



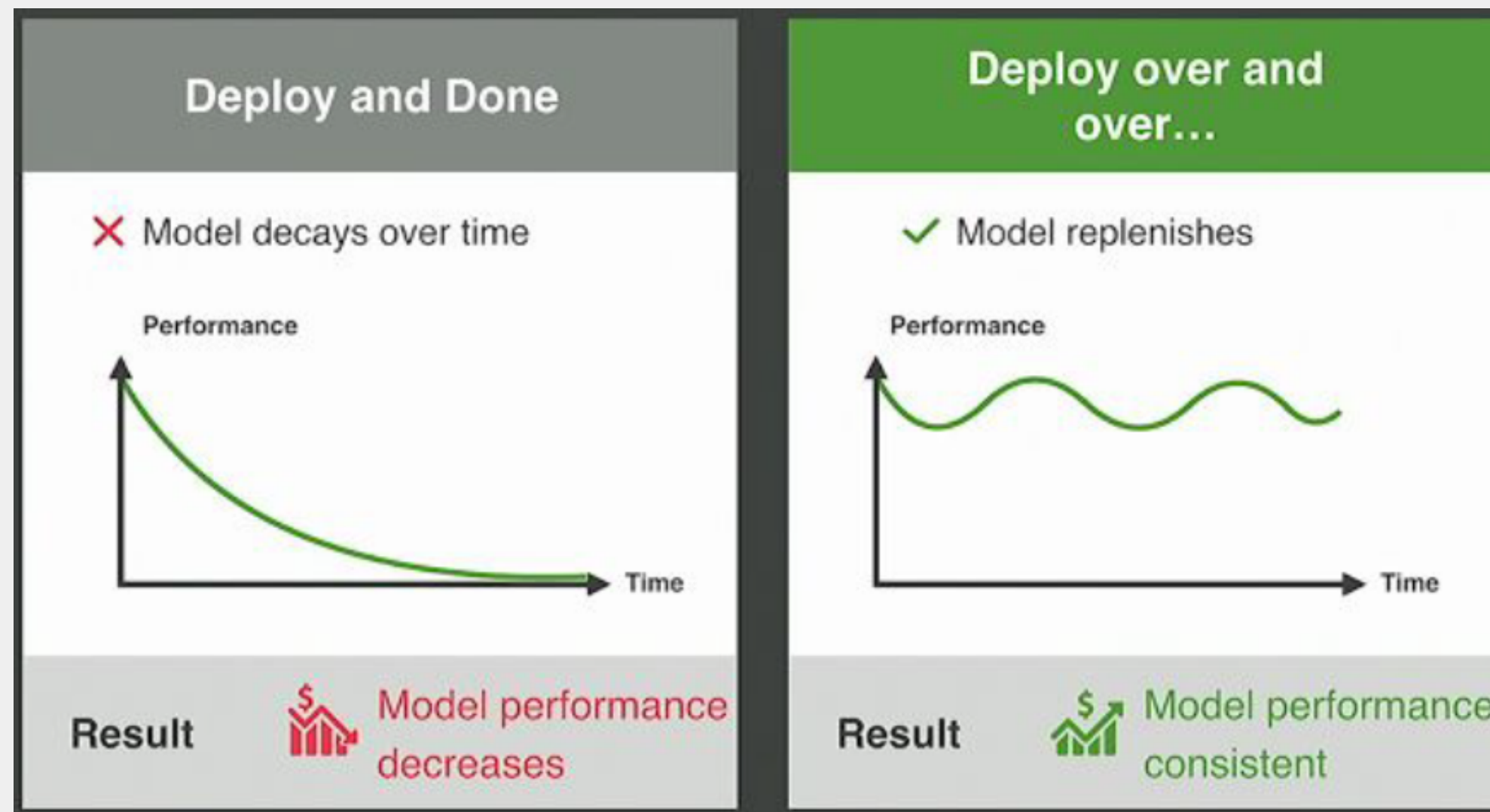
ITI | SEMINARIO INTERNO

Herramientas de MLOps

by Armando Gomis Maya

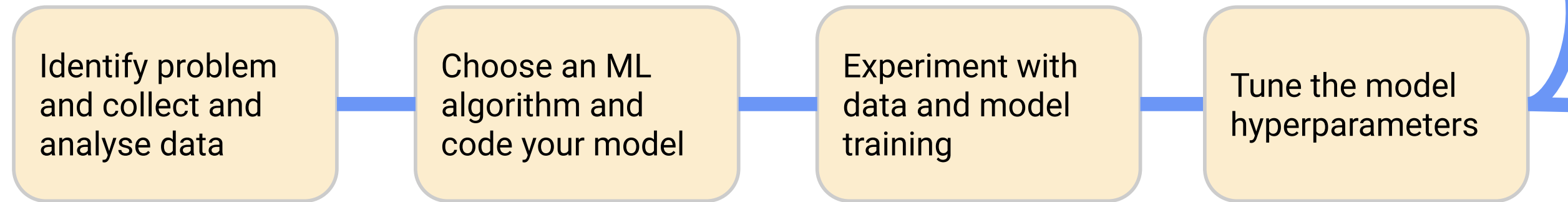


Introducción

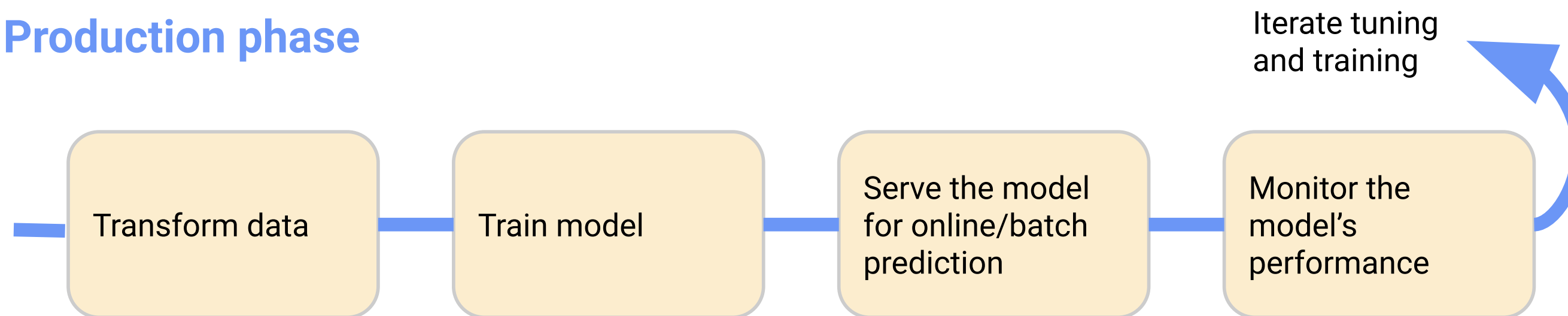




Experimental phase



Production phase





MLFlow

- No permite realizar el ciclo completo de MLOps. Requiere del apoyo de otras herramientas.
- Herramienta ligera y con una curva de aprendizaje sencilla.
- El compute se realiza en la máquina local.
- No permite gestionar permisos ni roles.
- Versionado tradicional mediante números.

