



# Data-science at scale at Mercadolibre

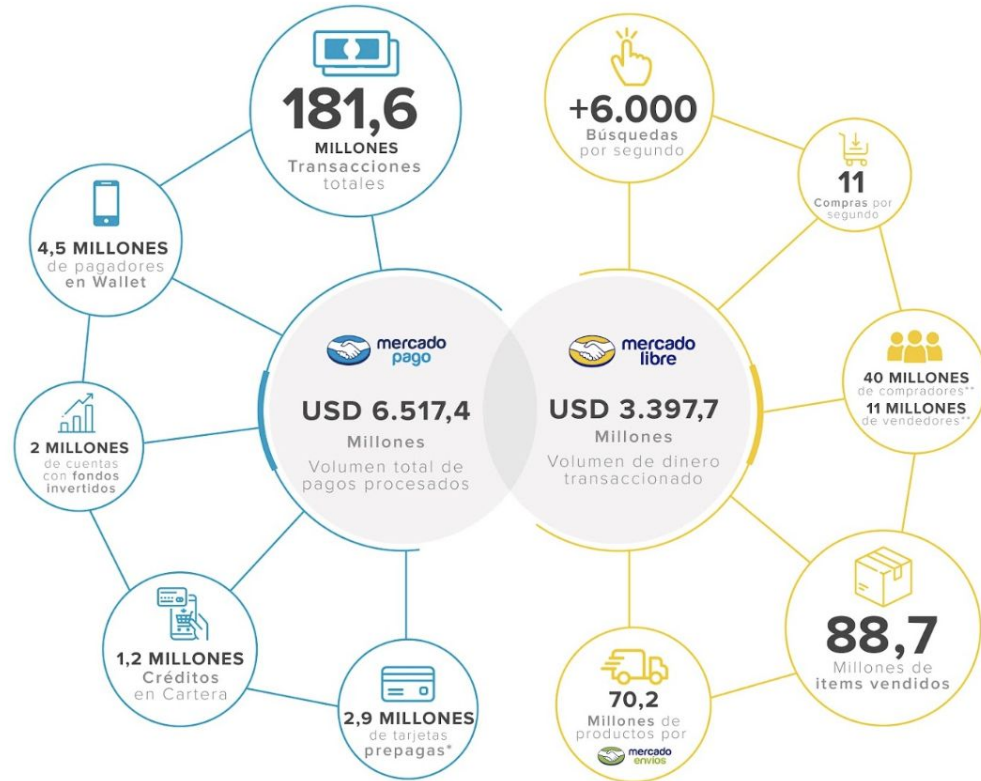
Carlos Matías de la Torre (a.k.a. *Litox*)

[carlos.delatorre@mercadolibre.com](mailto:carlos.delatorre@mercadolibre.com)

