# OpenShift for AI and ML

William Benton • @willb • willb@redhat.com

**OpenShift** is **enterprise Kubernetes** with a **great developer experience.**

resource management
for apps and compute

**OpenShift** is **enterprise Kubernetes**
with a **great developer experience.**

efficient isolation,
secure by default

resource management
for apps and compute

**OpenShift** is **enterprise** **Kubernetes**

with a **great developer experience.**

efficient isolation,
secure by default

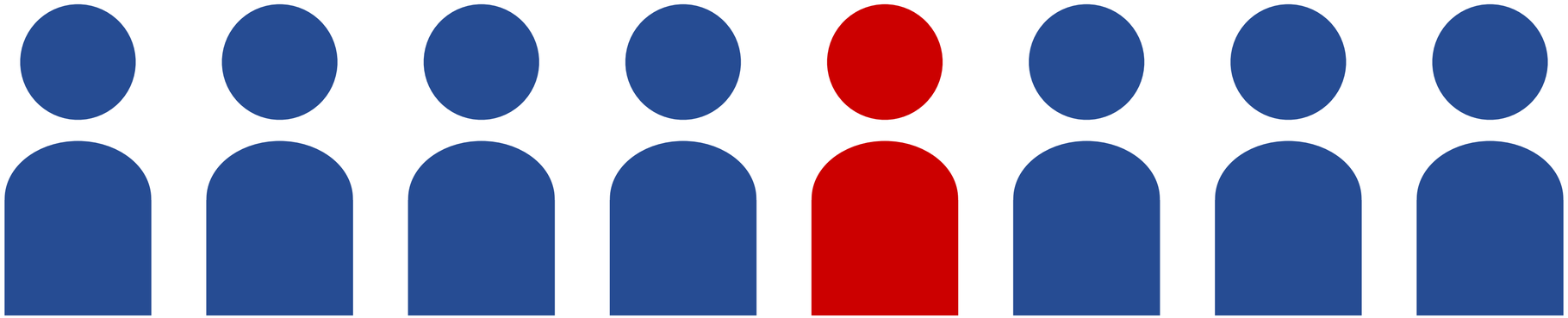
resource management
for apps and compute

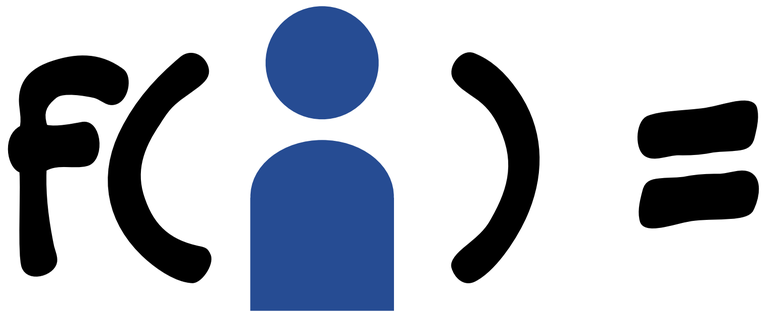
**OpenShift** is **enterprise Kubernetes**

with a **great developer experience.**

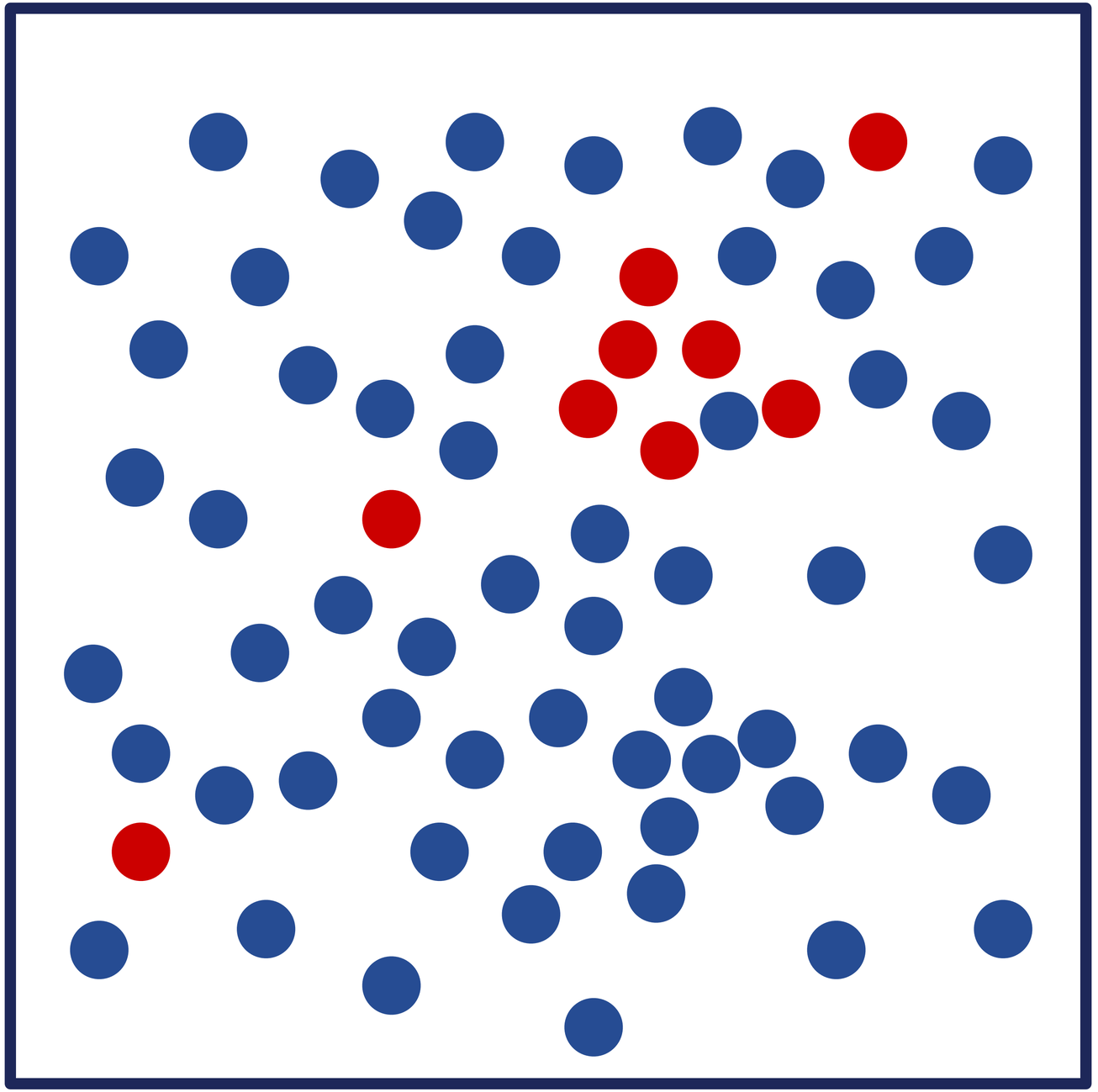
workflows to accelerate discovery

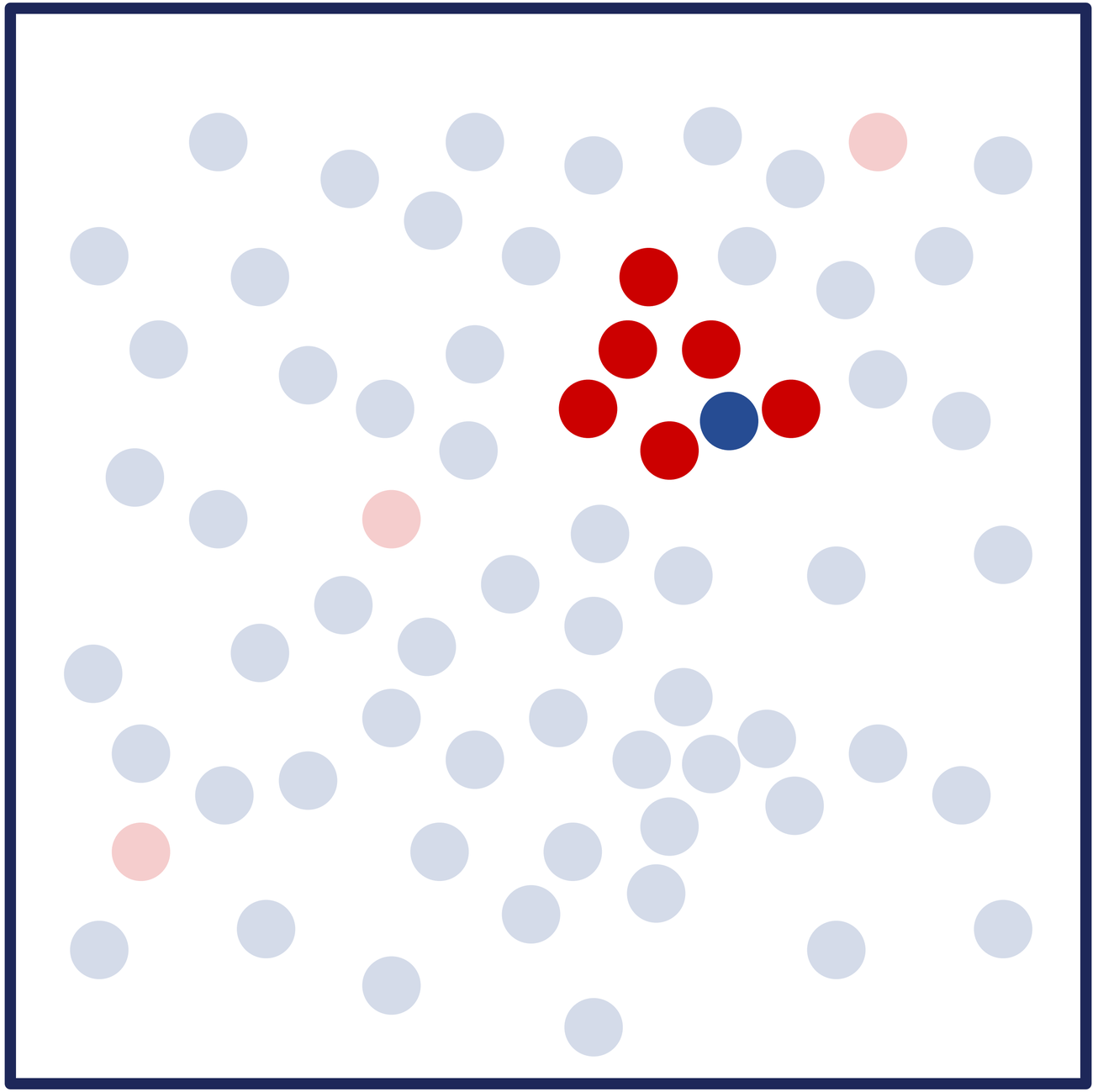
# What do machine learning workflows look like?

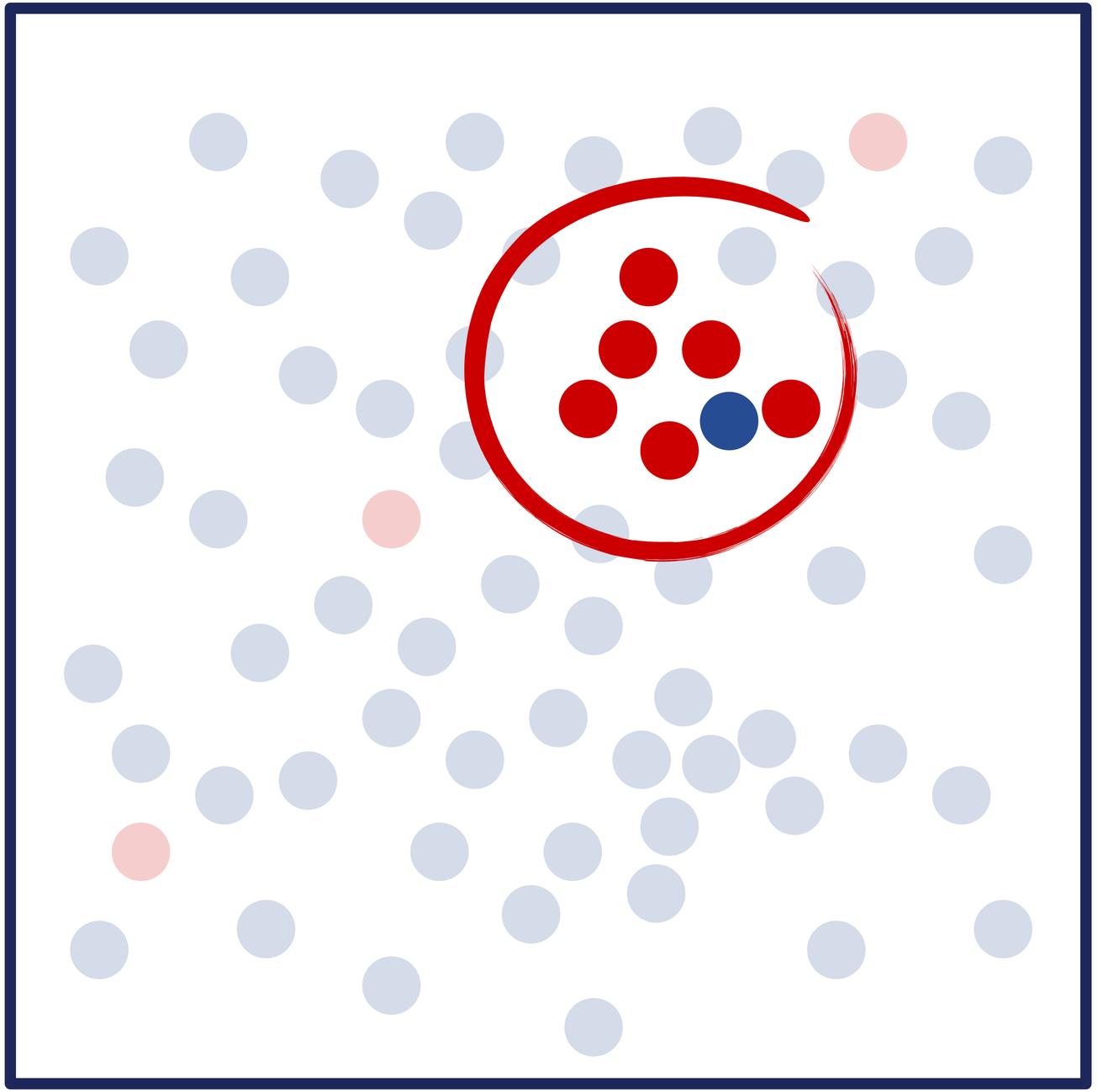


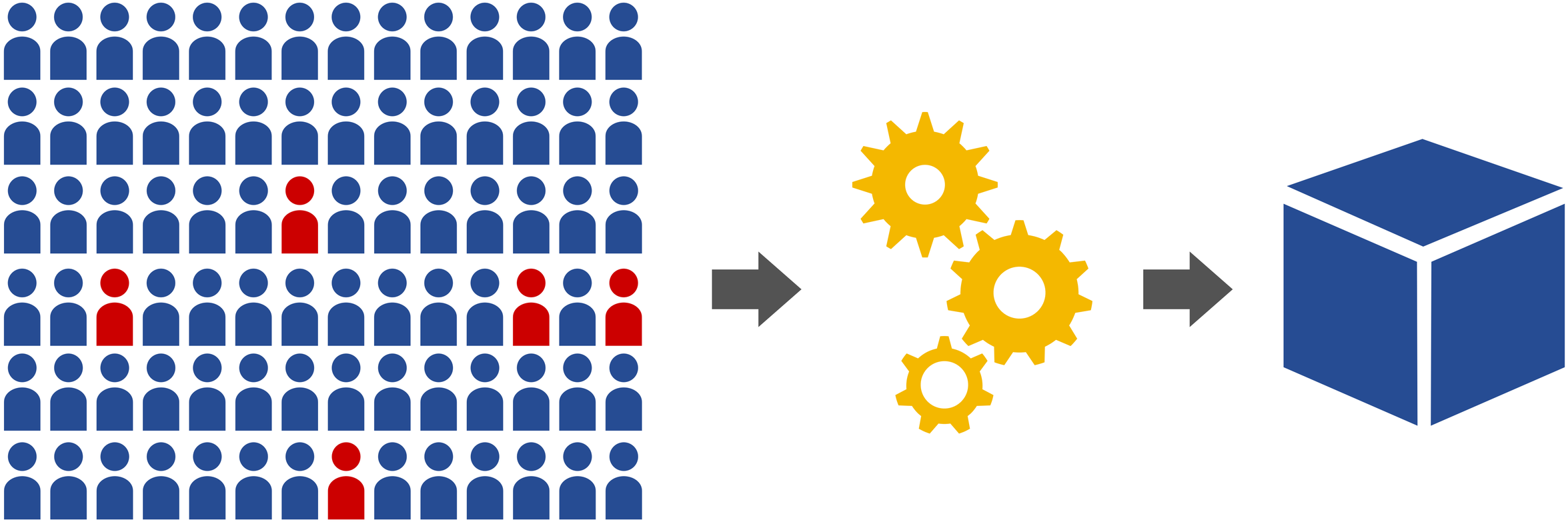


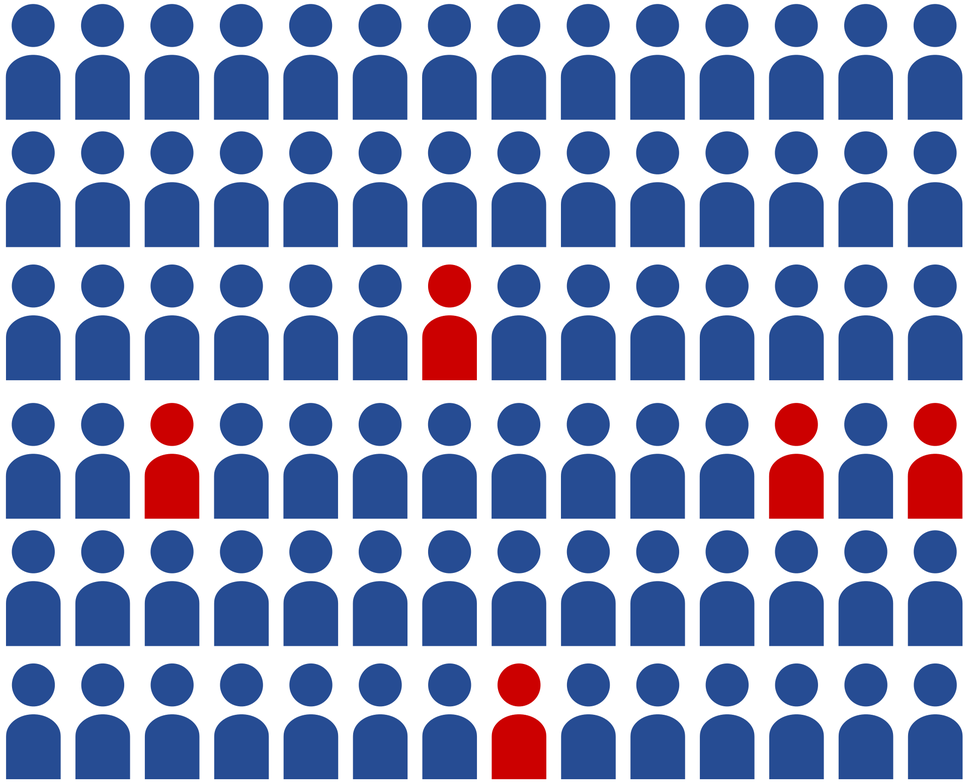
| 0.67 | 0.57 | 0.84 | • • • | 0.08 | 0.42 | 0.01 |
|---|---|---|---|---|---|---|

