Kubeflow Repositories

Weiqiang Zhuang
wzhuang@us.ibm.com
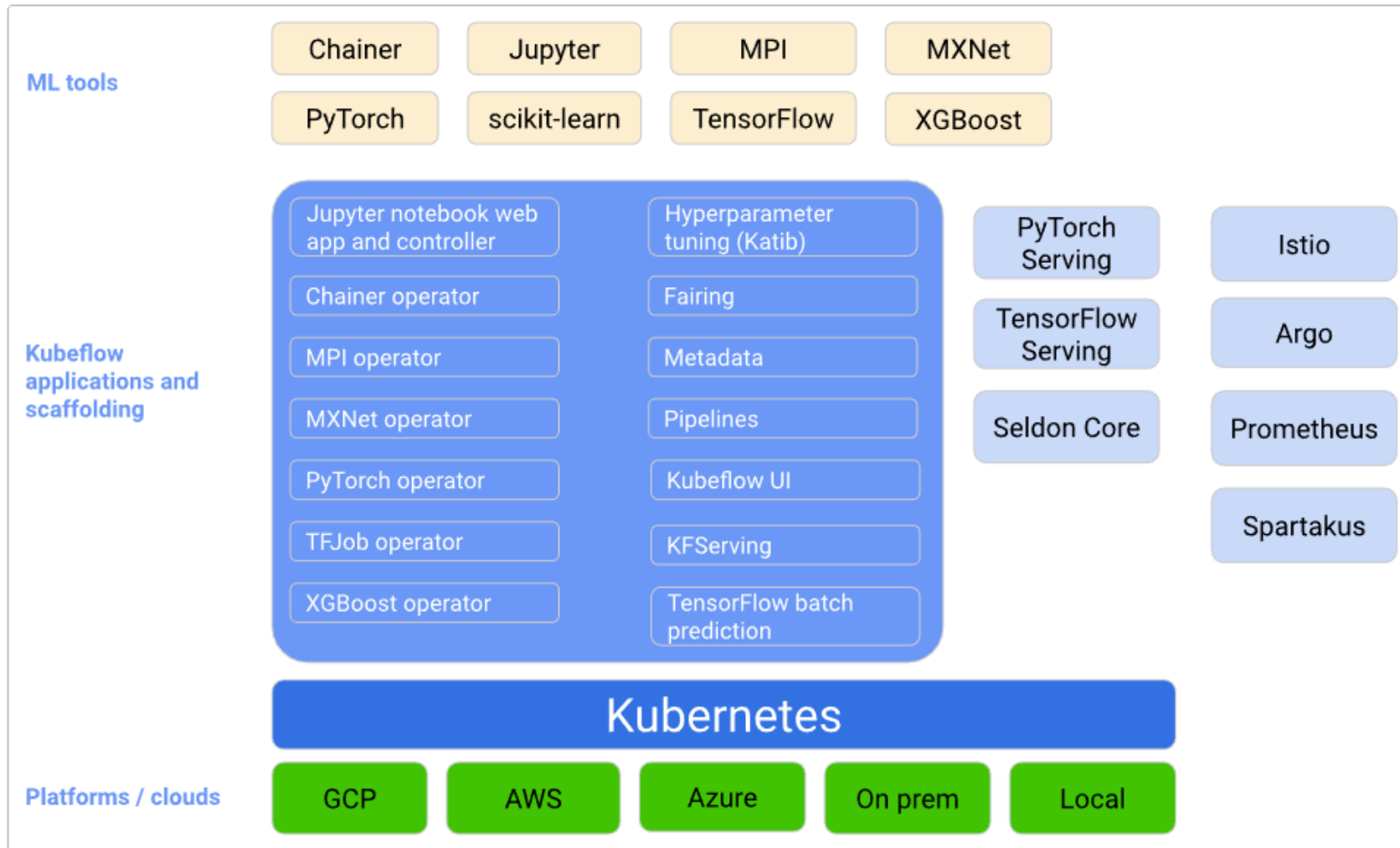IBM CODAIT
Github id: adrian555

Overview of Kubeflow Repositories

- Kubeflow is an open, community driven project to make it easy to deploy and manage an ML stack on Kubernetes
  - https://kubeflow.org
- Kubeflow org (github.com/kubeflow) is a collection of many repositories

Overview of Kubeflow Repositories

ML tools

| Chainer | Jupyter | MPI | MXNet |
| PyTorch | scikit-learn | TensorFlow | XGBoost |

Kubeflow applications and scaffolding

Jupyter notebook web app and controller	Hyperparameter tuning (Katib)
Chainer operator	Fairing
MPI operator	Metadata
MXNet operator	Pipelines
PyTorch operator	Kubeflow UI
TFJob operator	KFServing
XGBoost operator	TensorFlow batch prediction

PyTorch Serving

TensorFlow Serving

Seldon Core

Istio

Argo

Prometheus

Spartakus

Kubernetes

Platforms / clouds

| GCP | AWS | Azure | On prem | Local |

https://www.kubeflow.org/docs/started/kubeflow-overview/

Overview of Kubeflow Repositories

- **kubeflow** – machine learning toolkit for Kubernetes
  - Main repo, provides access management, central dashboard, jupyter web app, notebook controller, profile controller and more cloud native deployments and services

- **website** – Kubeflow's public website
  - Documentation

- **community** – information about the Kubeflow community including proposals and governance information

- **examples** – a repository to host extended examples and tutorials
  - end-to-end, component-focused and demos

- **code-intelligence** – ML-powered developer tools, using Kubeflow

- **manifests** – a repository for Kustomize manifests
  - Kubeflow installs with Kustomize, this repo provides KfDef manifests containing the Kustomize applications

- **kfctl** – a CLI for deploying and managing Kubeflow
  - Build, deploy and manage Kubeflow

Overview of Kubeflow Repositories

- **metadata** – repository for assets related to Metadata
  - Tracking and managing metadata of machine learning workflows
  - Includes info about executions, models, datasets and other artifacts
- **testing** – test infrastructure and tooling for Kubeflow
- **common** – common APIs and libraries shared by other Kubeflow operator repositories
  - Common controller code helpers for operators generated by kube-builder or operator-sdk
- **frontend** – repository for Kubeflow frontend

- **pipelines** – machine learning pipelines for Kubeflow
  - Reusable end-to-end ML workflows built using the Kubeflow Pipelines SDK
  - Workflow engine through Argo, Pipelines SDK and UI
- **kfp-tekton** – experimental project exploring Tekton and KFP integration
  - Kubeflow Pipelines with Tekton

Overview of Kubeflow Repositories

- tf-operator – tools for ML/Tensorflow on Kubernetes

- mpi-operator – Kubernetes operator for allreduce-style distributed training

- mxnet-operator – a Kubernetes operator for mxnet jobs

- pytorch-operator – tools for ML/Tensorflow on Kubernetes

- katib – repository for hyperparameter tuning

- fairing – Python SDK for building, training, and deploying ML models

- kfserving – serverless inferencing on Kubernetes

- Others