

Sugestões de como colocar ML em produção

Ricardo Manhães Savii



2M

visits daily

11M

customers

450,000+

SKUs online

6,000+

brands

THE BIGGEST FASHION AND LIFE-STYLE E-COMMERCE GROUP IN LATIN AMERICA

Dafiti was founded in 2011.

Part of Global Fashion Group (GFG) since 2014.

Became Dafiti Group in 2015 when Kanui,
Tricae and Dafiti merged.

GFG GLOBAL
FASHION
GROUP



Dafiti Group R&D Team

Marcio



Rafael



Georg



Ramon



Albert



We are hiring. :)

Will



Ricardo



drop us an email: **dftech@dafiti.com.br**



Um pouco sobre mim

Ricardo Manhães Savii

Eu cresci em São José dos Campos, Brasil e em Vitória-Gasteiz, Espanha.

Eu trabalhei e me especializei em Gestão Hoteleira e Turismo por 7 anos.

Em 2014 eu mudei de caminho e comecei o Bacharel em Ciência e Tecnologia na UNIFESP.



- Pesquisador e Desenvolvedor na **Dafiti Group**.
- Estudante eterno.
- Revisor Nanodegrees na **Udacity** de Machine Learning e Deep Learning.
- Organizador Meetup **PyData São Paulo**.



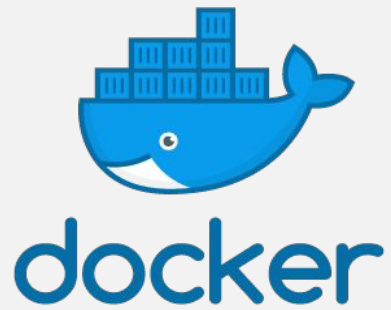
Eu gosto de

- Viajar (o mundo todo e talvez Marte xD).
- Estudar outras línguas.
- Tocar violão.
- Jogos (xadrez, Starcraft2, cs-go...).

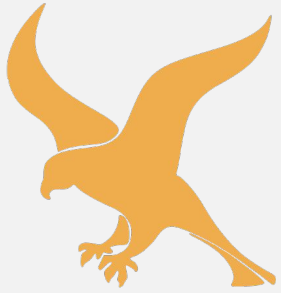
INTENÇÃO DE HOJE

1. Tecnologias para API
2. Sugestões de uso de API
3. Na prática
4. Aprendizado e cuidados

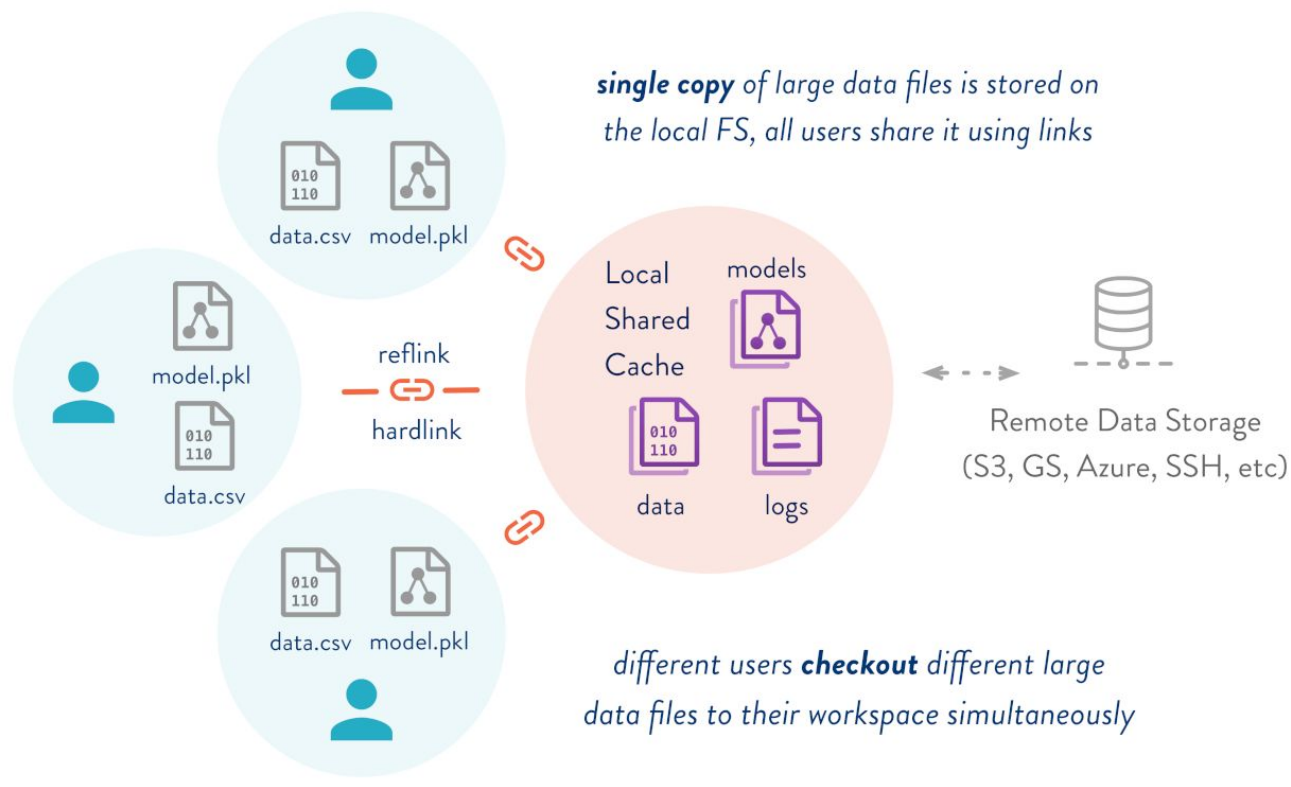
1. TECNOLOGIAS PARA API



TECNOLOGIAS
ENVOLVIDAS



DATA VERSION CONTROL



DATA VERSION CONTROL

ADD YOUR DATA

```
dvc add images.zip
```

ADD REPOSITORY LOCATION (E.G. AWS S3)

```
dvc remote add myrepo s3://mybucket
```