Kubeflow

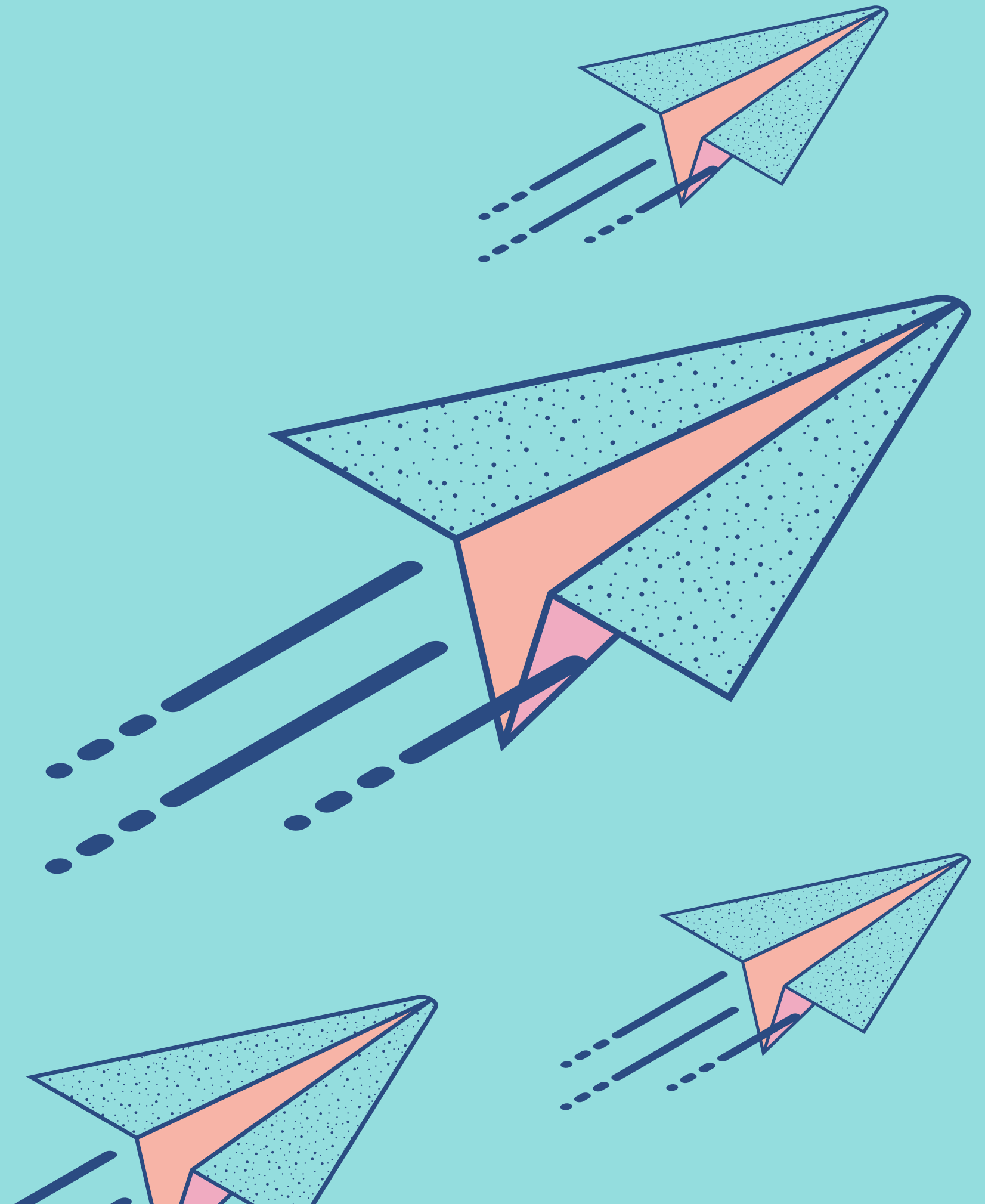
- Plataforma única para realizar el ciclo completo de MLOps.
- Plataforma que requiere una gran cantidad de recursos y con una curva de aprendizaje más compleja.
- El compute se realiza en clusters remotos.
- Kubernetes permite gestionar permisos y roles.
- No utiliza versionado tradicional sino que funciona mediante snapshots.

DVC

Data Version Control: Herramienta para el control de versiones en datos

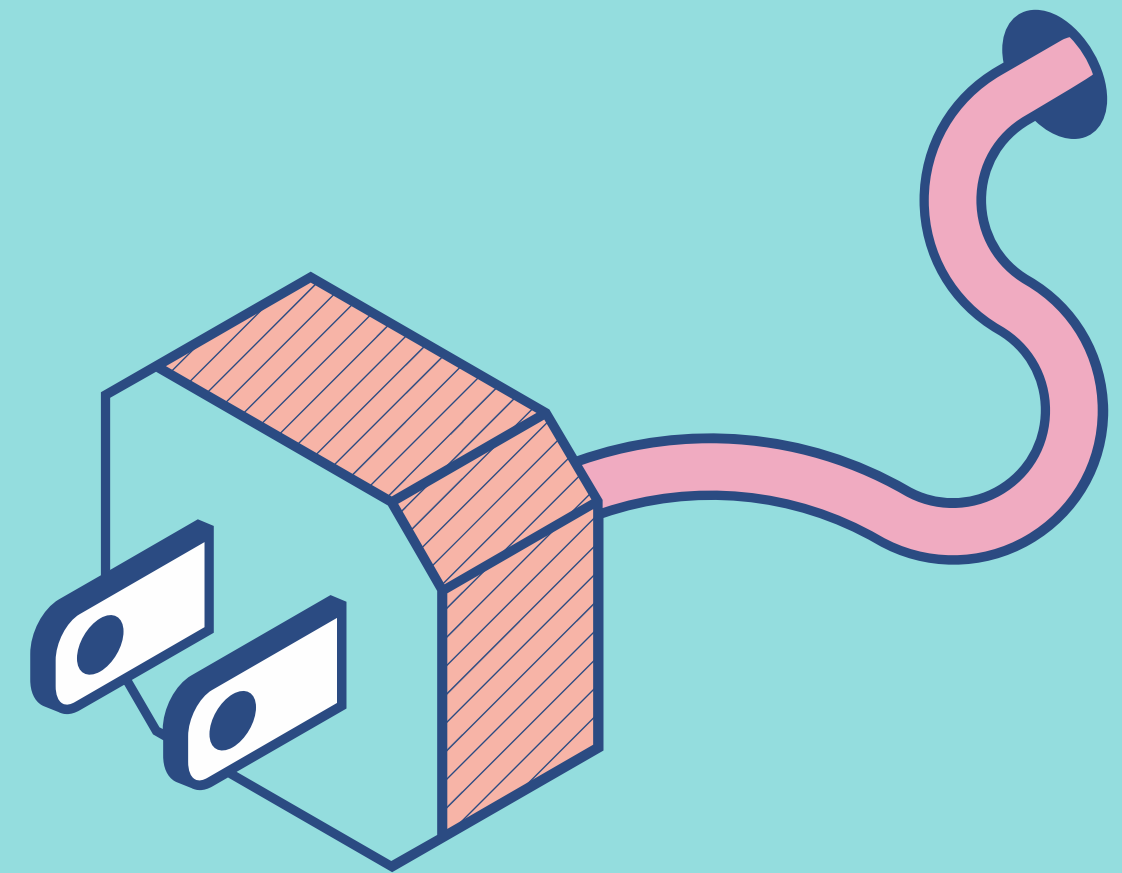
Divide el almacenamiento pesado en un repo remoto.

Trabaja con archivos temporales sobre git.