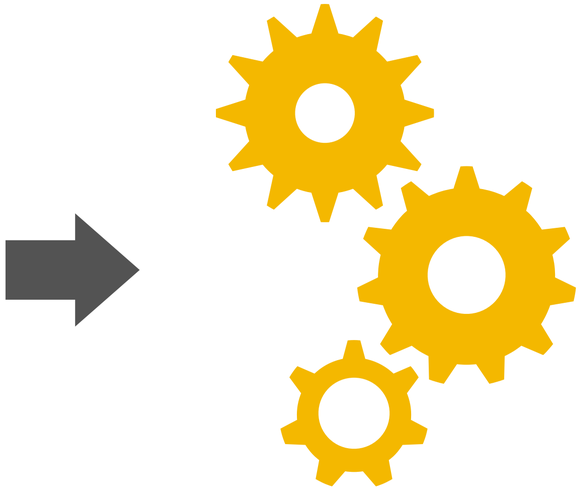




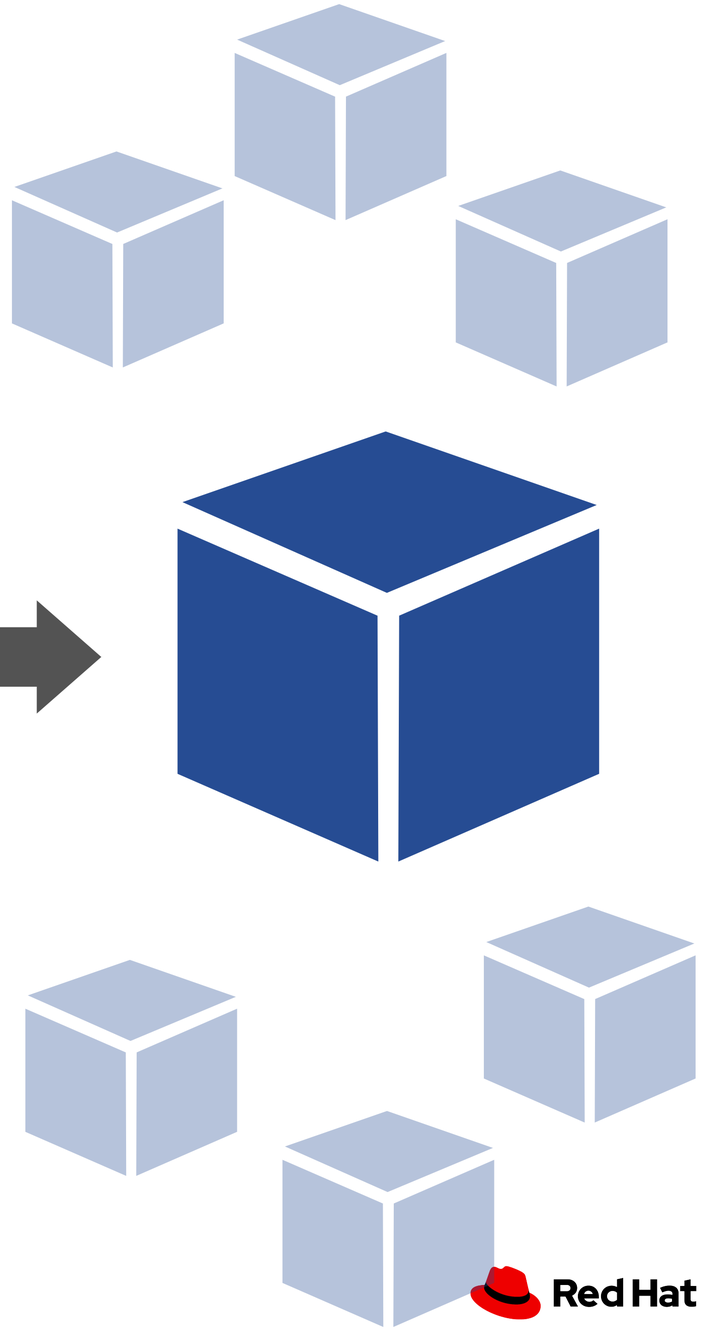






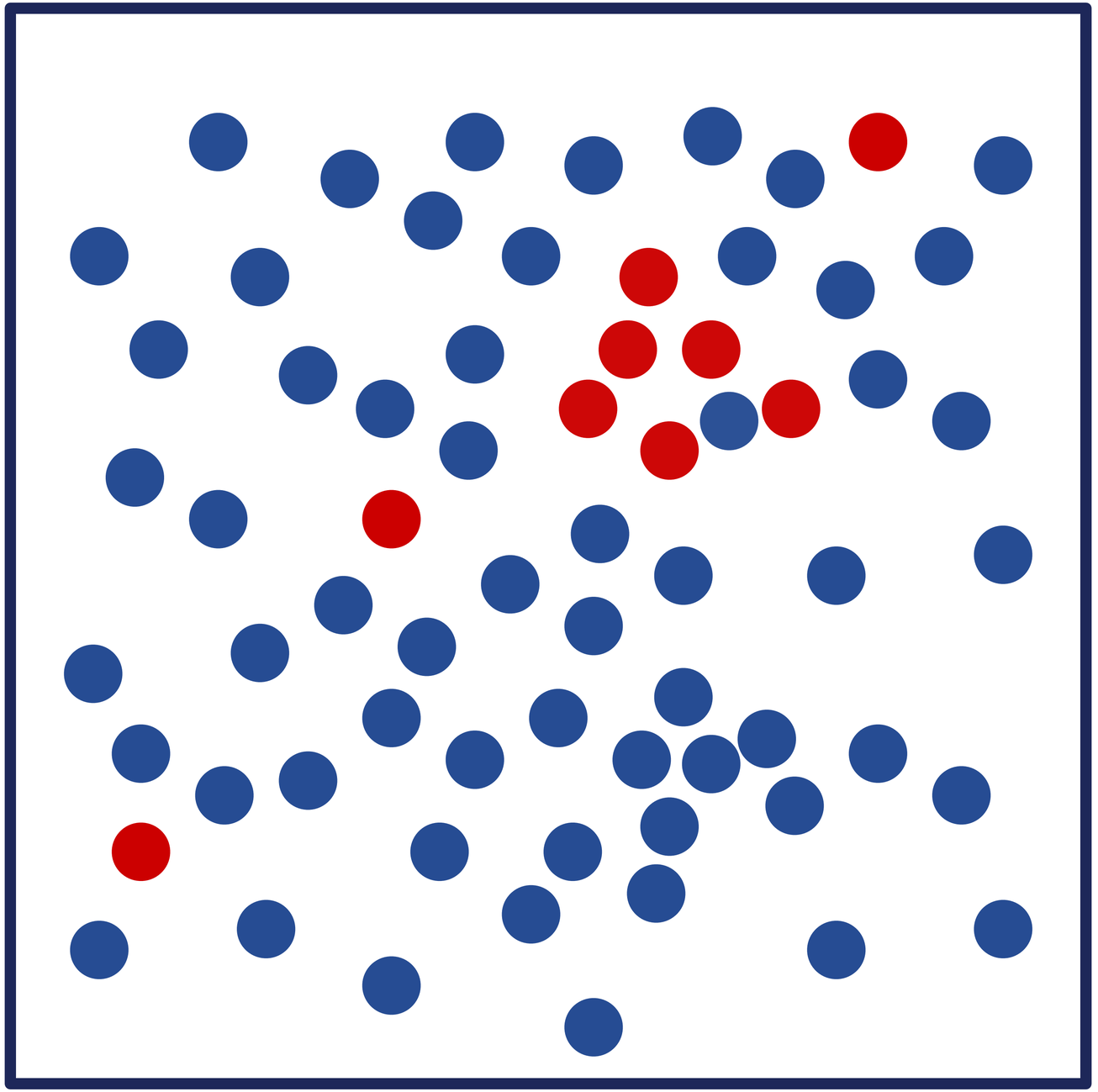


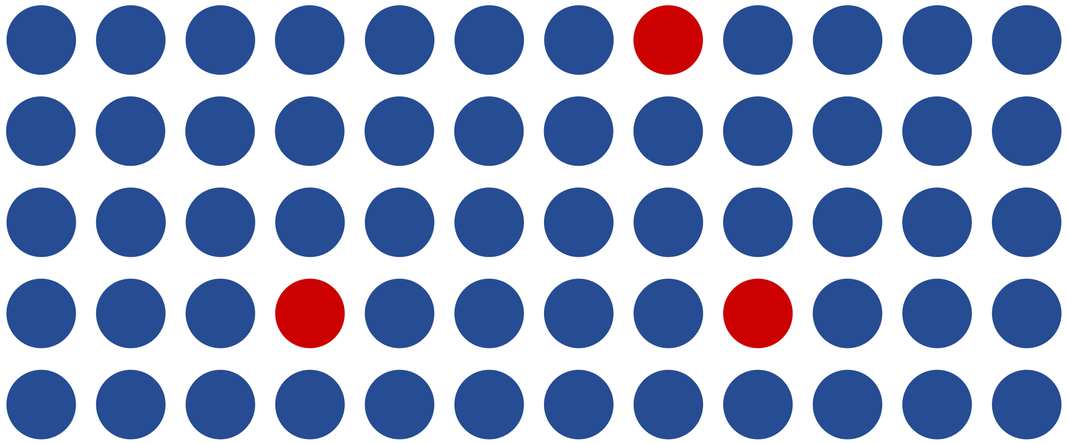




@willb

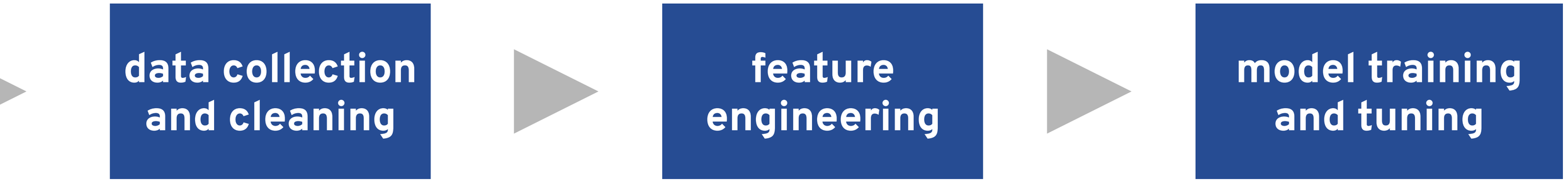
Red Hat

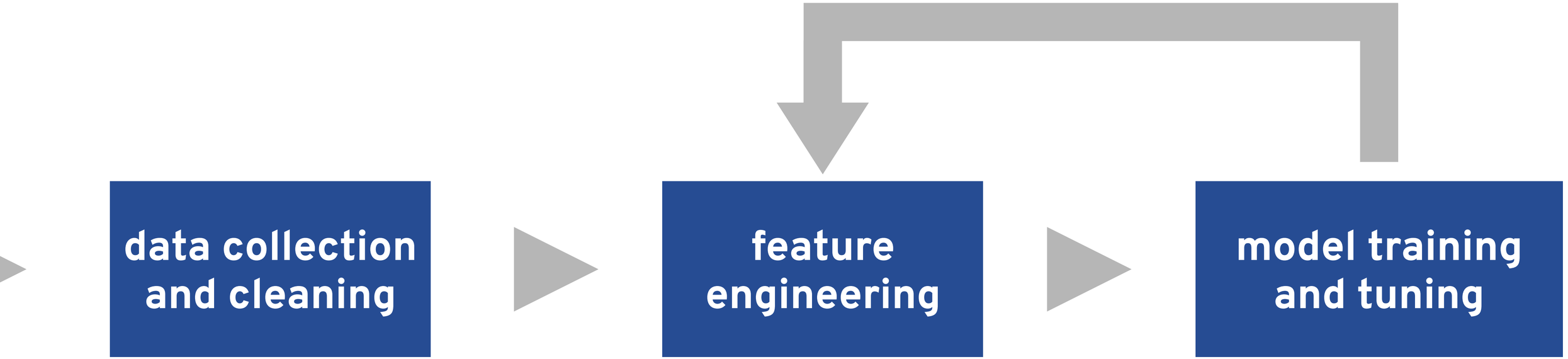


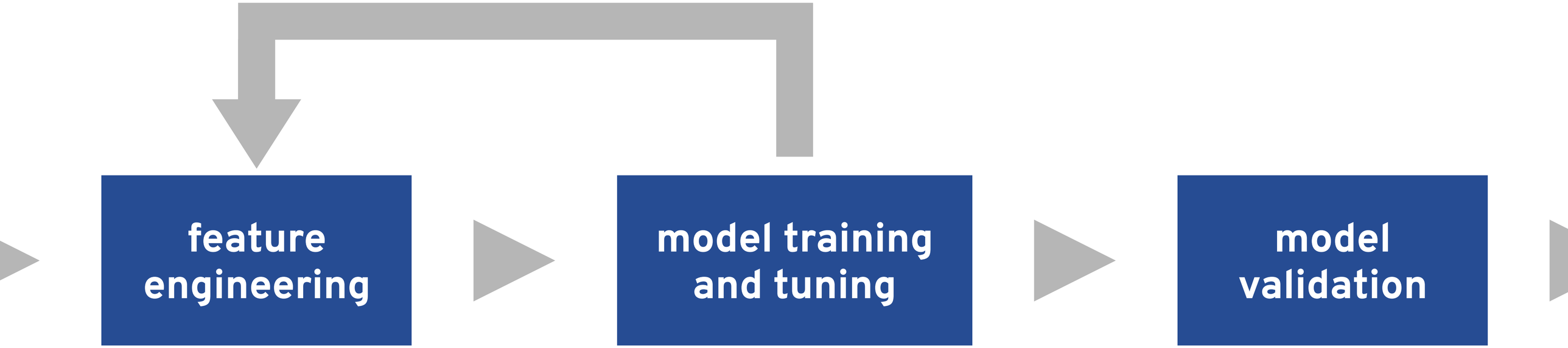


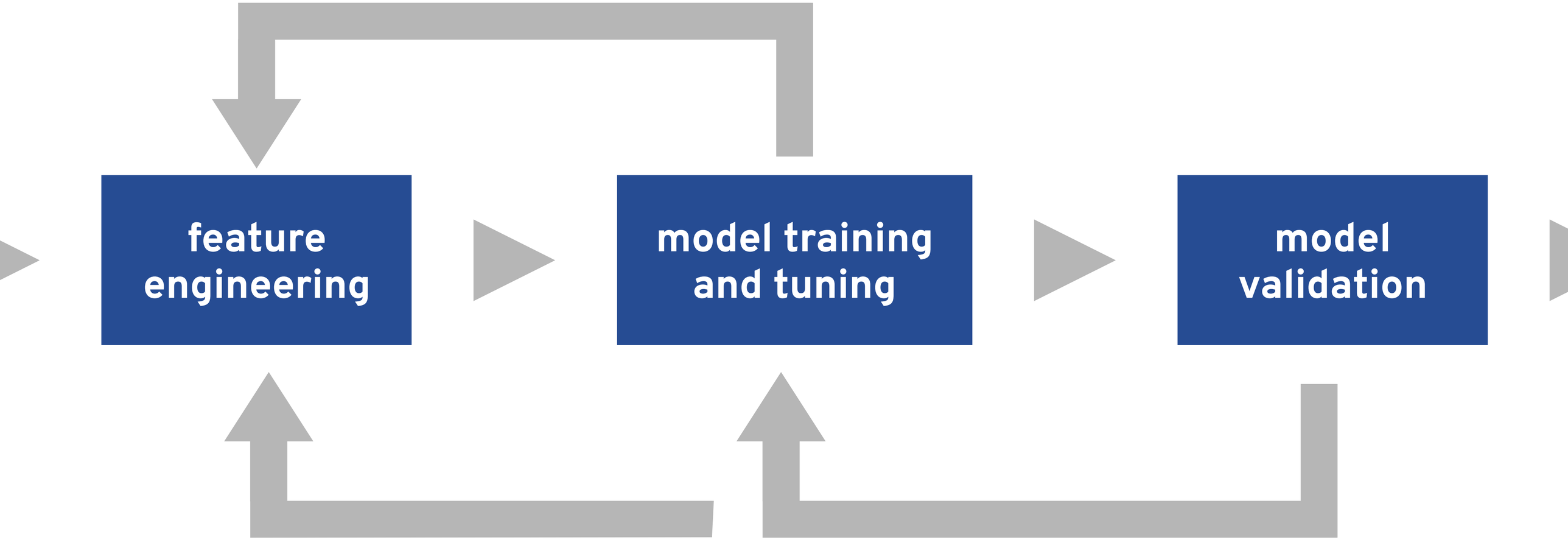
codifying problem and metrics

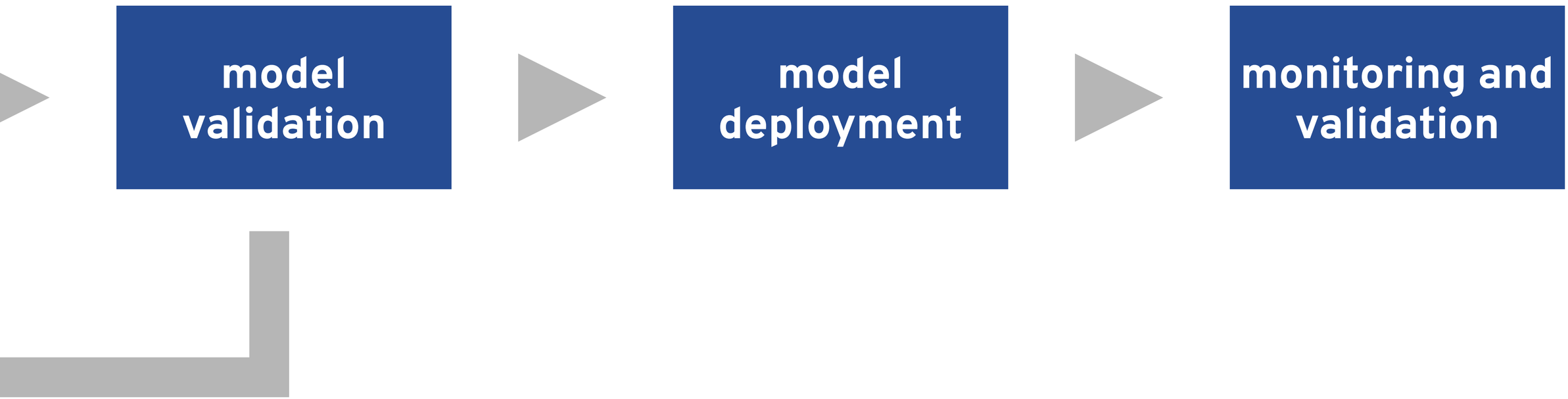
data collection and cleaning

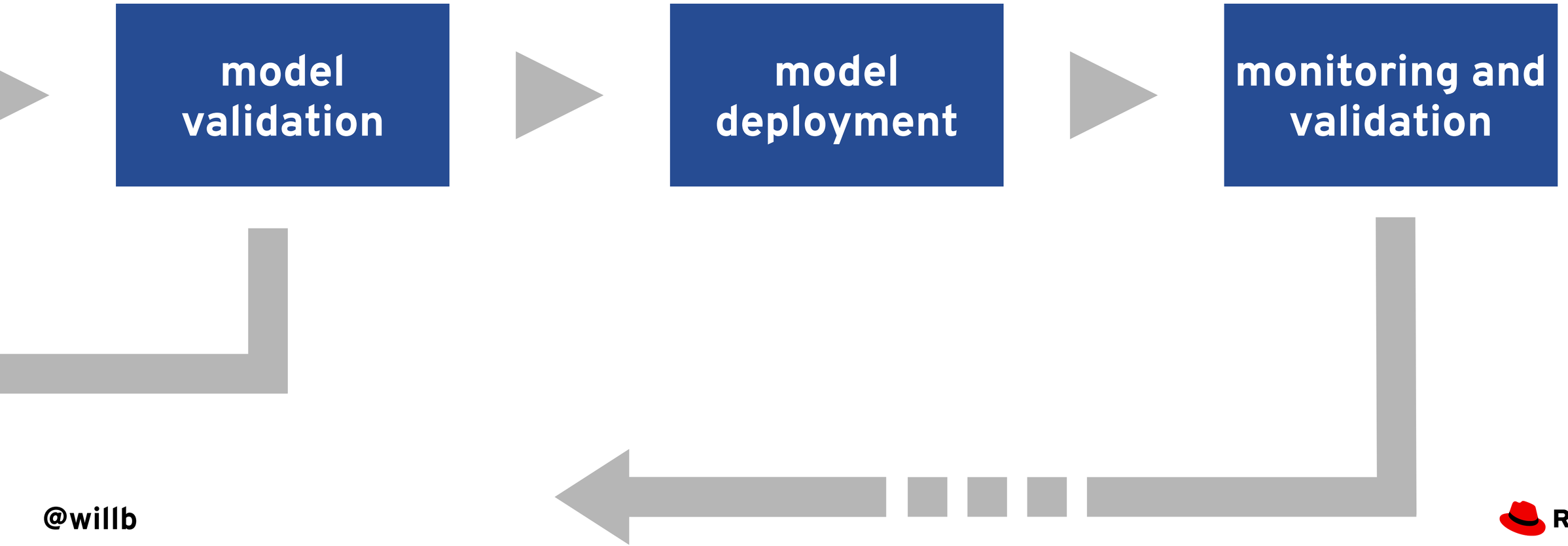

data collection and cleaning
feature engineering
model training and tuning

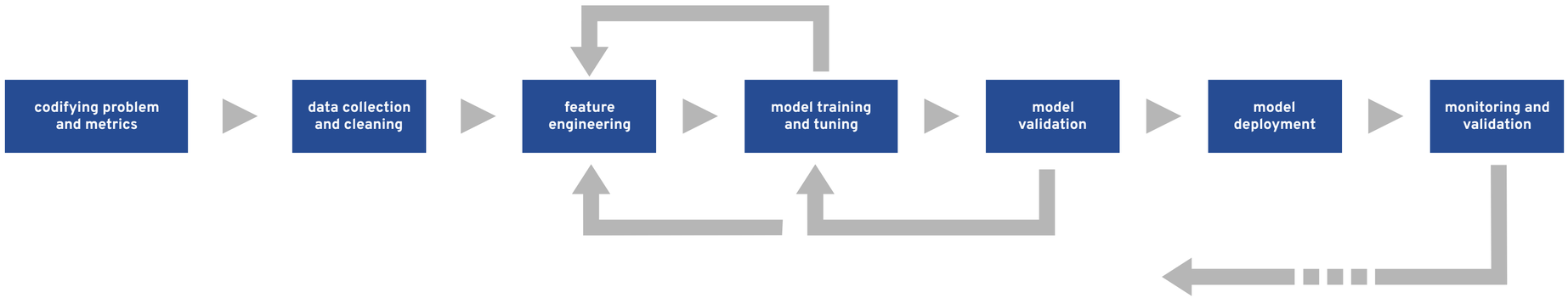

data collection and cleaning
feature engineering
model training and tuning

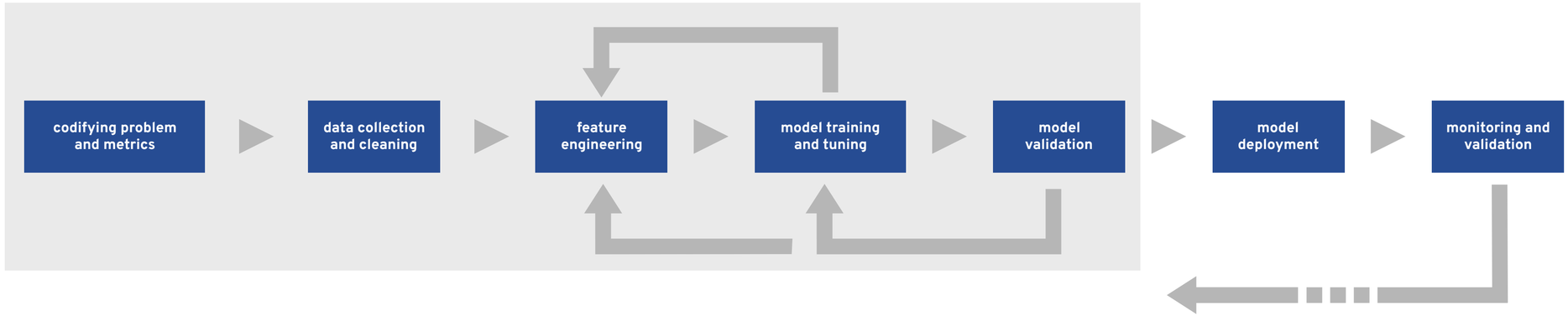

feature engineering
model training and tuning
model validation


feature engineering
model training and tuning
model validation


model validation
model deployment
monitoring and validation

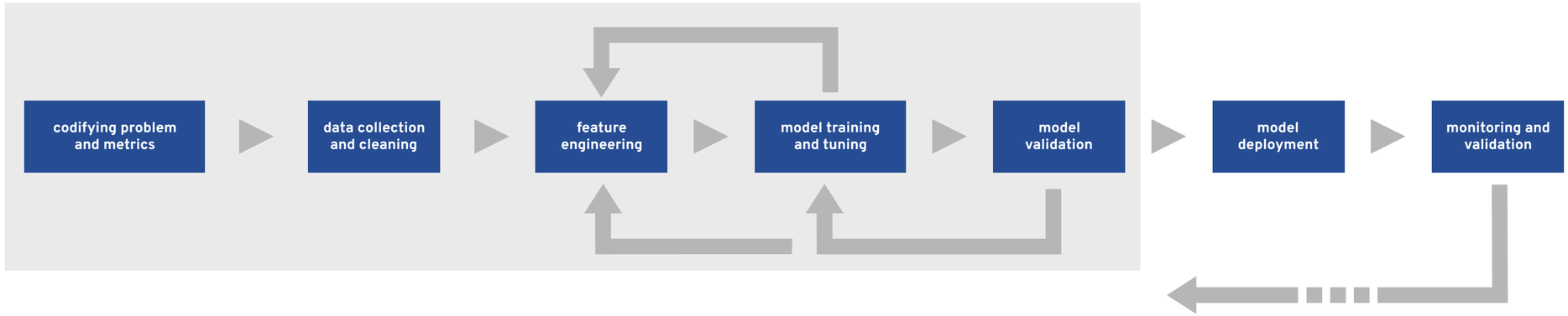

model validation
model deployment
monitoring and validation


codifying problem and metrics
data collection and cleaning
feature engineering
model training and tuning
model validation
model deployment
monitoring and validation



