

DP06

KUBEFLOW

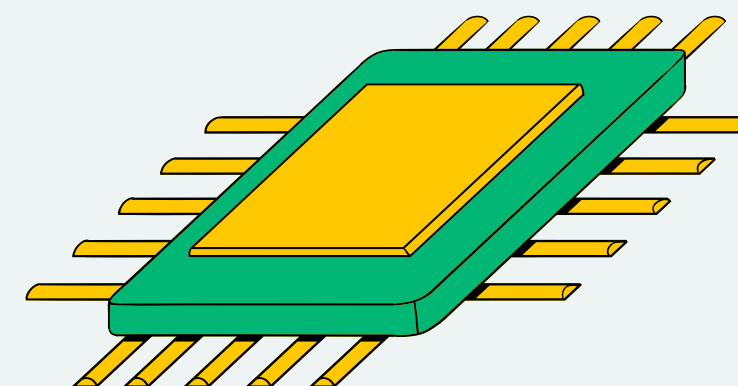
21KHMT2

Cao Hoài Yến Vy – 21127205

Trần Thanh Ngân – 21127115

Âu Dương Khang – 21127621

Nguyễn Bùi Mẫn Nhi – 21127662





PRESENTATION OUTLINE

- Machine Learning Landscape
- Kubeflow
- Advantages
- Components
- Comparison with similar solutions
- Demo



MACHINE LEARNING LANDSCAPE 2013 - 2017

- **Challenges:**

- **Lack of great notebooks** for data exploration.
- **Difficulty embedding** libraries in notebook frameworks.
- Python version **compatibility issues**.
- Complex installation of ML frameworks like TensorFlow, Torch, etc.

- **Transition:**

- Google's donation of **TensorFlow** to the community.
- Experimentation limited due to **installation complexities**.

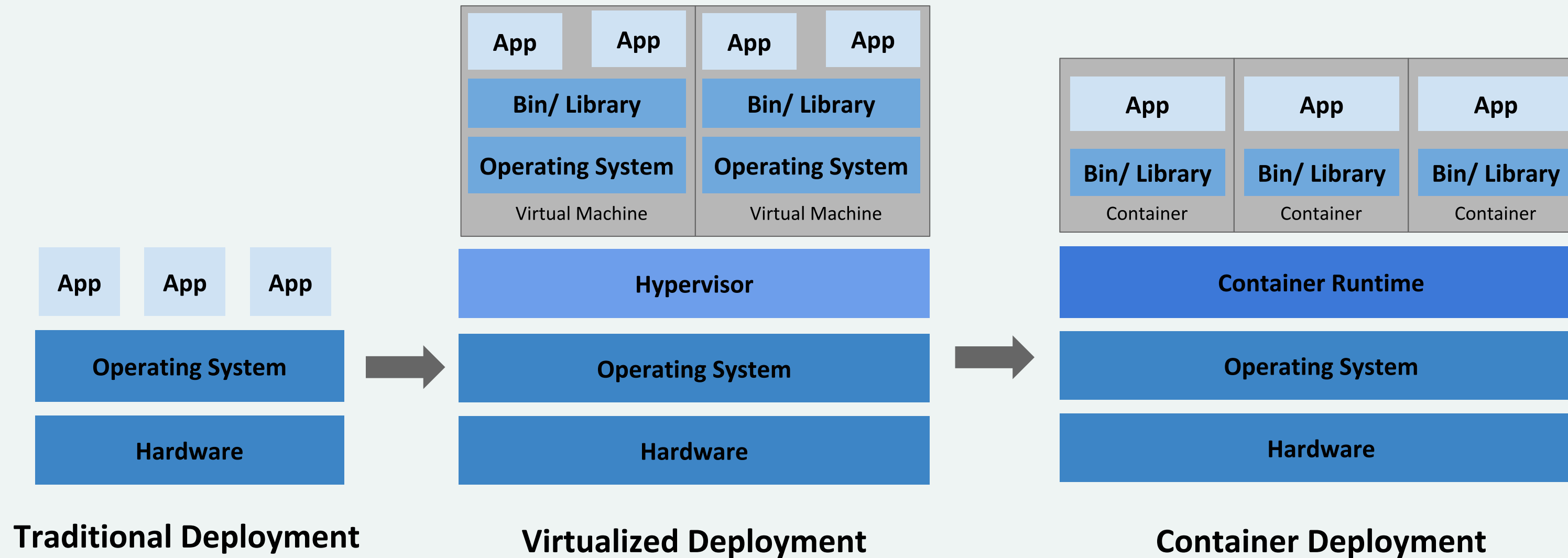


WHERE ARE WE NOW? AND WHAT IS ABOUT TO **CHANGE**?

- Containers ubiquitous.
- **Kubernetes** powerful orchestration engine.
- Spark natively supports Kubernetes.
- Kubeflow **simplifies ML environment** deployment on Kubernetes.
- Overall, **significant improvement** for IT departments supporting data scientists.



KUBERNETES



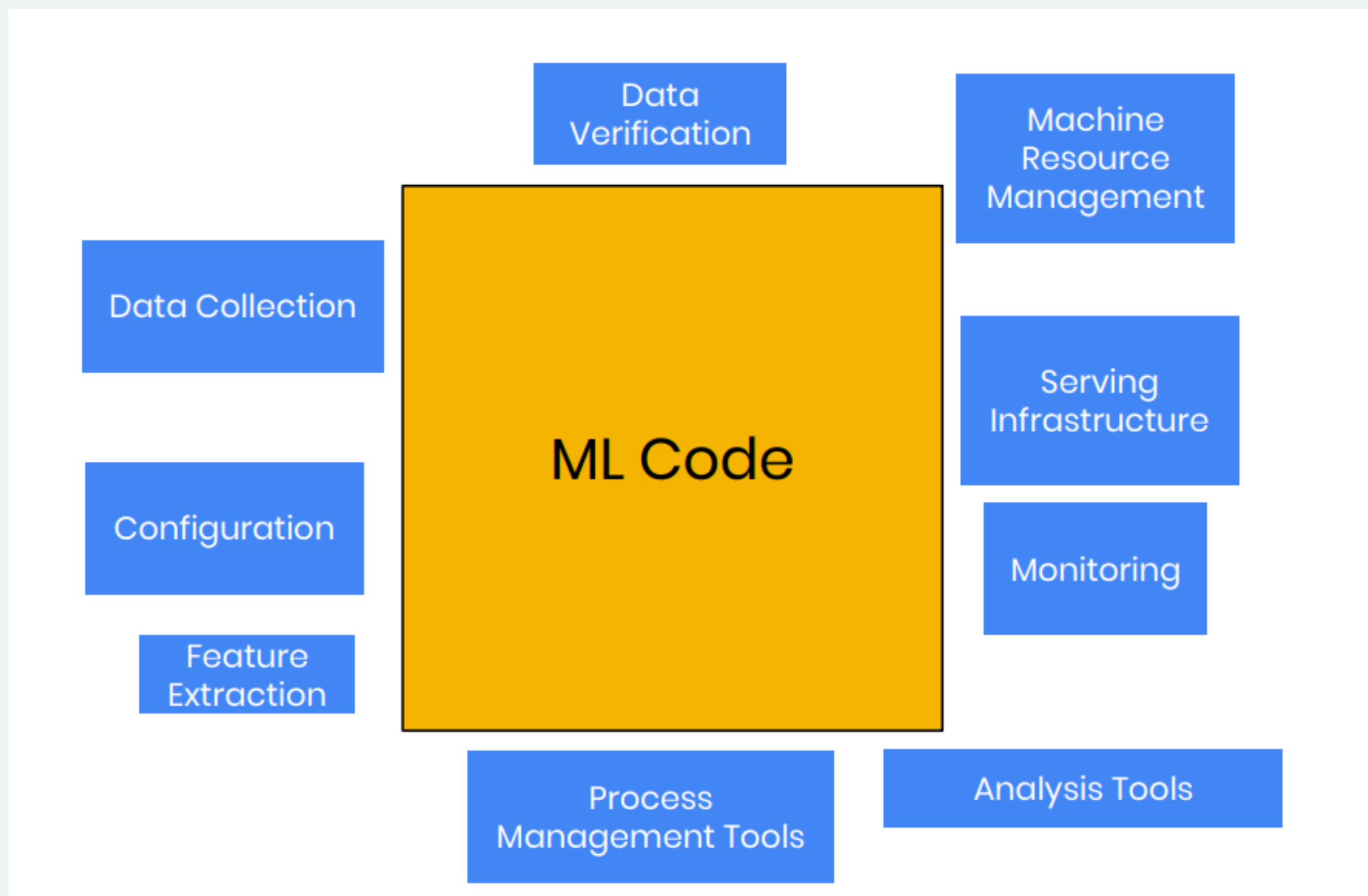
WHAT MAKES ML/AI STILL SO **CHALLENGING**?