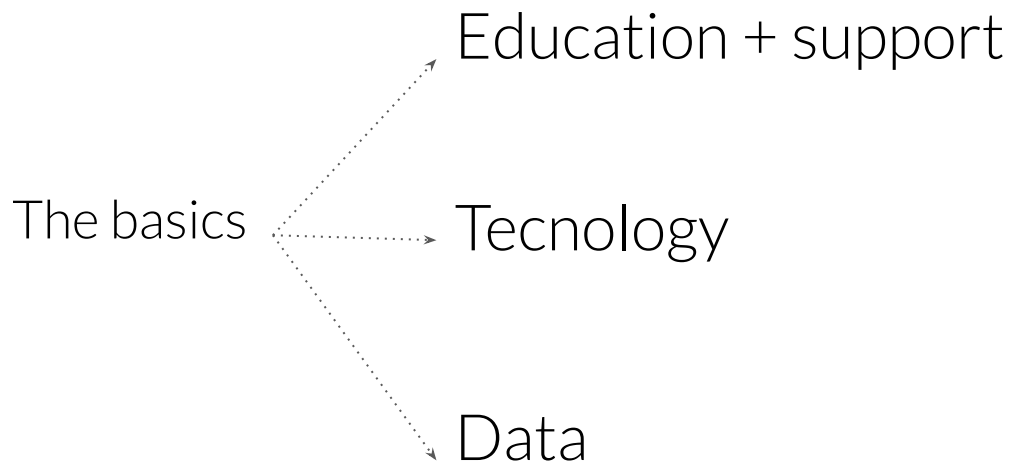
# One ecosystem



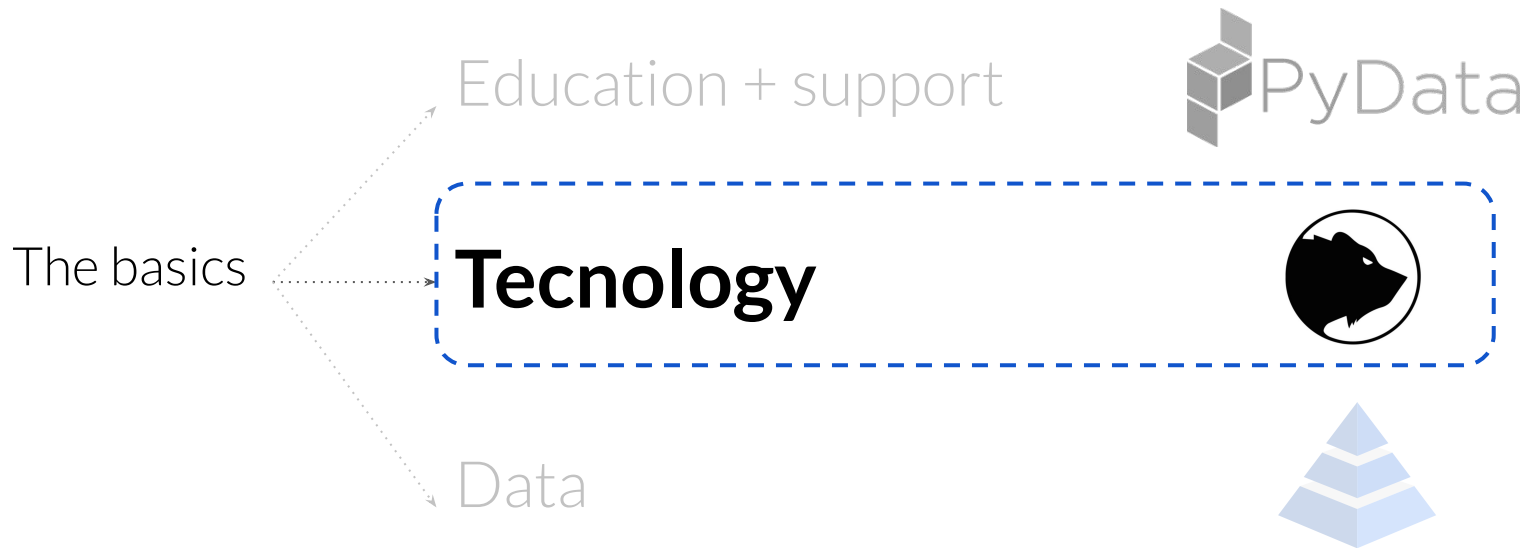
# Large-scale operations



# Large-scale ML strategy



# ML Large-Scale Strategy



# Fury



Fury - fda-test-app

Carlos de la Torre ▾



## Admin

Summary

Scopes

MPI  
Versions

CI-CD

Alerts

Collaborators

## Summary



Github!



Kibana!



Datadog!



Jenkins!

🔧 Start-up information

📜 Policies



# Data-science at scale

# Fury Data Apps

# Objectives

- Promote the practice of Machine Learning in MELI
- Encourage good practices
- Sharing knowledge
- Data governance and security
- Systematize the development process
- Speed-up the models' deployment



