



openexpo
europe

Cómo hemos construido una plataforma de IA basada en open source

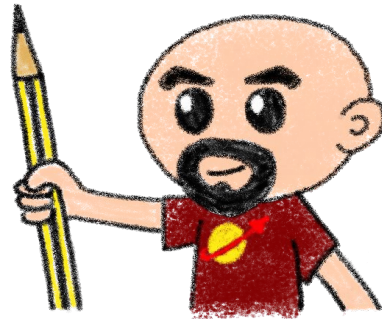
Horacio González
2022-06-30



Horacio González
@LostInBrittany

¿Quiénes somos?

Presentándome y presentando OVHcloud

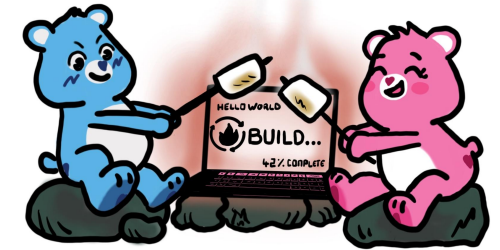
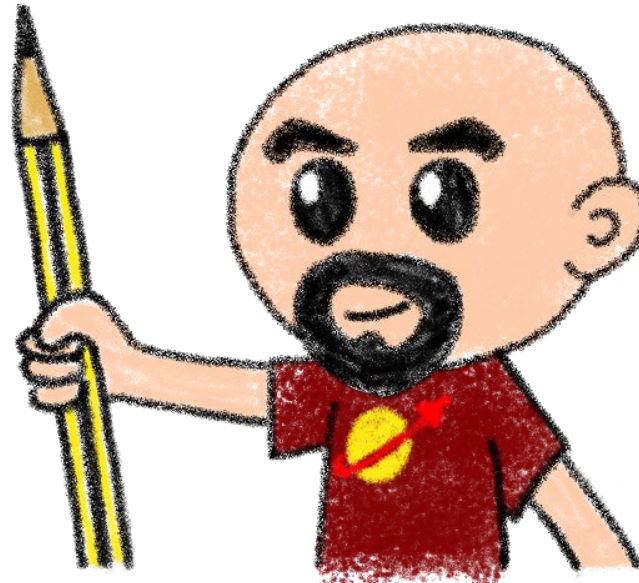


Horacio Gonzalez



@LostInBrittany

Español perdido en la Bretaña francesa, desarrollador de software, cuentacuentos, rolista y geek en general...