Instalación

- Trabaja sobre Git, por tanto es necesario tenerlo instalado previamente.
- Ejecutar 'pip install dvc' en un entorno python.
- Alternativas según sistema:
<https://dvc.org/doc/install>



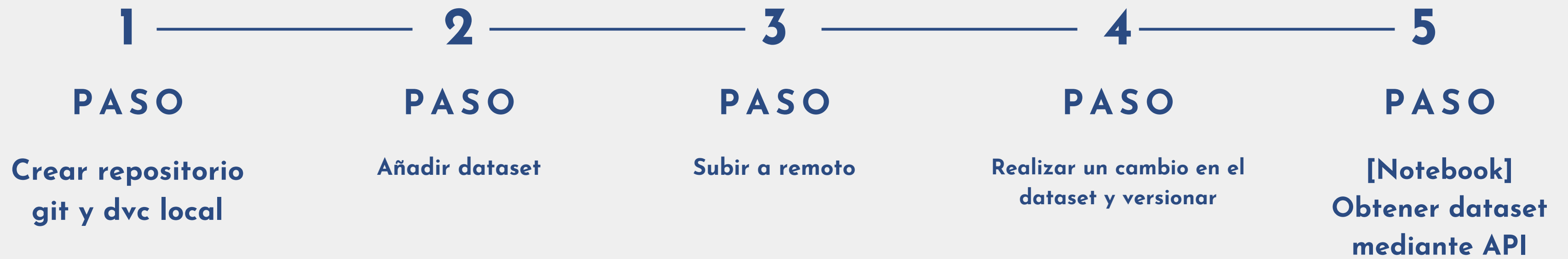


WIKI INDDIS

DVC tracks ML models and data sets

<https://gitlab.iti.upv.es/dataspace/2021-inddis/01-execution/-/wikis/Gu%C3%ADa-ML-Ops-%5Bin-progress%5D>

Demo DVC

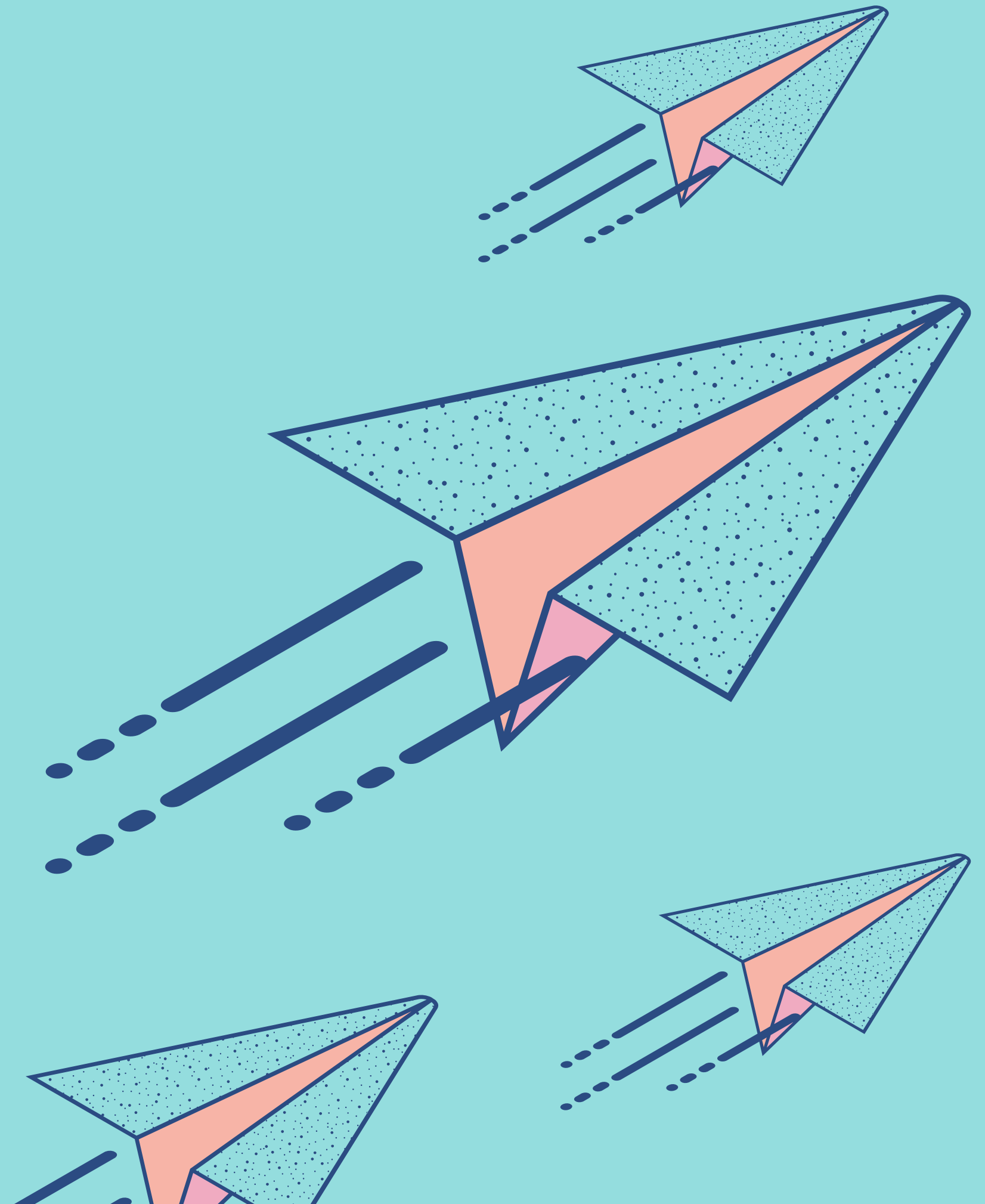


MLFlow

Plataforma open source para el ciclo de vida en proyectos de machine learning.

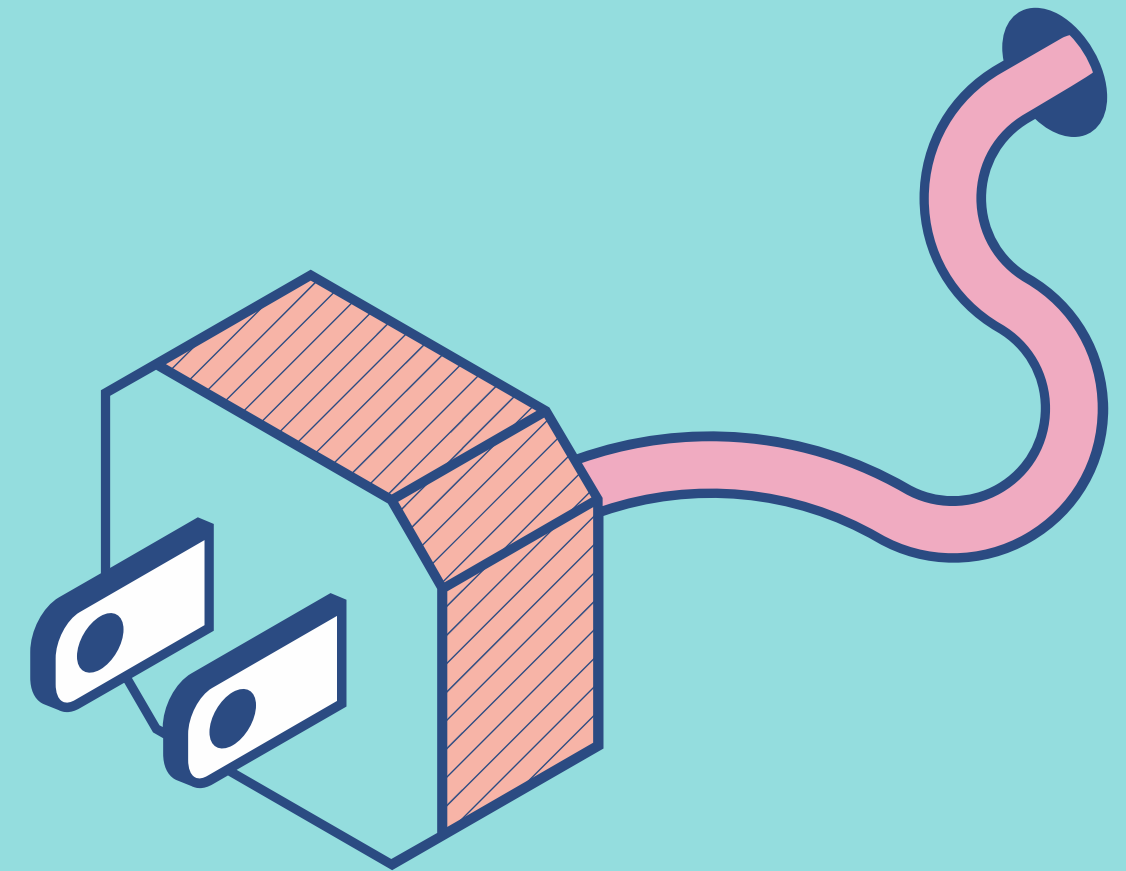
Componentes:

