Machine Learning as a Service
Application of Google Cloud Platform to Machine Learning problems

Marco Landoni

INAF – National Insitute of Astrophysics

Brera Astronomical Observatory

marco.landoni@inaf.it

But what if it's Cloudy?

Warning!

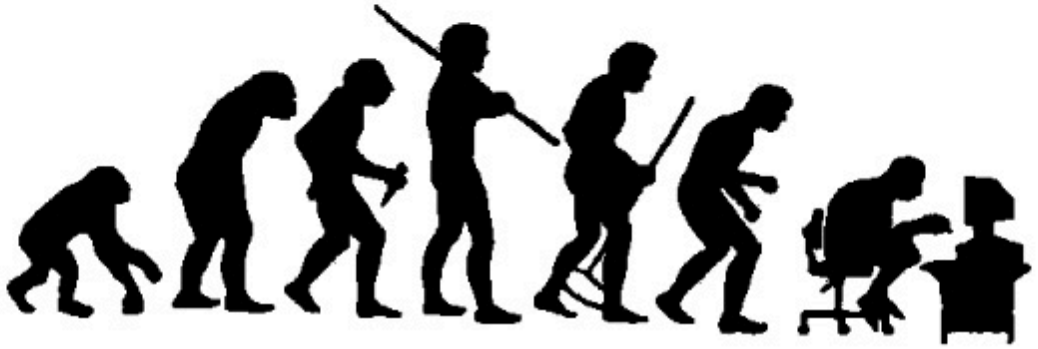
- **One size does not fit all.**

- Each problem has its **proper computational model** and architecture that maximise the return, the **cost efficiency** and the proper **exploitation of common resources**.

- I will try to give you an overview of just 1 platform and the related main services (this should serve as the "1st page of the book").

Cloud computing paradigm

- **Cloud Computing** is a style of **computing paradigm** in which typically real-time scalable resources can be accessible via Internet to users. Pay as-you-go for resource utilisation. (Wikipedia)



Various providers ….

Main Cloud Computing Services

- Computational power available for a reasonable price.

Compute Engine

Virtual Machines, Disks, Network

- Storage with high availability, virtual infinite storage and durability

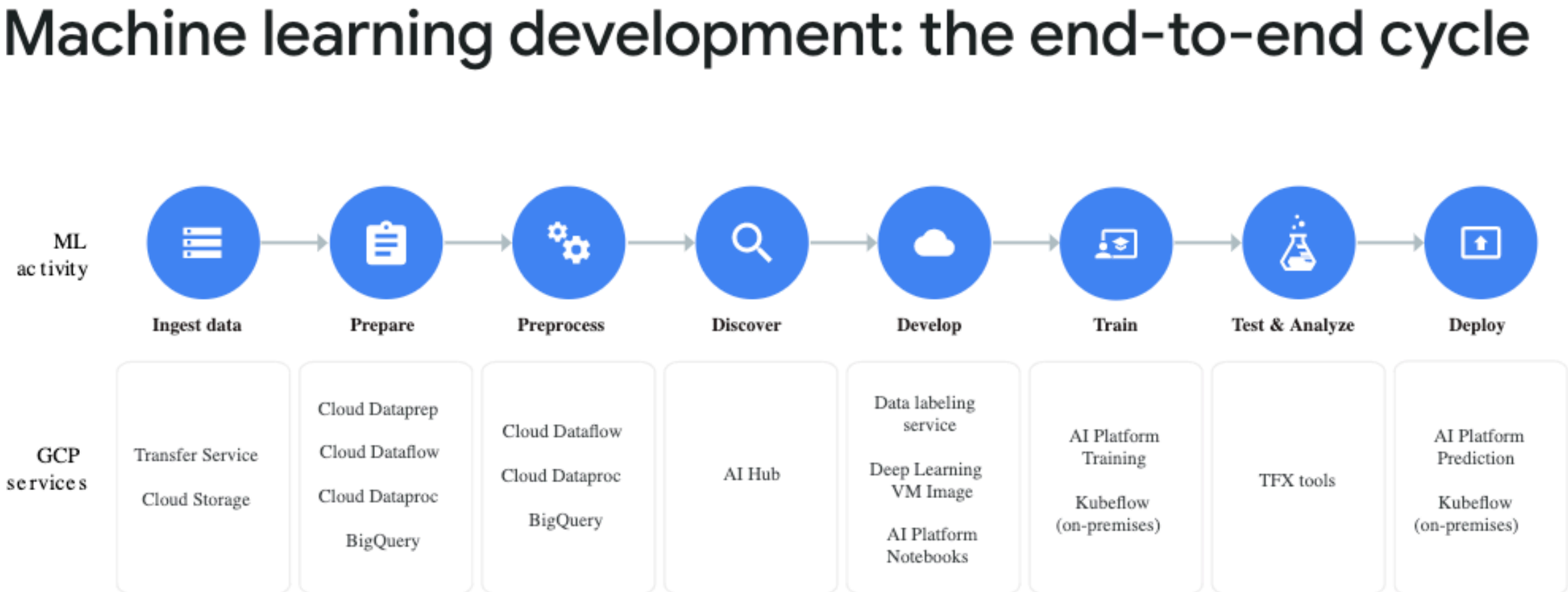
- A large set of services for data handling and analytics
  - Streaming
  - Data ingestion from various sources (e.g. sensor networks)
  - Messages and queue managements

Machine Learning As A Service (MLaaS)

- Set of services that offer machine learning tools as part of cloud computing services.
- MLaaS providers offer tools including:
  - Data visualization
  - APIs, face recognition, natural language processing, predictive analytics and deep learning, data preparation and cleaning…
  - **The provider's data centers handle the actual computation.**