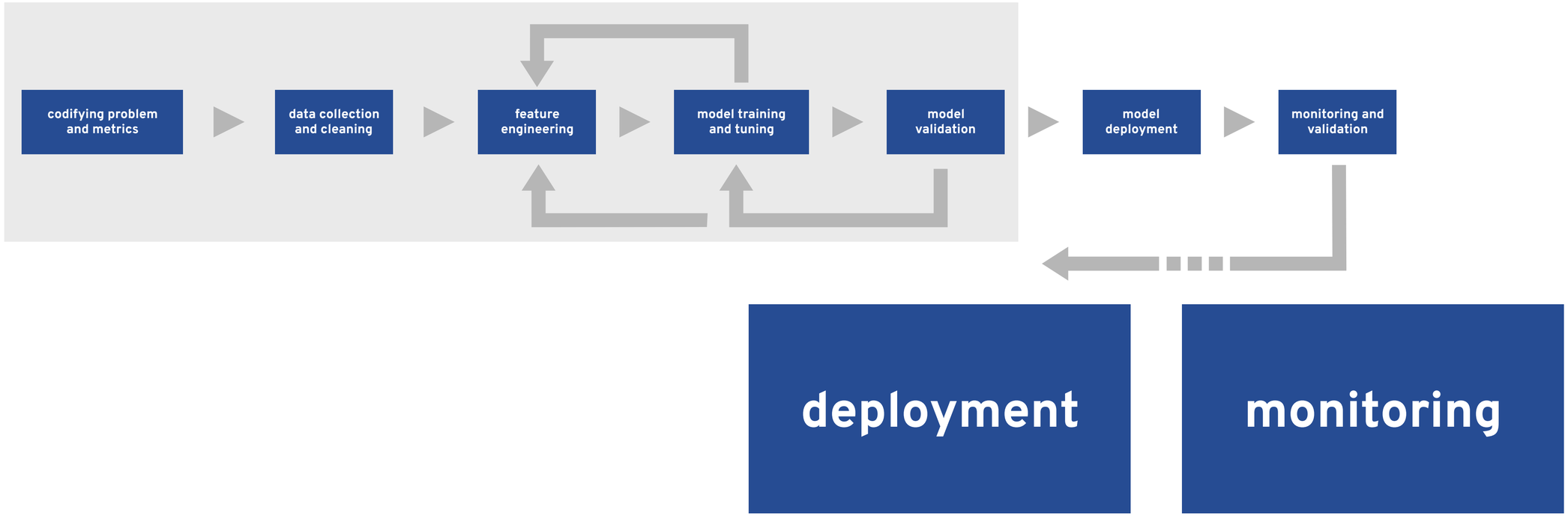

**defining types and interfaces**

**prototyping**

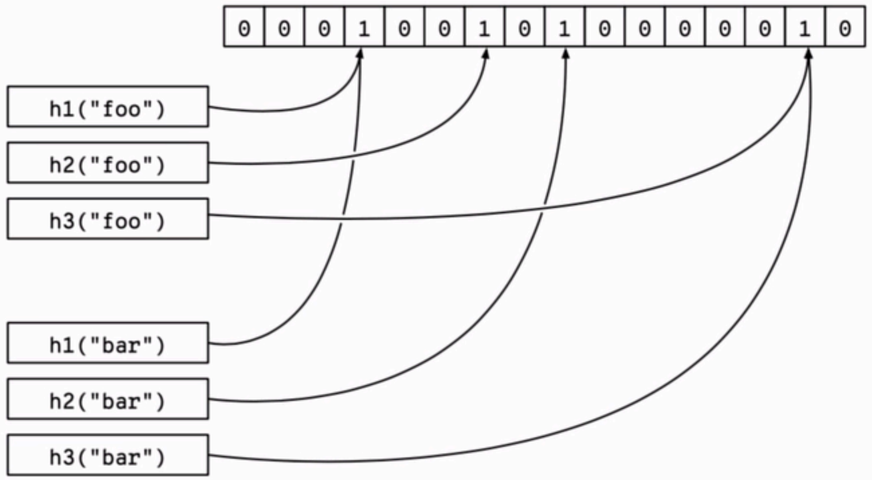



**unit, behavioral, and integration testing**

**formal verification**


codifying problem and metrics
data collection and cleaning
feature engineering
model training and tuning
model validation
model deployment
monitoring and validation
deployment
monitoring

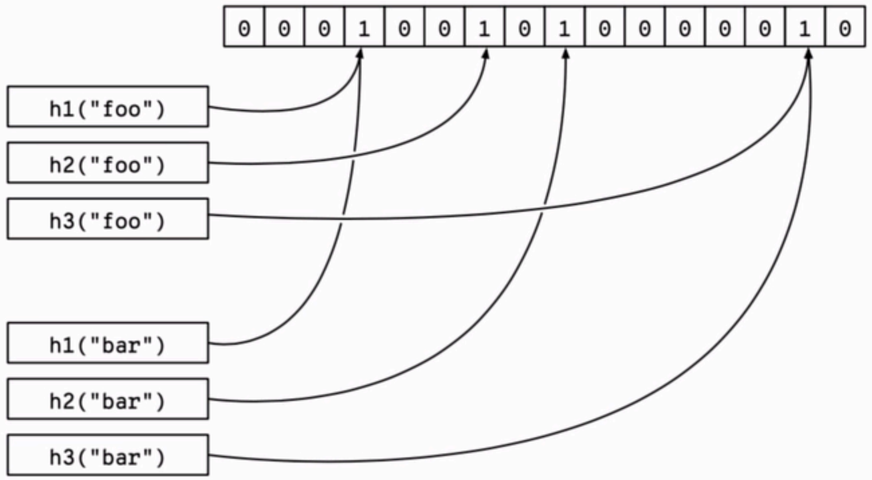
# Bloom filter ¶




A conventional hash table (or hash table-backed set structure) consists of a series of *buckets*. Hash table insert looks like this:

1. First, use the hash value of the key to identify the index of the bucket that should contain it.
2. If the bucket is empty, update the bucket to contain the key and value (with a trivial value in the case of a hashed set).
3. If the bucket is not empty and the key stored in it is not the one you've hashed, handle this *hash collision*. There are several strategies to handle hash collisions precisely; most involve extra lookups (e.g., having a second hash function or going to the next available bucket) or

# Bloom filter ¶

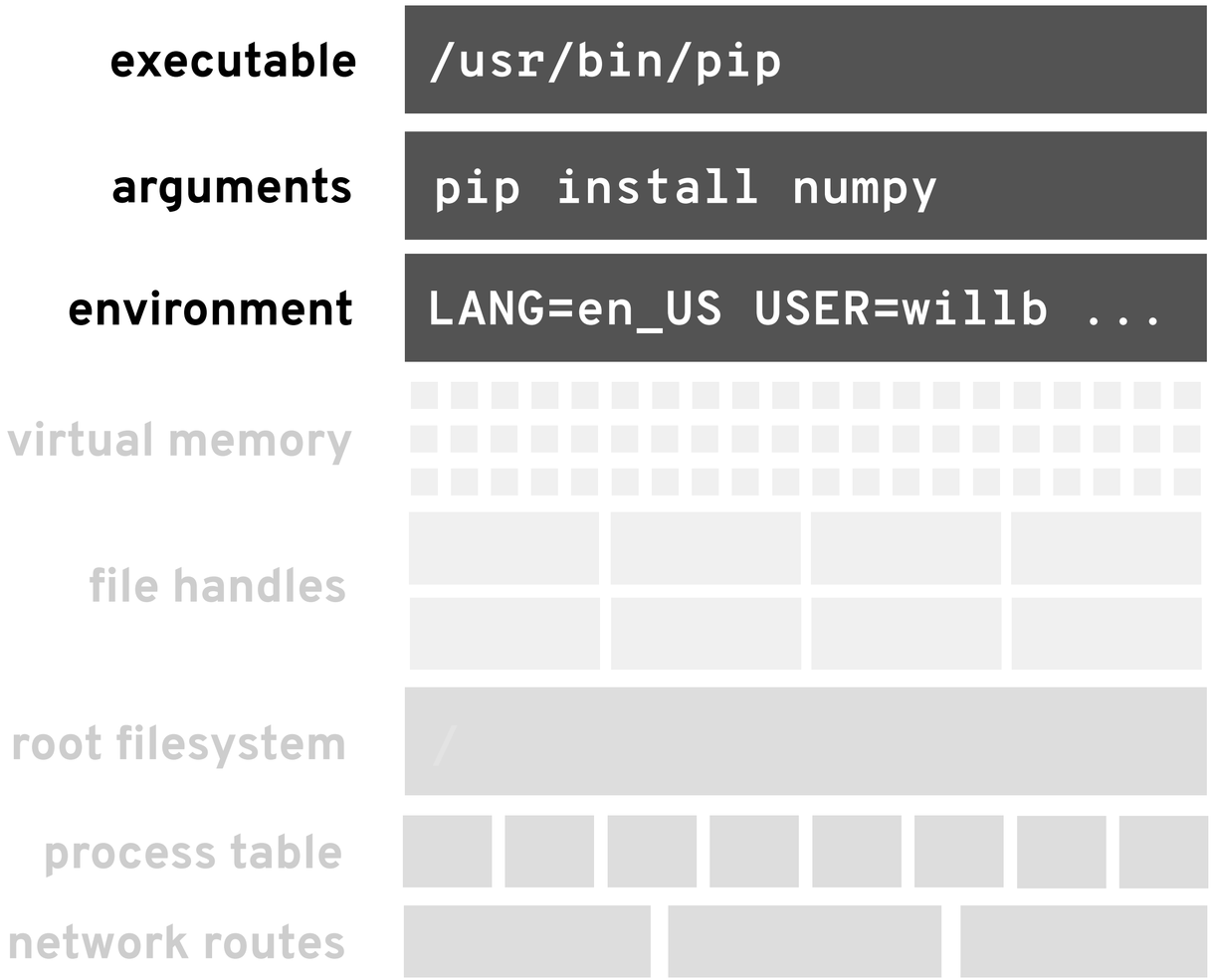


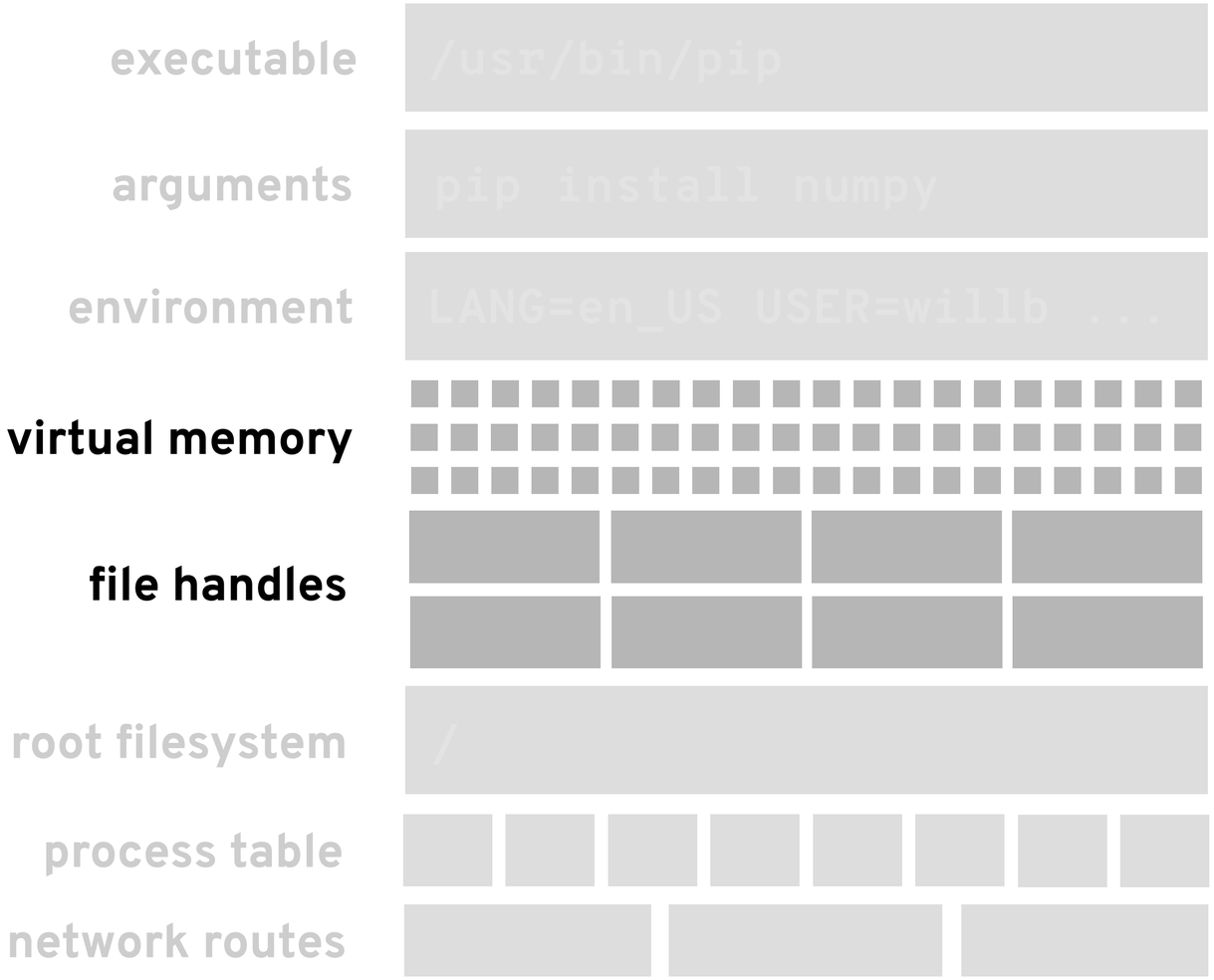

A conventional hash table (or hash table-backed set structure) consists of a series of *buckets*. Hash table insert looks like this:

1. First, use the hash value of the key to identify the index of the bucket that should contain it.
2. If the bucket is empty, update the bucket to contain the key and value (with a trivial value in the case of a hashed set).
3. If the bucket is not empty and the key stored in it is not the one you've hashed, handle this *hash collision*. There are several strategies to handle hash collisions precisely; most involve extra lookups (e.g., having a second hash function or going to the next available bucket) or

# Why is OpenShift great for machine learning workflows?

```
%pip install numpy
```


executable
/usr/bin/pip
arguments
pip install numpy
environment
LANG=en_US USER=willb ...
virtual memory
file handles
root filesystem
/
process table
network routes


executable
/usr/bin/pip
arguments
pip install numpy
environment
LANG=en_US USER=willb ...
virtual memory
file handles
root filesystem
/
process table
network routes