OVHcloud: un líder mundial



Web Cloud & Telecom



Private Cloud



Public Cloud



Storage



Network & Security



30 Data Centers
in 12 locations



34 Points of Presence
on a 20 TBPS Bandwidth Network



2200 Employees
worldwide



115K Private Cloud
VMS running



300K Public Cloud
instances running



380K Physical Servers
running in our data centers



1 Million+ Servers
produced since 1999



1.5 Million Customers
across 132 countries



3.8 Million Websites
hosting



1.5 Billion Euros Invested
since 2016



P.U.E. 1.09
Energy efficiency indicator

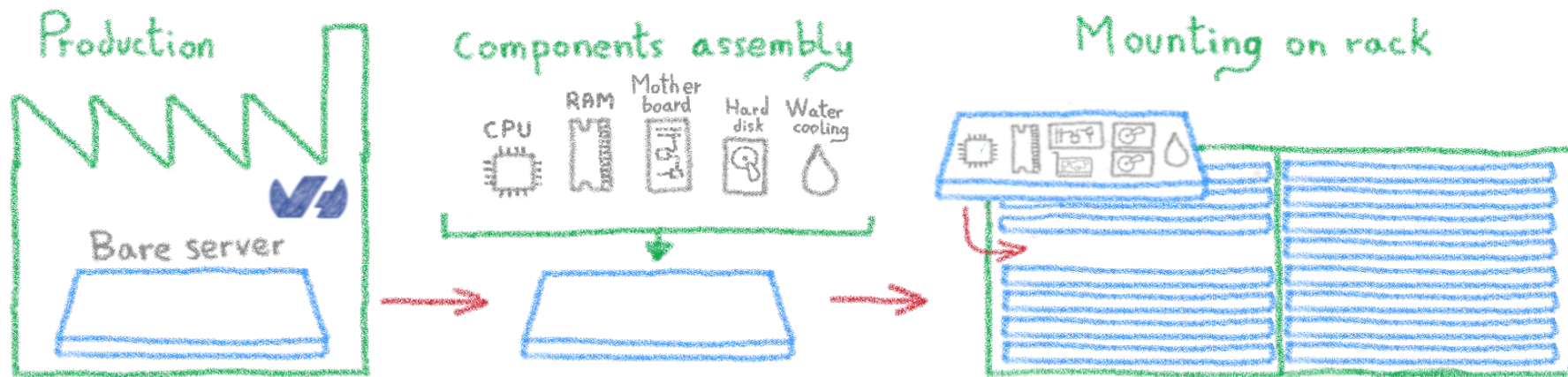


20+ Years in Business
Disrupting since 1999

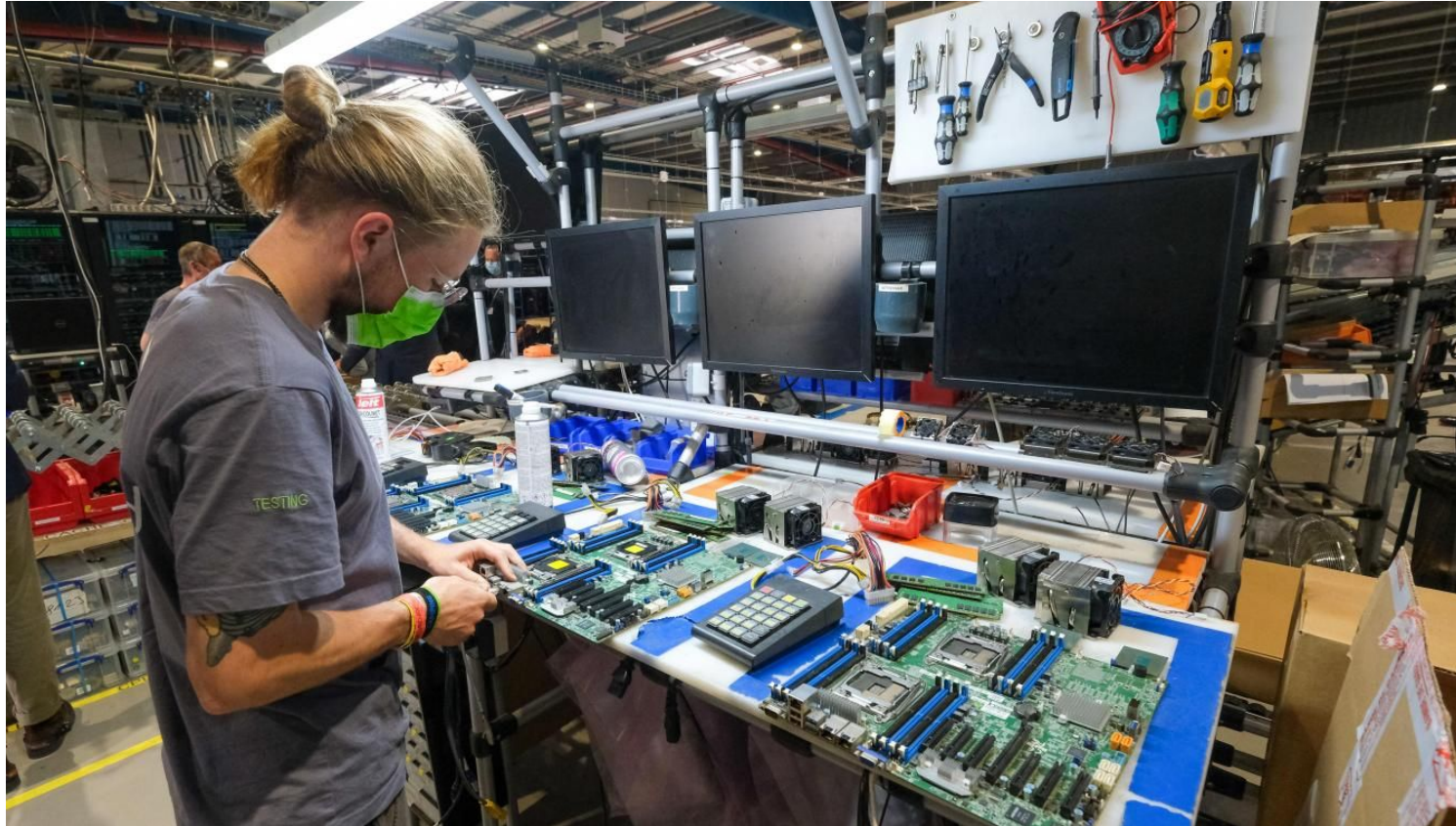


OVHcloud: una historia industrial

Primero Roubaix, luego el Mundo



OVHcloud comenzó con hardware



Ensamblamos nuestros servidores a partir de los componentes



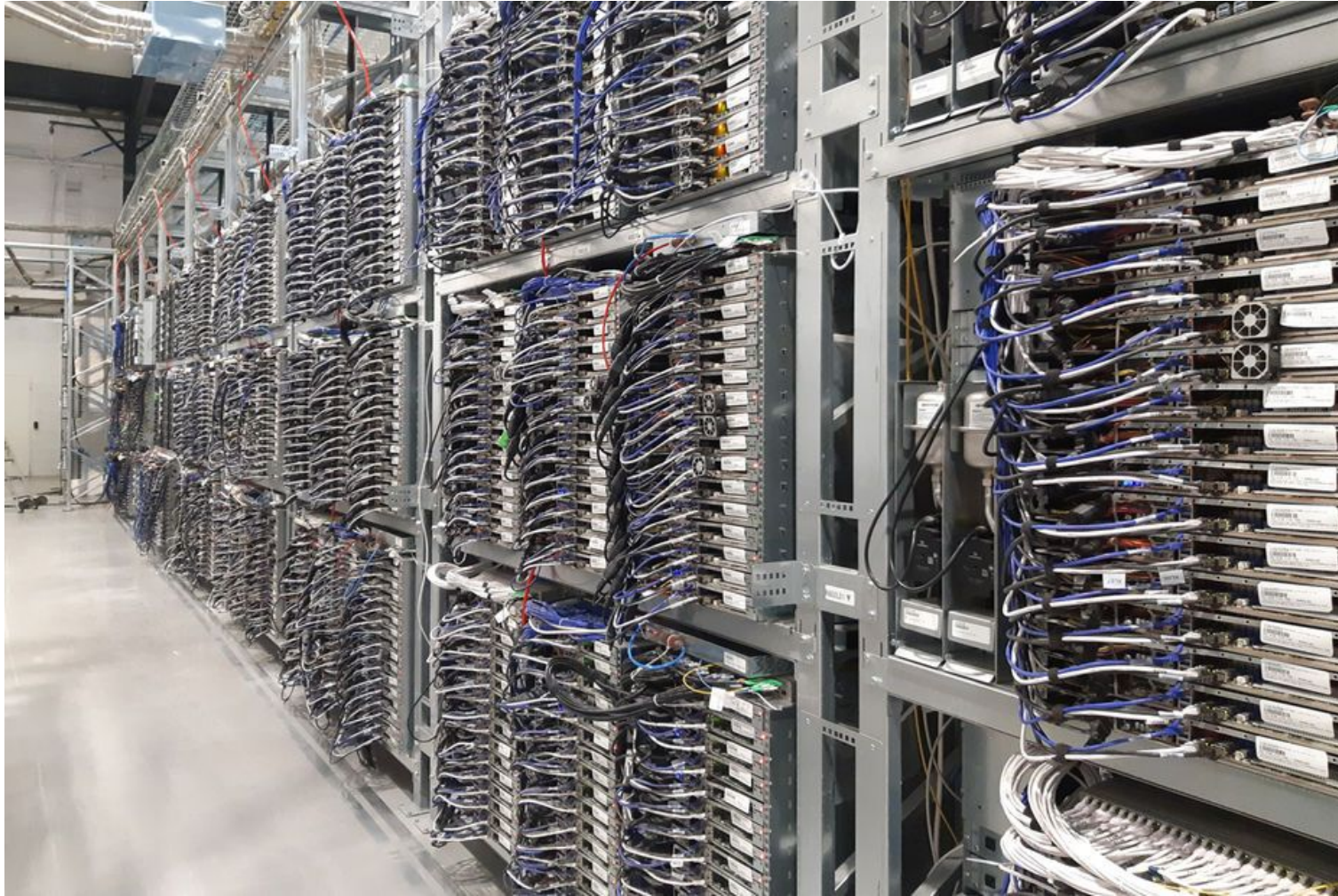
En busca de la eficacia máxima



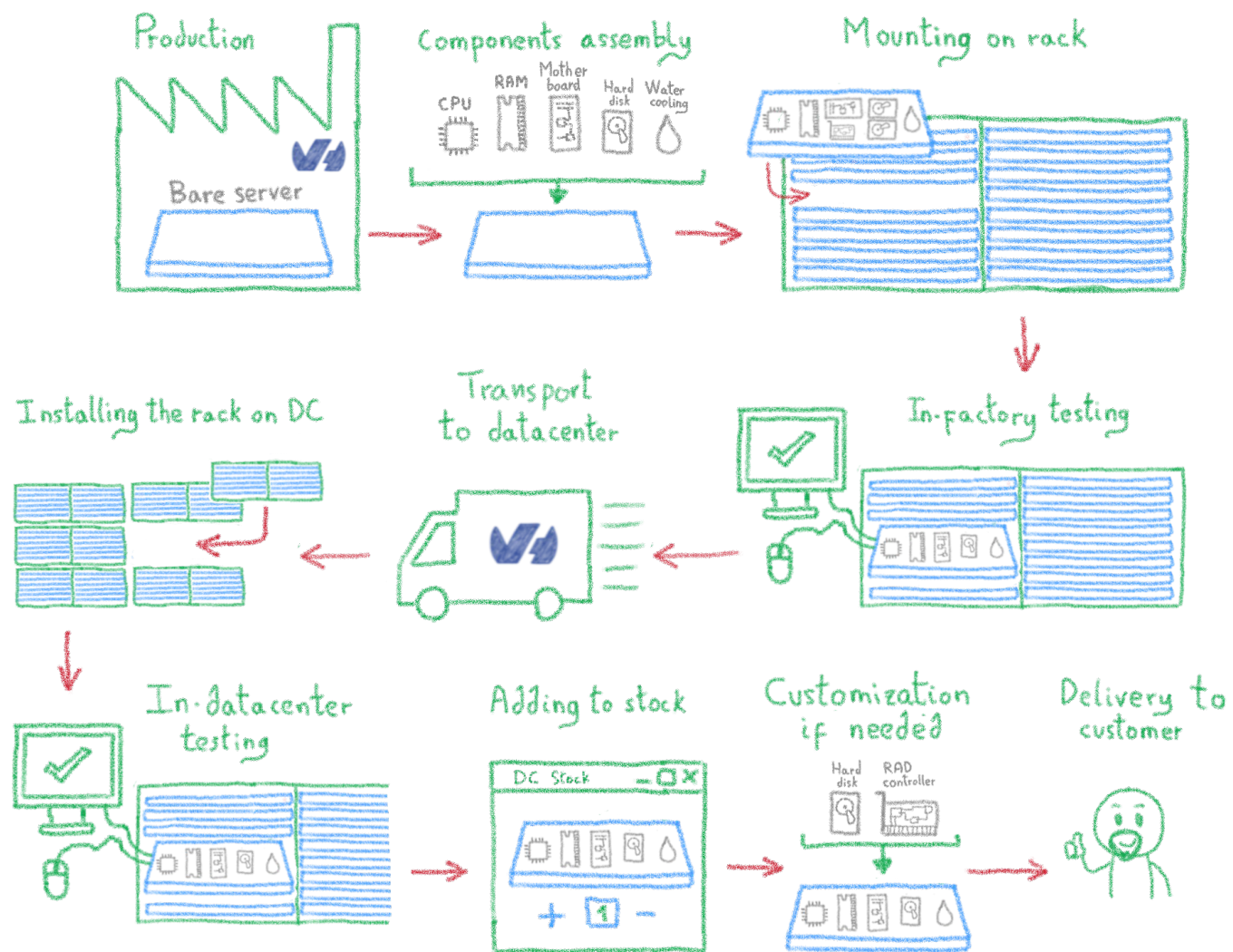
Construimos nuestros propios chasis (horizontales)



Desde componentes hasta centros de datos



Controlamos nuestra cadena industrial

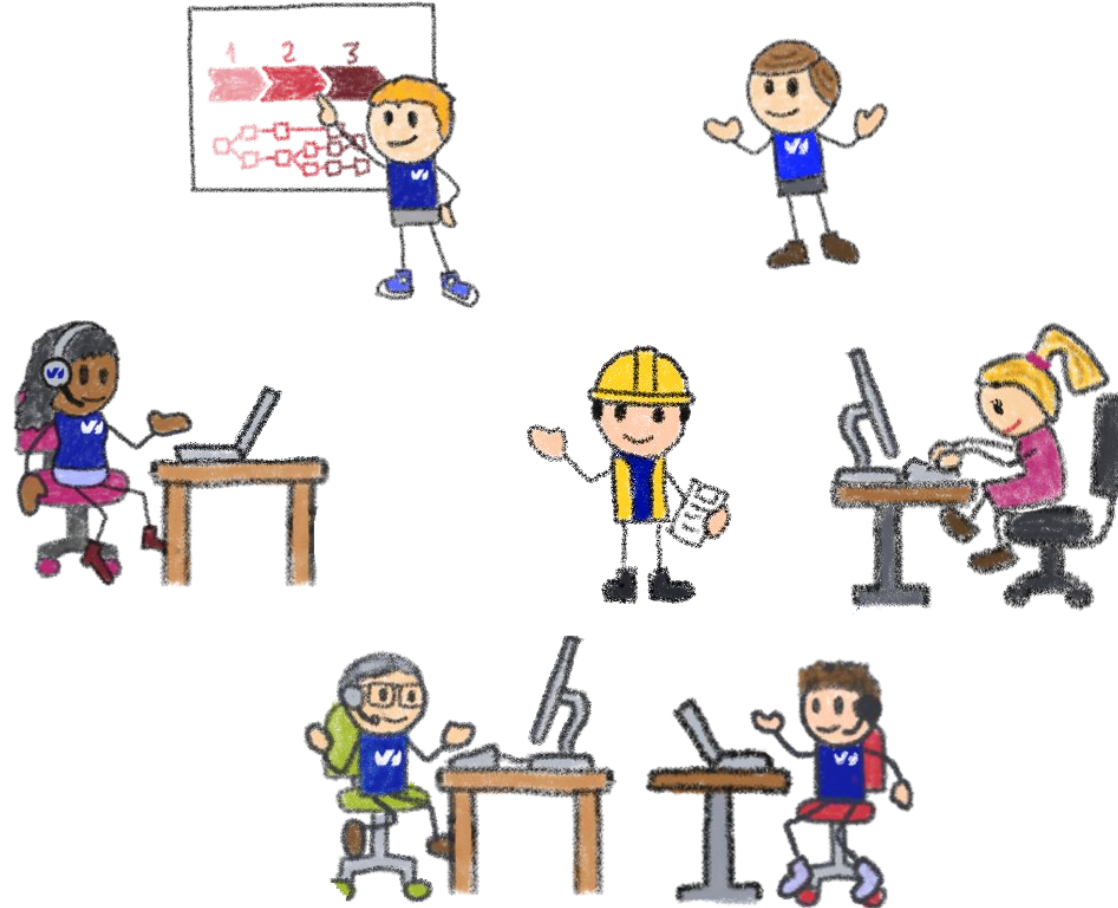


¿Pero tío, por qué nos cuentas esto?

Esta es una charla sobre IA y open source, ¿no?



