

Cloud for Data Analytics

4 - Introduction to the Google Cloud Platform

Juan Ángel Lorenzo del Castillo

École Internationale des Sciences du Traitement de l'Information

ING3 - Ingénierie Mathématique et Simulation Informatique

2019-2020



juanangel.lorenzodelcastillo@eisti.eu

Table of Contents

- 1 The Google Cloud Platform
- 2 Products and Services
- 3 Accessing the Cloud Platform
- 4 Machine Learning in the Google Cloud Platform
- 5 Final remarks

Table of Contents

- 1 The Google Cloud Platform
- 2 Products and Services
- 3 Accessing the Cloud Platform
- 4 Machine Learning in the Google Cloud Platform
- 5 Final remarks

The Google Cloud Platform

- The Google Cloud Platform (GCP) was originally built to power, ingest, manage and serve all data from Google's own applications (Gmail, Google Maps, Chrome, YouTube, Android, Play, Drive).
- Numerous big data and Machine Learning (ML) products are built on top of this infrastructure.
- GCP provides lots of compute power to run big data jobs.
- Google trains its production ML models on its network of data centres and then deploys smaller trained versions of these models to the hardware on your phone for accurate predictions.

The Google Cloud Platform

Example: Automatic Video Stabilisation on the Pixel smartphone with Machine Learning.



[Link on YouTube](#)

Data sources as input for training the ML model:

- Video data (frames ordered by timestamps)
- Context data: series data on the camera's position and orientation from the on-board gyroscope and motion on the camera lens.

The Google Cloud Platform

For an eight-Megapixel camera:

- 8 million pixels per image \times 3-channel layers \simeq 23 million data points per image frame.
- 30 frames per second of video \rightarrow quickly a short video becomes over a billion data points to feed into the model.

Conclusion:

We need a huge infrastructure to feed and train our ML models.

Table of Contents

- 1 The Google Cloud Platform
- 2 Products and Services**
- 3 Accessing the Cloud Platform
- 4 Machine Learning in the Google Cloud Platform
- 5 Final remarks

Products and Services

■ Main GCP products and services



Products and Services

Some products and services in detail:

■ Compute

- ▶ **Google Compute Engine (GCE):** IaaS component. It enables users to launch Windows or Linux virtual machines (VMs) on demand.
- ▶ **Google App Engine (GAE):** PaaS to deploy (developing and hosting) Java, PHP, Node.js, Python, C#, .Net, Ruby and Go web applications.
- ▶ **Google Kubernetes engine (GKE):** Environment for deploying and orchestrating containerized applications.
- ▶ **Google Cloud Functions:** Serverless execution environment for building and connecting cloud services. Single-purpose functions are attached to events emitted from the cloud infrastructure, and they are triggered when an event being watched is fired.

Products and Services

Some products and services in detail:

■ Storage and Databases

- ▶ **Cloud Storage:** Integrates storage into apps with a single unified API. All data are held into basic containers called *Buckets*.
- ▶ **Persistent disk:** Reliable, high-performance block storage which can be attached to virtual machine instances running in either GCE or GKE.
- ▶ **Cloud SQL:** fully-managed database service to set up, maintain, manage, and administer relational PostgreSQL and MySQL databases in the cloud.
- ▶ **Cloud Bigtable:** A managed NoSQL database service for large analytical and operational workloads on non-structured key/value data.
- ▶ **Cloud Datastore:** Highly-scalable NoSQL database for applications that rely on highly available structured data at scale.

Products and Services

Some products and services in detail:

■ Big Data and Data Analytics

- ▶ **Big Query:** It allows storing and querying massive datasets by enabling super-fast SQL queries.
- ▶ **Cloud Dataflow:** A service for executing a wide variety of data processing patterns in stream (real time) and batch (historical) modes.
- ▶ **Cloud Dataproc:** To run Apache Spark and Apache Hadoop clusters.
- ▶ **Cloud Datalab:** Interactive tool, built on Jupyter, to explore, analyse, transform and visualize data and build machine learning models.

■ AI and machine learning

- ▶ **Cloud TPU:** To train and run Machine Learning workloads on Tensor Processing Units (TPUs).
- ▶ **Cloud Machine Learning Engine:** Training and prediction Machine Learning services. Available with TensorFlow or scikit-learn & XG Boost.
- ▶ **Cloud Vision:** Offers pretrained machine learning models to classify images into categories.

