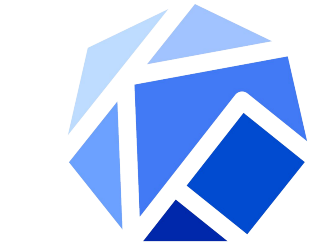



# Kubeflow v0.7
# Feature Review & Update

Kubeflow Community Presentation
December 2019

# Agenda

1. Kubeflow and Community Overview
2. Kubeflow User Survey
3. Kubeflow v0.7 Features
4. Wrap-up

# Make it Easy for Everyone to Develop, Deploy & Manage Portable, Composable, Distributed ML on Kubernetes

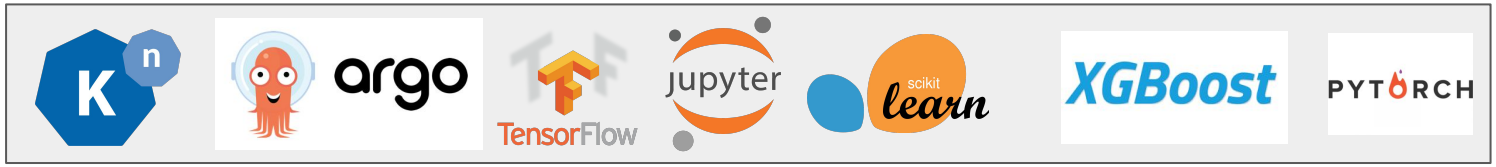
# Kubeflow

Platform

Lyft Learn | Bloomberg | Stripe Railyard | AirBnB BigHead | Google TFX | Many Others ..

Kubeflow

Applications




Infrastructure

Kubernetes | Spark | Borg

# From Single Apps to Complete Platform

**Introduce Kubeflow**
JupyterHub
TFJob
TFServing

**Kubeflow 0.2**
Katib -HP Tuning
Kubebench
PyTorch

**Kubeflow 0.4**
Pipelines
JupyterHub UI refresh
TFJob, PyTorch beta

**Kubeflow 0.6**
Metadata
Kustomize
Multi-user support

**Dec 2017**


**May 2018**

**Kubeflow 0.1**
Argo
Ambassador
Seldon

**Aug**

**Sep**

**Contributor Summit**

**Oct**

**Kubeflow 0.3**
kfctl.sh
TFJob v1alpha2

**2019 Jan**

**April**

**Kubeflow 0.5**
KFServing
Fairing
Jupyter WebApp + CR

**Jul**

**Nov**

**Kubeflow 0.7**
Pipelines+
KFServing v0.2
kfctl refactor

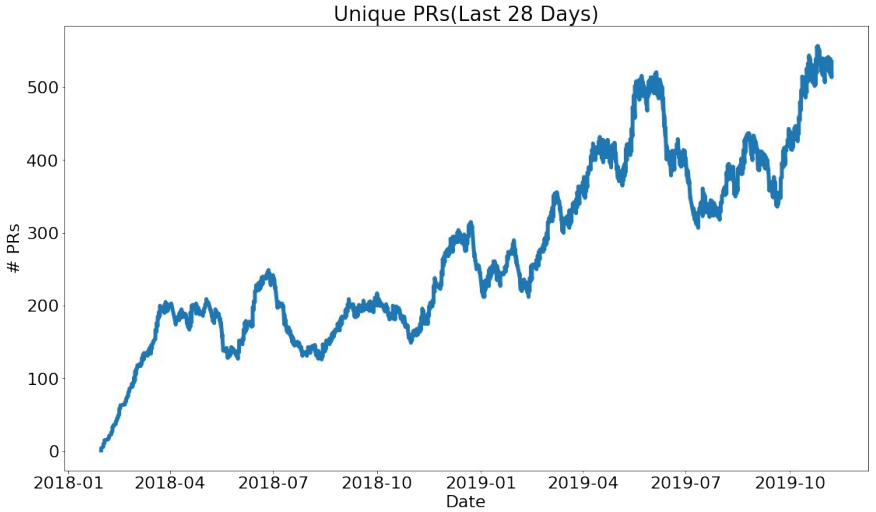
**Individual Applications**

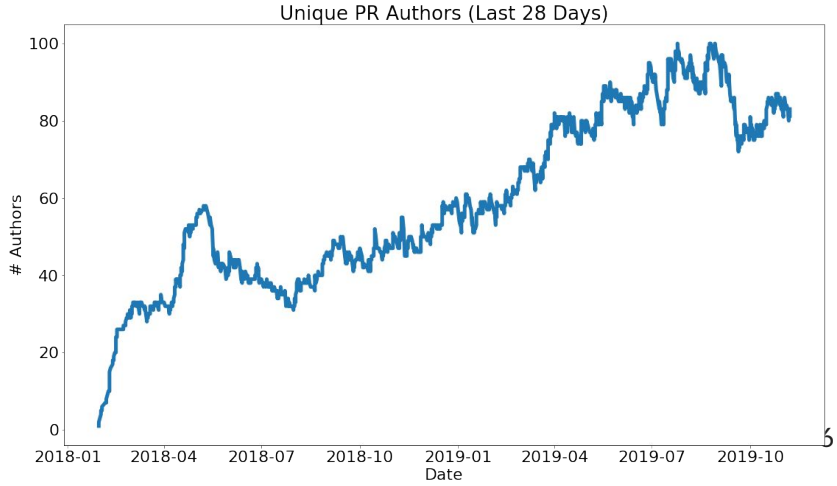
**Connecting Apps And Metadata**

**Productionization & Hardening**

# Community!


Arrikto
Alibaba Cloud
蚂蚁金服 ANT FINANCIAL
argo
aws
Bloomberg
Caicloud
CANONICAL
cisco
DATAWIRE
DELL
GitHub
GO-JEK
Google
heptio
IBM
iguazio
intel
Microsoft
MOMENTA
One Convergence
Pachyderm
PRIMER
jupyter
redhat
SELDON
shopify
tusimple
UBER
云知声 Unisound
weaveworks


Unique PRs(Last 28 Days)
# PRs
0
100
200
300
400
500
2018-01
2018-04
2018-07
2018-10
2019-01
2019-04
2019-07
2019-10
Date


Unique PR Authors (Last 28 Days)
# Authors
0
20
40
60
80
100
2018-01
2018-04
2018-07
2018-10
2019-01
2019-04
2019-07
2019-10
Date

# **Just a** SMALL sample of Community contributions

**Arrikto**

- Jupyter manager UI
- Pipelines volume support
- MiniKF
- Auth with Istio + Dex
- On-premise installation