Data Scientists are the **worst IT customers**.

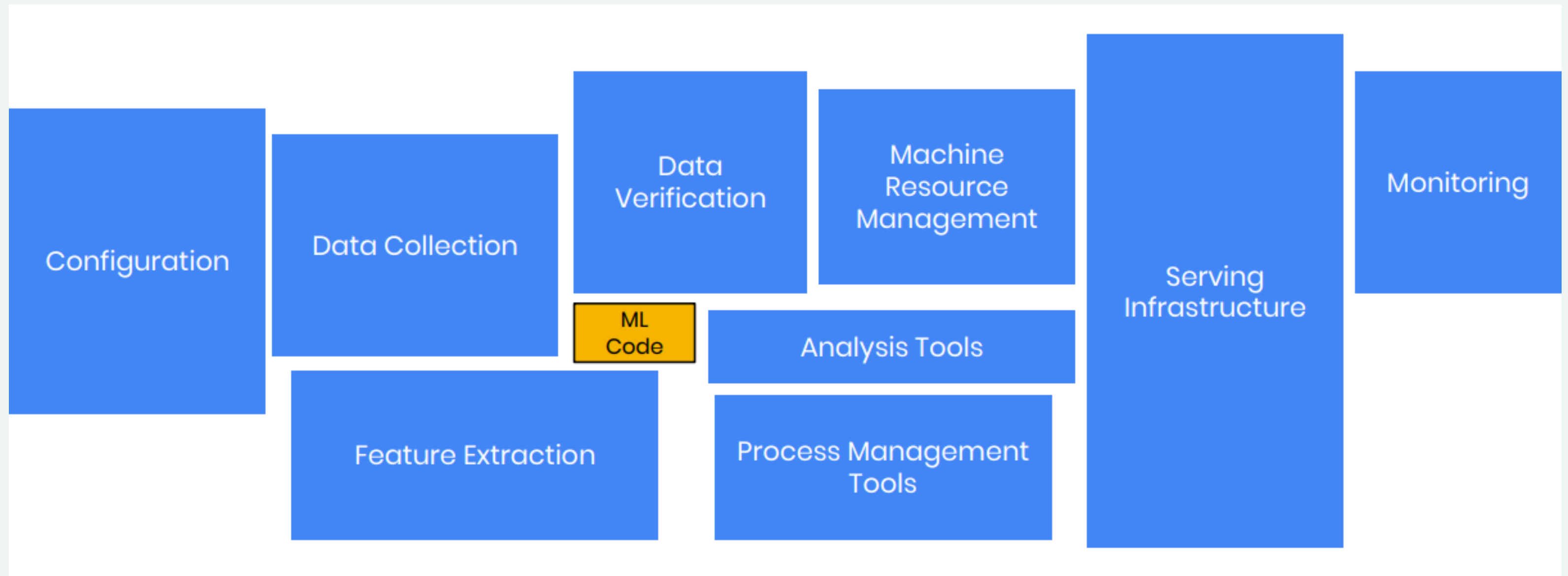
- Setting up an **ML stack/pipeline** is **incredibly hard**
- Setting up a production ML stack/pipeline is **even harder**
- Setting up an ML stack/pipeline that works across the 81% of enterprises that use multi-cloud* environments is **EVEN HARDER**