You can focus only on Data Science

The case of Google Cloud Platform

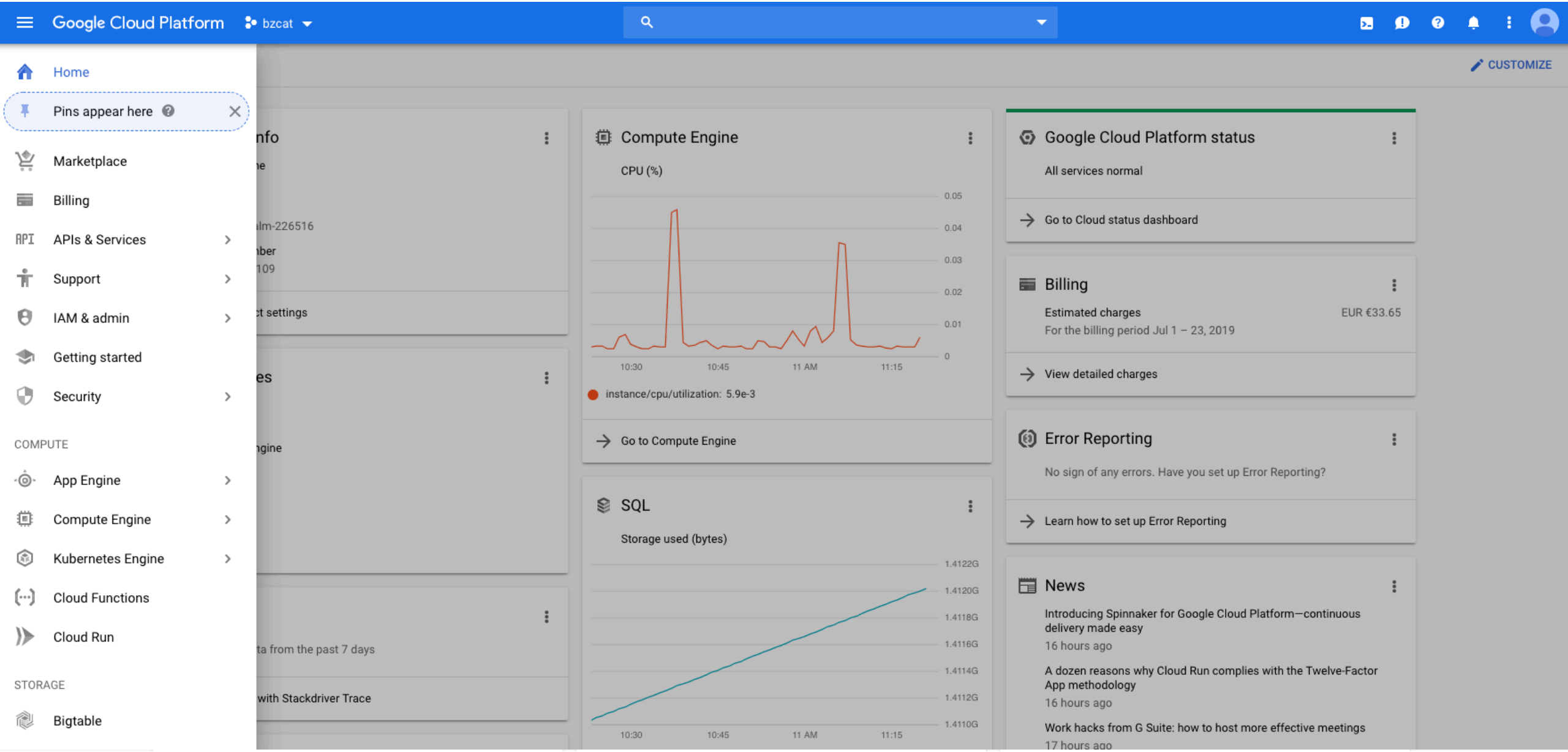


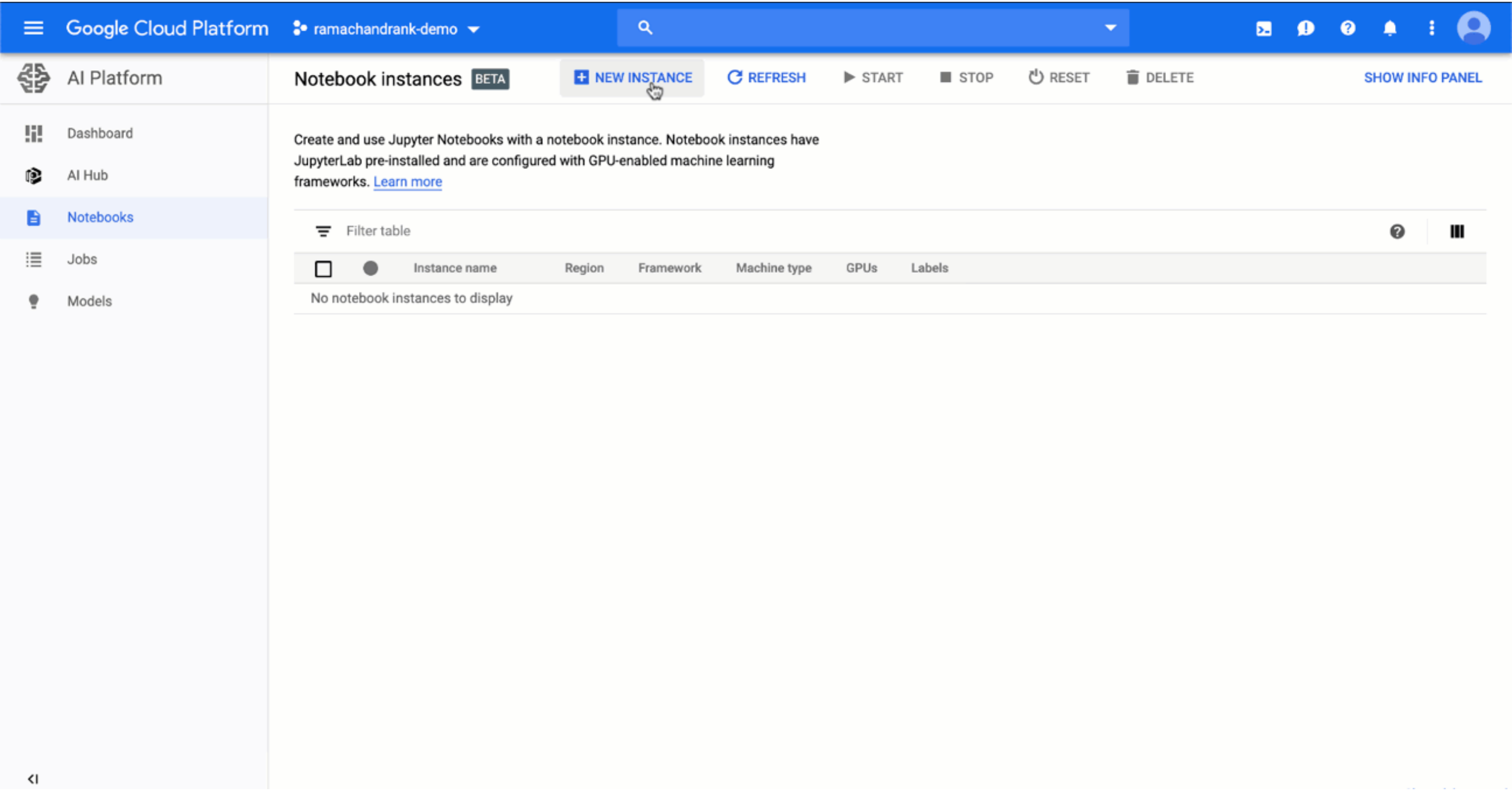
Freely took from https://codelabs.developers.google.com/codelabs/end-to-end-ml/index.html?index=..%2F..index#2