PUSH/PULL THE LOCATION SPECIFIED

```
dvc push
```

```
dvc pull
```

COMMIT DATA INPUT, MODEL OUTPUT AND CODE AND REPRODUCE

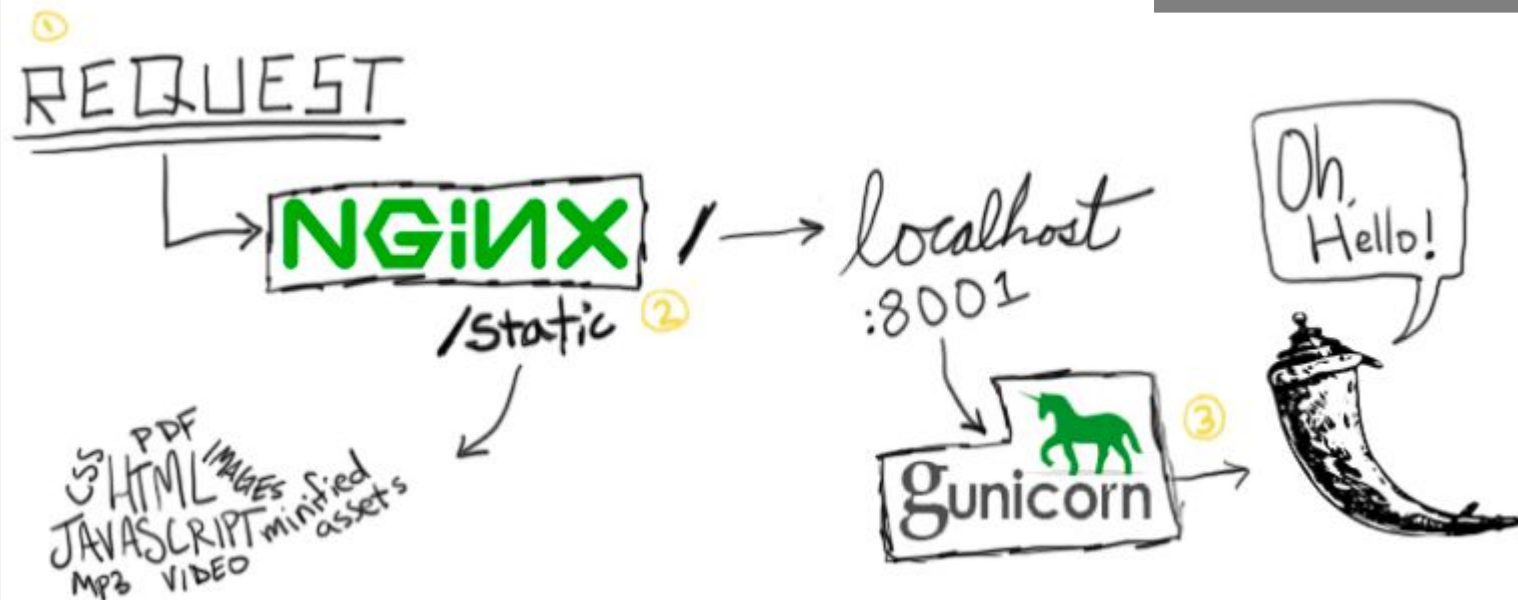
```
dvc run -d images.zip -o model.p ./cnn.py
```

```
dvc repro
```

PYTHON API



PYTHON API

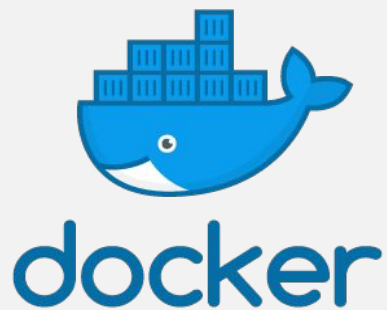




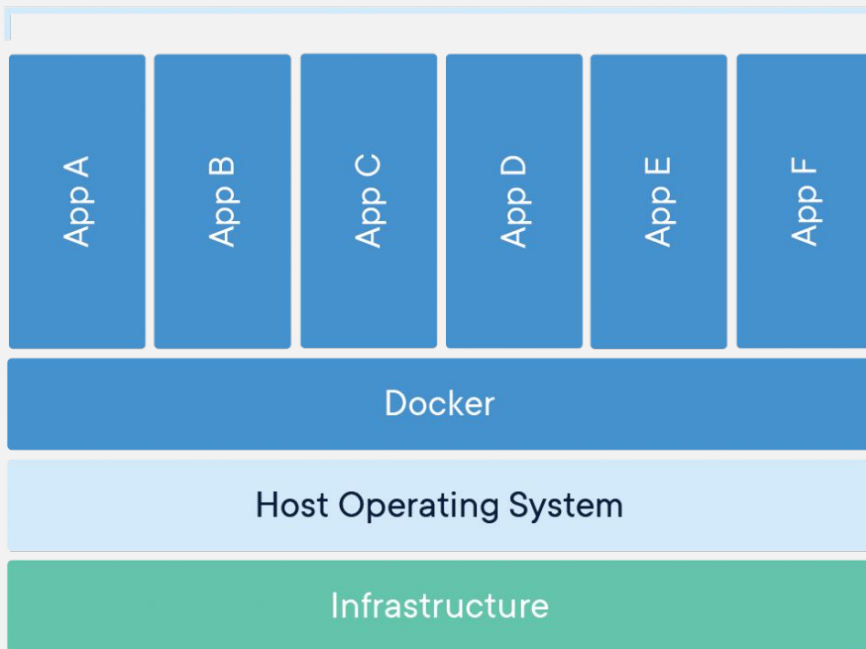
Vs



?





