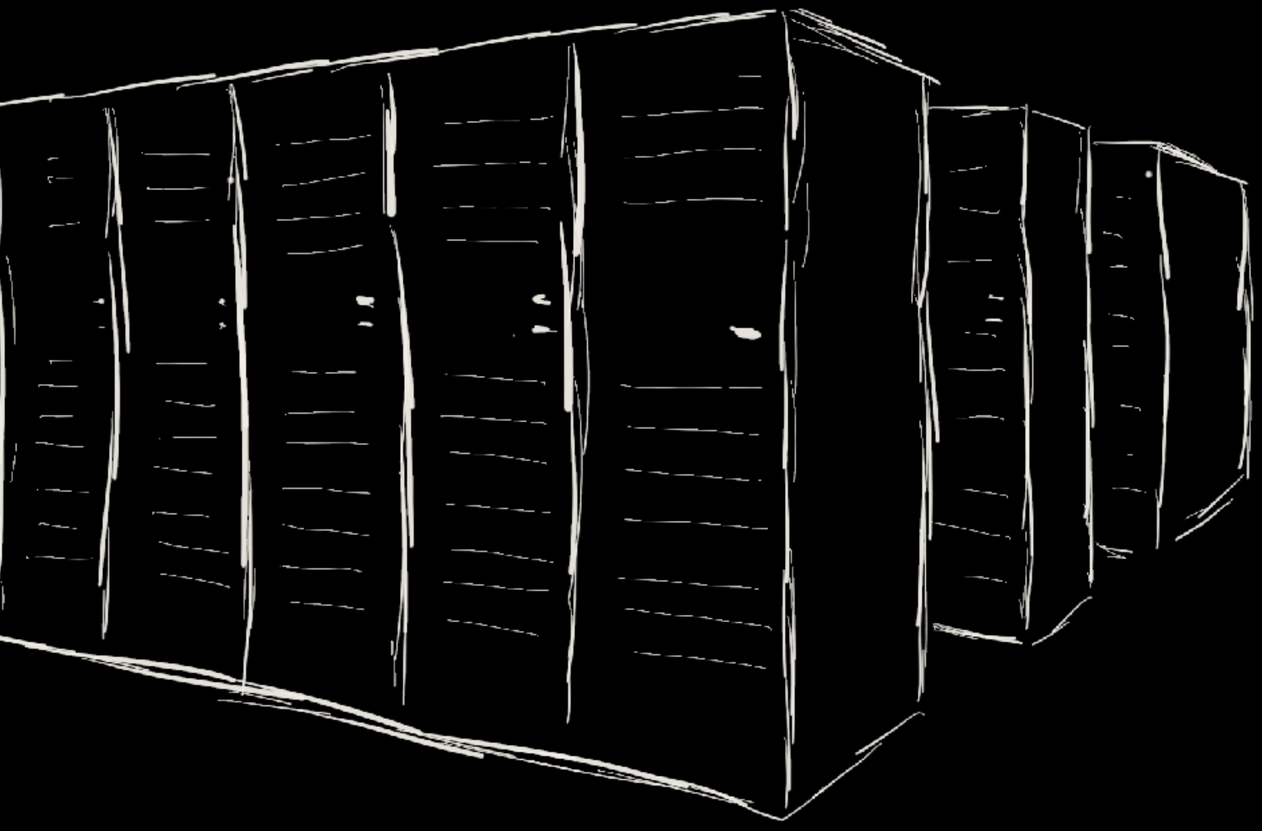


data centres



aviation

data centres

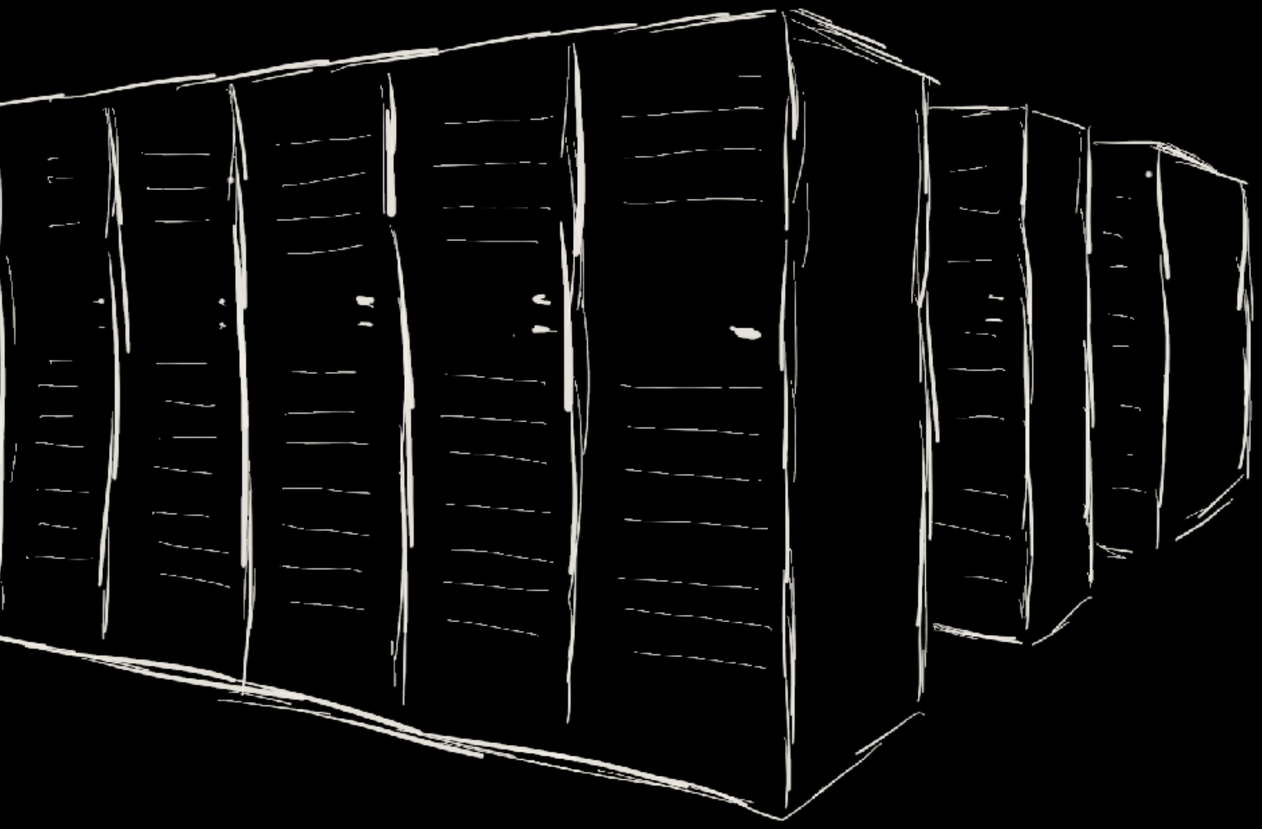


aviation

2.5%

data centres

1-2%

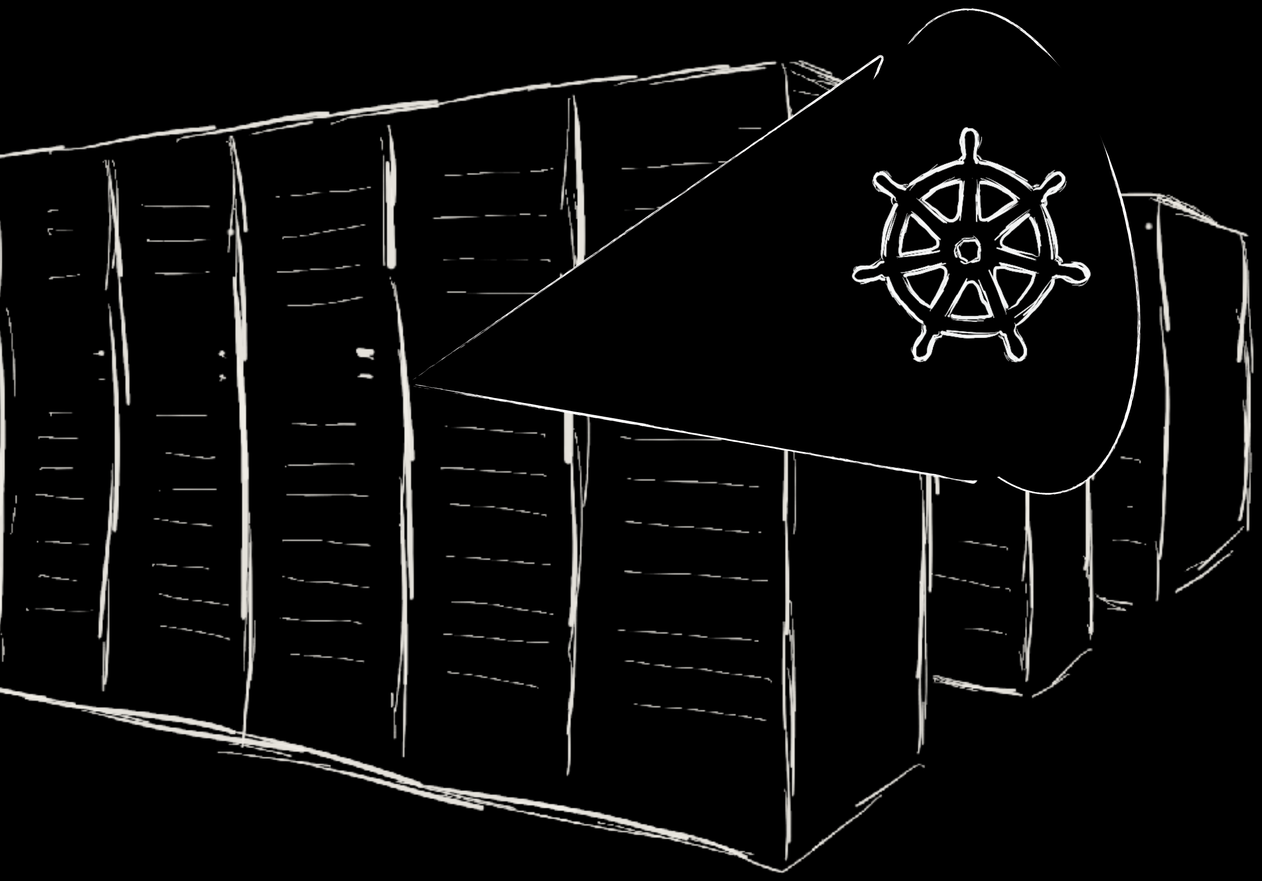


aviation

2.5%

data centres

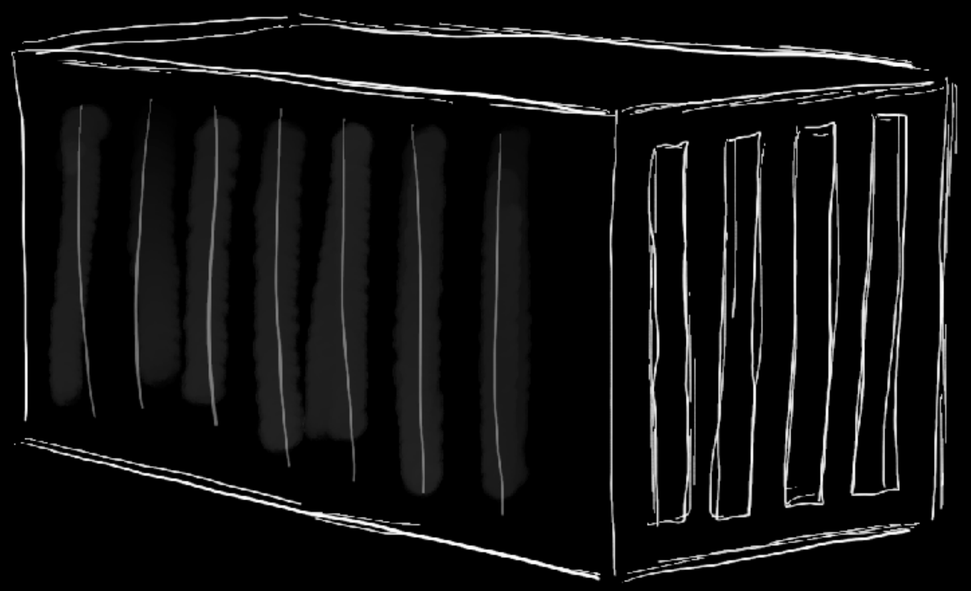
1-2%



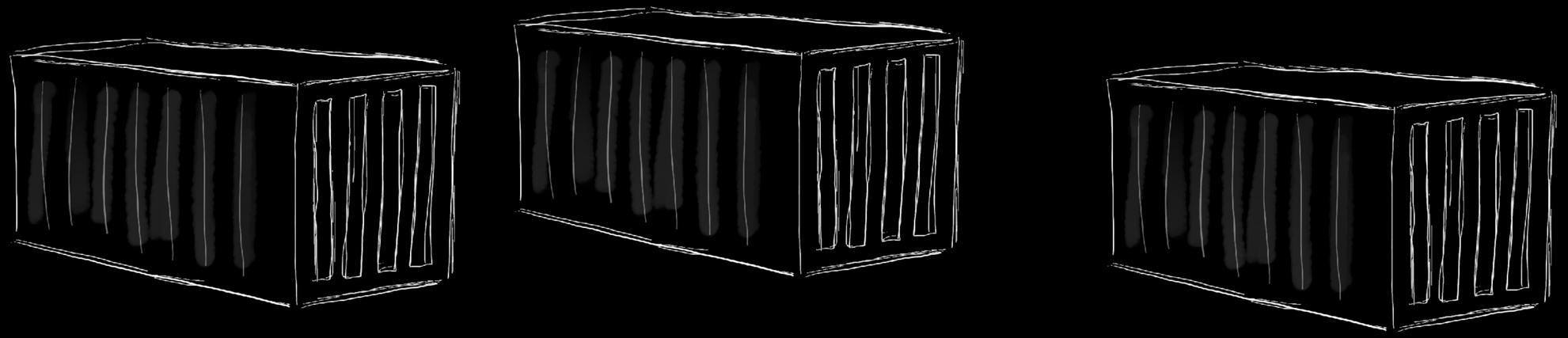
aviation

2.5%

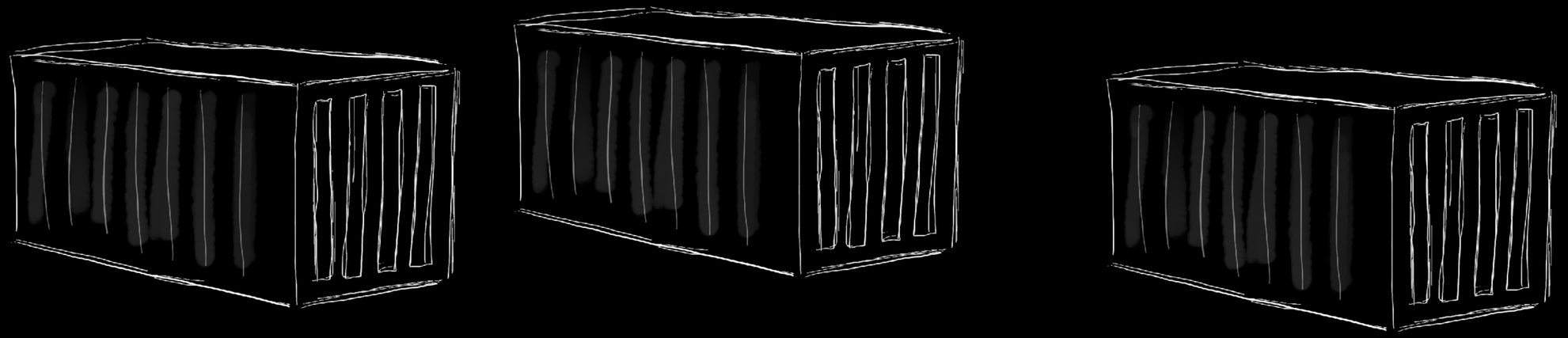
the dream

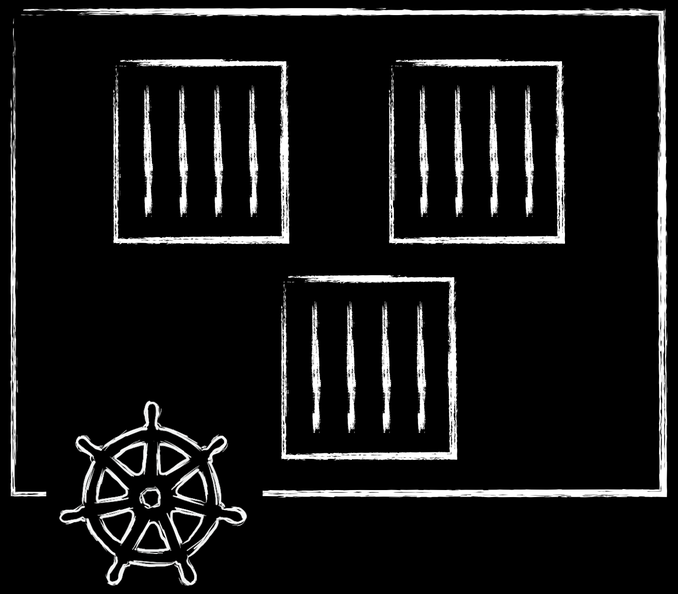


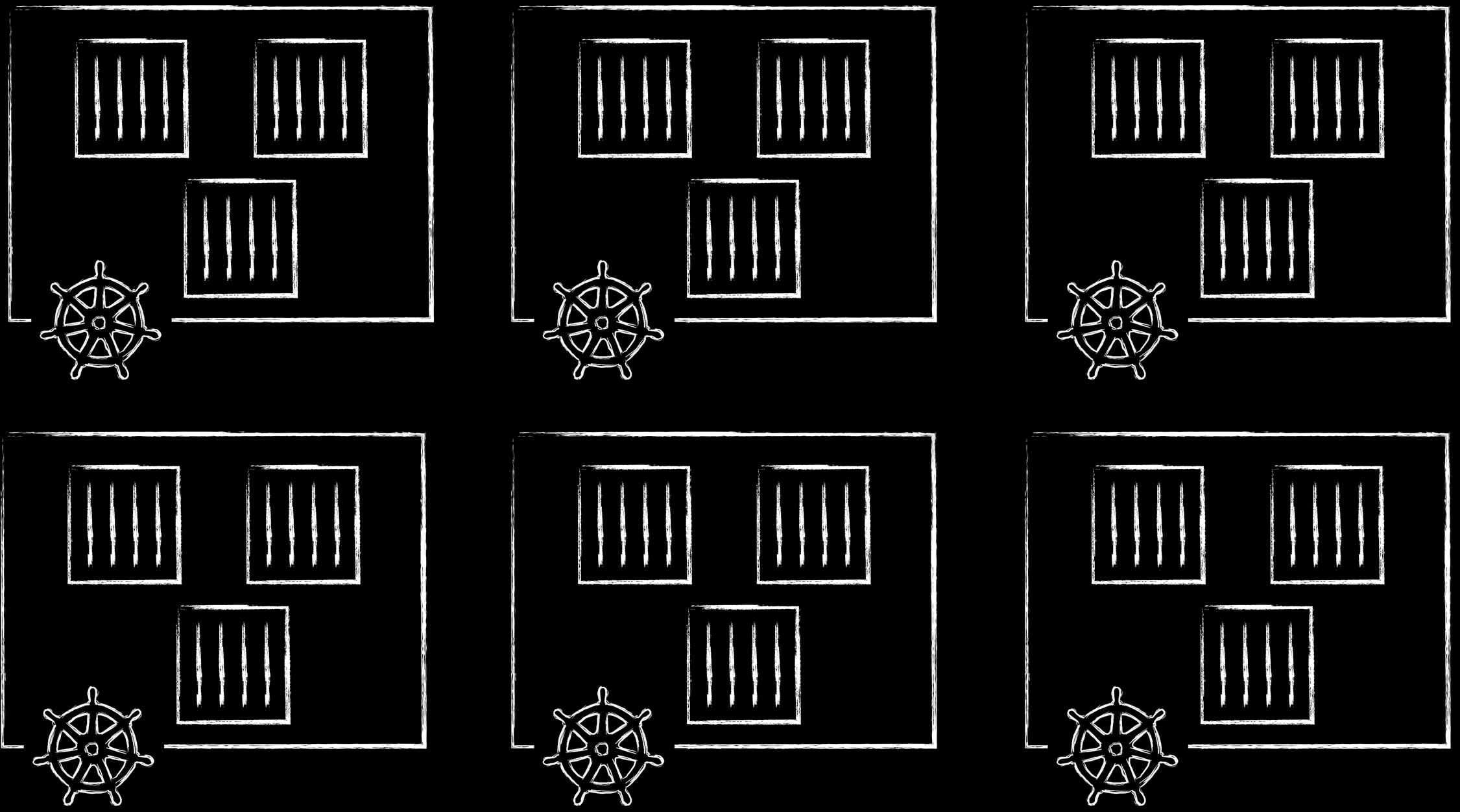
the dream

