



Large-scale analytics and distributed machine learning with TensorFlow, Cloud ML, and Dataflow (Apache Beam)

Kaz Sato & Amy Unruh



Kaz Sato

Staff Developer Advocate
Tech Lead for Data and Analytics
Google Cloud Platform



Amy Unruh

Staff Developer Program Engineer
Google Cloud Platform



What we'll cover

Deep Learning and Distributed Training

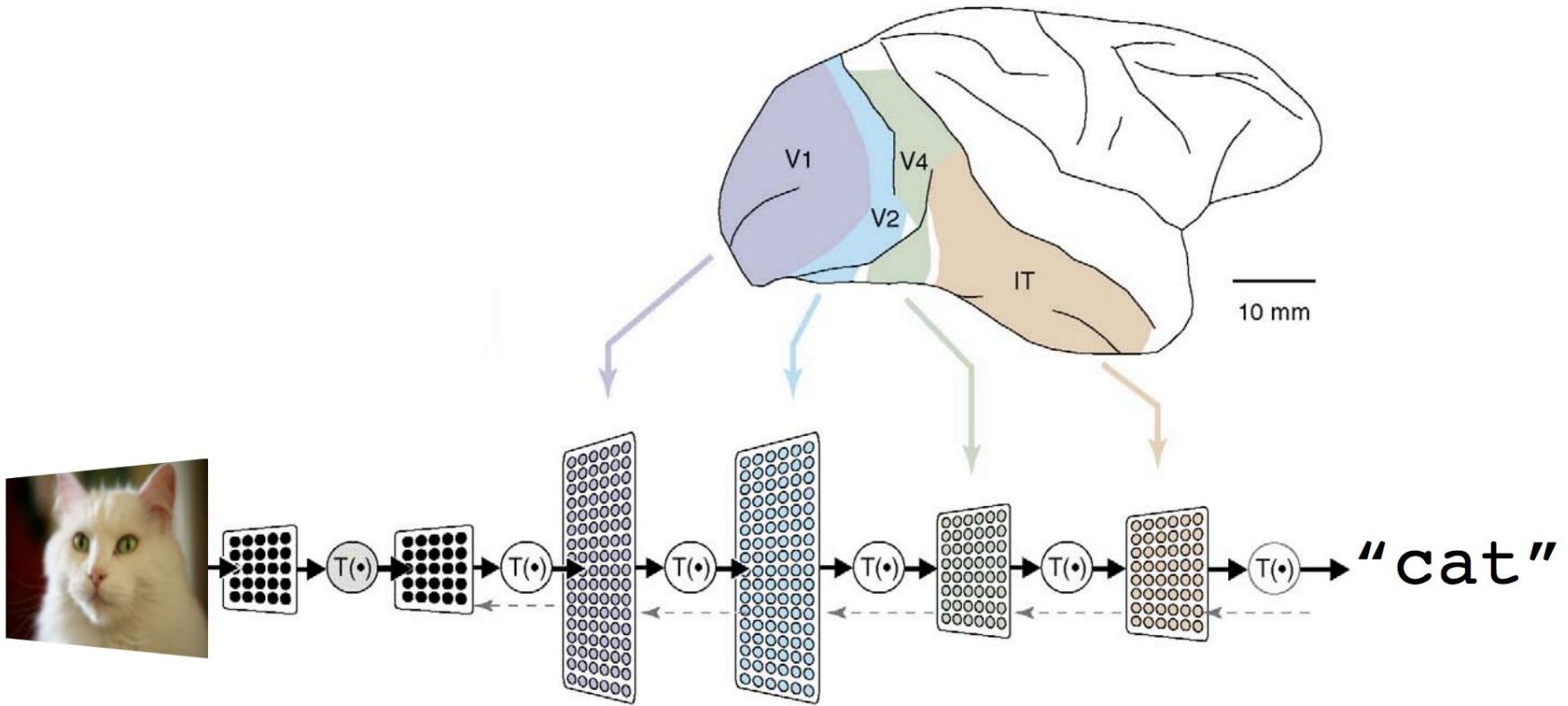
Large scale neural network on Google Cloud

