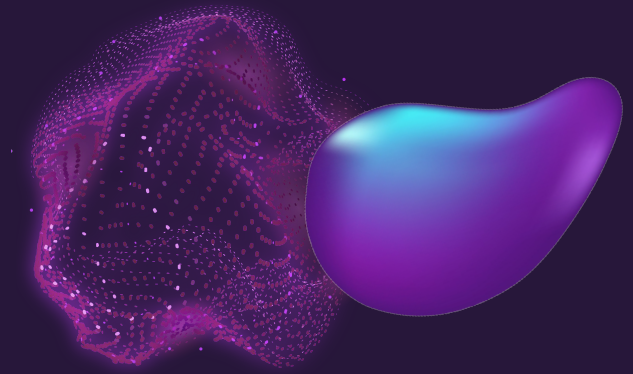


De 0 a héroe con MLOps

#pivoandcode25

Juan Isern Ghosn | Innovasur

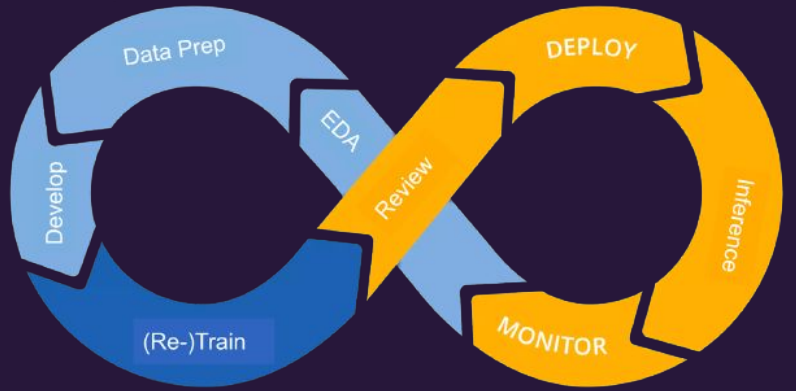
juan@isern:/ $ whoami

- Graduado en Ing. Informática + ADE
- Máster en Ciencia de Datos e Ingeniería de Computadores
- Doctorando en sistemas bioinspirados (UGR)
- Experto en visión artificial
- Optimización de modelos ML para sistemas de bajas prestaciones
- IA Manager en Innovasur

MLOps

La combinación de prácticas de DevOps y aprendizaje automático (Machine Learning ~ ML). Facilita la colaboración y comunicación entre desarrolladores y equipos de operaciones.

1. Datos
2. Entrenamiento / Validación
3. Despliegue



"All-in-one"

Data

Versioning
Pachyderm
DVC
LIQUIDATA

Labeling
figure eight
scale
Aquarium

Processing
Spark
DAGSTER

Exploration
pandas
RAPIDS
dbt

Data Lake / Warehouse
snowflake
databricks

Sources
S3
LOG
Parquet

Training/Evaluation

Frameworks & Distributed Training
fast.ai
PYTORCH
RAY
K
HOROVOD

Hyperparameter Tuning
SIGOPT
Determined AI
Weights & Biases
tune

Experiment Management
Weights & Biases
comet
TensorBoard
Neptune
Machine Learning Lab
mlflow

Resource Management
Determined AI
PY
slurm

Software Engineering
jupyter
Streamlit
git

Compute
CoreWeave
Lambda

Deployment

Feature Store
tecton

Monitoring
fiddler

Edge
ML
NVIDIA TensorRT
TensorFlowLite
ONNX

Web
ALGORITHMIA
SELDON

CI / Testing
Buildkite
great_expectations

or
or

¿Qué vamos a ver?

Conjunto de herramientas básicas (Toolkit) a incluir en nuestros desarrollos de ML: versionado, gestión de los datos, etiquetado, etc.

