



Introduction to Kubeflow Pipelines

Michelle Casbon
Kubecon Europe
Barcelona
May 21, 2019

Google Cloud

TAs

Abhishek Gupta



Jeremy Lewi



Dan Sanche



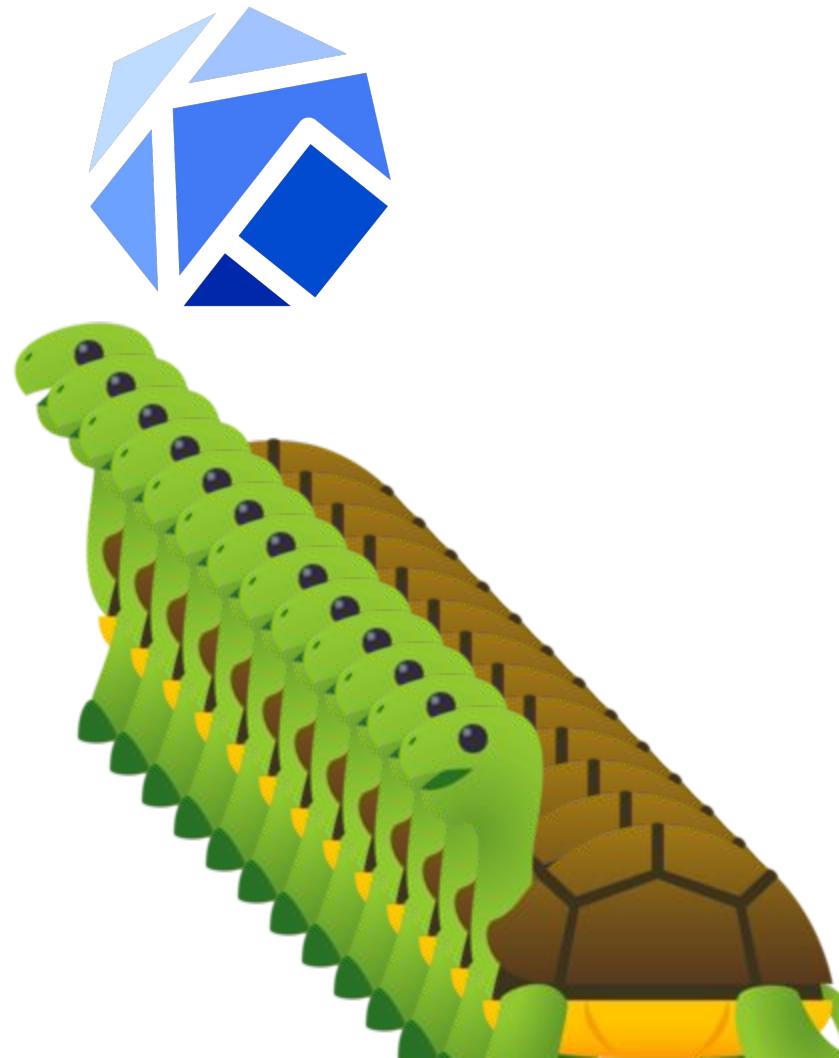
Dan Anghel



Michal Zylinski



@texasmichelle





Kubeflow is a curated set of compatible tools and artifacts that lays a foundation for running production ML apps

Enables consistency across deployments by providing Kubernetes object templates that bring together disparate components



Contributors

- Kubeflow is open
 - Open community
 - Open design
 - Open source
 - Open to ideas
- Get involved
 - github.com/kubeflow
 - kubeflow.slack.com 
 - @kubeflow 
 - kubeflow-discuss@googlegroups.com
 - Community call Tuesdays alternating 8:30am and 5:30pm Pacific
 - **Kubeflow Contributor Summit**
 - Q1 2019