**Bloomberg**

- KFServing

**Cisco**

- Auth with Istio + Dex
- Katib
- KubeBench
- PyTorch
- On-premise installation

**GoJEK**

- Feast feature store

**IBM**

- Pipeline components for spark, ffdl

**Intel**

- kfctl (CLI & library) & kustomize
- OpenVino

**Intuit**

- Argo

**RedHat + NVIDIA**

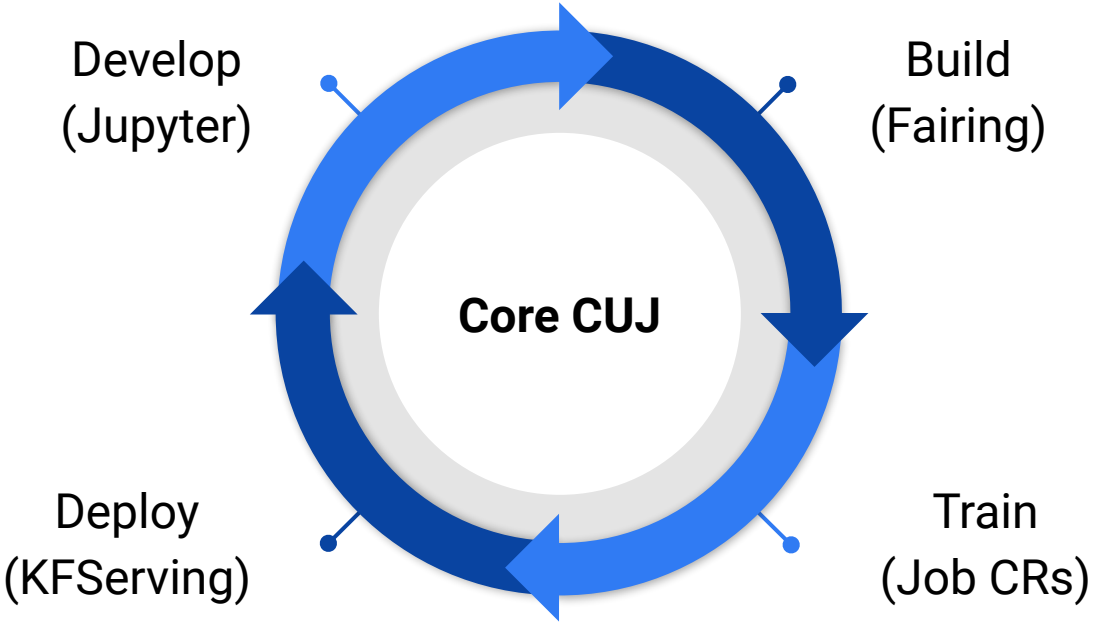
- TensorRT for notebooks

**Seldon**

- Seldon core

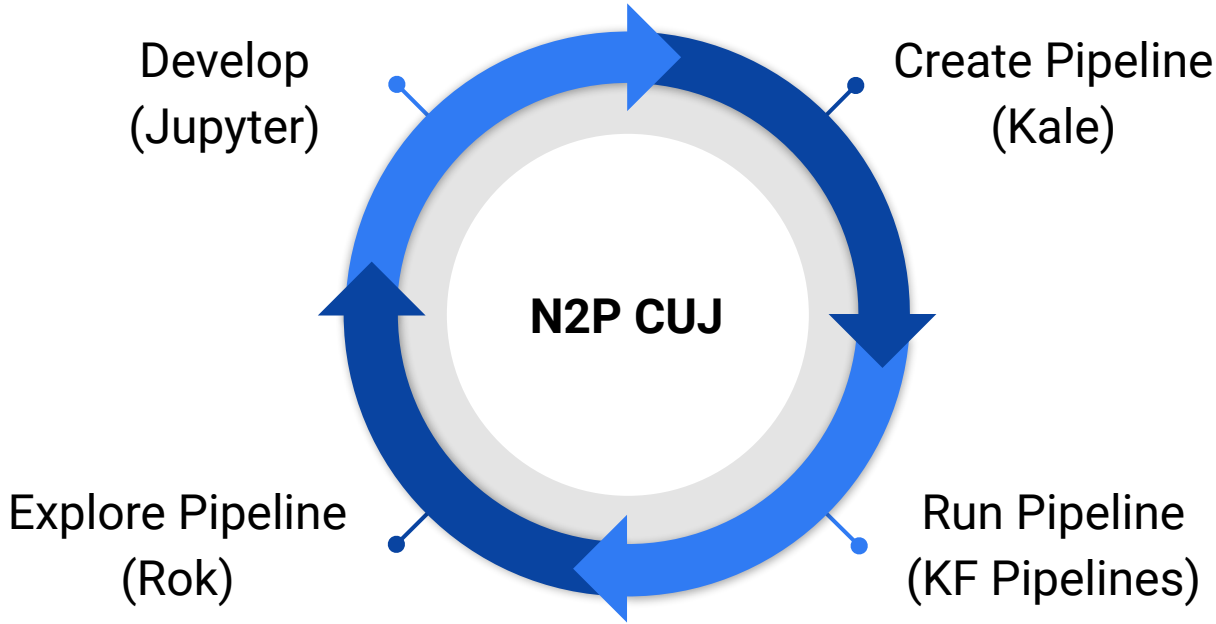
Kubeflow

# Kubeflow 1.0 Arriving January 2020



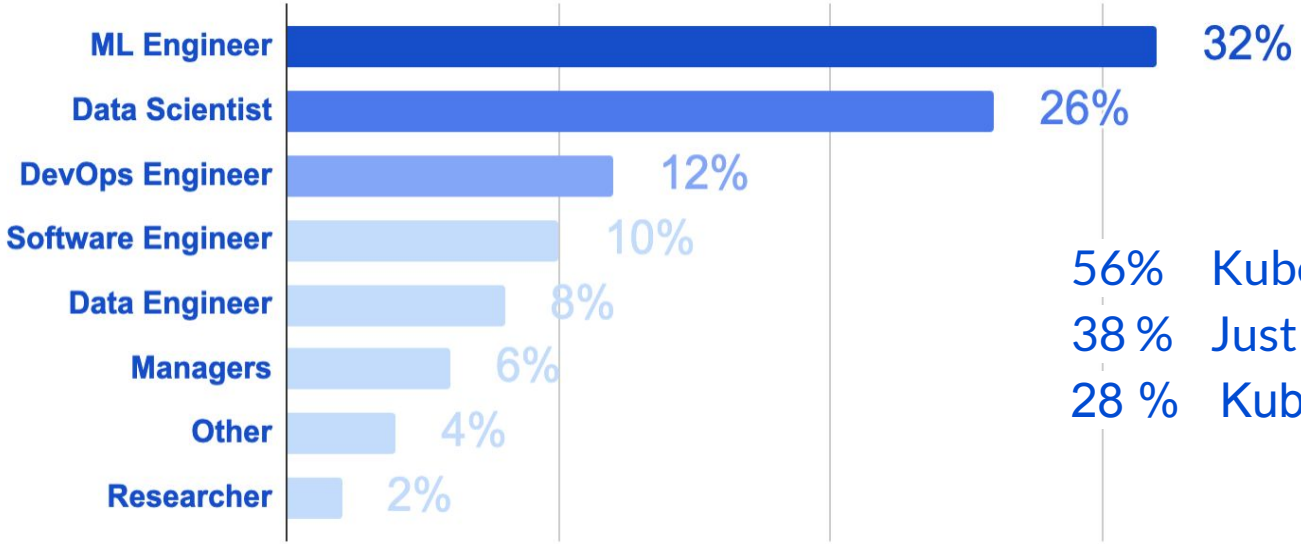