PERCEPTION: ML PRODUCTS ARE **MOSTLY ABOUT ML**



REALITY: ML REQUIRES **DEVOPS**; LOTS OF IT



WHY



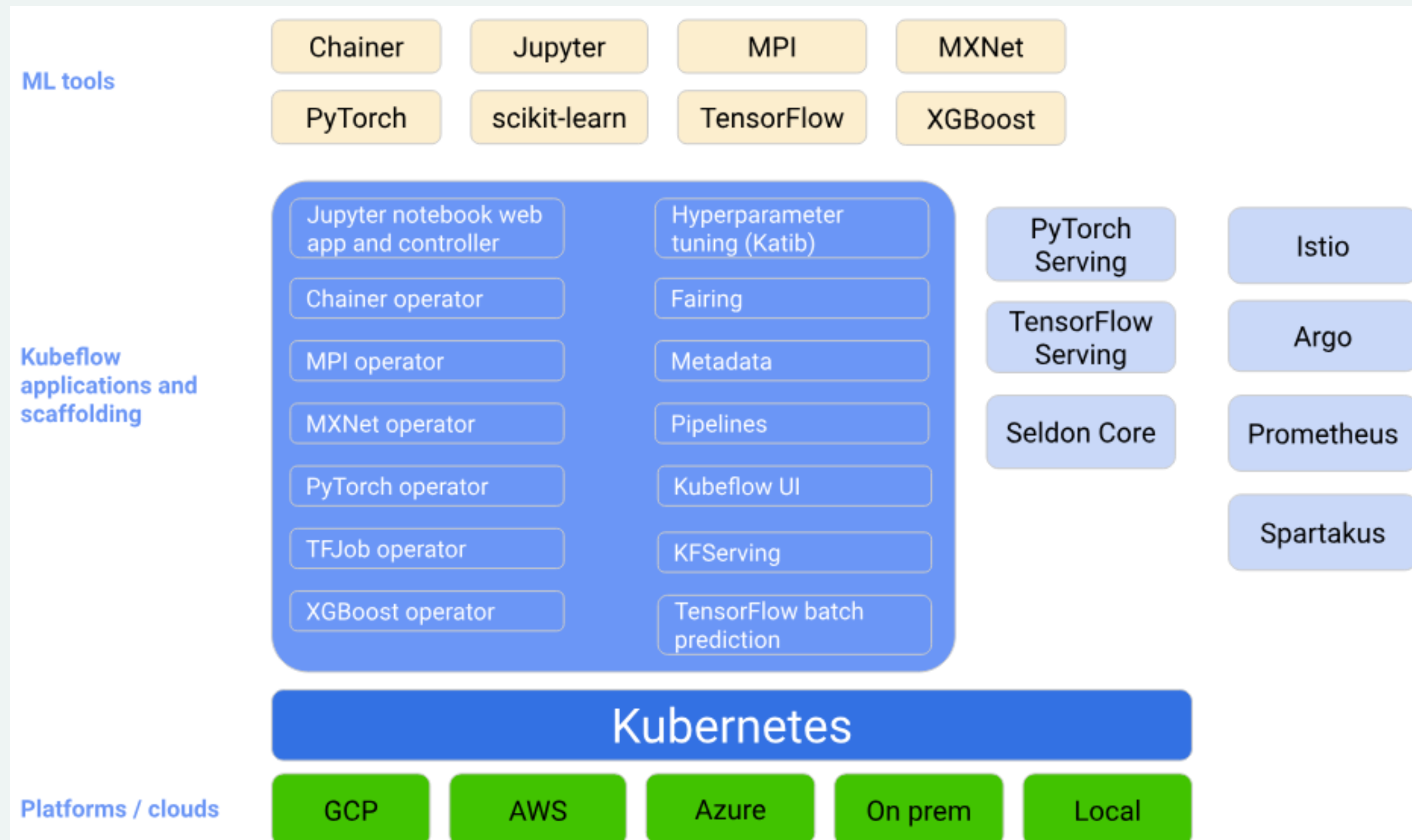
?

Kubeflow

- **End-to-end solution** for ML on Kubernetes
- **Containerized workload**
- **Experiment exploration** with **state-of-art AI** technologies
- Easy **on-boarding**
- Outstanding community and industry support



KUBEFLOW IS THE ML TOOLKIT FOR KUBERNETES.

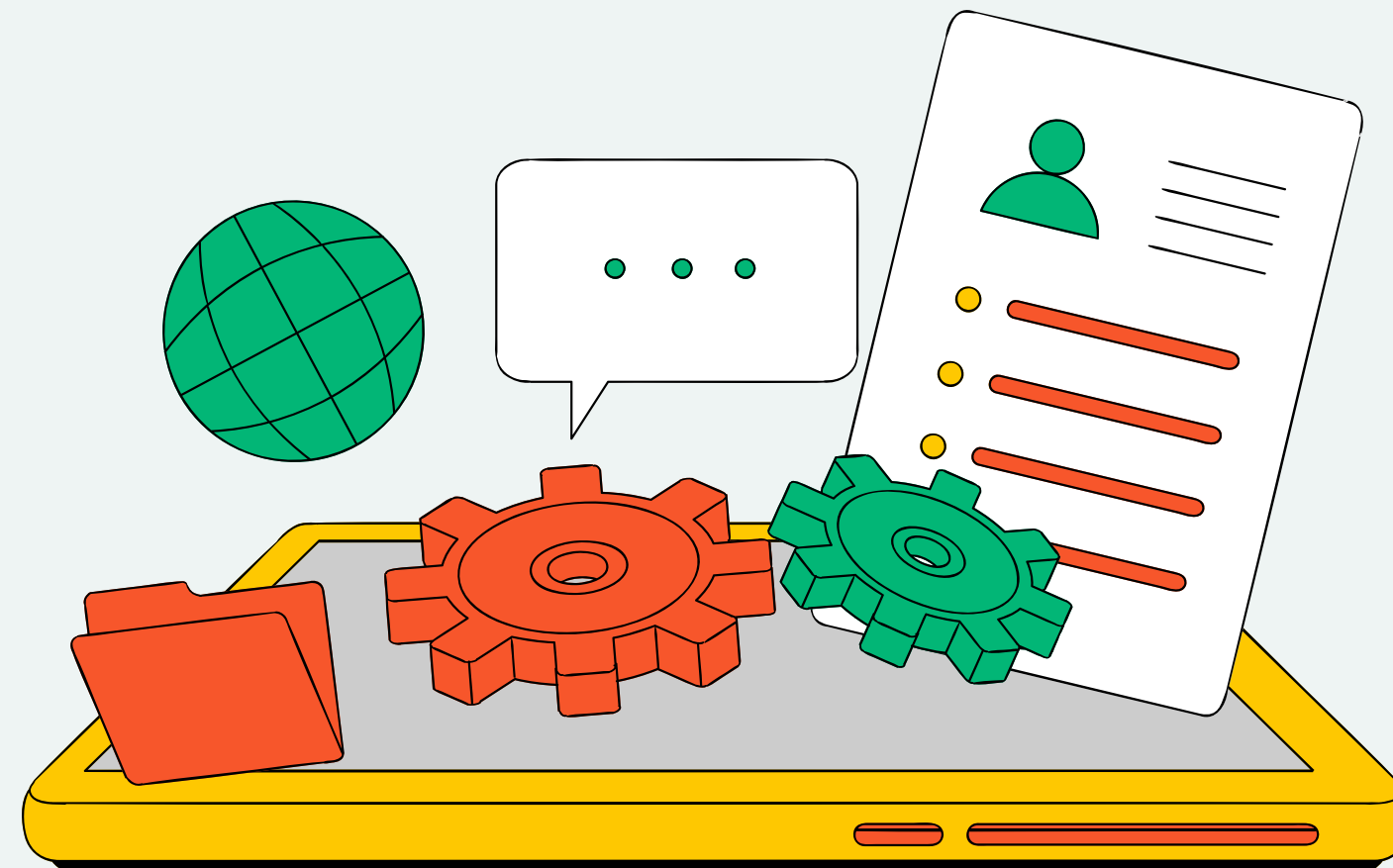


DEPLOYMENT ON CLOUD

Google Cloud Platform (GCP)

Amazon Web Services (AWS)