executable /usr/bin/pip

arguments pip install numpy


root filesystem /

process table

network routes

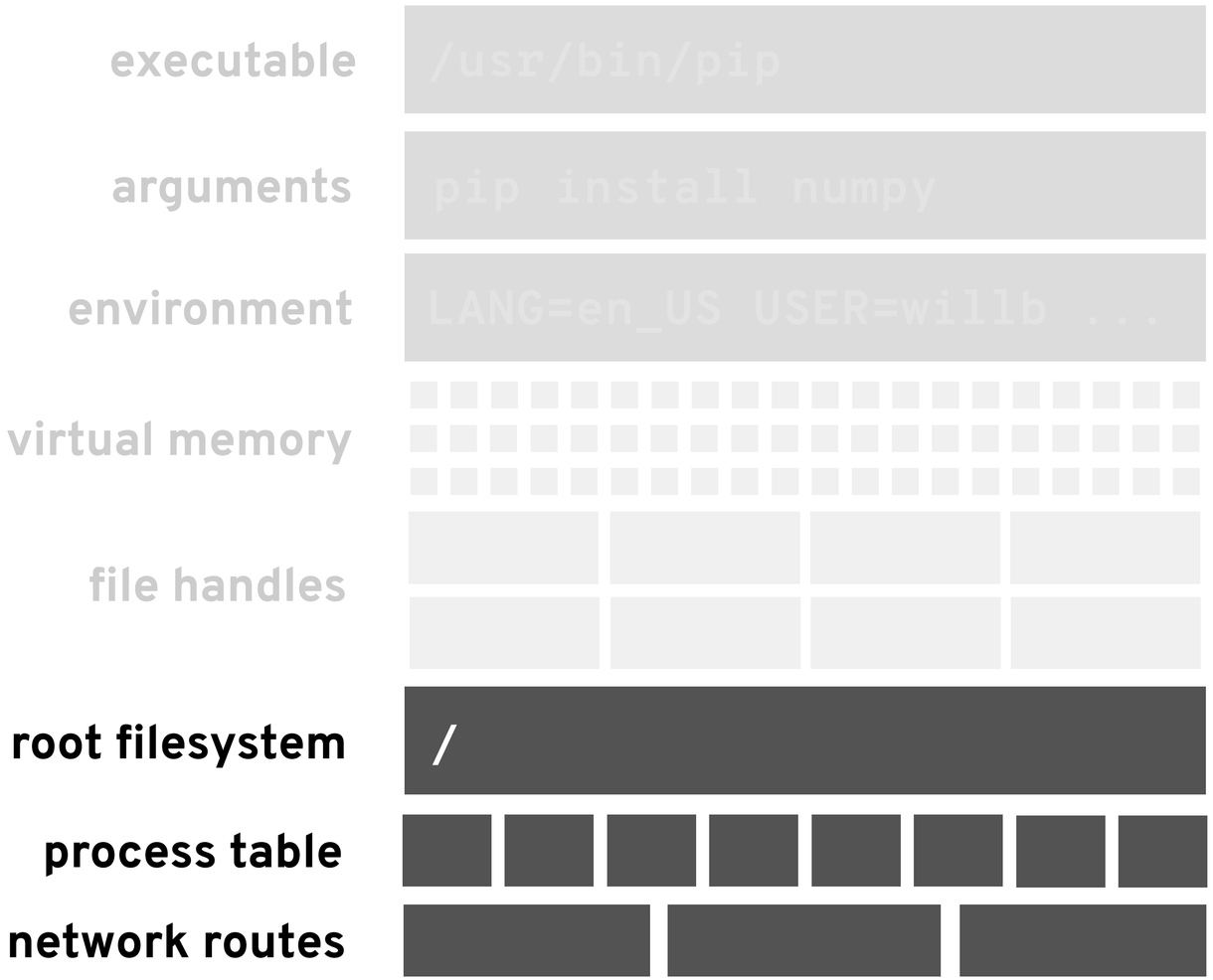
executable /usr/bin/pip

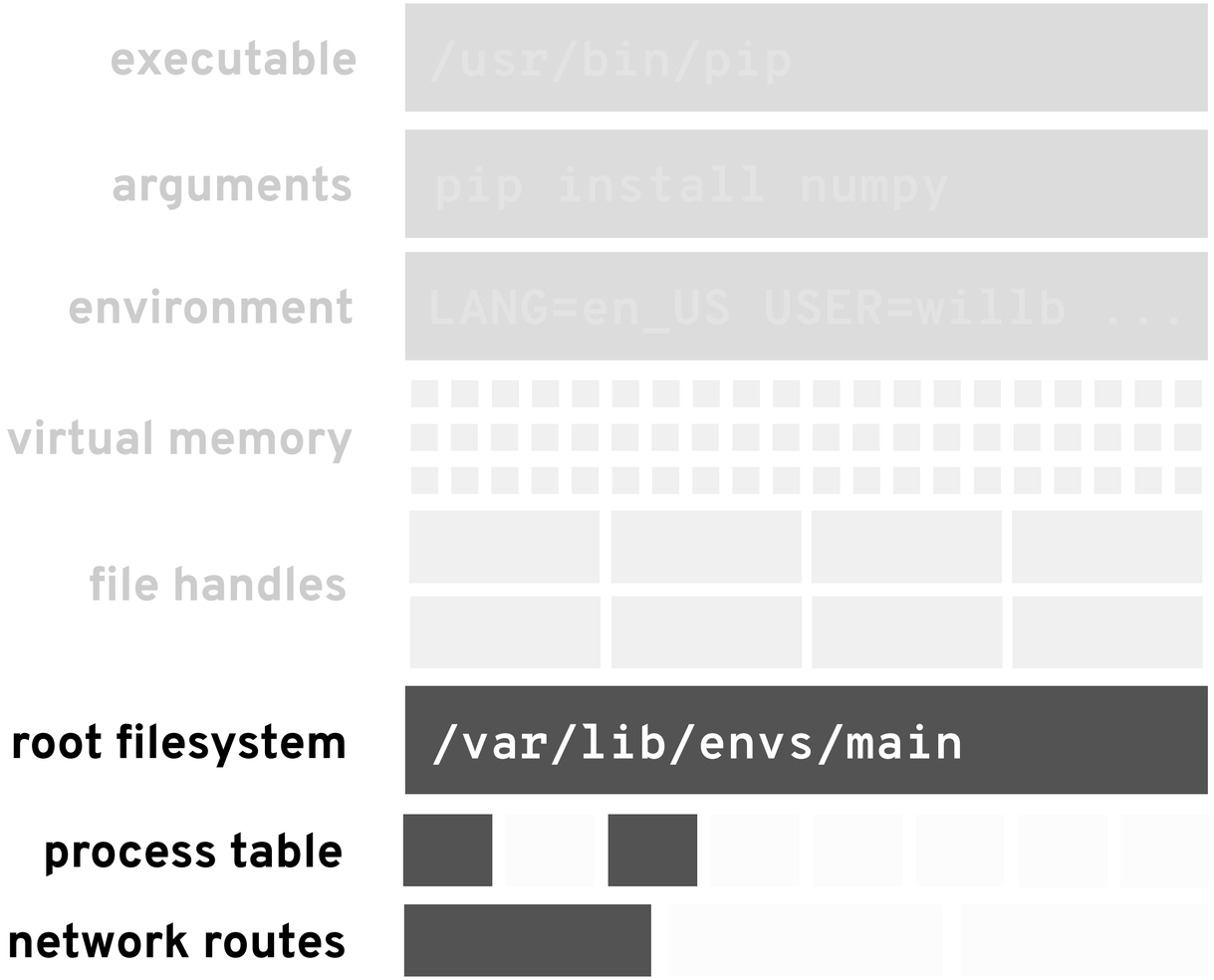
arguments pip install numpy

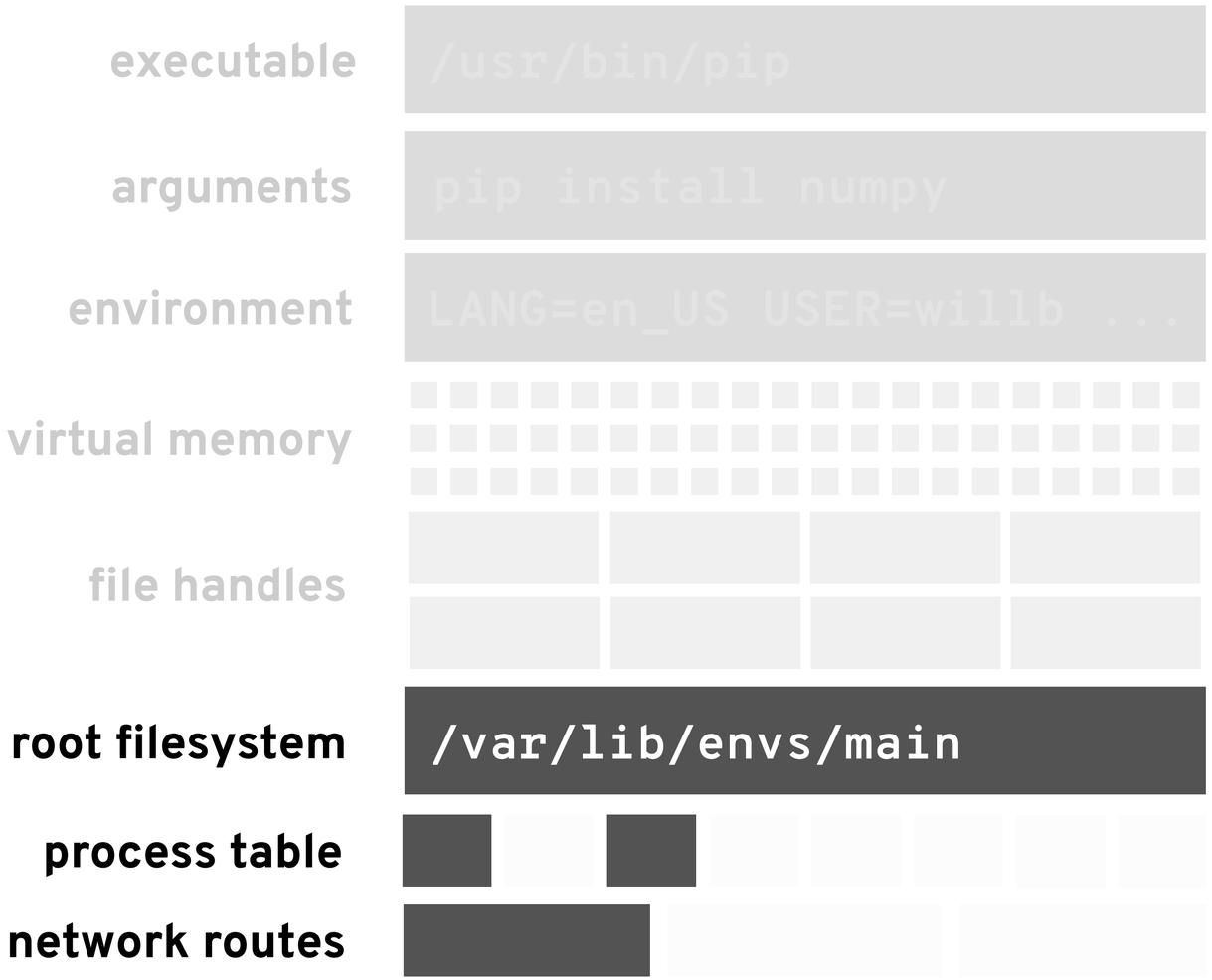

root filesystem /

process table

network routes


executable
/usr/bin/pip
arguments
pip install numpy
environment
LANG=en_US USER=willb ...
virtual memory
file handles
root filesystem
/
process table
network routes


executable
/usr/bin/pip
arguments
pip install numpy
environment
LANG=en_US USER=willb ...
virtual memory
file handles
root filesystem
/var/lib/envs/main
process table
network routes

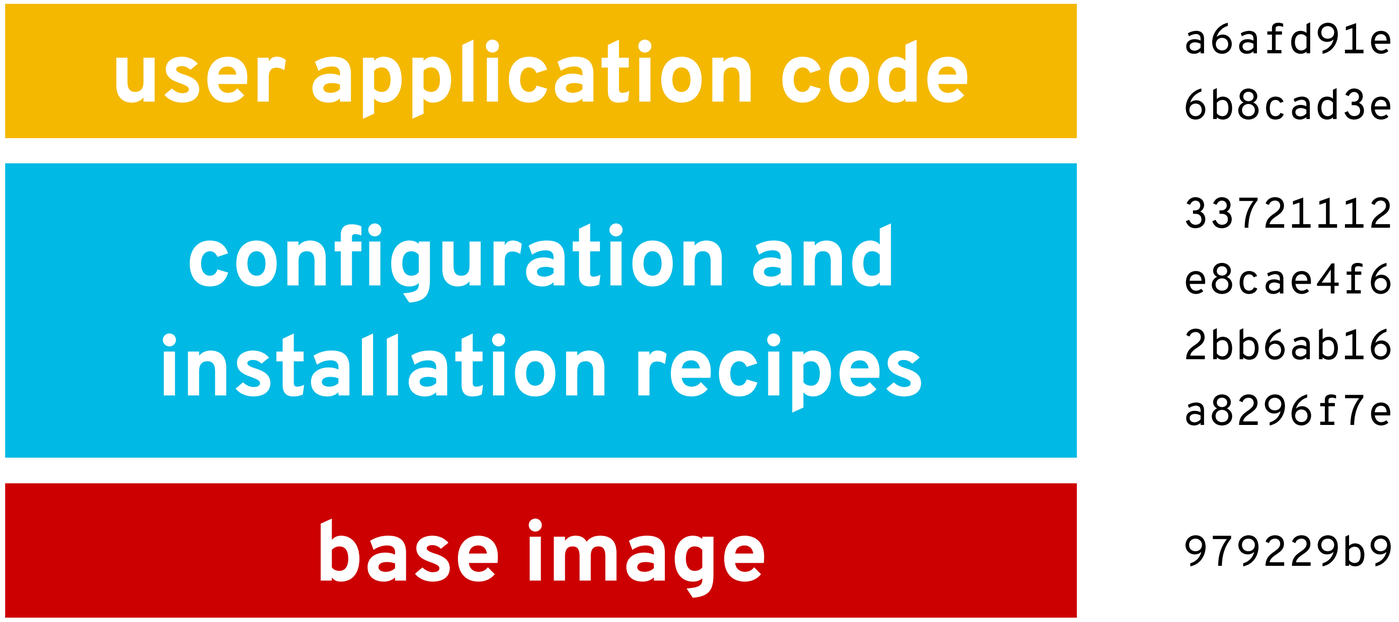

executable
/usr/bin/pip
arguments
pip install numpy
environment
LANG=en_US USER=willb ...
virtual memory
file handles
root filesystem
/var/lib/envs/main
process table
network routes


SPEED
LIMIT
55

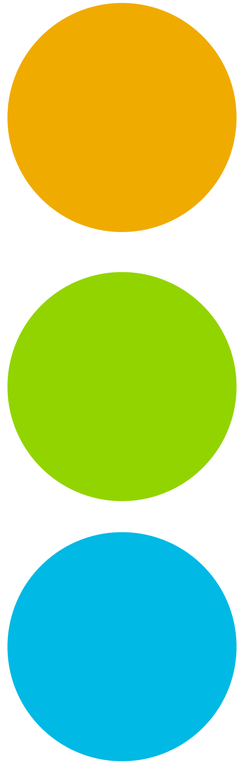
efficient isolation,
secure by default

**OpenShift** is ~~enterprise~~ Kubernetes with a **great developer experience.**

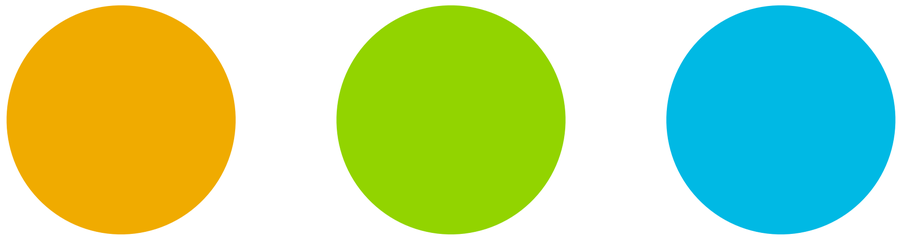
# Immutable images



# Stateless microservices

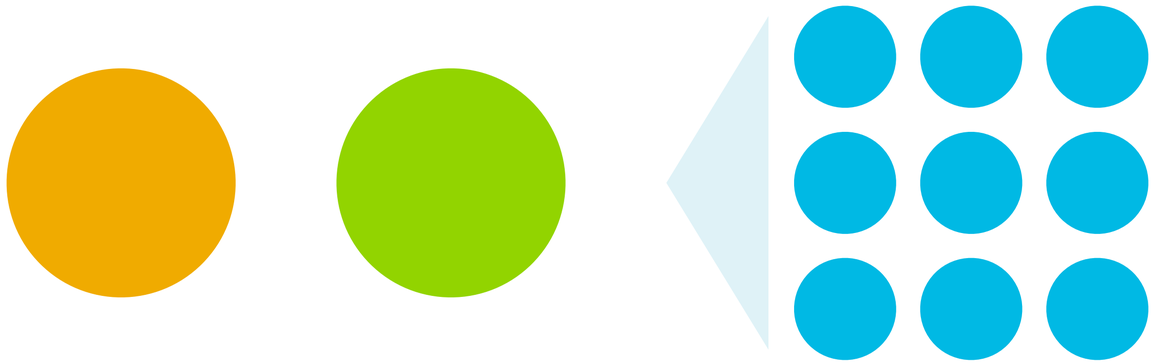


# Stateless microservices

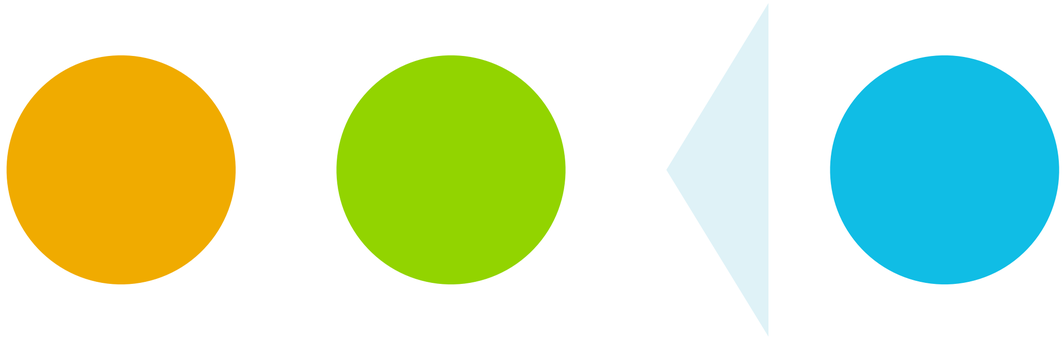


# Stateless microservices

# Stateless microservices



# Stateless microservices

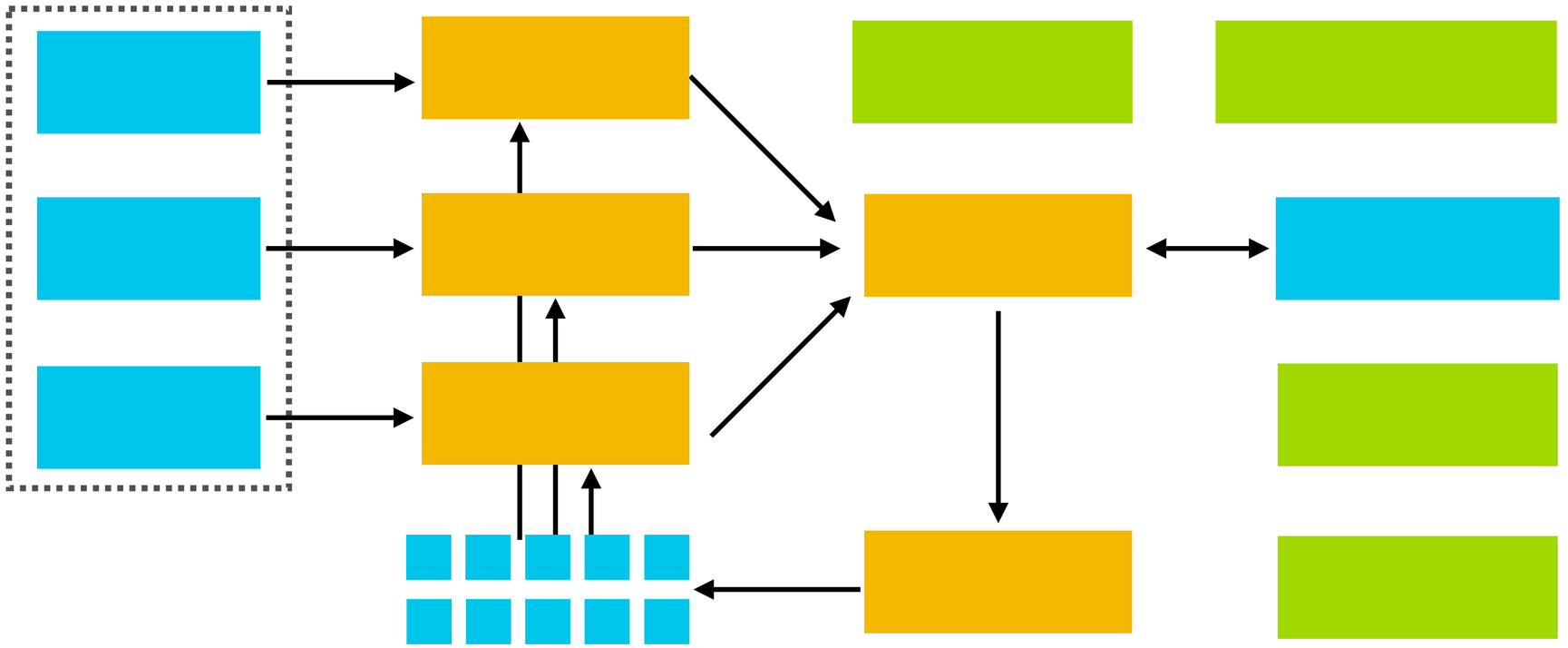


# Stateless microservices

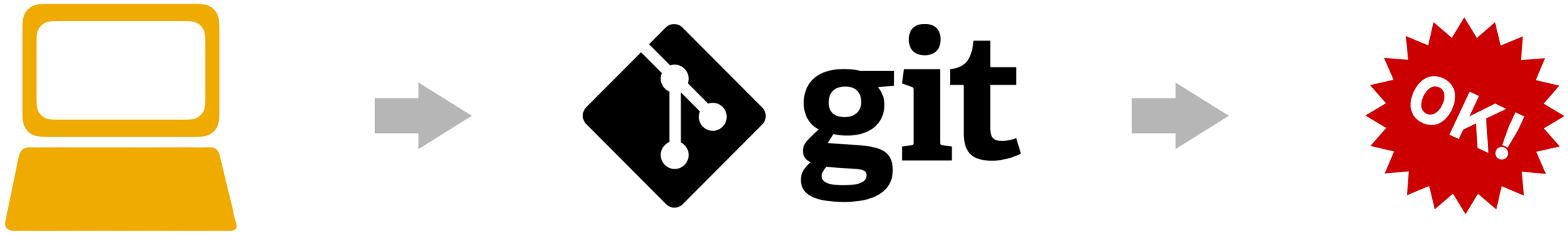
# Stateless microservices

# Stateless microservices

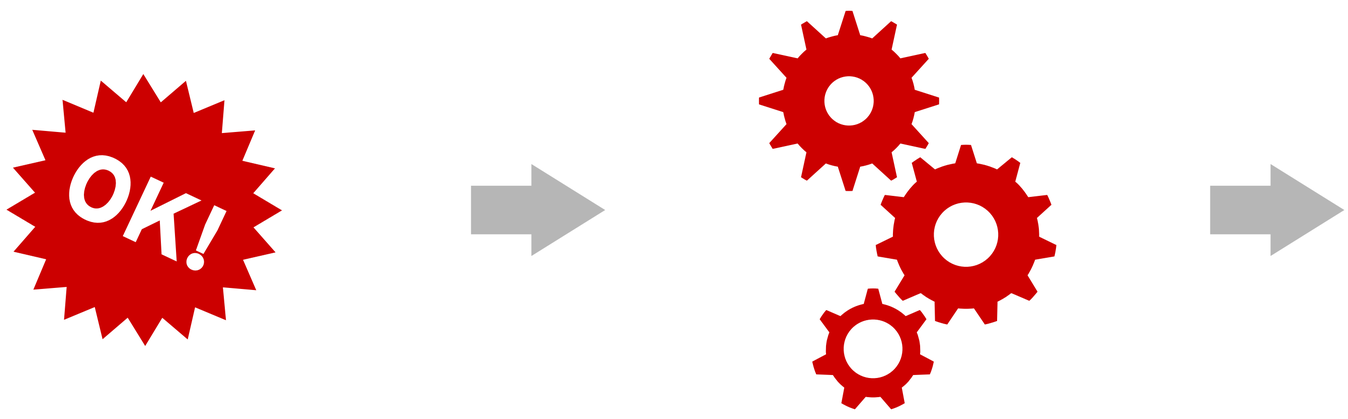
# Declarative app configuration



# Integration and deployment



# Integration and deployment

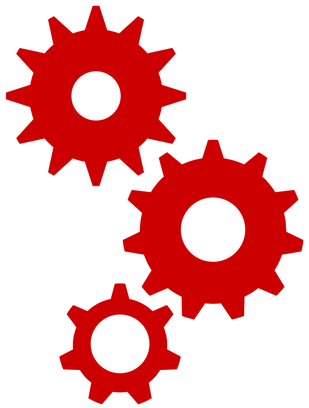



application code

configuration and installation recipes

base image

# Integration and deployment

application code

configuration and installation recipes

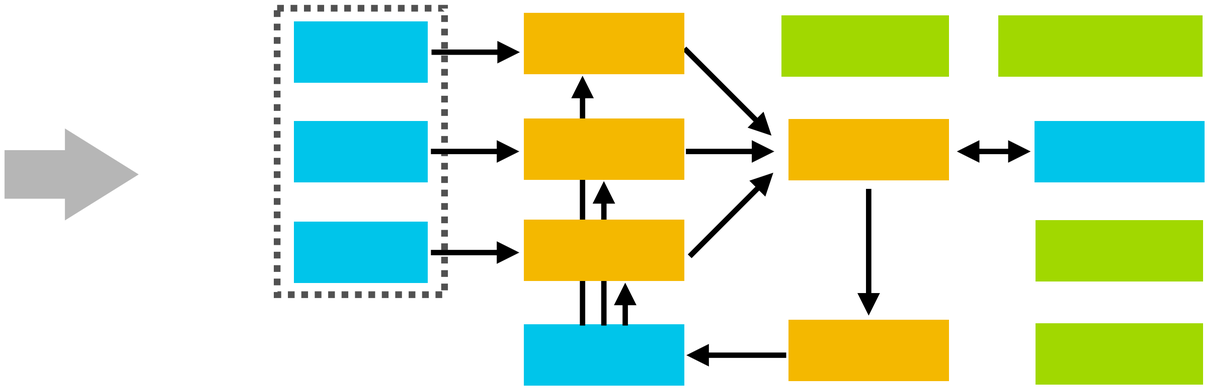
base image

# Integration and deployment

application code

configuration and installation recipes

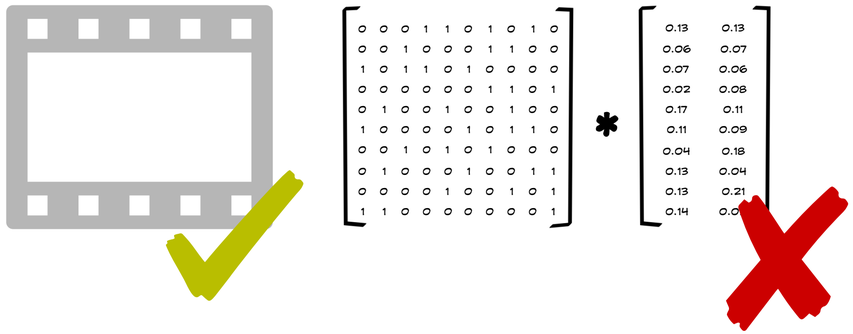
base image



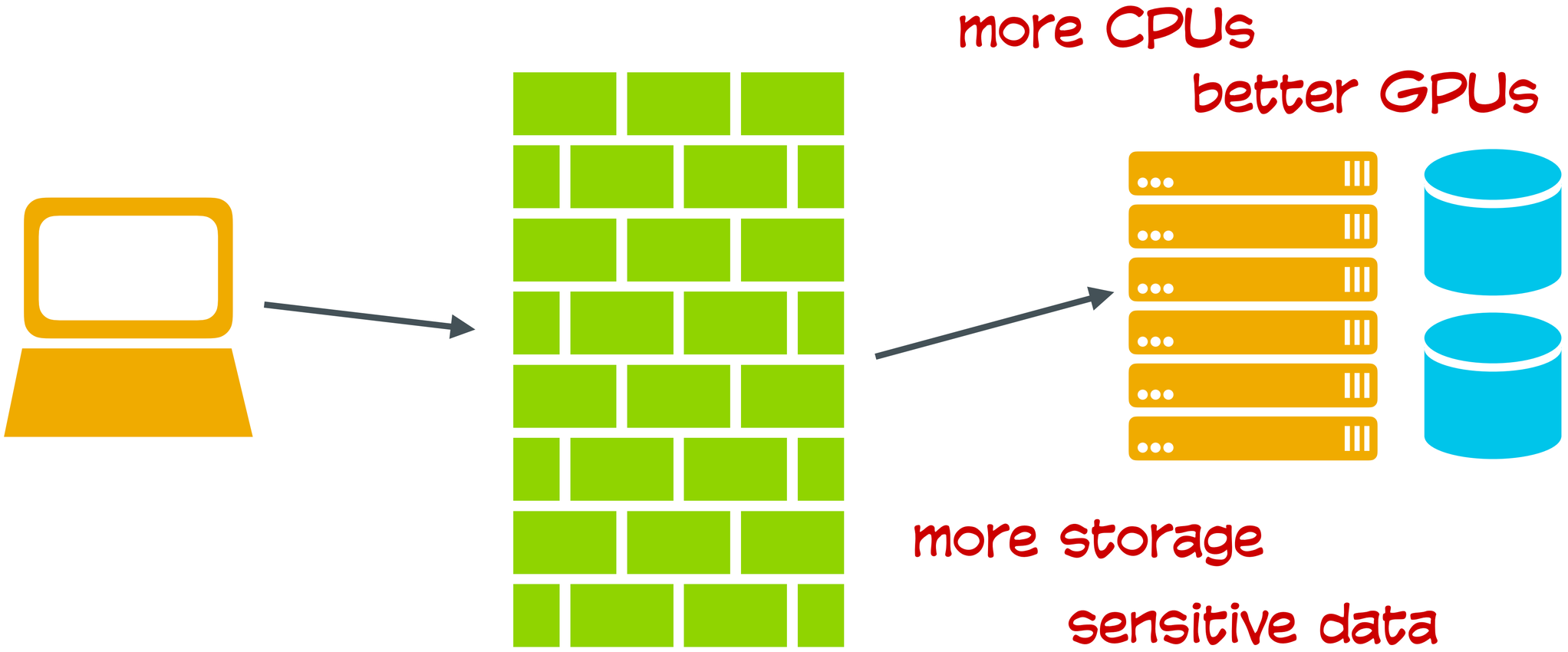
# What OpenShift offers working data scientists