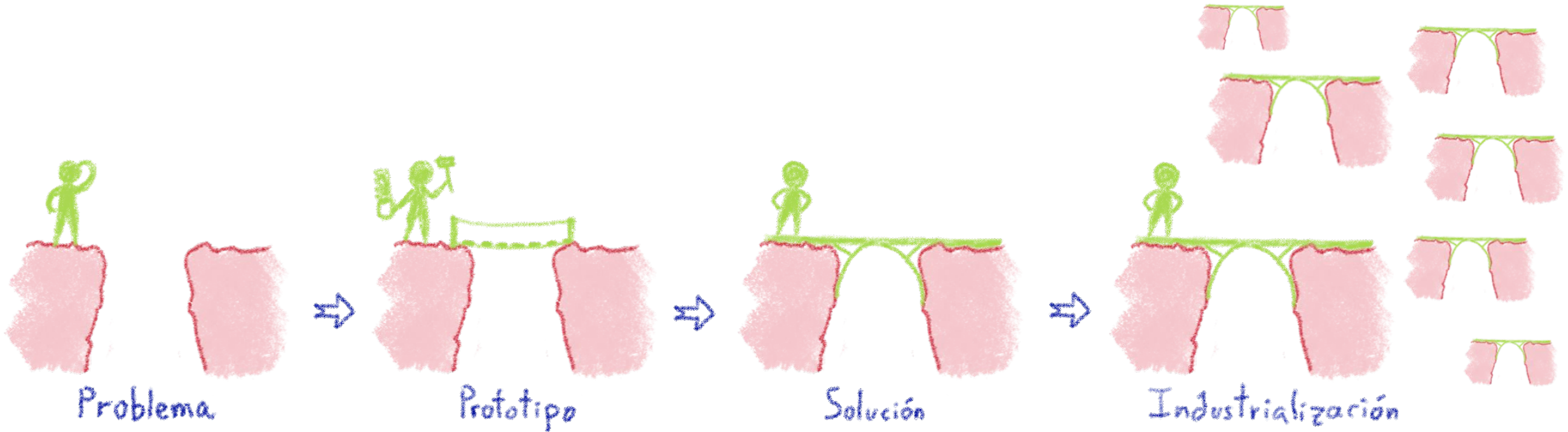
Una cultura industrial



En toda la empresa, incluyendo el software



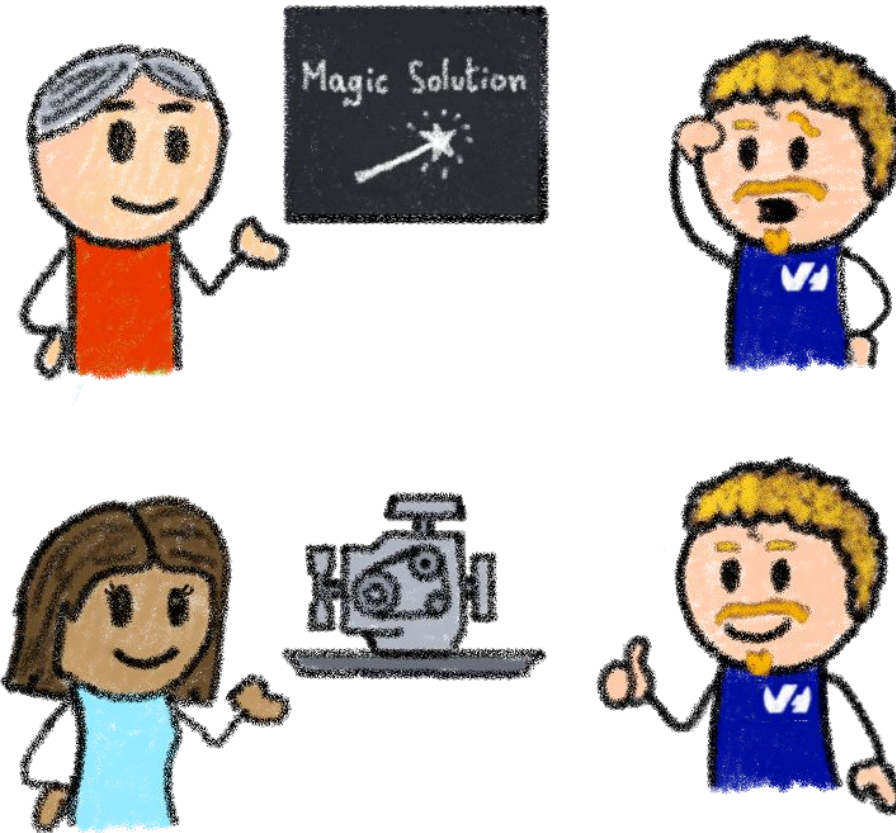
Atacamos los problemas de manera industrial



Buscando cómo generalizar, como industrializar



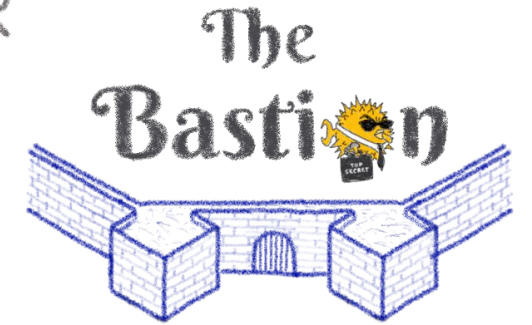
No nos gustan las cajas negras



Nos gusta ver lo que hay dentro...



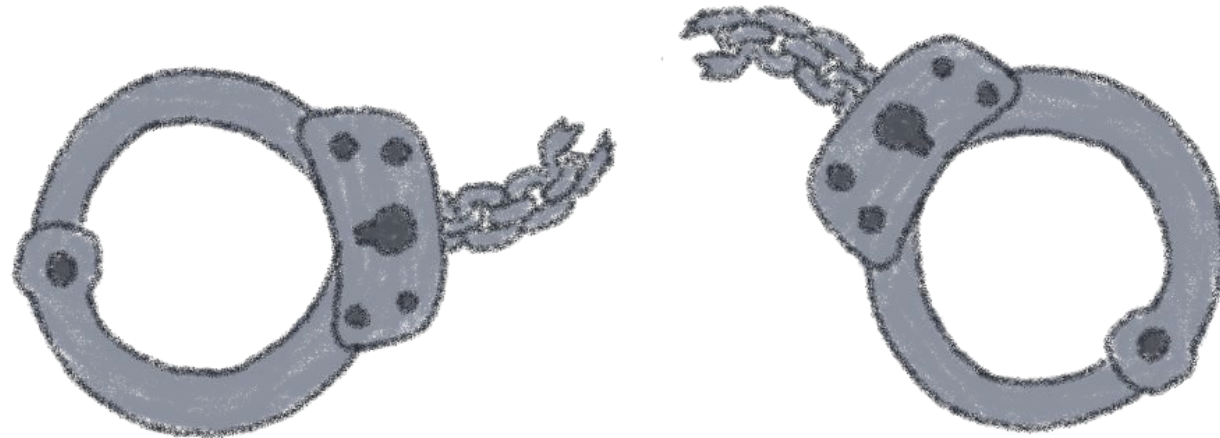
Creemos a fondo en el software libre



Como usuarios, colaboradores, mantenedores y creadores



Para nosotros y para nuestros clientes



La reversibilidad es uno de nuestros valores fundamentales



Y cuando la IA se cruzó en nuestro camino

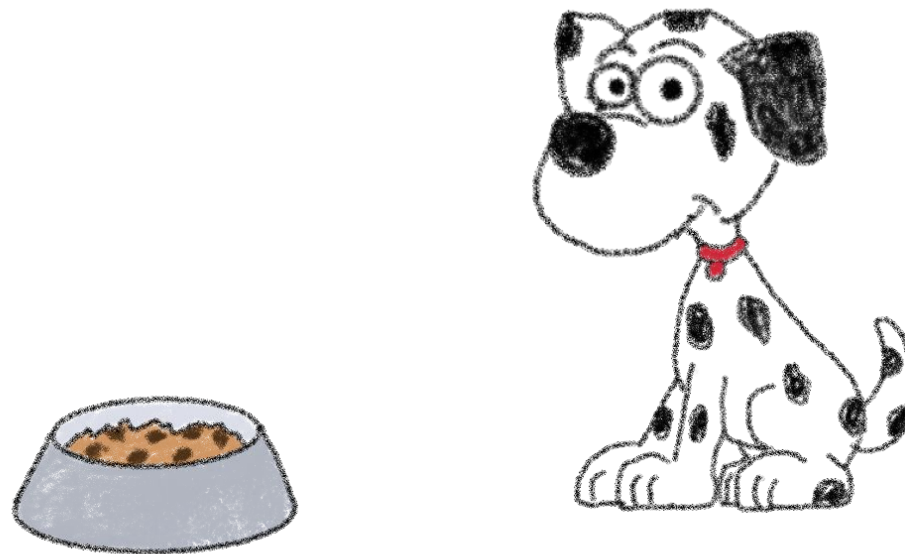


Aplicamos los mismos principios



El principio de la IA en OVHcloud

Eat your own dogfood, como dicen los americanos...



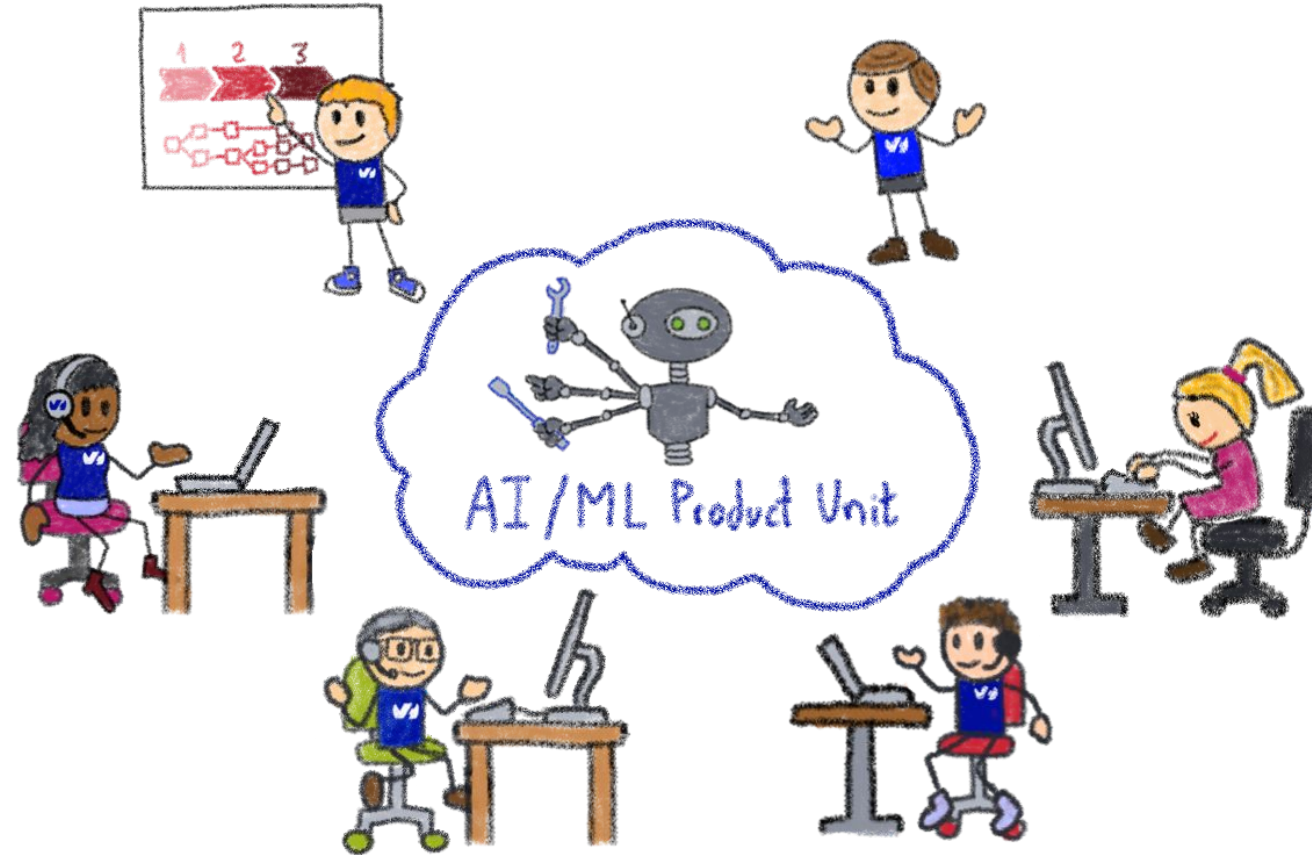
Problemas que pedían soluciones de IA



O al menos un poco de Machine Learning...