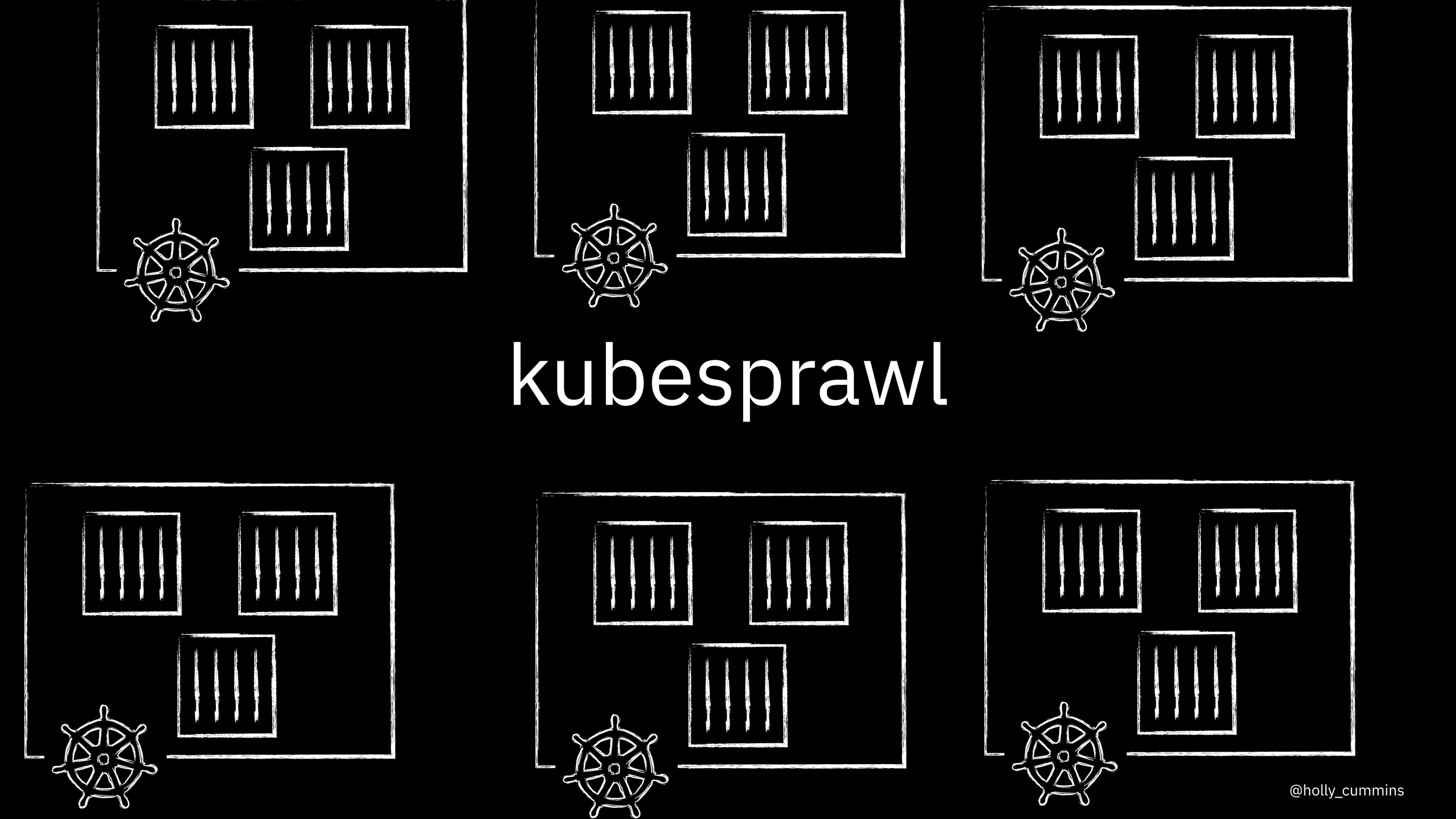


the dream









@holly_cummins

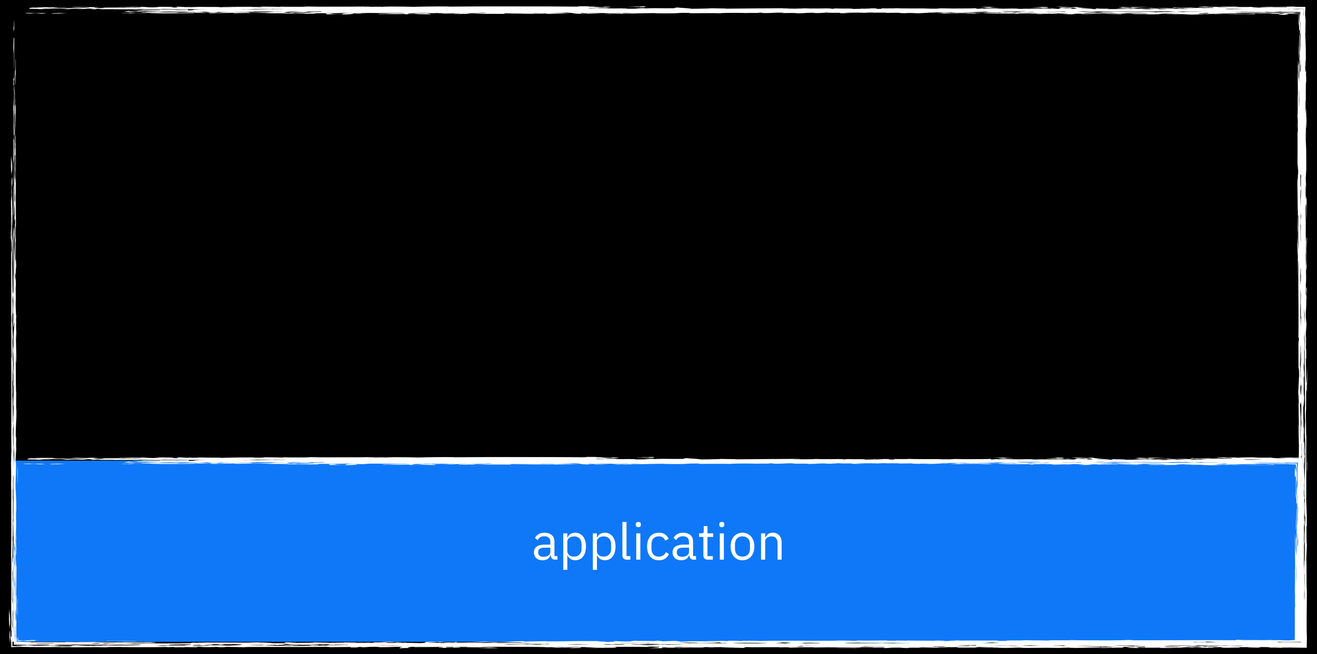
the cluster is the unit of deployment

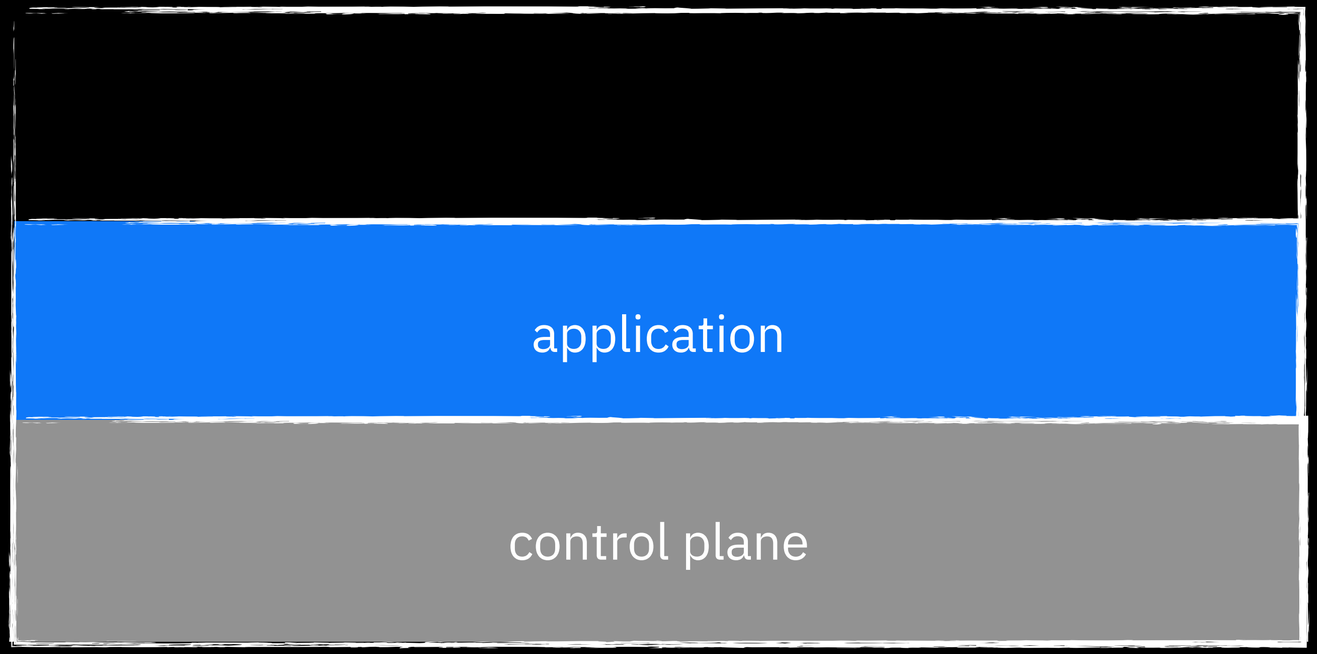
clusters per
IBM Cloud
account

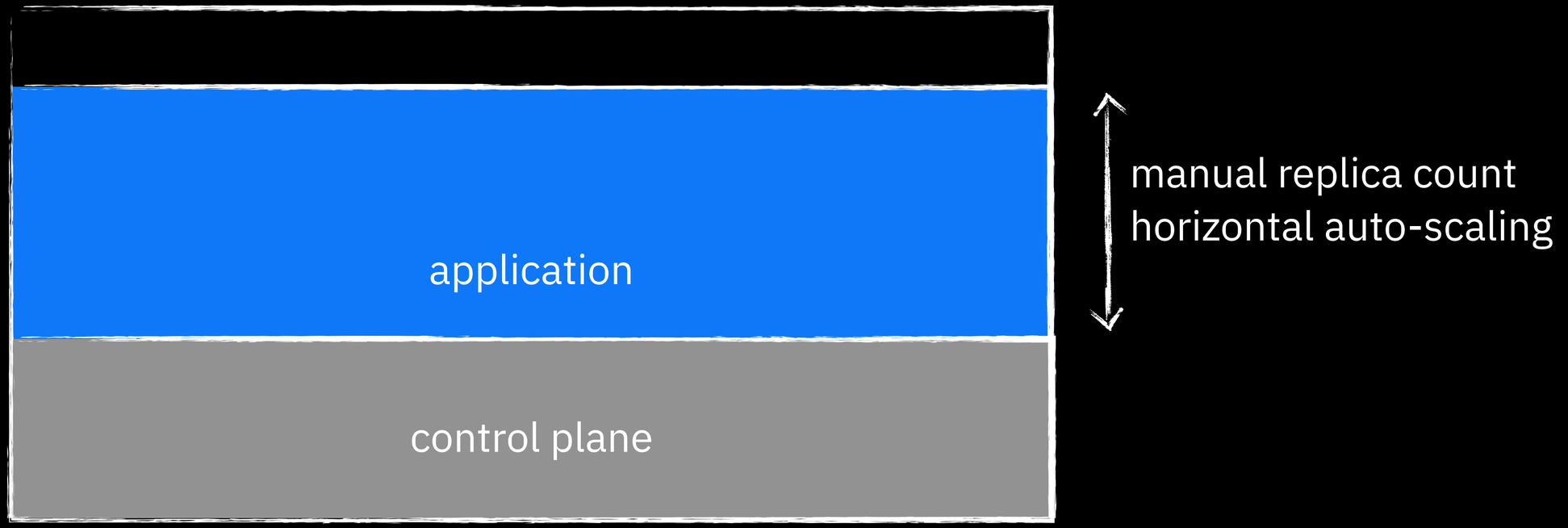
21

clusters per
IBM Cloud
account

utilisation
elasticity

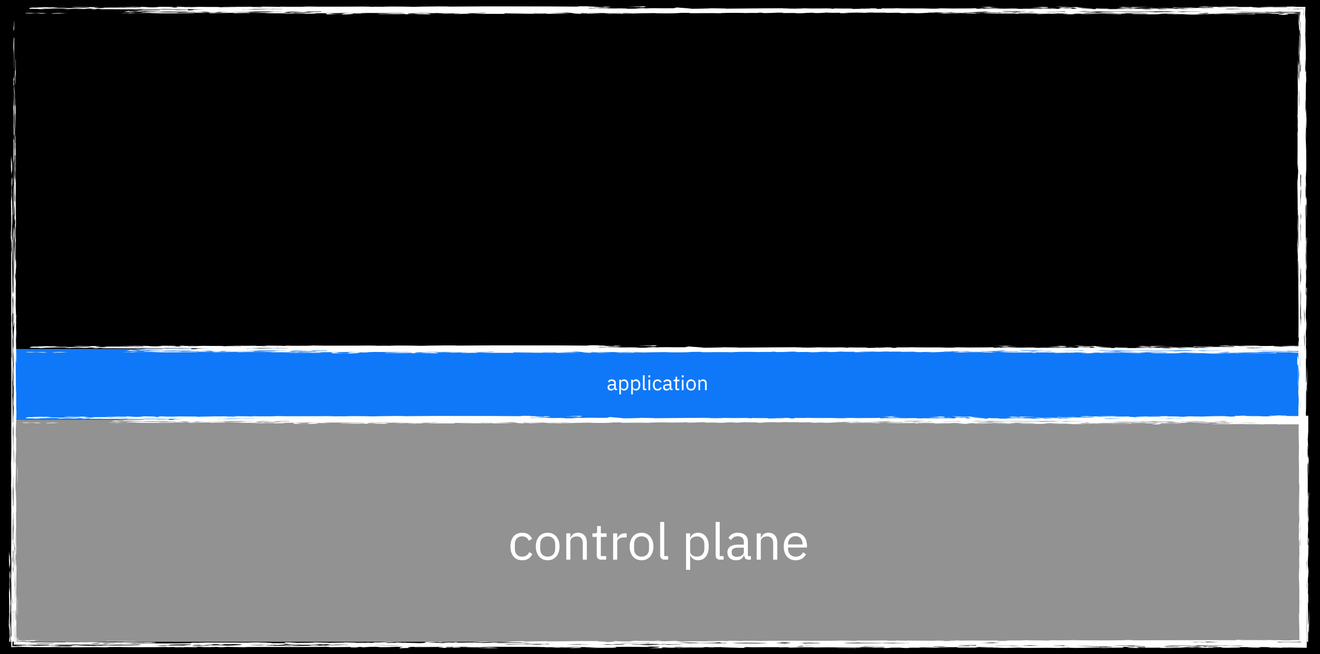

application


application
control plane


application
control plane
manual replica count
horizontal auto-scaling

!