01

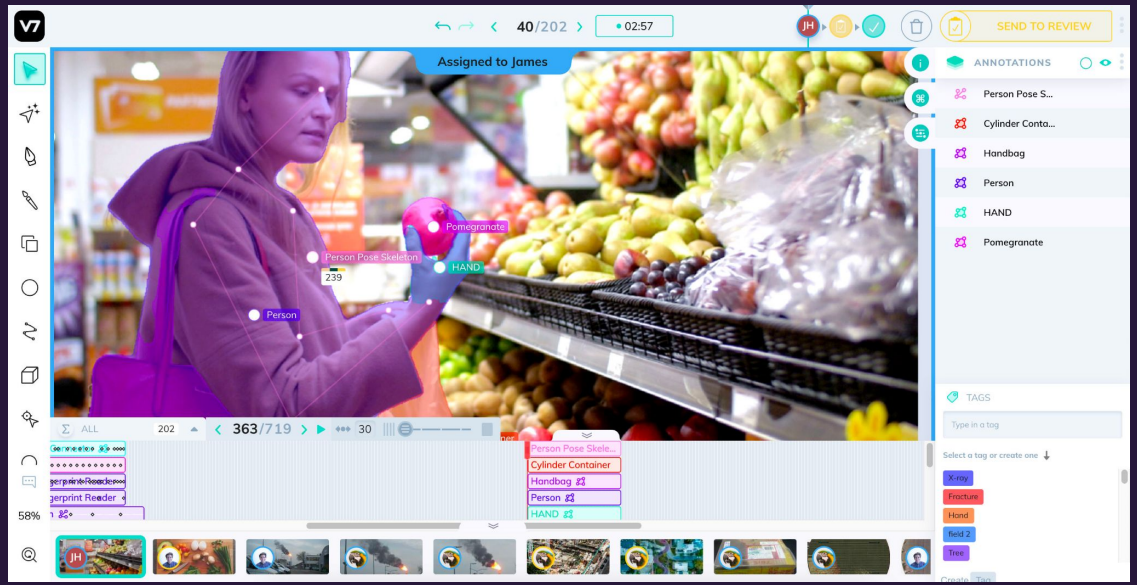
LOS DATOS

MLOps en los Datos

Garantía del rendimiento de los modelos ML respecto al buen hacer en la ingesta, transformación y almacenamiento de los datos.

01. Datos: Etiquetado

El etiquetado de datos implica asignar etiquetas o etiquetas a datos sin procesar



01. Datos: Etiquetado

Amazon
SageMaker
GroundTruth

Coste por etiqueta

500 objetos/mes gratis

Amazon MTurk

01. Datos: Etiquetado

Amazon
SageMaker
GroundTruth

Coste por etiqueta

500 objetos/mes gratis

Amazon MTurk

LabelStudio

Comunidad más amplia

Datos múlti-dominio:

- Texto
- Audio
- Video

Oh oh! Open source detected

01. Datos: Etiquetado

Amazon SageMaker GroundTruth

Coste por etiqueta

500 objetos/mes gratis

Amazon MTurk

LabelStudio

Comunidad más amplia

Datos múlti-dominio:

- Texto
- Audio
- Video

CVAT

Más completo para CV

Integración con ML para etiquetado asistido

V. Community gratuita

01. Datos: Control de versiones

Mantener un registro de las distintas versiones de los datos utilizados para nuestros experimentos

- Introducción de nuevos datos
- Reducción de sesgos evidenciados
- Registro de datos

01. Datos: Control de versiones

DVC

interactive.ai

Git para datos / modelos

Conecta el almacenamiento con el repositorio de código

Registro

01. Datos: Control de versiones

DVC
interactive.ai

Git para datos / modelos

Conecta el almacenamiento con el repositorio de código

Registro

Git LFS

Extensión de Git

Datos de gran tamaño

[Cuidado] No incluir versionado de código conjuntamente

01. Datos: Control de versiones

DVC
interactive.ai

Git para datos / modelos

Conecta el almacenamiento con el repositorio de código

Registro

Git LFS

Extensión de Git

Datos de gran tamaño

[Cuidado] No incluir versionado de código conjuntamente

Pachyderm

Version community “apañá”

Pipelines de validación

Integración con frameworks principales

02

ENTRENAMIENTO / VALIDACIÓN

Entrenamiento

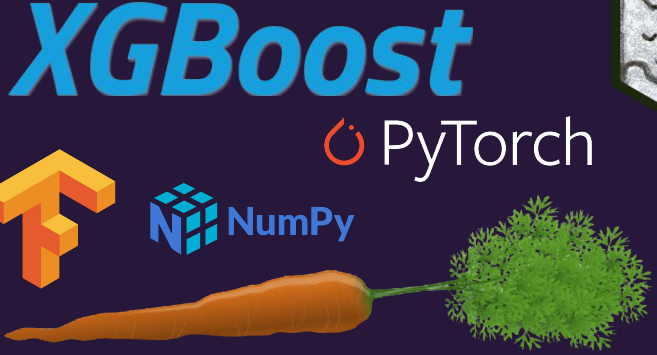
Aprender patrones o relaciones a partir de datos.

