

Machine Learning as a Service

Application of Google Cloud Platform to Machine Learning problems

Marco Landoni

INAF – National Institute of Astrophysics

Brera Astronomical Observatory

marco.landoni@inaf.it



But what if it's Cloudy?

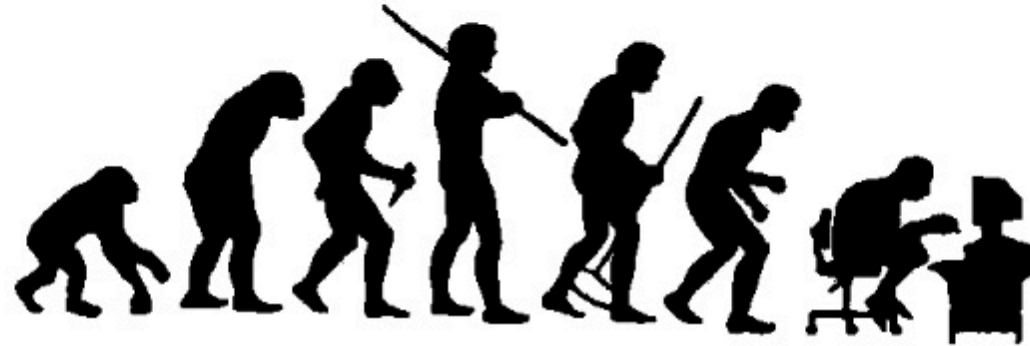
Warning!

- **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)



Mainframe
Computing



Personal
Computing



Client/Server
Computing



Mobile
Computing



Cloud
Computing

Various providers



Main Cloud Computing Services

- Computational power available for a reasonable price.
- 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



Compute Engine

Virtual Machines, Disks, Network



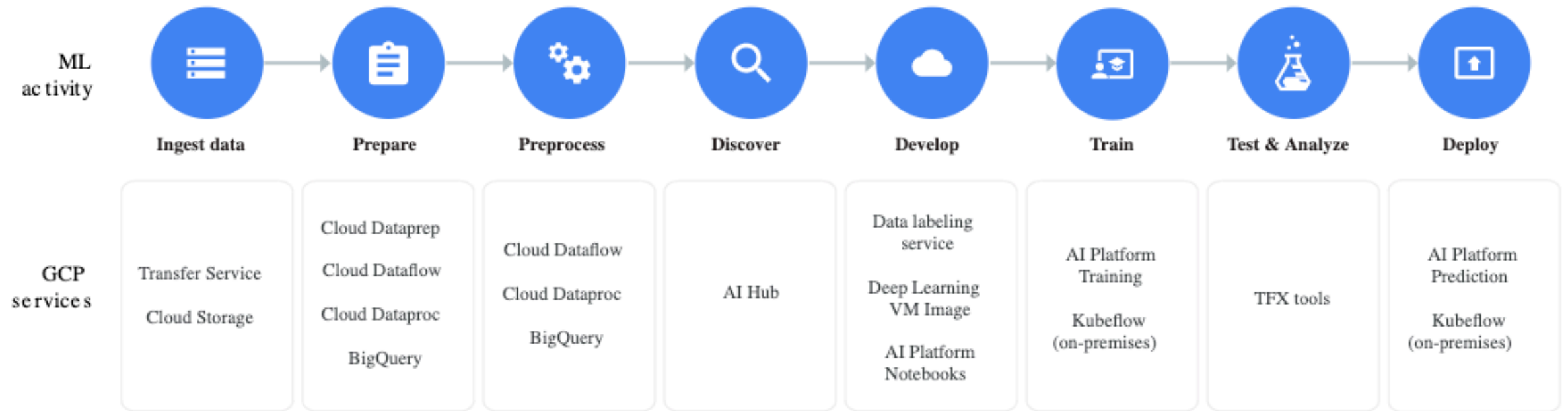
Machine Learning As A Service (MLaaS)

- 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

Machine learning development: the end-to-end cycle



The 1° page of the book

- Understand the console and then...
- Prepare the data

- Home
- Pins appear here
- Marketplace
- Billing
- APIs & Services
- Support
- IAM & admin
- Getting started
- Security
- COMPUTE
 - App Engine
 - Compute Engine
 - Kubernetes Engine
 - Cloud Functions
 - Cloud Run
- STORAGE
 - Bigtable