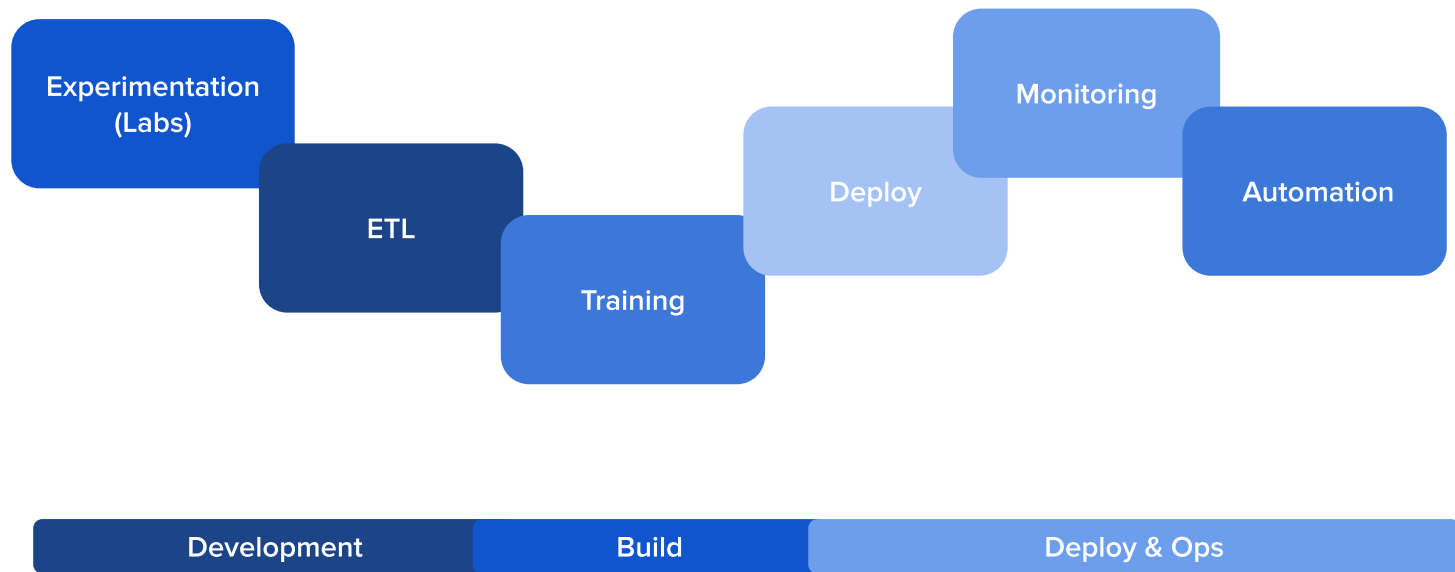




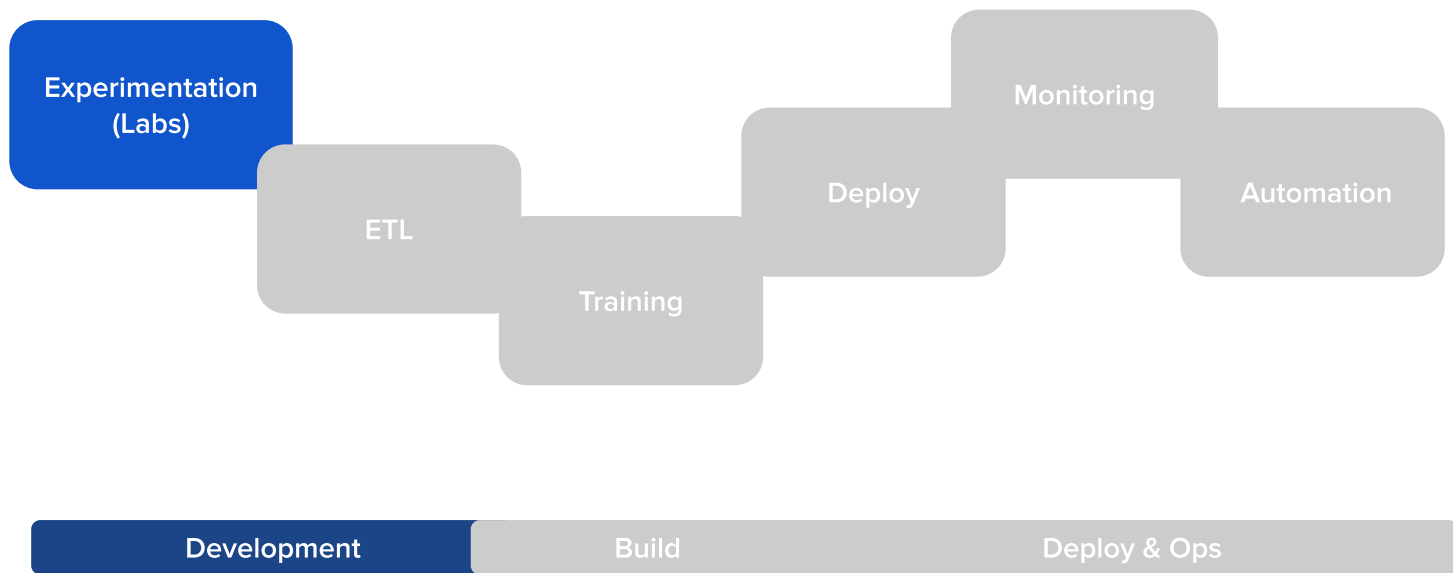
¿What is FDA?

*An in-house made framework  
for data-processing and machine-learning development,  
embedded in Fury*