**OpenShift** is **enterprise Kubernetes** with a **great developer experience.**

workflows to accelerate discovery

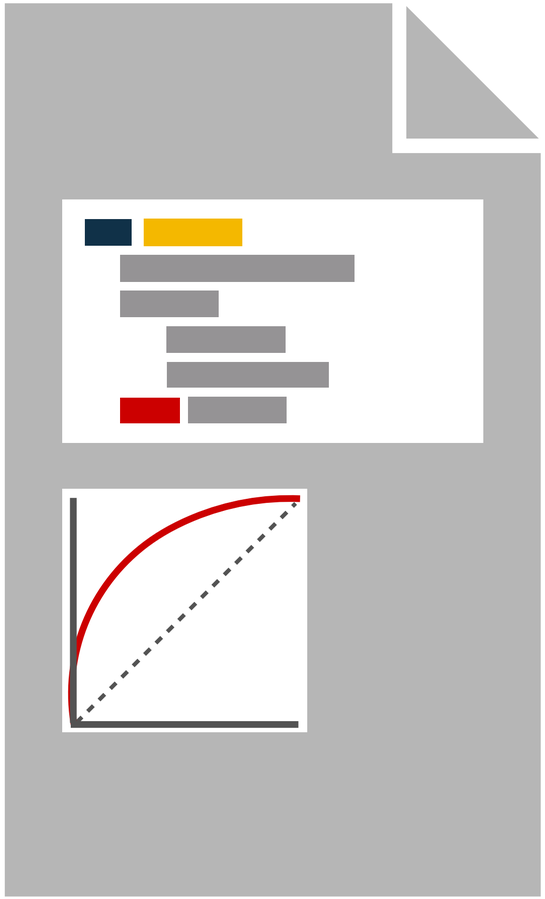

0 0 0 1 1 0 1 0 1 0
0 0 1 0 0 0 1 1 0 0
1 0 1 1 0 1 0 0 0 0
0 0 0 0 0 0 1 1 0 1
0 1 0 0 1 0 0 1 0 0
1 0 0 0 0 1 0 1 1 0
0 0 1 0 1 0 1 0 0 0
0 1 0 0 0 1 0 0 1 1
0 0 0 0 1 0 0 1 0 1
1 1 0 0 0 0 0 0 0 1
*
0.13 0.13
0.06 0.07
0.07 0.06
0.02 0.08
0.17 0.11
0.11 0.09
0.04 0.18
0.13 0.04
0.13 0.21
0.14 0.0

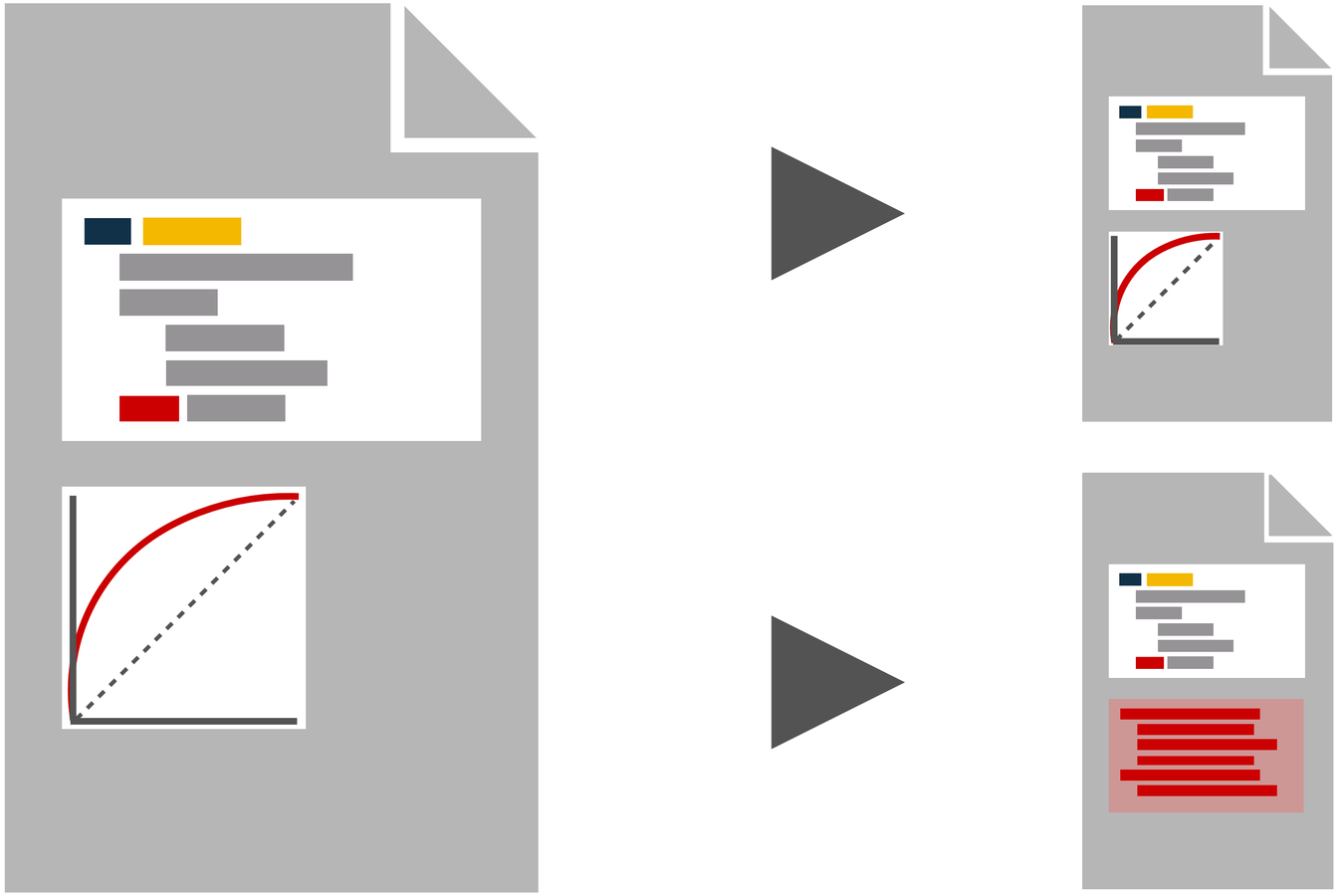
# Self-service environments

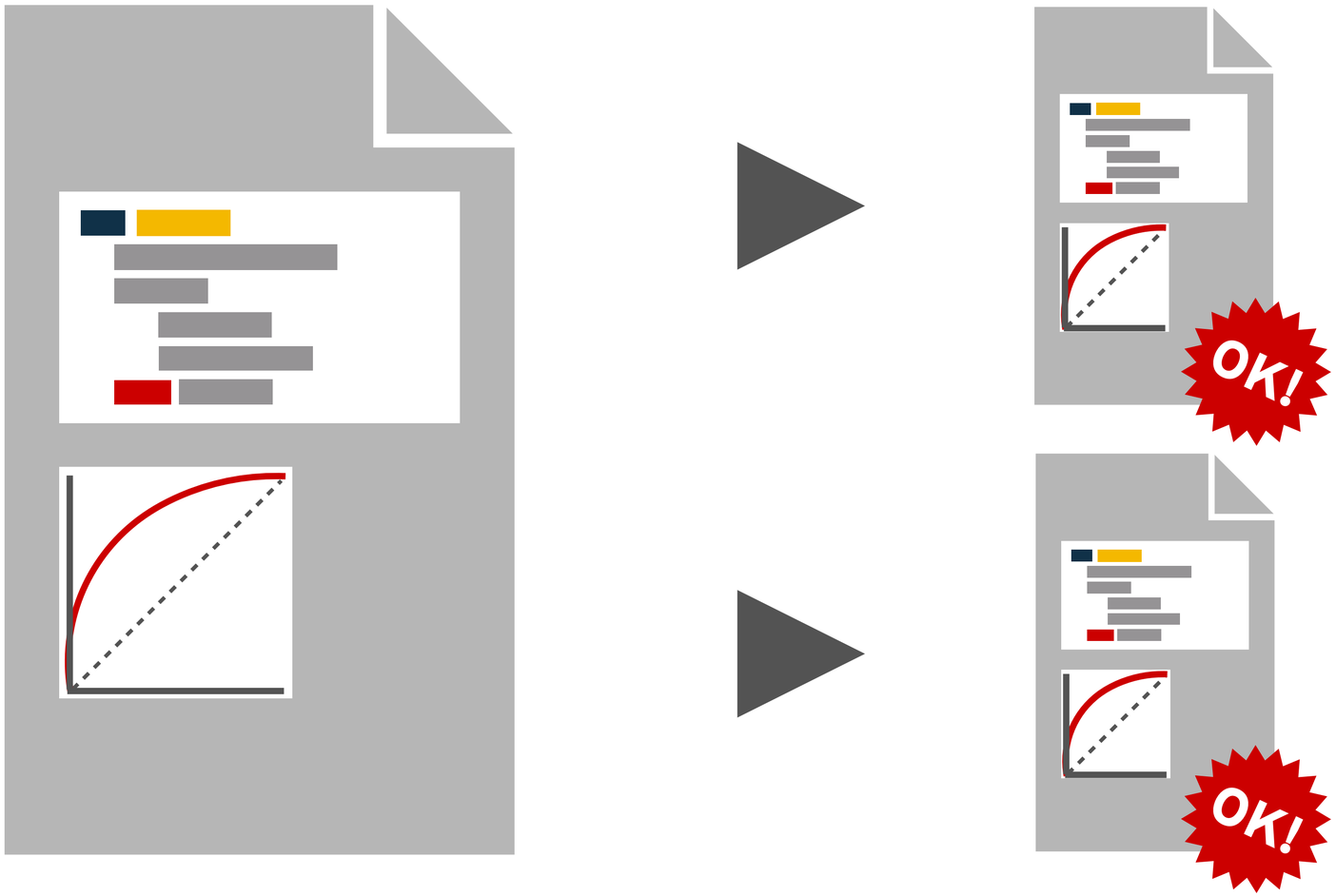


```
willb@echo % oc project workspace && oc create -f https://radanalytics.io/resources.yaml
```

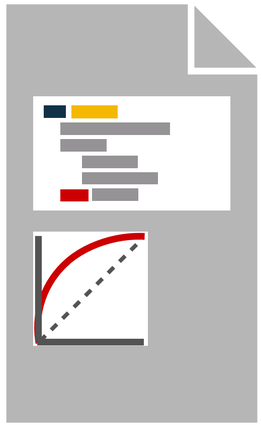
```
willb@echo % oc project workspace && oc create -f https://radanalytics.io/resources.yaml
```

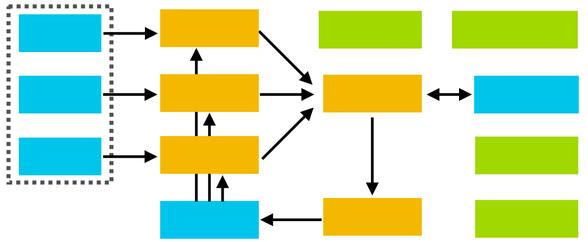




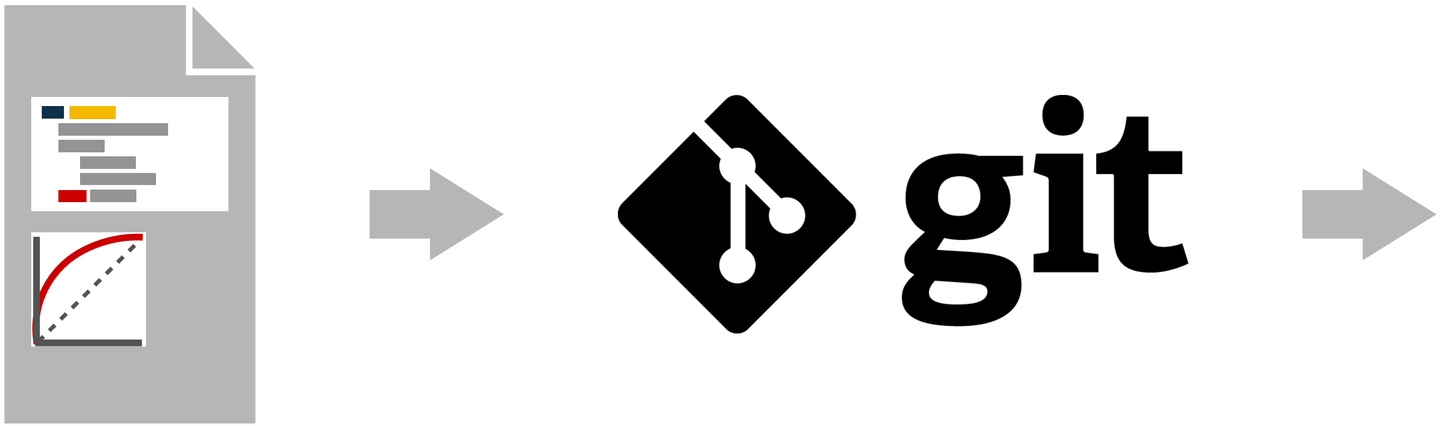

OK!
OK!

# No friction:  mybinder.org

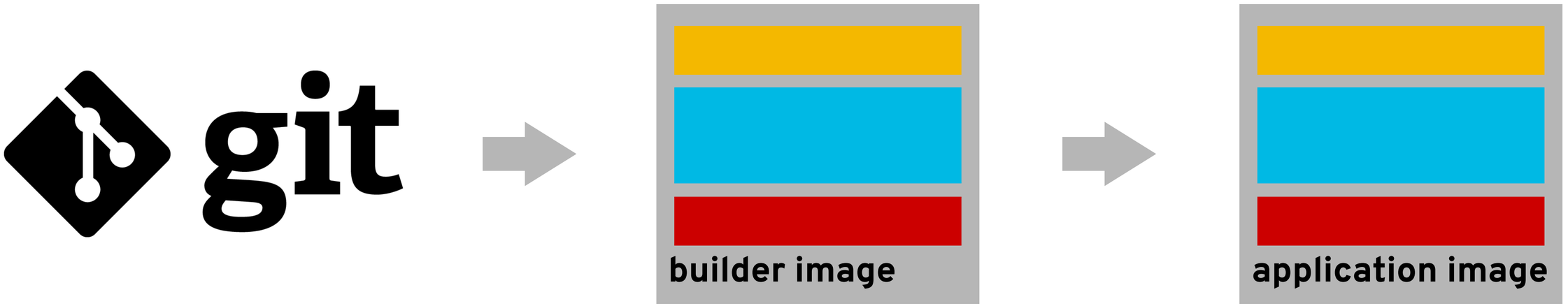


# More flexible:  source-to-image



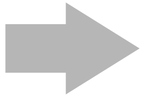
# More flexible:  source-to-image



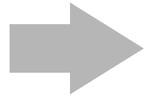

https://github.com/openshift/source-to-image

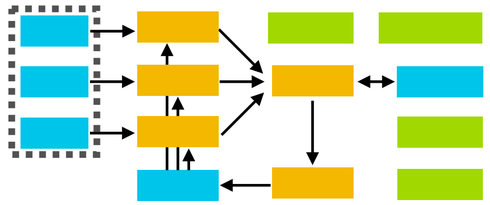
```
willb@echo % oc new-app \
getwarped/s2i-minimal-notebook:latest~https://github.com/willb/probabilistic-structu
res \
-e JUPYTER_NOTEBOOK_PASSWORD=developer
```

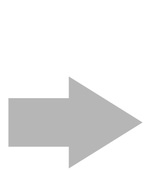
```
willb@echo % oc new-app \
getwarped/s2i-minimal-notebook:latest~https://github.com/willb/probabilistic-structu
res \
-e JUPYTER_NOTEBOOK_PASSWORD=developer
```



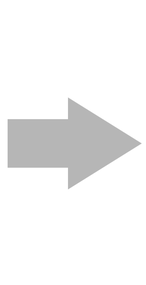

git

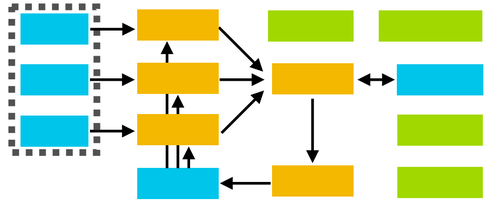


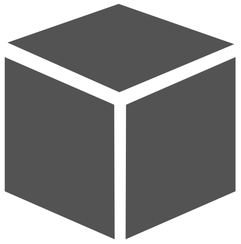


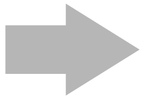



git

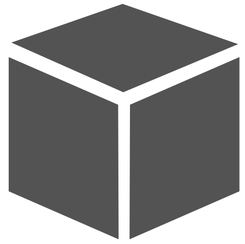


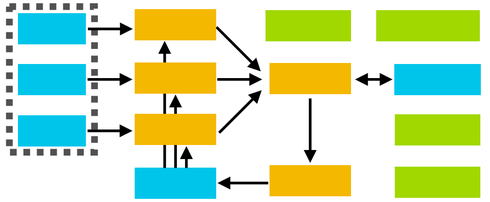






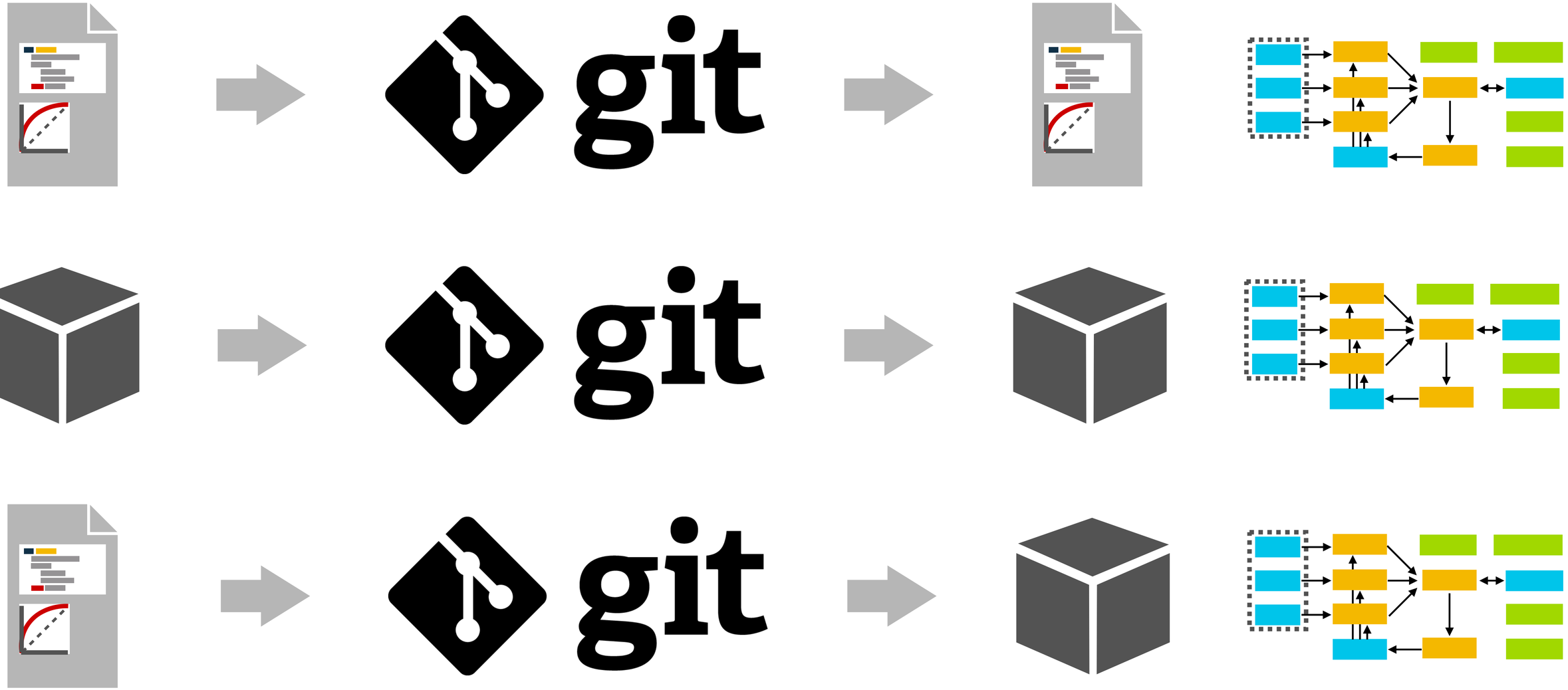

git



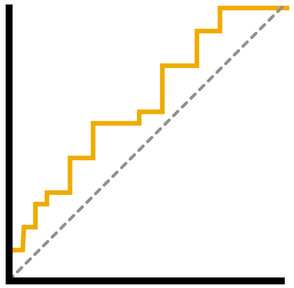


```
willb@echo % oc new-app --name model \
quay.io/willbenton/simple-model-s2i:demo\
~https://github.com/willb/example-model-s2i-notebook
```

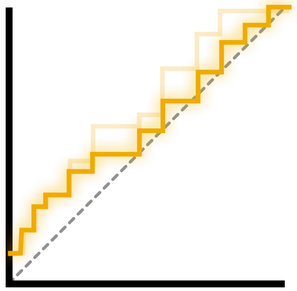
```
willb@echo % oc new-app --name model \
quay.io/willbenton/simple-model-s2i:demo\
~https://github.com/willb/example-model-s2i-notebook
```


git
git
git


mA

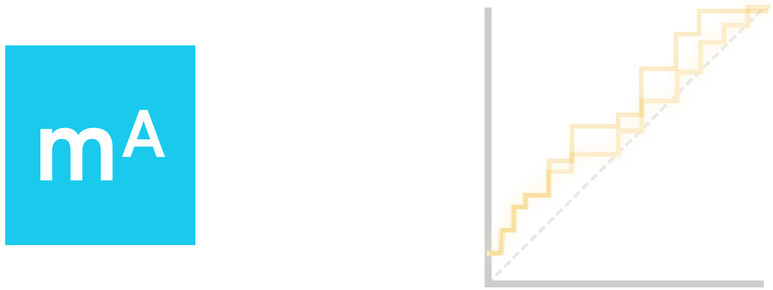



mA

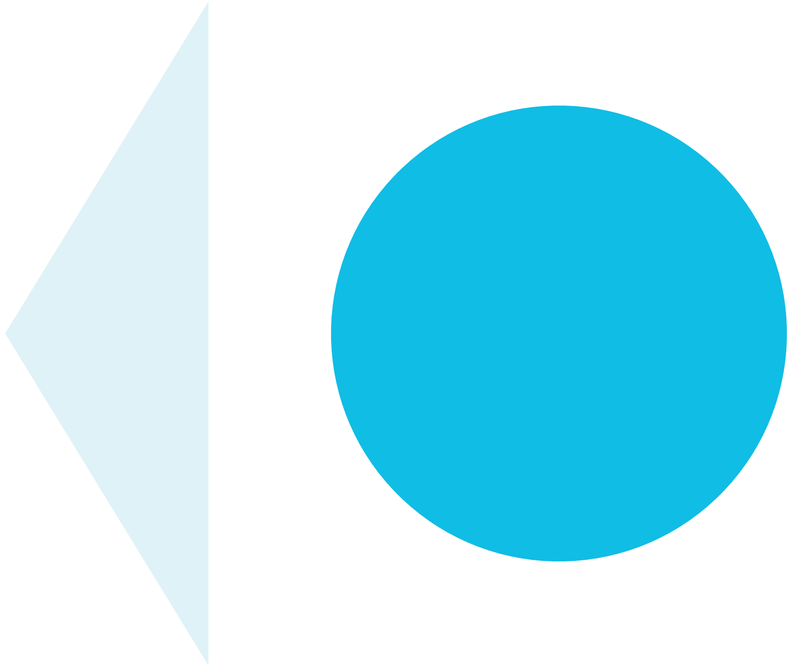


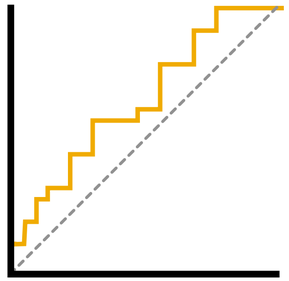
(joint) distribution of input data?