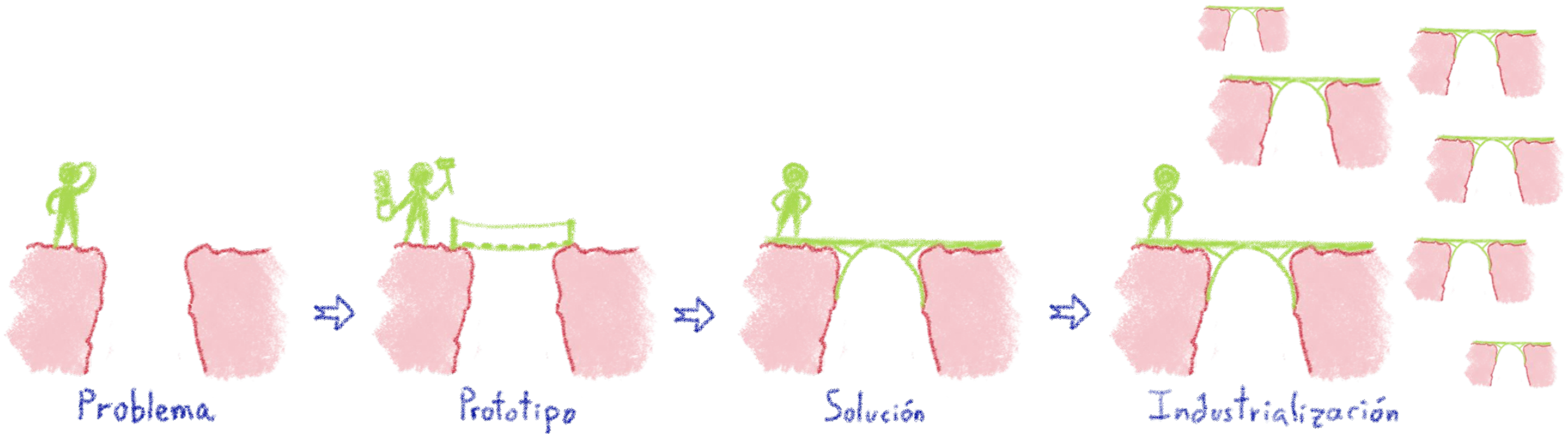
Creamos un equipo IA/ML



Y empezaron a atacar los problemas

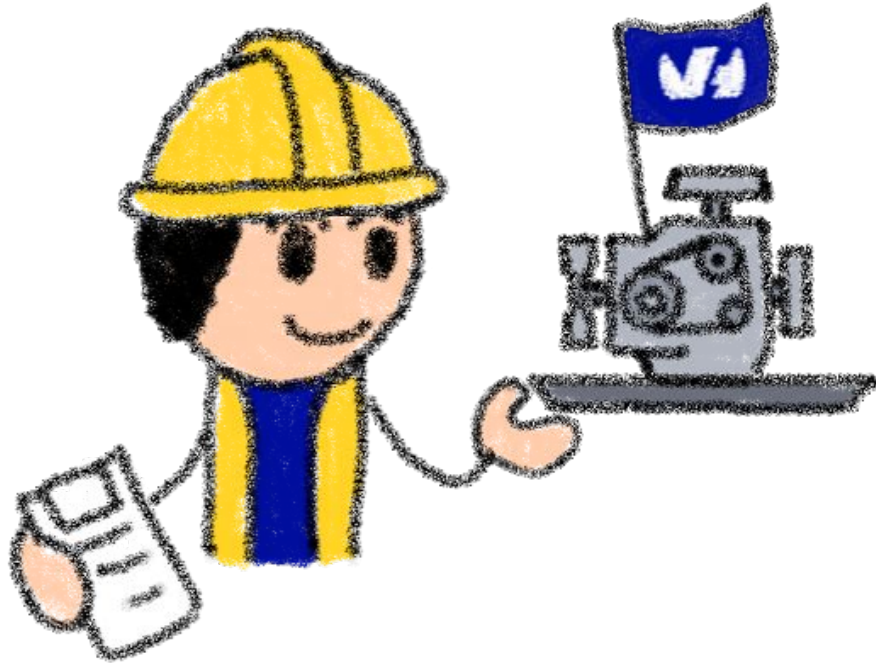


La solución debía ser industrializable



Si no, no nos serviría para nada

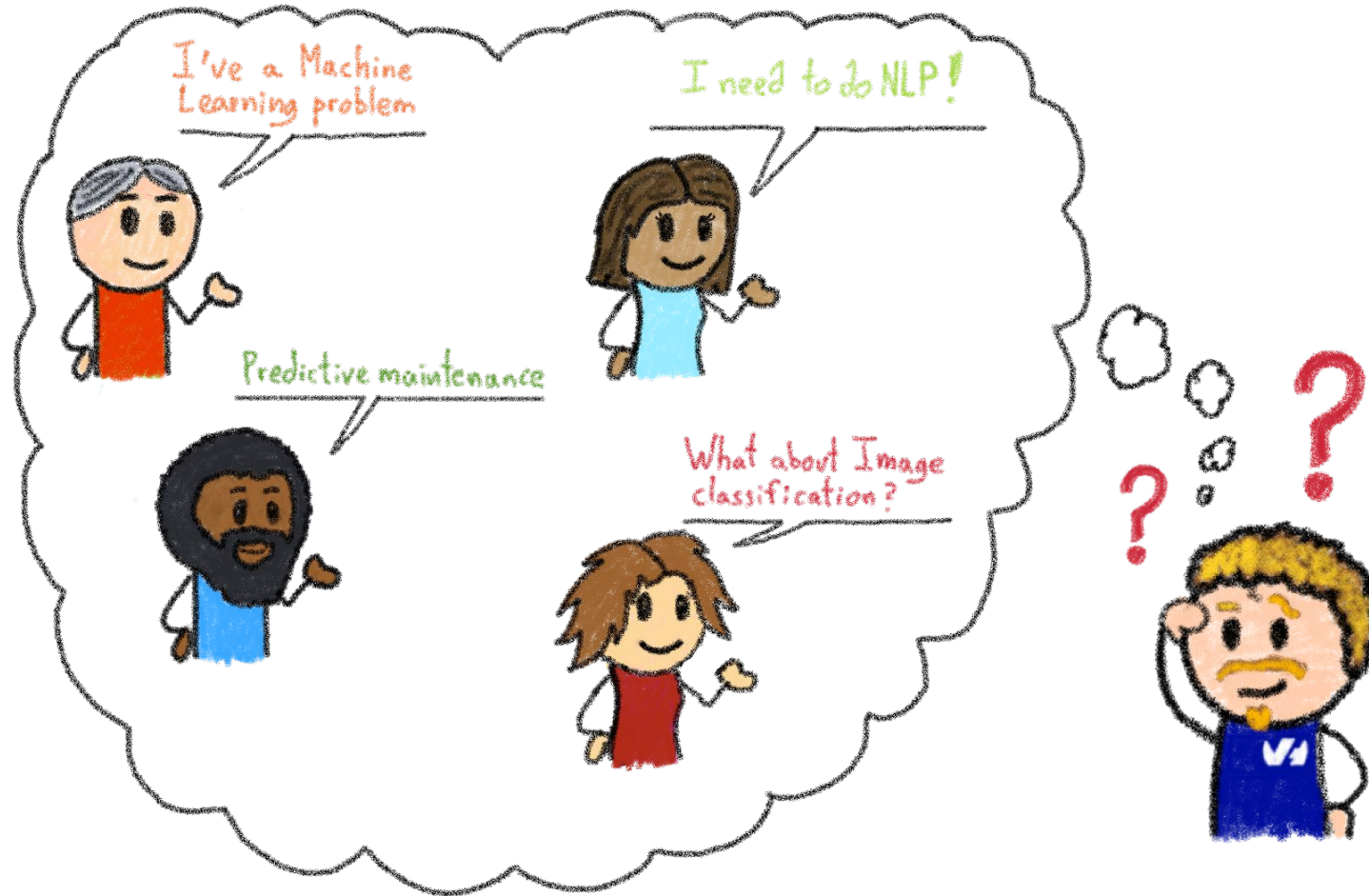
Las primeras soluciones en producción



¡Y funcionaban!



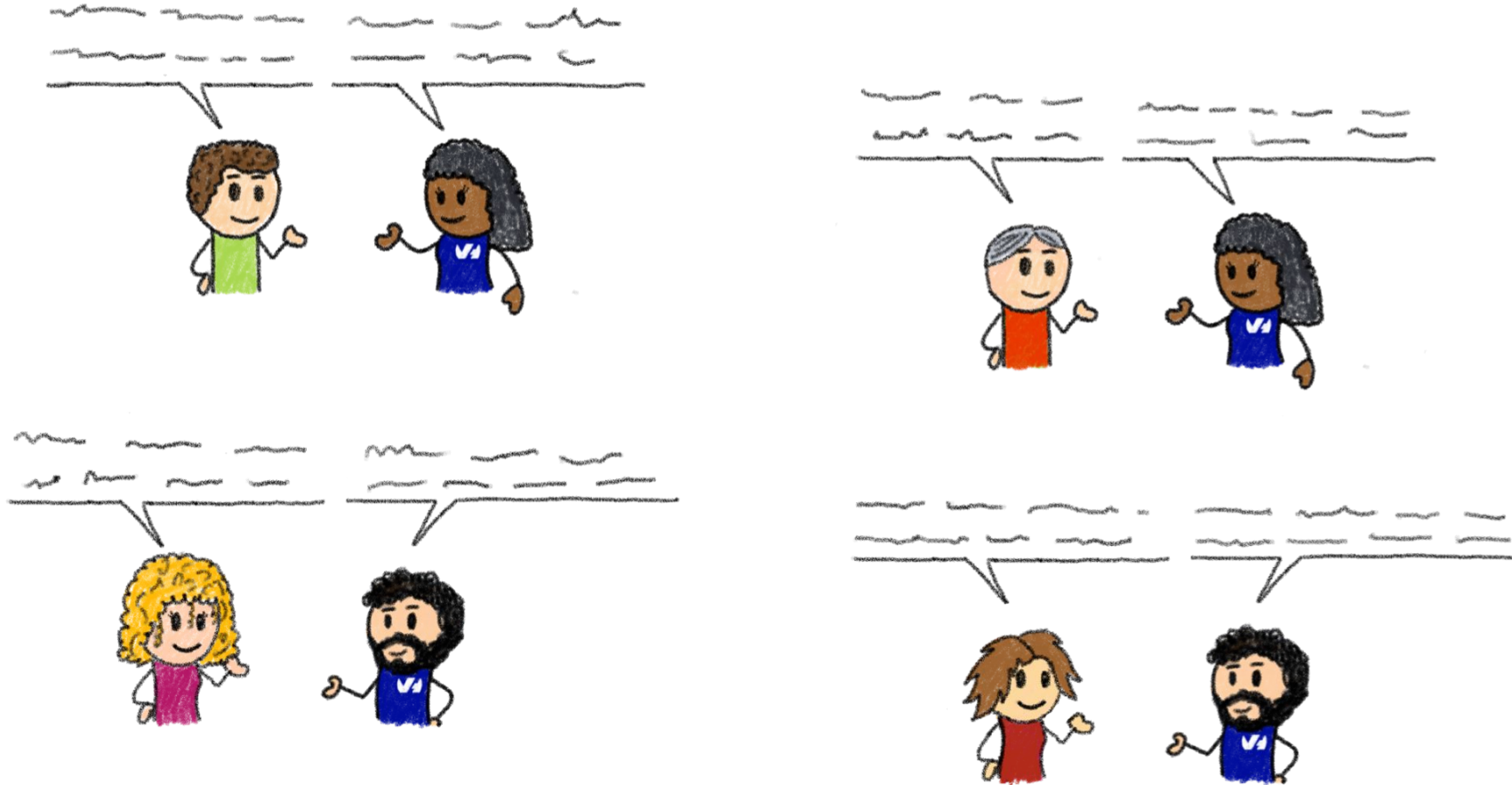
¿Y ahora qué?