# Machine-learning pipeline

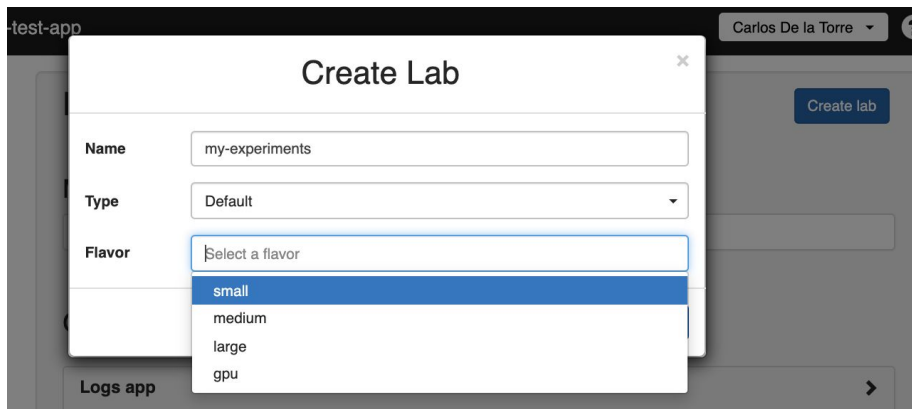


# Experimentation (Labs)

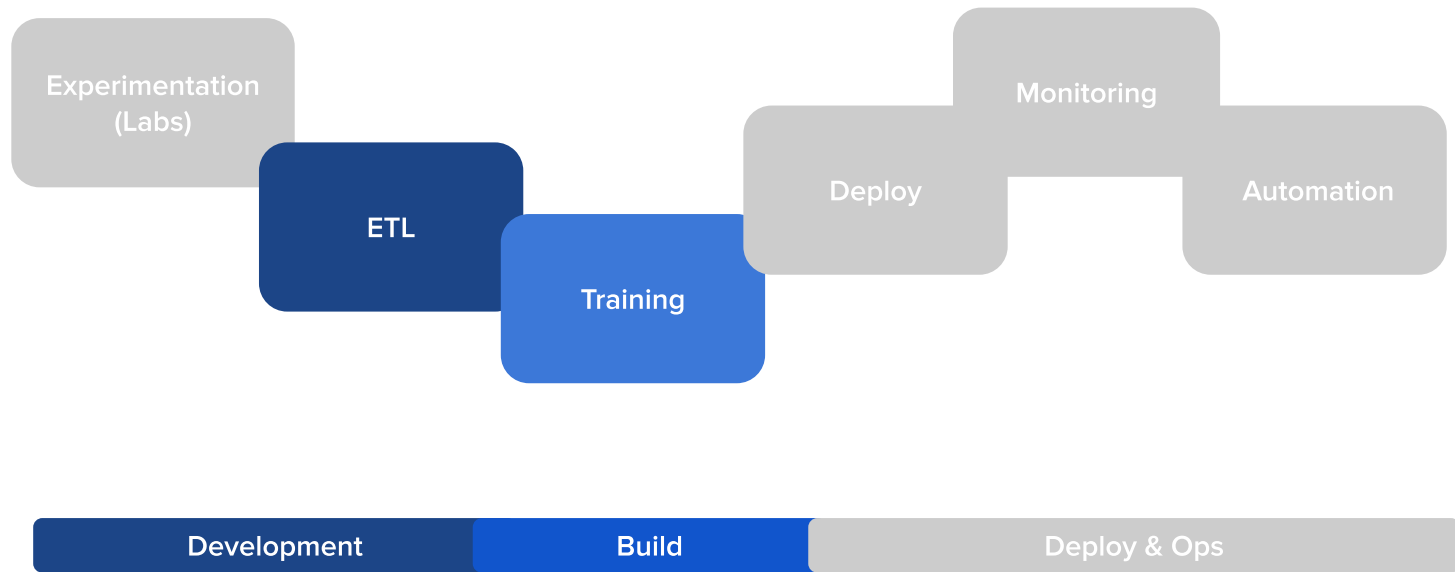


# LABs

- Dedicated instance running Jupyter Lab
- Repo with user credentials
- Access to data sources
- Different hardware flavours to choose



# ETL & Train



# ETL & Train

## ETL

[Create ETL](#)













Search:

version	branch	status		user	created	valid	
<a href="#">1.0.0-new-init</a>	test-new-init	finished	 	cdelatorre	11 days ago		
<a href="#">2.0.0-dev1-2</a>	master	finished	 	cdelatorre	a month ago		
<a href="#">2.0.0-dev1</a>	master	error	 	cdelatorre	a month ago	-	

## Trainings

[Create training](#)

Search:

Version	Branch	ETL	Status		User	Created	
<a href="#">1.0.1-new-init</a>	test-new-init	<a href="#">1.0.0-new-init</a>	finished	 	cdelatorre	11 days ago	
<a href="#">1.0.0-new-init</a>	test-new-init	<a href="#">1.0.0-new-init</a>	error	 	cdelatorre	11 days ago	
<a href="#">2.0.0-dev2-3</a>	master	<a href="#">2.0.0-dev1-2</a>	finished	 	cdelatorre	a month ago	
<a href="#">2.0.0-dev2-2</a>	master	<a href="#">2.0.0-dev1-2</a>	error	 	cdelatorre	a month ago	

# ETL & Train

## ETL

Create ETL

Search:

Filter text

version	branch	status		user	created	valid	
1.0.0-new-init	test-new-init	finished	≥	cde	11 days ago	👍	📄
2.0.0-dev1-2	master	finished	≥	cde	a month ago	👍	📄
2.0.0-dev1	master	error	≥	cde	a month ago	-	📄