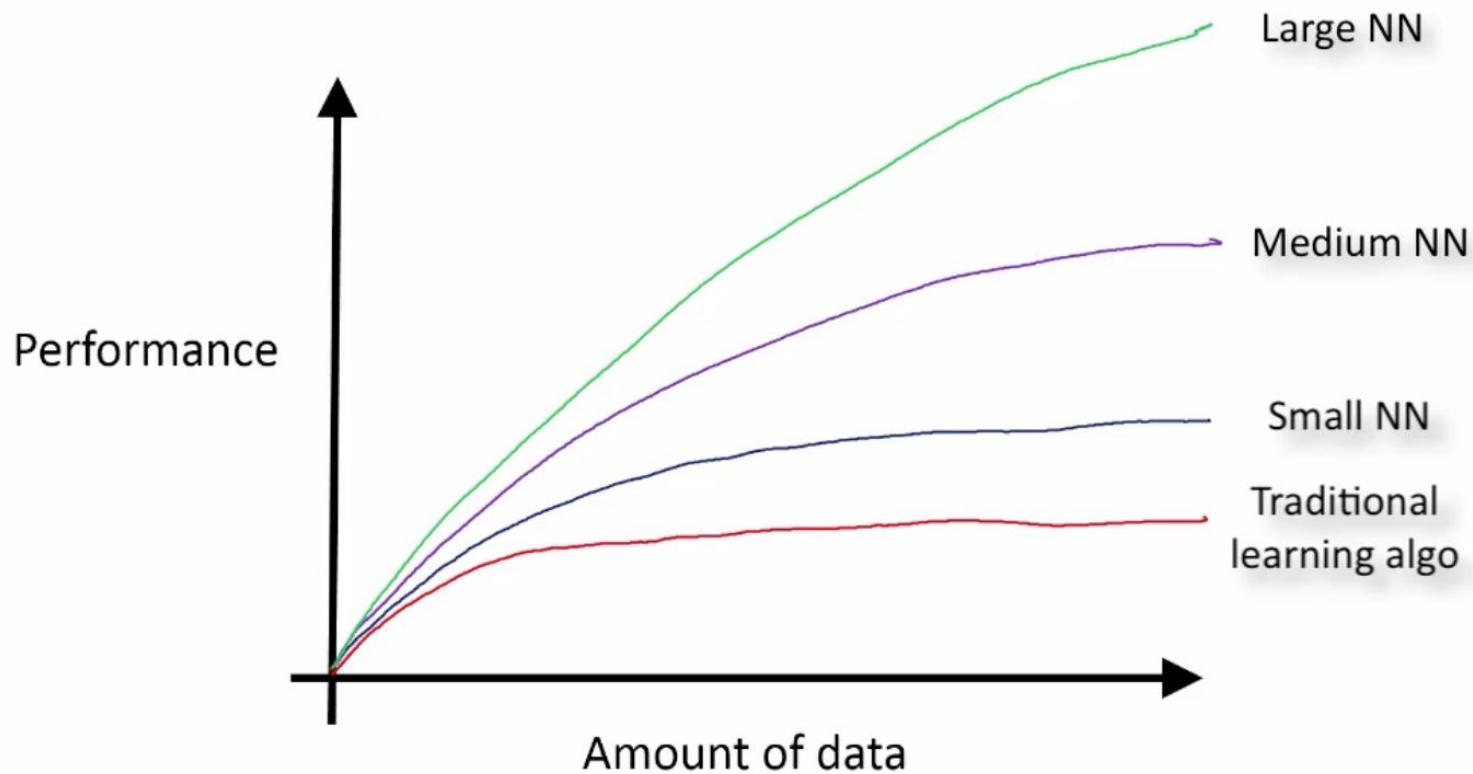
Cloud Vision API and Speech API

TensorFlow and Cloud Machine Learning



Deep Learning and Distributed Training





From: [Andrew Ng](#)



How to train a large NN?
It's tons of **matrix** ops.

a few GPUs >> CPU
But it still takes **days** to train.