- MLflow Tracking
- MLflow Projects
- MLflow Models
- Model Registry



Instalación

Levantar máquina en Docker



DEMO MLFLOW

1 ————— 2 ————— 3 ————— 4 ————— 5

PASO

Levantar docker compose

- Abrir Jupyter indicado en el terminal.
- Abrir interfaz mlflow en localhost:5000

PASO

Abrir notebook MLFlow base y hacer tour

- Trastear añadiendo logs
- Adjuntar notebook como artifact
- Ver metadatos en MLFlow

PASO

Abrir notebook MLFlow Hyper Param y hacer tour

- Crear nuevo experimento e utilizar.
- Realizar búsqueda entre modelos.
- Realizar comparativa entre modelos.

PASO

Seleccionar experimento y crear modelo

- Crear nuevo modelo.
- Subir varios modelos y observar como cambia la versión.
- Añadir tags, descripción. notas, ...
- Cambiar el Stage.

PASO

Desplegar modelo en una API

- Desplegar con comandos wiki.
- Hacer una petición con Postman.

MLflow Components

mlflow Tracking

Record and query
experiments: code,
data, config, results

mlflow Projects

Packaging format
for reproducible runs
on any platform

mlflow Models

General model format
that supports diverse
deployment tools

Consideraciones

Si se desea implementarlo a escala.

1. Necesaria incorporación de herramientas de seguridad y auditoría.
2. Necesaria la incorporación de bases de datos robustas y almacenamiento de archivos con respaldos.
3. Necesaria la incorporación de control de versiones y un flujo de CI/CD, Ansible, etc.
4. Despliegue del modelo sobre infraestructura de microservicios de kubernetes para garantizar su alta disponibilidad.
5. Necesaria Incorporación de herramienta descriptiva de API mediante swagger.
6. Necesaria incorporación de un API manager para llevar la adecuada contabilidad y acceso al modelo.

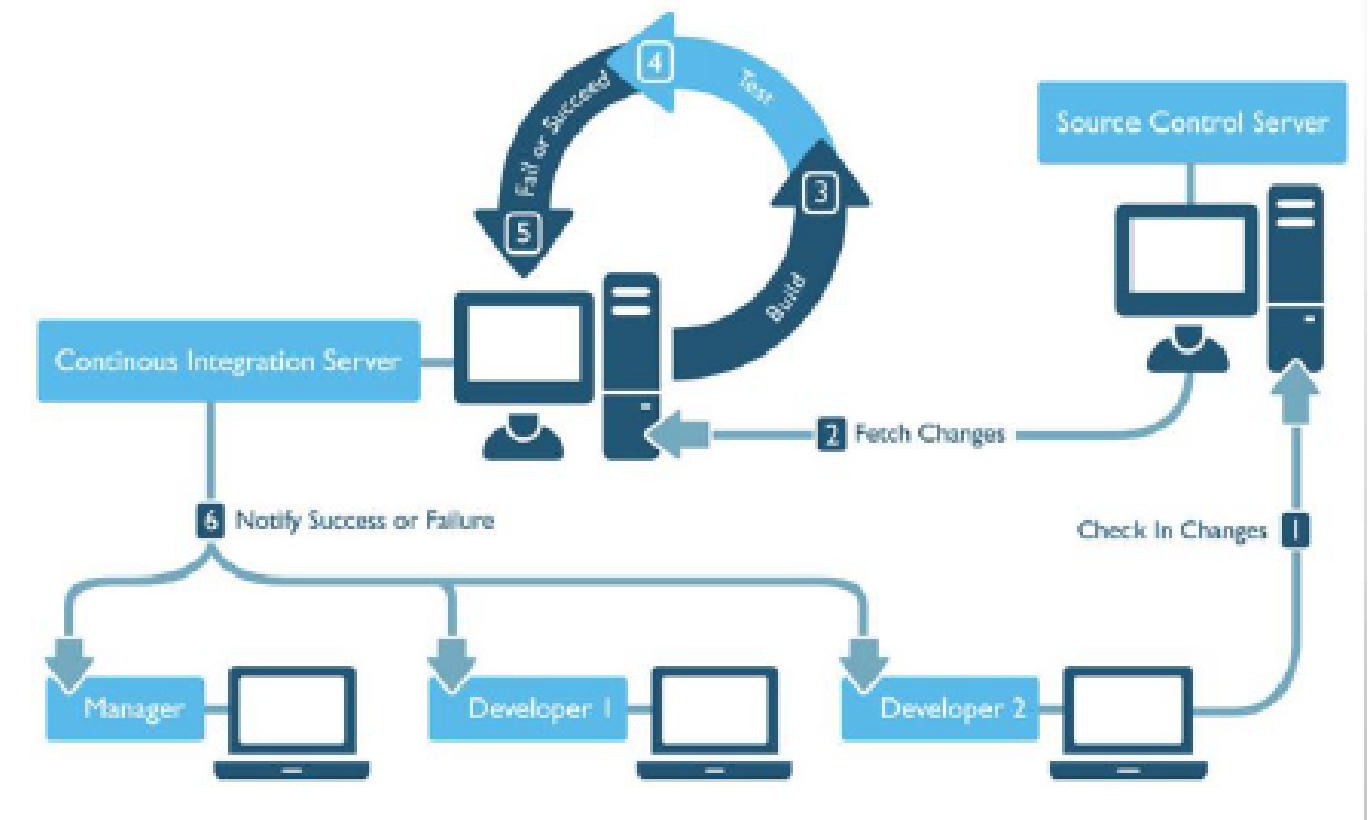
Alternativa

VERSIÓN CLOUD SAAS NEPTUNE
[HTTPS://NEPTUNE.AI/](https://neptune.ai/)

Model Training

Automatization best practices:

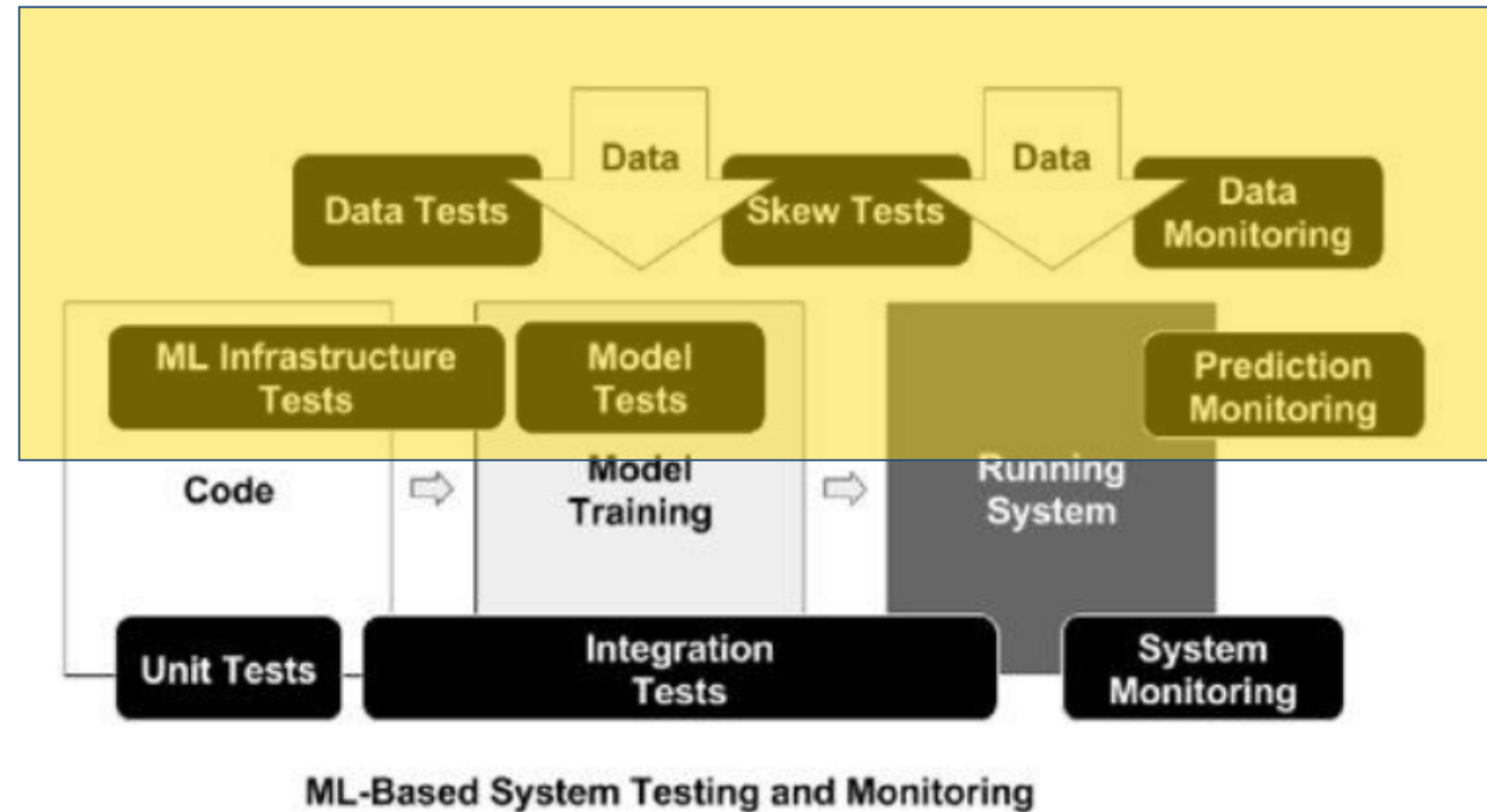
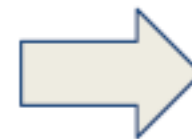
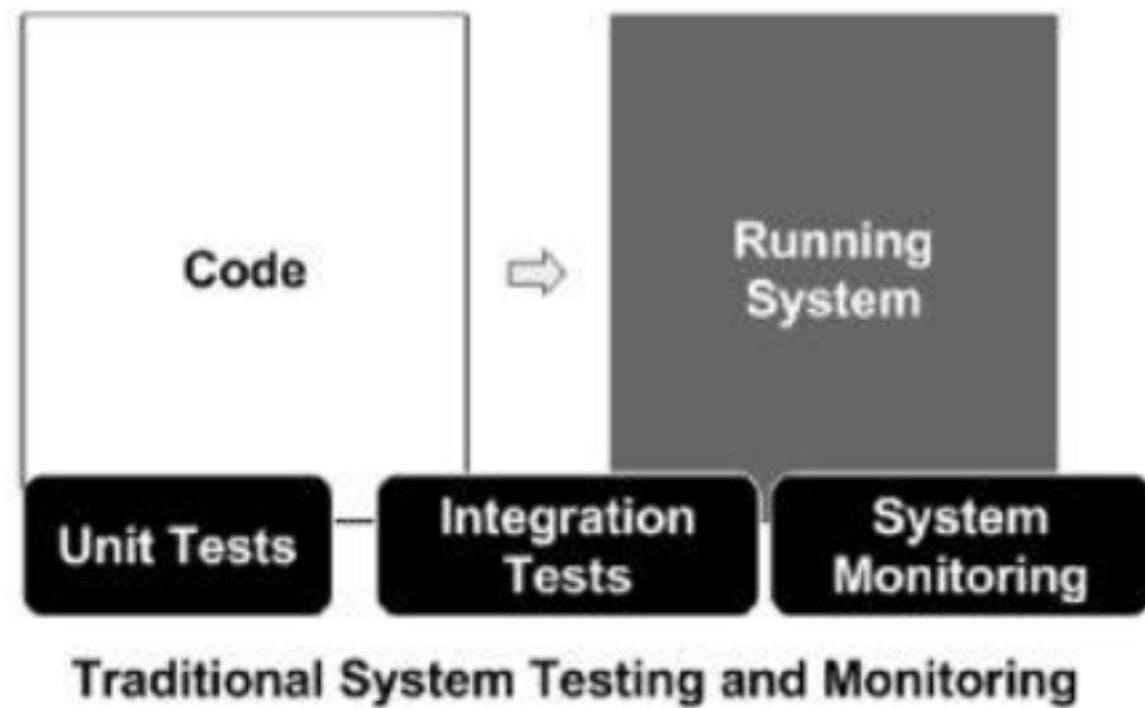
- Versioning the data
- Retrain model
- Accuracy detection
- Log training



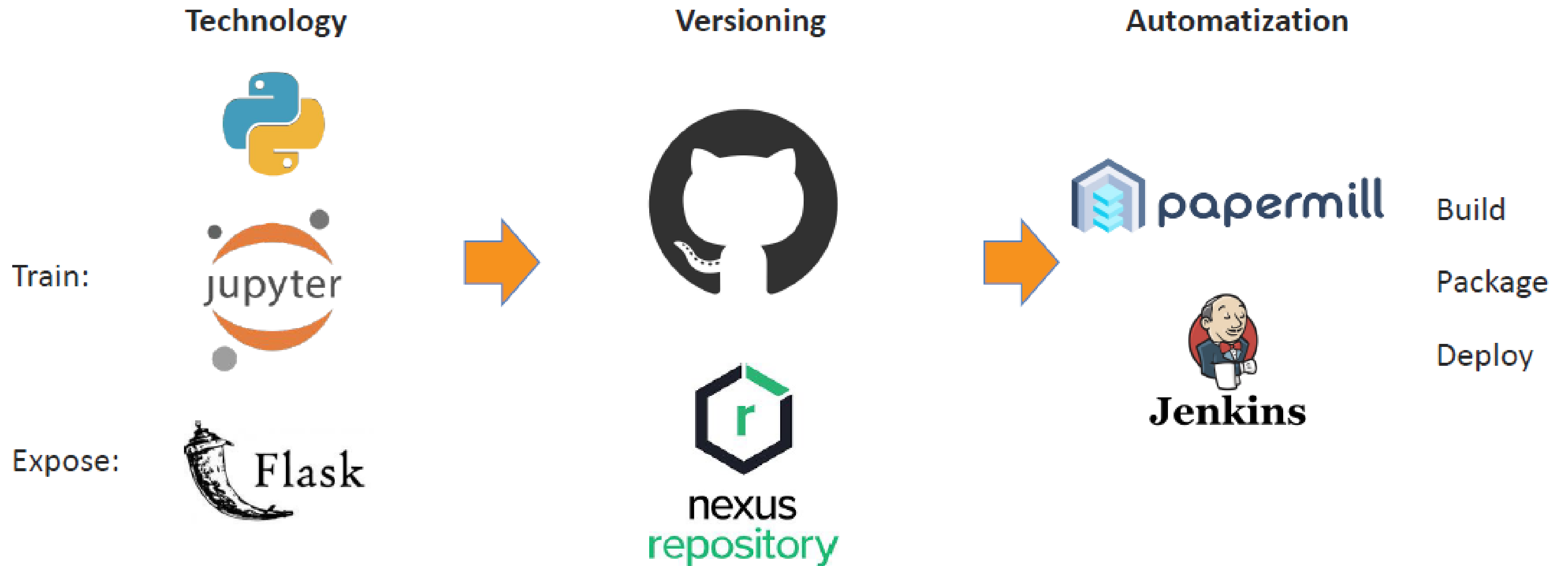
Model	M. Ver.	Data Ver.	Exec.	Accuracy
Churn_pred	0.9	1.2	16:03:15 01/03/18	76.5%
Churn_pred	1.0	1.2	17:45:00 05/03/18	81.3%
Churn_pred	1.0	1.3	09:14:05 10/03/18	89.5%
Churn_pred	0.9	1.3	09:15:00 10/03/18	92.3%