Ajuste de los parámetros internos del modelo.

Dependiente de una validación

02. Entrenamiento: Frameworks

El de tu elección: Depende de tu caso de uso de aplicación ML



02. Entrenamiento: IDE

El de tu elección: Depende de framework / lenguaje

python™ Haskell R

C++ julia Ruby

Java C# ...

02. Entrenamiento: Control

WandB

Cloud & self-hosted

Vistas personalizadas

Gestión de equipos & roles

[Importante] Caro

02. Entrenamiento: Control

WandB

Cloud & self-hosted

Vistas personalizadas

Gestión de equipos & roles

[Importante] Caro

MLFlow

Self-hosted

Lo justo para hacer de todo

Control de versiones de modelos & registro

02. Entrenamiento: Control

WandB

Cloud & self-hosted

Vistas personalizadas

Gestión de equipos & roles

[Importante] Caro

MLFlow

Self-hosted

Lo justo para hacer de todo

Control de versiones de modelos & registro

Comet

Despliegues automatizados

Accesibilidad a terceras partes

Entorno colaborativo

02. Entrenamiento: Frameworks

Diversidad de frameworks

Distintos formatos de modelos

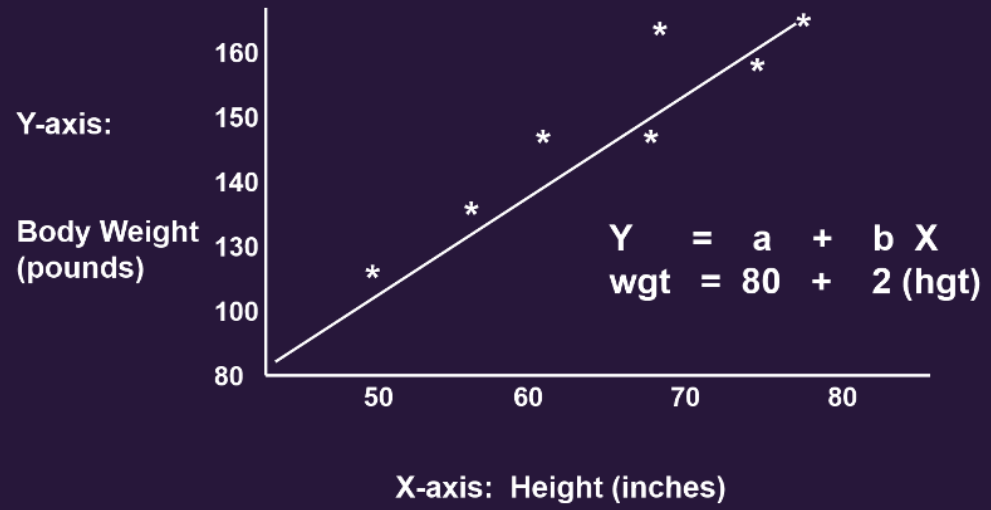
02. Entrenamiento: Interoperabilidad

Necesidad de un estándar:

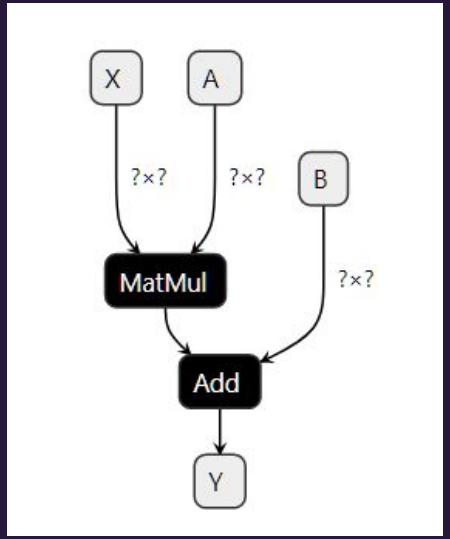
- Interoperabilidad entre sistemas/frameworks
- Adaptado a la innovación/ Soporte de la comunidad