Quizás nuestros clientes tenían problemas semejantes...



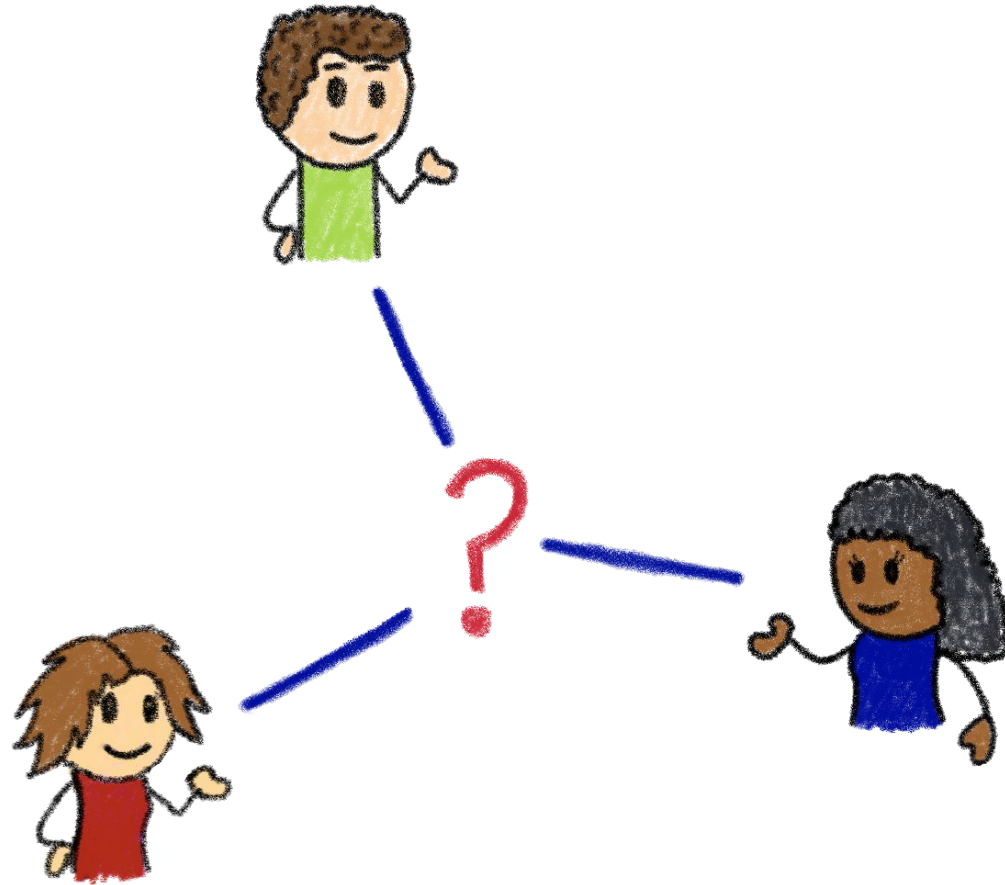
¡Hablemos con el ecosistema!



Clientes, partners, startups...



El eslabón perdido

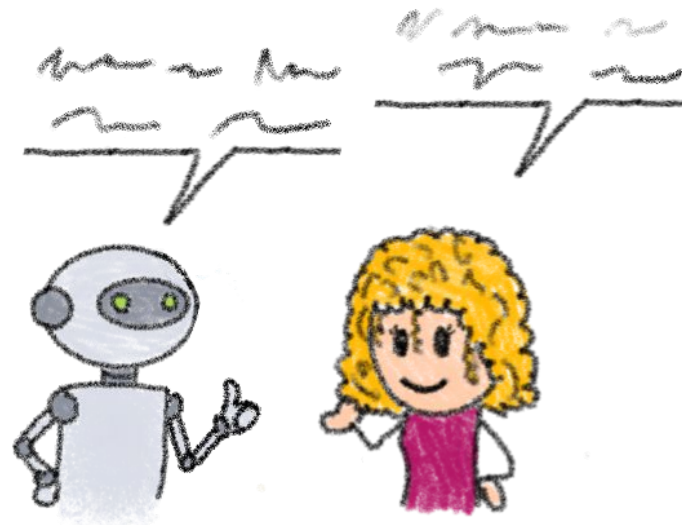


¿Y si fuera ahí donde podíamos innovar?



Las múltiples facetas de la IA

Et de la gente que trabaja con ella



Dos roles clásicos en el mundo de la IA

A



Doing AI

Startups, empresas tech
Sector público, gobiernos,
Provedores de servicio

Tengo científicos de datos en plantilla
Necesito herramientas de IA para desarrollar mi proyecto

B



Using tools
"powered by IA"

Busco productos listos para usar
impulsados por la IA. Un chatbot,
la detección de fraude, un motor de búsqueda

Y, claro está:

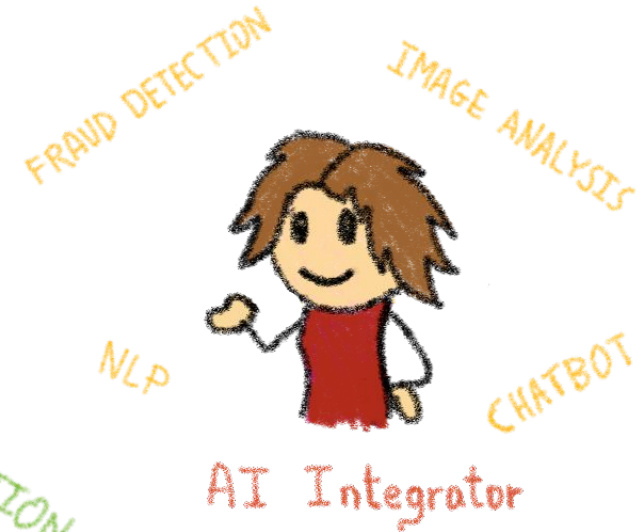
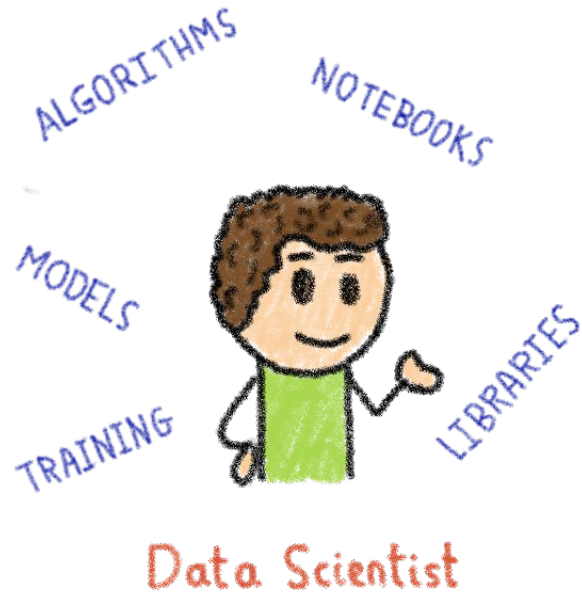
- Simple
- Seguro
- Respetando privacidad
- A un precio controlado



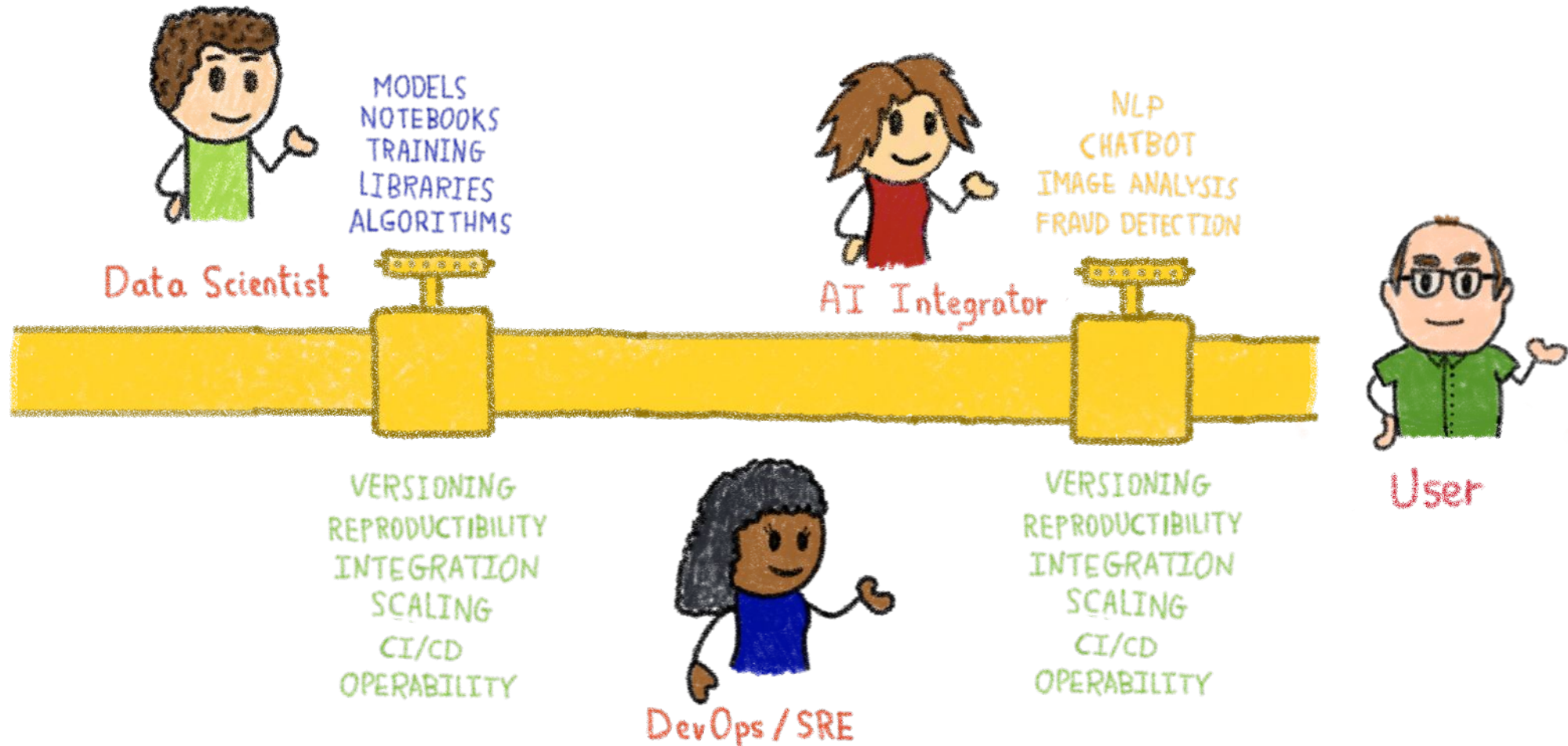
Pero hay un tercero: el DevOps/SRE/DataOps



Los tres hablan idiomas diferentes...