http://bit.ly/kf_roadmap

# Notebook-to-Pipeline CUJ




**Ecosystem-supported CUJ with Notebook-to-Pipeline CUJ Tutorial**

# 3Q Kubeflow User Survey - Job titles




56% Kubeflow Users

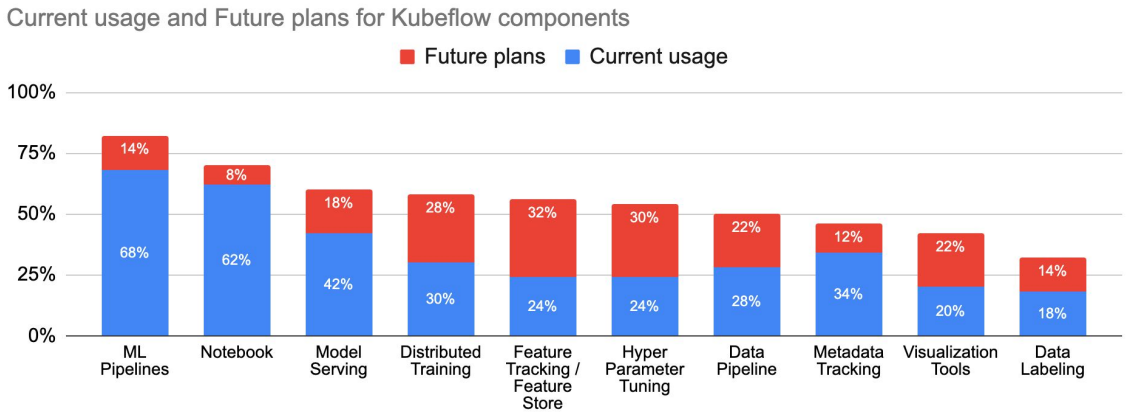
38 % Just learning

28 % Kubeflow Integrators

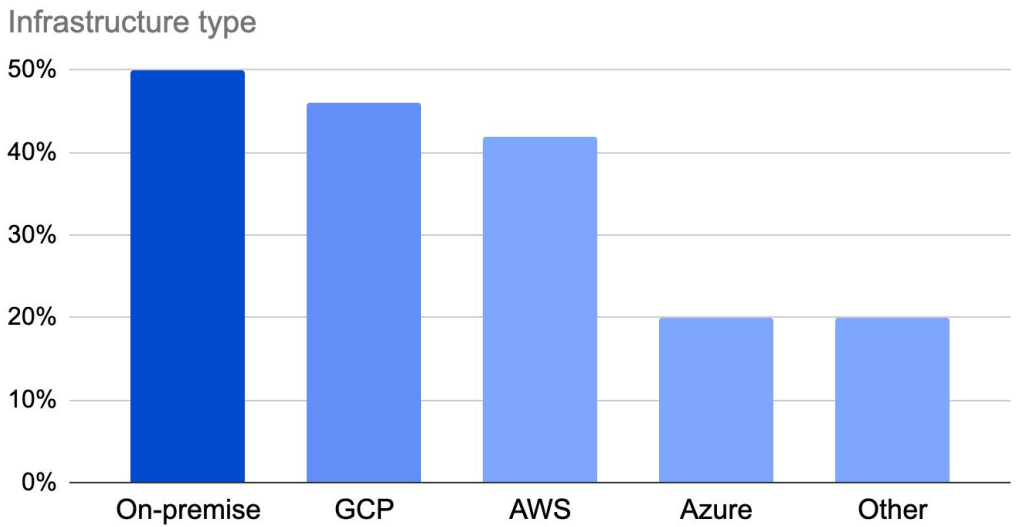
# Kubeflow Survey - Components usage

Top 5 components:

82% Pipelines
70% Notebooks
60% Model Serving
58% Distributed Training
56% Feature Tracking/Store