Containerized Applications




CONTAINERS,
SIMPLY
DOCKER

SHOW ME THE
CODE





**ricoms** ▼

Repositories

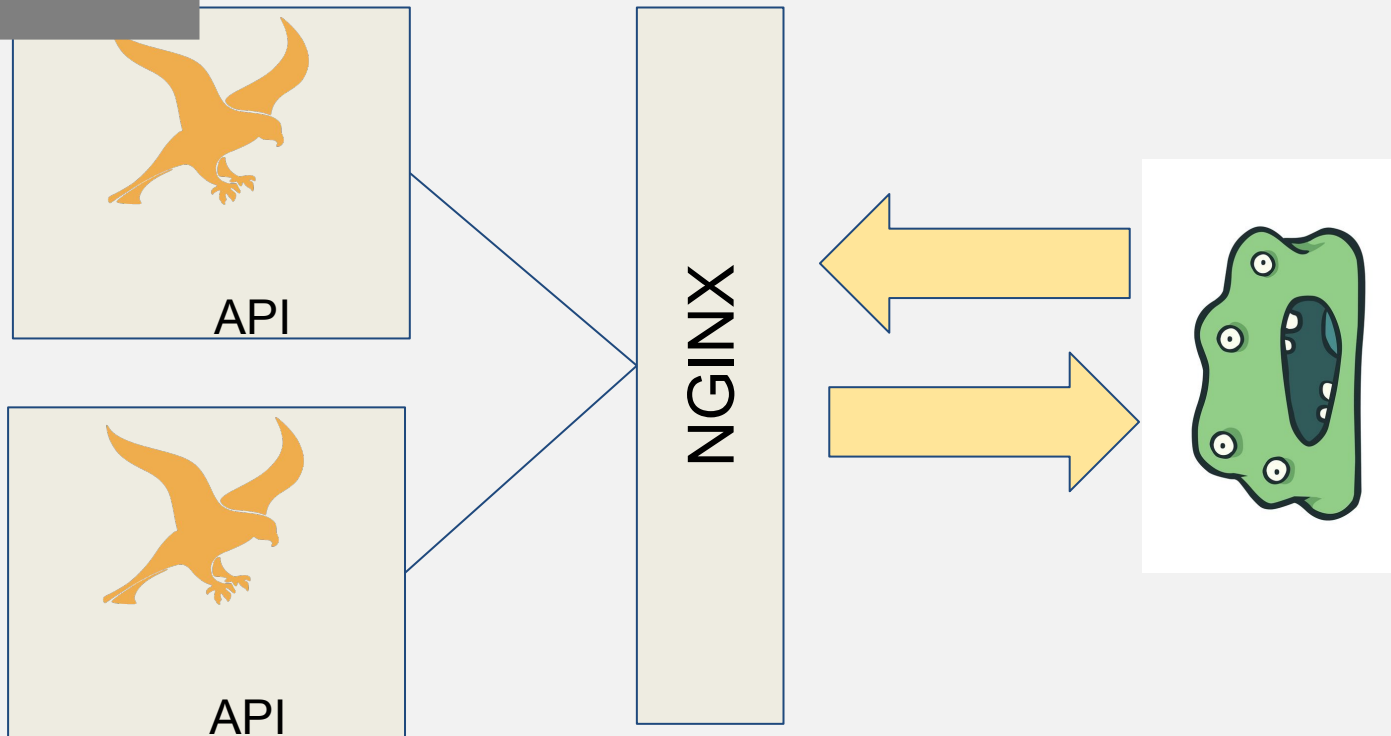
ml-api-template

Machine learning api template

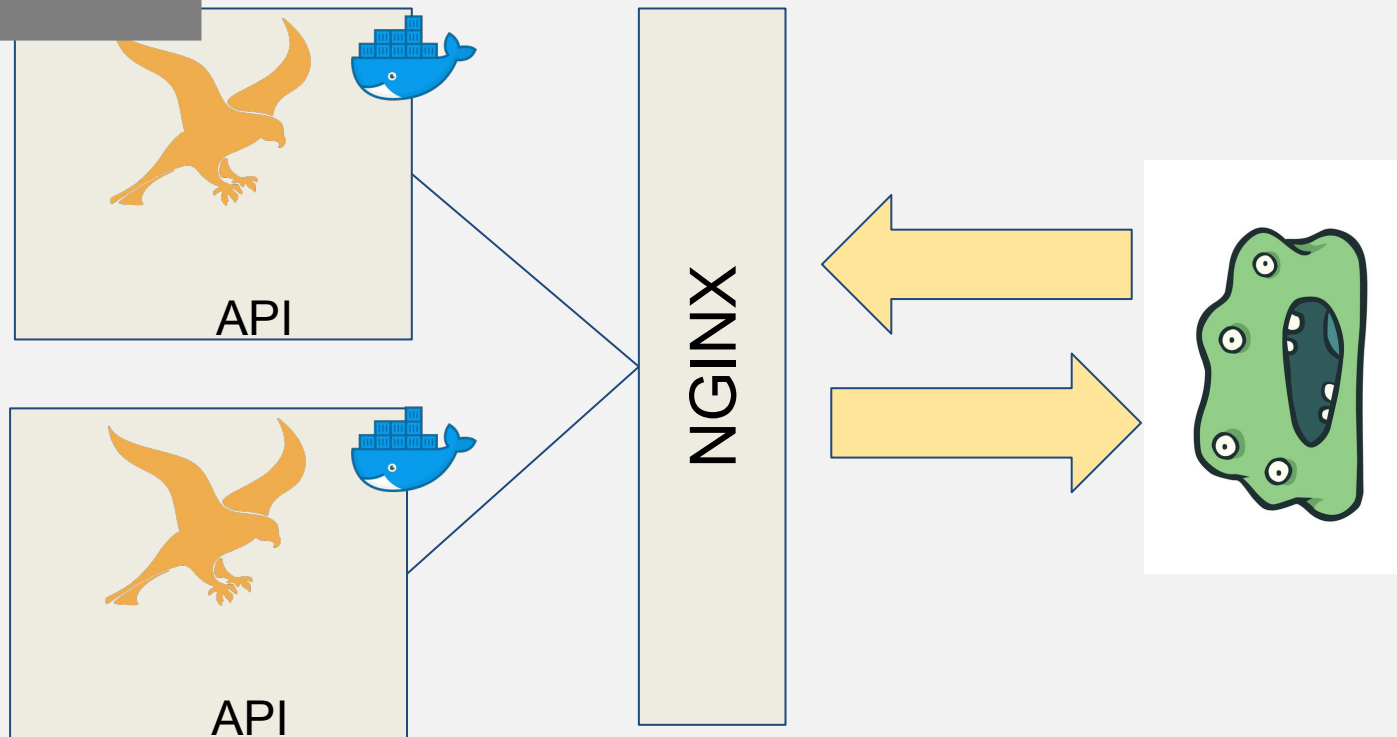
 Python  MIT License Updated on Jun 21

2. SUGESTÕES DE API