Pero necesitan trabajar juntos



El reto de la integración



Integrar los equipos, procesos y herramientas
de IA/ML & DevOps/SRE/DataOps



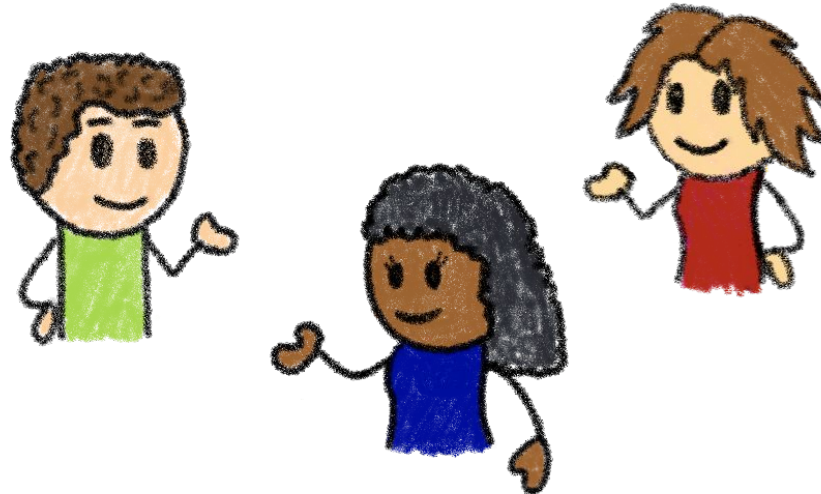
Muchas cuestiones a las que responder

We need to setup / use
AI environments
easily & quickly.
Infrastructure?
Not our job.

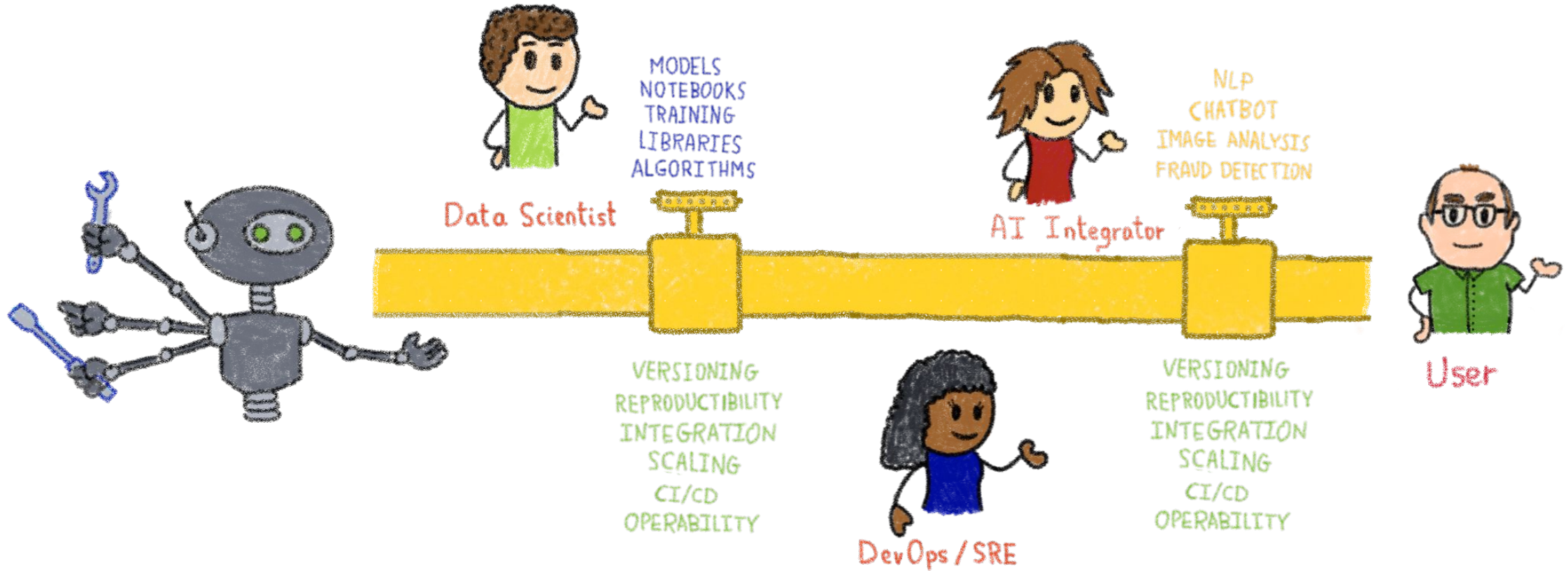
Our team needs to
collaborate and have
guaranteed access to
resources.

Our needs for resources
are evolving.
We need scalability and
pay as we grow.

How do we guarantee
the security and
compliance required
for critical or sensitive
data?



Automatizar el pipeline de un extremo al otro

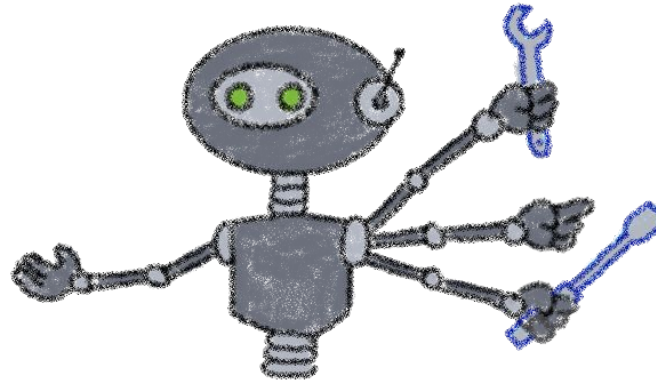


De la idea a la producción



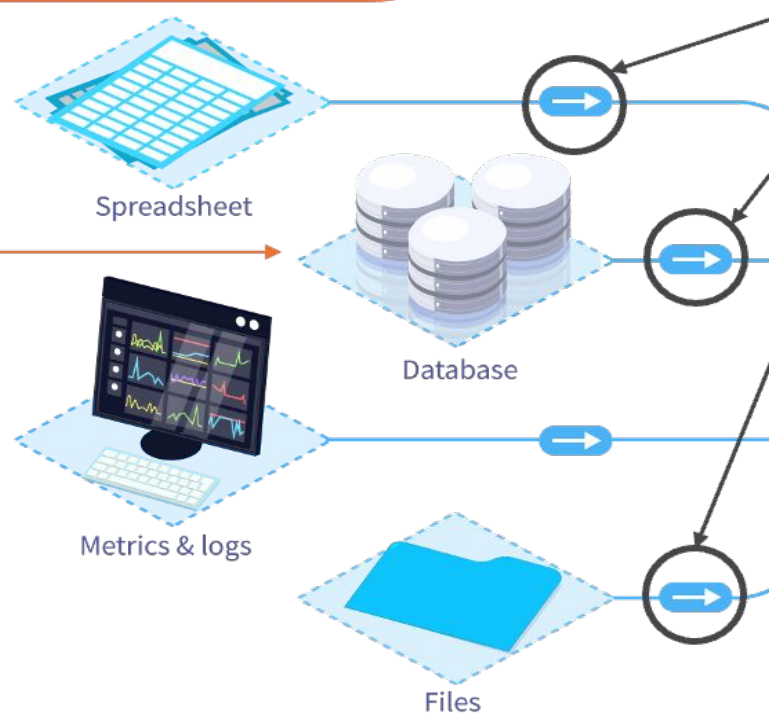
OVHcloud & IA

Nuestra respuesta a la automatización del pipeline



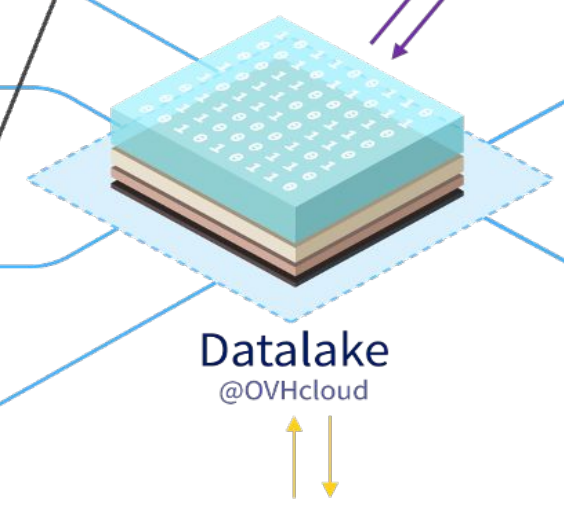
La IA con parte de nuestra gama Data

New Database as a Service



New Ingestion
kafka

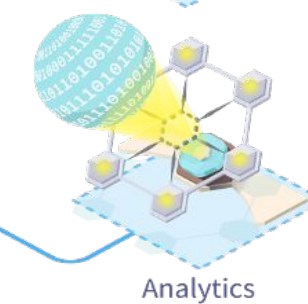
Data Processing
APACHE Spark



High Perf Object Storage



AI Tools

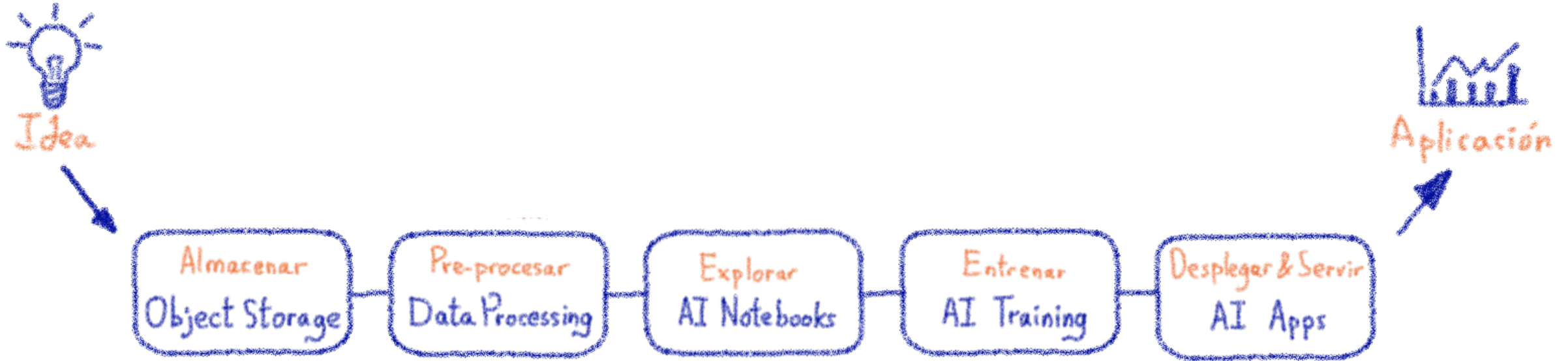


COMING SOON

AI Notebooks
AI Training
AI Apps



¿Qué es OVHcloud AI Solutions?