Info

Instance: `...-226516`

Number: `...109`

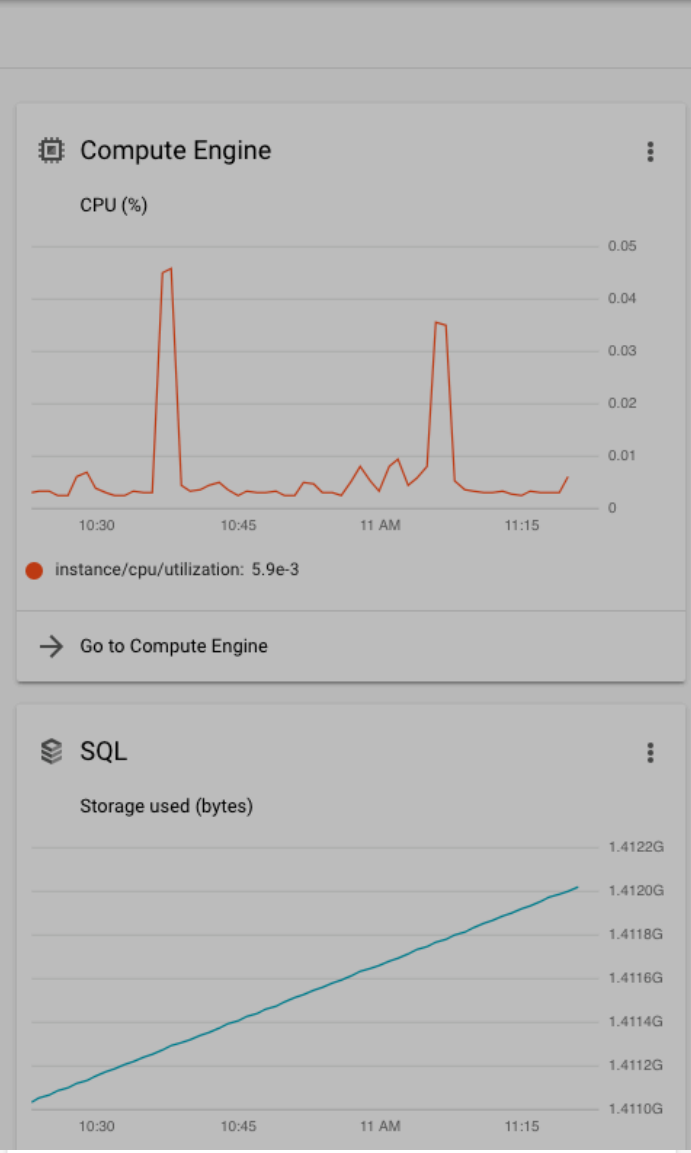
Instance settings

Resources

Compute Engine

Data from the past 7 days

with Stackdriver Trace



Google Cloud Platform status

All services normal

Go to Cloud status dashboard

Billing

Estimated charges EUR €33.65

For the billing period Jul 1 – 23, 2019

View detailed charges

Error Reporting

No sign of any errors. Have you set up Error Reporting?

Learn how to set up Error Reporting

News

- Introducing Spinnaker for Google Cloud Platform—continuous delivery made easy 16 hours ago
- A dozen reasons why Cloud Run complies with the Twelve-Factor App methodology 16 hours ago
- Work hacks from G Suite: how to host more effective meetings 17 hours ago



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



<input type="checkbox"/>	<input type="radio"/>	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



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

The screenshot displays the Google BigQuery web interface. At the top left is the 'Google bigquery' logo. On the right, there is a user profile for '+younes', a grid icon, a notification badge with the number '1', a plus sign, and a 'WATO' logo. Below the logo is a 'COMPOSE QUERY' button. The left sidebar contains 'Query History', 'Job History', and a section for 'GA-DEMO' which is currently empty, with a message: 'No datasets found in this project. Please create a dataset or select a new project from the menu above.' Below this, a list of datasets is shown: 'githubarchive:day', 'githubarchive:github' (selected with a mouse cursor), 'language_correlation', 'timeline', 'githubarchive:month', 'githubarchive:year', and 'publicdata:samples'. The main area is titled 'New Query' and contains a SQL query:

```
1 SELECT repository_url, repository_forks
2 FROM githubarchive:github.timeline
3 ORDER BY repository_forks DESC
4 LIMIT 10
```

 Below the query editor are buttons for 'RUN QUERY', 'Save Query', 'Save View', and 'Show Options'. A status message indicates 'Query complete (2.3s elapsed, cached)' with a green checkmark. The 'Query Results' section shows a table with 10 rows and 2 columns: 'repository_url' and 'repository_forks'. The results are sorted by 'repository_forks' in descending order. At the top right of the results section are buttons for 'Download as CSV' and 'Save as Table'.