Table of Contents

- 1 The Google Cloud Platform
- 2 Products and Services
- 3 Accessing the Cloud Platform**
- 4 Machine Learning in the Google Cloud Platform
- 5 Final remarks

The GCP Console

► <http://console.cloud.google.com/>

The screenshot shows the Google Cloud Platform console interface. Callouts identify the following elements:

- Main menu:** The hamburger menu icon in the top left corner.
- Current project:** The text "Account Automation" in the top navigation bar.
- Search bar:** The magnifying glass icon in the top navigation bar.
- Utilities:** A group of icons (notifications, help, etc.) in the top navigation bar.
- Context menu:** The "Home" link in the left sidebar.
- Current user:** The user profile icon in the top right corner.
- Minimizes menu:** The hamburger menu icon in the bottom left corner.

The main content area displays a "Dashboard" with several cards:

- Try Compute Engine:** A blue card with a "Get started" button.
- Learn to use Cloud Storage:** A white card with a "Get started" button.
- Learn Google Cloud Platform:** A white card with a "Get started" button.
- Use Google APIs:** A white card with a "Get started" button.

Source: Google

Creating and Managing Projects

- A project consists of a set of users, a set of APIs, and billing, authentication, and monitoring settings for those APIs.
- <https://cloud.google.com/resource-manager/docs/creating-managing-projects>

The Google Cloud Platform

Demo: Creating a VM on Computer Engine

- spin up a Compute Engine instance,
- run some software on it,
- copy files from that Compute Engine instance to cloud storage,
- publish those files to make them public.

Use case:

Earthquake activity from the U.S. Geological Survey website.

The Google Cloud Platform

Procedure:

1. Go to `cloud.google.com`
2. Select the *Compute Engine* section, go to *VM instances* and create a VM called "earthquake-VM".
3. Pick a zone, one virtual CPU and 10 GB of space.
4. Allow full access to all the cloud APIs, so that we can write to Cloud Storage from the VM.
5. Create the VM.
6. Access the VM console through SSH.
7. Install git.
8. Do `git clone https://www.github.com/GoogleCloudPlatform/training-data-analyst`.
9. Go into `courses/bdml_fundamentals/demos`, and there is an `earthquakevm` directory.
10. Inspect the content of the `ingest.sh` file.
11. Inspect the content of the `transform.py` file.
12. Install the missing Python packages listed in `install_missing.sh` by running this file.

The Google Cloud Platform

Procedure:

13. Run `ingest.sh` to download `earthquakes.csv`.
14. Inspect the beginning of `earthquakes.csv` with `head` and check the first few earthquake occurrences in the US.
15. Run `transform.py` to create an image (`earthquakes.png`) out of the csv file.
16. Create an `earthquake` bucket (Storage/browser) and verify with `gsutil ls gs://your-bucket-name` that it is empty.
17. Copy the produced data off the VM into cloud storage so that we can delete the VM.
18. Copy `earthquakes.*` to the bucket with `gsutil cp` and refresh the bucket in the GCP page.
19. Stop the VM (will be paused, I will keep paying for the storage but I will not pay for the computing) or delete it.
20. Make the data public so that we can access it from everywhere: go to *Permissions* and add members called *allUsers*, and give those users a *Storage Object Viewer* role.
21. Check the public links associated with our files and click on the htm's file link.

Table of Contents

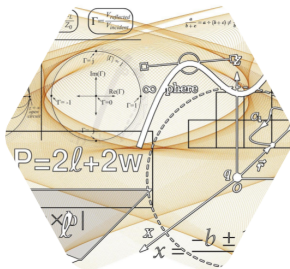
- 1 The Google Cloud Platform
- 2 Products and Services
- 3 Accessing the Cloud Platform
- 4 Machine Learning in the Google Cloud Platform**
- 5 Final remarks

Machine Learning

The popular imagination of what ML is



Lots of data



Complex mathematics in multidimensional spaces



Magical results

Source: Google

Machine Learning

In reality ML is...



Define objectives



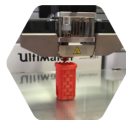
Collect data



Understand and prepare the data



Create and evaluate the model



Refine the model

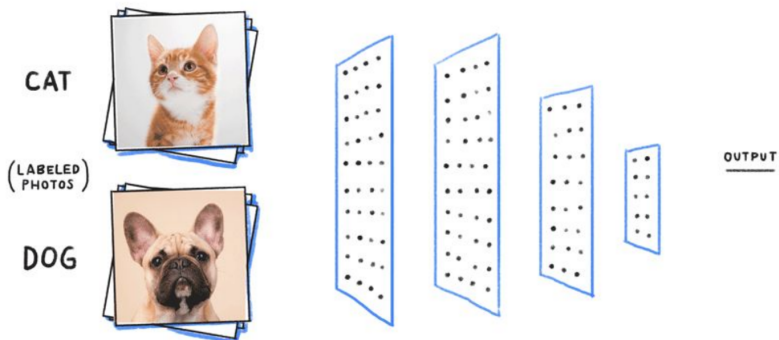


Serve and monitor the model

Source: Google

Neural Networks

A Neural Network is a function that can learn



Source: Google

Neural Networks

Demonstration of neural networks:

► <http://playground.tensorflow.org>

Neural Networks

Puppy or muffin?