ONNX

02. Entrenamiento: ONNX

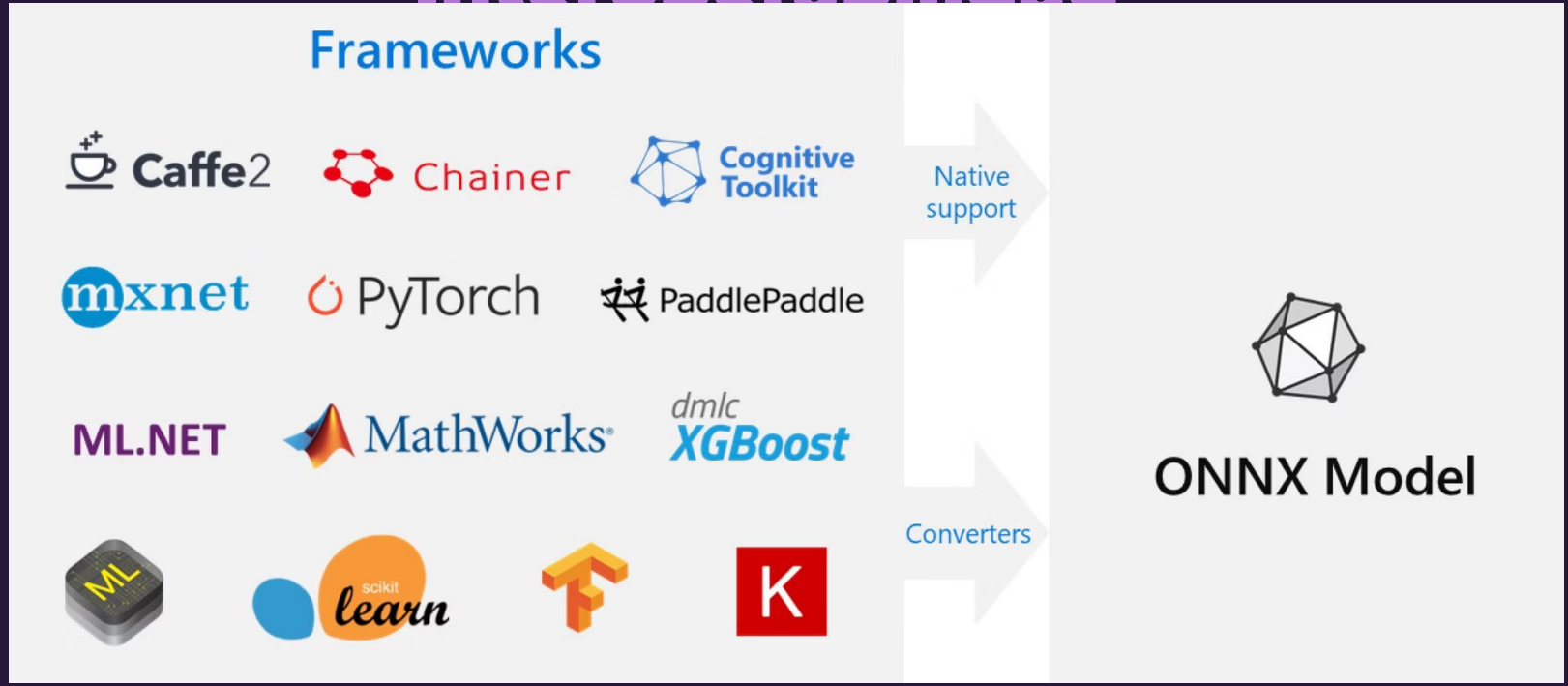


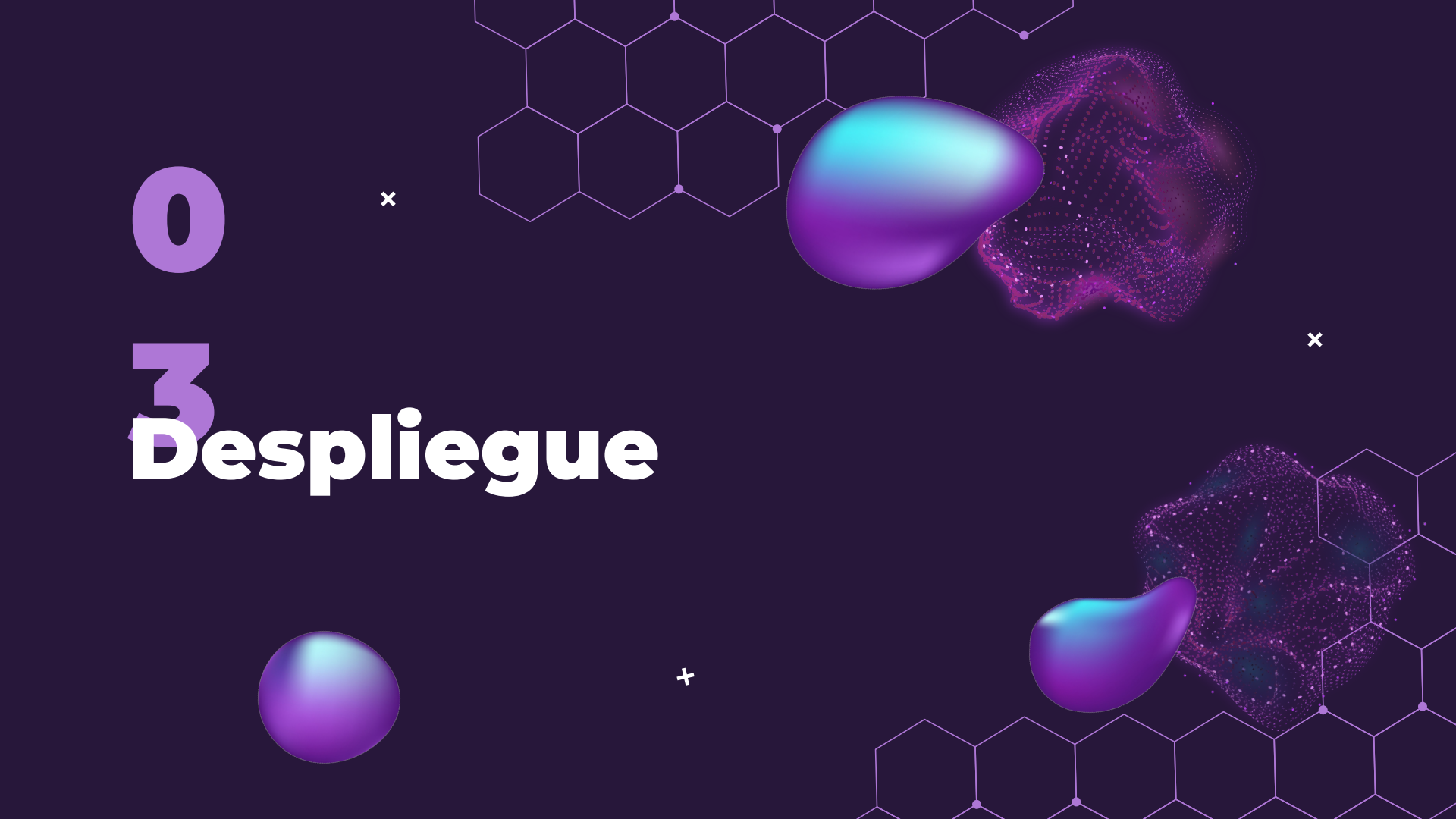
ONNX




```
def onnx_linear_regressor(X):
    "ONNX code for a linear regression"
    return onnx.Add(onnx.MatMul(X, coefficients), bias)
```

02. Entrenamiento: Interoperabilidad




03
Despliegue

Despliegue

Poner el modelo ML a disposición del usuario final / grupos de interés.