Google bigquery

+younes

COMPOSE QUERY

Query History

Job History

GA-DEMO

No datasets found in this project.
Please create a dataset or select a new project from the menu above.

githubarchive:day

githubarchive:github

language_correlation

timeline

githubarchive:month

githubarchive:year

publicdata:samples

New Query

```
1 SELECT repository_url, repository_forks
2 FROM githubarchive:github.timeline
3 ORDER BY repository_forks DESC
4 LIMIT 10
```

RUN QUERY Save Query Save View Show Options Query complete (2.3s elapsed, cached)

Query Results

Download as CSV Save as Table

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

FORECAST ANALYSIS > orders Initial Sample

Recipe

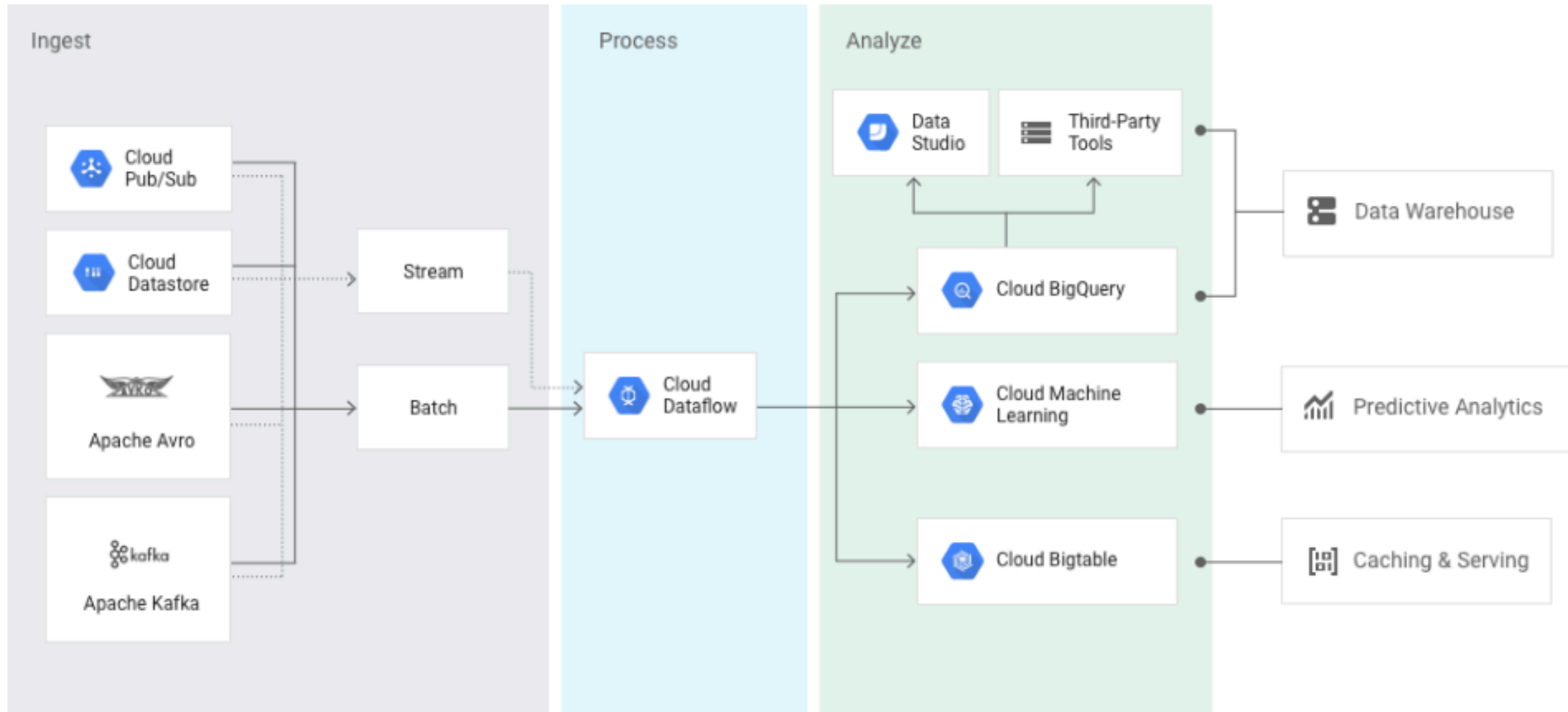
category	#	product_id	#	order_id	#	qty	order_date
	587.87k - 63.43M		10.93k - 6.1M		10 - 998		Feb 2005 - Jun 2015
		587869		46071.55		176	2006-Nov-11
		777087		165000		112	1/24/14
		793801		106828		201	4/19/10
		827808		70000		357	5/24/12
		854431		275000		275	1/21/06
		1086454		91000		945	9/25/12
		1193455		320000		331	6/25/15
		1204611		105000		460	7/7/12
		1210322		72000		024	2006-Nov-30
		1221074		58000		300	1/17/12
		1250631		117000		305	12/16/11
		1275365		90000		945	2006-Nov-30
		1287259		45000		339	3/22/12
		1310730		65000		221	1/27/06
		1311757		56000		191	6/24/14
		1314258		39707		751	7/23/14
		1314904		90000		330	12/12/14
		1346728		65774		115	12/22/10
		1347716		80973		782	10/22/05
		1367982		89500		847	1/24/14
		1374233		150000		600	2/28/10
		1379788		109000		328	9/25/12
		1386439		54000		024	12/22/10
		1403676		125000		166	2006-Nov-27
		1416290		56000		112	8/24/12