The 1° page of the book

- Understand the console and then…
- Prepare the data

https://cloud.google.com




Google Cloud Platform
ramachandrank-demo
AI Platform
Notebook instances
BETA
NEW INSTANCE
REFRESH
START
STOP
RESET
DELETE
SHOW INFO PANEL
Dashboard
AI Hub
Notebooks
Jobs
Models
Create and use Jupyter Notebooks with a notebook instance. Notebook instances have JupyterLab pre-installed and are configured with GPU-enabled machine learning frameworks. Learn more
Filter table
Instance name
Region
Framework
Machine type
GPUs
Labels
No notebook instances to display

You can also customize the instance

Compute Engine

Virtual Machines, Disks, Network

← New notebook instance BETA

Instance name *
tensorflow-20190723-112814

Region *
us-west1 (Oregon)

Zone *
us-west1-b

Requests to your instance from the Datalab/Jupyter interface may be routed through a different region than selected above depending on service availability.

Framework *
TensorFlow 1.x

Machine type *
4 vCPUs, 15 GB RAM

Machine types determine the specifications of your machines, such as the amount of memory, virtual cores, and persistent disk limits an instance will have.

LEARN MORE

GPUs

Based on the zone, framework, and machine type selected above, the available GPU types and the minimum number of GPUs that can be selected may vary. Learn more

GPU type
None

Boot disk

Boot disk type *
Standard Persistent Disk

Boot disk size in GB *
100

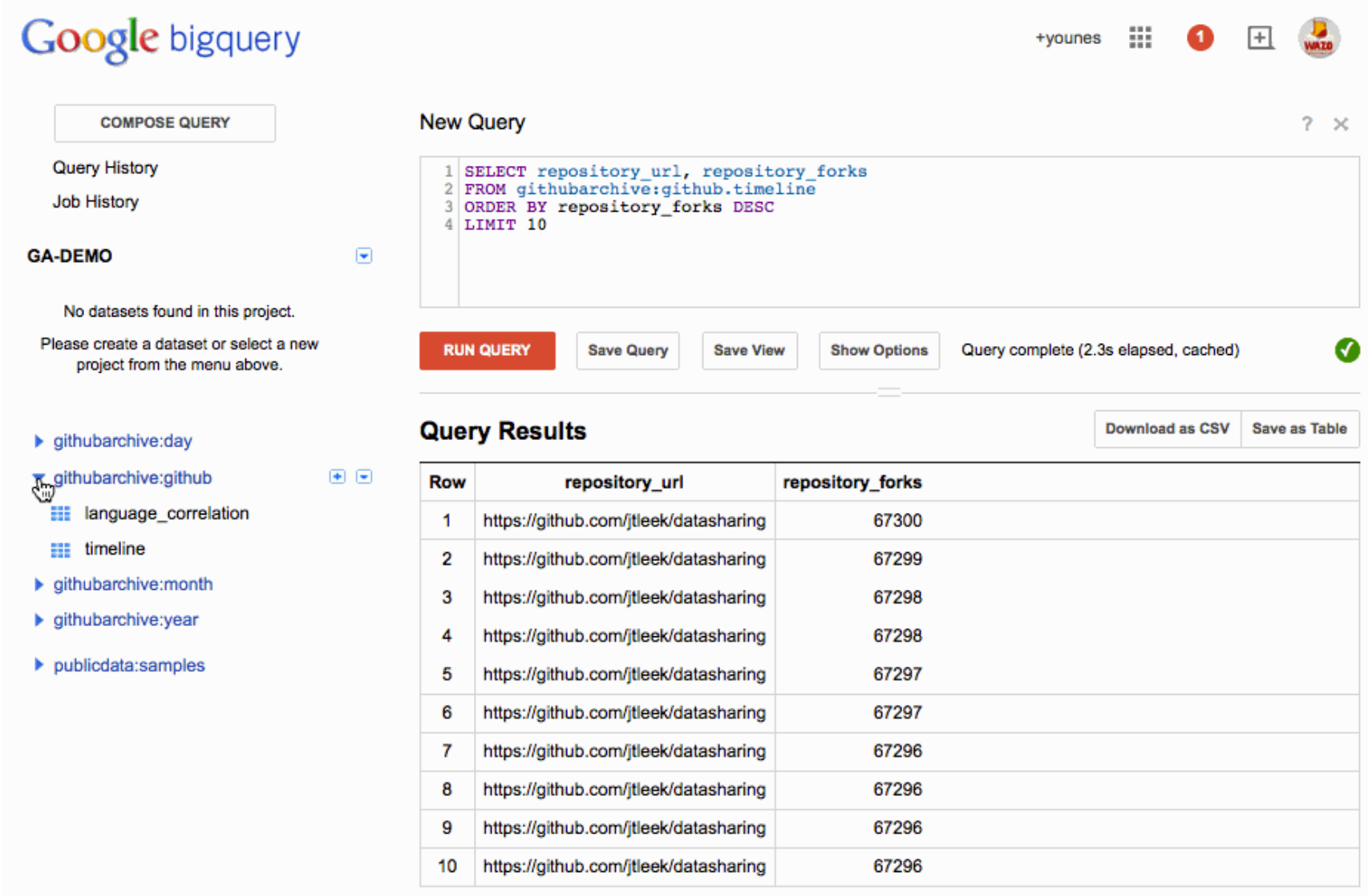
Prepare, explore and pre-process the data