## Trainings

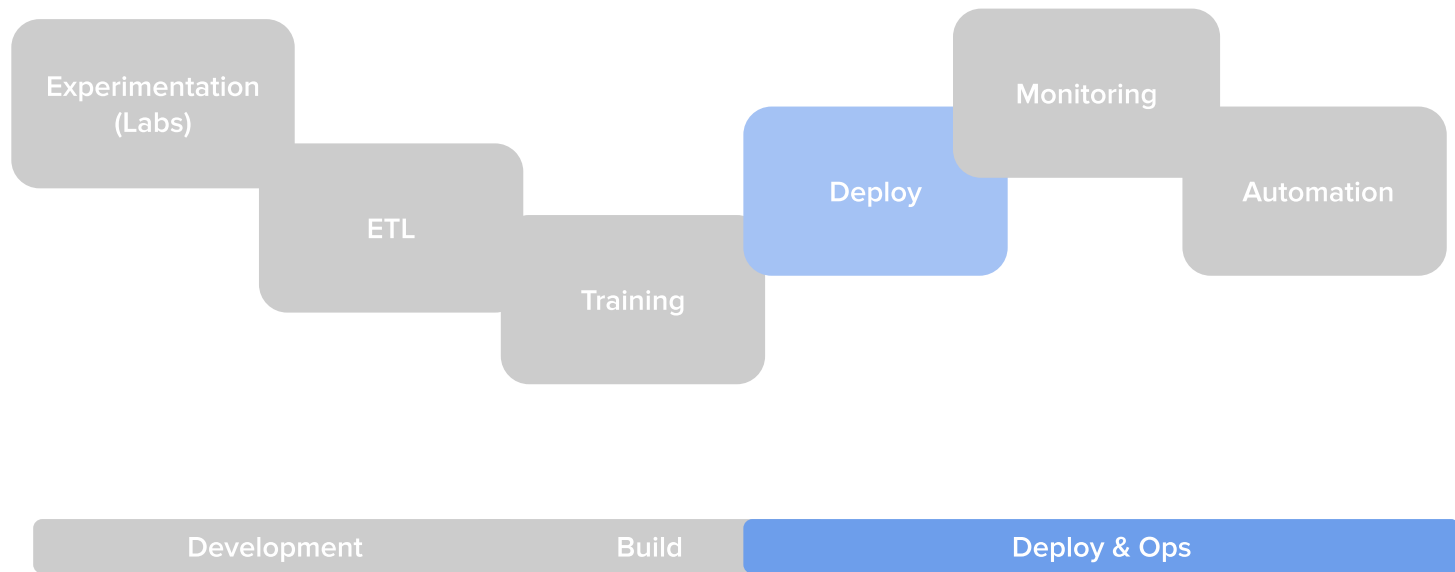
Create training

Search:

Filter text

Version	Branch	ETL	Status	User	Created	
1.0.1-new-init	test-new-init	1.0.0-new-init	finished	cde	11 days ago	📄
1.0.0-new-init	test-new-init	1.0.0-new-init	error	cde	11 days ago	📄
2.0.0-dev2-3	master	2.0.0-dev1-2	finished	cde	a month ago	📄
2.0.0-dev2-2	master	2.0.0-dev1-2	error	cde	a month ago	📄

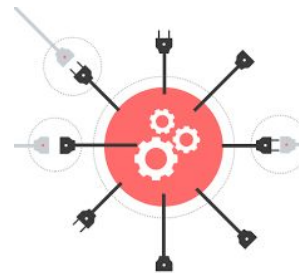
# Deploy





# Build & Deploy

- Semi-automatic REST API generation  
(serving model predictions)
- Fury operations features
  - autoscaling
  - logs
  - metrics
  - audits
  - costs
  - ...




# Build & Deploy

## MPI Versions

Search:  Status:

Create Version

Version	Branch	Training	Status		User	Created	Deployable
1.0.0-new-init	test-new-init	1.0.1-new-init	finished	≥ ⓘ	odelatorre	11 days ago	
0.0.1-dev	master	2.0.0-dev2-3	disabled	≥ ⓘ	odelatorre	a month ago	