Azure Cloud Platform



DEPLOY LOCALLY



Minikube

deployKF

MiniKF

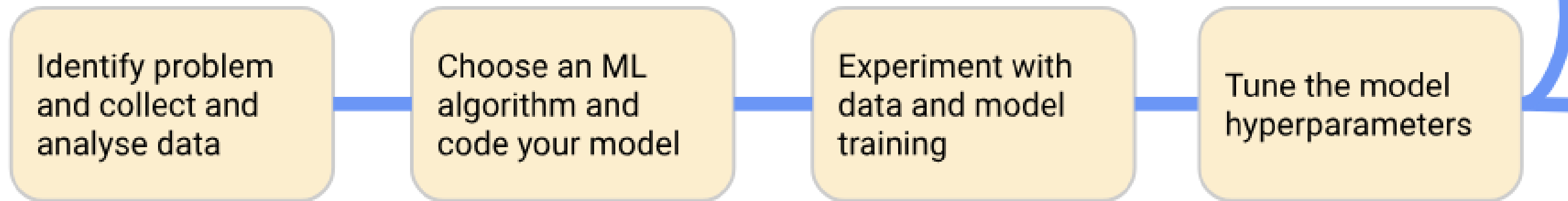
Kubernetes on Docker Desktop

Local VMs or Bare Metal

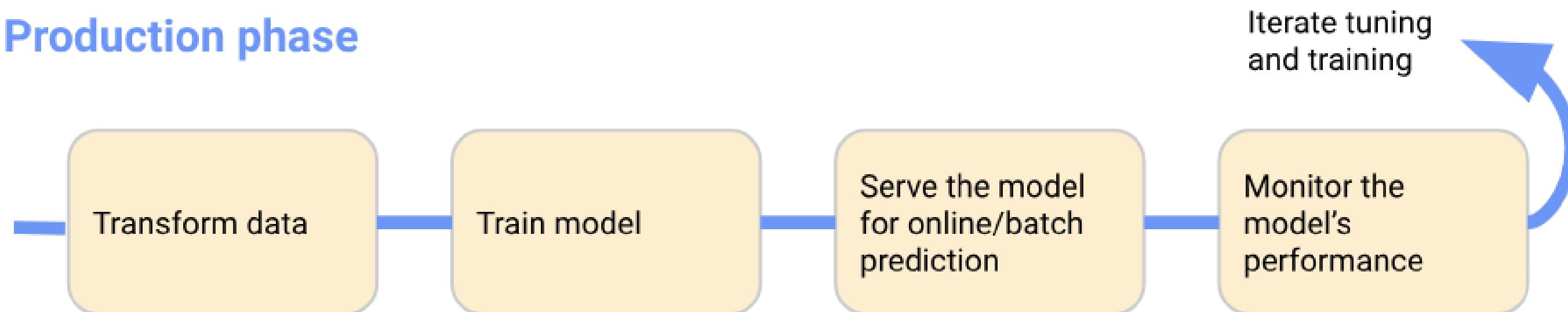


INTRODUCING THE ML WORKFLOW

Experimental phase



Production phase



KUBEFLOW COMPONENTS IN THE ML WORKFLOW

Experimental phase

with Kubeflow

Identify problem
and collect and
analyse data

Choose an ML
algorithm and
code your model