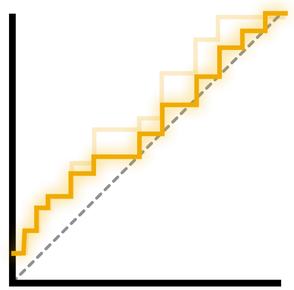
distribution of predictions?



distribution of acyclic paths
taken through scoring code?

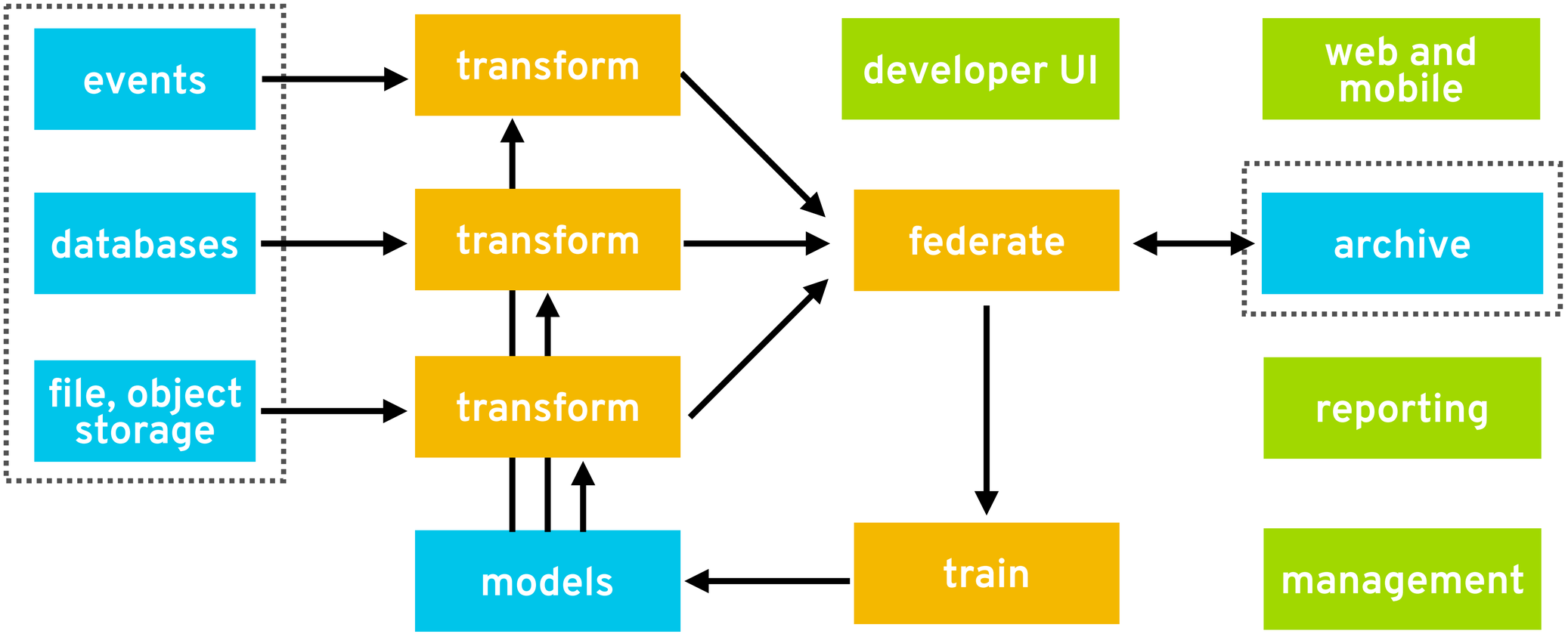


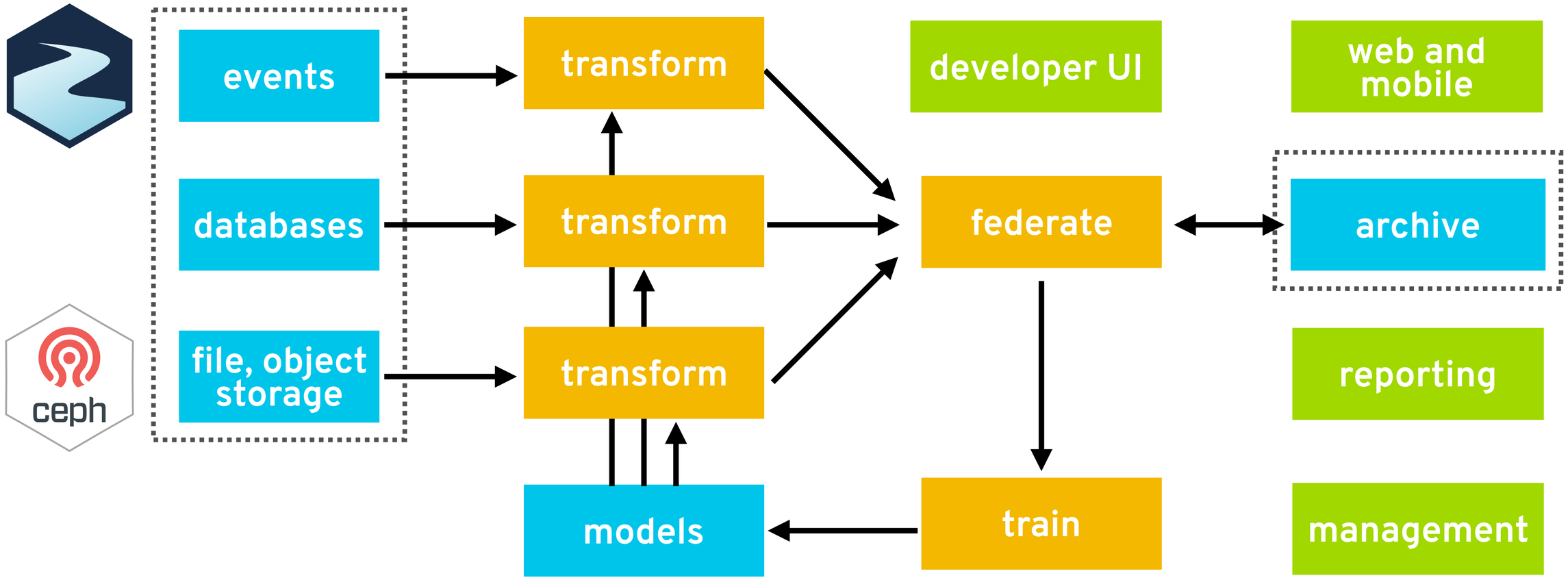


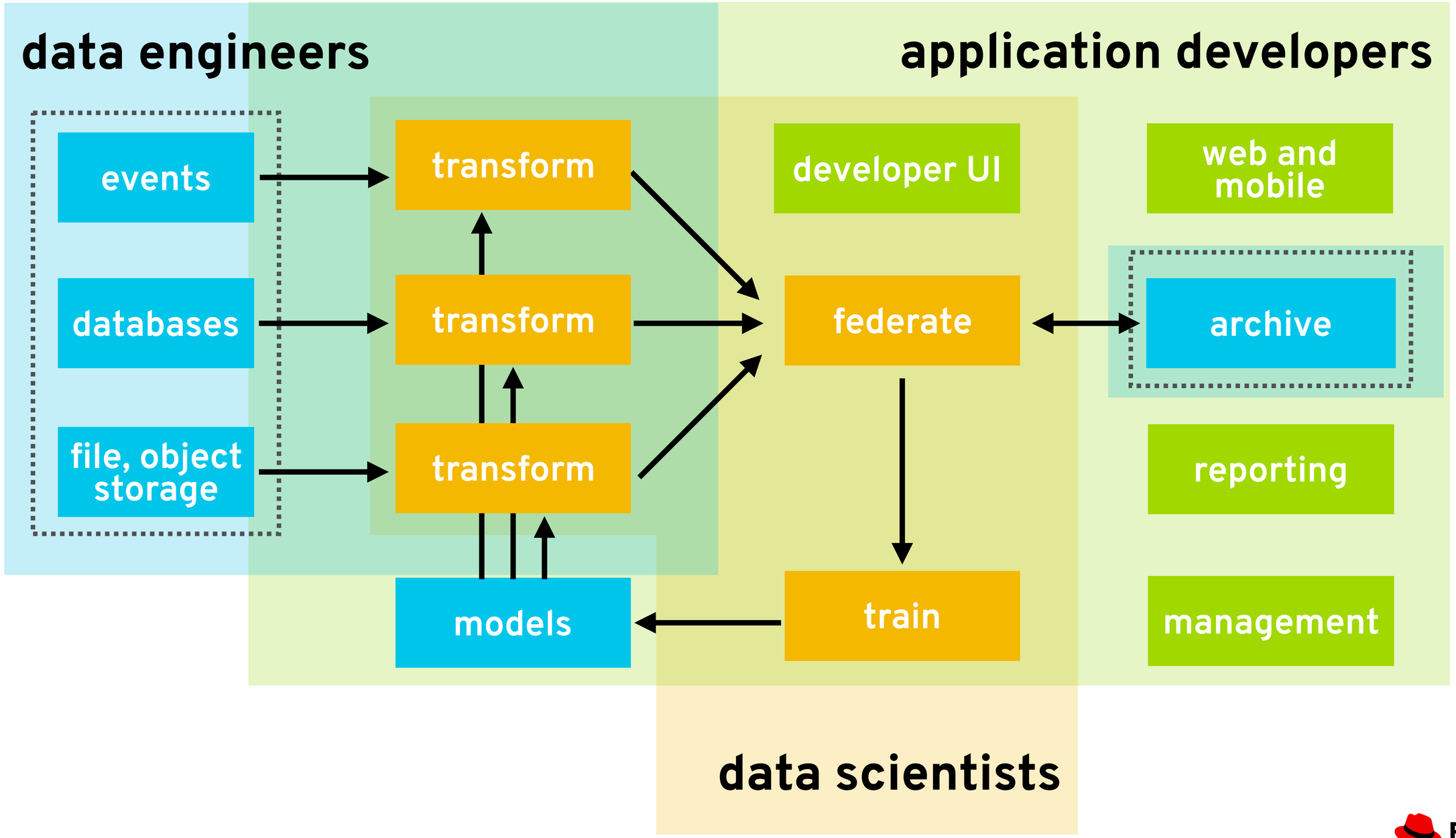


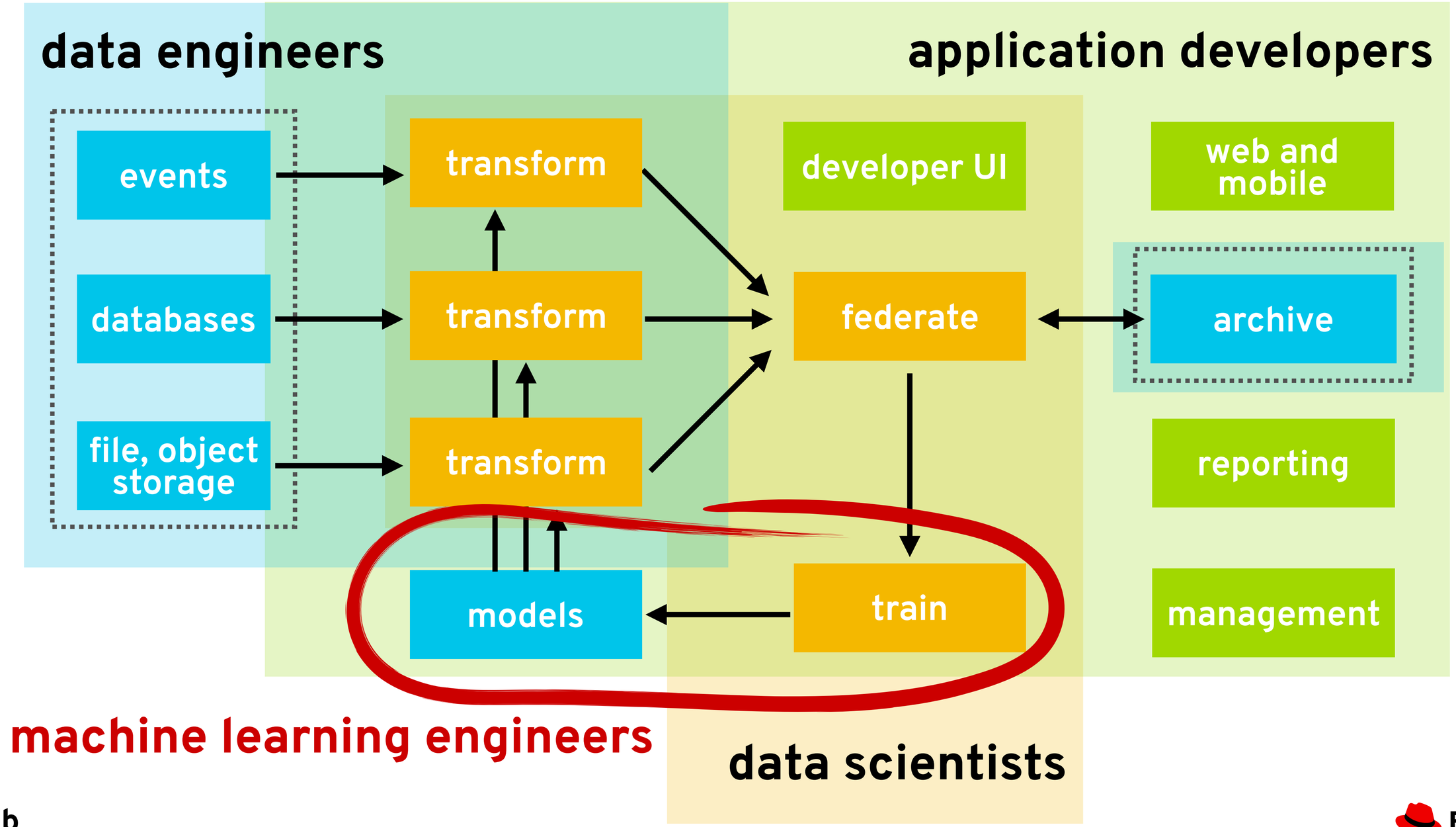
# Where from here?

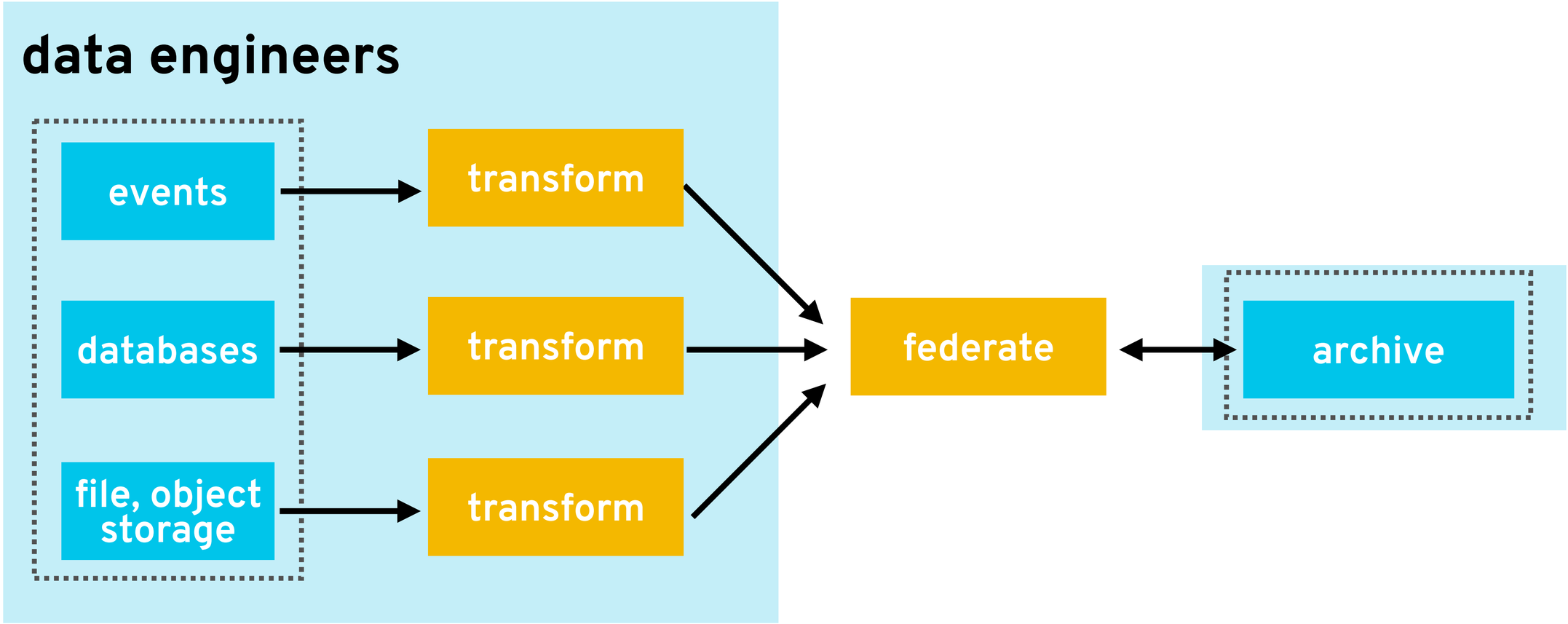
Intelligent applications continuously collect and learn from data to support essential functionality

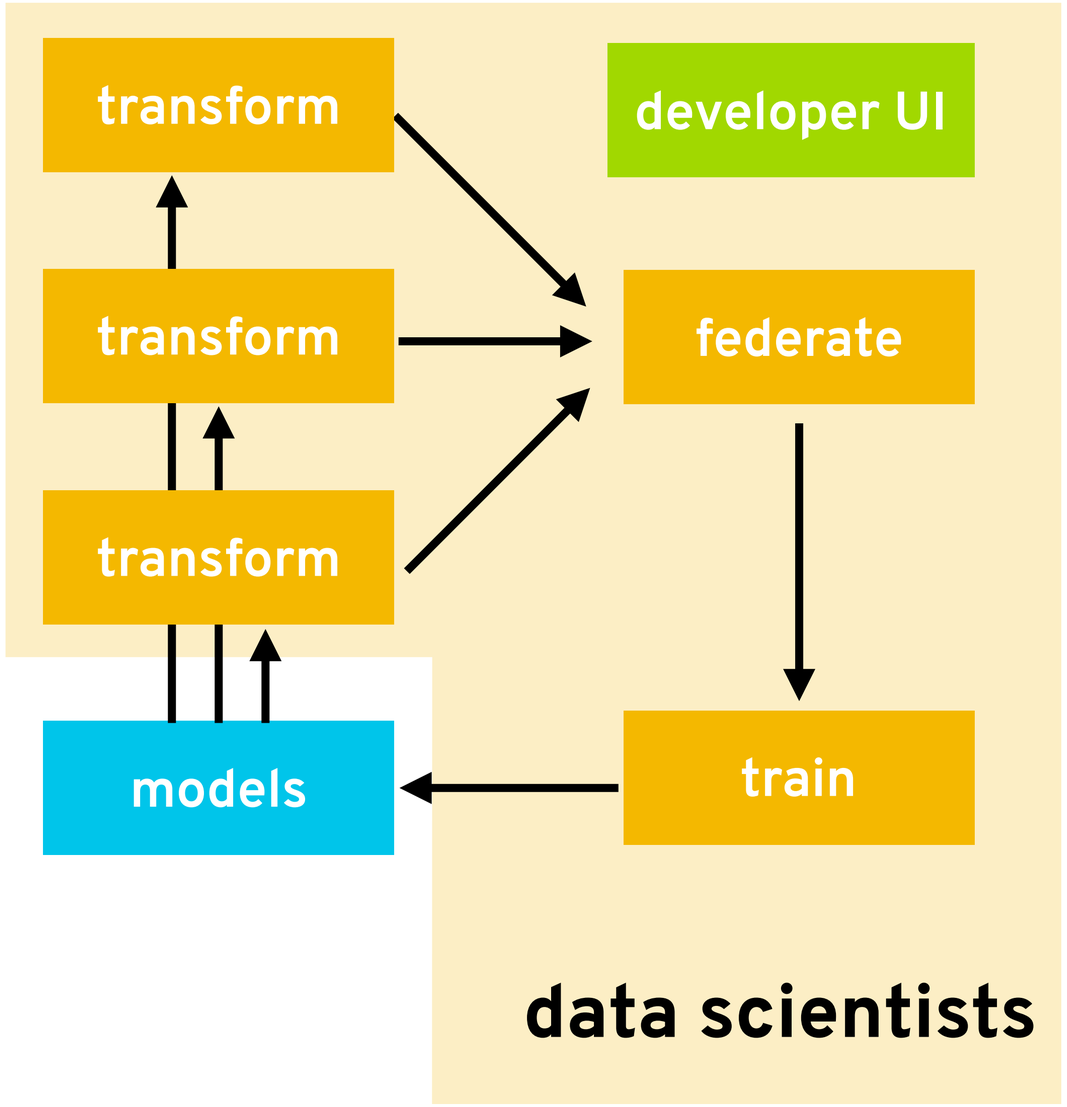

events
databases
file, object storage
transform
transform
transform
models
developer UI
federate
train
web and mobile
archive
reporting
management