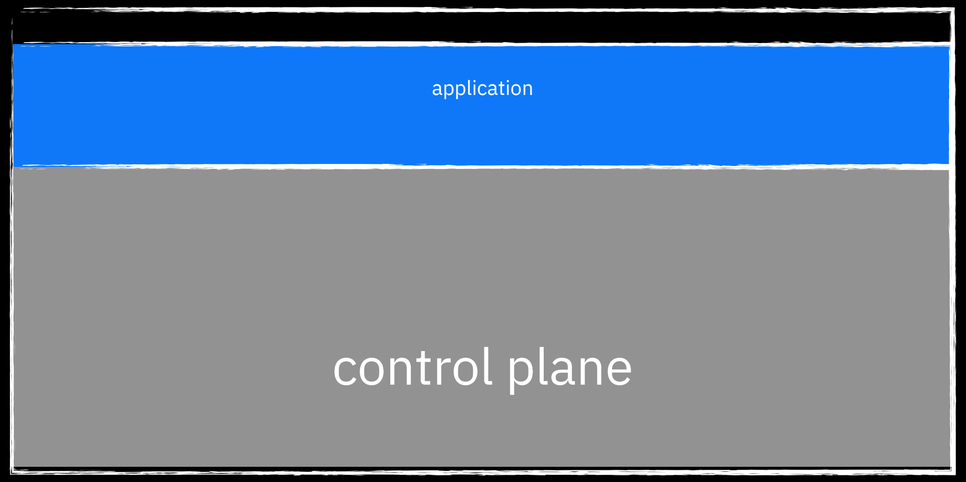
clusters are less elastic than applications



manual replica count
horizontal auto-scaling

!

clusters are less elastic than applications



!

every cluster has overhead

Navigation Menu

Orchestration service

The container platform type and version for the cluster. Choose Kubernetes for a native Kubernetes experience. Learn more about the options

Kubernetes

1.17.7 (Stable, Default)

Get even more features with

RedHat OpenShift

Infrastructure

Choose which network and compute environment to run your cluster on. Learn more about the differences.

Classic
Run your cluster with native subnet and VLAN networking on our classic infrastructure.

VPC
Create a fully customizable, software-defined virtual network with superior isolation using IBM Cloud VPC.

Location

Choose your location and configure your VLANs. Learn more about this.

Resource group

bofa-maas

Geography

Asia Pacific

Europe

North America

South America

Availability

Single zone

Multizone

Worker zone

Amsterdam 03

Private VLAN
2870350-1387-ber01a.ams03

Public VLAN
2870388-1222-ber01a.ams03

Worker pool

Set up a worker pool with the flavor and number of worker nodes that you want to run your first workload. At any time later, you can add more worker pools with different flavors, or resize your worker pools to fit the resource needs of your workloads.

Virtual - shared, Ubuntu 18

4
vCPUs

16 GB
Memory

$0.29 / hr
Cost

Change flavor

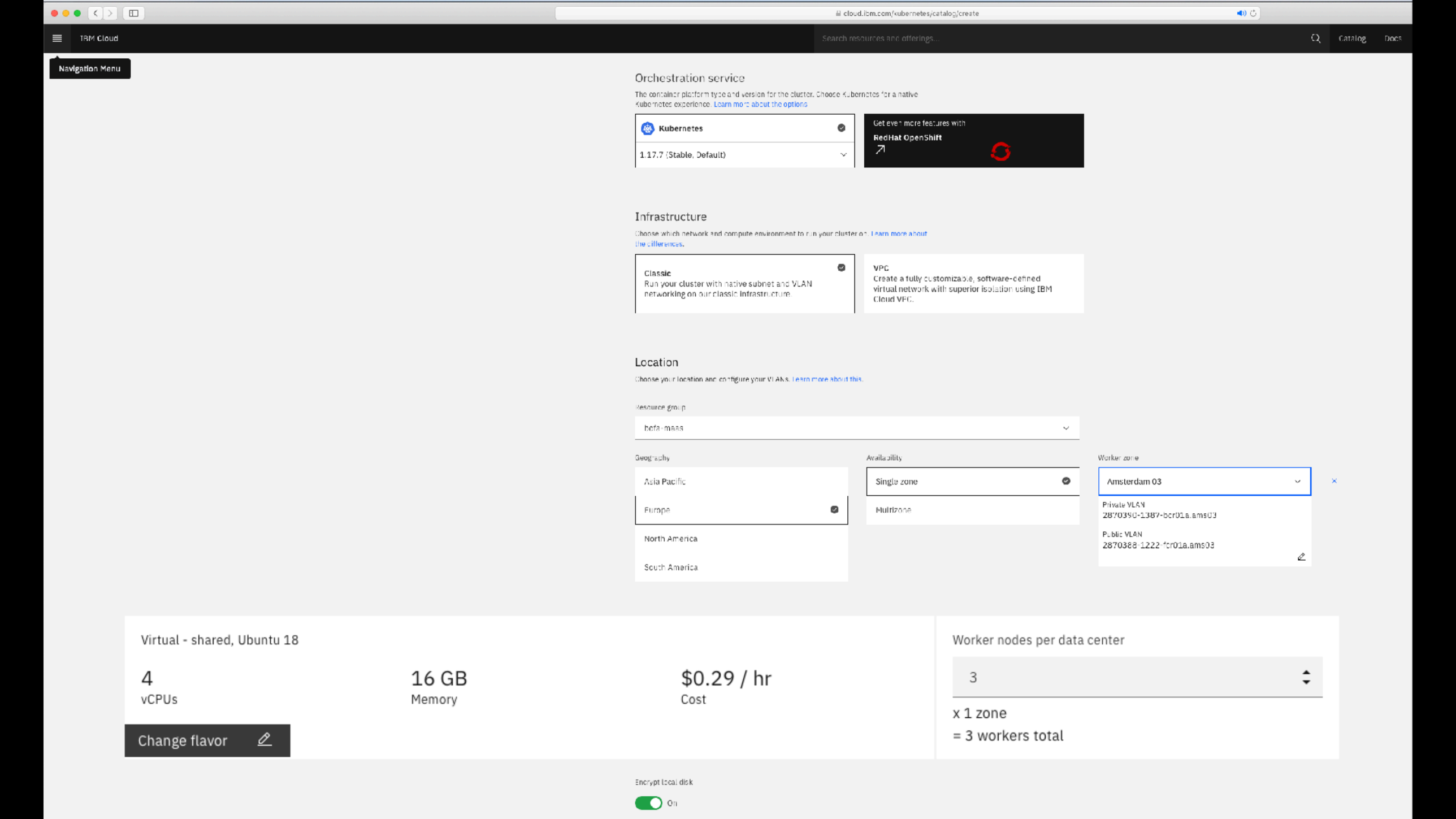
Worker nodes per data center

3

x 1 zone
= 3 workers total

Encrypt local disk

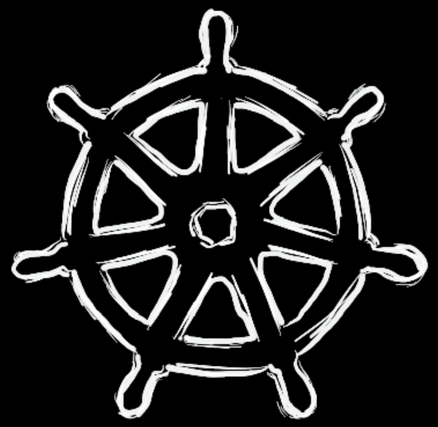
On

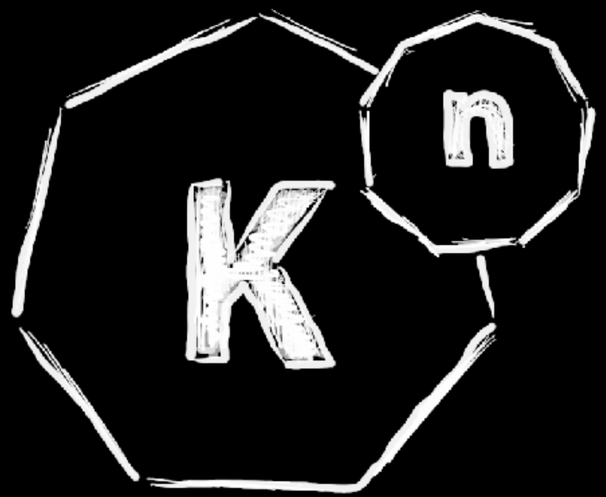

cloud.ibm.com/kubernetes/catalog/create
IBM Cloud
Search resources and offerings...
Catalog
Docs
Navigation Menu
Orchestration service
The container platform type and version for the cluster. Choose Kubernetes for a native Kubernetes experience. Learn more about the options
Kubernetes
1.17.7 (Stable, Default)
Get even more features with
RedHat OpenShift
Infrastructure
Choose which network and compute environment to run your cluster on. Learn more about the differences.
Classic
Run your cluster with native subnet and VLAN networking on our classic infrastructure.
VPC
Create a fully customizable, software-defined virtual network with superior isolation using IBM Cloud VPC.
Location
Choose your location and configure your VLANs. Learn more about this.
Resource group
bofa-maas
Geography
Asia Pacific
Europe
North America
South America
Availability
Single zone
Multizone
Worker zone
Amsterdam 03
Private VLAN
2870390-1387-ber01a.ams03
Public VLAN
2870388-1222-ber01a.ams03
Virtual - shared, Ubuntu 18
4
vCPUs
16 GB
Memory
$0.29 / hr
Cost
Change flavor
Worker nodes per data center
3
x 1 zone
= 3 workers total
Encrypt local disk
On

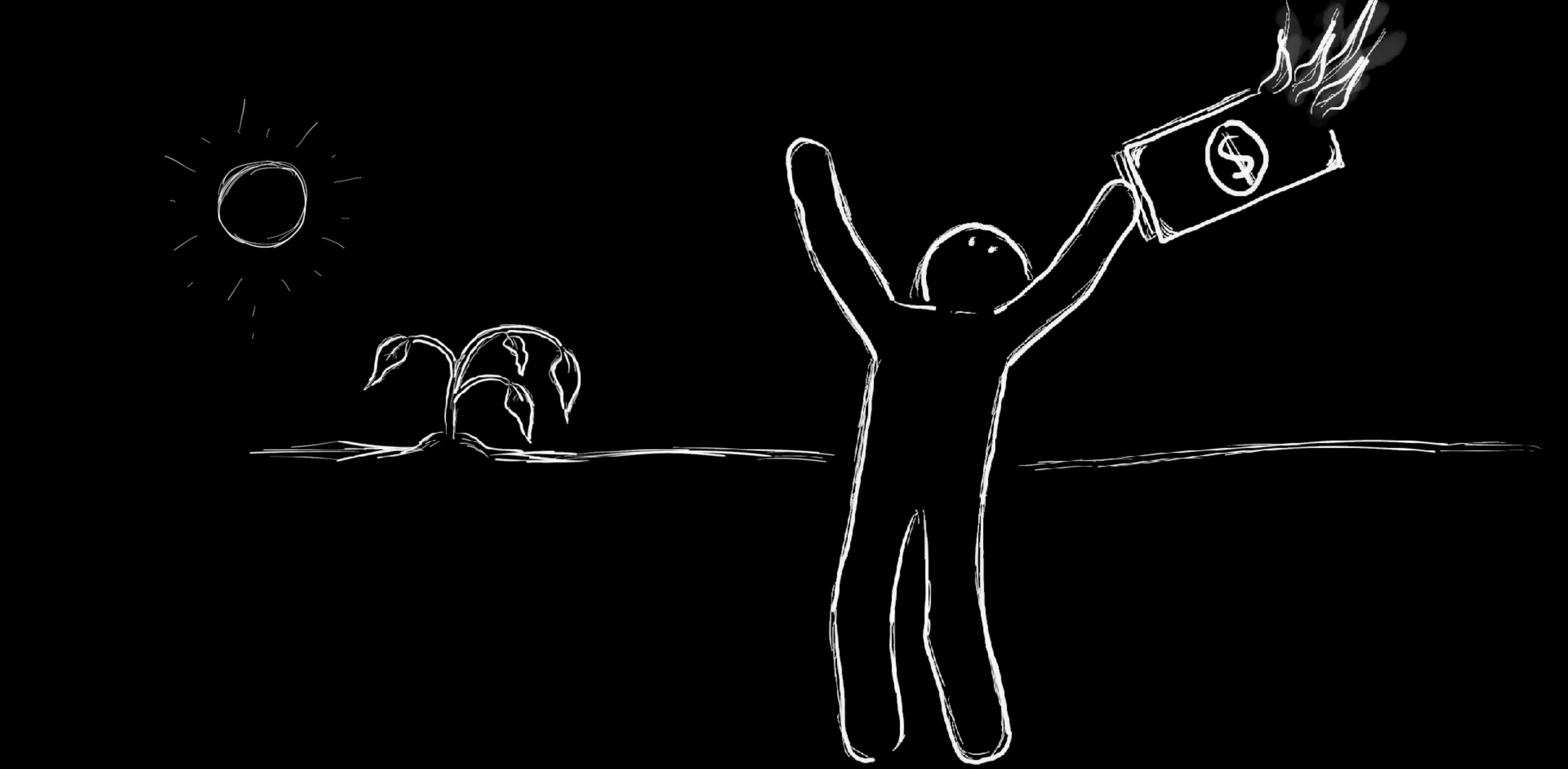
serverless?

serverless?

serverless?