# Current Usage : Platforms, Processor, Components




Hardware accelerators

78% CPU

72% GPU

Istio usage

40% Deployed

30% Planning

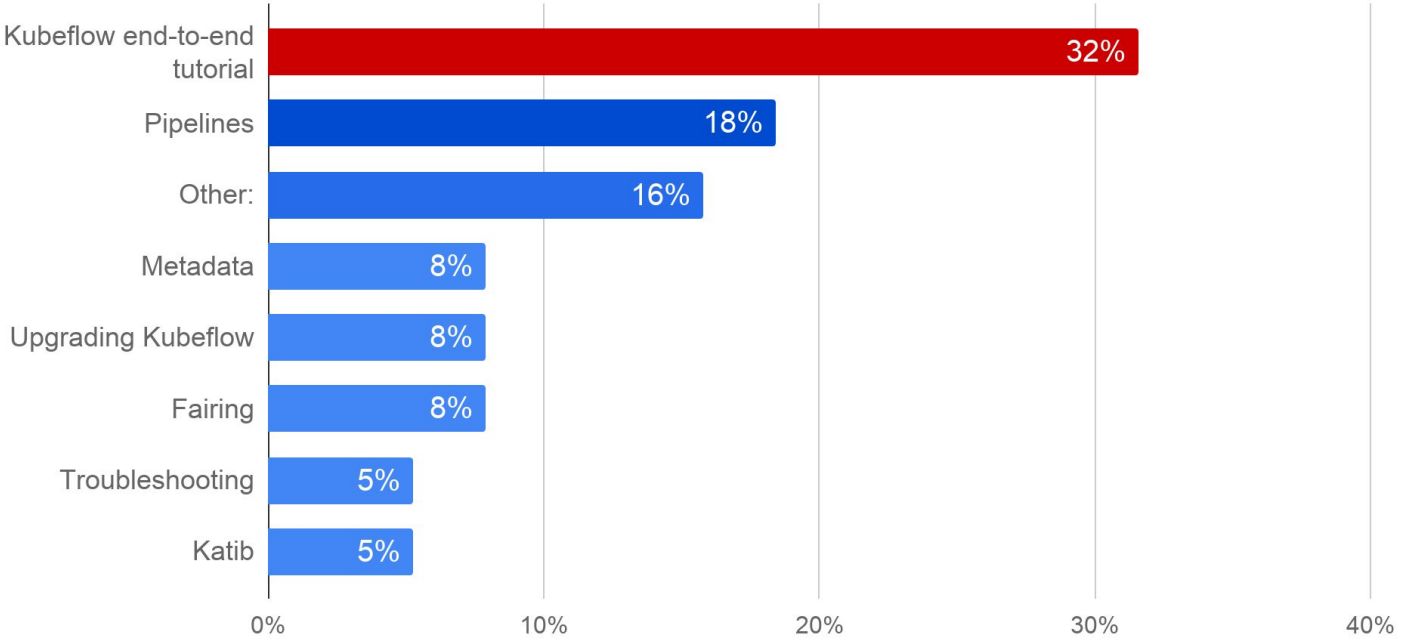
20% Not Planning

Identity providers

48% Github

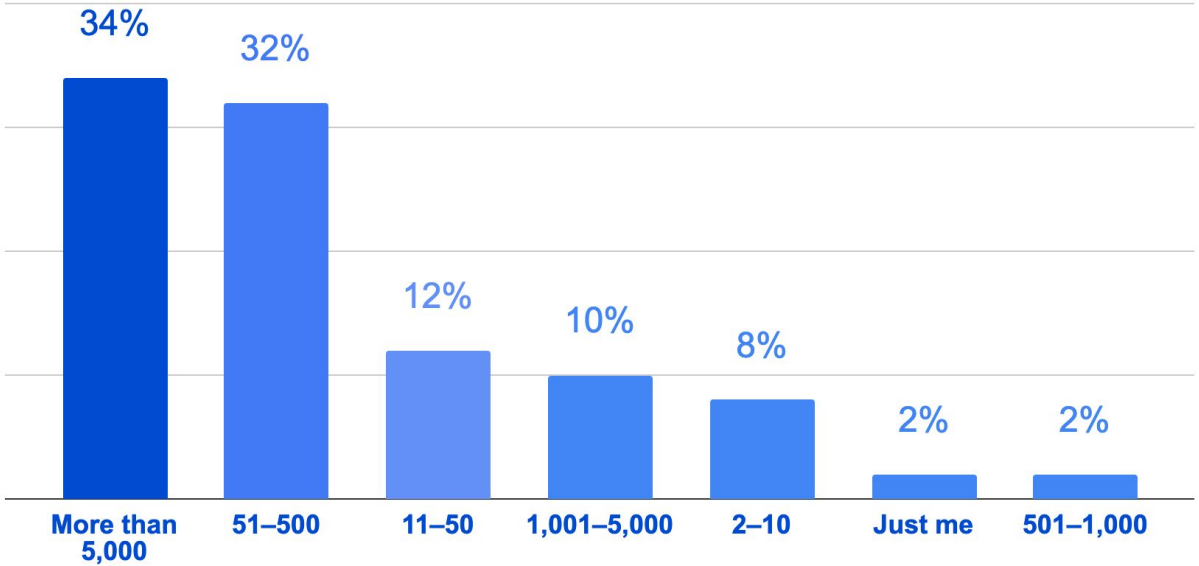
42% AD/LDAP

# Areas for Improvements




**Update:** Input needed, please take: [Survey on Training Needs]

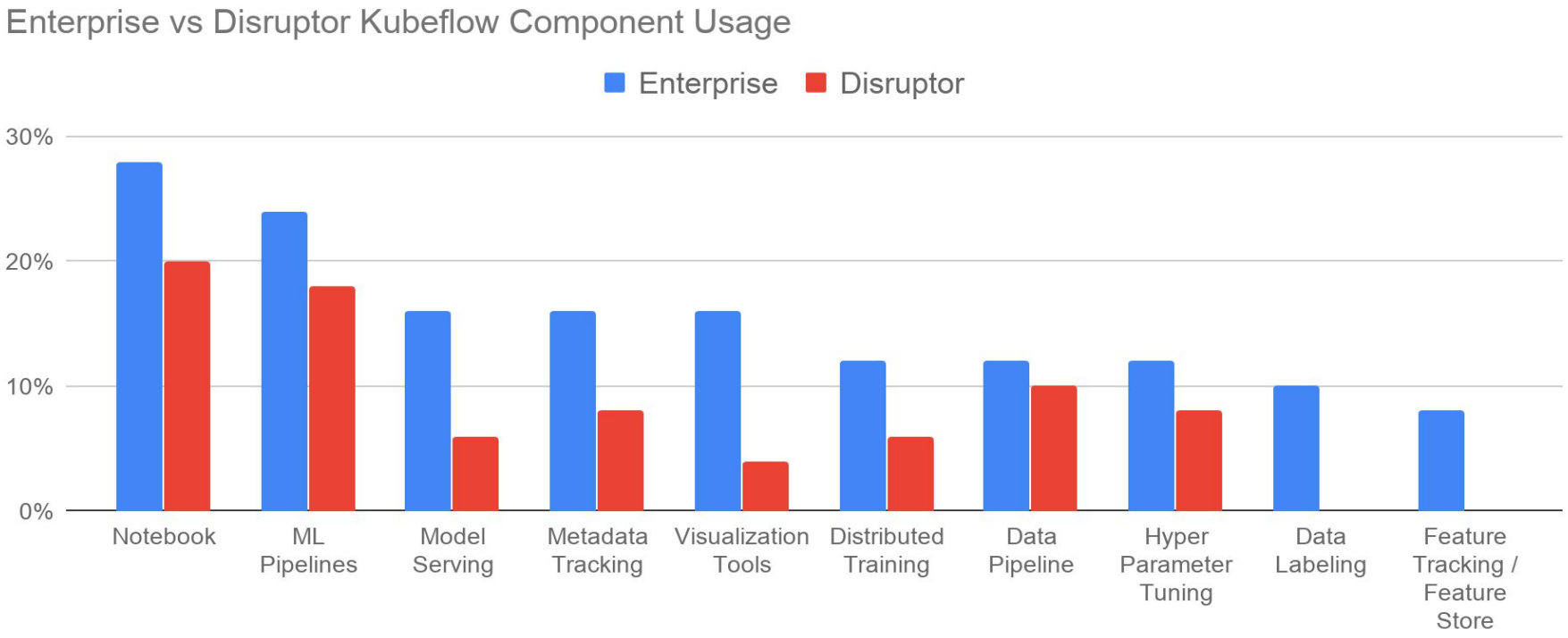
# Organization size



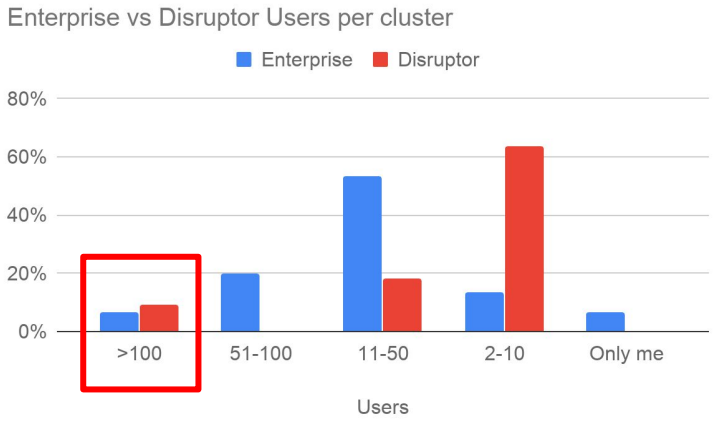
# Mixed messages & priorities

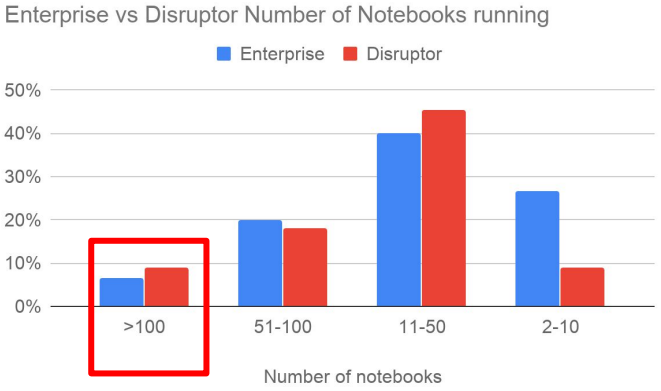
| Disruptor | Enterprise |