Experiment with
data and model
training

Tune the model
hyperparameters

Iterate tuning
and training

PyTorch

scikit-learn

TensorFlow

XGBoost

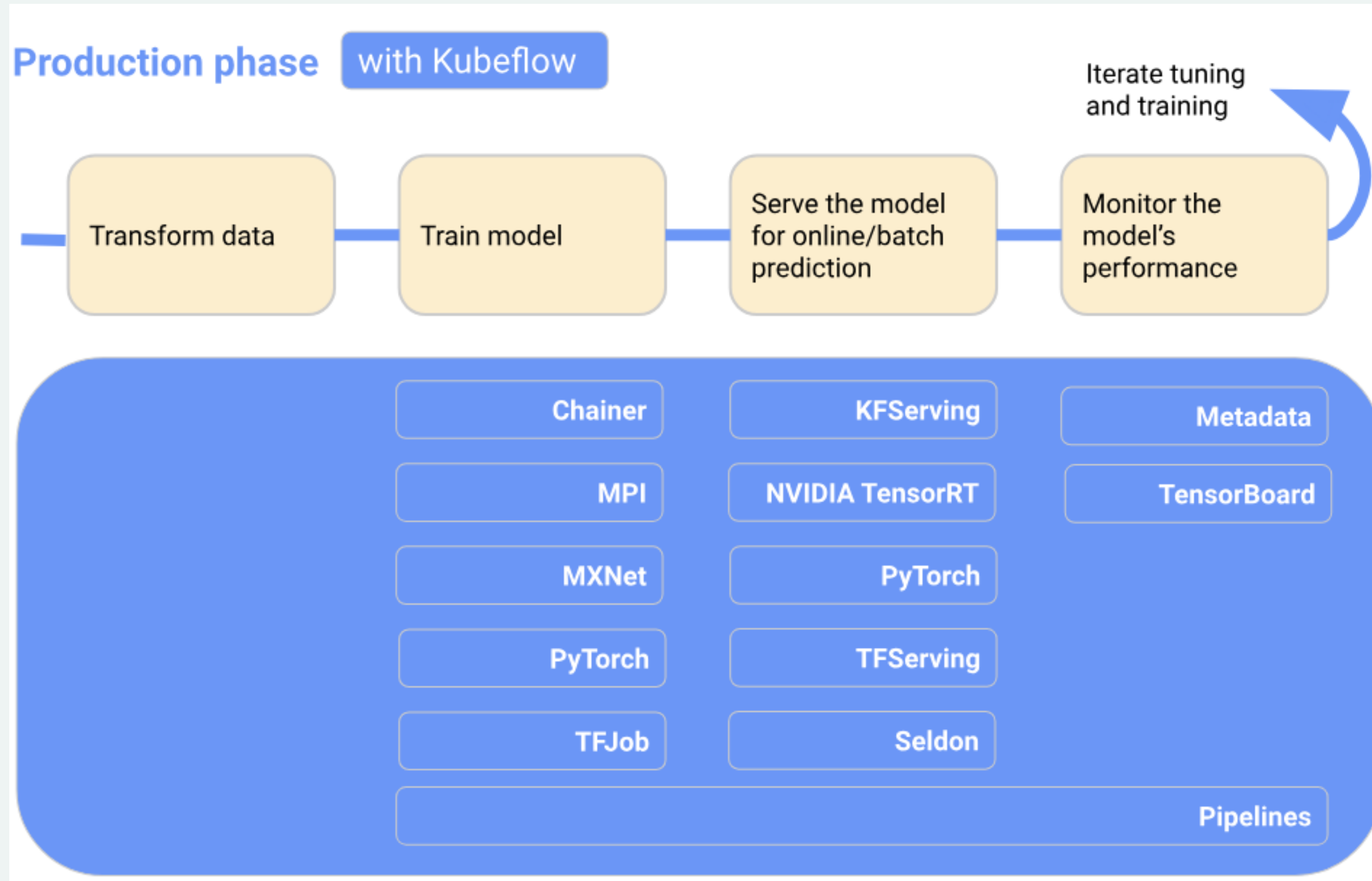
Jupyter Notebook

Fairing

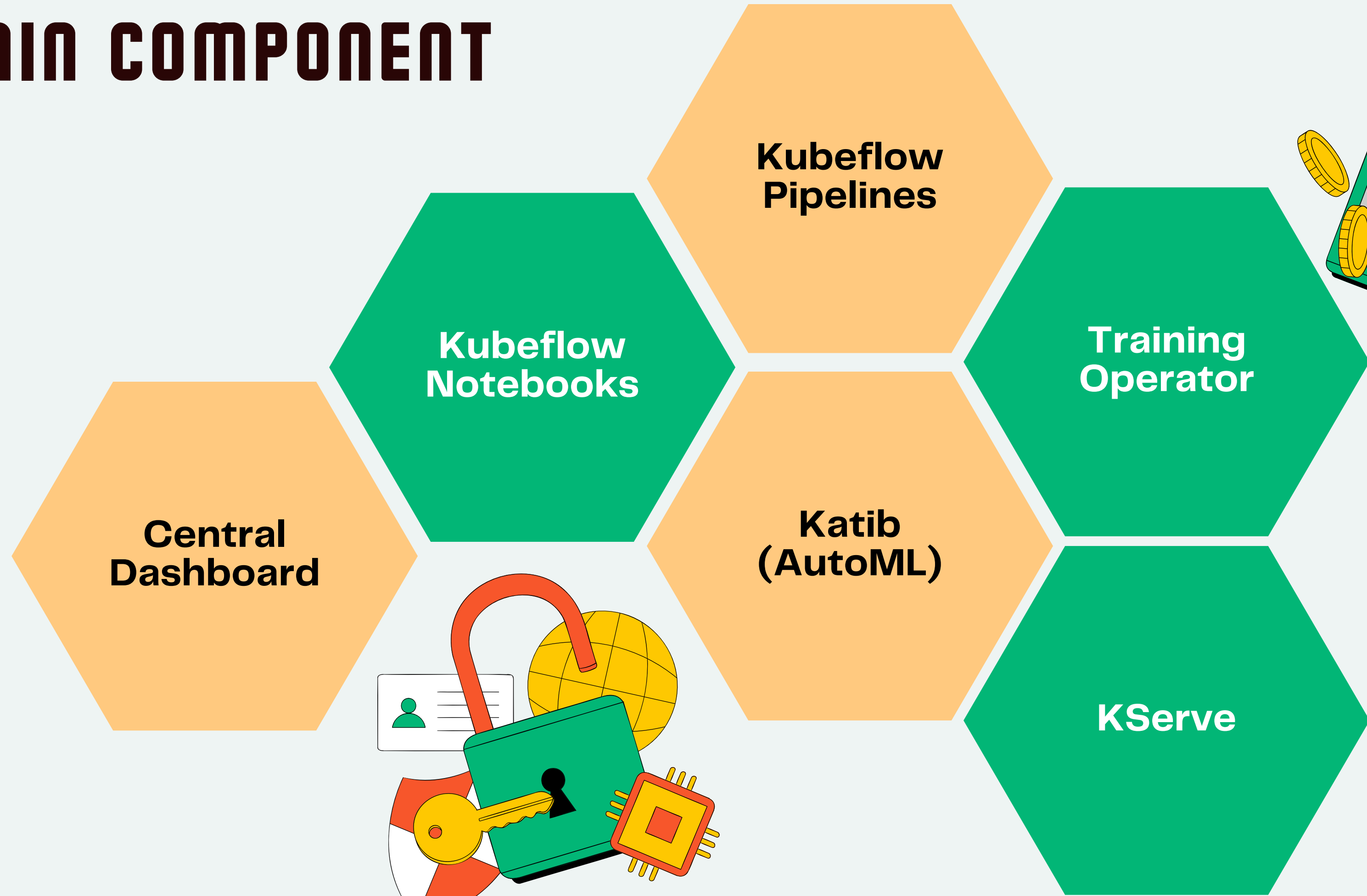
Pipelines

Katib


KUBEFLOW COMPONENTS IN THE ML WORKFLOW



MAIN COMPONENT



CENTRAL DASHBOARD

 Kubeflow

Home

Notebooks

Tensorboards

Models

Volumes

Experiments (AutoML)

Experiments (KFP)