5 Columns 10,651 Rows 3 Data Types

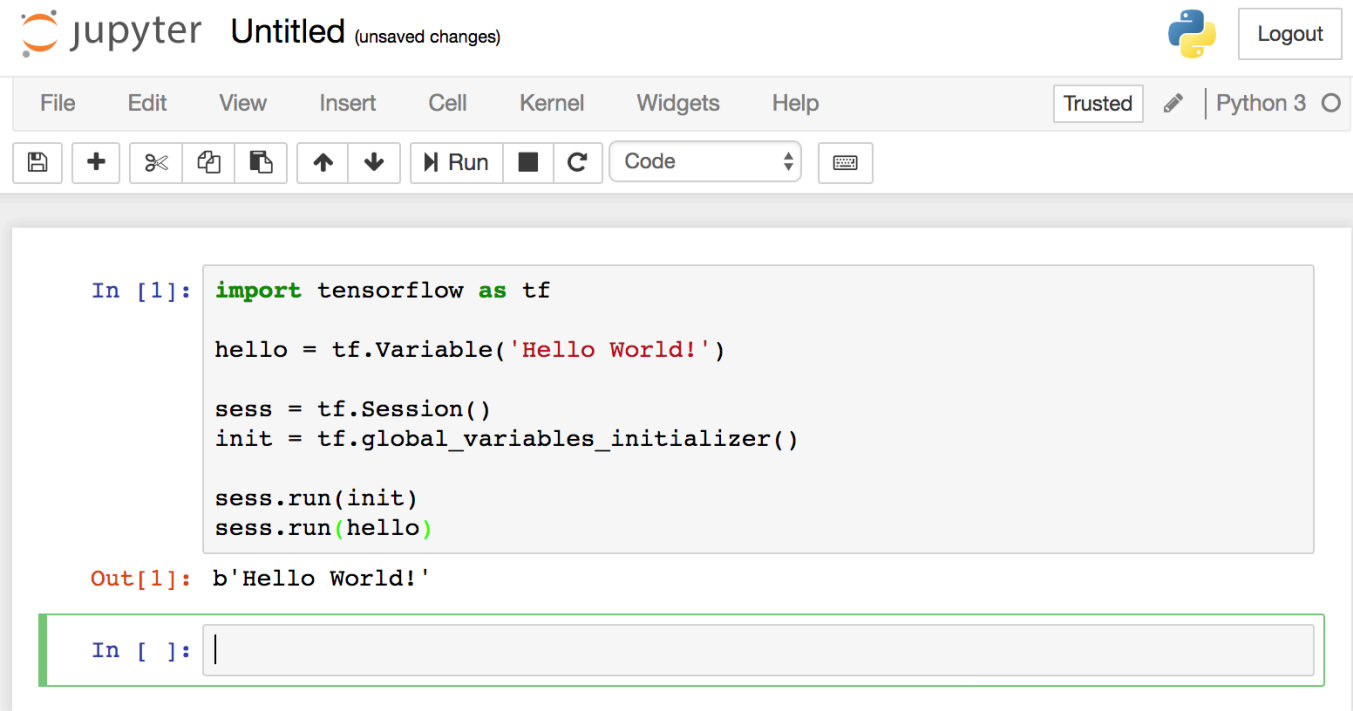
The recipe is empty

Add New Step

For more complex situations..