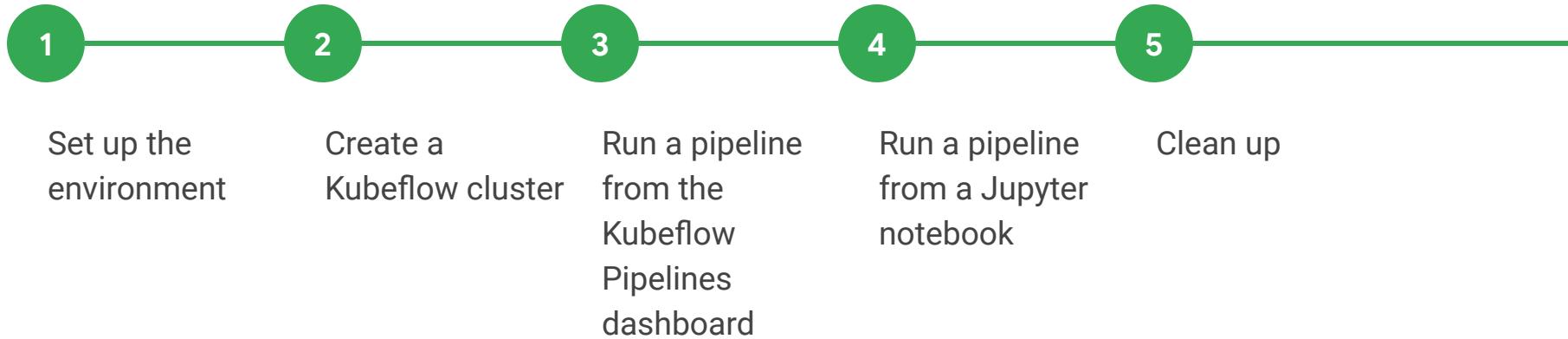


<https://github.com/kubeflow/kubeflow>



Agenda

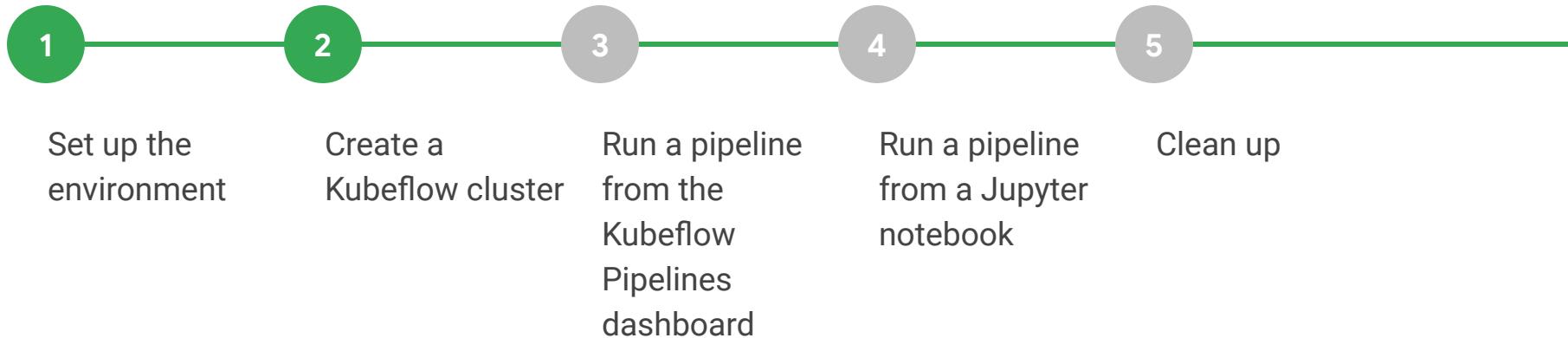
g.co/codelabs/kfp-gis



Agenda

g.co/codelabs/kfp-gis

Zone: europe-west1-b, europe-west1-d



PROBLEM

Moving from local to production



SOLUTION

Portability

Package infrastructure
components together





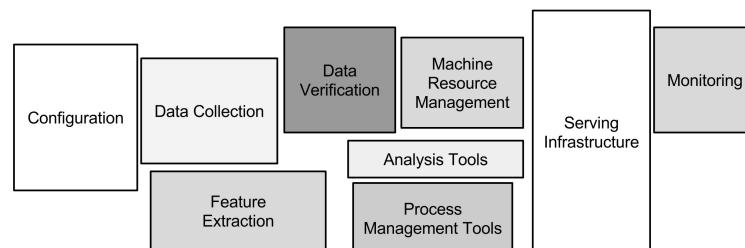
PROBLEM

Complexity

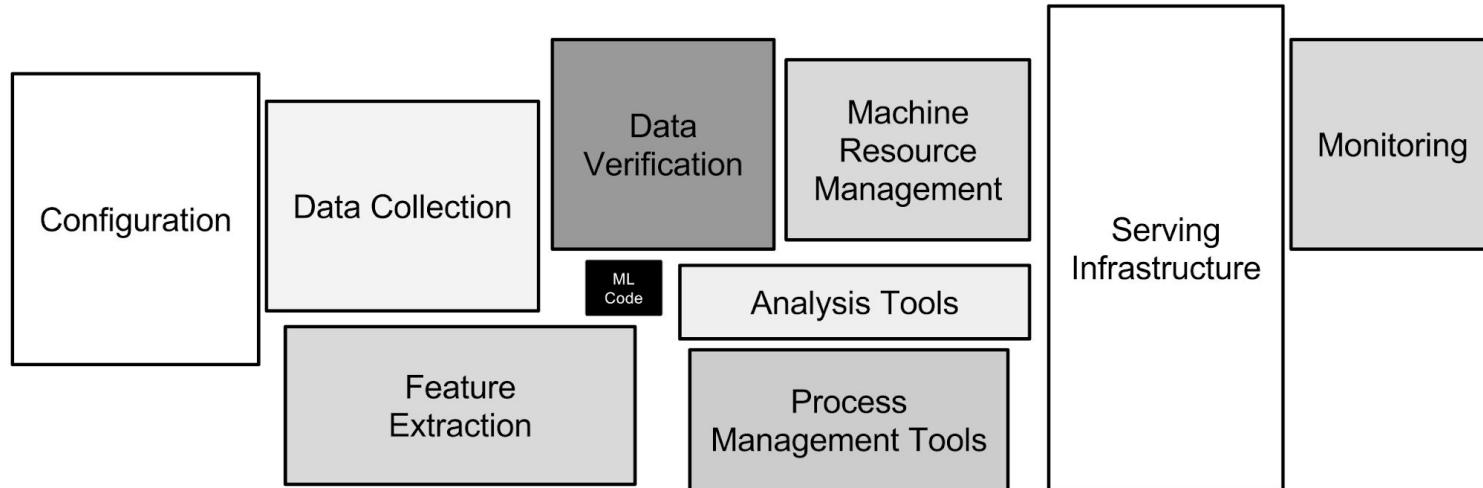


Perception

ML
Code



Reality



Data

Featurization

Training

Application

Platform

Data Ingestion

Feature Extraction

Model Building

Serving Infrastructure

Configuration

Data Exploration

Model Validation

Business Logic

Process Management

Data Transformation

Model Versioning

UI

Resource Management

Data Validation

Model Auditing

Load Balancing

Monitoring

Data Analysis

Distributed Training

Logging

Training Data Segmentation

Continuous Training

Continuous Delivery

PROBLEM

Complexity



SOLUTION

Composability

Logical groupings

Reusable components

PROBLEM

Maintainability

- Error resolution, recovery, & prevention
- Speed of iteration
- Versioning

SOLUTION

Composability

Shorten the development lifecycle

Automation

PROBLEM

Capacity Planning

- Usage patterns
- Demand spikes
- Efficient resource usage

SOLUTION

Scalability

Kubernetes

Autoprovisioning



Make it easy for everyone to develop,
deploy, and manage portable, scalable
ML everywhere

Kubeflow

Portability

Entire stack

Scalability

Native to k8s

Reduce variability
between services
& environments

Composability

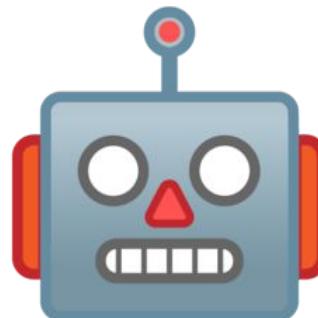
Single, unified tool
for common
processes