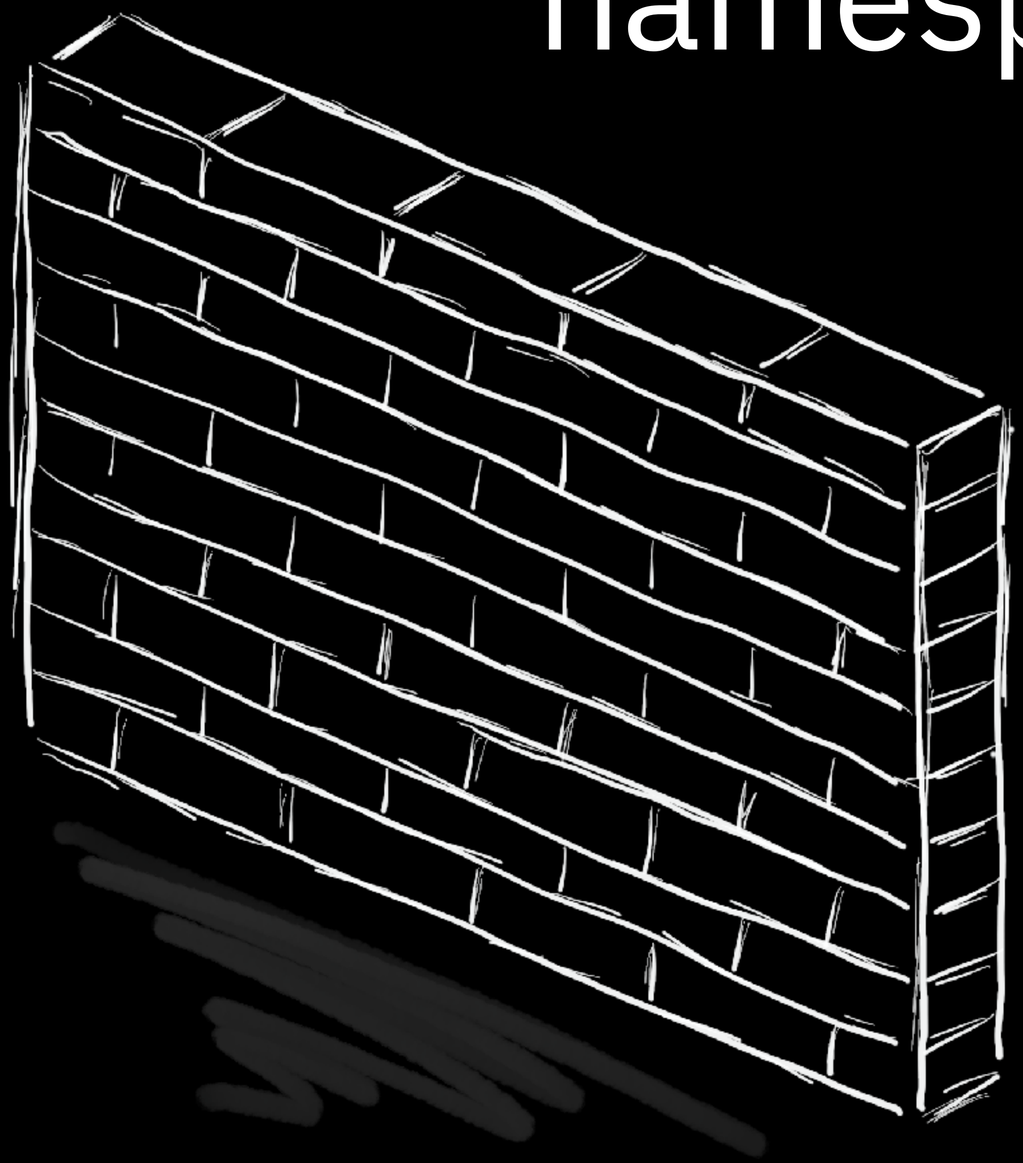




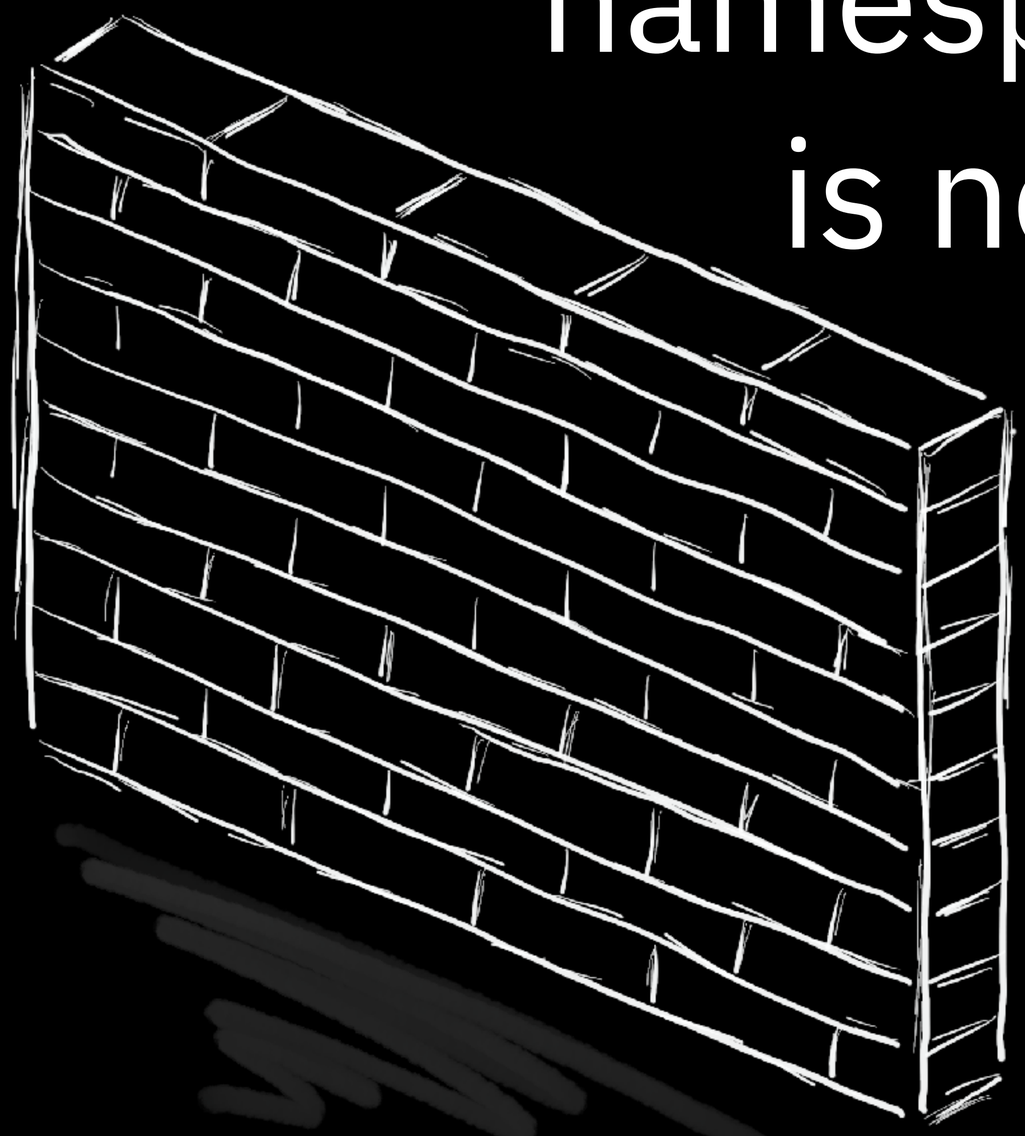


dr. malice

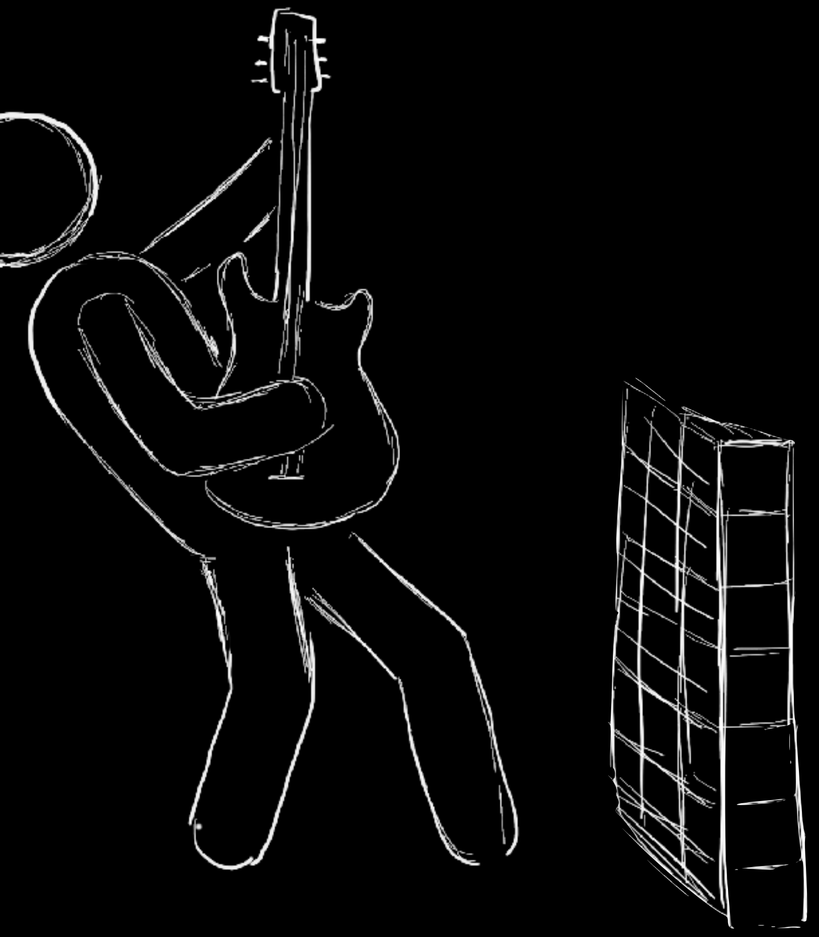
namespaces

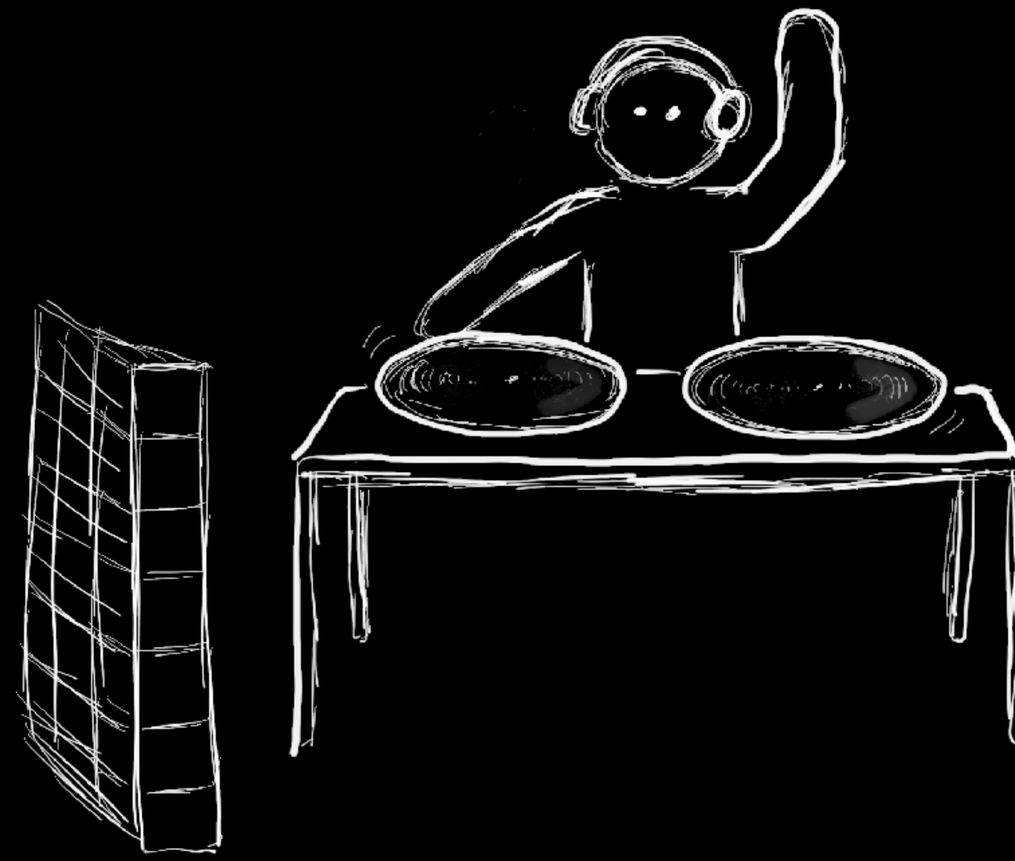


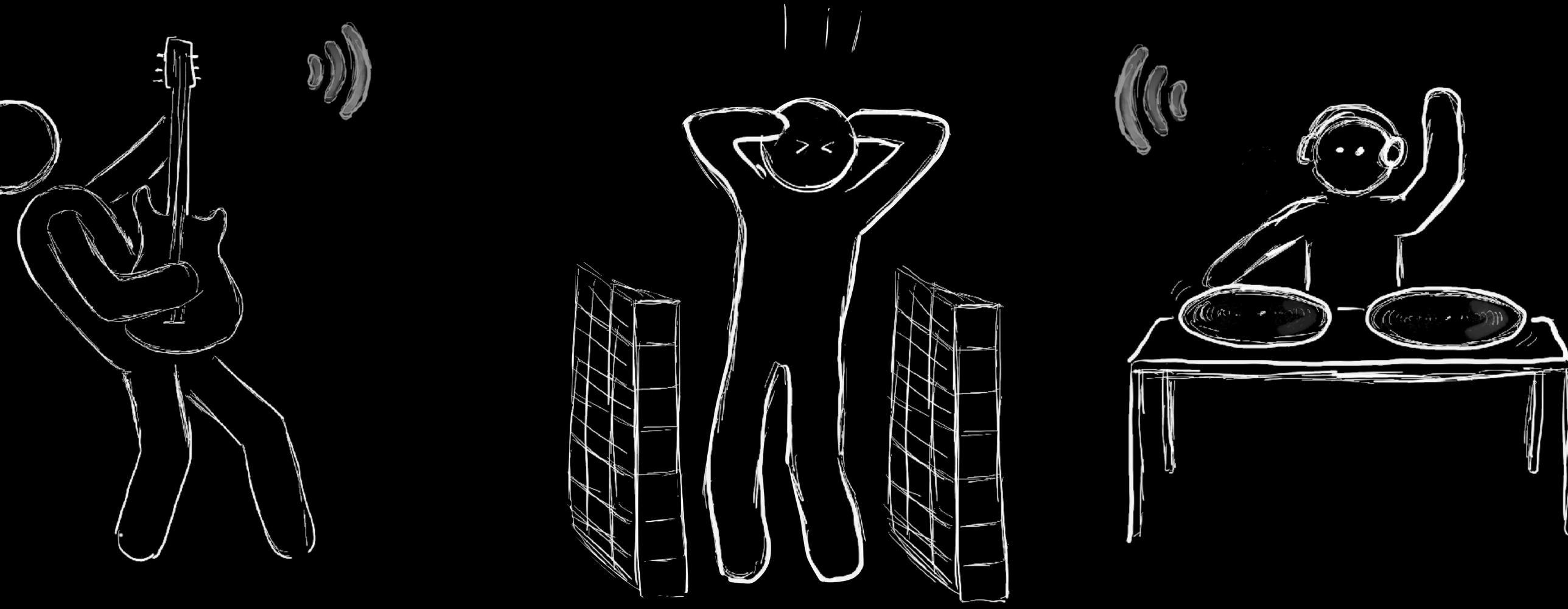
namespace isolation is not enough



conway’s law is for clusters, too

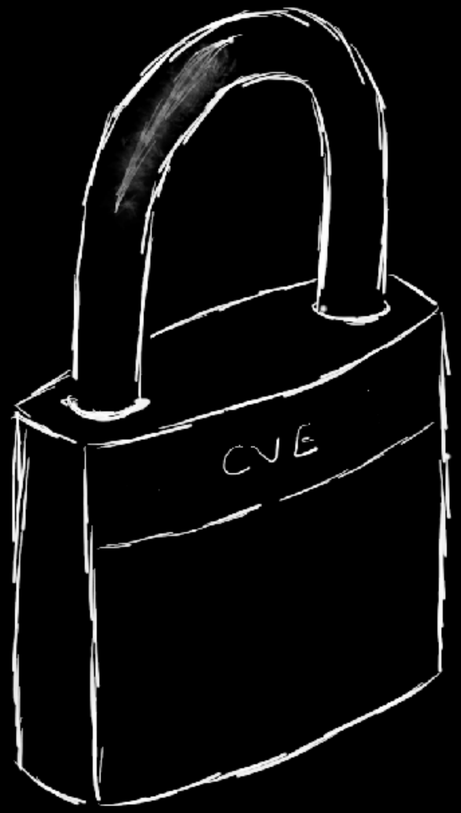


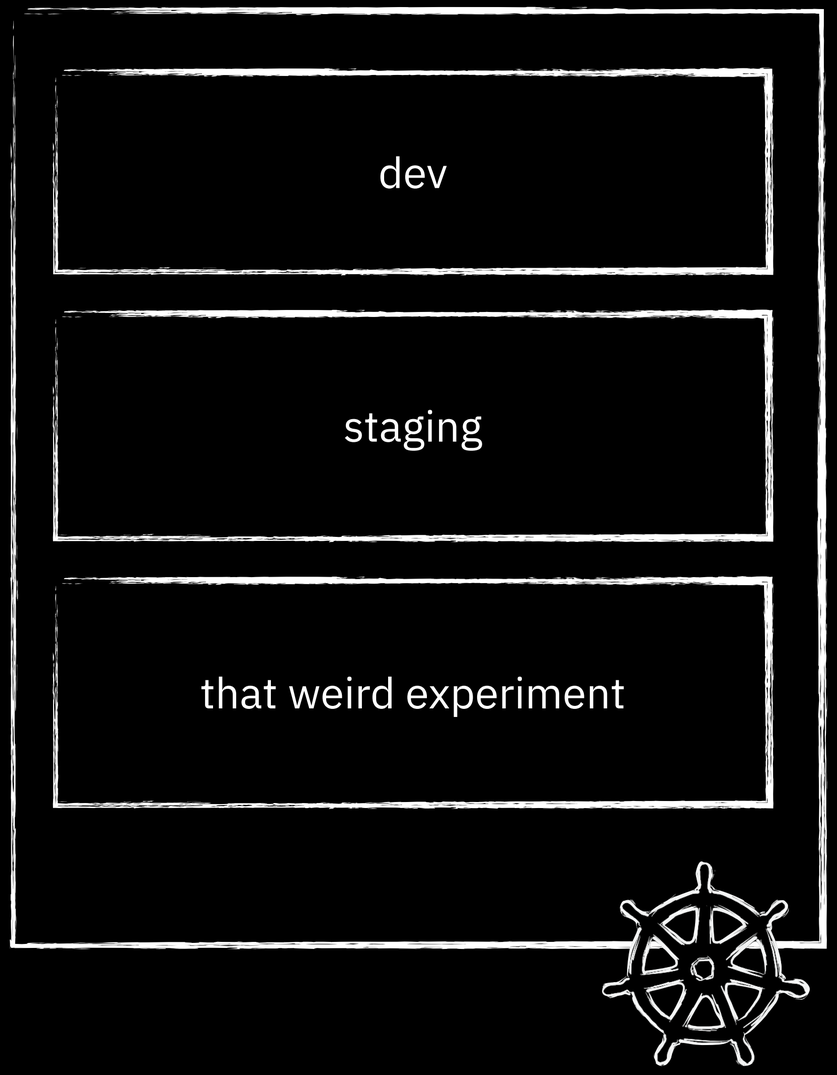


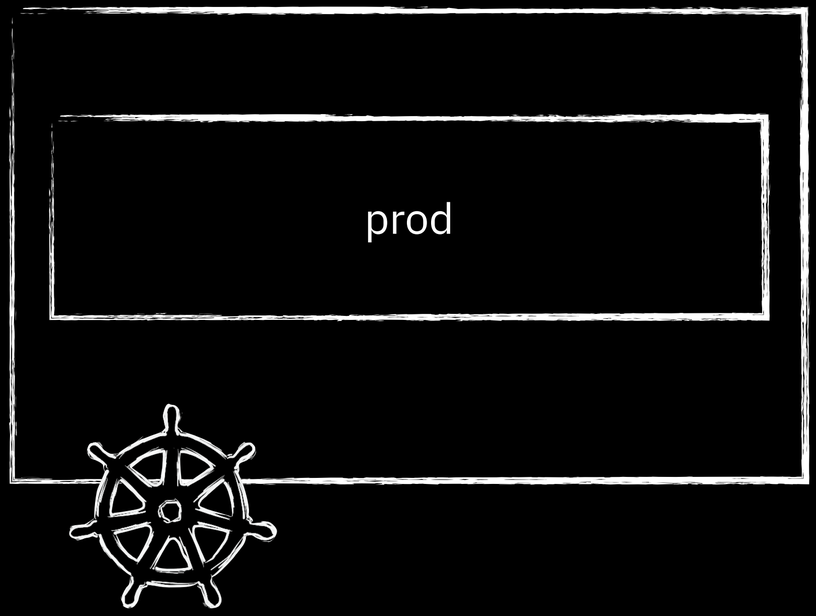


noisy neighbours

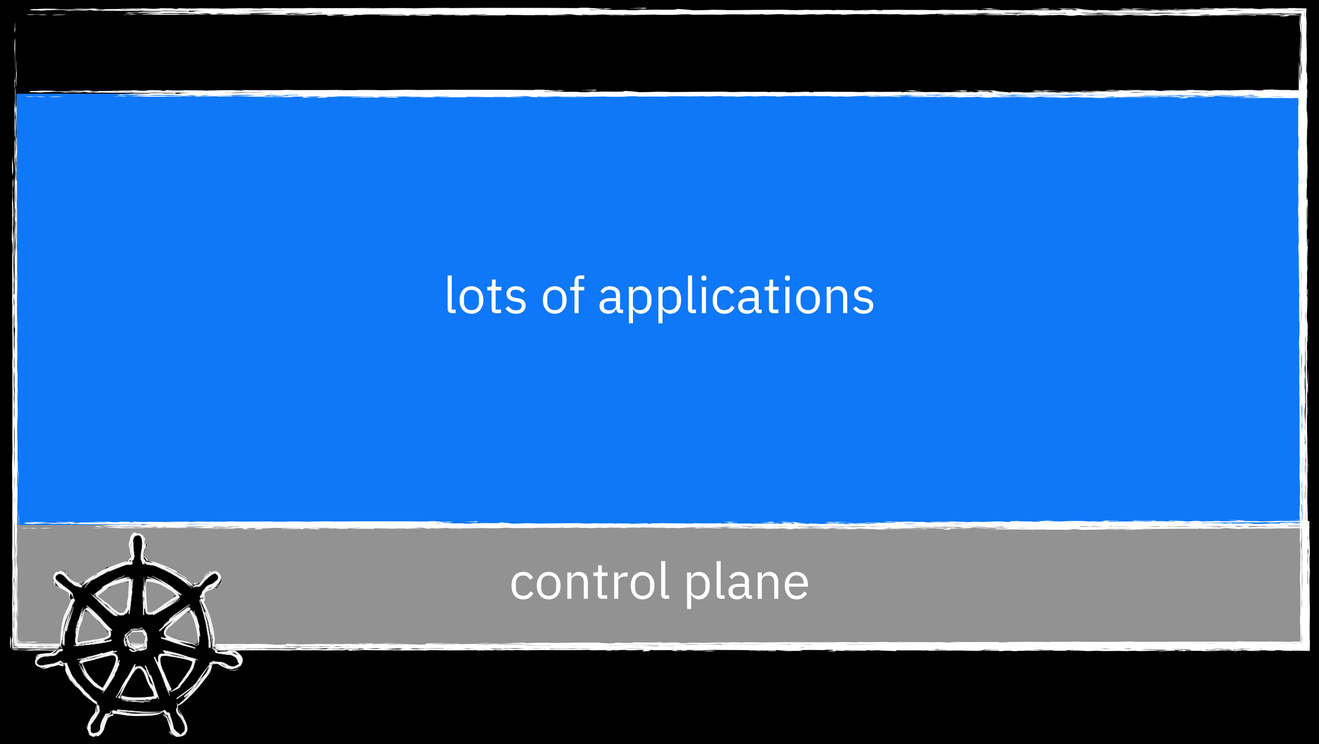
name collisions