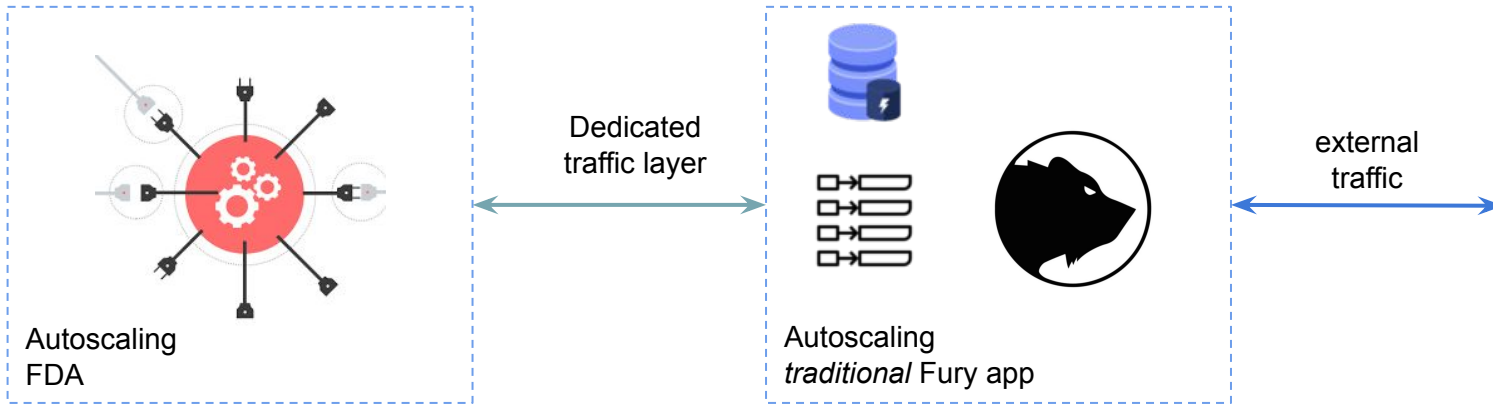
## Scopes

Search:  Productive:

Create

name ↑	region	version	branch	cname	alerts	middleware	test	
dev	us-east-1-si...	0.0.1-dev	demo-fake	dev.app-fake.data.furyapps.io		✓	✓	Activate ⓘ ⌚ ⌘ + ⌵
mla	us-east-1-si...	1.0.0-new-init	master	mla.app-fake.data.furyapps.io	✓	✓		⌵ ⓘ ⌚ ⌘ + ⌵
mlb	us-east-1-si...	2.0.2	master	mlb.app-fake.data.furyapps.io	✓	✓		⌵ ⓘ ⌚ ⌘ + ⌵
mlm	us-east-1-si...	2.0.1	master	mlm.app-fake.data.furyapps.io	✓	✓		⌵ ⓘ ⌚ ⌘ + ⌵

# Deploy: Serving app



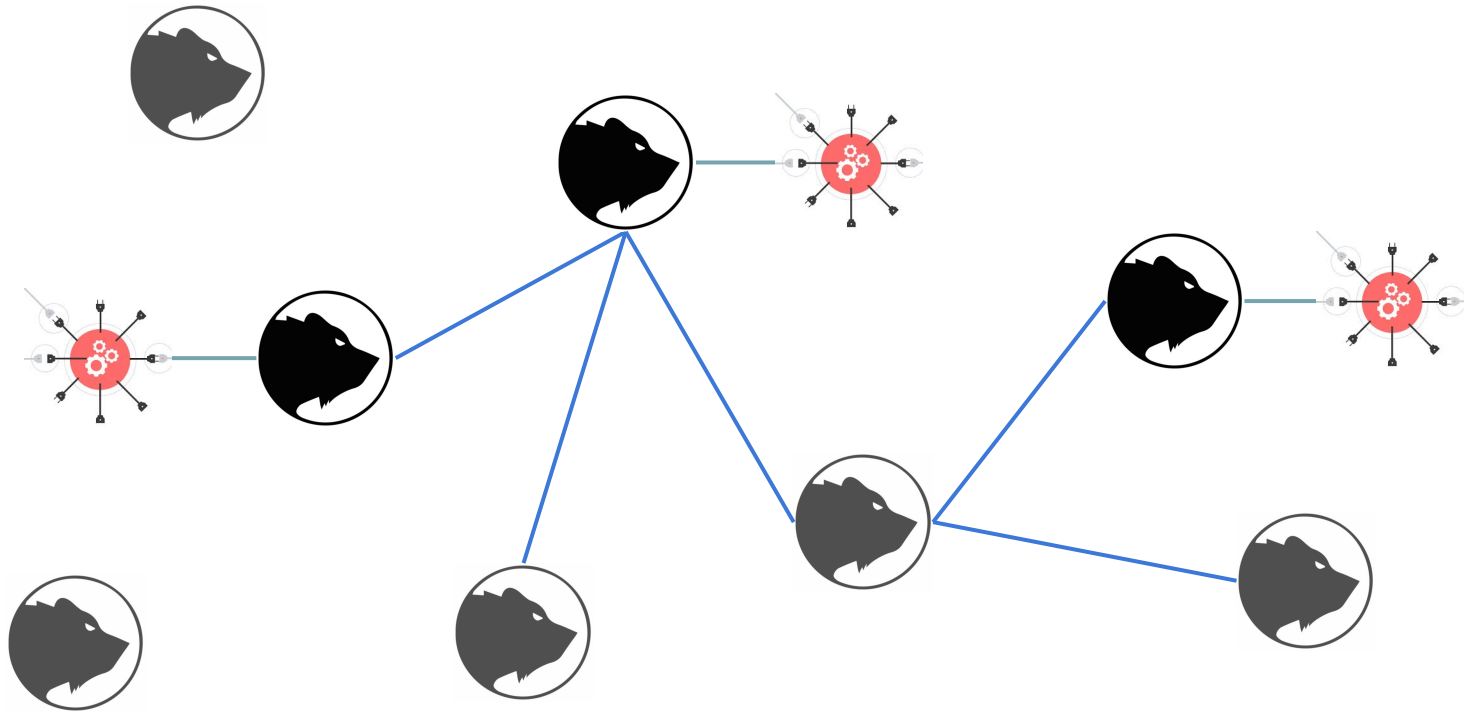
## ***FDA***

- Predictions As A Service
- CPU-bound (model execution)
- Standard *and machine-learning'* specific metrics and monitoring