- There are two basic services for this purpose:
  - The first one is BigQuery
  - The second one is Cloud Dataprep

- BigQuery
  - A fast Hadoop base system to perform SQL like query on very very large dataset

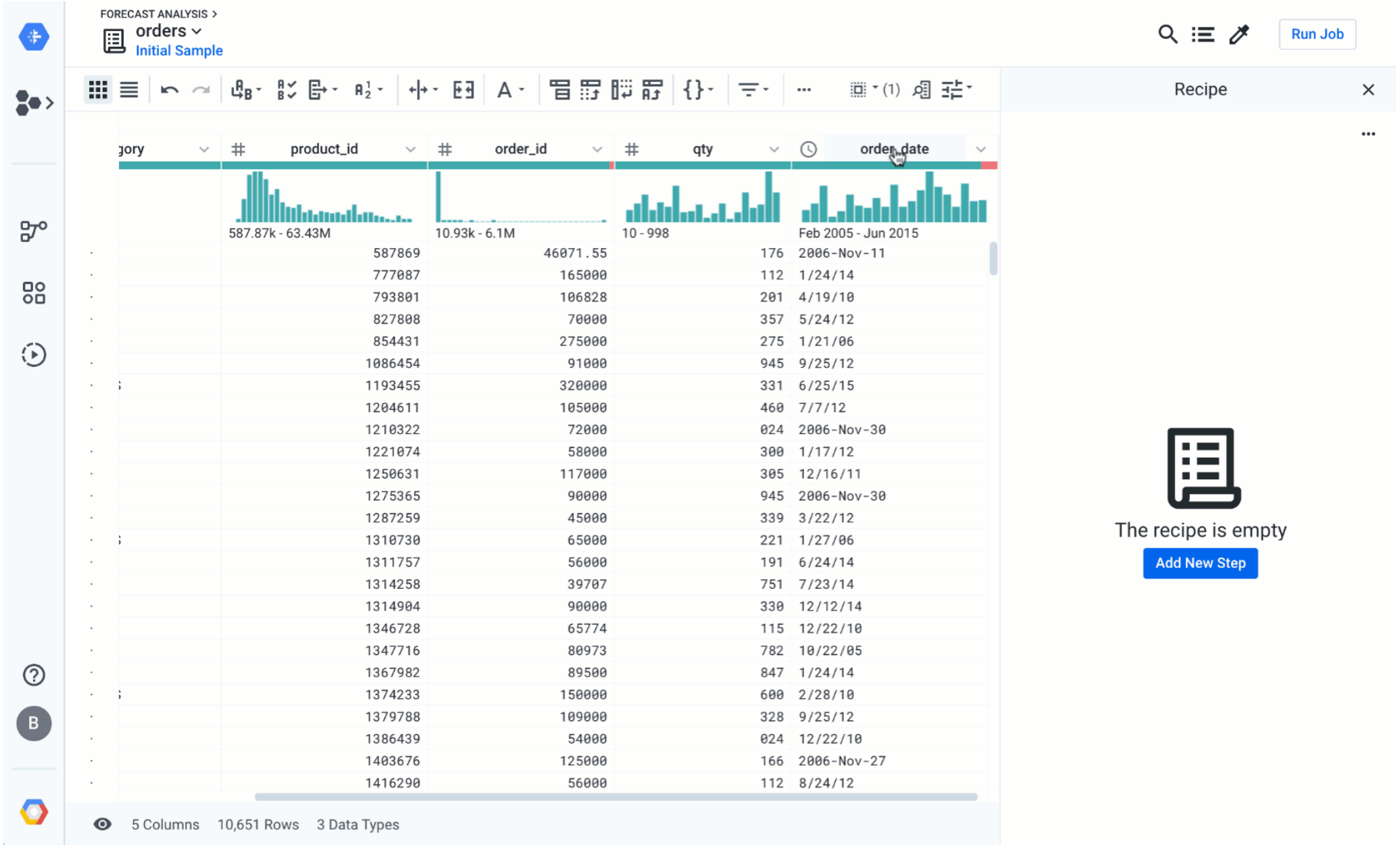
- Cloud Dataprep
  - A fast Hadoop based system to perform SQL like query on very very large dataset