Full product lifecycle

Support specialized hardware, like GPUs & TPUs

Reduce costs

Improve model
performance



Kubeflow

Who

Data scientists

ML researchers

Software engineers

Product managers

What

Portable ML products on k8s

v0.5.0 release

Why

Because building a platform is too big of a problem to tackle alone

<https://github.com/kubeflow/kubeflow>

Kubeflow

Kubernetes-native platform for ML

Run wherever k8s runs

Use k8s to manage ML tasks

CRDs for distributed training

Adopt k8s patterns

Microservices

Manage infra declaratively

Package infrastructure components together

Kustomize & Ksonnet

Move between local -> dev -> test -> prod -> onprem

Support multiple ML frameworks

TensorFlow

Pytorch

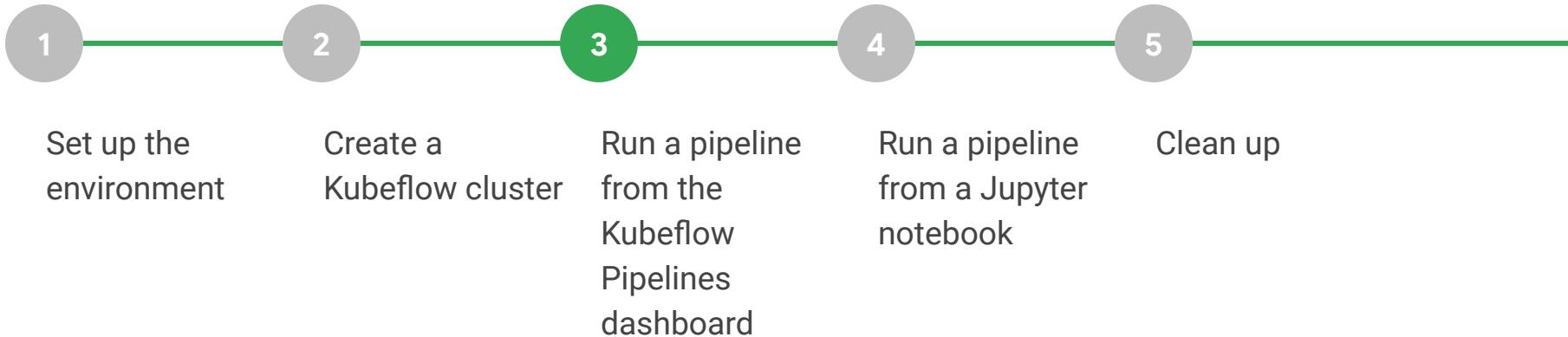
Scikit

Xgboost

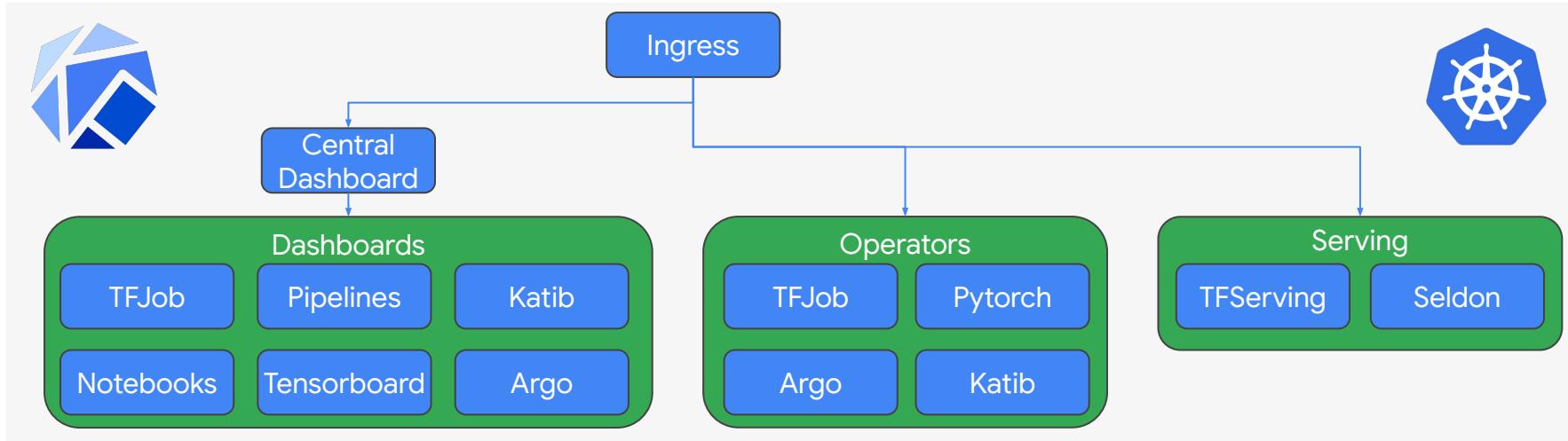
Et al.