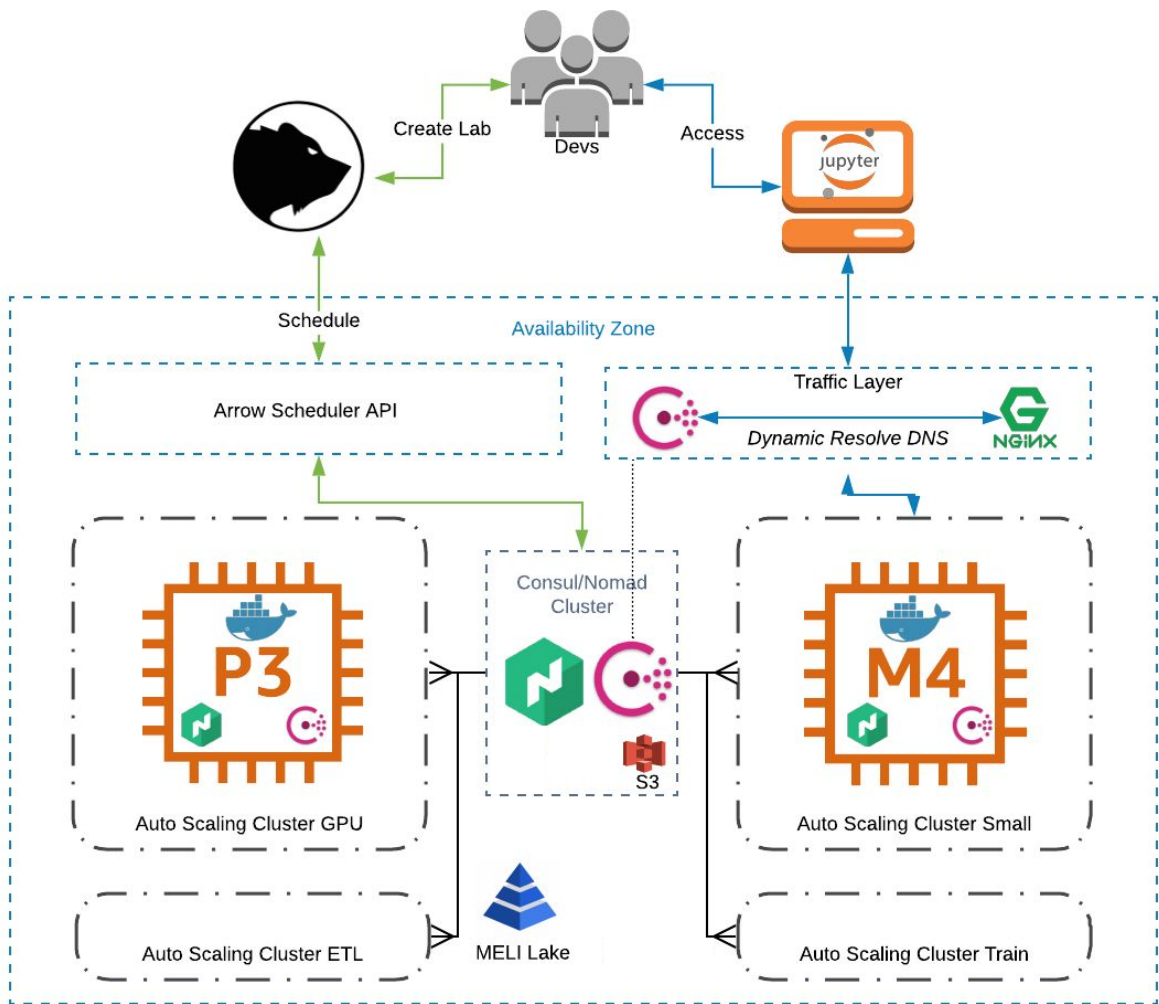
## ***Serving App***

- Public-facing
- Business logic
- IO-bound
- Services (DB, KVS, cache...)
- Standard metrics and monitoring