| --- | --- |
| 51-500 employees | 5k+ employees |
| On-prem & AWS | On-prem & GCP |
| GitHub Auth | LDAP/AD |
| RBAC | + Security,<br>Air-gapped |
| Notebooks,<br>Pipelines | + Serving,<br>Distributed Training,<br>Katib, Metadata,<br>Feature Store |

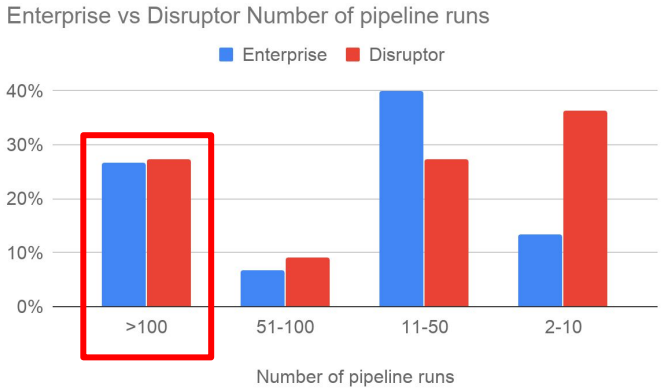
# Enterprise & Disruptor - Components in use



# Max users, notebooks, pipelines per cluster



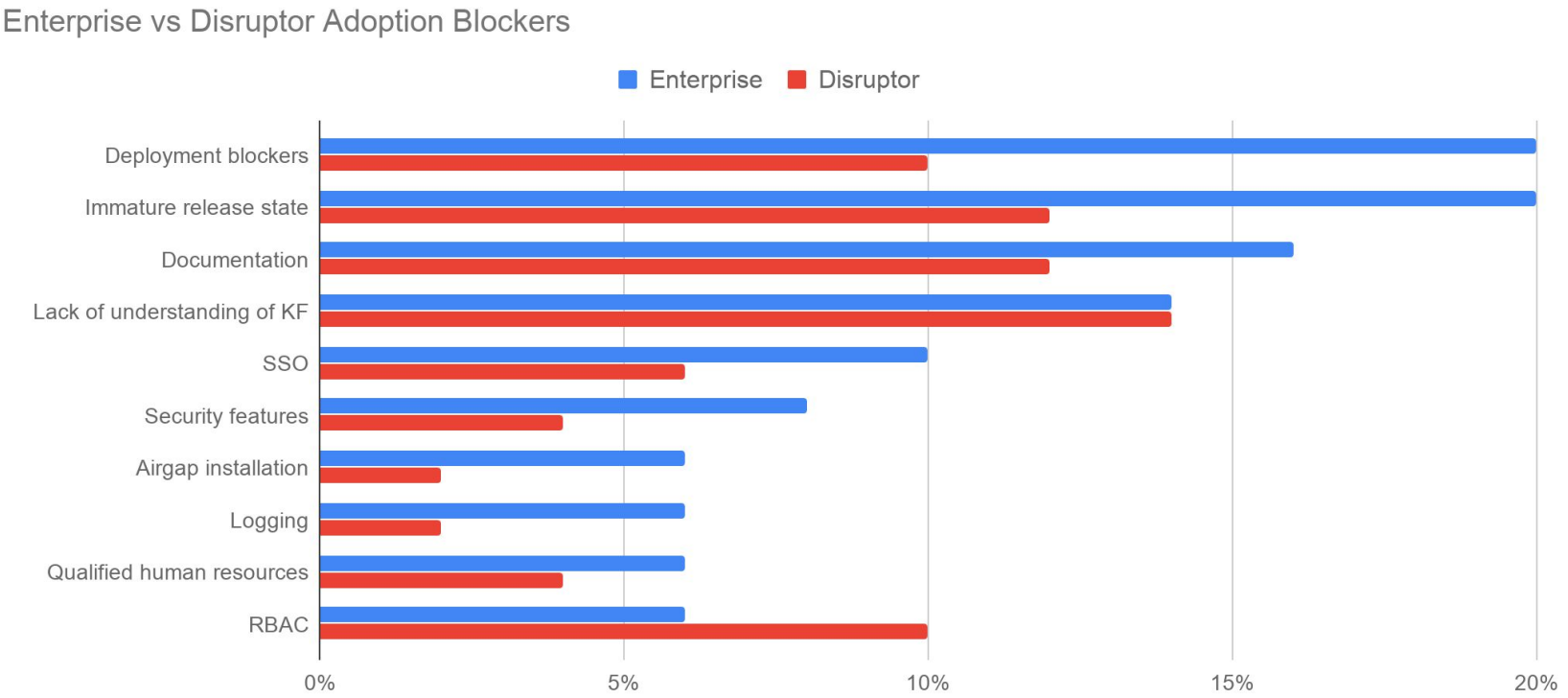




## Scale - Quick Summary

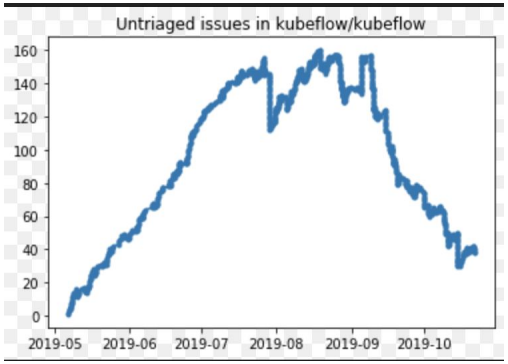
- # of Pipelines > Notebooks > Users
- 20%+ need 100+ pipelines