Monitorizar el rendimiento del modelo

Automatizar el despliegue

03. Despliegue: Parcial

gradio

Ejemplos

Evaluación por parte del equipo de desarrollo

Difusión interna/ marketing del modelo

Servicio web

03. Despliegue: Control

Monitorizar modelo en producción

WandB

Cloud & self-hosted

Vistas personalizadas

Gestión de equipos & roles

[Importante] Caro

MLFlow

Self-hosted

Lo justo para hacer de todo

Control de versiones de modelos & registro

Comet

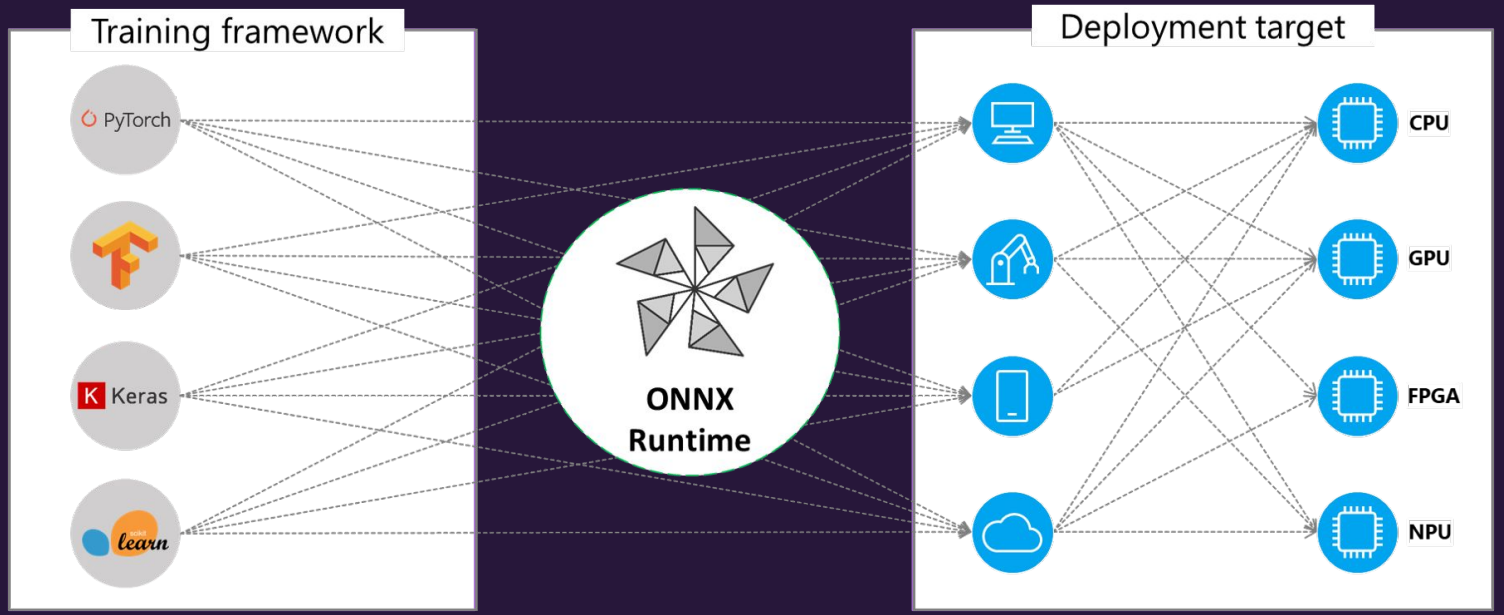
Despliegues automatizados

Accesibilidad a terceras partes

Entorno colaborativo

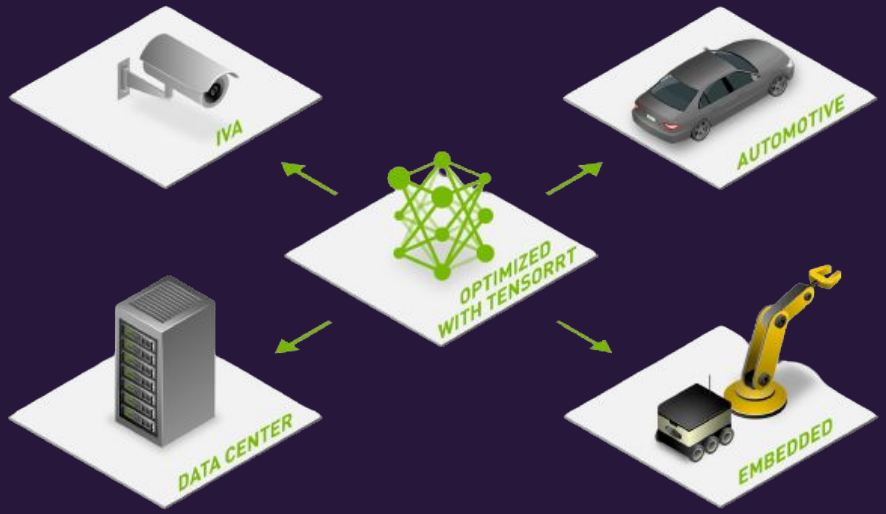
03. Despliegue: Arquitectura

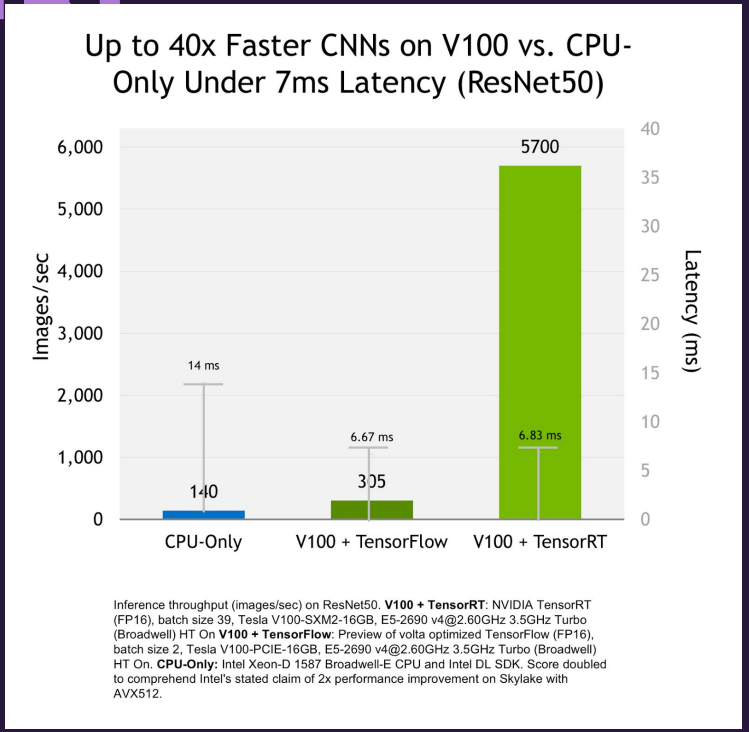
Elección del dispositivos de procesamiento. ONNX como nexo.