scope errors

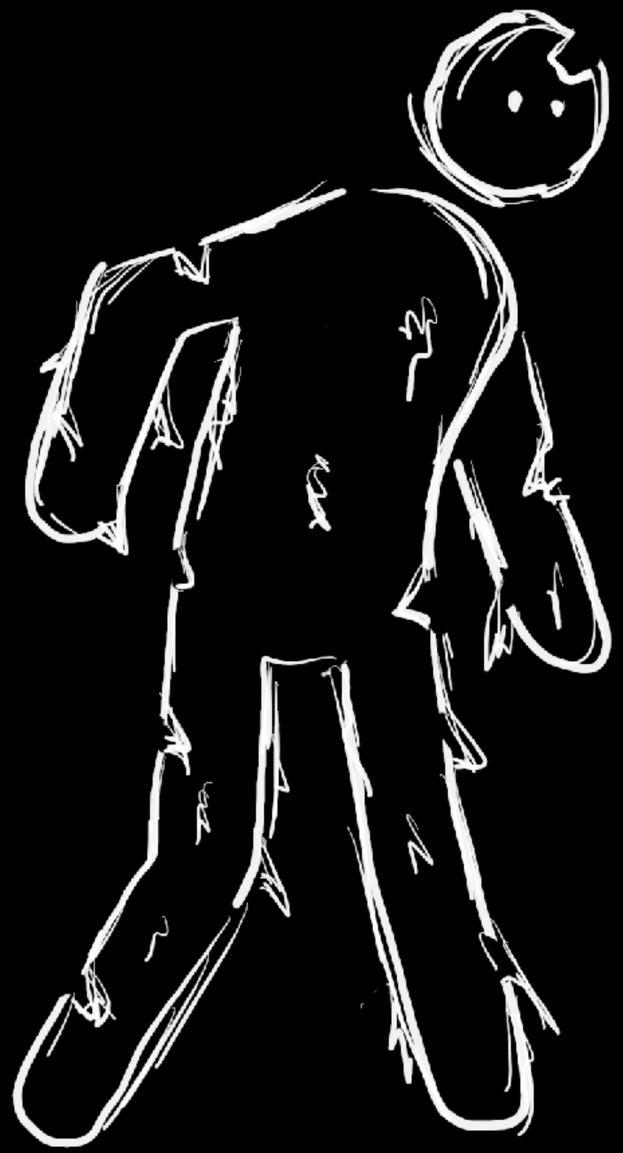

CVE


dev
staging
that weird experiment


prod

is this a win?





zombie workload

2017 survey

25%

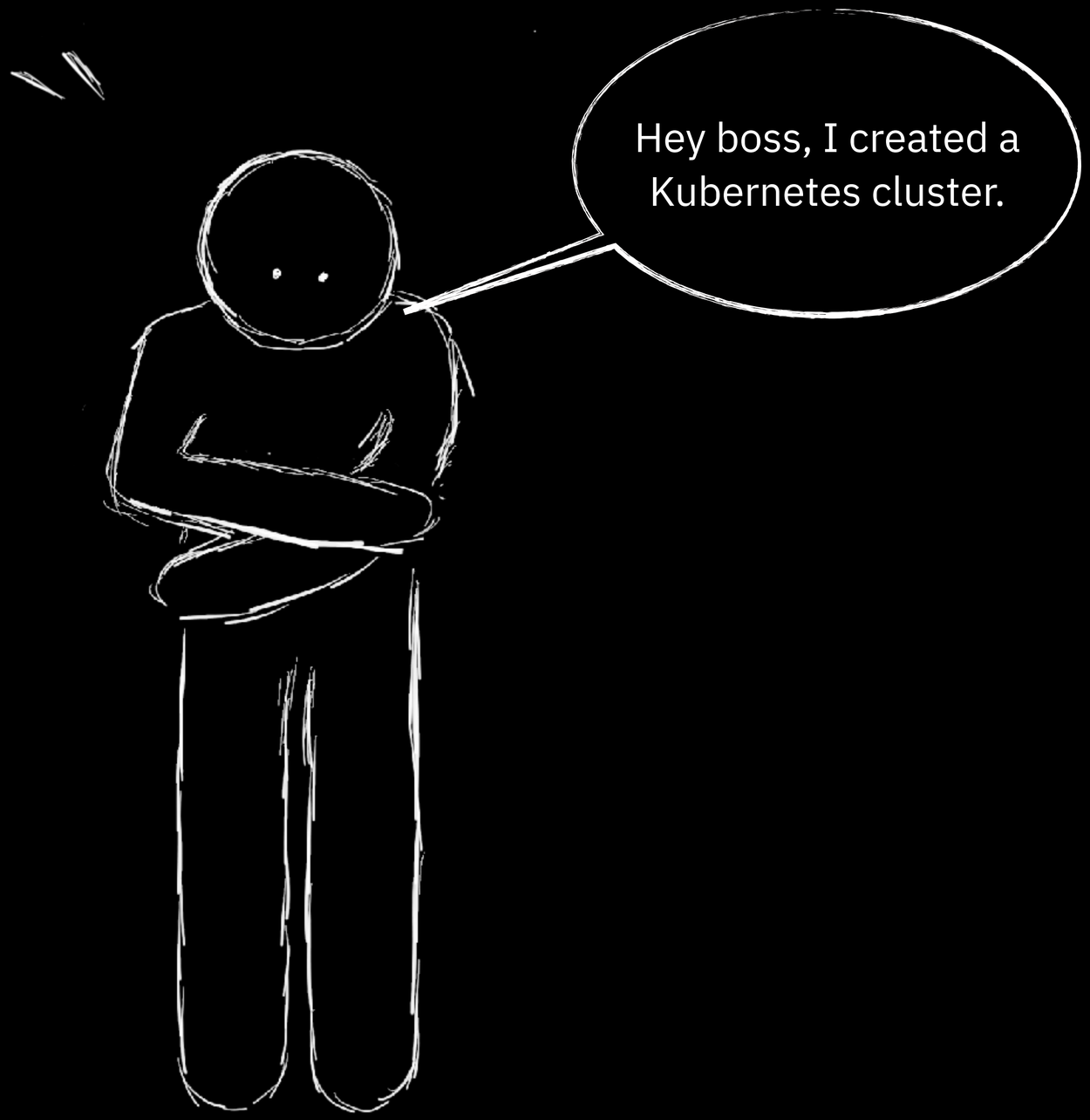
of 16,000 servers
doing no useful work

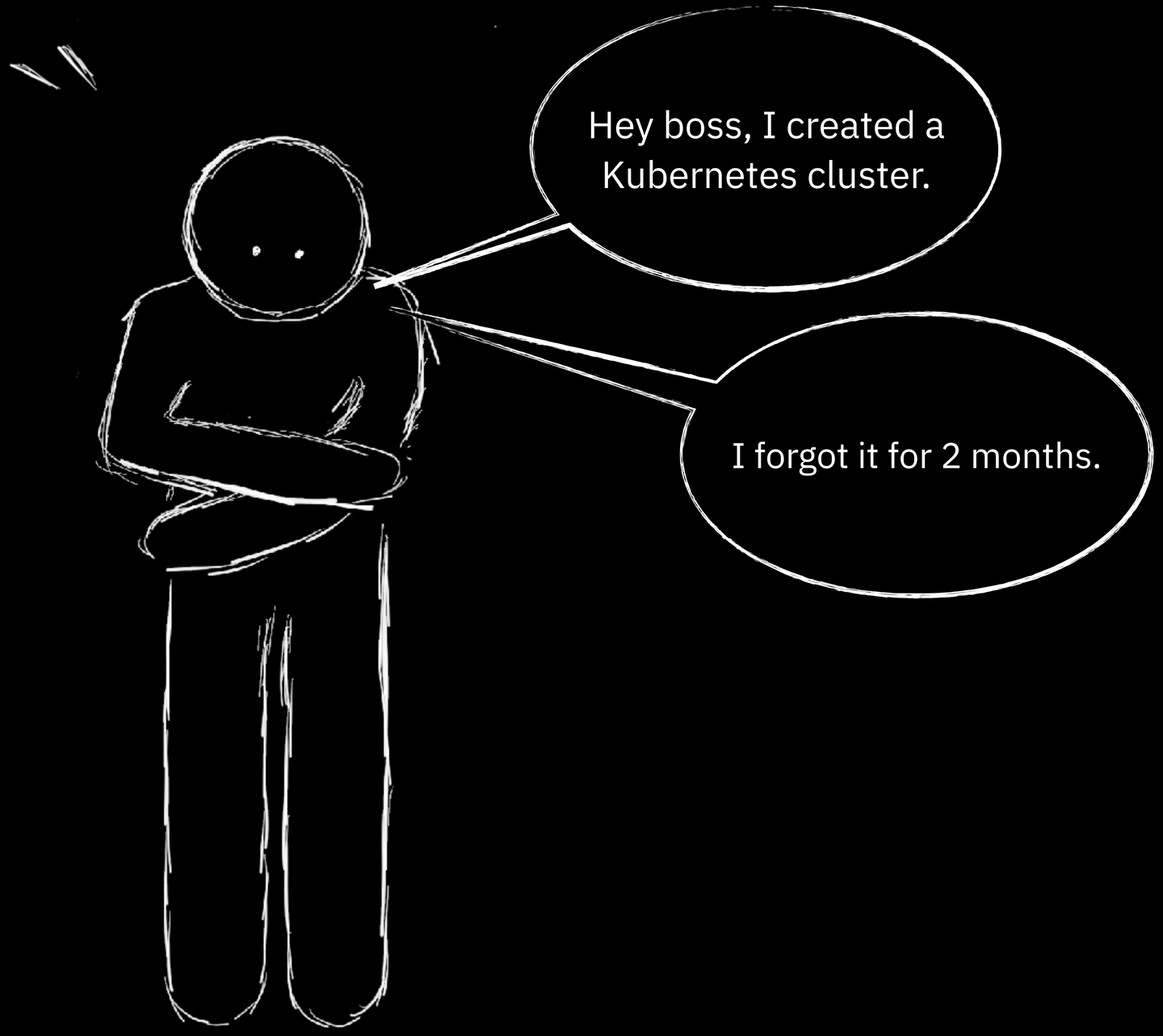
2017 survey

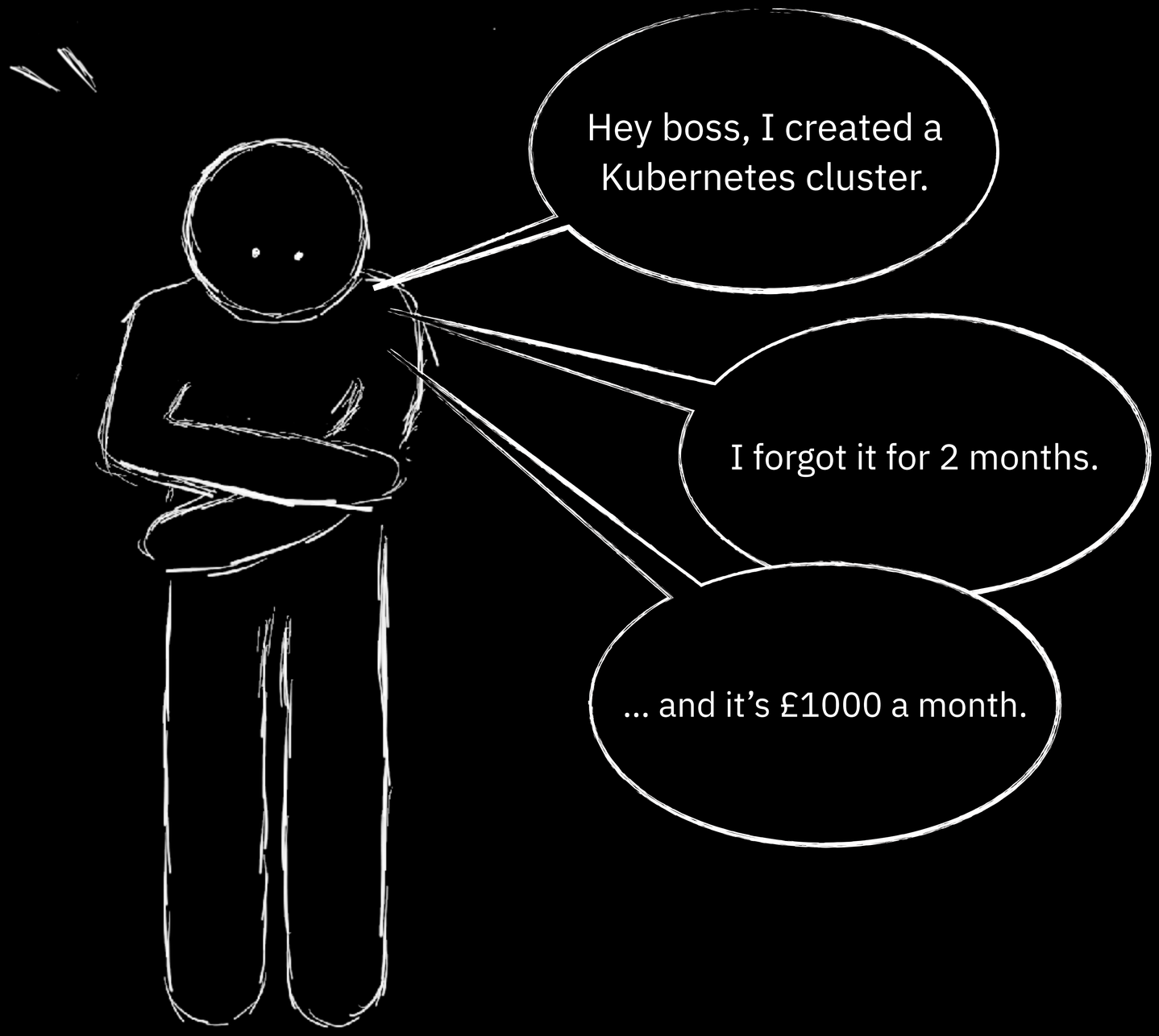
25%

of 16,000 servers
doing no useful work

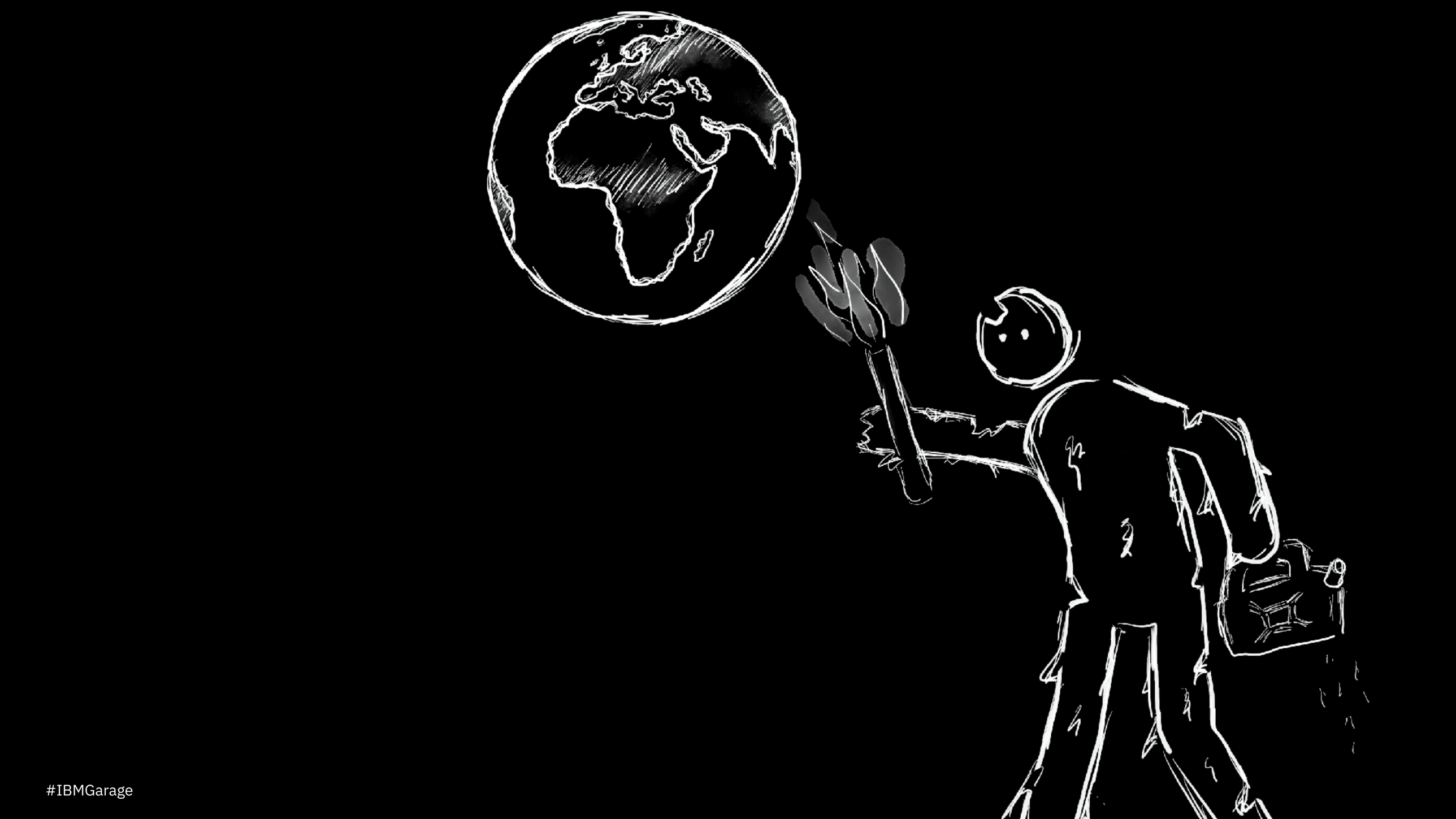
“perhaps someone forgot to turn them off”


Hey boss, I created a Kubernetes cluster.