# Kubeflow adoption blockers



# New Issue Triage

- Prioritization...a mountain  of work ->
- Response Times
- Defining and Closing Stale issues
- Appeals




| User Survey | Recent Deliveries |
|---|---|
| Deployment Blockers | Kfctl refactoring |
| Release Maturity | 1.0 Criteria Defined / Betas |
| Improve Docs | 1.0 Design Doc, 2nd Doc Sprint |
| Lack of KF understanding | E2E Tutorials, Workshops |
| SSO, RBAC, Monitoring | Istio, Profiles, logging |

# Features v0.7 - kfctl & Notebooks

- Kfctl
  - A new simplified syntax to build and deploy Kubeflow configurations.
  - Deploy a KFDef spec, build profiles and role bindings.
  - Further integration with: kustomize, istio and knative.

```
CONFIG=https://raw.githubusercontent.com/kubeflow/manifests/v0.7-branch/kfdef/kfctl_existing_arrikto.0.7.0.yaml
kfctl apply -f ${CONFIG}
```

- Notebooks
  - 1.0 spec defined & beta implemented in 0.7
    - **Kubeflow 1.0 Application Requirements**
  - Improves operations, scalability, and isolation.
  - Added and improved the notebook images: i.e.  tensorflow 2.0 images with GPU

# Features v0.7 - Pipelines

- Metadata & uDSL
  - Introduced uDSL with TFX sample pipeline
  - Enabled metadata-driven orchestration for TFX on KFP
    - Built-in MLMD support through a deployed MLMD gRPC server
    - Automated logging of artifacts and executions in the metadata store
    - Default cache-based execution of pipeline
- UI
  - Retry button to re-run the failed pipeline from failure point
  - Metadata UI to show Artifacts & Executions
- Deployment/Perf/Scale
  - Introduced standalone KFP deployment so that it's easier to manage the deployment
  - Perf/scale improvements: more efficient API calls by PA and UI, Cloud SQL support, garbage collection on logs & argo workflows, etc

# Features v0.7 - Pipelines

- KFP DSL
  - Loop Support
    - WithItems support user defined loop over a static set of parameter values in DSL
    - WithParams support user can define a loop over a dynamic set of values, i.e. assigned at runtime or on container op.
  - Simple python component which enables a consistent and easy way to author a python component
  - File I/O simplification for platform independent data exchange
- Integration and samples
  - Airflow component for any airflow operations in KFP
  - AutoML pipeline,
  - dataflow pipeline,
  - CMLE TPU,
  - xgboost samples with GCP, TFX, and etc

# Features v0.7 - KFServing, Katib

- Katib (HP Tuning)
  - “Suggestion CR” tracks the suggestions for the corresponding algorithm service,
  - Enhanced event support with knative
  - Support for multiple DB backends.
  - Prometheus metrics for controller runtime metrics and Experiment/Trial counters.
  - New metric collector:  StdOutCollector(default), FileCollector and TFEventCollector,

- KF Serving
  - CRD for serving models on popular frameworks
  - Performant, high abstraction interfaces
  - Initial Frameworks: Tensorflow, XGBoost, ScikitLearn, PyTorch, ONNX.
  - Server Config, Networking, Health checks, Autoscaling (GPU), Scale to Zero, A/B,  Canaries

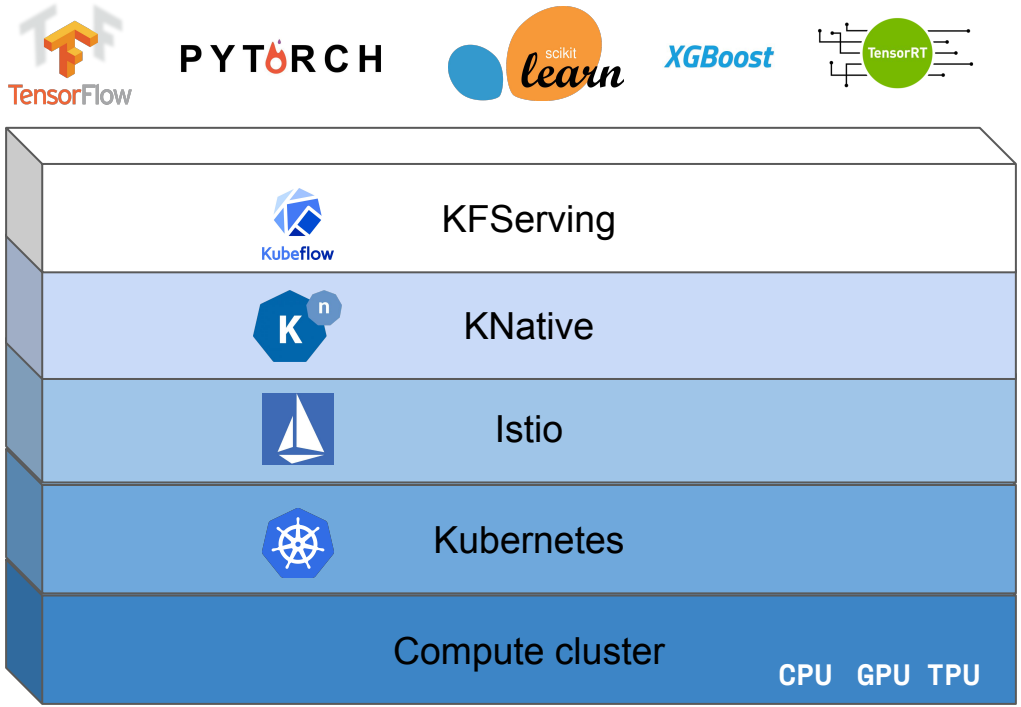
# KFServing Interface

```yaml
apiVersion: "serving.kubeflow.org/v1alpha2"
kind: "InferenceService"
metadata:
  name: "sklearn-iris"
spec:
  default:
    sklearn:
      storageUri: "gs://kfserving-samples/models/sklearn/iris"
```