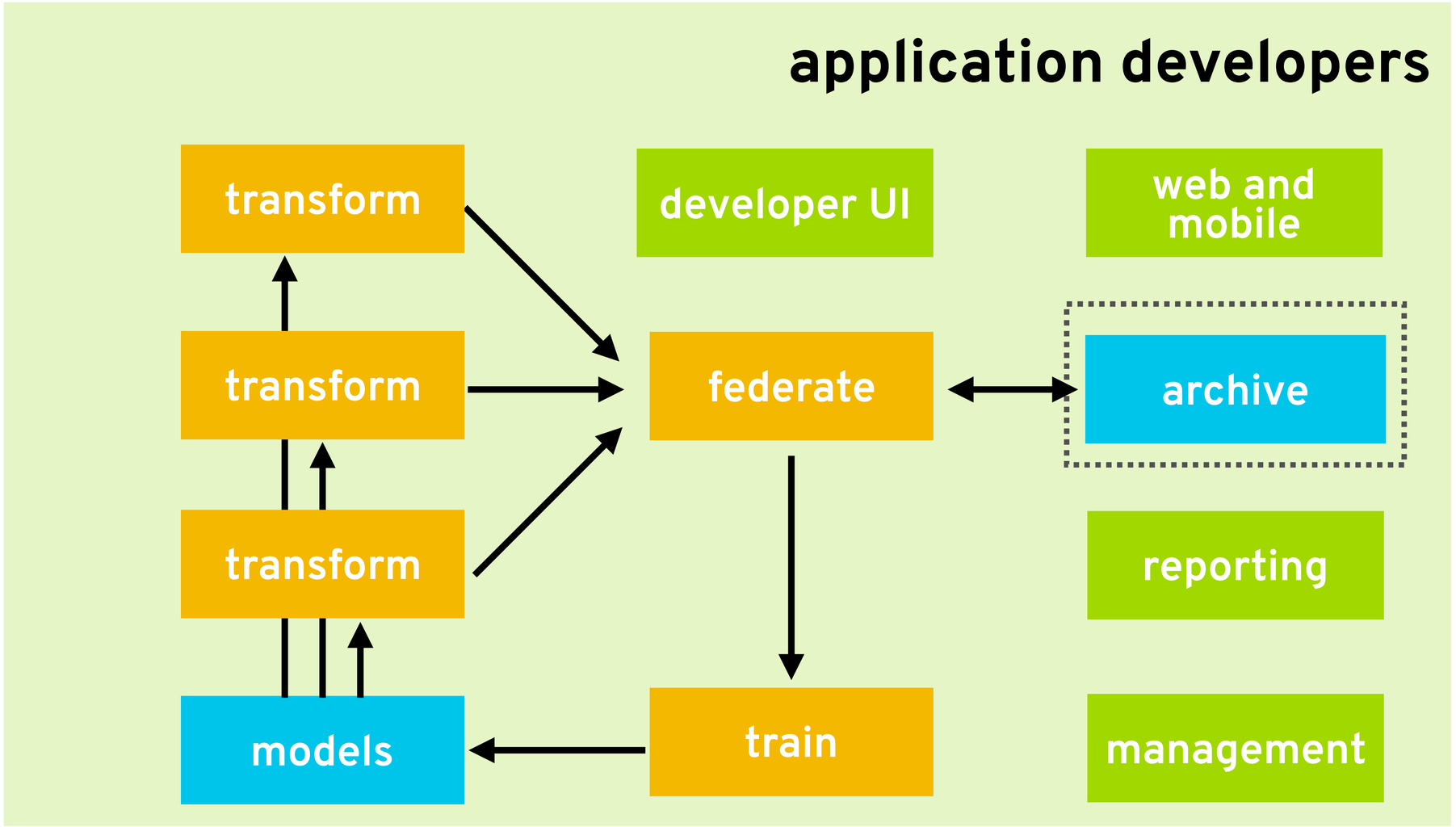

events
databases
file, object storage
ceph
transform
transform
transform
models
developer UI
federate
train
web and mobile
archive
reporting
management


data engineers
application developers
events
databases
file, object storage
transform
transform
transform
models
developer UI
federate
train
web and mobile
archive
reporting
management
data scientists


data engineers
application developers
events
databases
file, object storage
transform
transform
transform
developer UI
federate
web and mobile
archive
reporting
management
models
train
machine learning engineers
data scientists


data engineers
events
databases
file, object storage
transform
transform
transform
federate
archive

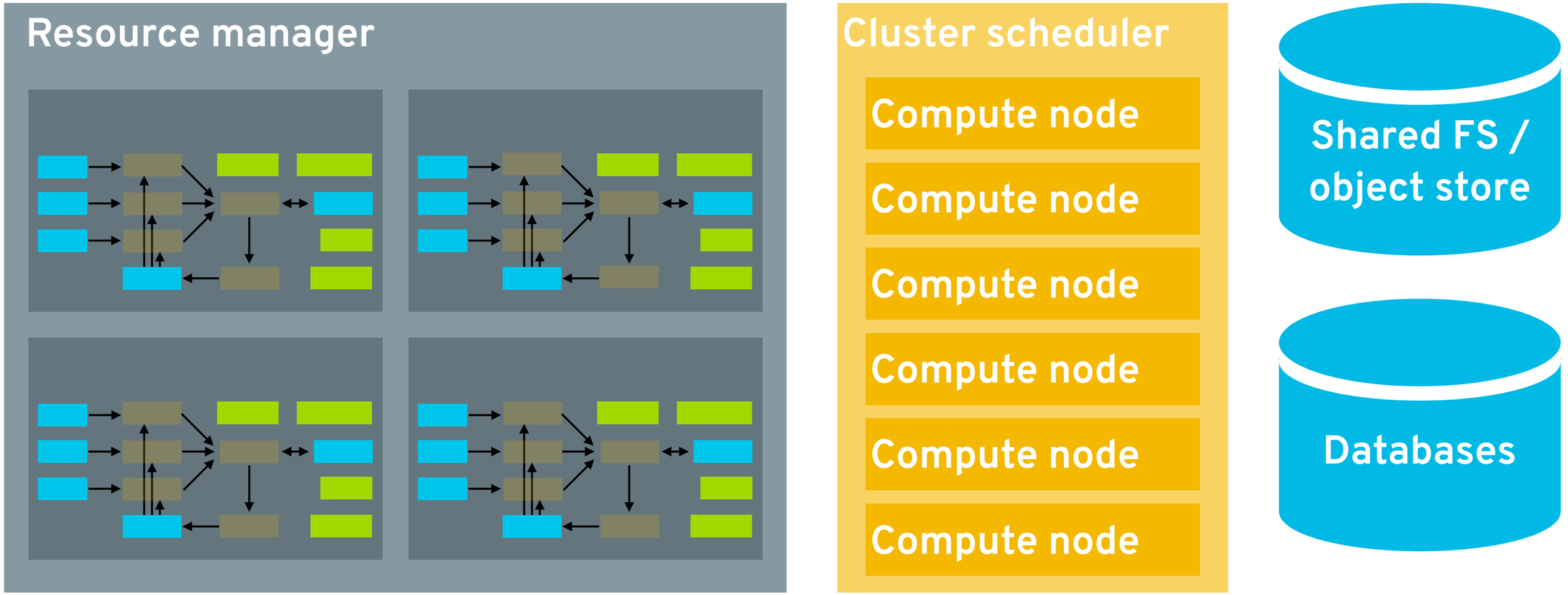

transform
developer UI
transform
federate
transform
models
train
data scientists


application developers
transform
transform
transform
models
developer UI
federate
train
web and mobile
archive
reporting
management

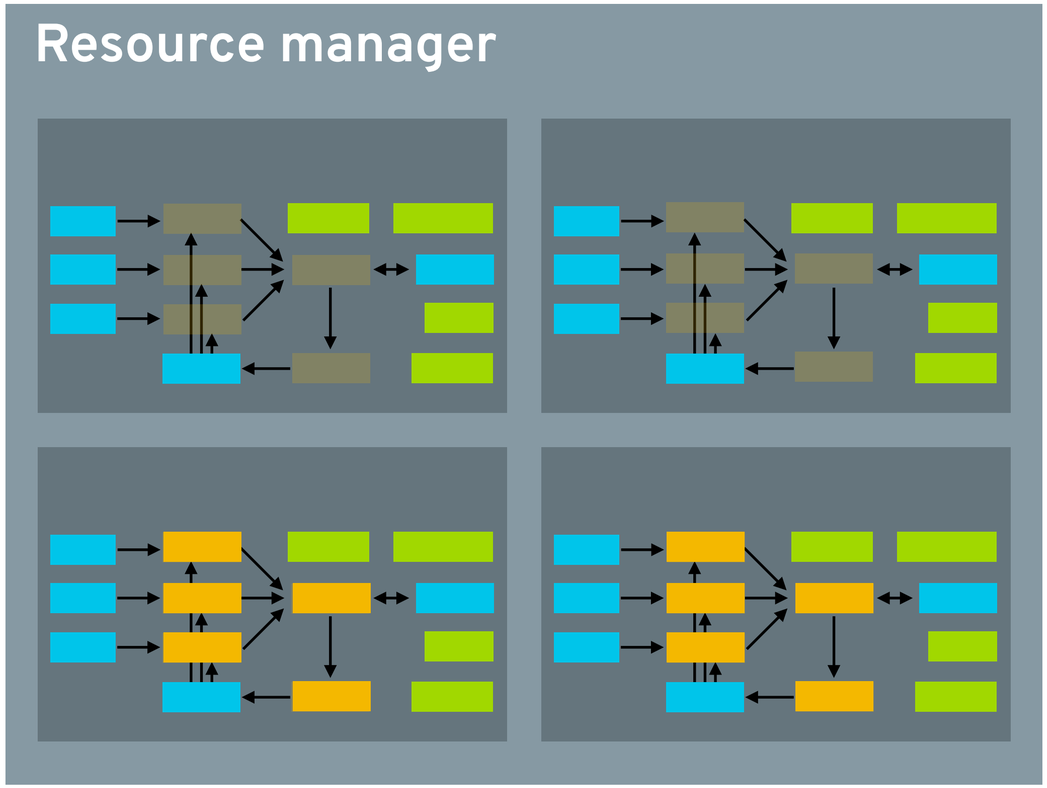
resource management
for apps and compute

**OpenShift** is **enterprise Kubernetes**
with a **great developer experience.**

# Multitenant compute clusters



# Multitenant compute clusters



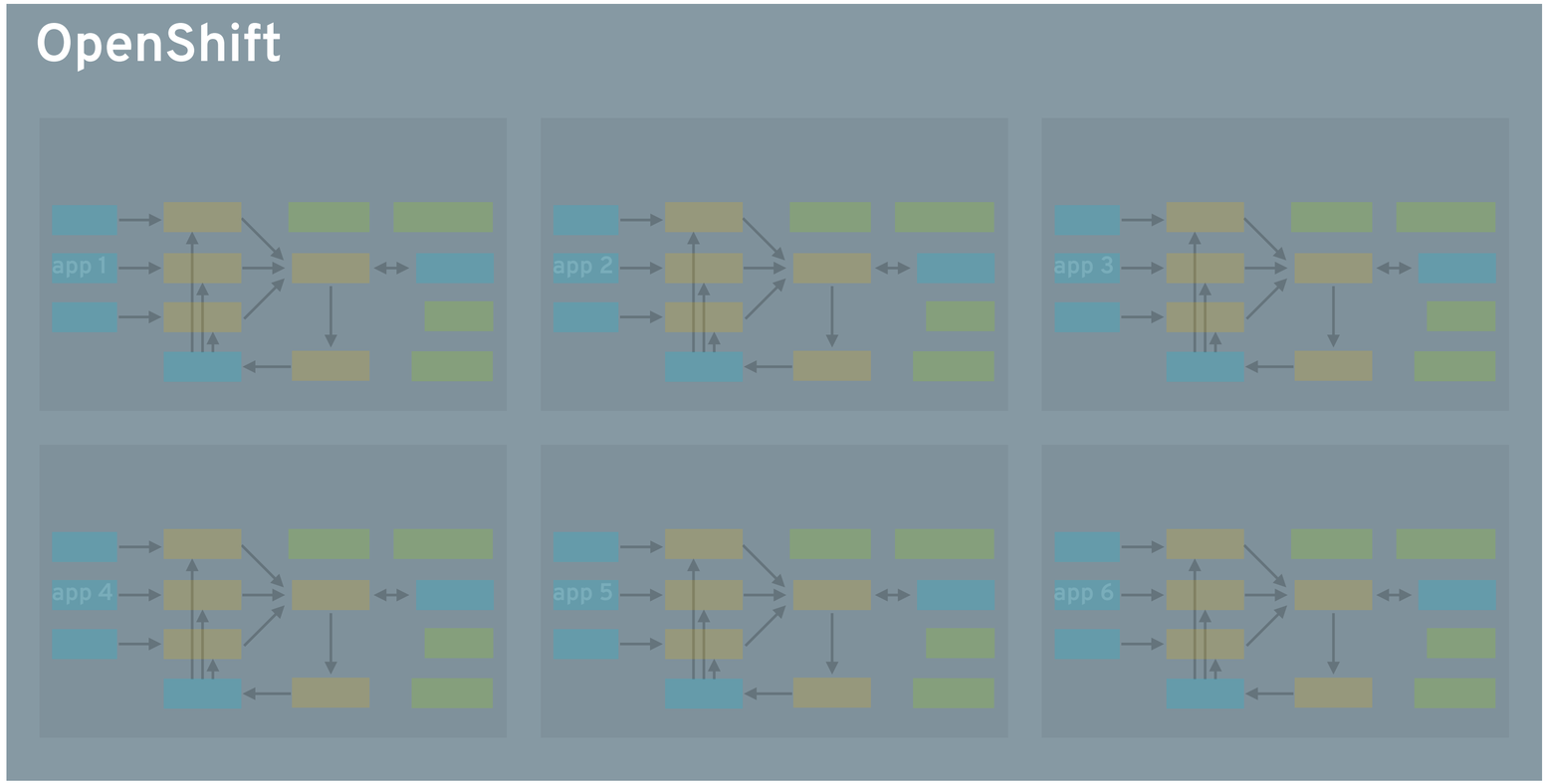

Cluster scheduler

- Compute node
- Compute node
- Compute node
- Compute node
- Compute node
- Compute node

Shared FS / object store

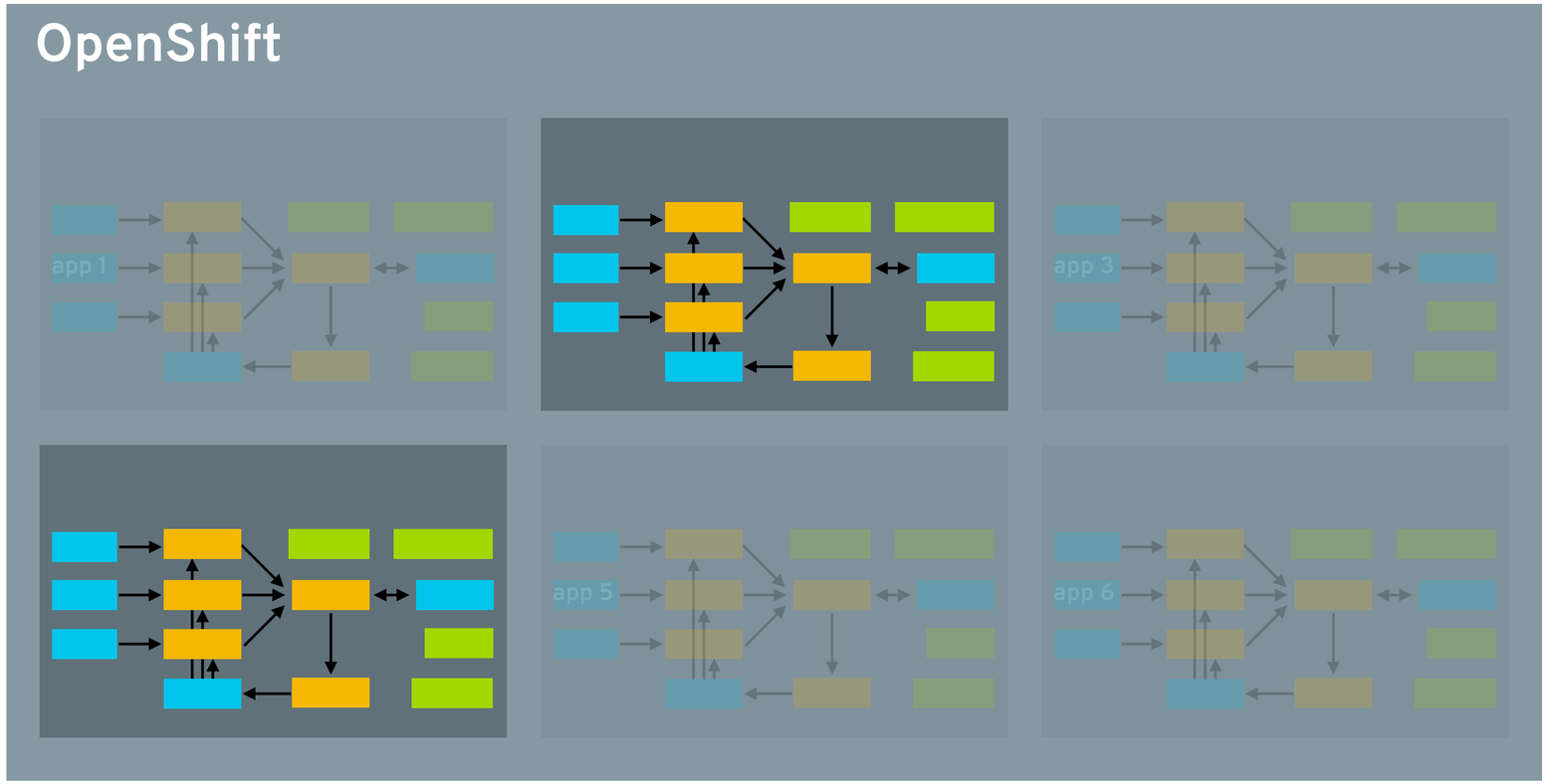
Databases

# One cluster per application





# One cluster per application

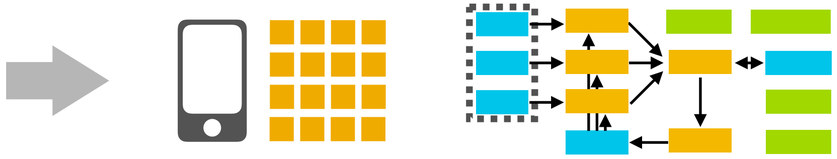



Shared FS / object store

Databases

# radanalytics.io



# opendatahub.io


# Kubeflow


# Other technologies

$$\mathbf{x} \cdot \mathbf{y} = x_1 \cdot y_1 + \cdots + x_n \cdot y_n$$

# Other technologies

$$\mathbf{x} \cdot \mathbf{y} = x_1 \cdot y_1 + \cdots + x_n \cdot y_n$$

```
def dot(xs, ys):
  return sum([x * y for x, y in zip(xs, ys)])
```

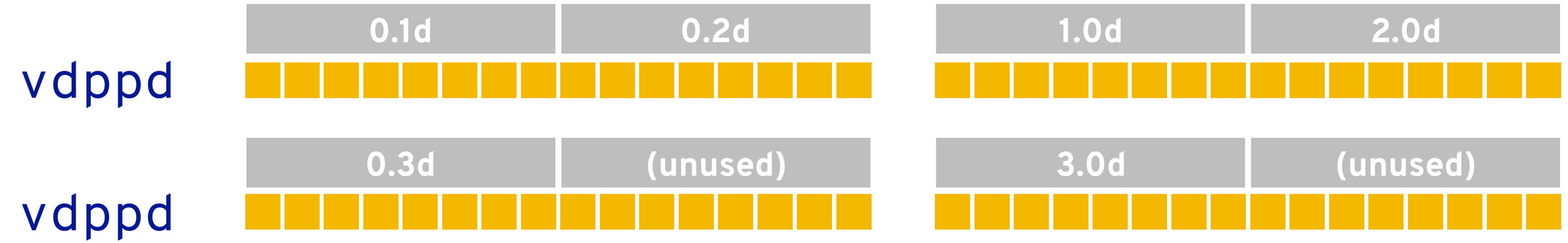
# Other technologies

$$\mathbf{x} \cdot \mathbf{y} = x_1 \cdot y_1 + \cdots + x_n \cdot y_n$$

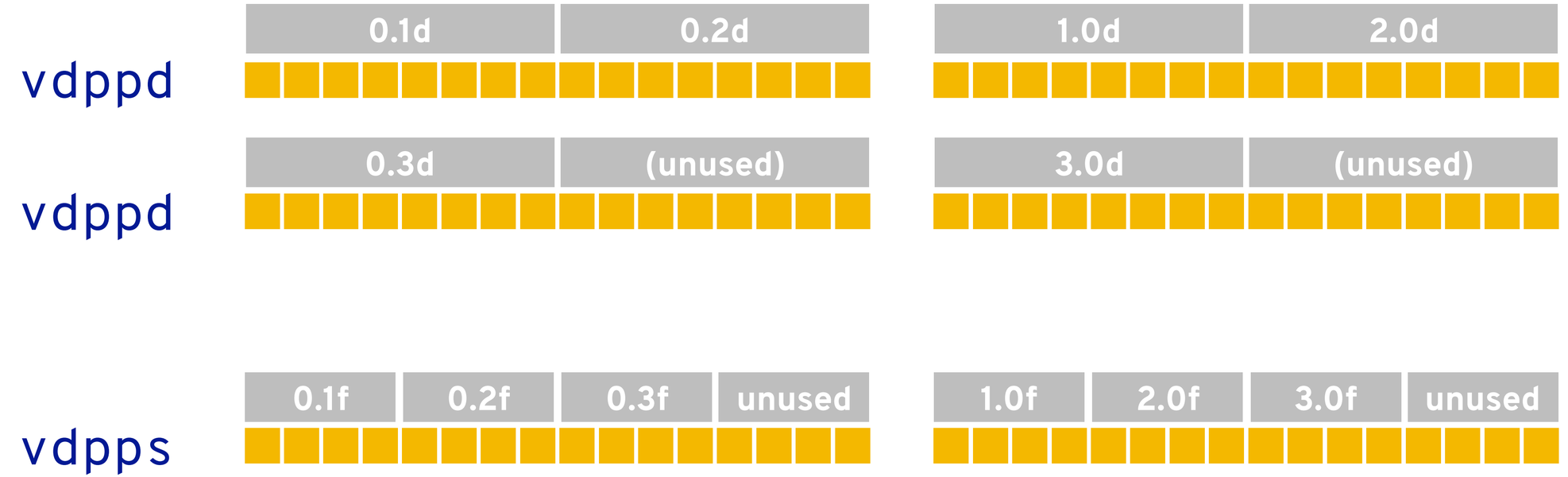
```
def dot(xs, ys):
  return sum([x * y for x, y in zip(xs, ys)])

dot([0.1, 0.2, 0.3], [1.0, 2.0, 3.0])
```