Your model in Tensorflow – Using managed Jupyter notebooks



The screenshot shows a Jupyter Notebook titled "Untitled (unsaved changes)". The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help), a "Trusted" status indicator, and a "Python 3" kernel. The code cell contains the following Python code:

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

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

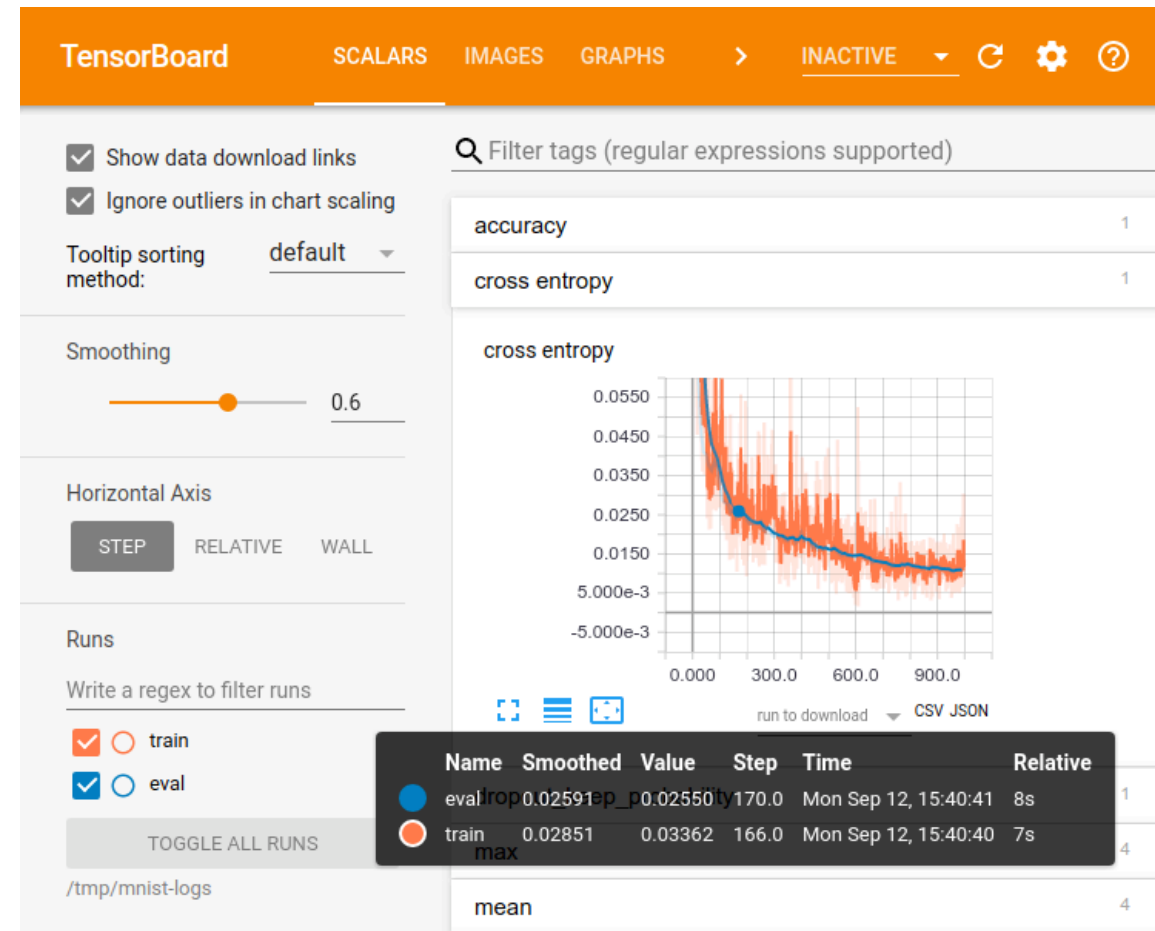
sess = tf.Session()
init = tf.global_variables_initializer()

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

The output of the code cell is:

```
Out[1]: b'Hello World!'
```

Below the output, there is an empty input cell for the next command.