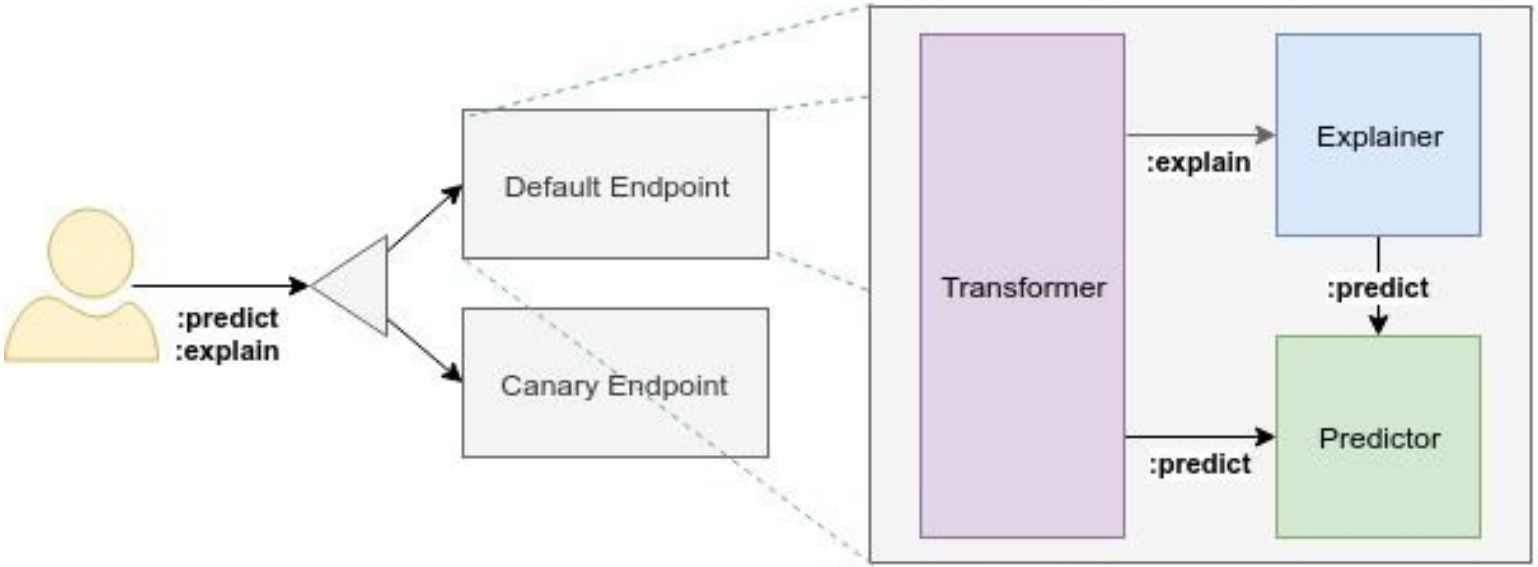

```yaml
apiVersion: "serving.kubeflow.org/v1alpha2"
kind: "InferenceService"
metadata:
  name: "flowers-sample"
spec:
  default:
    tensorflow:
      storageUri: "gs://kfserving-samples/models/tensorflow/flowers"
```


```yaml
apiVersion: "serving.kubeflow.org/v1alpha2"
kind: "InferenceService"
metadata:
  name: "pytorch-cifar10"
spec:
  default:
    pytorch:
      storageUri: "gs://kfserving-samples/models/pytorch/cifar10"
      modelClassName: "Net"
```