AI Solutions are a complete set of **managed services** that provide a **quick and easy way** for your data scientists and developers to get **ML models from idea to production**



Las etapas del pipeline de IA



Las etapas del pipeline de IA

Almacenar
Object Storage

- **Centralize your data** within a data lake such as **Object Storage**, to benefit from high throughput with reduced latency



Las etapas del pipeline de IA



- **Centralize your data** within a data lake such as **Object Storage**, to benefit from high throughput with reduced latency
- **Clean and process your data** with **Data Processing** solution, powered by Apache Spark



Las etapas del pipeline de IA



- **Centralize your data** within a data lake such as **Object Storage**, to benefit from high throughput with reduced latency
- **Clean and process your data** with **Data Processing** solution, powered by Apache Spark
- **Develop your AI models** with **AI Notebooks**, by instantly starting a Jupyter or VS Code notebook with GPU/CPU resources



Las etapas del pipeline de IA



- **Centralize your data** within a data lake such as **Object Storage**, to benefit from high throughput with reduced latency
- **Clean and process your data** with **Data Processing** solution, powered by Apache Spark
- **Develop your AI models** with **AI Notebooks**, by instantly starting a Jupyter or VS Code notebook with GPU/CPU resources
- **Train your AI models** with **AI Training**, without complex setup configuration, allowing CPU/GPU parallelization



Las etapas del pipeline de IA



- **Centralize your data** within a data lake such as **Object Storage**, to benefit from high throughput with reduced latency
- **Clean and process your data** with **Data Processing** solution, powered by Apache Spark
- **Develop your AI models** with **AI Notebooks**, by instantly starting a Jupyter or VS Code notebook with GPU/CPU resources
- **Train your AI models** with **AI Training**, without complex setup configuration, allowing CPU/GPU parallelization
- **Deploy your AI models** in production or pre-trained models or even applications very easily with **AI Apps**



Las etapas del pipeline de IA



- **Centralize your data** within a data lake such as **Object Storage**, to benefit from high throughput with reduced latency
- **Clean and process your data** with **Data Processing** solution, powered by Apache Spark
- **Develop your AI models** with **AI Notebooks**, by instantly starting a Jupyter or VS Code notebook with GPU/CPU resources
- **Train your AI models** with **AI Training**, without complex setup configuration, allowing CPU/GPU parallelization
- **Deploy your AI models** in production or pre-trained models or even applications very easily with **AI Apps**

These services are built on our **trusted and secured Public Cloud** based on European values & law and powered by Openstack open source cloud



Levantemos el capó

Con qué y cómo construimos AI Solutions



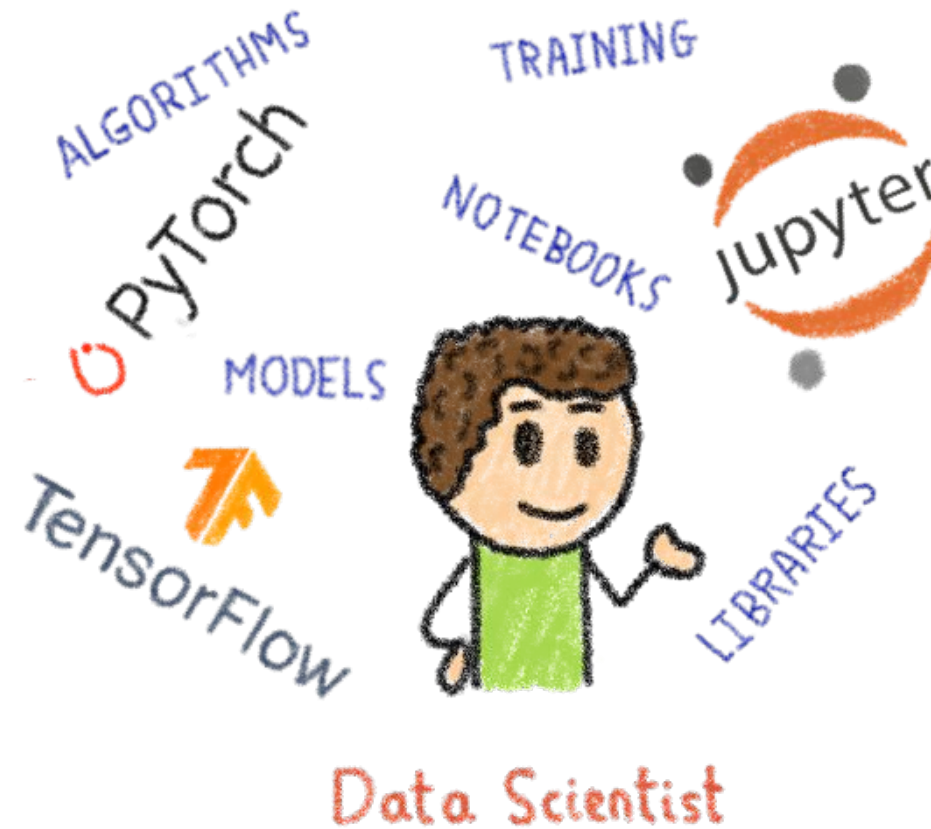
AI Training: entrenar modelos de IA



En nuestra nube, sobre CPUs y GPUs a la demanda



Con las herramientas habituales



La misma manera de hacer que en el laptop





Llegó justo después de ML Serving...



Qué no era lo suficientemente genérico



¿Cómo ser lo suficientemente genéricos?

User tells me they don't want  TensorFlow,
they want to execute  PyTorch with some
custom libraries...

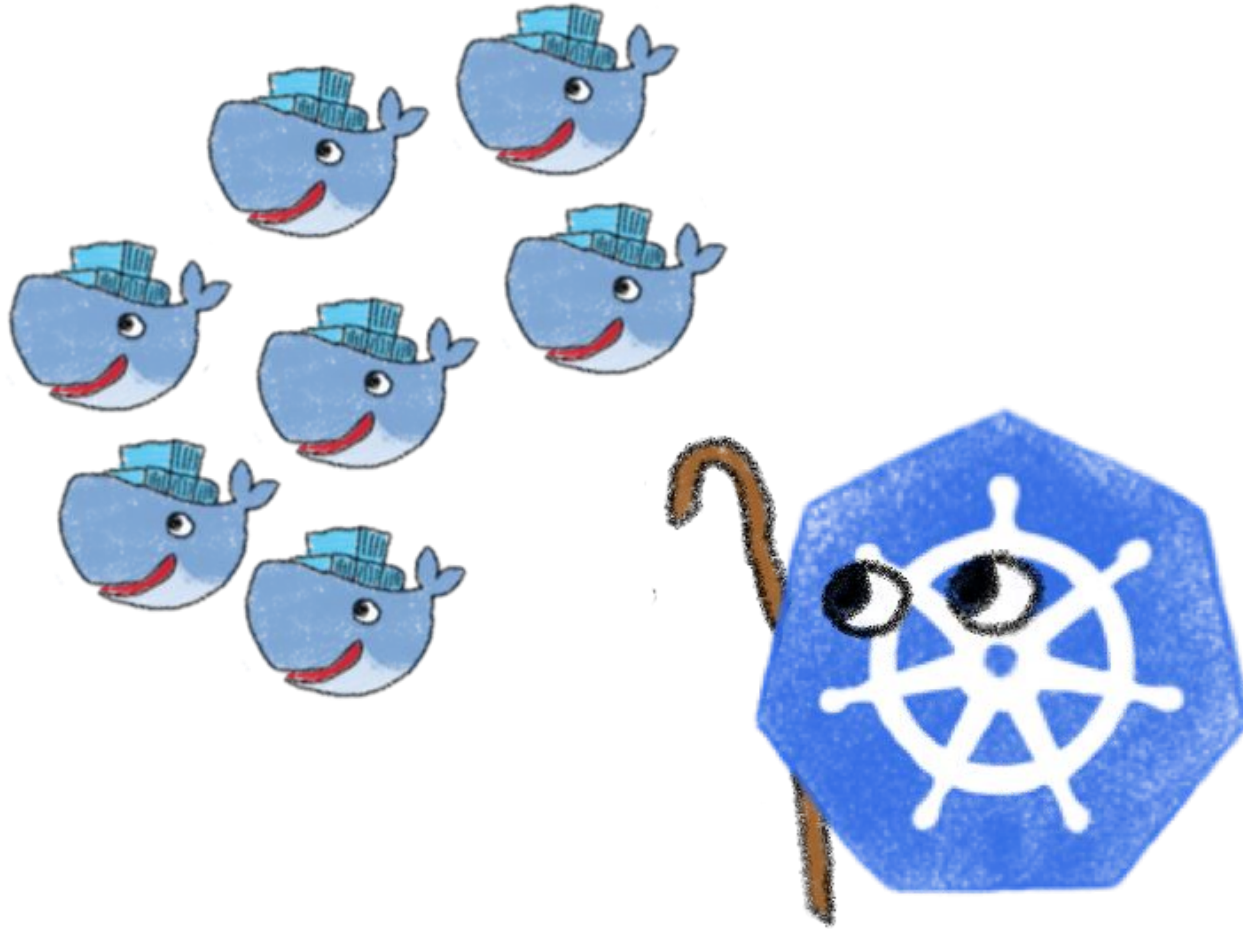


Hmmm... We could simply allow
them to execute  docker

I agree!