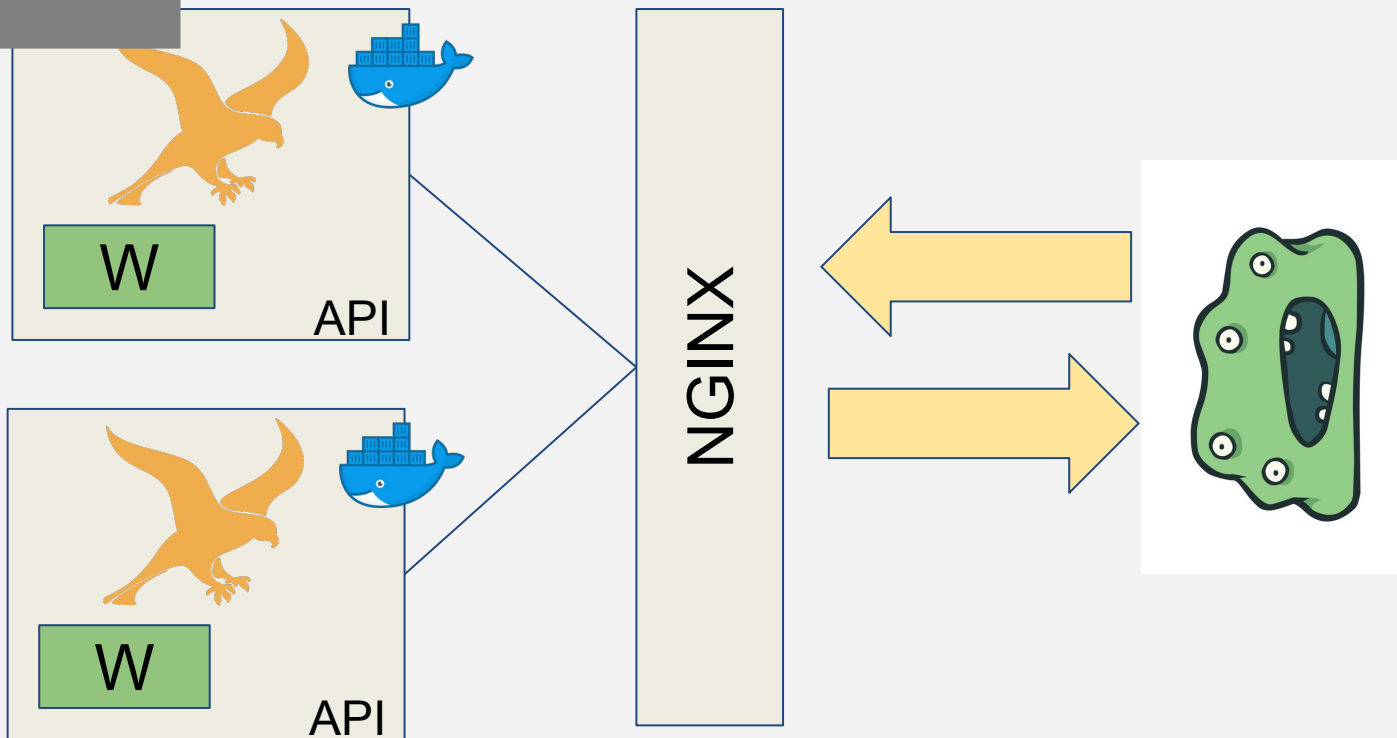
USE-CASE API SIMPLE



USE-CASE API SIMPLE

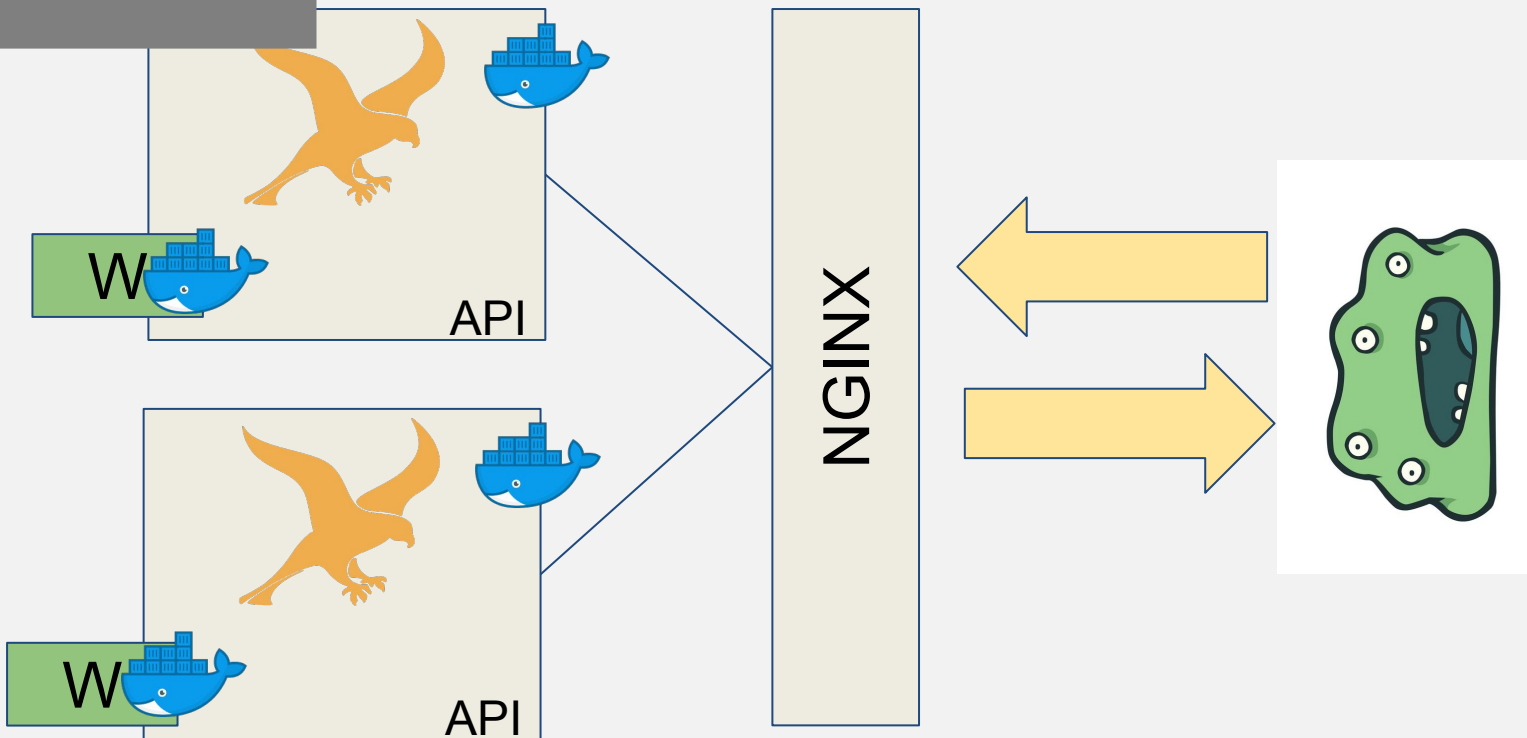


USE-CASE
API ~ SIMPLE



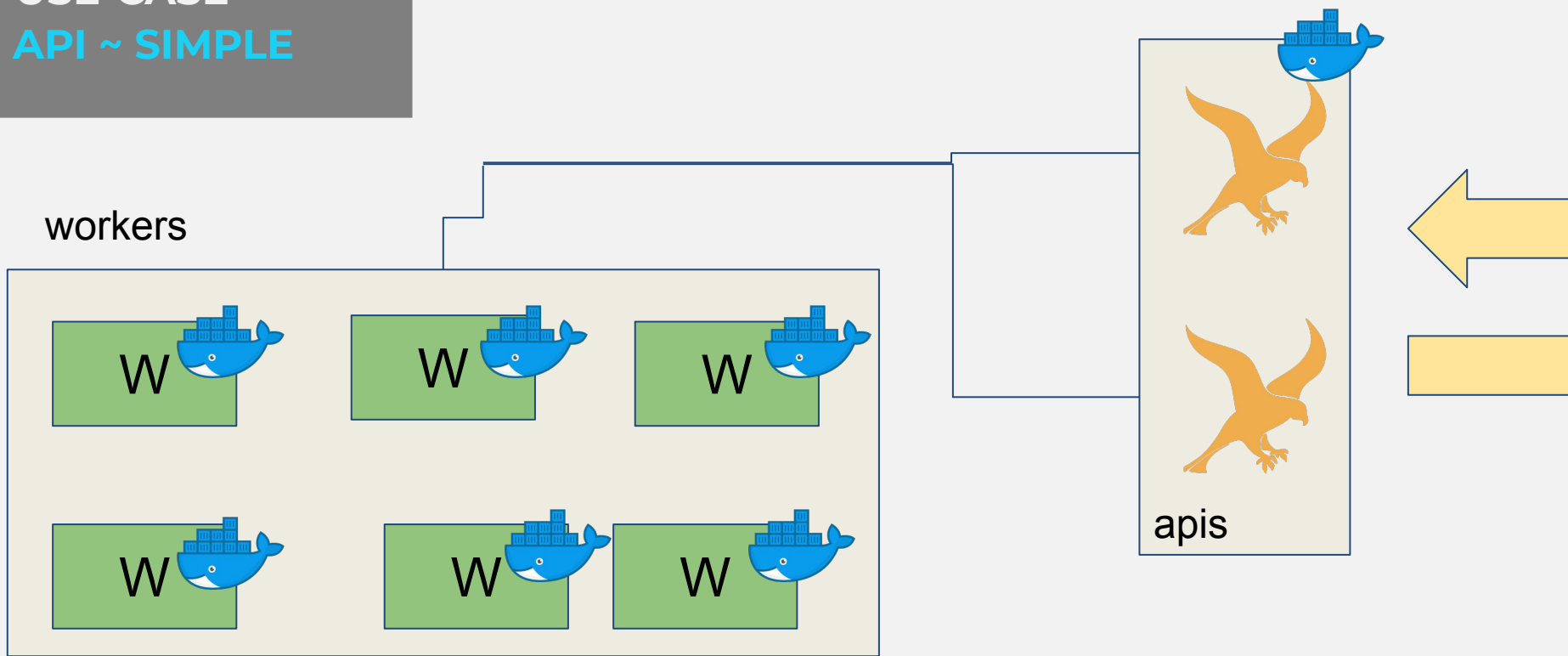
USE-CASE

API ~ SIMPLE



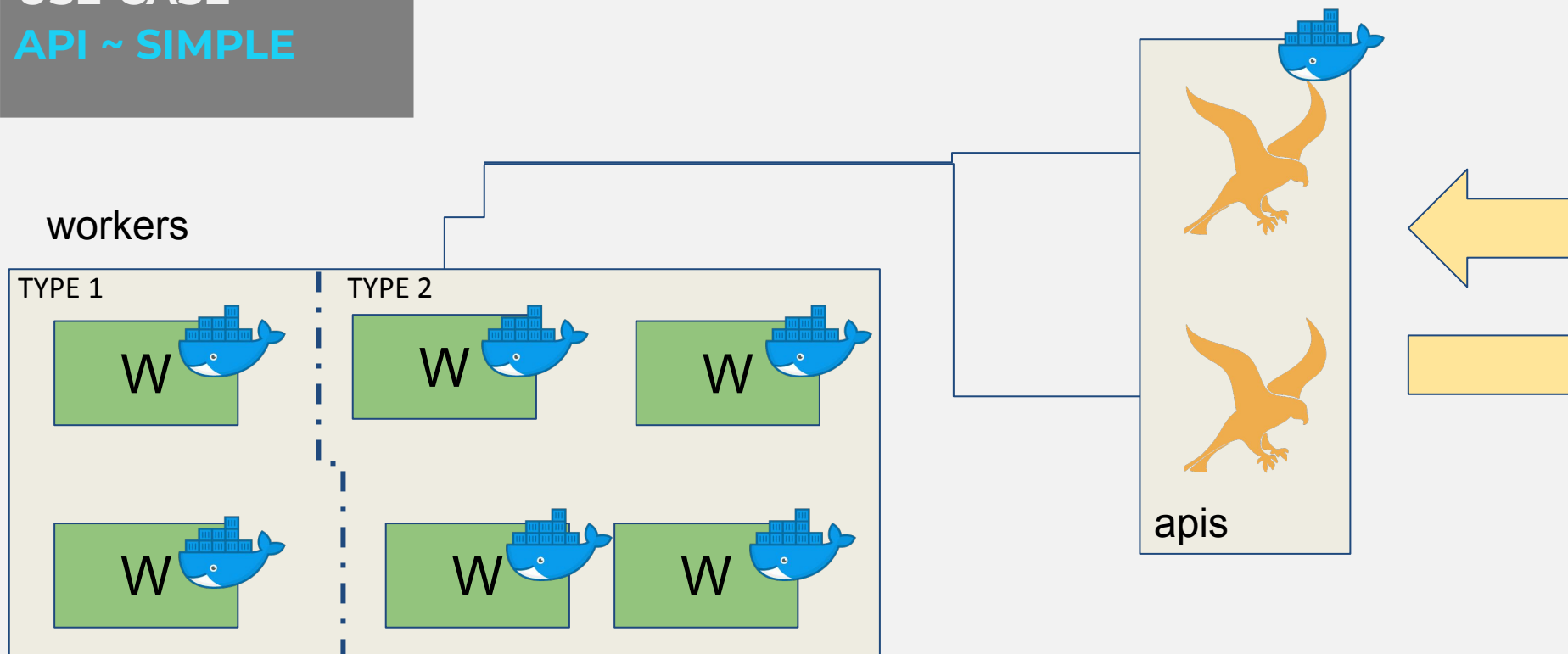
USE-CASE

API ~ SIMPLE



USE-CASE