# Microservices Ecosystem



# Behind the scenes



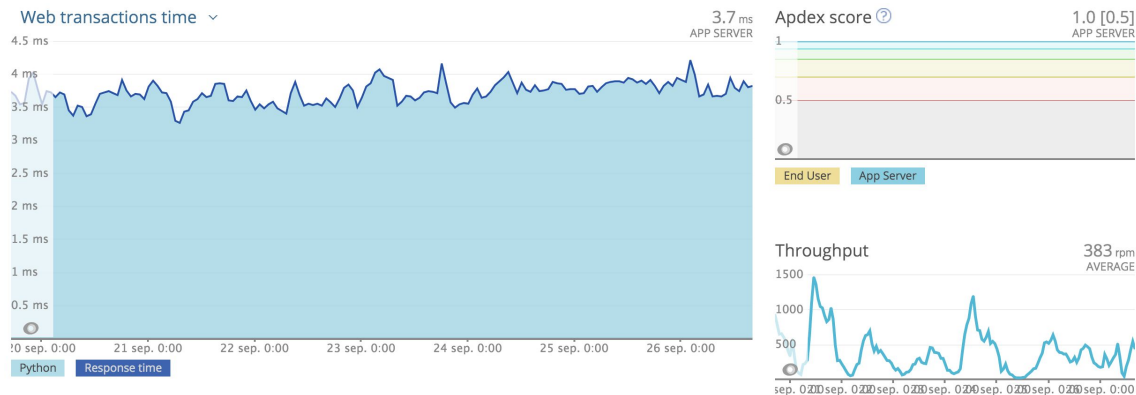


# Production ready

# Stock questions

Detect “do you have in stock?” questions, processing items title and description

- API response time: 3.5 ms
- Throughput: 300 - 1500 rpm
- Infra: 2 instances

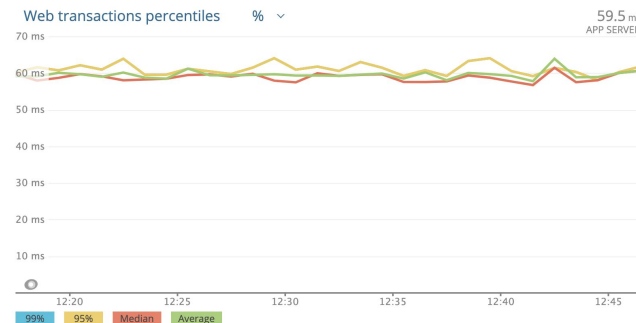
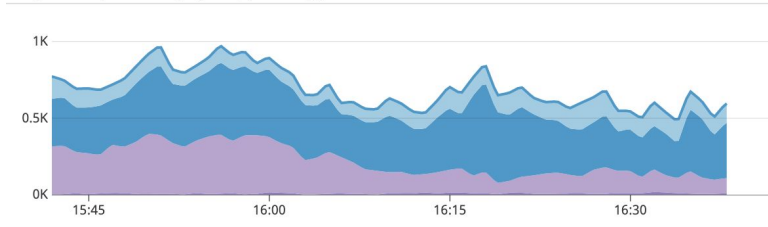


# Fakes

Detect fake items processing images.

- API response time: 60 ms
- Throughput: 0.5 ~ 10 krpm
- Infra: 2 instances

Highest api throughput (quantity)

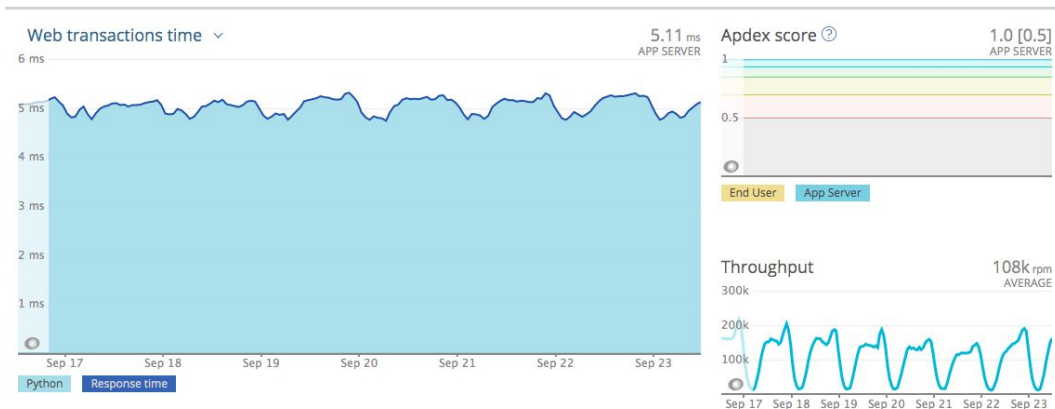




# SHARP

Predict how many days will a package spend in the carrier (one of the factors to predict delivery-time).

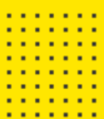
- API response time: 5~8 ms
- Throughput: **up to ~200 krpm**
- Infra: 12 instances



## To be continued...

- Automation
- Predictions cache
- Serve on GPU
- Advanced metrics and model monitoring
- Spark on-demand
- Further adoption
- and more...





# ¡Gracias! :)

Carlos Matías de la Torre

[carlos.delatorre@mercadolibre.com](mailto:carlos.delatorre@mercadolibre.com)