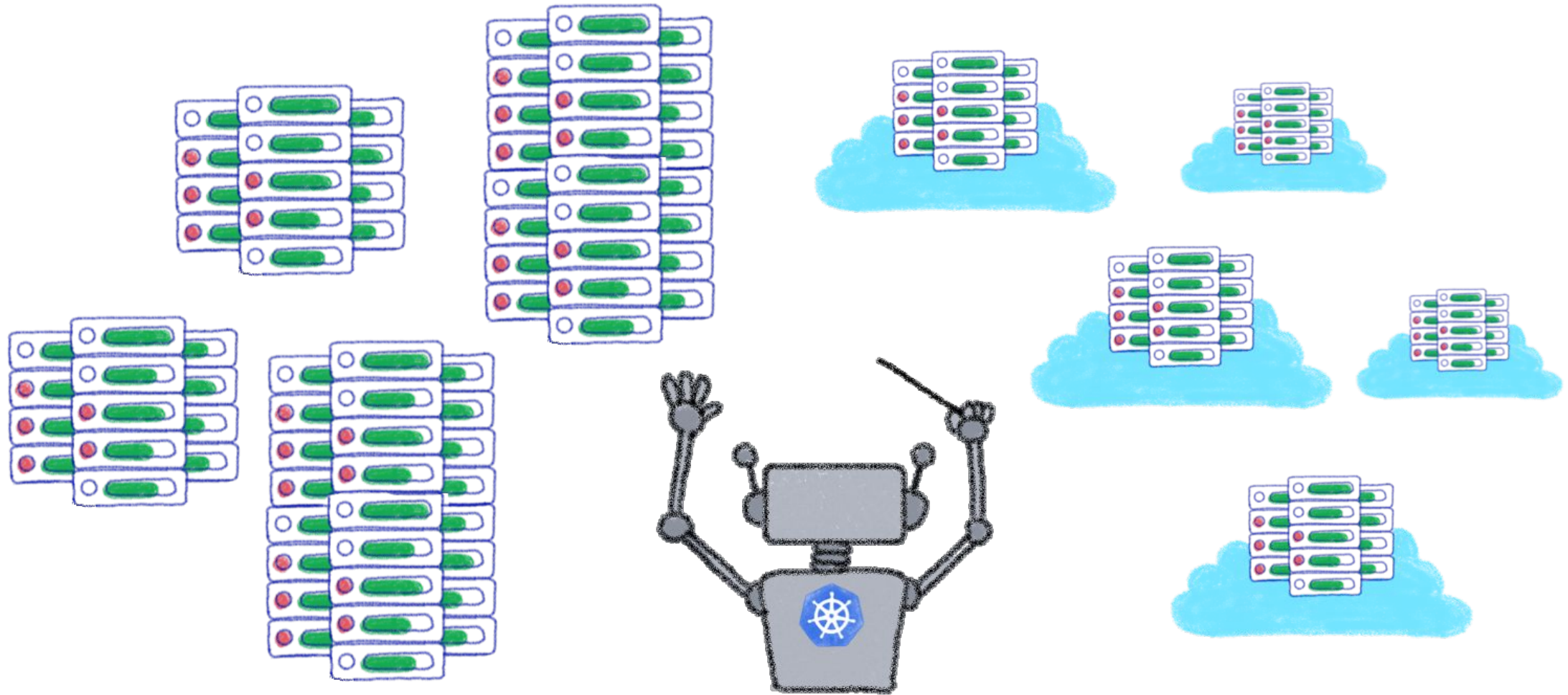
An orchestrator is needed



We simply cannot admin
all those  manually!



Orquestando nodos físicos con CPUs & GPUs



Desplegando los clusters con Terraform

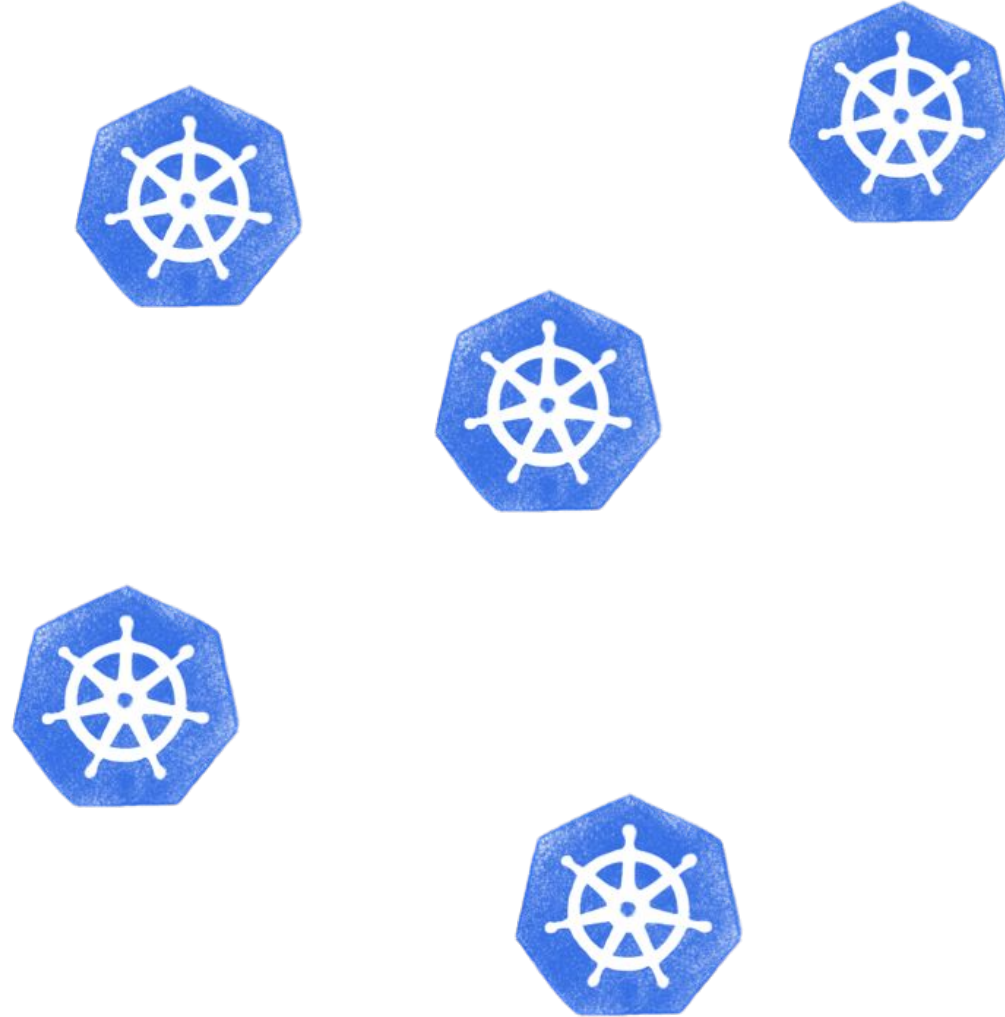
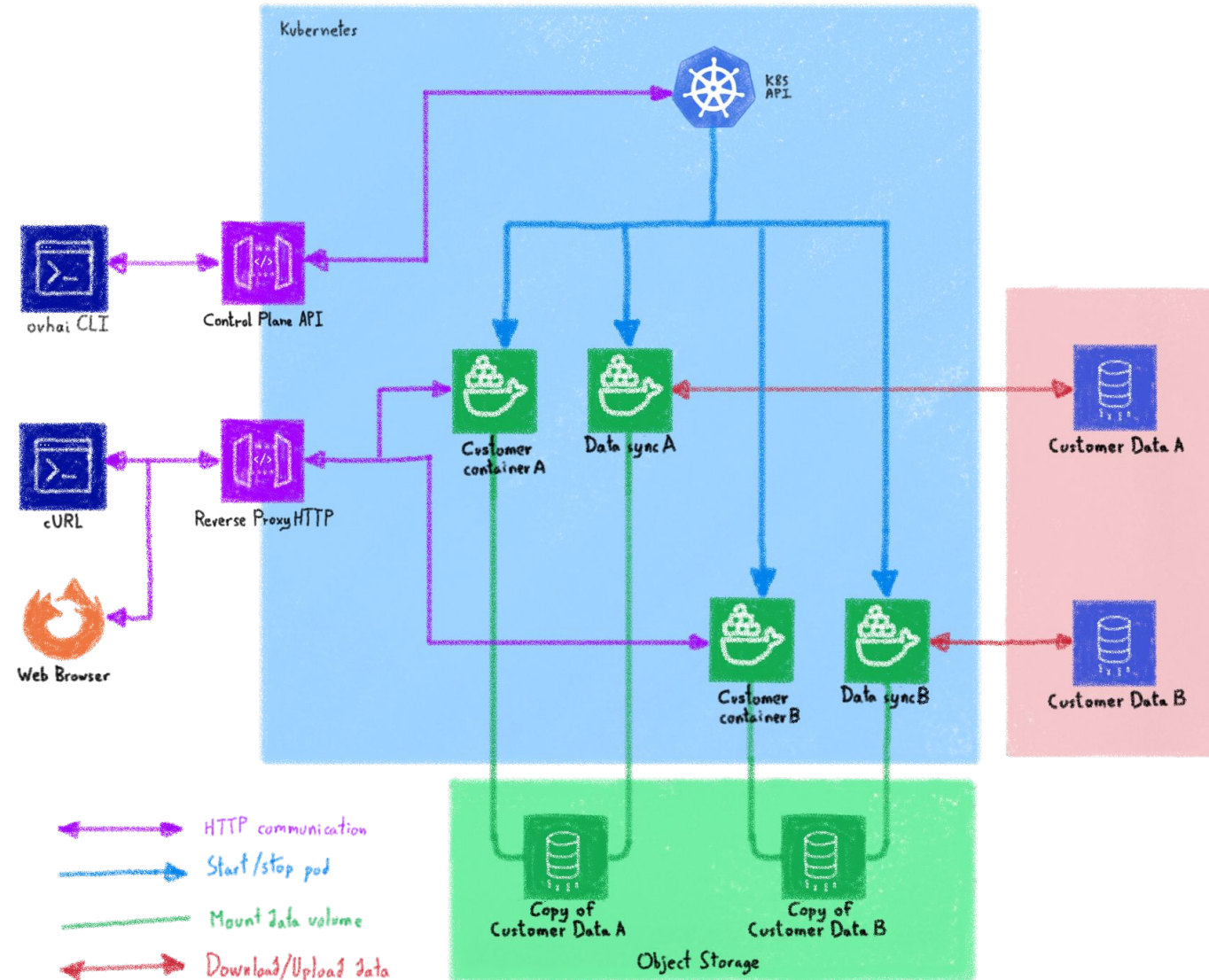


Diagrama de arquitectura



Algunos de nuestros clientes se dieron cuenta

Your AI Training is so flexible
we are using it to run generic batch jobs!



So AI Training under the hood
is a Container as a Service, isn't it?

El producto es tan flexible que puede usarse para otras cosas...