Quality Testing

DEVOPS VS MLOPS

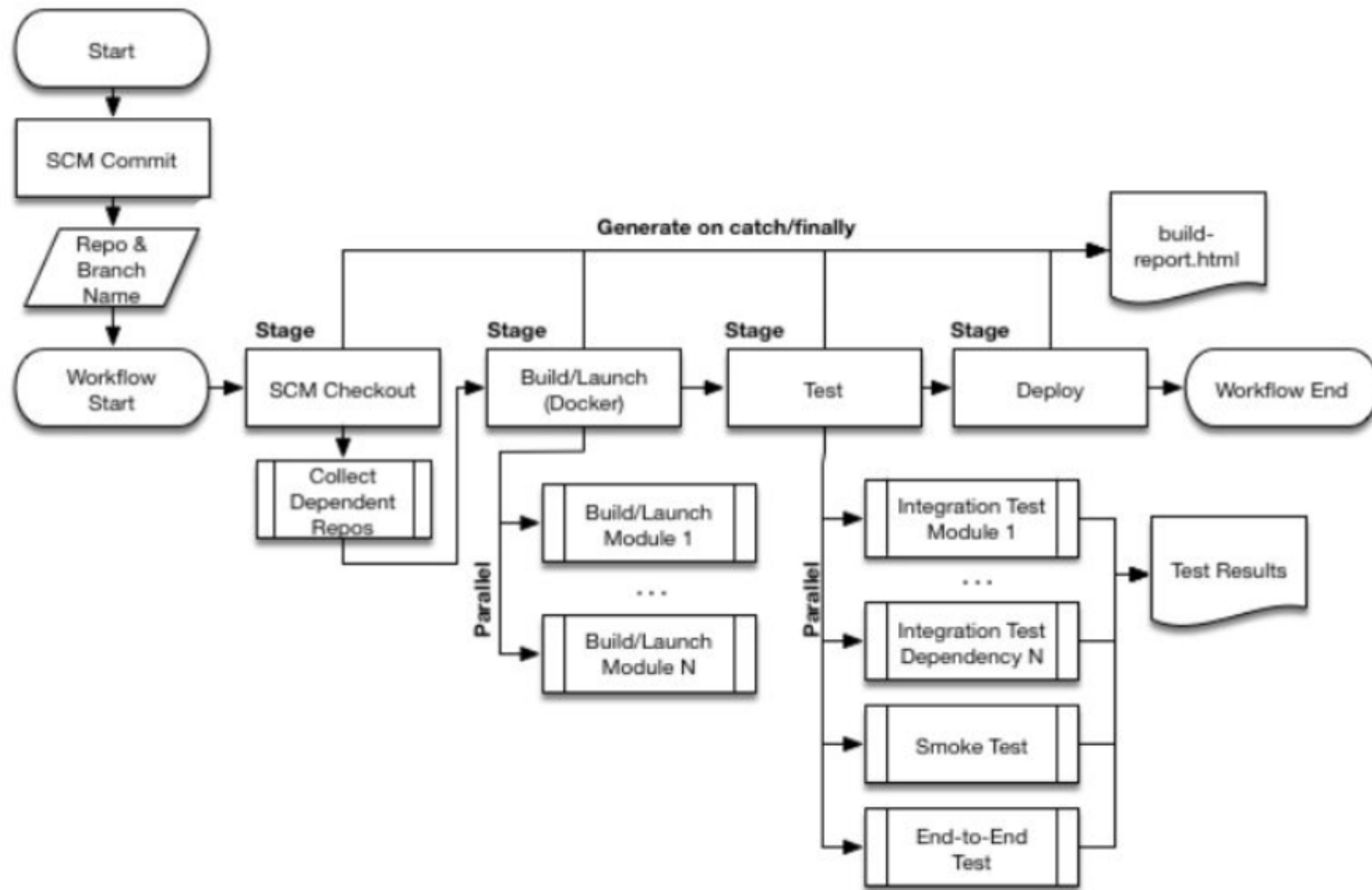


Continuous Integration



Continuous Integration

Automation

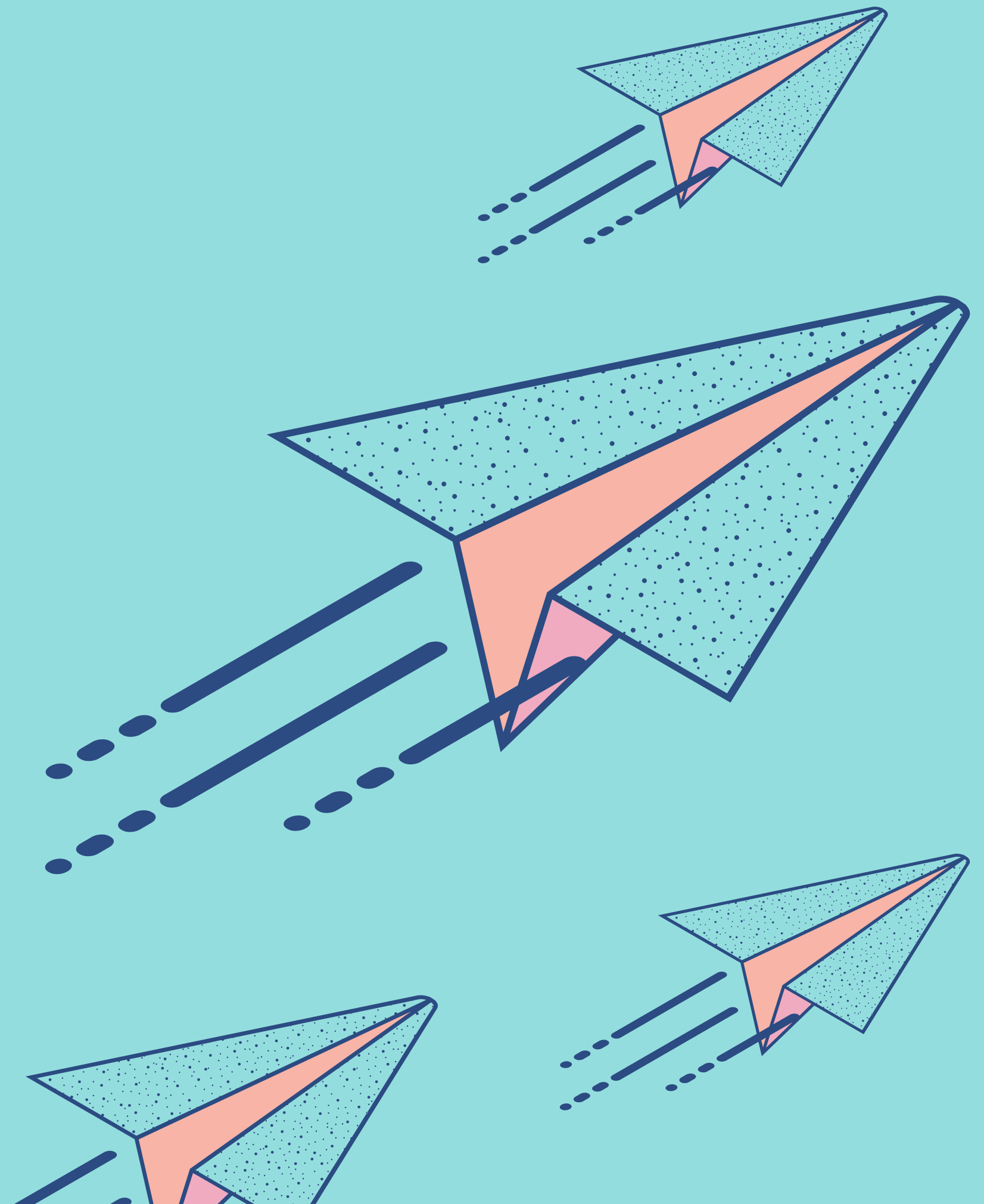


Kubeflow

The Machine Learning Toolkit for
Kubernetes

Componentes:

- Notebooks
- Pipelines
- Training
- Serving



ML tools

Chainer

Jupyter

MPI

MXNet

PyTorch

scikit-learn

TensorFlow

XGBoost

Kubeflow applications and scaffolding

Jupyter notebook web app and controller

Kubeflow UI

Training operators: MPI, MXNet, PyTorch, TFJob, XGBoost

Hyperparameter tuning (Katib)