API ~ SIMPLE

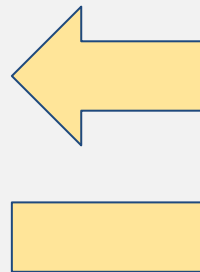
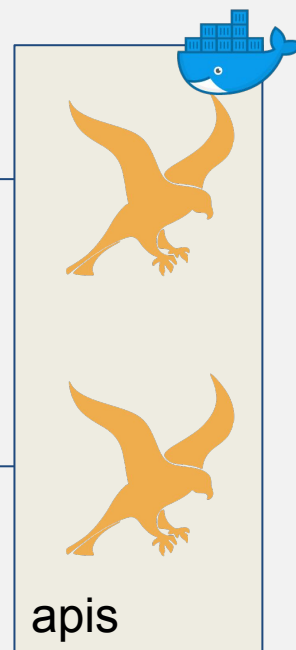
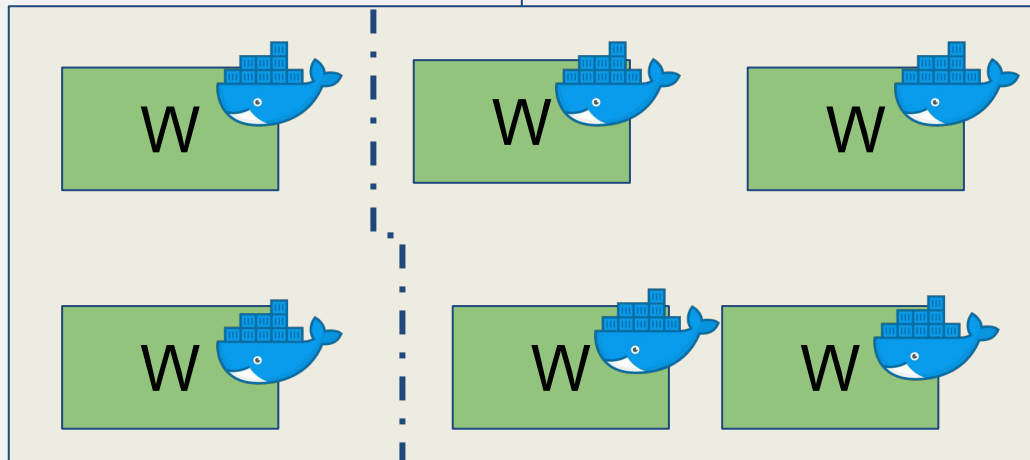


USE-CASE
API ~ SIMPLE



Queues

workers



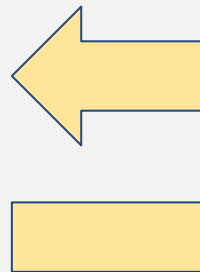
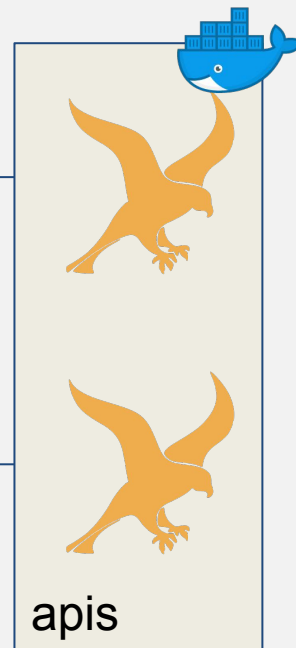
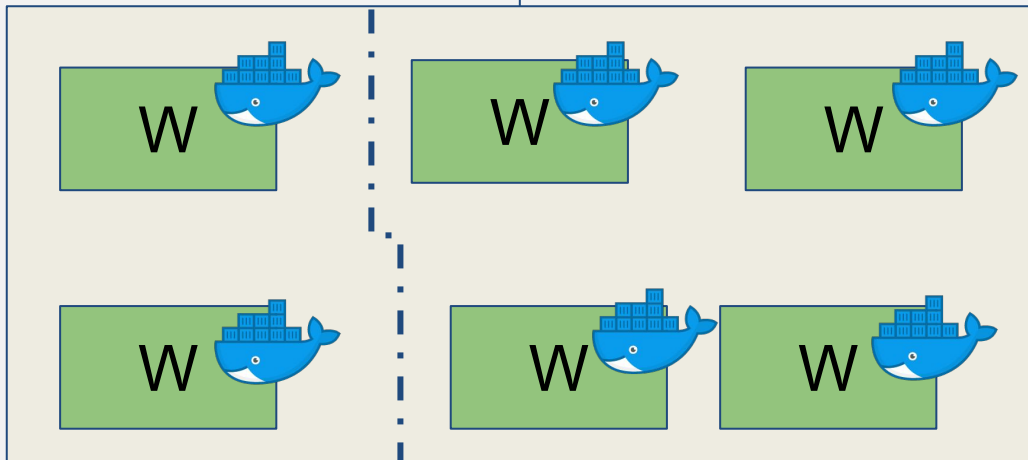