Google BigQuery

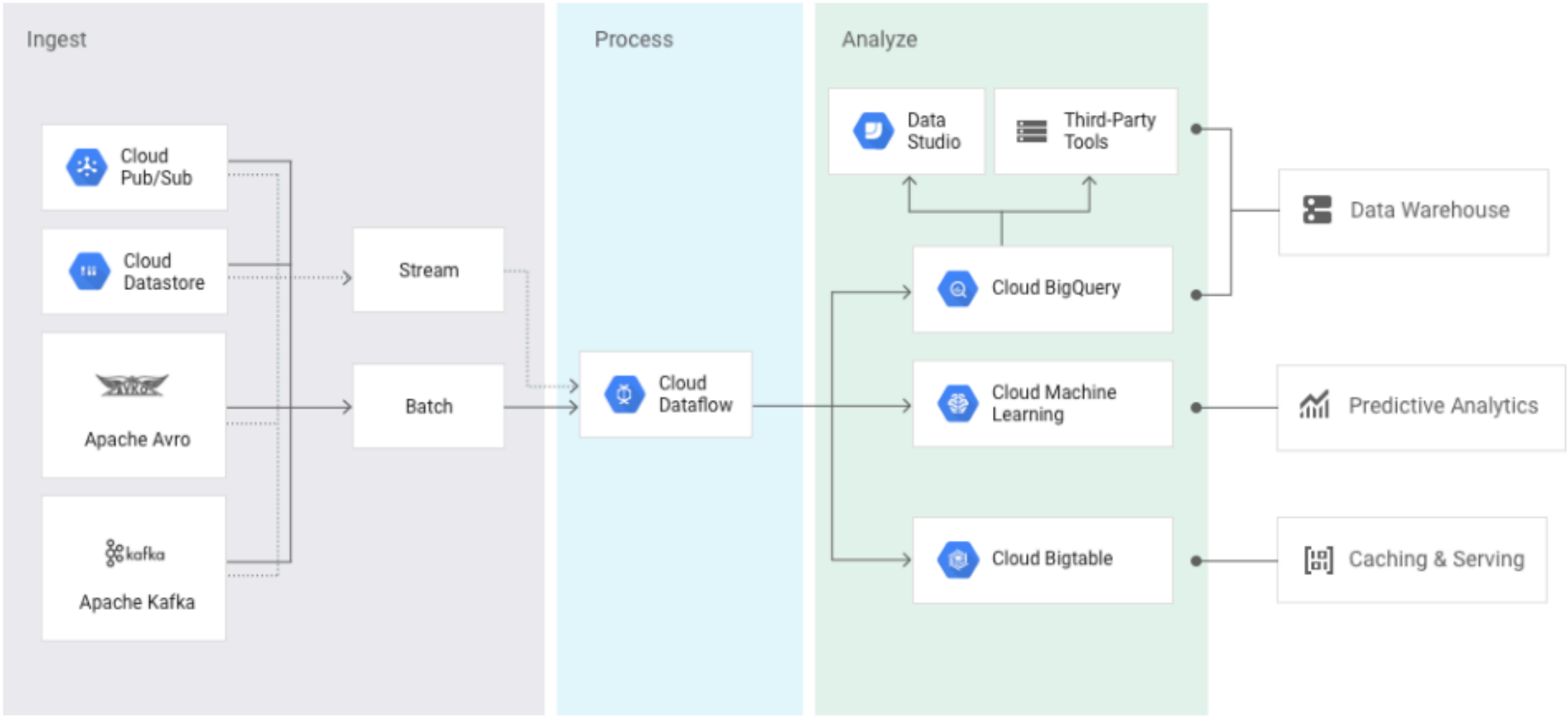


Row	repository_url	repository_forks
1	https://github.com/jtleek/datasharing	67300
2	https://github.com/jtleek/datasharing	67299
3	https://github.com/jtleek/datasharing	67298
4	https://github.com/jtleek/datasharing	67298
5	https://github.com/jtleek/datasharing	67297
6	https://github.com/jtleek/datasharing	67297
7	https://github.com/jtleek/datasharing	67296
8	https://github.com/jtleek/datasharing	67296
9	https://github.com/jtleek/datasharing	67296
10	https://github.com/jtleek/datasharing	67296

Cloud Dataprep



For more complex situations..



Google Cloud

Your model in Tensorflow – Using managed Jupyter notebooks

jupyter Untitled (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [1]:
```
import tensorflow as tf

hello = tf.Variable('Hello World!')

sess = tf.Session()
init = tf.global_variables_initializer()

sess.run(init)
sess.run(hello)
```

Out[1]: b'Hello World!'

In [ ]:

TensorBoard SCALARS IMAGES GRAPHS INACTIVE

Show data download links

Ignore outliers in chart scaling

Tooltip sorting method: default

Smoothing 0.6

Horizontal Axis

STEP RELATIVE WALL

Runs

Write a regex to filter runs

train

eval

TOGGLE ALL RUNS

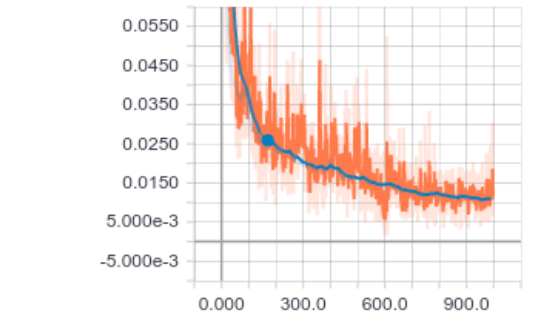
/tmp/mnist-logs

Filter tags (regular expressions supported)