Kale

Metadata

Pipelines

KFServing

PyTorch Serving

TensorFlow Serving

Seldon Core

Istio

Argo

Prometheus

Spartakus

Platforms / clouds

Kubernetes

GCP

AWS

Azure

IBM Cloud

OpenShift

On prem

ML workflow

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

Serve the model
for online/batch
prediction

Jupyter
Notebook

PyTorch

scikit-learn

TensorFlow

XGBoost

Jupyter
Notebook

Kale

Pipelines

Katib

KFServing

NVIDIA TensorRT

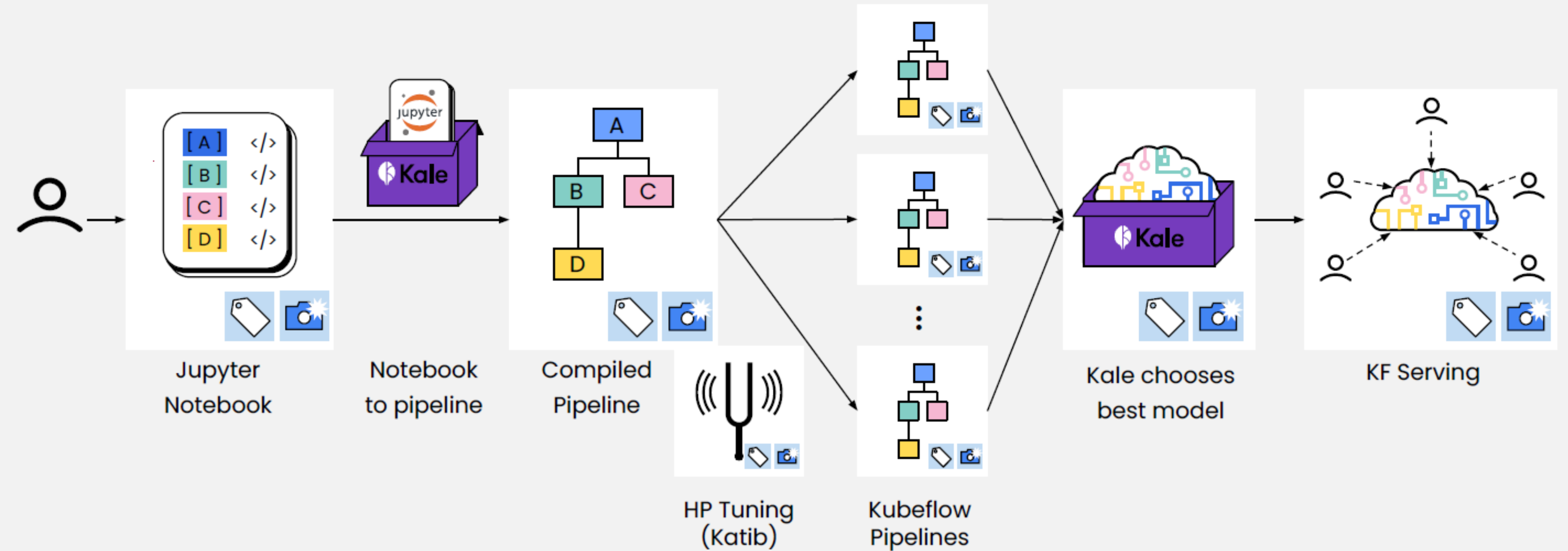
PyTorch

TF Serving

Seldon Core



Data Science Workflows



- A** Step 1
- B** Step 2
- C** Step 3
- D** Step 4



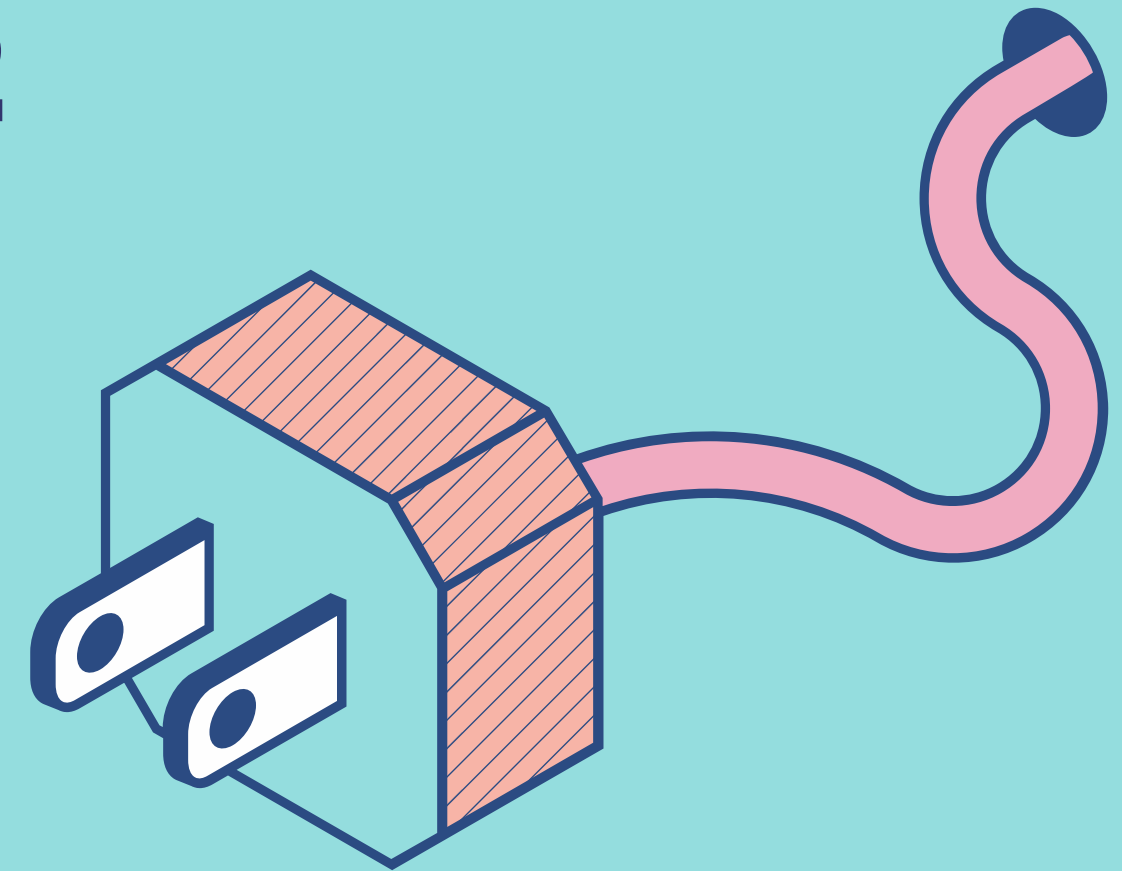
MLMD

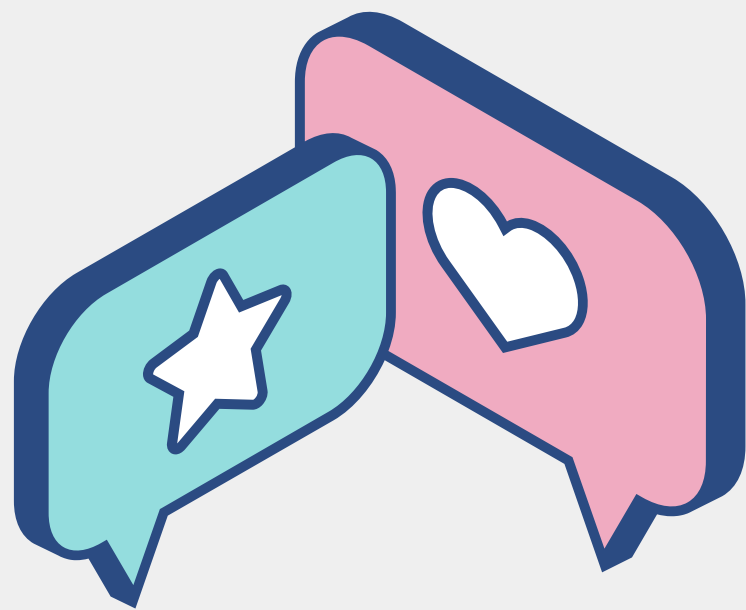


Rok
Data Management

Instalación: wiki instalación 2

- Instalar VirtualBox y Vagrant.
- Es posible que se requiera desinstalar HYPER-V.
- Ejecutar vagrant init arrikto/minikf
- Editar vagrantfile para aumentar recursos
- Ejecutar vagrant up y abrir minikfc en el navegador
- Arrancar MiniKfc con el espacio y abrir kubeflow en





Arrikto: kale + rock

Tutoriales: <https://www.arrikto.com/tutorials/>

DEMO KUBEFLOW [TODO]

1 ————— 2 ————— 3 ————— 4 ————— 5

PASO

Levantar docker
compose

- Abrir minikf en el navegador.
- Abrir kubeflow en el navegador.

PASO

Demo 1:
Creación de un