USE-CASE API ~ SIMPLE



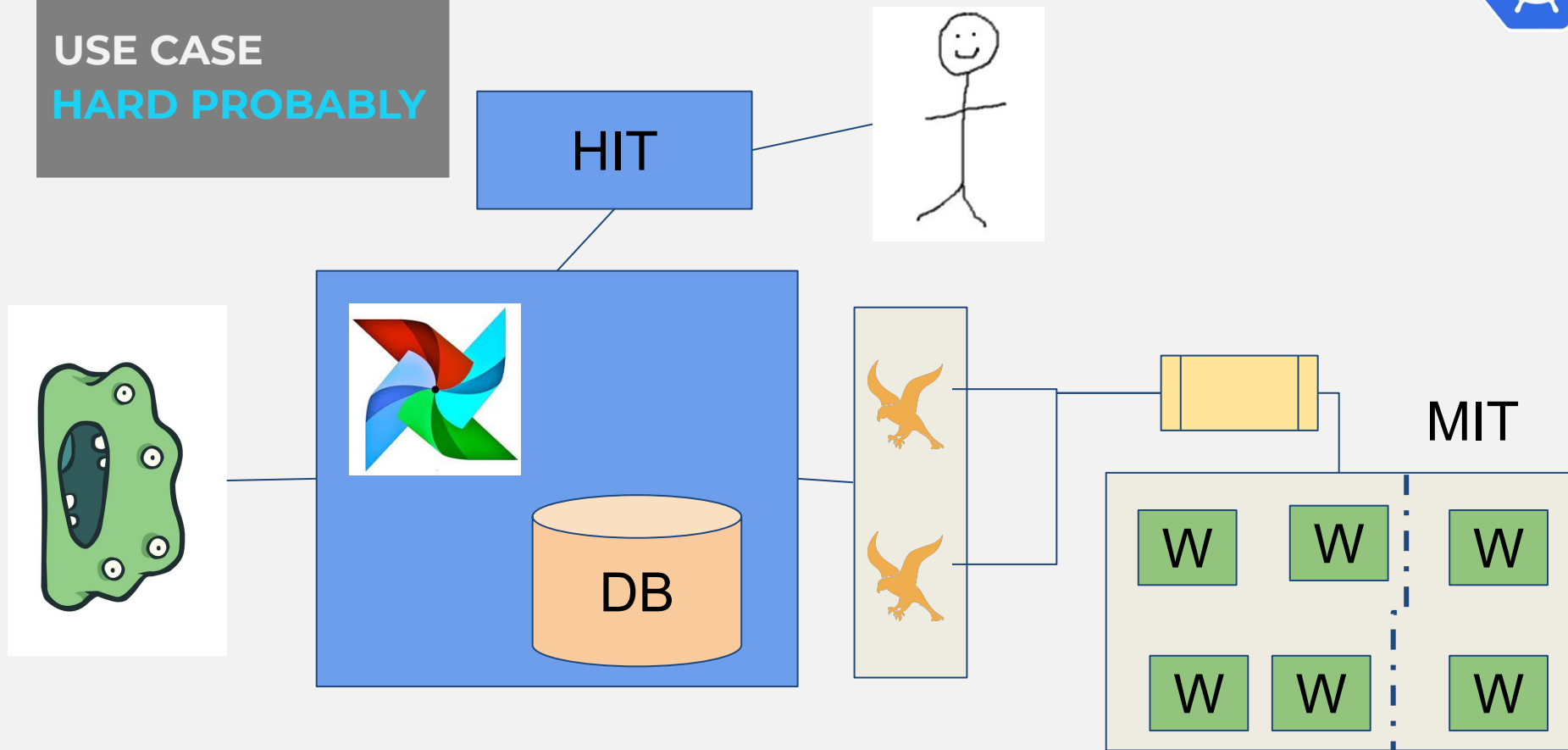
Queues

workers



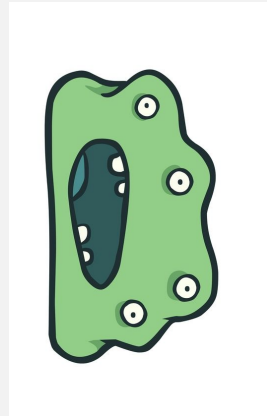


USE CASE
HARD PROBABLY

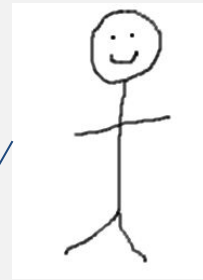




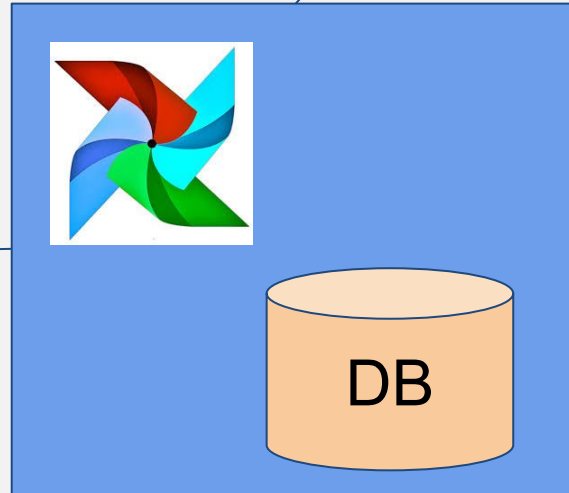
USE CASE
HARD PROBABLY



HIT



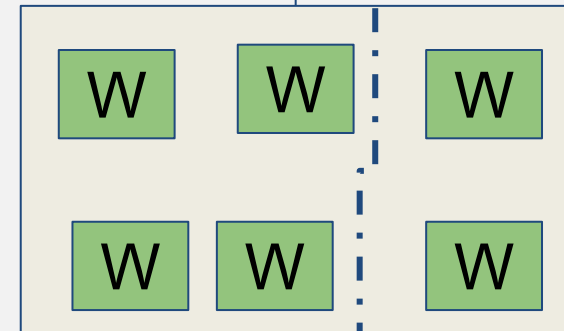
ADMIN



dj

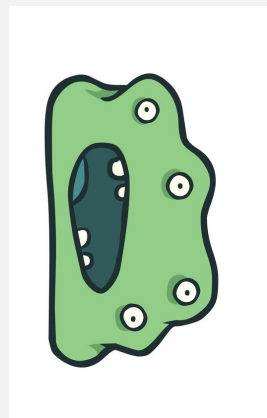


MIT

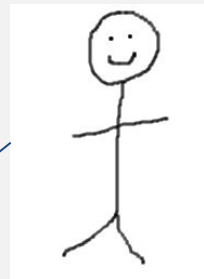
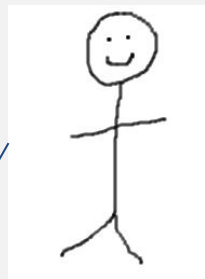