accuracy 1

cross entropy 1

cross entropy



run to download CSV JSON

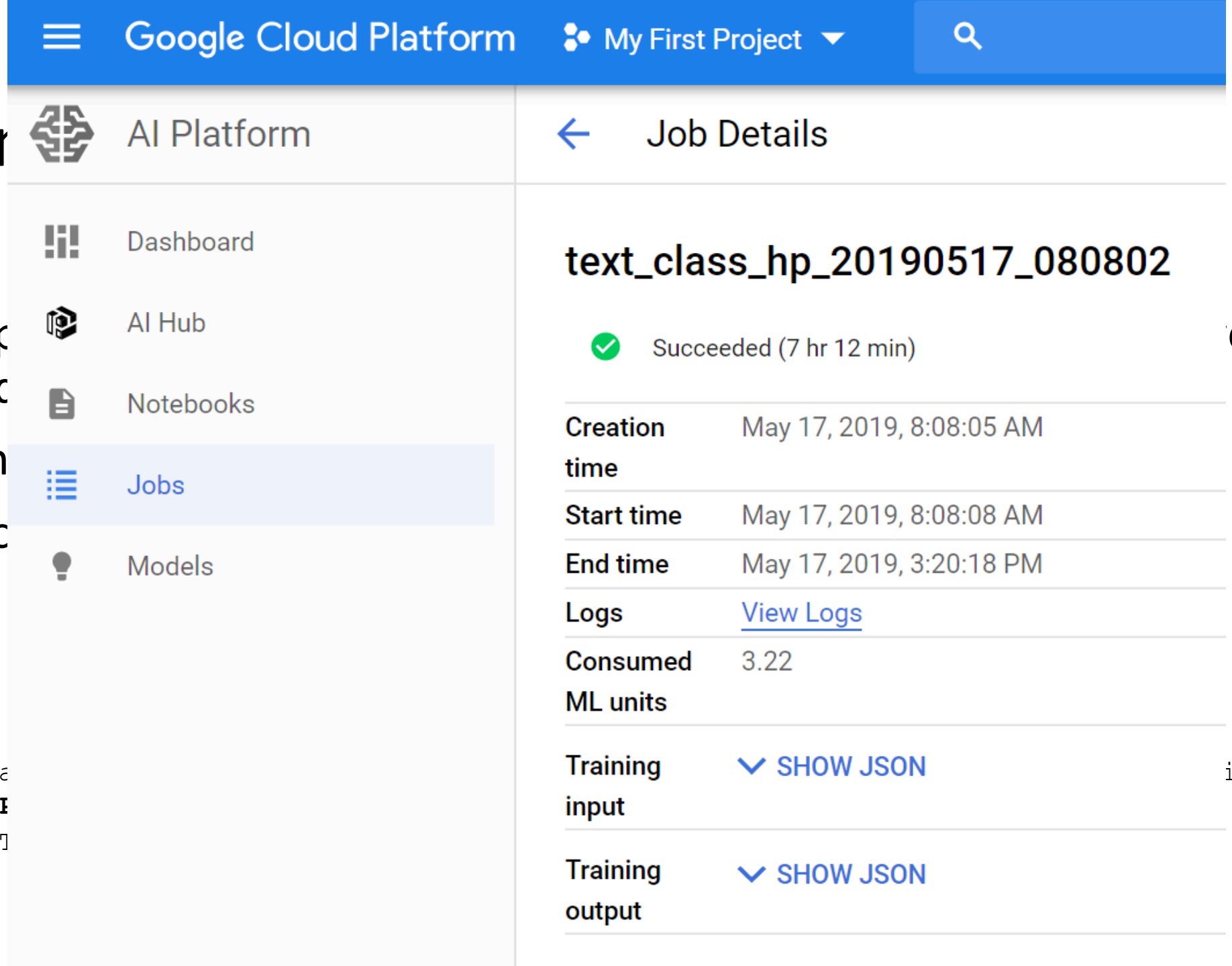
Name	Smoothed	Value	Step	Time	Relative
eval	0.02591	0.02550	170.0	Mon Sep 12, 15:40:41	8s
train	0.02851	0.03362	166.0	Mon Sep 12, 15:40:40	7s