pipeline

- Comparación entre modelos.
- Reestablecer snapshot mejor modelo.
- Diferencia entre pipeline con Kale y sin Kale.

PASO

Demo 2:
Hyperparameter
Tunning

- Hyperparameter Tunning con Katib.
- Navegación entre experimentos.
- Despliegue del mejor modelo.

PASO

Demo 3:
Auto ML

- Ejecutar proyecto AutoML.
- Navegación experimental y selección del mejor modelo.
- Creación de Tensorboard y monitorización.

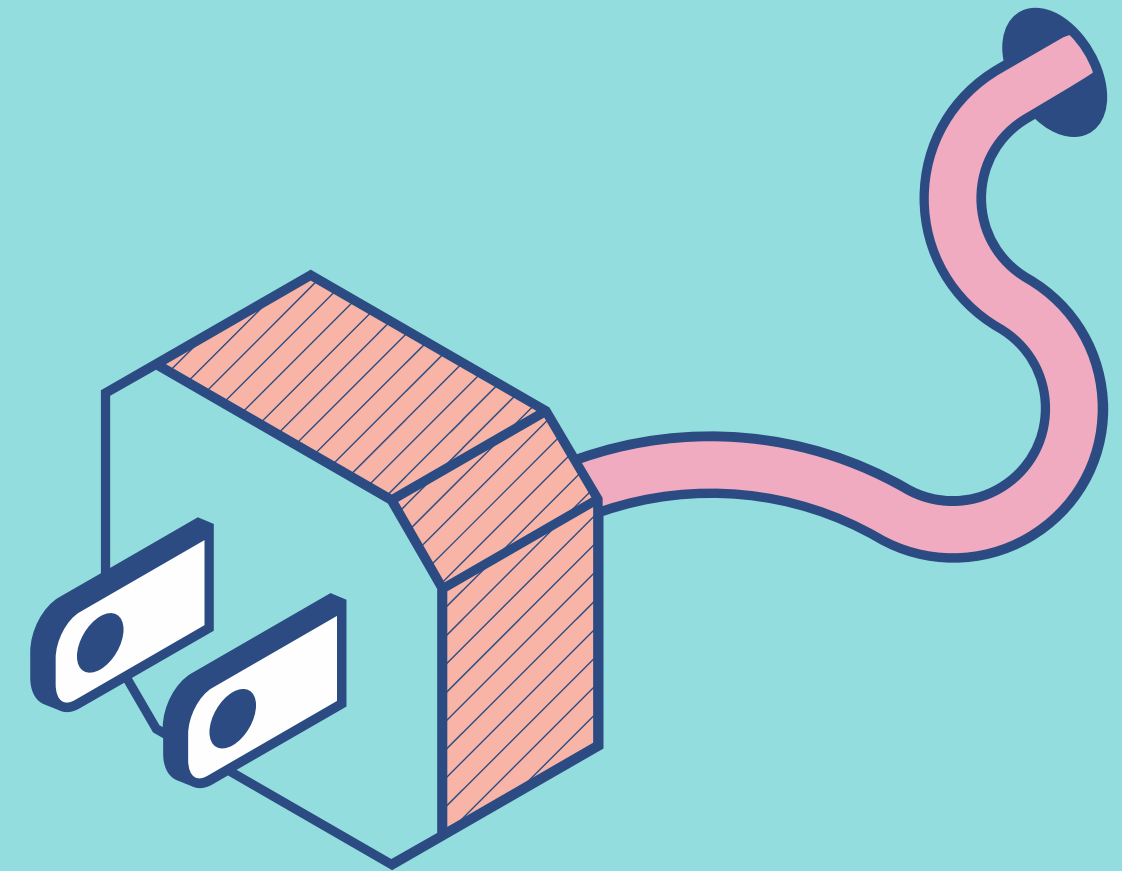
PASO

ADD-ON Feast
[Opcional]

Extras: Add-On Feast

Feature Store

- <https://docs.feast.dev/getting-started/quickstart>
- <https://www.kubeflow.org/docs/external-add-ons/feature-store/>





MLFlow

- No permite realizar el ciclo completo de MLOps. Requiere del apoyo de otras herramientas.
- Herramienta ligera y con una curva de aprendizaje sencilla.
- El computo se realiza en la máquina local.
- No permite gestionar permisos ni roles.
- Versionado tradicional mediante números.

Kubeflow

- Plataforma única para realizar el ciclo completo de MLOps.
- Plataforma que requiere una gran cantidad de recursos y con una curva de aprendizaje más compleja.
- El computo se realiza en clusters remotos.
- Kubernetes permite gestionar permisos y roles.
- No utiliza versionado tradicional sino que funciona mediante snapshots.

The End