# Cloud Native Layers



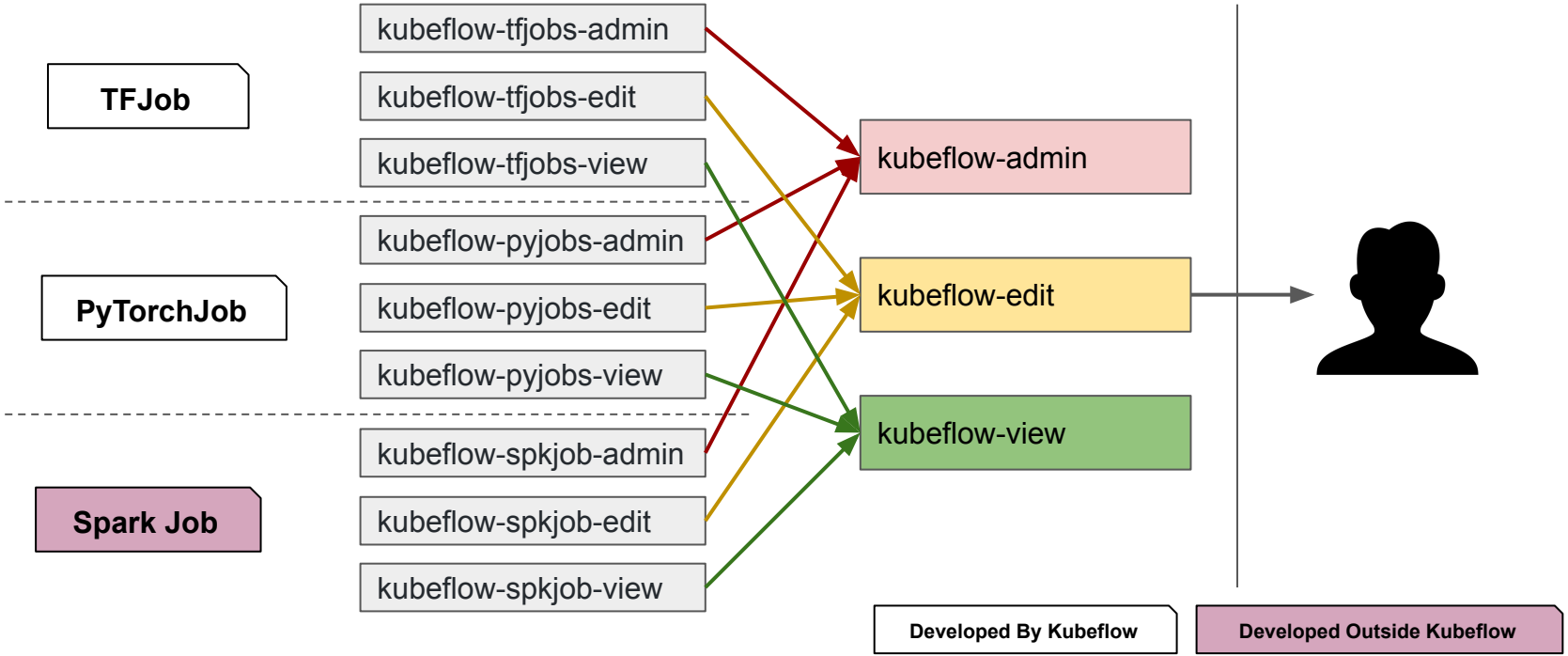

# Opinionated ML Microservices




# Features v0.7 - RBAC, Operators, Monitoring, SDKs

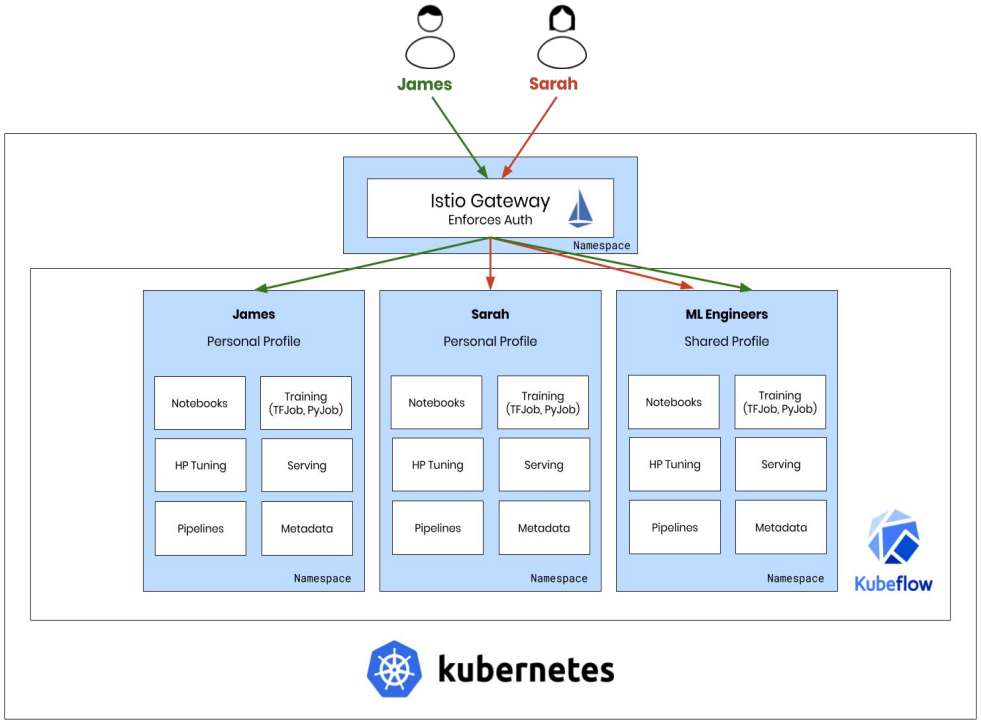
- Aggregated RBAC
  - Aggregated cluster roles grant admin, edit,  view permissions across all kubeflow apps (kubeflow-admin, kubeflow-edit or kubeflow-view)
  - Use UI or Kubectl to add grant permissions (role bindings to namespaces)
- ML Operators
  - Graduating  to 1.0 beta with kustomize manifests: TFJob, Pytorch, XGBoost, MXNet
- Monitoring
  - kubectl get applications -
  - applications resource map  K8s resources to  app
- SDKs
  - A new import functionality using kubeflow.fairing and kubeflow.medadata

# Kubeflow Aggregated ClusterRoles



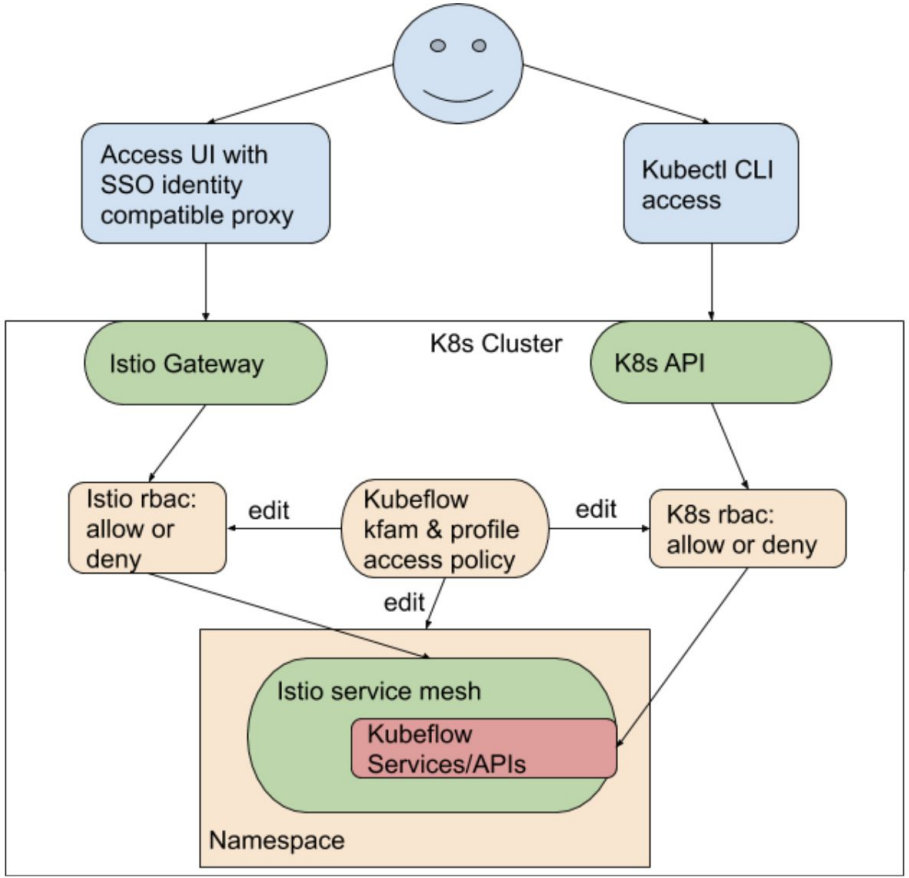
# Features v0.7 - Multi Tenancy for End Users

- Users will operate on same k8s cluster while each user has their own workspace hosting their services. Workspaces are logically isolated: each user can only access services to their own workspace.
- Fine-grained access control to public cloud APIs are supported through profile plugins: an interface for managing user-level resources that exist outside k8s cluster.



# Kubeflow Access Control

- User access through kubectl: controlled by k8s rbac policy.
- User access through browser: controlled by istio rbac policy.
- Kubeflow multi-tenancy is implemented k8s-native way, new services can be integrated easily.



# Features v0.7 - Docs

- Docs
  - Kubeflow 1.0 Design - Doc & Process
  - Kubeflow 1.0 KANBAN - Docs
  - New Python-based visualizations in Kubeflow Pipelines
  - An initial guide to KFServing
  - Metadata API reference
  - Improved deployment guides:
    - Existing k8s clusters (on-prem)
    - GCP
    - AWS
    - Azure
    - IBM
  - **Volunteer for DocSprint**

# Kubeflow - User Reviews

- “Multi-user functionality is a foundational building block, especially for on-prem environments, and we are excited to integrate this enhancement into our deployment.”

  ***Jeff Fogarty, Innovation Engineer at US Bank***

- “Data versioning, especially for Kubeflow Pipelines, allows us to snapshot datasets and recreate models quickly. This significantly simplifies workflows and improves productivity.”

  ***Laura Schornack, who is using Kubeflow as a part of the City Scholars partnership between Chase and the University of Illinois***

# Kubeflow - User Reviews

- “Kubeflow v0.7 provides valuable enhancements that simplify our ML operations, especially on hyperparameter tuning with the new suggestion CR.”

  ***Jeremie Vallee, AI Infrastructure Engineer, BabylonHealth***

- “We appreciate the Kubeflow Community’s on-going efforts to mature the project, and v0.7 enables us to deliver a wide range of ML models to our diverse customers and their analytical use cases.”

  ***Tim Kelton, Co-Founder, Descartes Labs, Inc.***

# Come Help!

- website: https://kubeflow.org
- github: https://github.com/kubeflow/kubeflow
- slack: kubeflow (http://kubeflow.slack.com)
- twitter: @kubeflow


# Thank You
# www.kubeflow.org