Neural Networks

Sometimes is not that easy...



Source: boredpanda.com

Neural Networks

And can get worse...



Source: boredpanda.com

Machine Learning

Machine Learning is learning from
rules **plus experience.**

Machine Learning and GCP

Neural Networks aren't new. The compute power they need is...

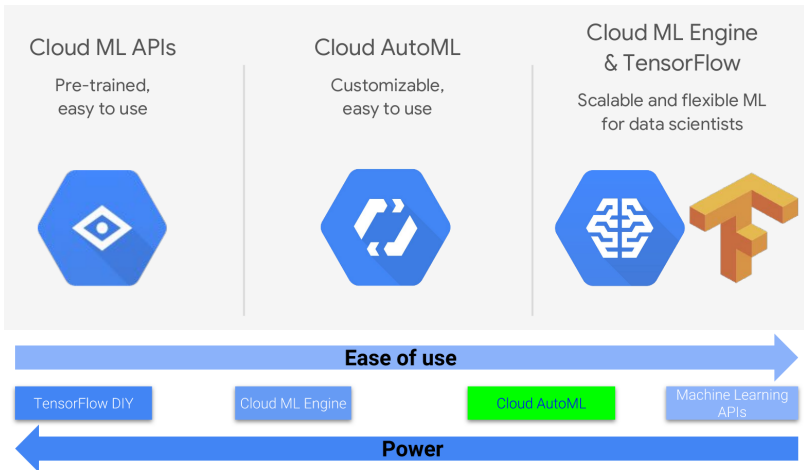
- Neural Networks have been around since the 70s
- But they fell out of favour because the compute power needed to train them was unrealistic at the time

Machine Learning and GCP

Neural Networks aren't new. The compute power they need is...

- Even in 2018 deep learning algorithms are often bottlenecked by your compute power
- Some platforms, like Google Cloud Platform or Microsoft Azure give anyone access to a supercomputer (or a few)

Machine Learning Services from Google Cloud Platform



Source: google.com

Machine Learning Services from Google Cloud Platform

Machine Learning as an API



Cloud
Vision API



Cloud
Speech API



Cloud
Natural Language
API



Cloud
Translation API



Cloud Video
Intelligence API

Use your own data to train models



TensorFlow

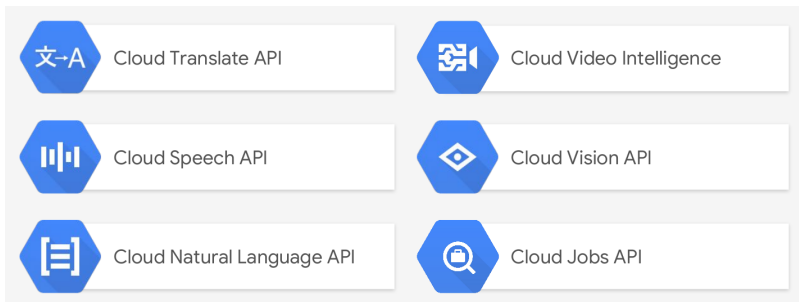


Cloud Machine
Learning Engine

Source: google.com

Cloud ML APIs

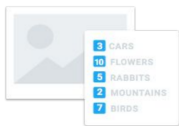
Cloud ML APIs: pre-trained ML models



Source: google.com

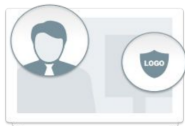
Cloud ML APIs

Example: Cloud Vision API



Object Recognition

Detect broad range of categories within an image, including landmarks, modes of transport, and animals.



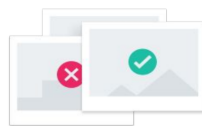
Facial Sentiment Detection
Logo Recognition

Detect faces and analyze facial features to detect emotions: joy, sorrow, anger.
Detect popular logos.



Optical Character Recognition

Detect and extract text within an image, with support for a broad range of languages, along with support for automatic language identification.