Pipelines

Runs

Recurring Runs

Artifacts


Privacy • Usage Reporting
build version dev_local


kubeflow-user (Owner) ▾


Dashboard


Activity

Quick shortcuts


 Upload a pipeline
Pipelines


 View all pipeline runs
Pipelines

 Create a new Notebook server
Notebook Servers


 View Katib Experiments
Katib


Recent Notebooks


 kale.log
Accessed 10/12/2021, 2:06:43 PM


 lost+found
Accessed 10/12/2021, 2:06:00 PM


Recent Pipelines

 open-vaccine-model
Created 5/6/2021, 12:32:25 PM

 [Tutorial] DSL - Control structures
Created 5/6/2021, 1:42:51 AM

 [Tutorial] Data passing in python components
Created 5/6/2021, 1:42:49 AM


 [Demo] TFX - Taxi tip prediction model trainer
Created 5/6/2021, 1:42:48 AM

 [Demo] XGBoost - Iterative model training

Documentation


Getting Started with Kubeflow

Get your machine-learning workflow up and running on Kubeflow




Minikf

A fast and easy way to deploy Kubeflow locally




Microk8s for Kubeflow

Quickly get Kubeflow running locally on native hypervisors




Minikube for Kubeflow

Quickly get Kubeflow running locally




Kubeflow on GCP

Running Kubeflow on Kubernetes Engine and Google Cloud Platform




Kubeflow on AWS

Running Kubeflow on Elastic Container Service and Amazon Web Services



Requirements for Kubeflow

Get more detailed information about using Kubeflow and its components



KUBEFLOW NOTEBOOKS

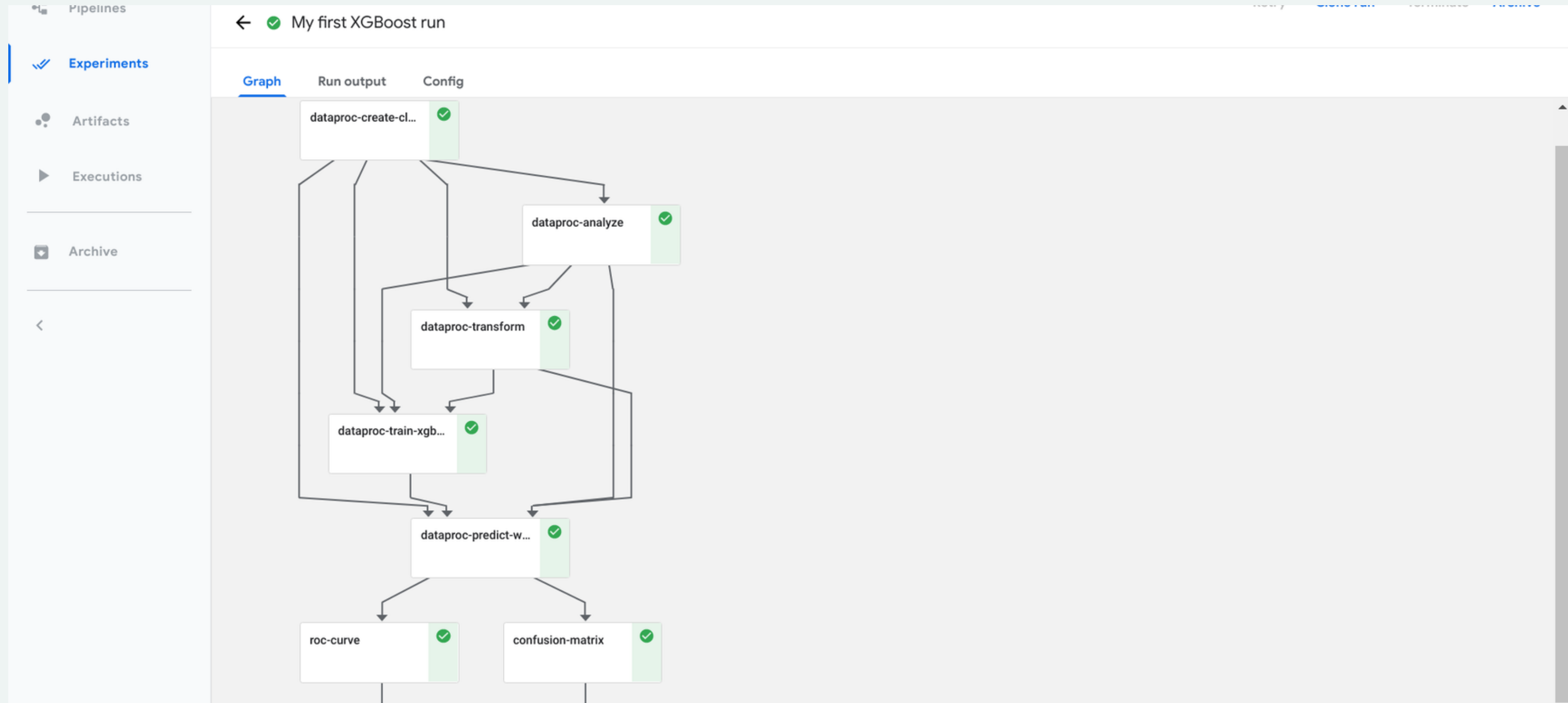
kubeflow-user (Owner)

Notebook Servers

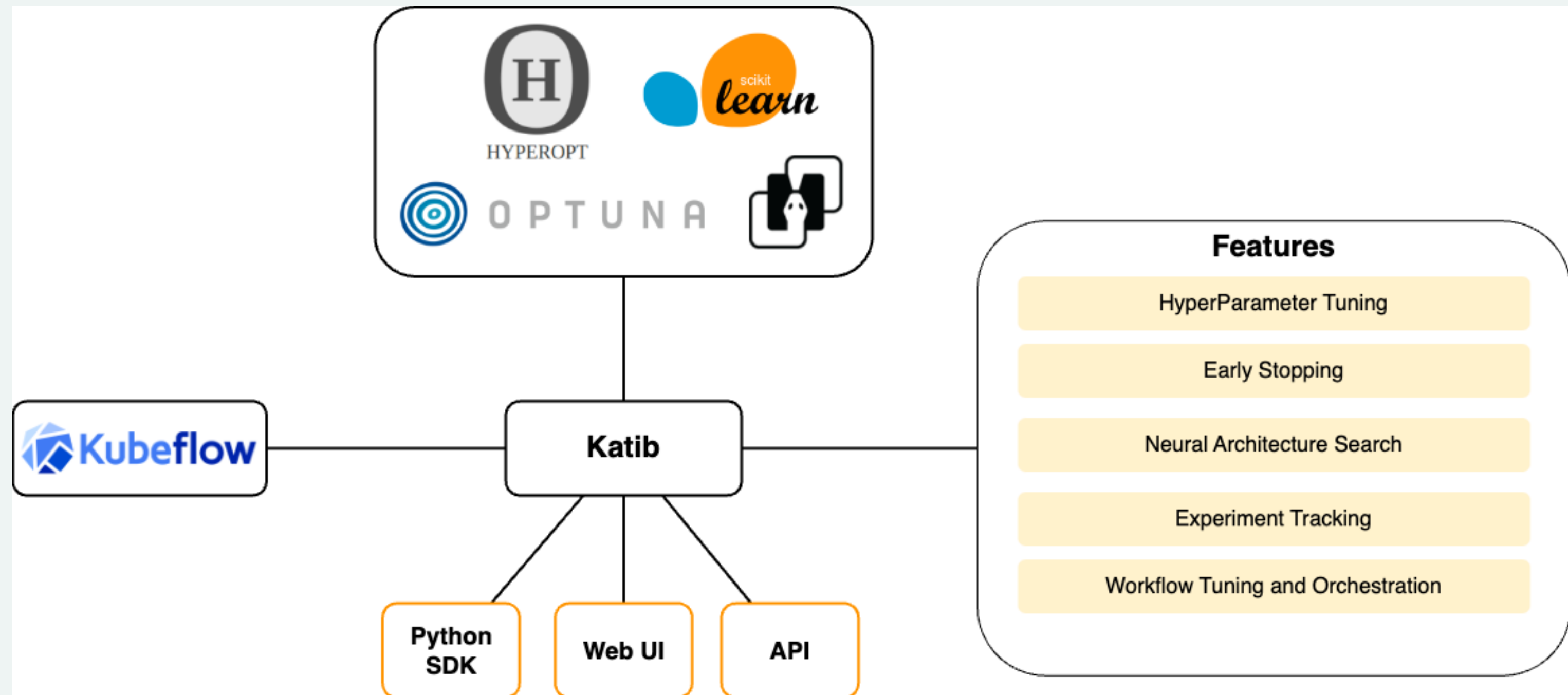
+ NEW SERVER

Status	Name	Type	Age	Image	GPUs	CPUs	Memory	Volumes			
	demo-35		42 days ago	jupyter-kale-py36:develop-l0-release-1.2-pre-29...	0	0.5	1Gi		CONNECT		
	dogbreed2-example		6 days ago	jupyter-kale-py36:develop-l0-release-1.2-pre-29...	0	0.5	1Gi		CONNECT		
	open-vaccine-1		42 days ago	jupyter-kale-py36:develop-l0-release-1.2-pre-29...	0	0.5	1Gi		CONNECT		
	open-vaccine-2		42 days ago	jupyter-kale-py36:develop-l0-release-1.2-pre-29...	0	0.5	1Gi		CONNECT		
	serve-best-open-vax-2		42 days ago	jupyter-kale-py36@sha256:5c30d30c0459b0d...	0	0.001	0.001Gi		CONNECT		
	titanic-example		6 days ago	jupyter-kale-py36:kubecon21eu-automl-nightly	0	0.5	1Gi		CONNECT		


KUBEFLOW PIPELINES



KATIB (AUTOML)



KATIB (AUTOML)

 Kubeflow

Home

Notebooks

Tensorboards

Models



Volumes


Experiments (AutoML)

Experiments (KFP)