max 4

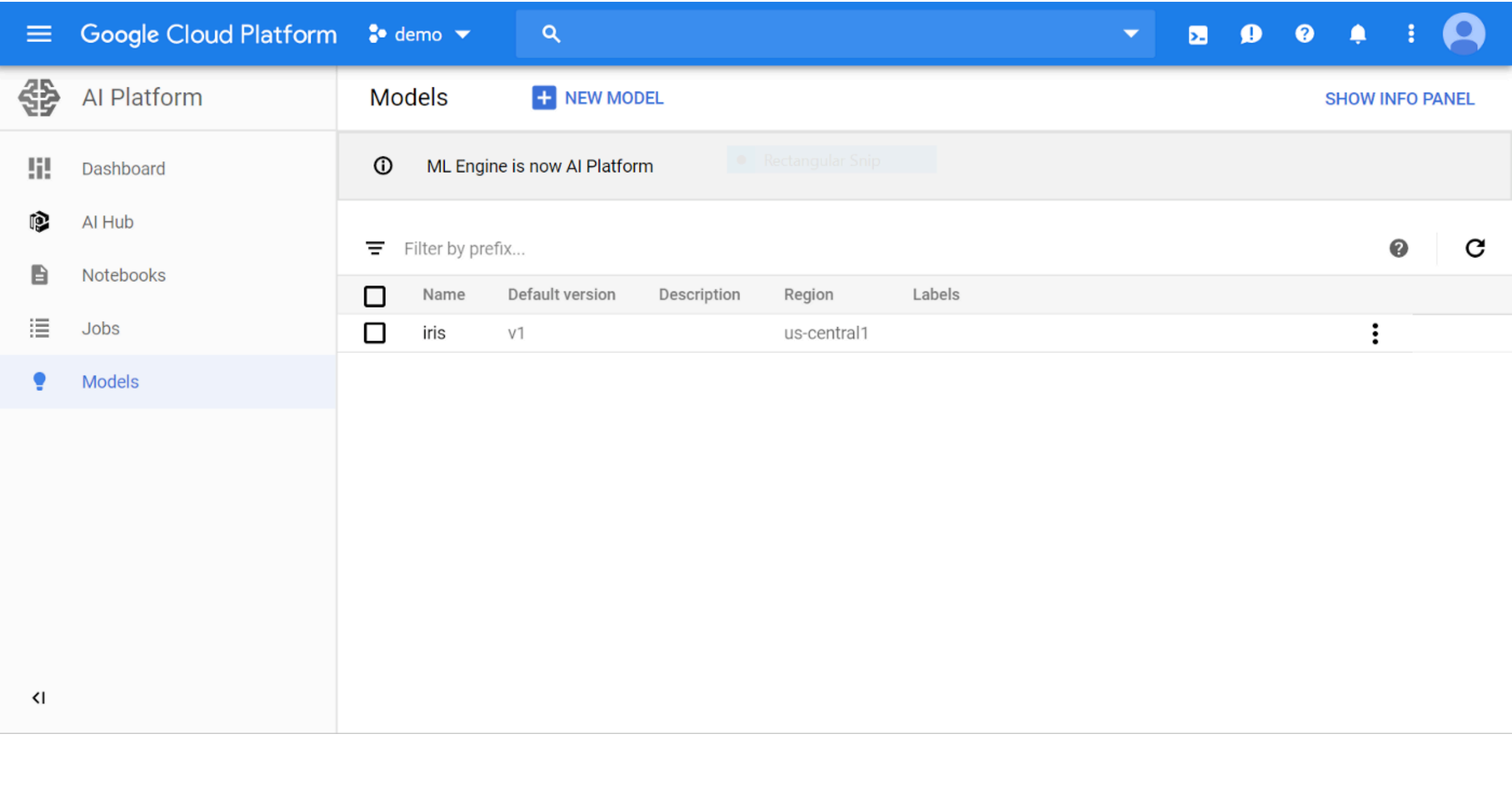
mean 4

Subr

- You p ork
  stanc
- Then
- Whic




```
gsutil gcloud a                                                    iner.task --
package-path=$I
bucket=${BUCKET
```


Google Cloud Platform
demo
AI Platform
Dashboard
AI Hub
Notebooks
Jobs
Models
Models
NEW MODEL
SHOW INFO PANEL
ML Engine is now AI Platform
Filter by prefix...
Name
Default version
Description
Region
Labels
iris
v1
us-central1

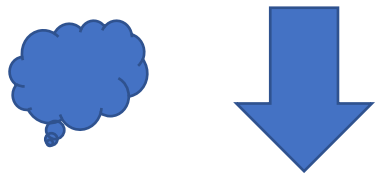
Deploy the model – A client for the MLaaS project

- You could deploy an **Endpoint** (via API) accessible from outside in order to use the trained model, fully running on GCP.
- The main advantage is that you may serve to end users the latest version of your model without caring about SW releases, servers and so on.
- REST API for online prediction
- … or batch prediction if you have thounsands preditions to be performed

How it works ?

JSON Input:

```
data = { 'instances': [ { 'key': 'b1', 'is_male': 'True', 'mother_age': 26.0,
'plurality': 'Single(1)', 'gestation_weeks': 39 },
{ 'key': 'g1', 'is_male': 'False', 'mother_age': 29.0, 'plurality': 'Single(1)',
'gestation_weeks': 38 },
{ 'key': 'b2', 'is_male': 'True', 'mother_age': 26.0, 'plurality': 'Triplets(3)',
'gestation_weeks': 39 },
{ 'key': 'u1', 'is_male': 'Unknown', 'mother_age': 29.0, 'plurality': 'Multiple(2+)',
'gestation_weeks': 38 }, ] }
```



response = requests.post(api, json=data)
https://ml.googleapis.com/v1/projects/**{}**/models/**{}**/versions/**{}**:predict'

print(response.content)

7.66, 7.22, 6.32

Take home messages

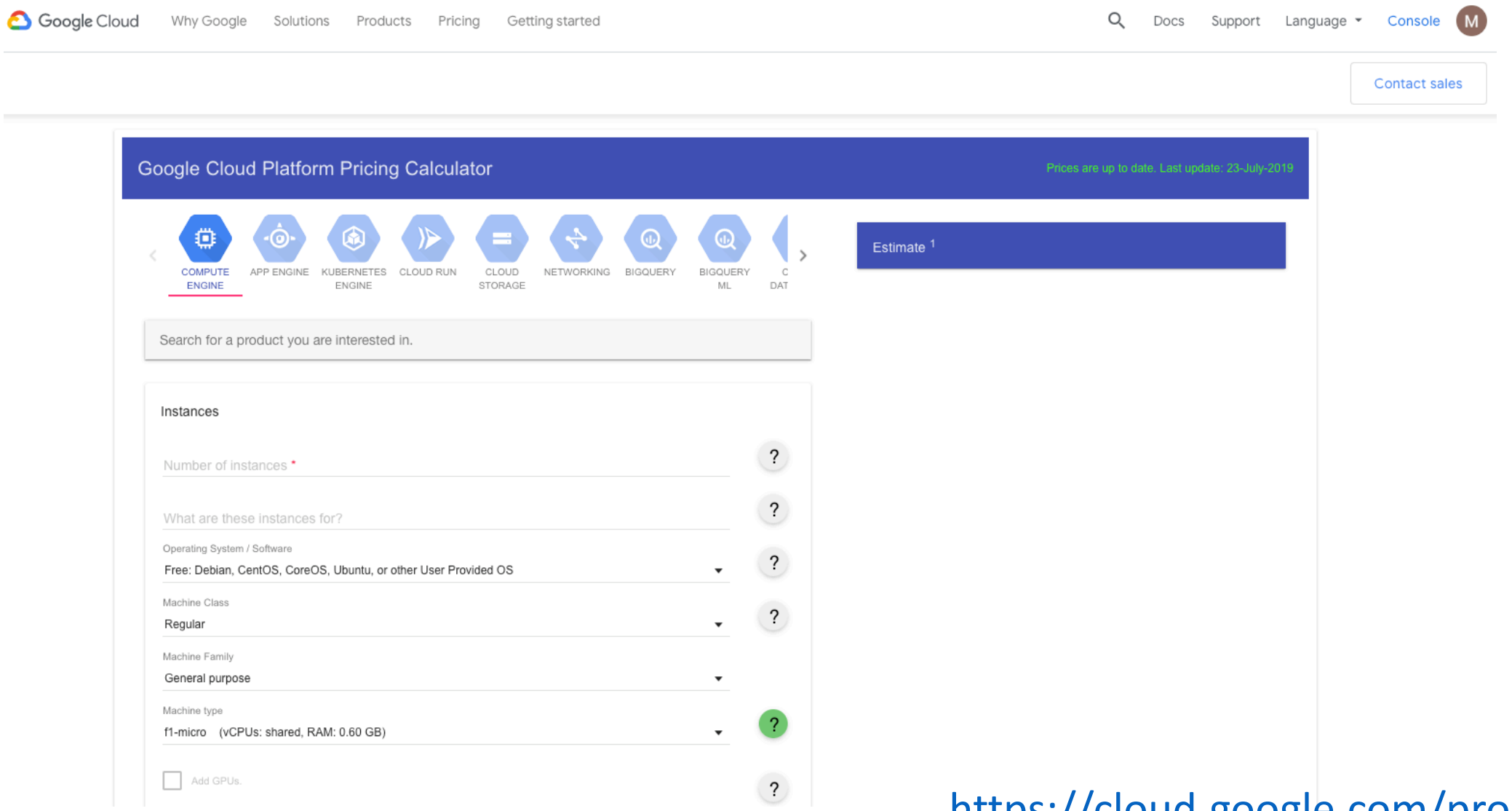
- Cloud platforms (Google, Amazon, ….) offer a HUGE variety of tools that will greatly help Data Scientists to perform their jobs without taking care of the infrastructures. **Try out and select the one you prefer!**

