



Build data pipelines, [the easy way](#)



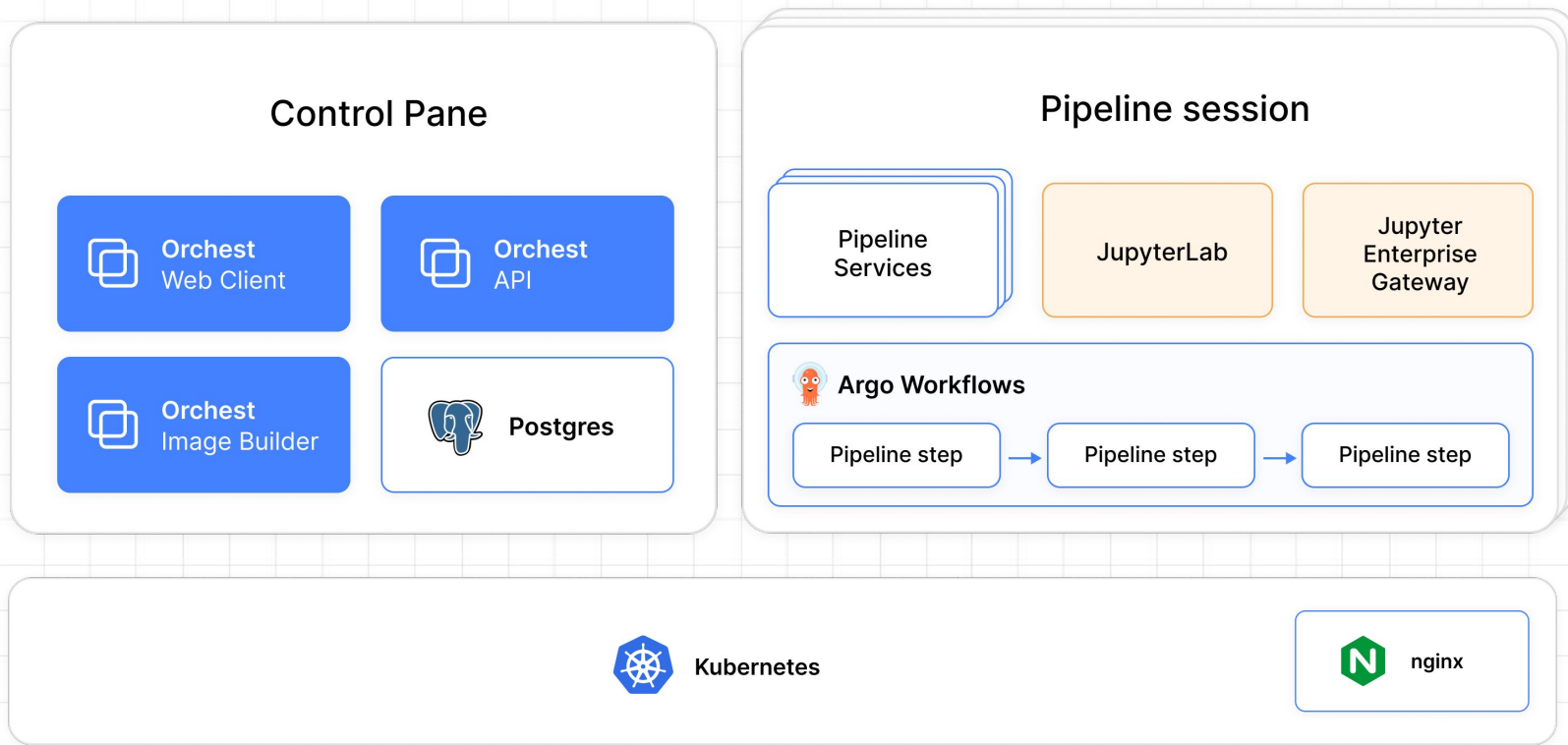
A data workflow orchestrator for the masses


The screenshot displays the Orchest web interface for a project named 'california-housing'. The interface is divided into several sections:

- Header:** Shows the project name 'california-housing' and buttons for 'STOP SESSION' and 'SWITCH TO JUPYTERLAB'.
- Navigation:** Includes a 'Project quickstart' dropdown, a '+ NEW STEP' button, and tabs for 'LOGS', 'SERVICES', and 'SETTINGS'.
- File Explorer:** A sidebar on the left shows the project file structure, including folders like '.git', '.orchest', and files like 'get-data.py', 'linear-regression.ipynb', etc.
- Workflow Canvas:** A central grid area displays a workflow graph. The steps are:
 - 'Get data' (get-data.py) - Ready
 - 'Data exploration' (explore-data.ipynb) - Ready
 - 'Preprocessing' (pre-processing.ipynb) - Ready
 - 'linear-regression' (linear-regression.ipynb) - Ready
 - 'decision-tree' (decision-tree.ipynb) - Ready
 - 'svm' (svm.ipynb) - ReadyThe flow starts with 'Get data', which branches into 'Data exploration' and 'Preprocessing'. 'Preprocessing' then branches into 'linear-regression', 'decision-tree', and 'svm'.
- Properties Panel:** On the right, a 'PROPERTIES' panel is open for the selected 'Get data' step. It shows:
 - Title: Get data
 - File path: get-data.py (with a green checkmark and the note 'File exists in the project directory.')
 - Environment: Python 3
 - Parameters: 1 {}
 - Buttons: 'EDIT IN JUPYTERLAB' and 'VIEW FILE'.
 - Footer: 'CLOSE' and 'DELETE' buttons.
- Footer:** A 'RUN SELECTED STEPS' button is located at the bottom center of the canvas.