Pipelines







Runs

 kubeflow-user (Owner) 

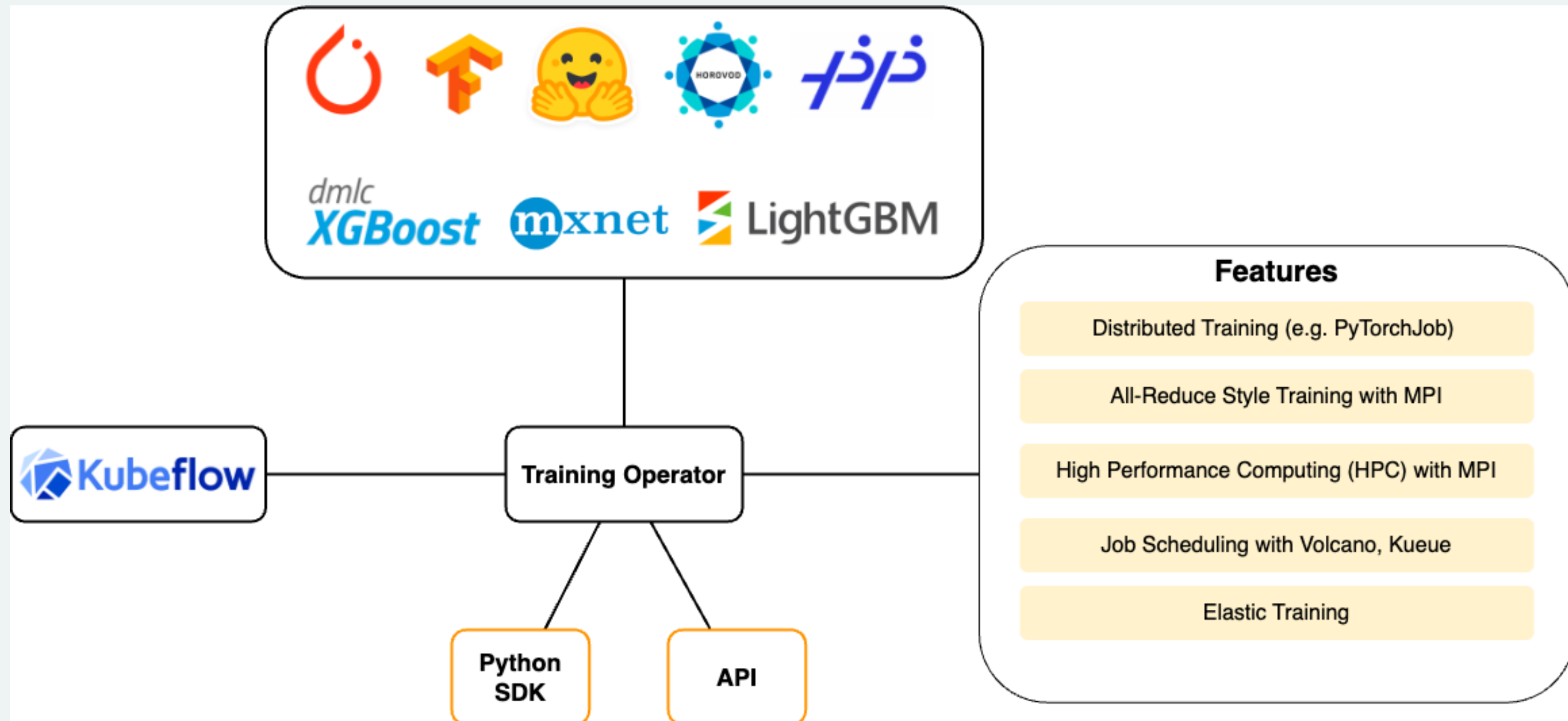


Experiments


 NEW EXPERIMENT

Status	Name	Age	Successful trials	Running trials	Failed trials	Optimal trial
	median-stop	1 hour ago	13	0	0	Validation accuracy: 0.95531 
	open-vaccine-l53wz	3 months ago	3	0	0	Validation loss: 0.62008 
	random-example	23 days ago	12	0	0	Validation accuracy: 0.9779 

TRAINING OPERATOR



KSERVE

 Kubeflow

Home

Notebooks

Tensorboards

Models

Volumes

Experiments (AutoML)


Experiments (KFP)


Pipelines

Runs


Recurring Runs
















Artifacts

kubeflow-user (Owner) 



Model Servers

 NEW MODEL SERVER

Status	Name	Age	Predictor	Runtime	Protocol	Storage URI		
	flowers	4 minutes ago	Tensorflow	1.14.0		gs://kfserving-samples/models/tensorflow/flo...		
	pmml-demo	4 minutes ago	PMML	v0.5.1		https://raw.githubusercontent.com/openscorin...		
	sklearn-iris	4 minutes ago	SKLearn	0.2.1	v2	gs://seldon-models/sklearn/iris		
	torchserve	4 minutes ago	PyTorch	0.3.0	v1	gs://kfserving-examples/models/torchserve/i...		
	xgboost-iris	4 minutes ago	XGBoost	0.2.1	v2	gs://kfserving-samples/models/xgboost/iris		

Comparison with similar solutions



Pricing

Free

Free

Free

Open source



Easy to use



Orientation

Built for deploying and managing ML on Kubernetes

Focus on project management and model performance tracking

Provides flexible workflow management tools for a variety of projects

Integration Capabilities

Powerful integration with Kubernetes and popular machine learning tools

Supports integration with various machine learning frameworks and tools like TensorFlow and PyTorch