Agenda

g.co/codelabs/kfp-gis



What's Inside v0.5?

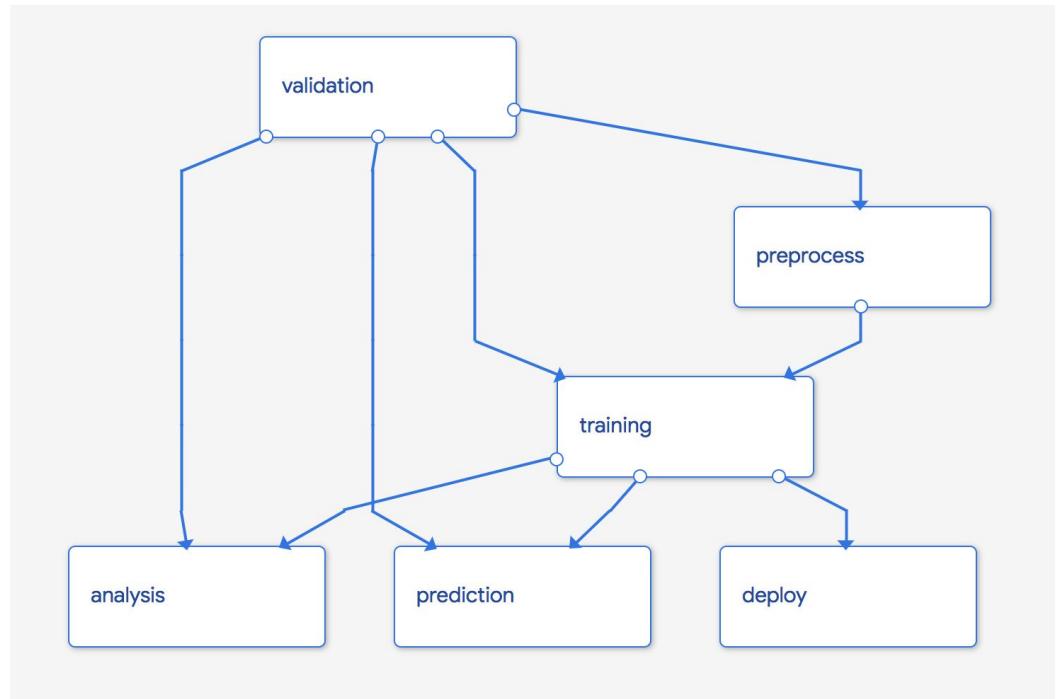


What's new in v0.5?

- Deploy
 - CLI: **kfctl** Go binary with bugfixes
 - Click-to-deploy: in-cluster auth
 - Istio support
- Develop
 - Central Dashboard overhaul
 - Notebooks UI
 - Fairing library
 - Build, train, & deploy directly from Python
 - Katib improvements