Explicit Content Detection

Detect different types of inappropriate content from adult to violent content. Powered by Google Safe Search

Source: google.com

Cloud ML APIs

Example: Cloud Vision API

- Example of the Cloud Vision API on <https://cloud.google.com/vision/>
- Try it by dragging different images from the Internet or from those contained in the Arel package *muffvsdog.zip*

Optional:

- Codelabs: [Detect Labels, Faces, and Landmarks in Images with the Cloud Vision API](#)

Cloud AutoML

State of the Industry: Complex & Time Intensive



Define objectives



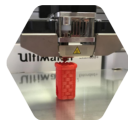
Collect data



Understand and prepare the data



Create and evaluate the model



Refine the model

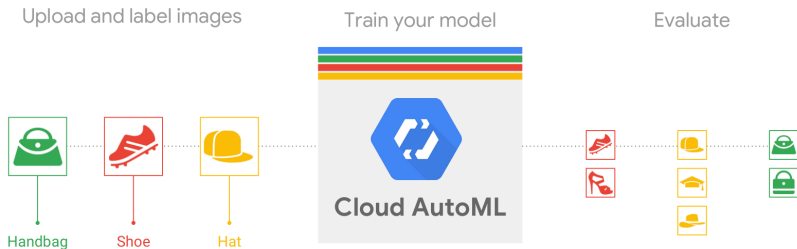


Serve and monitor the model

Source: google.com

Cloud AutoML

A technology that can automatically create a Machine Learning Model
Cloud AutoML Vision:

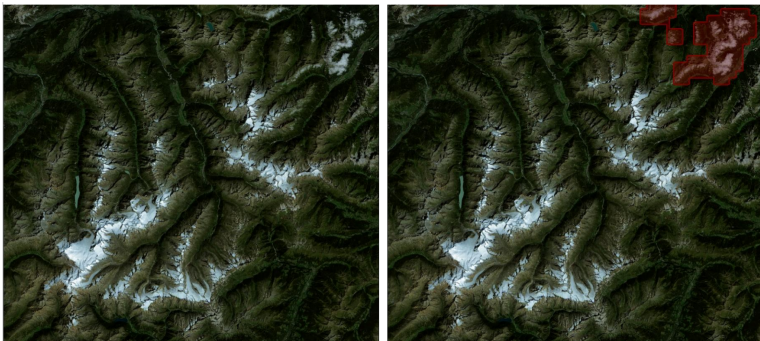


Source: google.com

Cloud AutoML

Exercise:

- Codelabs: [Classify Images of Clouds in the Cloud with AutoML Vision](#)



Source: google.com and Airbus

Cloud ML Engine & Tensorflow

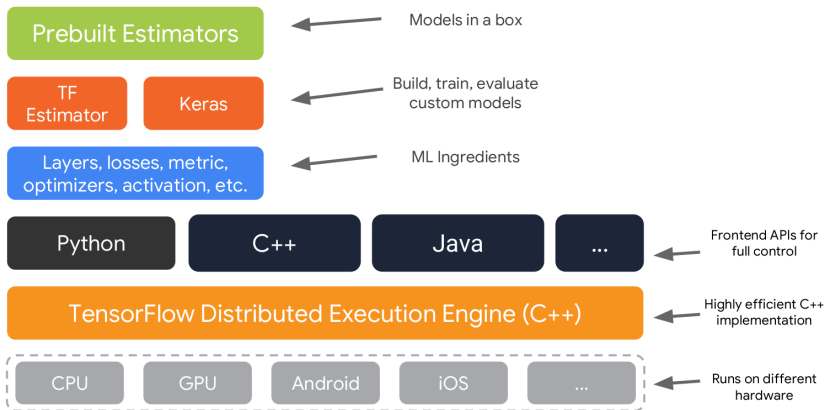
Tensorflow:

- Open source software for ML development
- tensorflow.org launched in November 2015
- Created by the Google Brain team
- Standard tool for ML development at Google



Cloud ML Engine & Tensorflow

Tensorflow layers:



Source: google.com

Cloud ML Engine & Tensorflow

Exercise:

- Codelabs: [End-to-end Machine Learning with Tensorflow on GCP](#)

Table of Contents

- 1 The Google Cloud Platform
- 2 Products and Services
- 3 Accessing the Cloud Platform
- 4 Machine Learning in the Google Cloud Platform
- 5 Final remarks**

Final remarks

More resources to start with the GCP:

- Quickstarts: [Getting started with the GCP](#)

Remember: Your GCP credit is limited! Remember to turn off services when they are not in use.