03. Despliegue: NVIDIA TensorRT

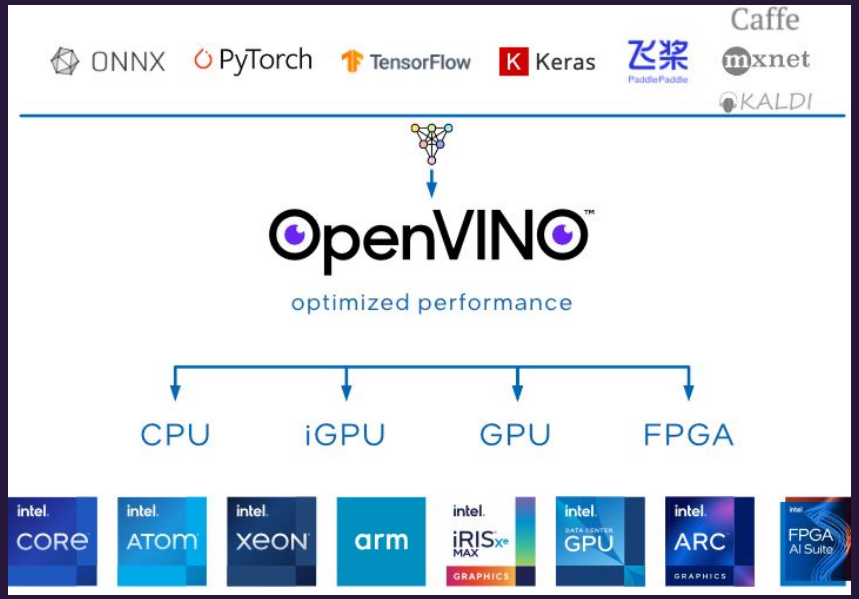
Framework de NVIDIA para NVIDIA

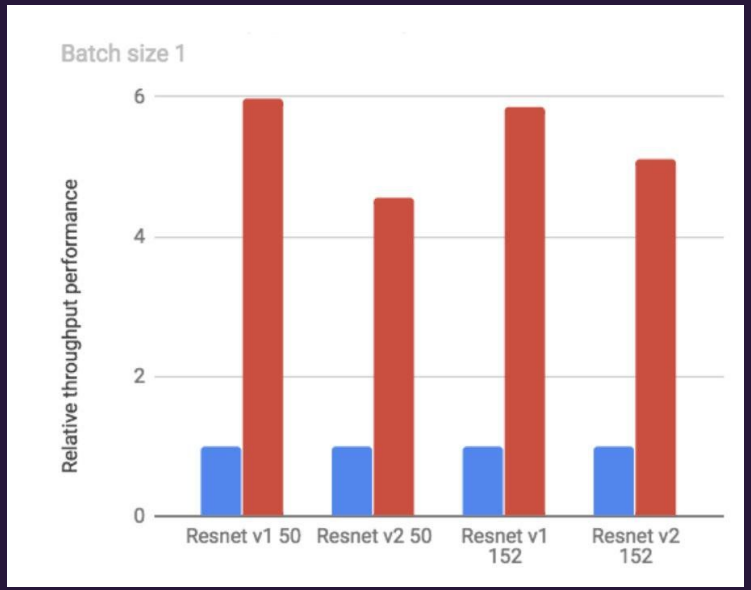




03. Despliegue: Intel OpenVino

Framework de Intel para Intel





03. Despliegue: CI/CD

Nueva tecnología, mismos conocidos:

Github actions

Gitlab CI/CD

Jenkins 

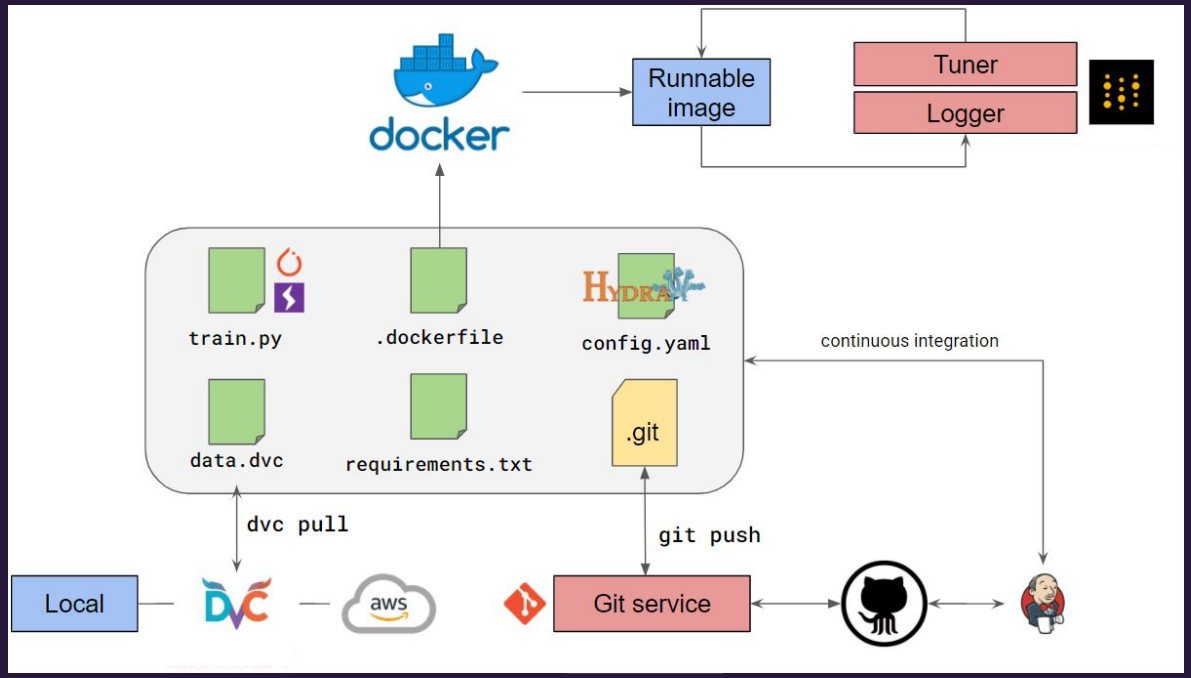
Pipelines de acciones ante eventos de desencadenado:

- Nuevos datos de entrenamiento
- Nuevo modelo más efectivo
- Nuevo modelo más eficiente

04

Flujo completo

04. Ejemplo operativo



Gracias!

¿Alguna pregunta?

juan.isern95@gmail.com

jisern.rocks

04. Otros: IaaS-ML