Integration with various technologies and services is possible through plugins

ADVANTAGES

Scalability

Portability

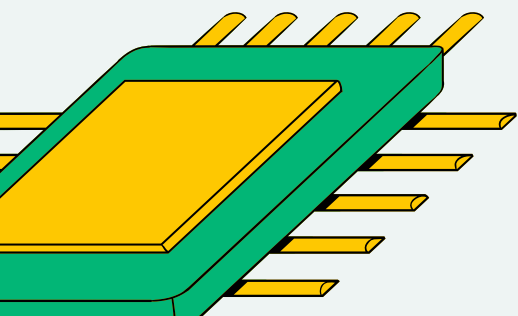
Automation

DISADVANTAGES

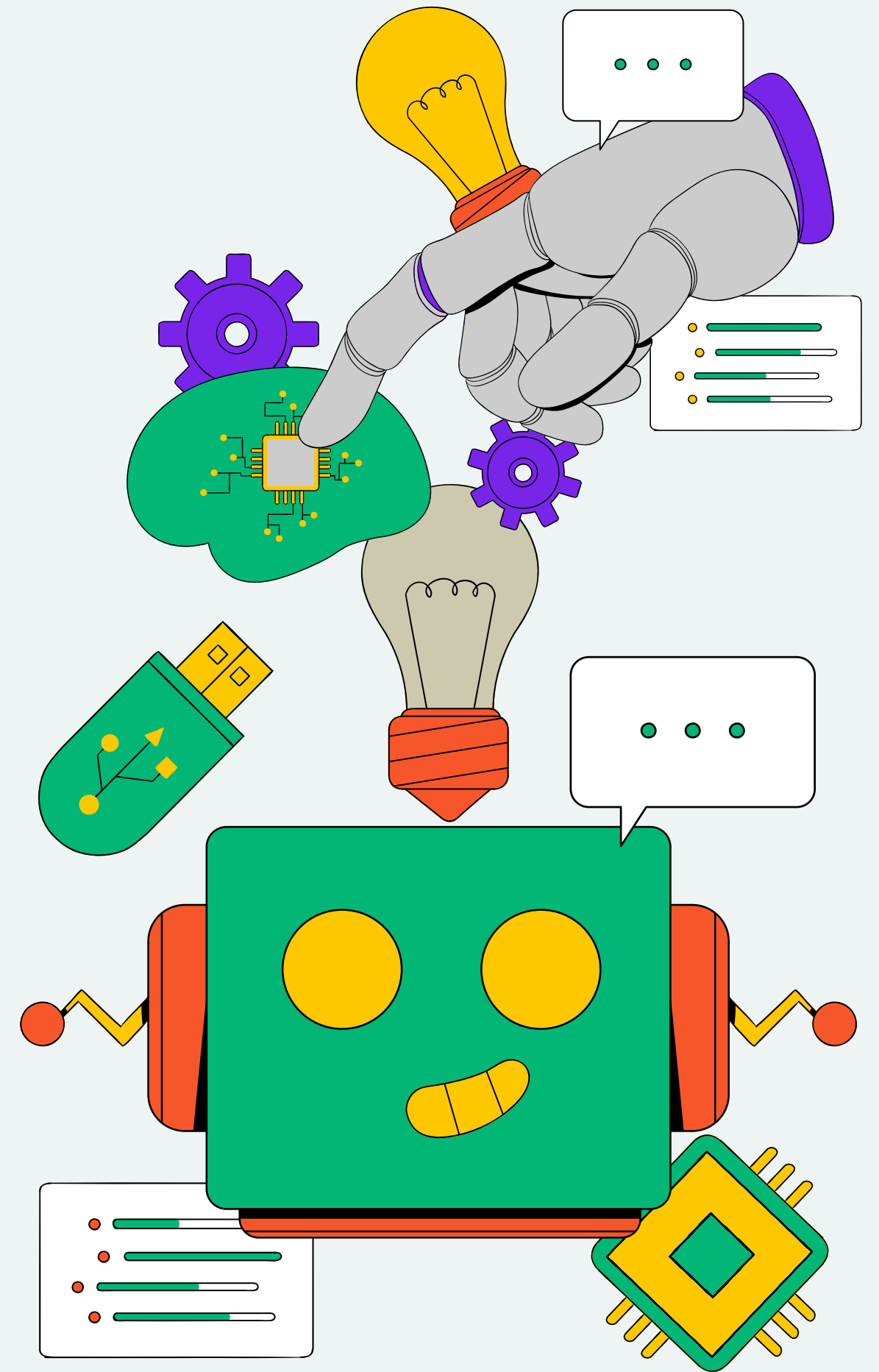
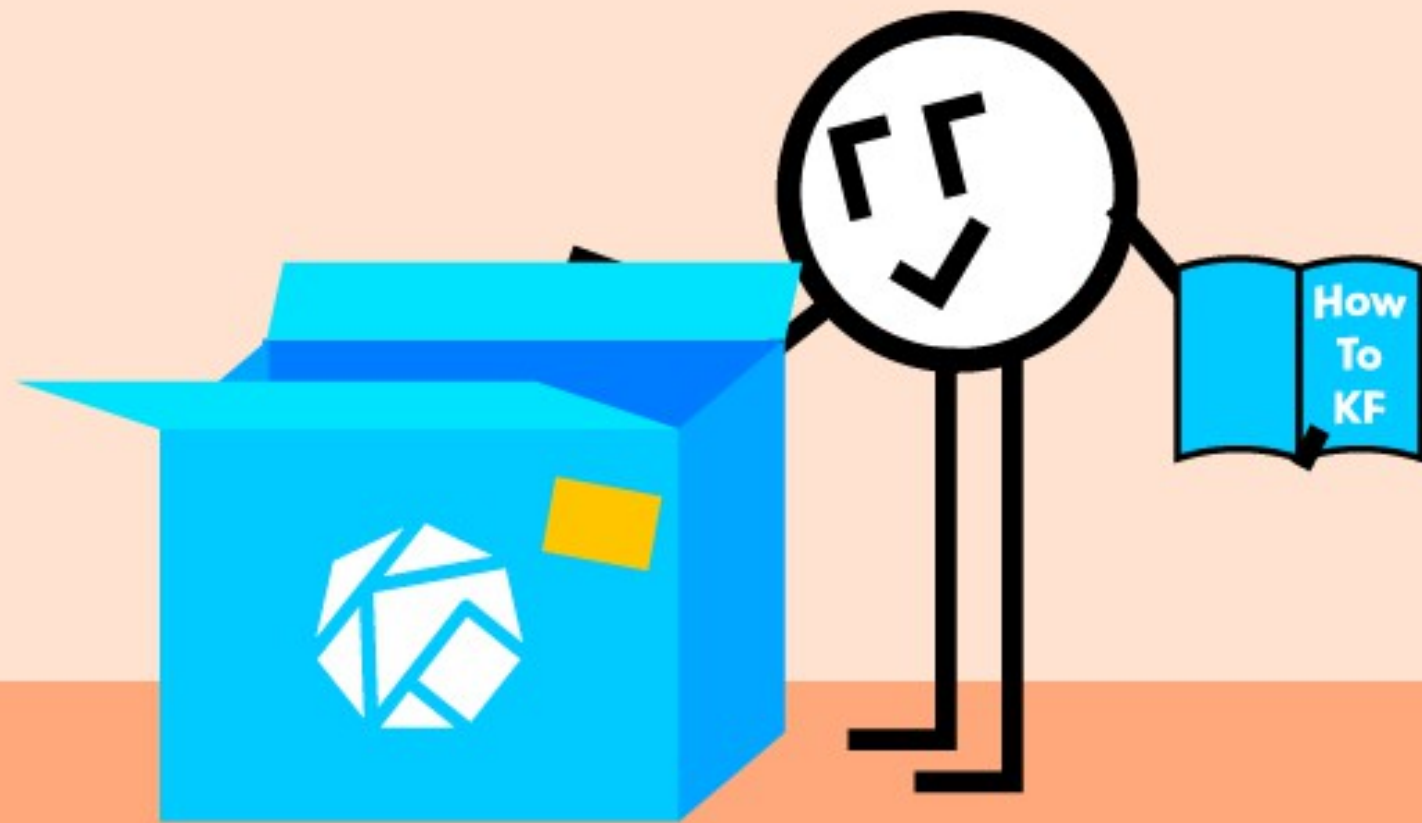
Complexity

High technical requirements

Integration challenges



DEMO TIME!



THANKS FOR LISTENING!