Architecture



 Optional

Feature Deep Dive Jupyter Kernels

```
explore-data.ipynb
Python 3

Data exploration

Explore the housing data that was passed by the previous step.

The packages orchest, matplotlib and sklearn are included by default, i.e. you don't have to install the dependencies using pip for example.

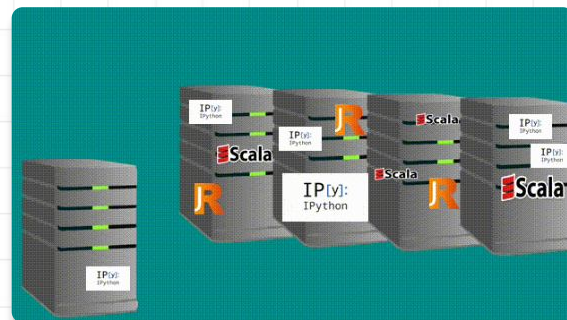
[1]: from matplotlib import pyplot as plt
import orchest
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler

[2]: # Retrieve the data from the previous step.
data = orchest.get_inputs()
data, target = data["data"]

[3]: data.head()

[3]:
```

	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude
0	8.3252	41.0	6.984127	1.023810	322.0	2.555556	37.88	-122.23
1	8.3014	21.0	6.238137	0.971880	2401.0	2.109842	37.86	-122.22
2	7.2574	52.0	8.288136	1.073446	496.0	2.802260	37.85	-122.24
3	5.6431	52.0	5.817352	1.073059	558.0	2.547945	37.85	-122.25
4	3.8462	52.0	6.281853	1.081081	565.0	2.181467	37.85	-122.25



Jupyter
Enterprise
Gateway



Feature Deep Dive Environment Builds

```
1 #!/bin/bash
2
3 # Install any dependencies you have in this shell script.
4
5 pip install tld pandas
6 sudo apt-get update
7 sudo apt-get install -y graphviz
8
```

 BUILD

Build successfully completed!

Logs

```
Setting up libx11-6:amd64 (2:1.6.9-2ubuntu1.2) ...
Setting up libharfbuzz0b:amd64 (2.6.4-1ubuntu4) ...
Setting up libtiff5:amd64 (4.1.0+git191117-2ubuntu0.20.04.2) ...
Setting up libfontconfig1:amd64 (2.13.1-2ubuntu3) ...
Setting up libsm6:amd64 (2:1.2.3-1) ...
Setting up libxft2:amd64 (2.9.10+dfsg-5ubuntu0.20.04.2) ...
Setting up fontconfig (2.13.1-2ubuntu3) ...
Regenerating fonts cache... done.
Setting up libxpm4:amd64 (1:3.5.12-1) ...
Setting up libxrender1:amd64 (1:0.9.10-1) ...
Setting up libpango-1.0-0:amd64 (1.44.7-2ubuntu4) ...
Setting up libxext6:amd64 (2:1.3.4-0ubuntu1) ...
Setting up libcairo2:amd64 (1.16.0-4ubuntu1) ...
Setting up shared-mime-info (1.15-1) ...
Setting up libgd3:amd64 (2.2.5-5.2ubuntu2.1) ...
```

 Kaniko

 docker

BuildKit



podman



We're proud to **announce** ...

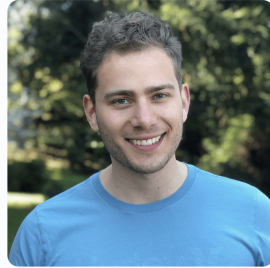


... that Kubernetes support is part of the **Open Core!**

Meet the team



Rick Lamers
CEO & Co-Founder



Yannick Perrenet
CTO & Co-Founder



Jacopo Gobbi
Founding Backend Engineer



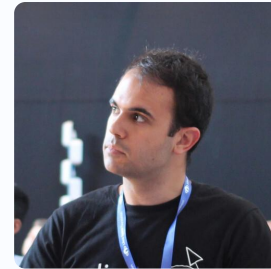
Huang-Ming Chang
Fullstack Engineer



Navid Haghighat
Software Engineer



Danique Sipkes
Operations Manager



Juan Luis Cano Rodríguez
Data Scientist Advocate

Project continuity

PROUDLY BACKED BY



Gradient Ventures



BASIS SET



script/

Deploy, explore and contribute



[orchest.io/github](https://github.com/orchest-io/github)

Fully Self-hostable!