Hey boss, I created a Kubernetes cluster.
I forgot it for 2 months.


Hey boss, I created a Kubernetes cluster.
I forgot it for 2 months.
... and it’s £1000 a month.

= £



#IBMGarage

is there a solution?

let’s figure out what all these cloud workloads are, since I’m **paying** for them

IT Department, UK Bank

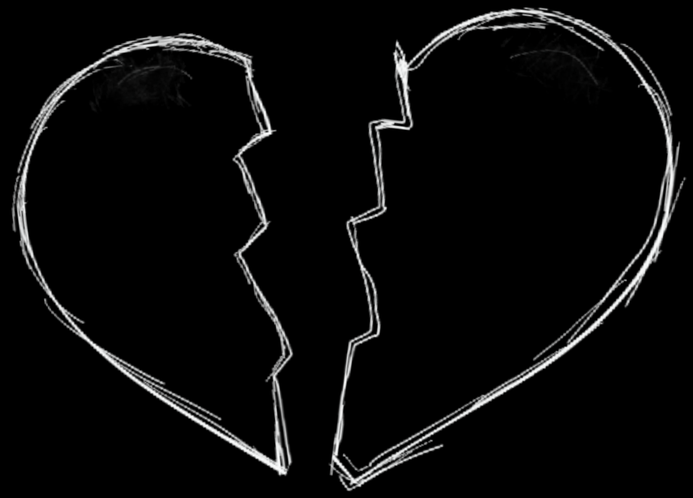
long meetings

let’s figure out what all these cloud workloads are, since I’m **paying** for them

IT Department, UK Bank

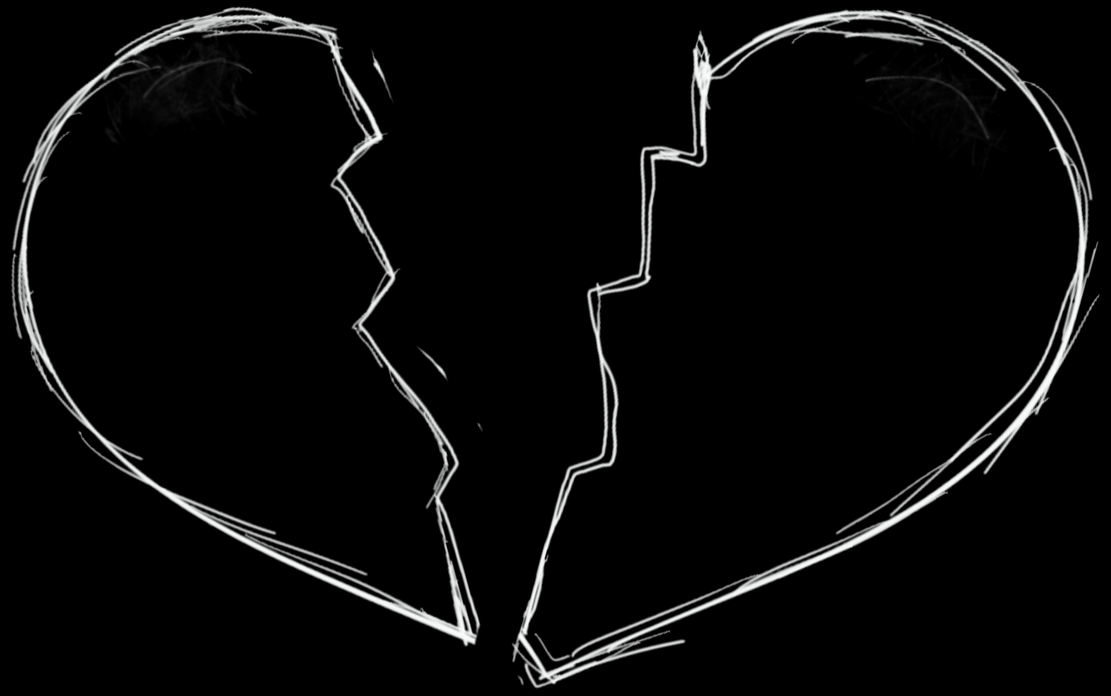
long meetings

tags



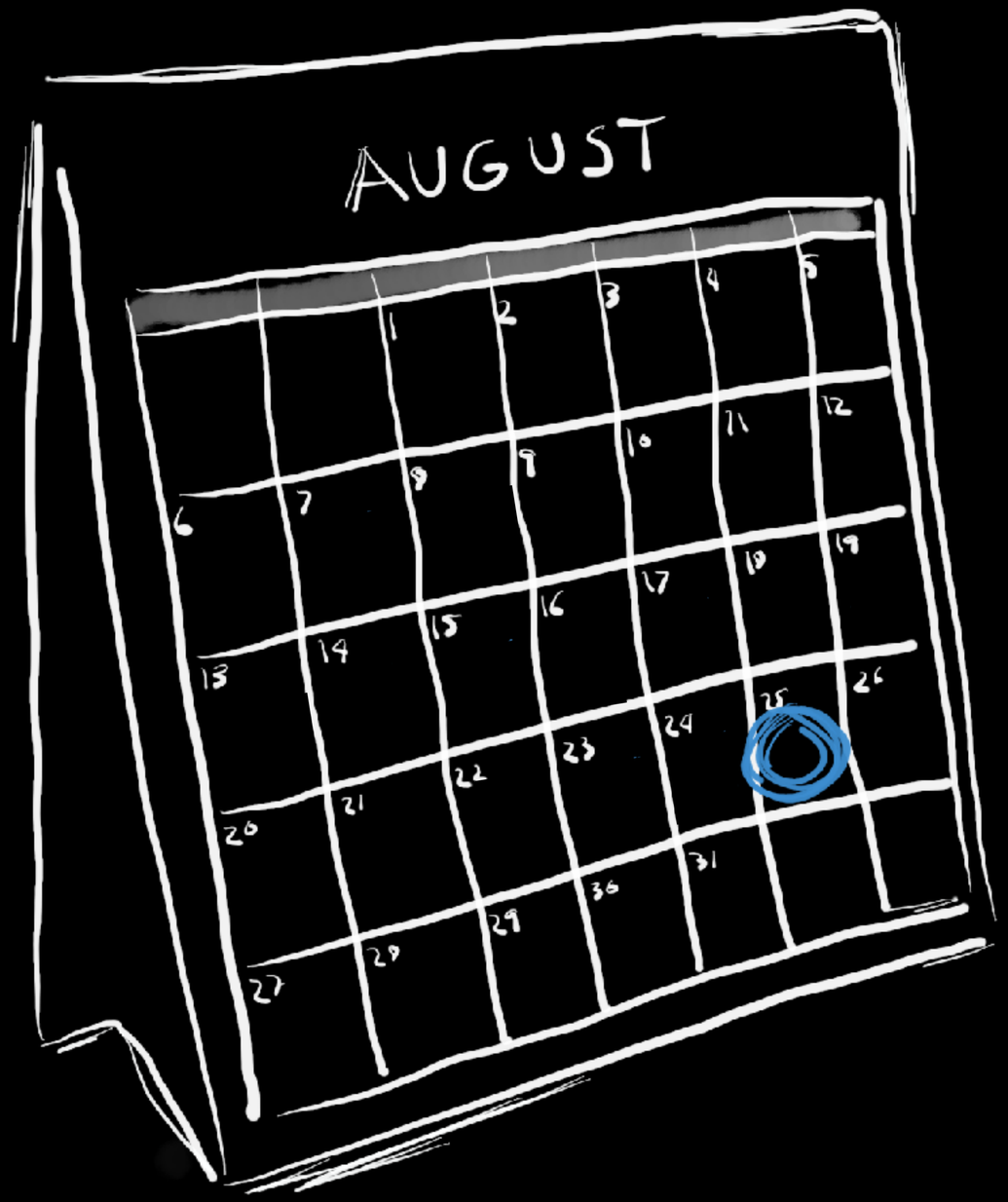
governance

make it **easiest** to do the right thing



make it **easiest** to
do the right thing

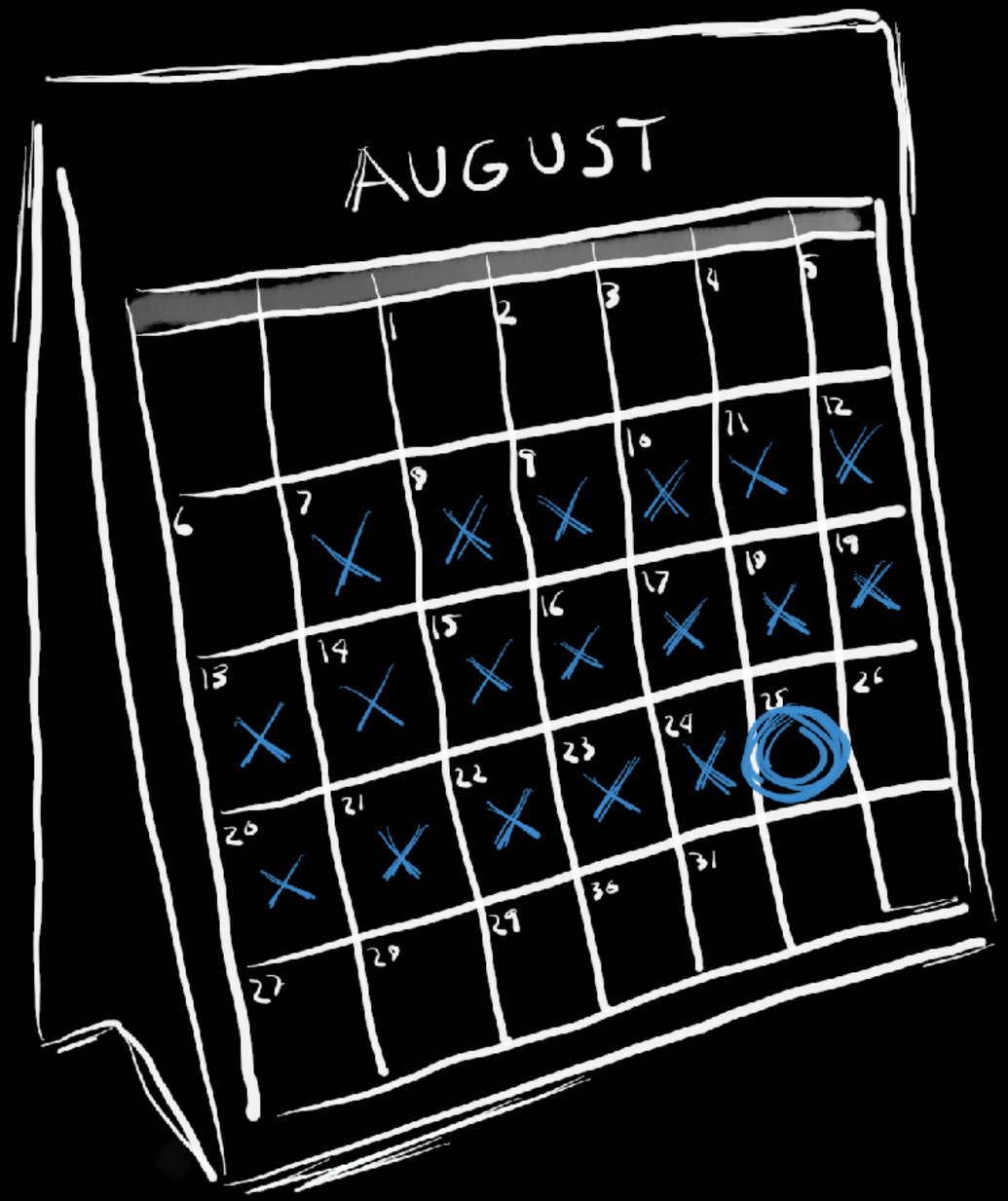
finops



large bank, 2013

50%

reduction in CPUs with a lease system



large bank, 2013

50%

reduction in CPUs with a lease system

“chaos testing”

(turning it off and waiting for shrieks)

multicloud management

traffic monitoring

is K8s the solution?

is K8s the solution?

(is K8s zombie-proof?)

is the cloud zombie-proof?

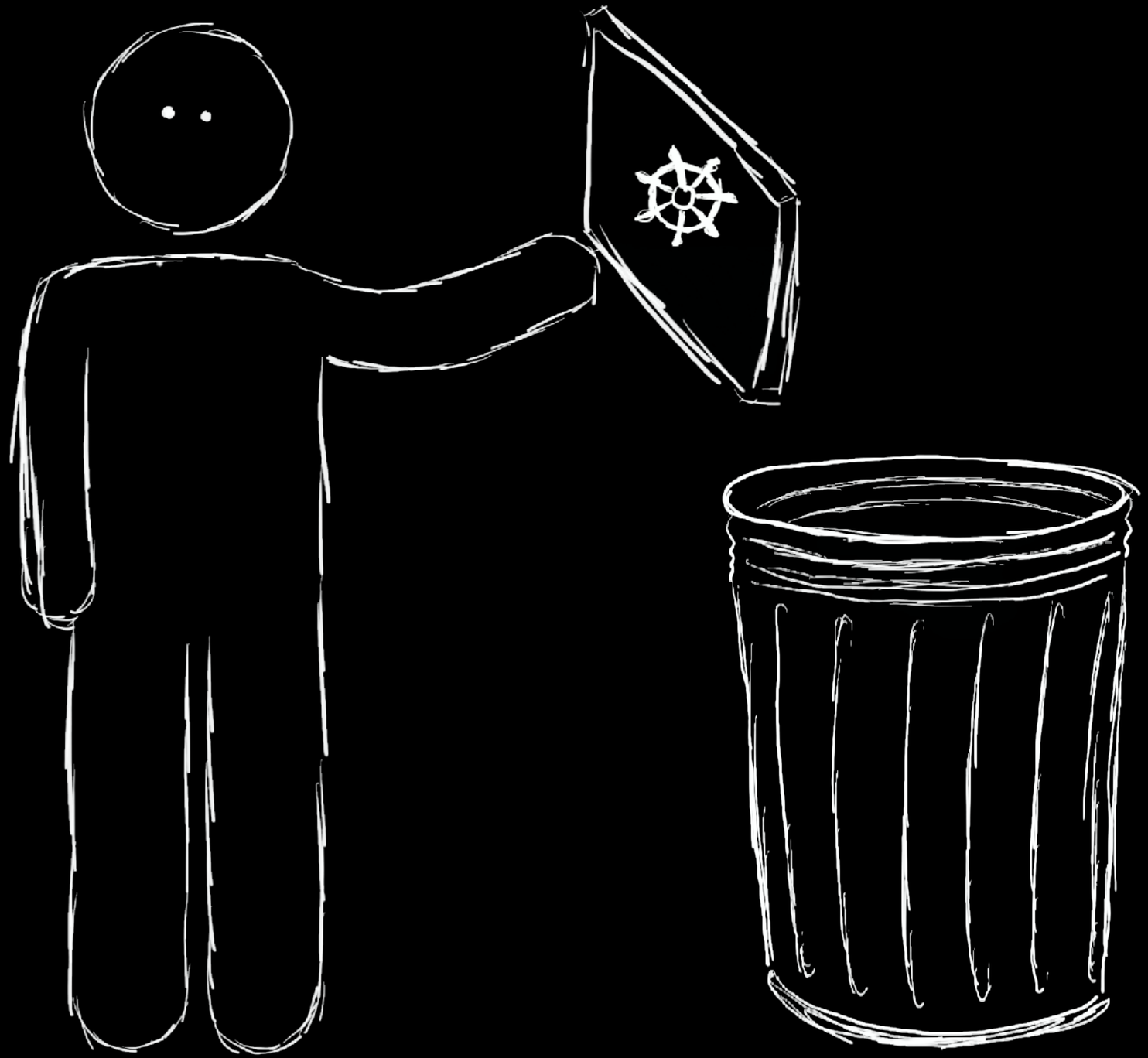

shut it down?
but ... what if I
need this
cluster later?