Subr

- AI Platform
- Dashboard
- AI Hub
- Notebooks
- Jobs**
- Models

- You p
- stanc
- Then
- Whic

Job Details

text_class_hp_20190517_080802

 Succeeded (7 hr 12 min)

ork

Creation time May 17, 2019, 8:08:05 AM

Start time May 17, 2019, 8:08:08 AM

End time May 17, 2019, 3:20:18 PM

Logs [View Logs](#)

Consumed 3.22 ML units

Training input [SHOW JSON](#)

iner.task --

Training output [SHOW JSON](#)

```
gsutil gcloud a
package-path=$E
bucket=${BUCKET
```



AI Platform



Dashboard



AI Hub



Notebooks



Jobs



Models

Models

+ NEW MODEL

SHOW INFO PANEL



ML Engine is now AI Platform

Rectangular Snip



Filter by prefix...



Name

Default version

Description

Region

Labels



iris

v1

us-central1



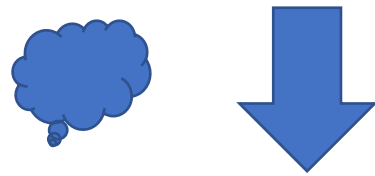
Deploy the model – A client for the MLaaS project

- 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 thousands predictions 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...

Contact sales

Google Cloud Platform Pricing Calculator

Prices are up to date. Last update: 23-July-2019



Estimate 1

Search for a product you are interested in.

Instances

Number of instances *



What are these instances for?



Operating System / Software

Free: Debian, CentOS, CoreOS, Ubuntu, or other User Provided OS



Machine Class

Regular



Machine Family

General purpose

Machine type

f1-micro (vCPUs: shared, RAM: 0.60 GB)



Add GPUs.



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

AI Hub

The screenshot shows the AI Hub web interface in a browser window. The browser's address bar displays `https://aihub.cloud.google.com`. The page features a navigation sidebar on the left with categories like 'Asset type', 'Data type', and 'ML Workflow'. The main content area includes a hero section with the text 'A hub just for your organization' and a carousel of Kubeflow pipelines. The first pipeline is 'Submitting a SparkSql Job to Cloud Dataproc', the second is 'Data preparation by using the General Purpose Preprocessing component', and the third is 'Batch predicting using Cloud Machine Learning Engine'. Below these is a 'Notebooks' section with the entry 'Text generation using a RNN with eager execution'.

Notebook instances - AI Platfo x AI Hub x +

https://aihub.cloud.google.com

Incognito

AI Hub Search Feedback M

Home

Asset type

- Kubeflow pipeline
- Notebook
- Service
- TensorFlow module
- VM image
- Trained model

Data type

- Image
- Text
- Video
- Other

ML Workflow

- Data gathering
- Data preparation
- Training
- Deploying

A hub just for your organization

Create a private collection for your team to share and deploy AI assets

[Learn more](#)

Kubeflow pipelines

Submitting a SparkSql Job to Cloud Dataproc By Google

Public Kubeflow pipeline Text data GCP Dataproc Kubeflow Pipeline

A Kubeflow Pipeline component to submit a SparkSql job to Google Cloud Dataproc service.

Data preparation by using the General Purpose Preprocessing component By Google

Public Kubeflow pipeline Other data Preprocessing Data transformation Cloud Dataflow TFT TFRecord CSV TFDV

The component gives you a standard way of preprocessing datasets. Use it to read datasets and serve raw data serving using a standard process. The output of this component is in the TFRecord format.

Batch predicting using Cloud Machine Learning Engine By Google

Public Kubeflow pipeline Text data GCP ML Engine Kubeflow Pipeline

A Kubeflow Pipeline component to submit a batch prediction job against a trained model to Cloud ML Engine service.

[View more Kubeflow pipelines](#)

Notebooks

Text generation using a RNN with eager execution By Google

Public Notebook Text data recurrent text charnn gru Seedbank

<https://aihub.cloud.google.com>

Thank you for your attention!

Marco Landoni

INAF – National Institute of Astrophysics

Brera Astronomical Observatory

marco.landoni@inaf.it