- The deploy-phase of trained model is really simple, for online and batch prediction. Very good for communicating applications (or to expose a service to the community).

- We didn’t discuss … but there are also tons of off-the-shelf trained algorithms to be used (especially for NLP or Image processing) or to be exploited for transfer learning

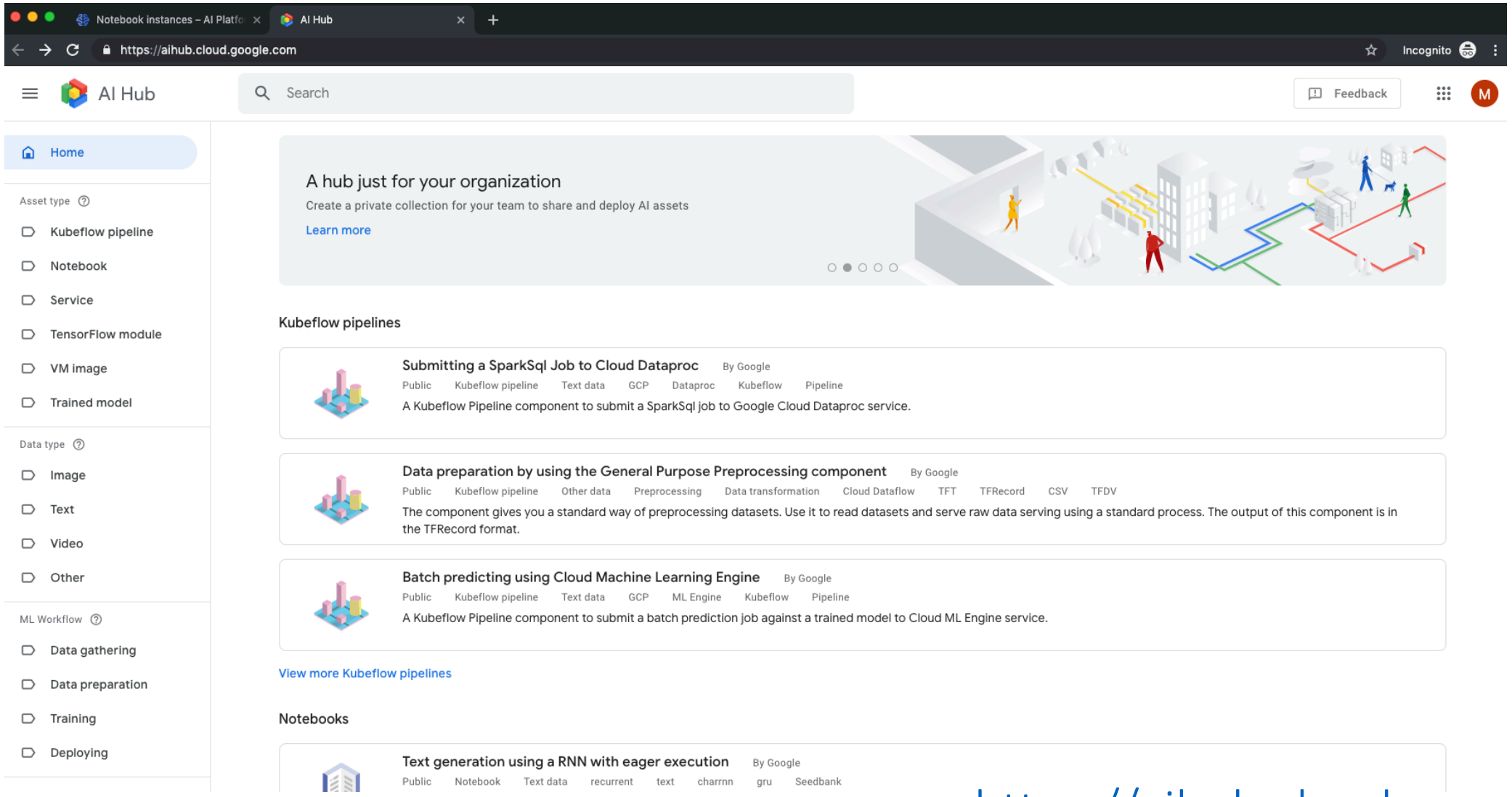
- Do not re-invent the wheel each time…

Always check the prices…



https://cloud.google.com/products/calculator/

AI Hub



https://aihub.cloud.google.com

Thank you for your attention!

Marco Landoni
INAF – National Insitute of Astrophysics
Brera Astronomical Observatory
marco.landoni@inaf.it

Osservatorio Astronomico di Brera
INAF

INAF
ISTITUTO NAZIONALE DI ASTROFISICA
20 ANNI DI RICERCA SCIENTIFICA DI ECCELLENZA