the IKEA cognitive bias

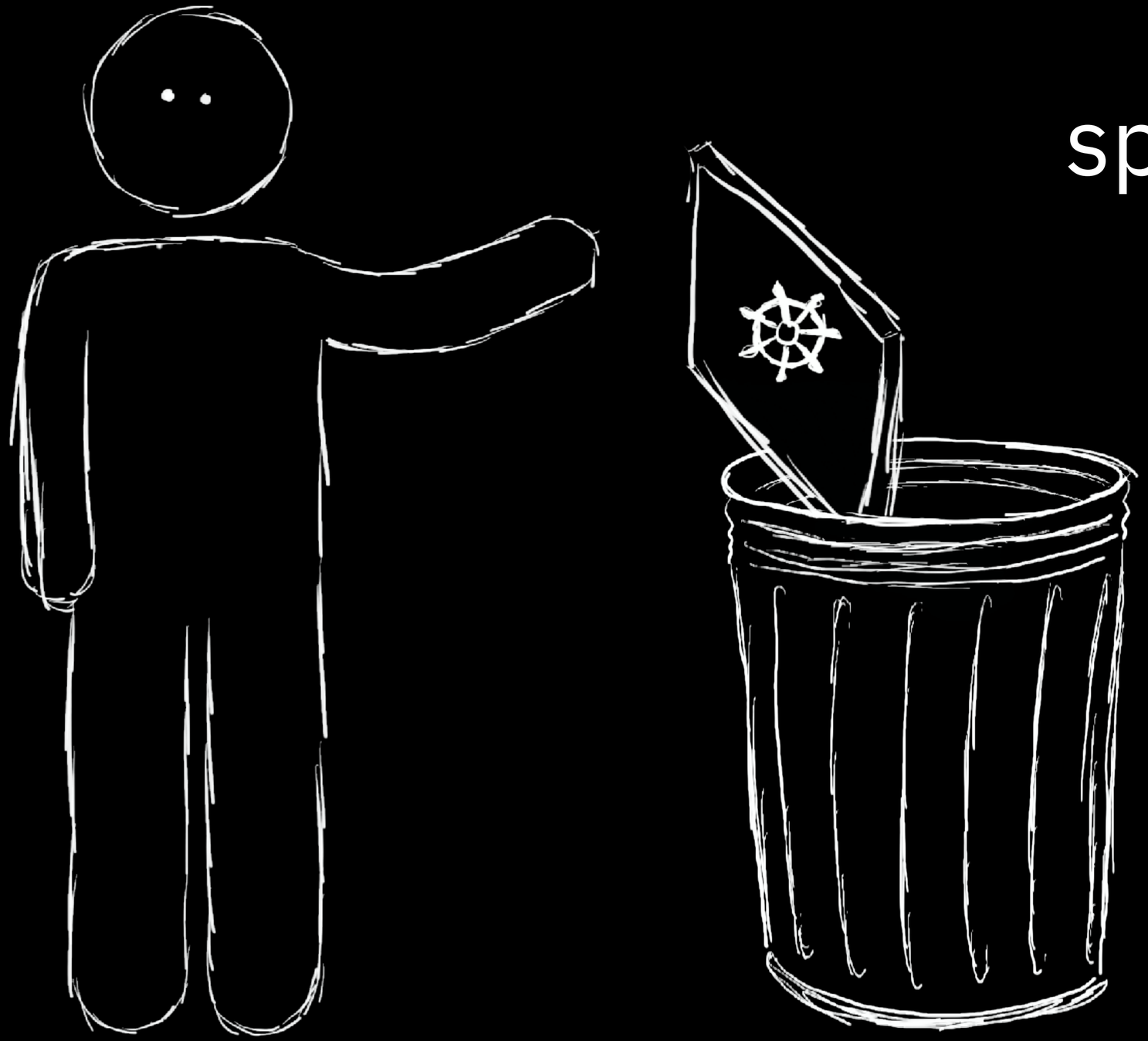
gitops

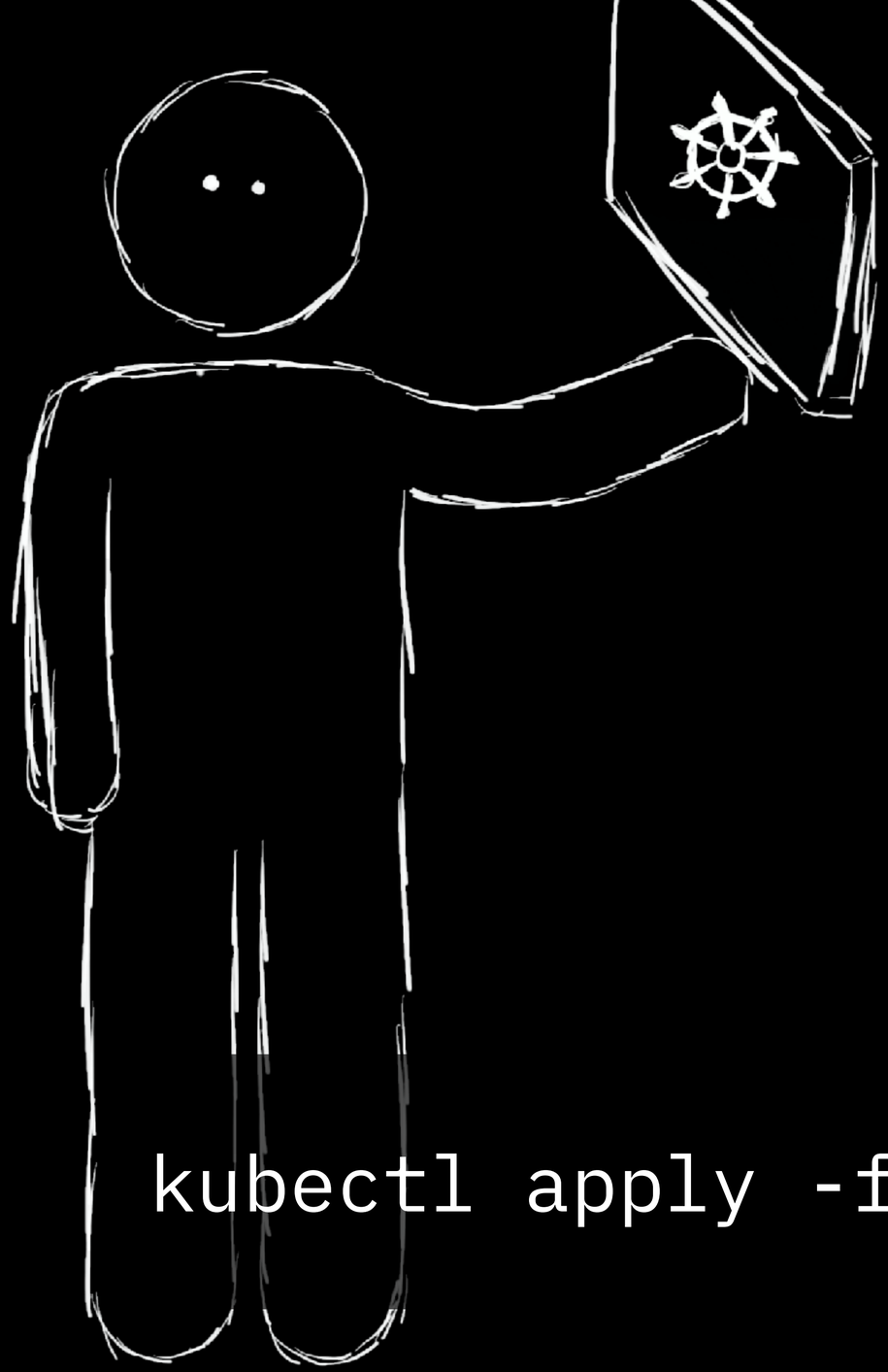
gitops

(infrastructure as code)



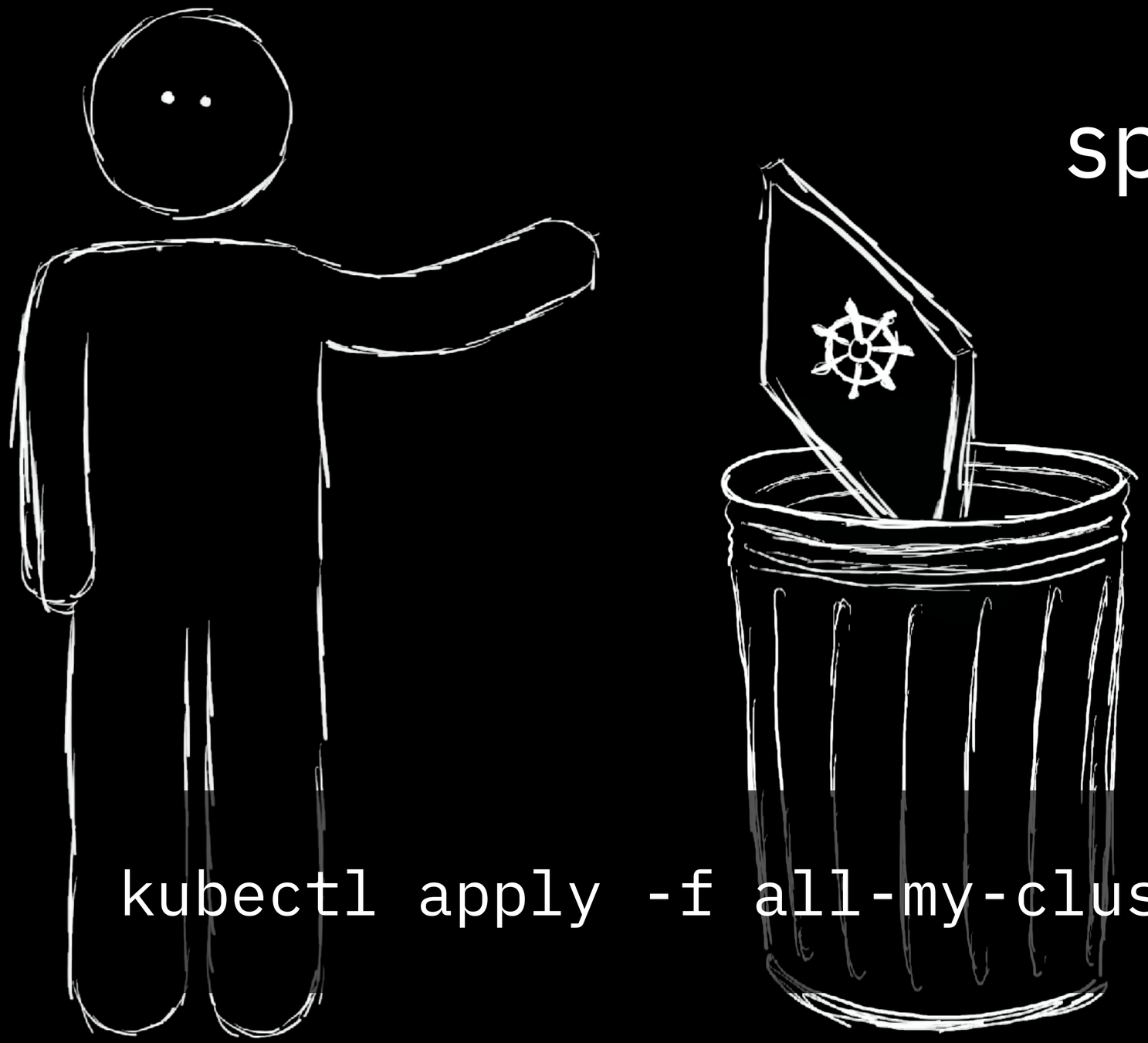
spin it down





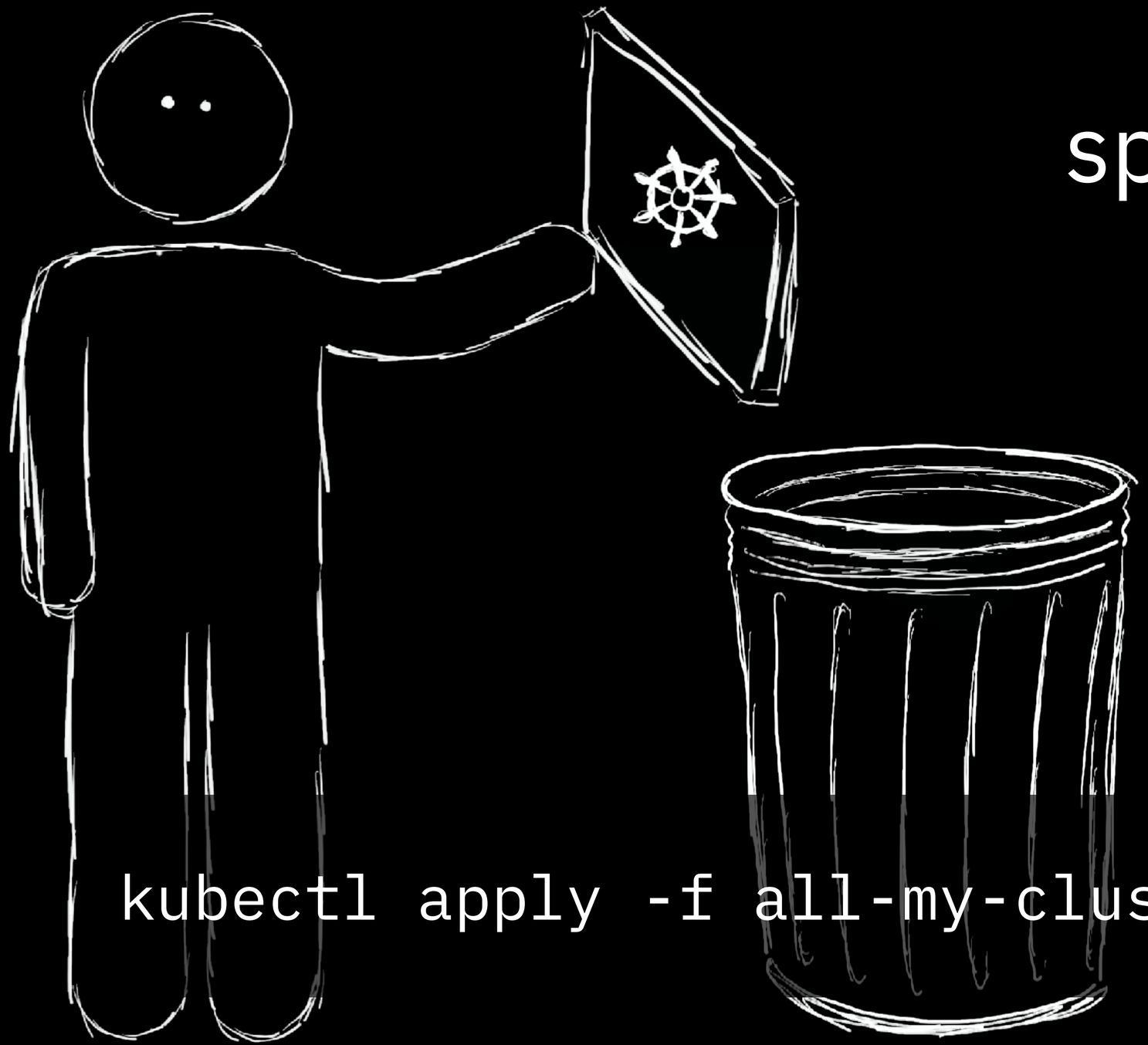
spin it down
spin it up

```
kubectl apply -f all-my-cluster/
```



spin it down
spin it up

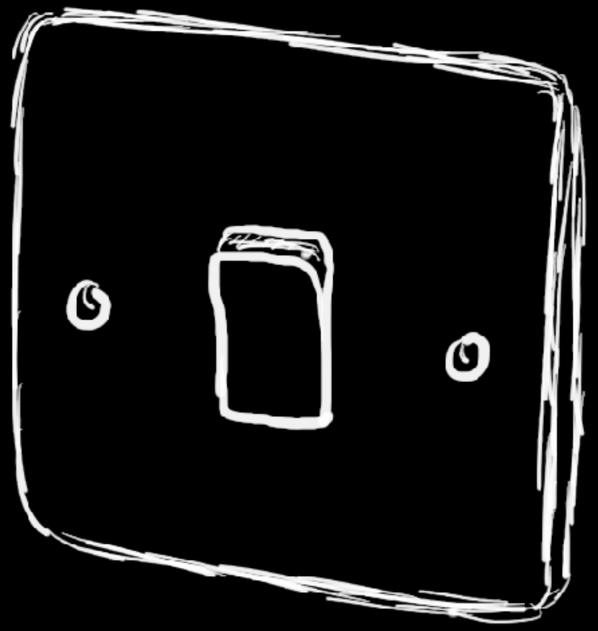
```
kubectl apply -f all-my-cluster/
```



spin it down
spin it up

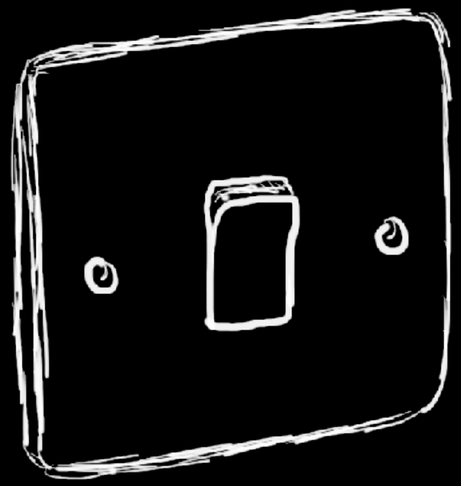
```
kubectl apply -f all-my-cluster/
```

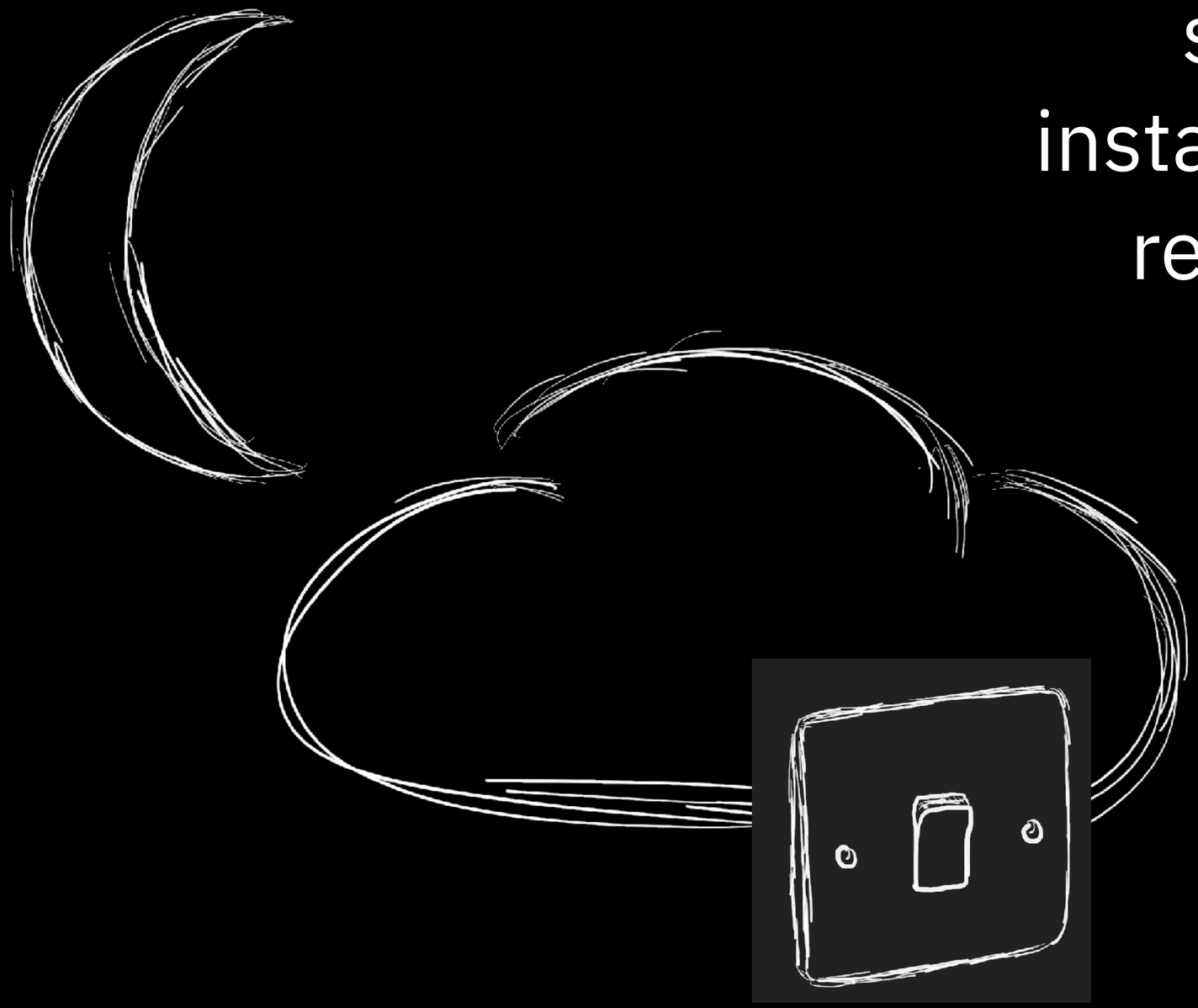
spinning down clusters: the new lights off?