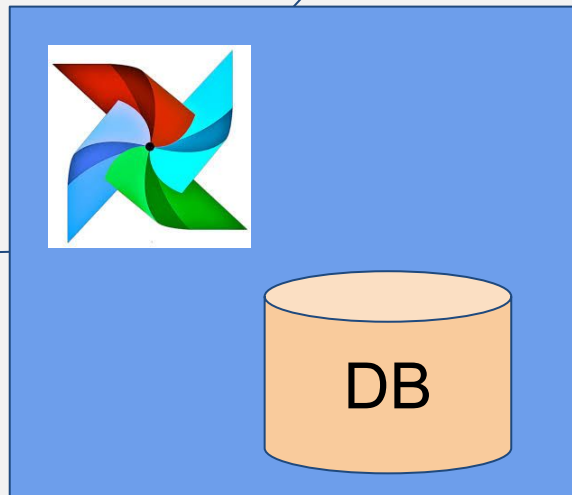
USE CASE
HARD PROBABLY



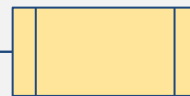
HIT



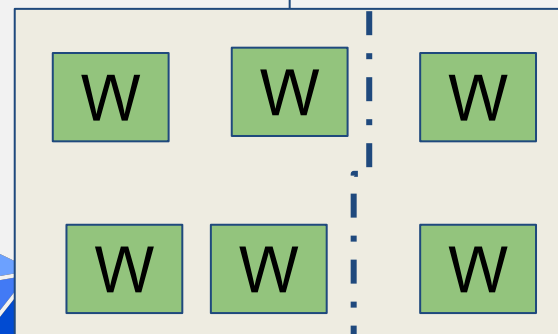
ADMIN



dj



MIT



Kubeflow



TECNOLOGIAS ENVOLVIDAS




3. NA PRÁTICA


ATENDIMENTO AJUDA

dafiti


OLÁ, RICARDO!  

FEMININO MASCULINO INFANTIL CASA ESPORTE MARCAS ESTILO **OUTLET**


O que você procura? 



LOOK COMPLETO COM TOP MARCAS
Já sabe o que usar na data especial?
CONFIRA ►



NOSSO LOOK
Inspire-se no estilo de Viih Rocha e Victor
CONFIRA ►



DOWNLOAD DO AMOR
Envie ao seu crush vale-gestos de carinho
BAIXAR ►





Is it important?

From Brightlocal:

- **91% of consumers rely on reviews made by other consumers** as much as personal recommendations,
 - a. Positive reviews **make** 68% of consumers more likely to buy
 - b. Negative reviews **stop** 40% of consumers wanting to buy

Machine Learning models are cool, but not so cool as to read text directly.

Word frequencies

Some data about us:

~11M clients

~700 reviews daily

~450K skus

Essa sandália é maravilhosa, so que so teve um pequeno probleminha, calço 34 e esse número ficou pequeno, tive que trocar por 35 que ficou perfeito, é muito linda se quiser ve-la na minha linha do tempo do facebook você pode verificar como é maravilhosa, amei ameeeeei !!!

vector with 62.635 positions (words dict size)

[0, 0, 0, 0,..., 2, ..., 2, ..., 3, ..., 1, ..., 0, 0]

amei*

maravilhosa

que

trocar

Machine Learning Approach



Naive Bayes

It's a popular (baseline) method for **text categorization**, with strong (naive) **independence assumption** between features.

Probabilistic approach

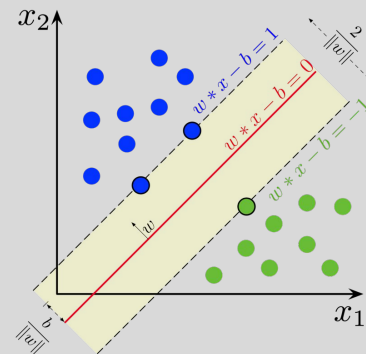
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

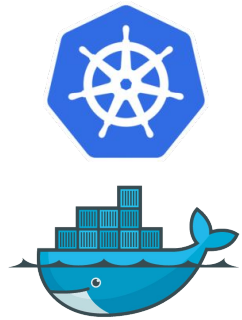
(Bayes' theorem)

Support Vector Machines (SVM)

Its **default** is a **non-probabilistic binary linear classifier**. Very powerful due to its **kernel trick** to construct **hyperplane(s)**.

Linear algebra approach





*

Intelligent System

ML API



Mediator

HIT

Predictions
Qs

Qs
Monitor

*

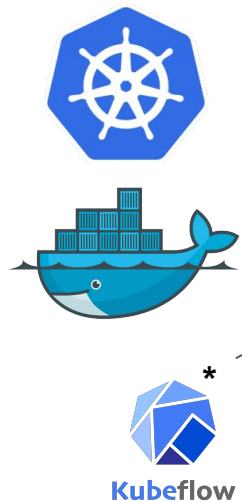
Production System

System
API

Ratings
Reviews



* Training new models.