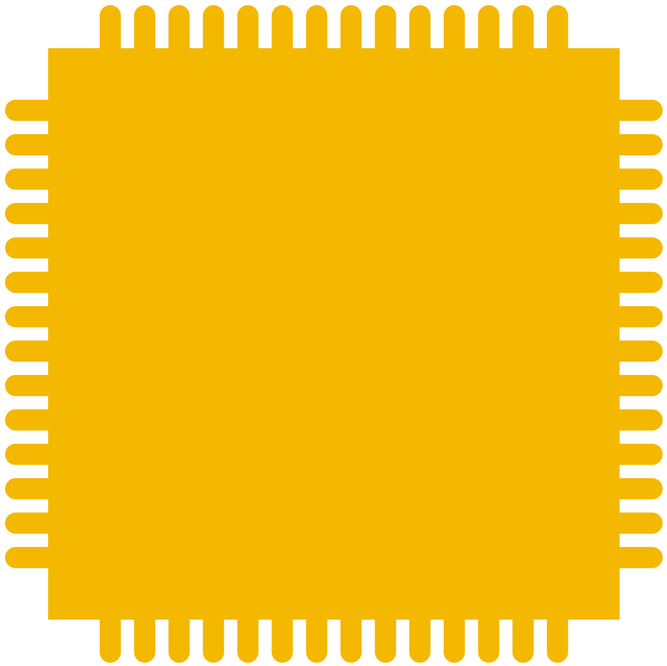
```
def dot(xs, ys):
    return sum([x * y for x, y in zip(xs, ys)])
```



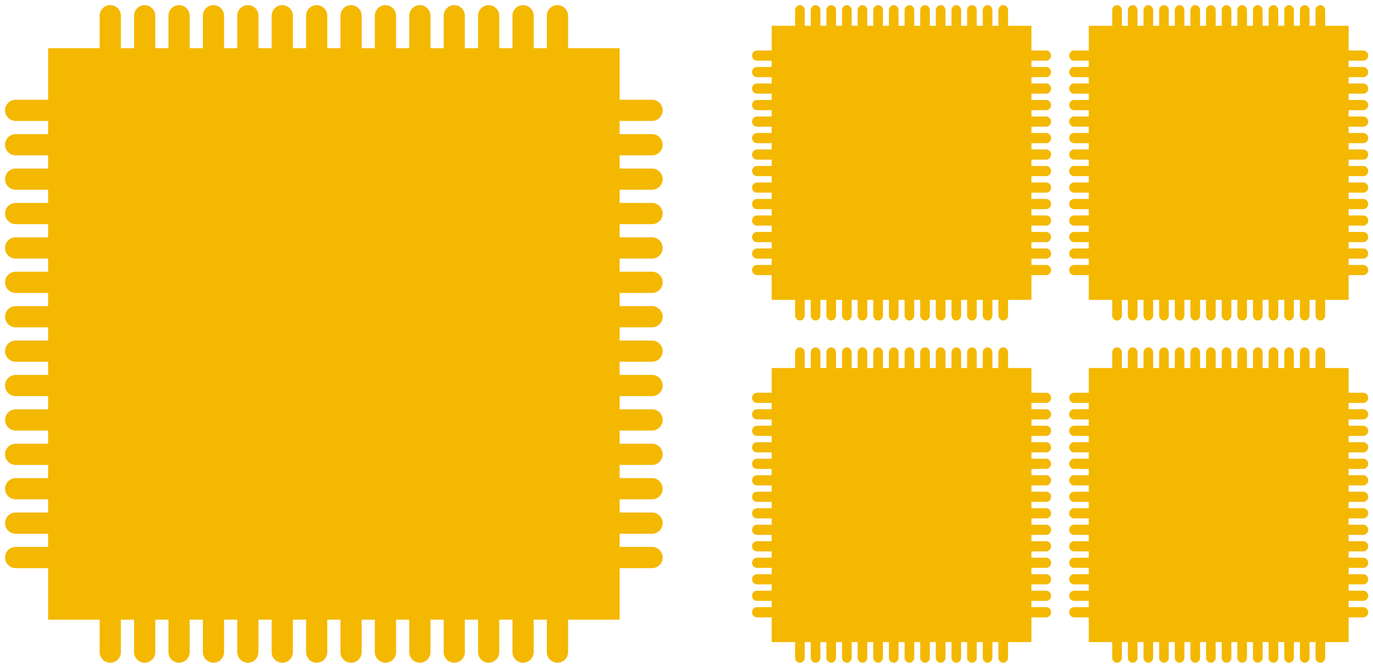
```
def dot(xs, ys):
    return sum([x * y for x, y in zip(xs, ys)])
```



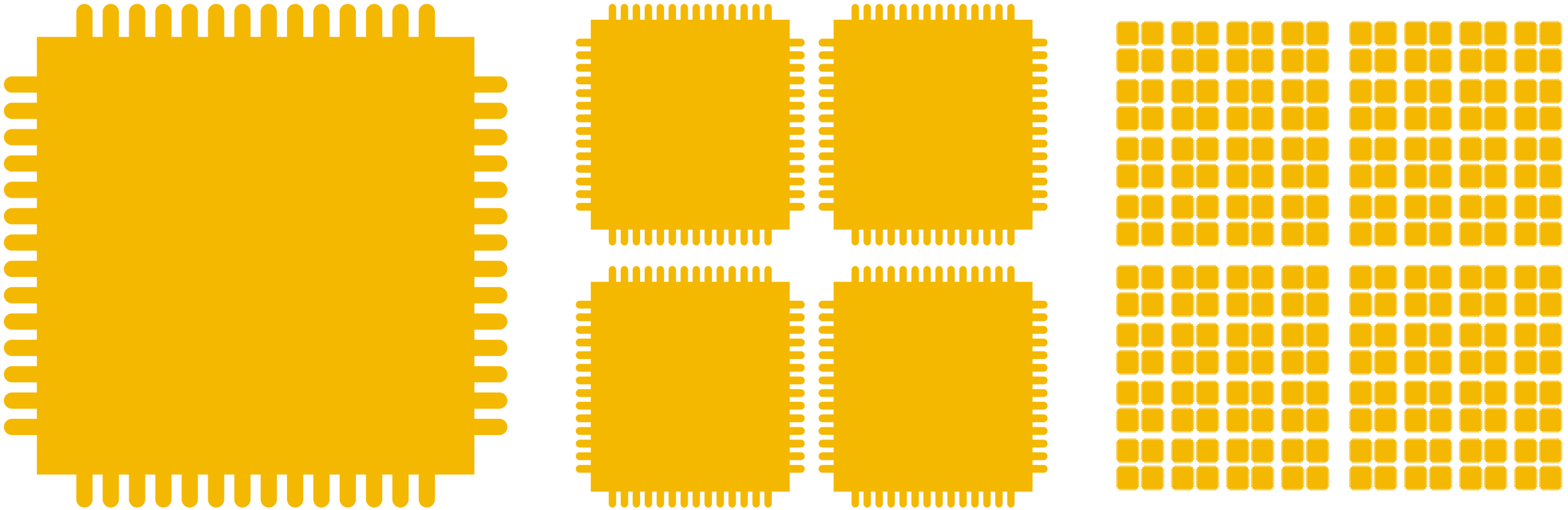
# Other technologies



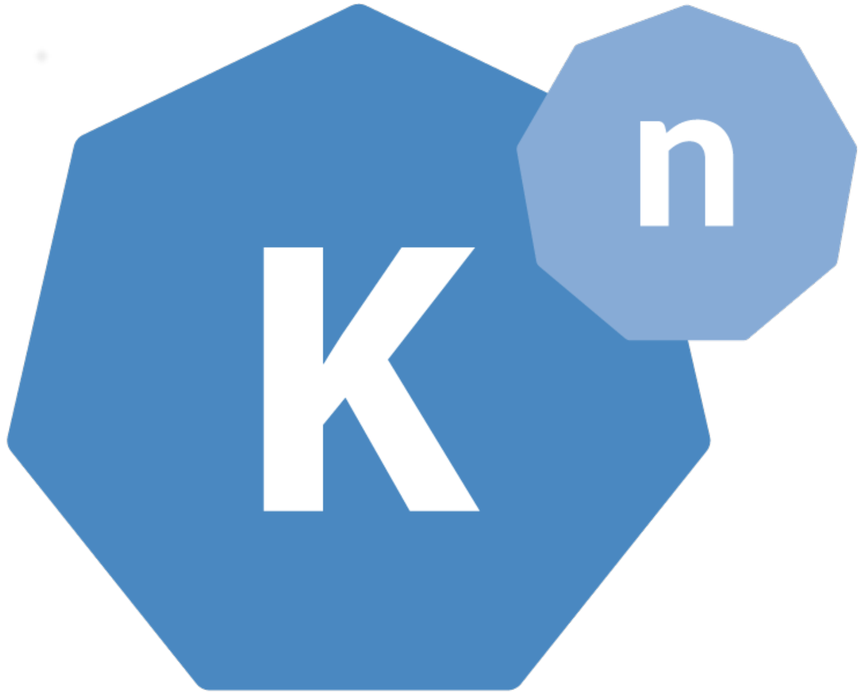
# Other technologies



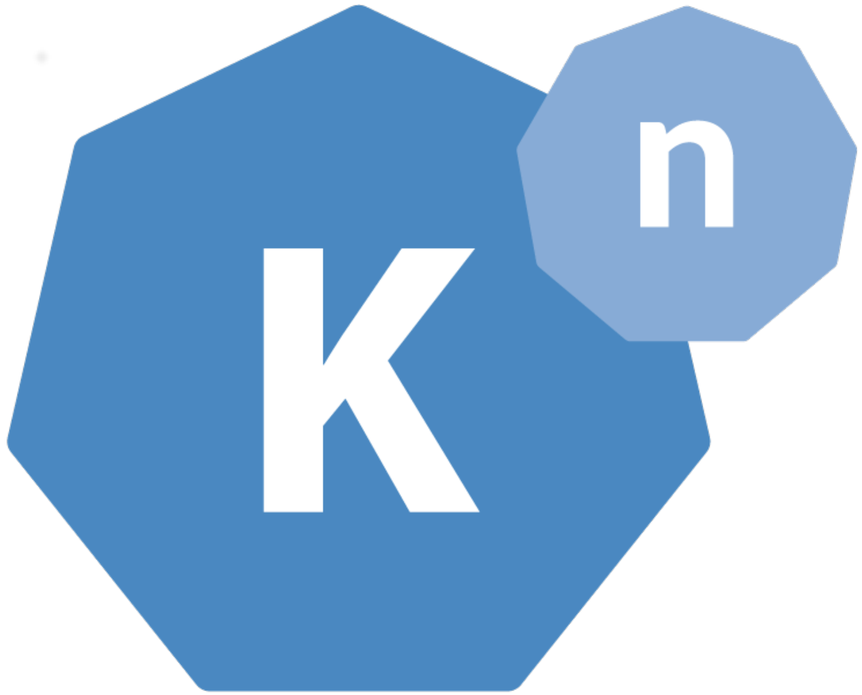
# Other technologies

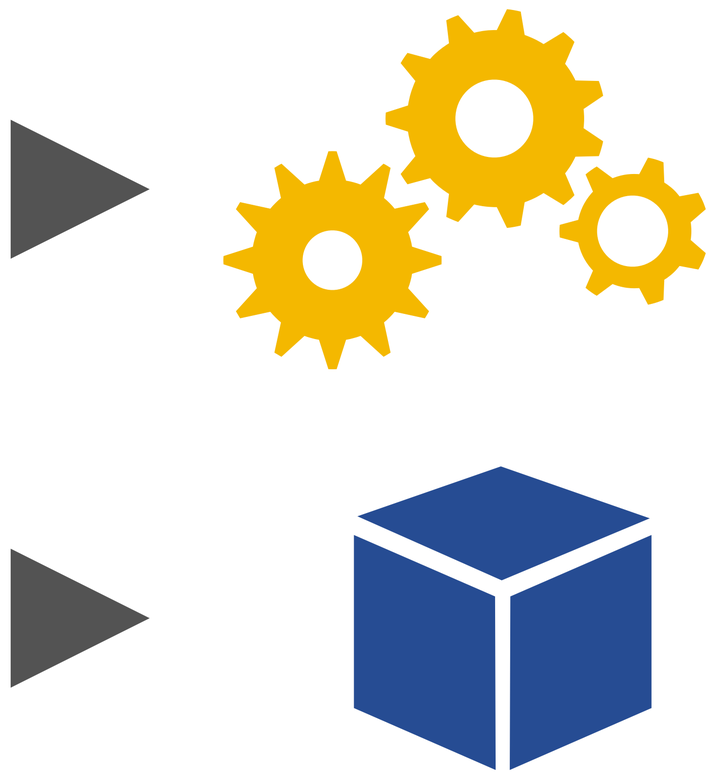


# Other technologies



# Other technologies



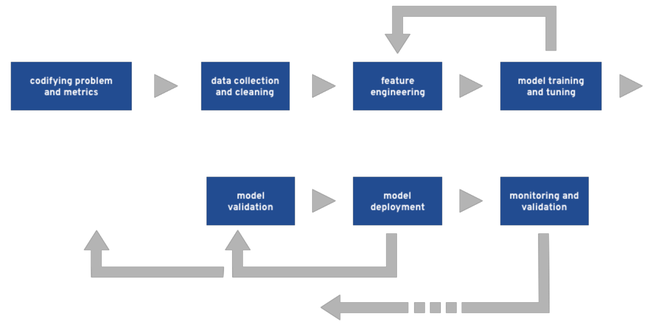



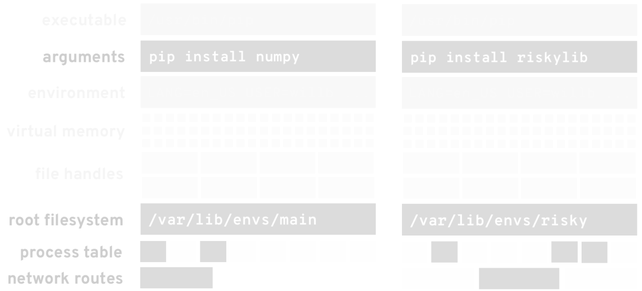
# Ecosystem demos at Summit



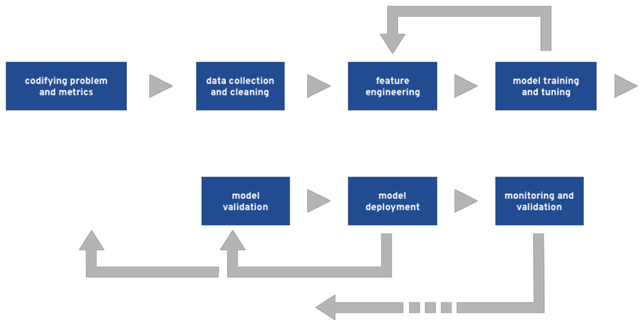


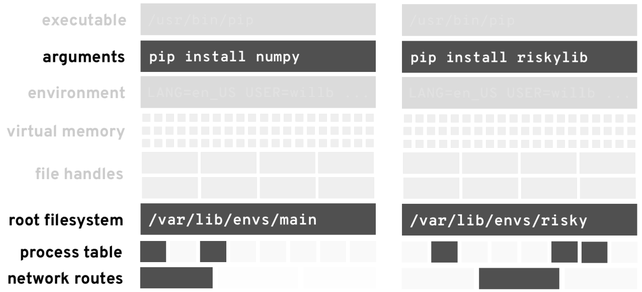

# What did we talk about today?


codifying problem and metrics
data collection and cleaning
feature engineering
model training and tuning
model validation
model deployment
monitoring and validation

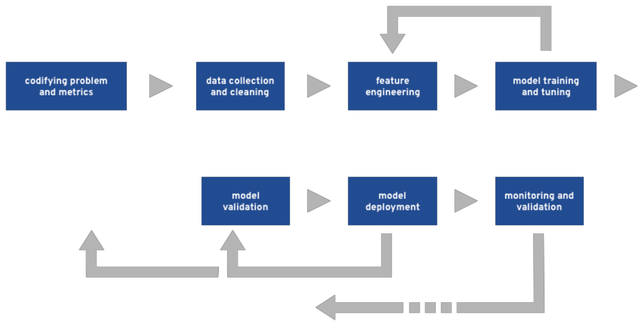

executable
arguments
pip install numpy
pip install riskylib
environment
virtual memory
file handles
root filesystem
/var/lib/envs/main
/var/lib/envs/risky
process table
network routes

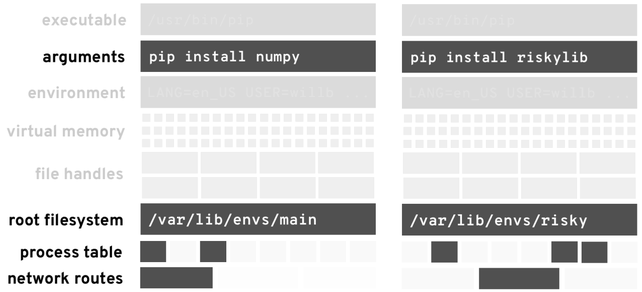

git
OK!


codifying problem and metrics
data collection and cleaning
feature engineering
model training and tuning
model validation
model deployment
monitoring and validation


executable
/usr/bin/pip
/usr/bin/pip
arguments
pip install numpy
pip install riskylib
environment
LANG=en_US USER=willb ...
LANG=en_US USER=willb ...
virtual memory
file handles
root filesystem
/var/lib/envs/main
/var/lib/envs/risky
process table
network routes


git
OK!



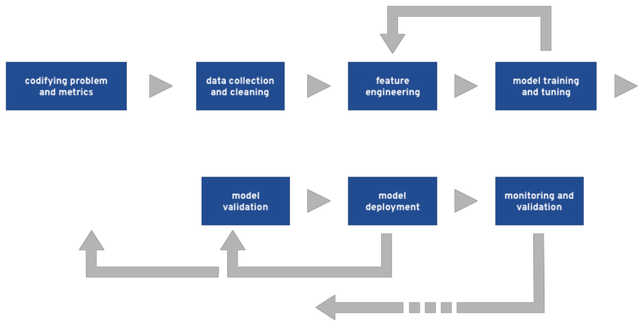



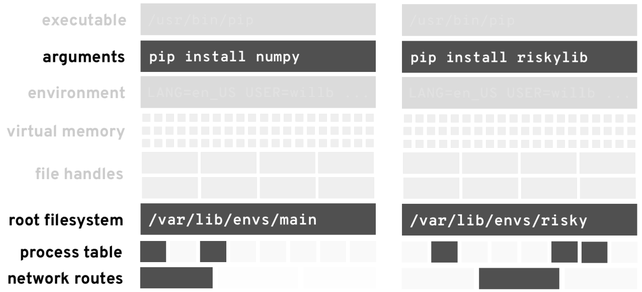




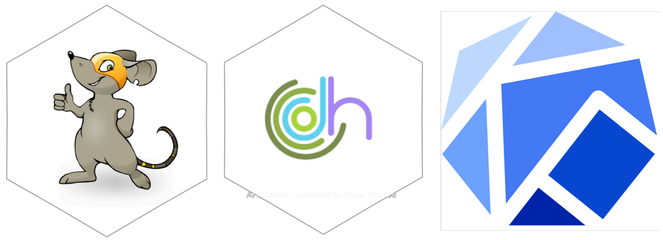
@willb

Red Hat





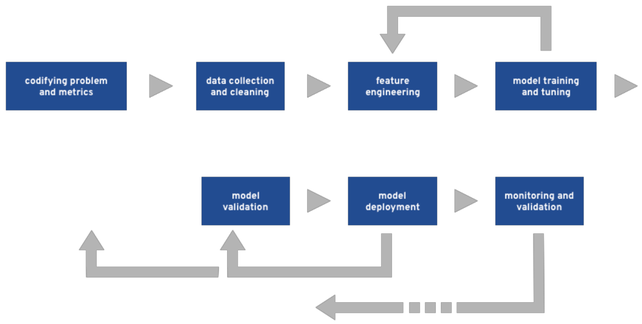


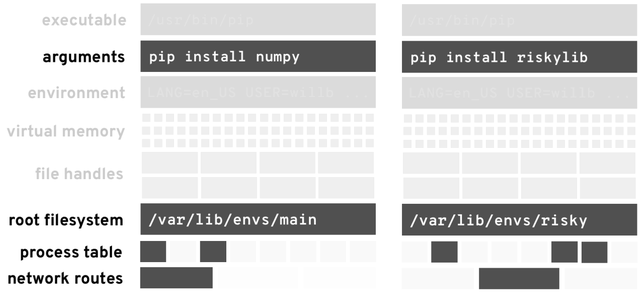




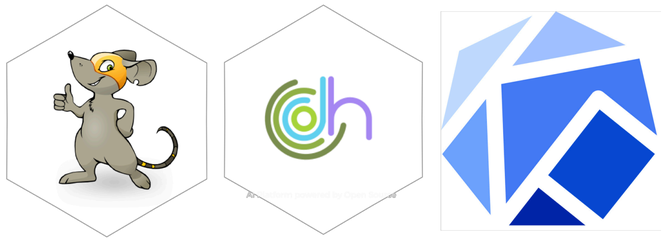
@willb

Red Hat







# THANKS

https://www.redhat.com

@willb

Red Hat