spinning down clusters:
the new lights off?

oh. it **is**.





shutting down
instances out of hours
reduced costs by

37%

(I made up the statistic, @darkandnerdy)

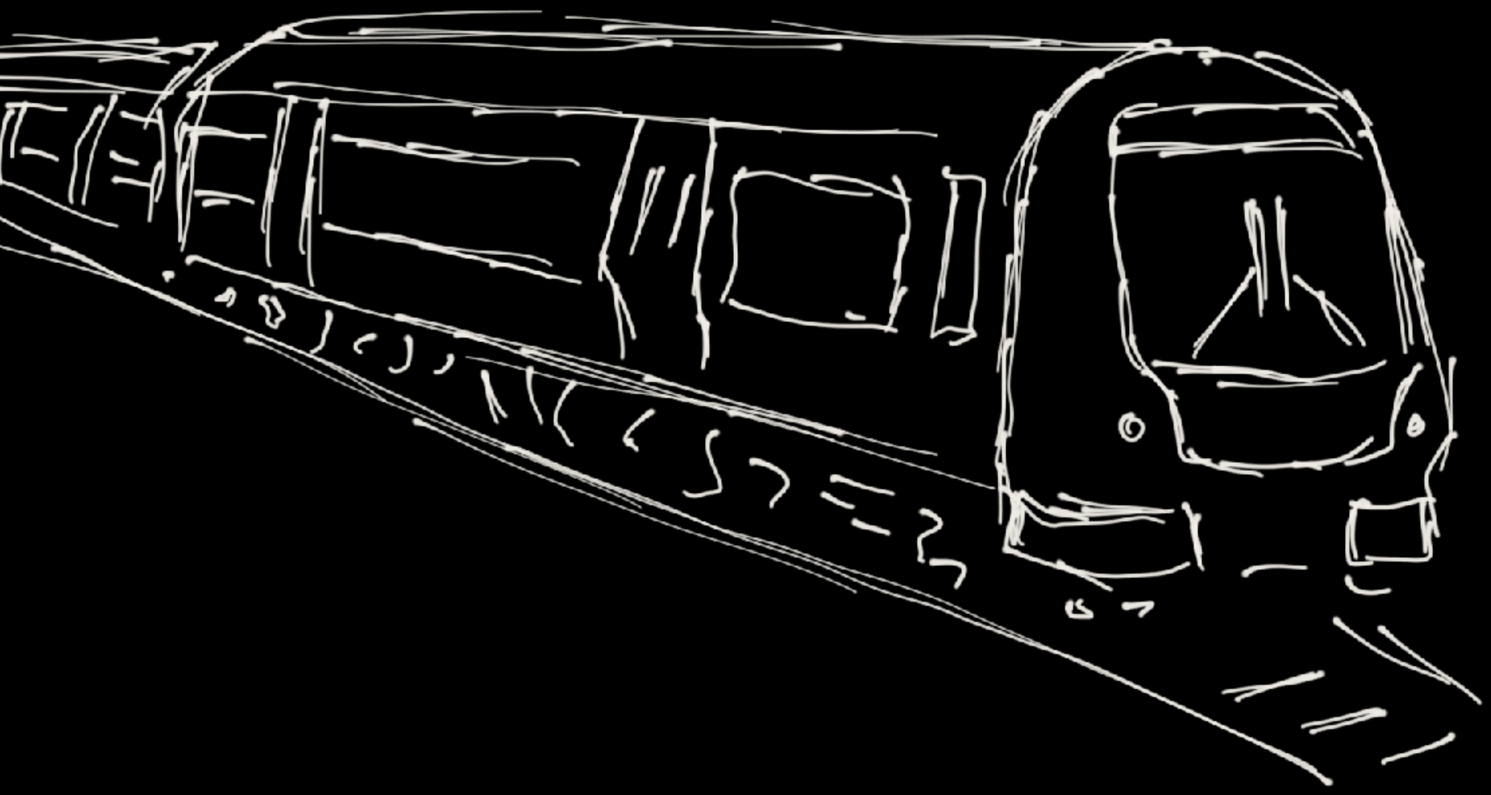
cautions

beware
micro-optimisation
theatre

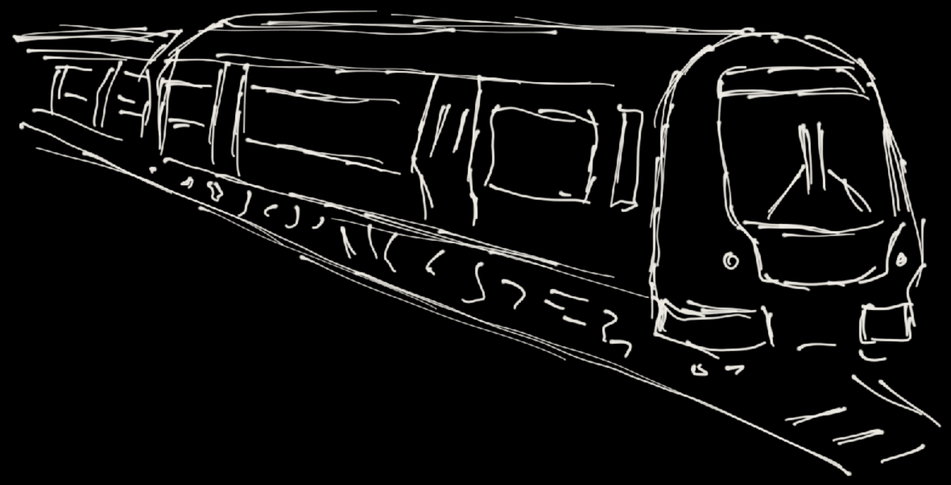

THE LEAN SERIES
ERIC RIES, SERIES EDITOR
Jez Humble, Joanne Molesky & Barry O'Reilly
LEAN
ENTERPRISE
How High Performance
Organizations
Innovate at Scale
O'REILLY

“unsustainable”

“sustainable”



“sustainable”

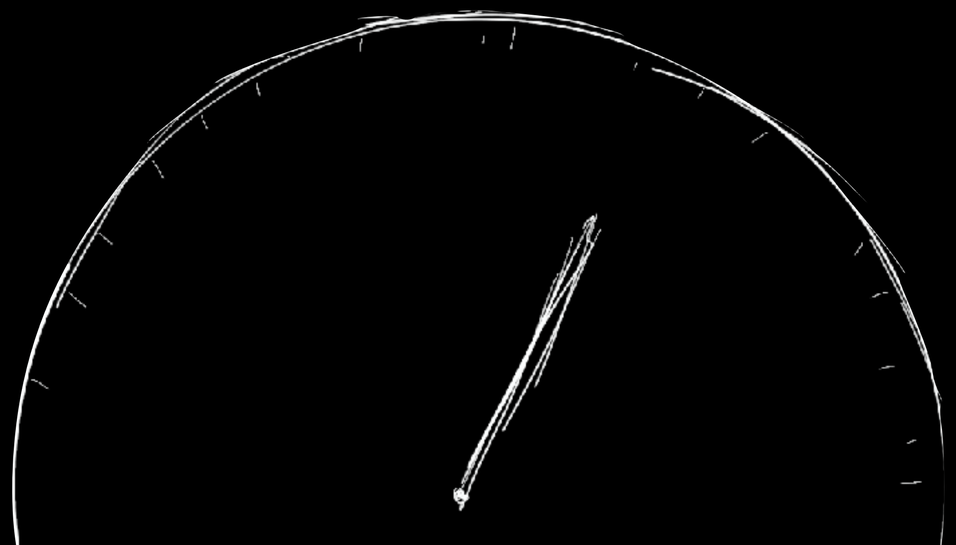


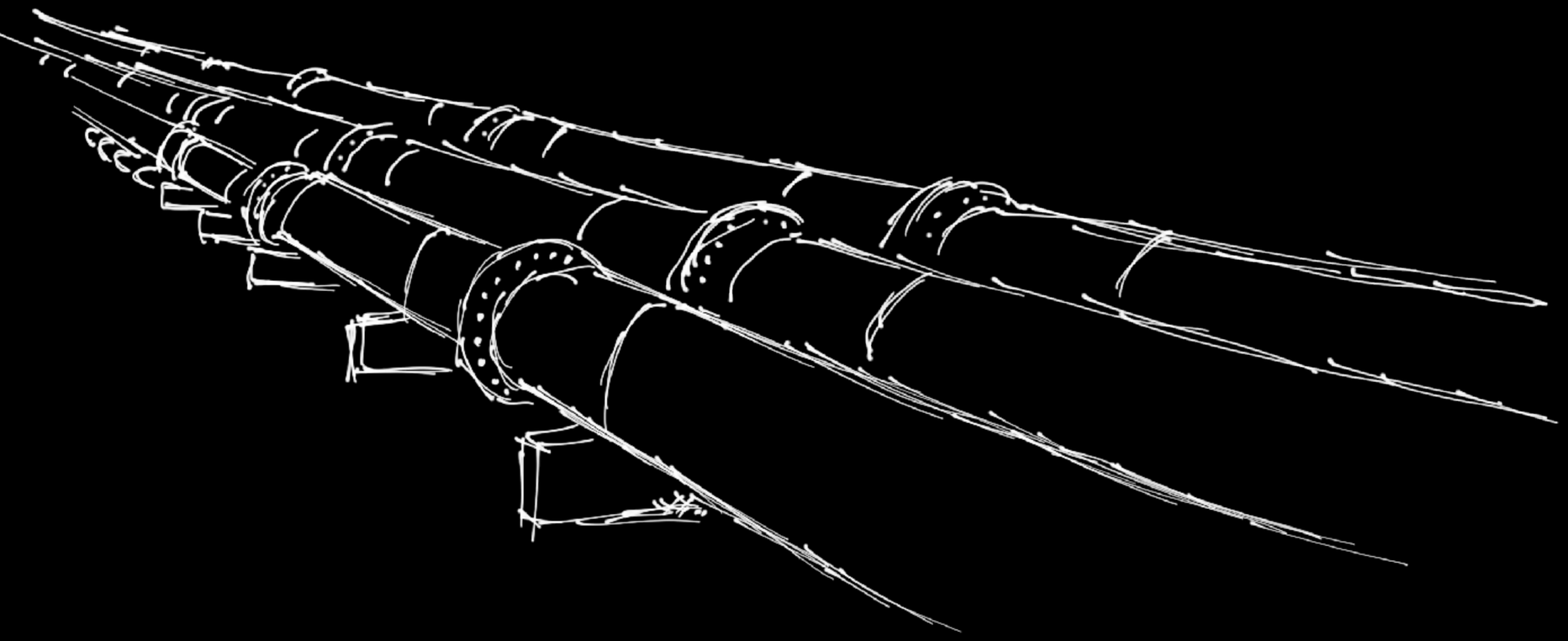
fixing the **wrong** problem

but ... every little helps...

think about the **other** things you're **not** optimising

knowledge helps us
focus on optimisations
that **matter**



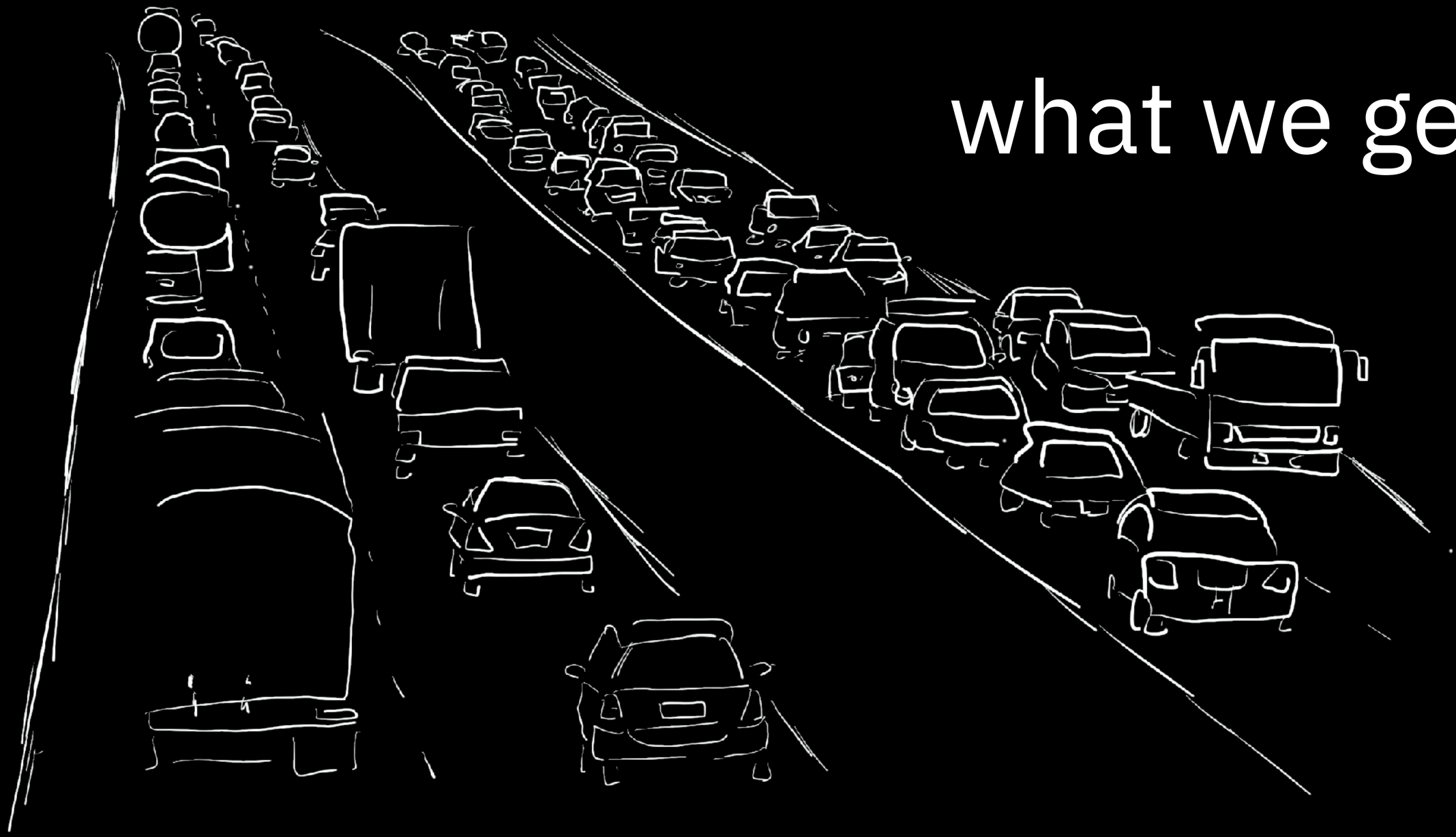


jevons’ paradox

the highway problem

what we imagine when we widen roads

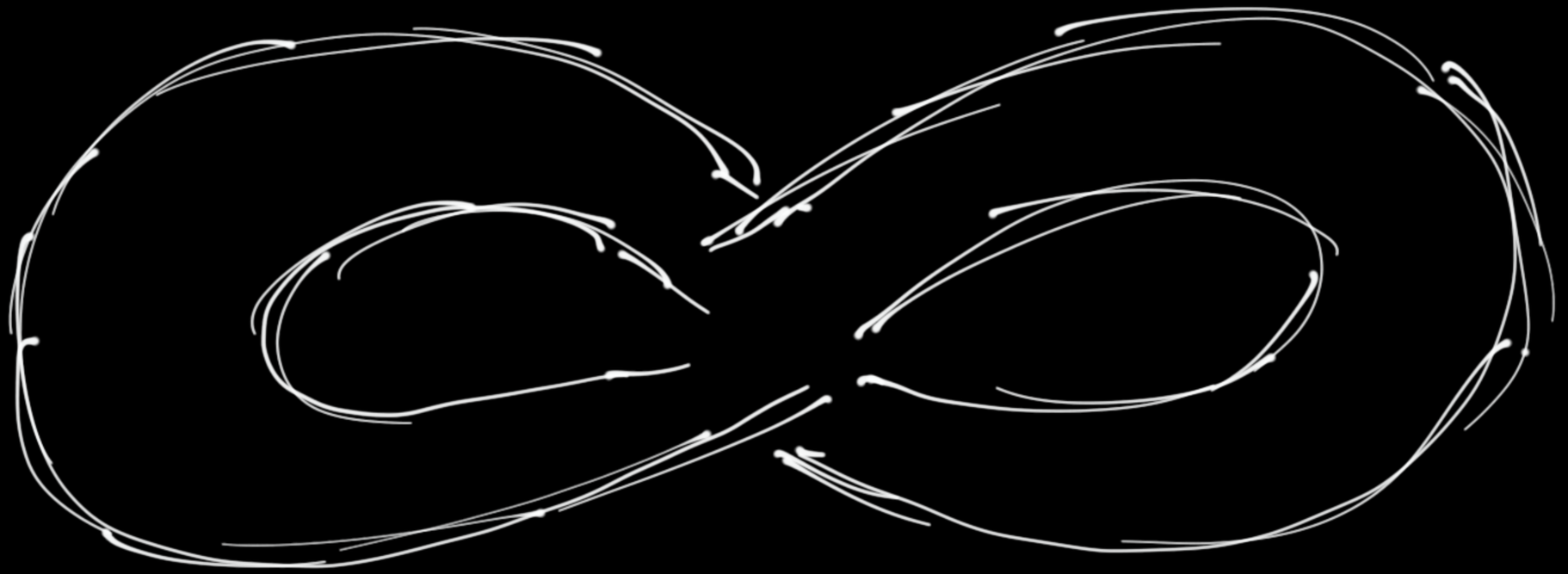
what we get



unsolved problem

opportunity

sustainability == continuous lifecycle



1-2%

tool creators, support

1-2%

tool creators, support

better utilisation

elasticity
multi-tenancy

de-zombification

visibility
disposability

1-2%

1-2%

users ...

1-2%

users ...

up utilisation

aim for elasticity
limit kubesprawl

de-zombification

know what you're using
turn it off

1-2%

@holly_cummins

questions