Intelligent System

ML API

Mediator

HIT

Predictions
Qs

System
API

Ratings
Reviews

Qs
Monitor

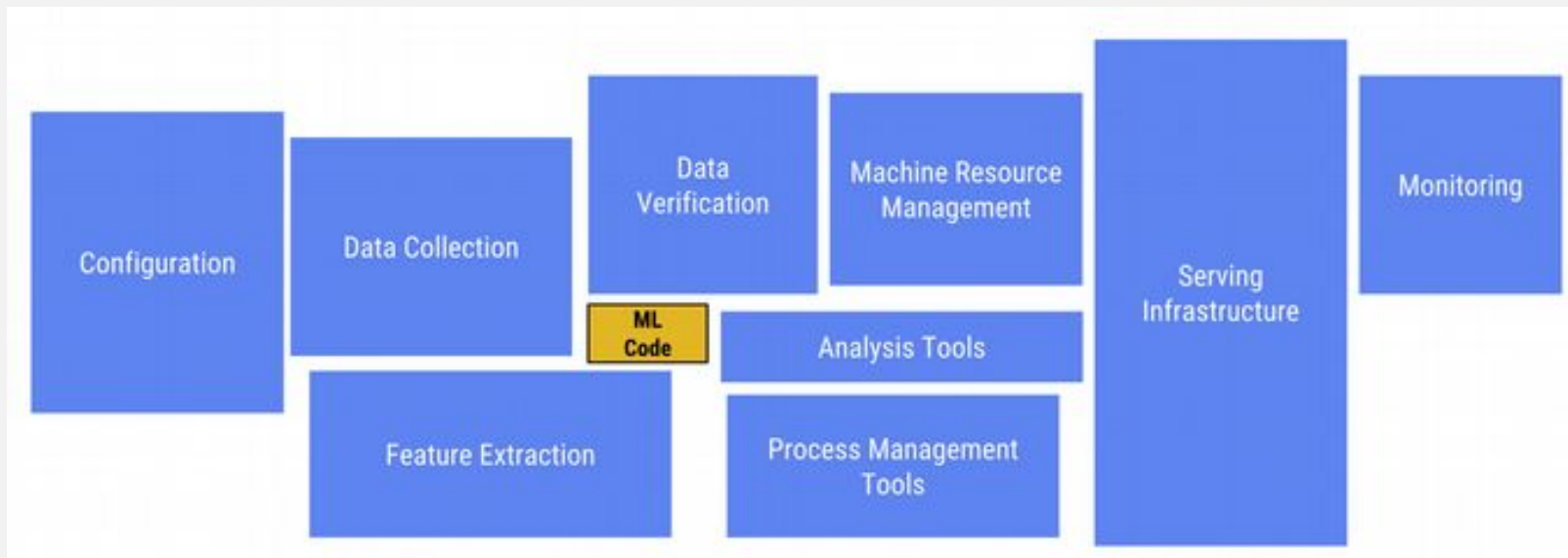
R&D
INFRA
ADMIN

Production System

* Training new models.

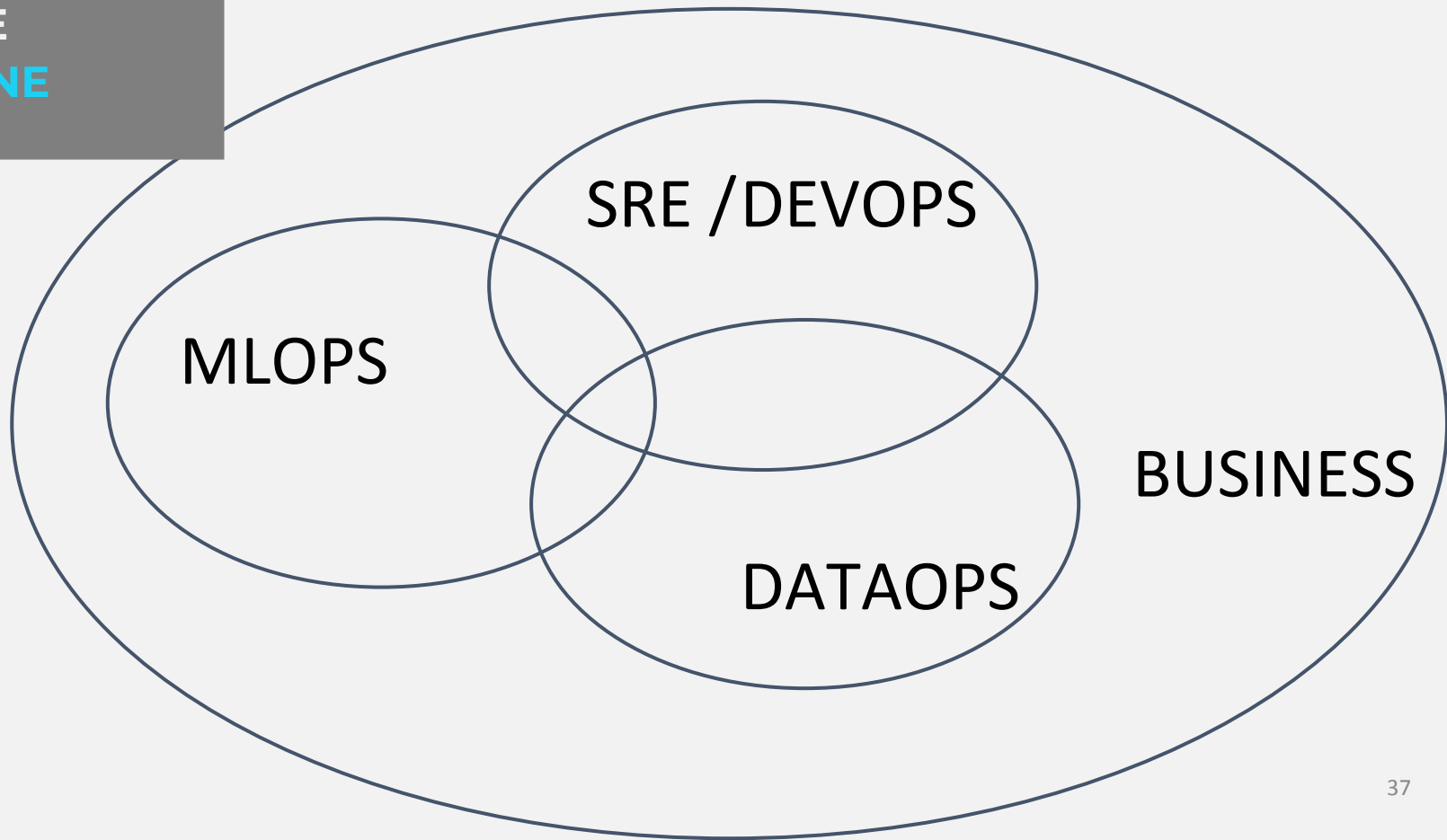
4. APRENDIZADOS E CUIDADOS

SIMPLY COMPLEX



CUSTOMER

INCLUDE
EVERYONE



Maturity Model

12factor app

1. **Transparência**
2. **Identificar / minimizar bias**
3. **Empoderamento**
4. **Reprodutibilidade**
5. **Colaborativo**
6. **Contínuo**
7. **Responsável**
8. **Explicável**

ÓTIMAS

REFERÊNCIAS



Kubeflow



[Home](#) > [Catalog](#) > [Scalable Microservices with Kubernetes](#)

FREE COURSE

Scalable Microservices
with Kubernetes
by Google

Master infrastructure management

START FREE COURSE

The DevOps for ML Manifesto.

awesome Maintained? YES Release PROD Languages MULTI License MIT Follow 2k

Awesome machine learning operations

This repository contains a curated list of awesome open source libraries that will help you deploy, monitor, version and your machine learning.

Quick links to sections in this page

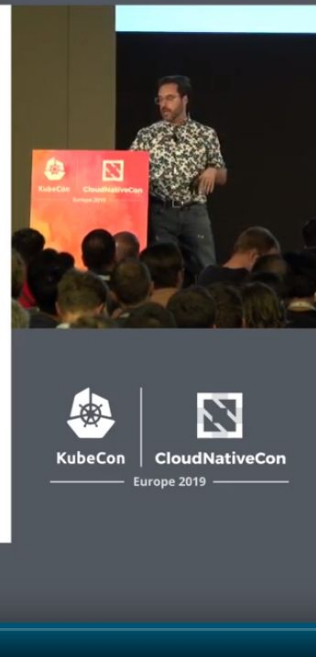
Explaining predictions & models	Privacy preserving ML	Model & data versioning
Model Orchestration	Feature engineering	Neural Architecture Search
Reproducible Notebooks	Visualisation frameworks	Industry-strength NLP
Data pipelines & ETL	Data Labelling	Data storage
Functions as a service	Computation distribution	Model serialisation
Compiler optimisation	Data Stream Processing	Commercial Platforms

ÓTIMAS

REFERÊNCIAS



11



▶ ⏮ 🔊 3:59 / 33:30

Videos brought to you by: Google Cloud



Como me encontrar:



github://ricoms



twitter://@Rico858



dftech@dafiti.com.br

Thanks!



RESEARCH &
DEVELOPMENT