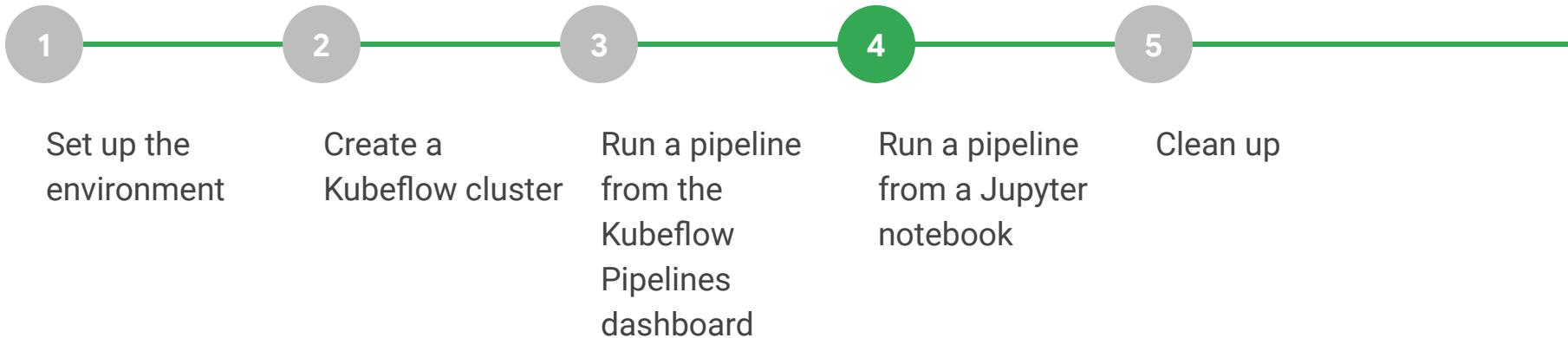
Pipelines

- End-to-end ML workflows
- Orchestration
- Service integration
- Components & sharing
- Job tracking, experimentation, monitoring
- Notebook integration



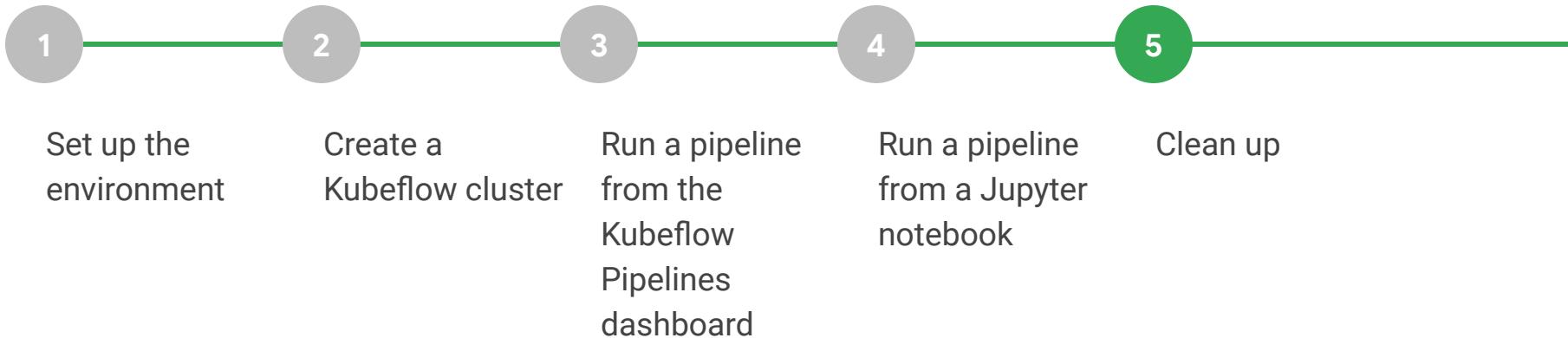
Agenda

g.co/codelabs/kfp-gis



Agenda

g.co/codelabs/kfp-gis



Roadmap

- v0.6 out in early summer
- Multi-user isolation
- Ksonnet replacement
- v1.0 Enterprise readiness
 - Model management
 - Hardened APIs
 - Clean deployments, upgrades
- **You tell us!** (Or better yet, help!)



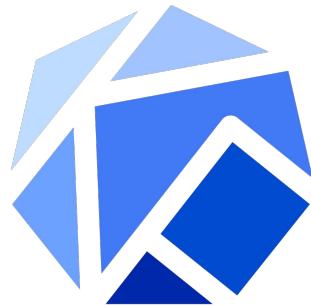
Contributors

- Kubeflow is open
 - Open community
 - Open design
 - Open source
 - Open to ideas
- Get involved
 - github.com/kubeflow
 - kubeflow.slack.com 
 - @kubeflow 
 - kubeflow-discuss@googlegroups.com
 - Community call Tuesdays alternating 8:30am and 5:30pm Pacific



<https://github.com/kubeflow/kubeflow>





Questions?