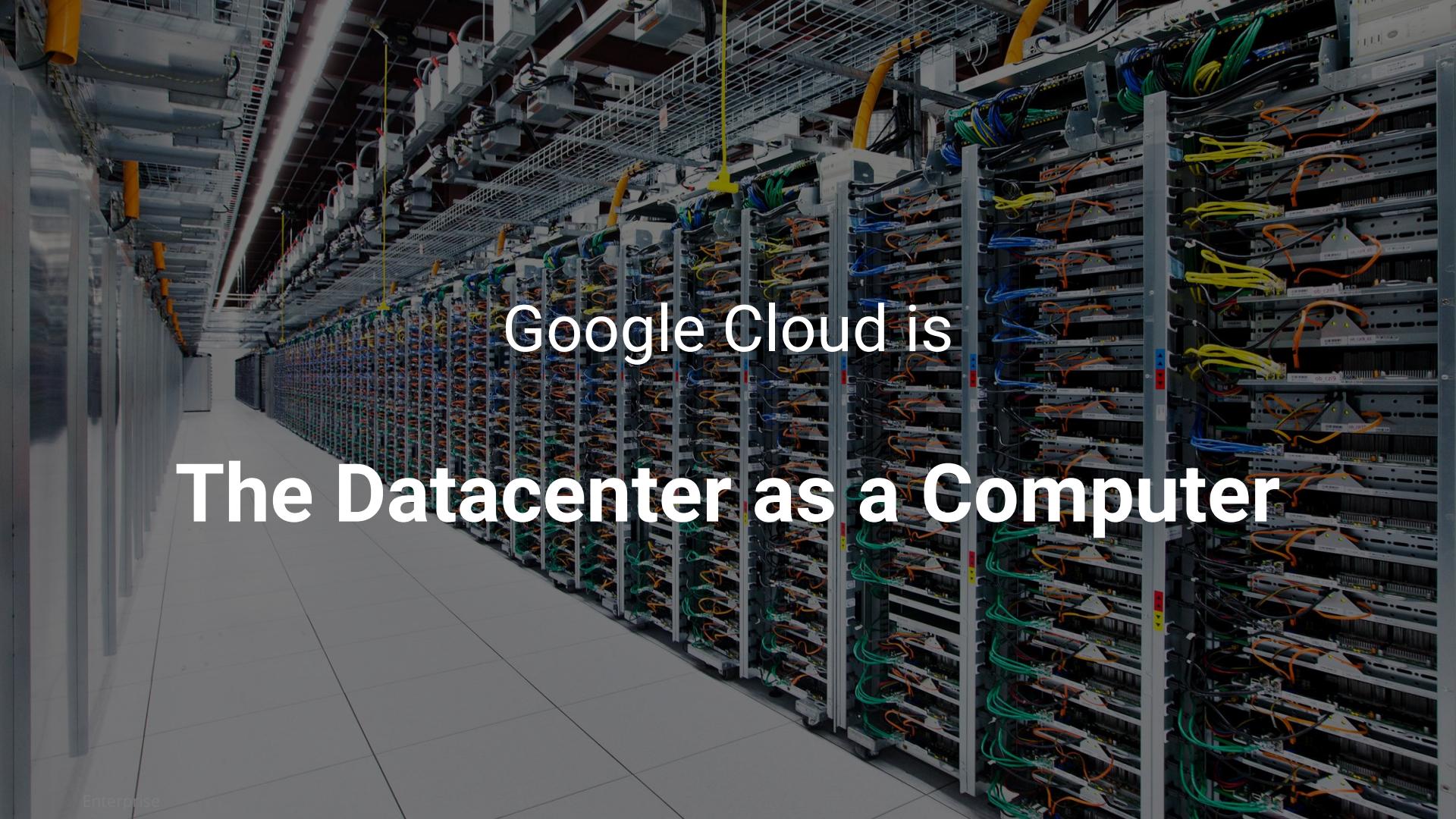
a supercomputer >> a few GPUs
But you don't have a supercomputer.

You need **Distributed Training** on the cloud.



Google Brain: Large scale neural network on Google Cloud





Google Cloud is The Datacenter as a Computer



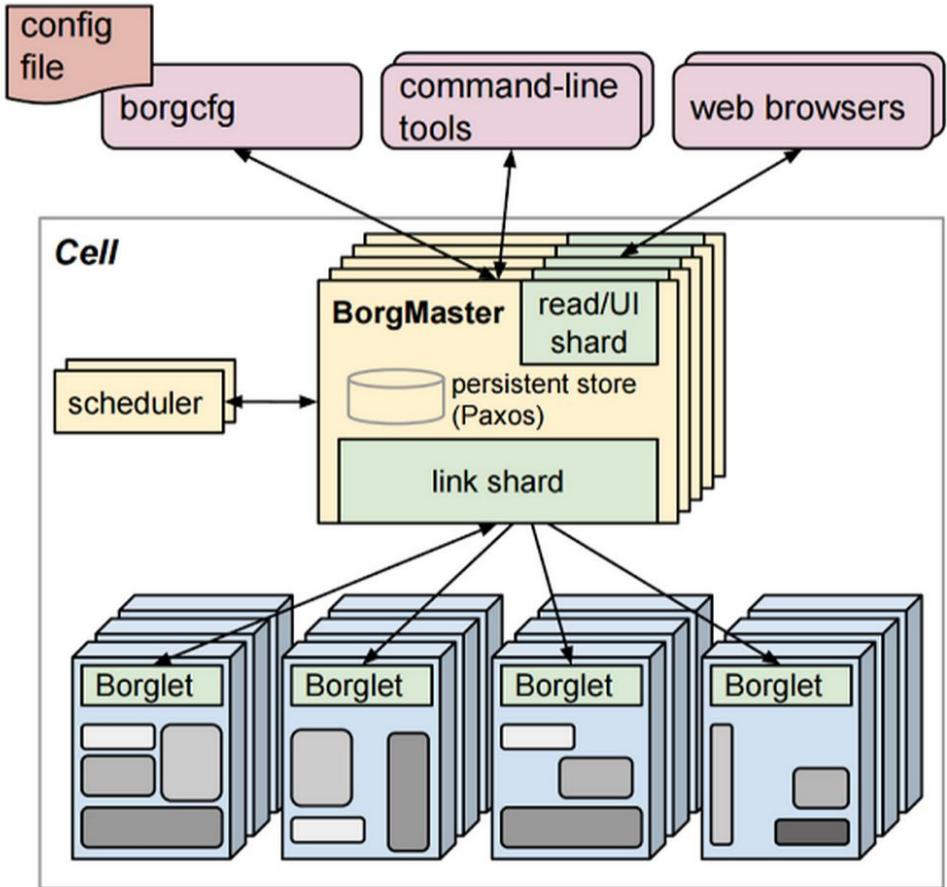
Jupiter network

10 GbE x 100 K = 1 Pbps

Consolidates servers with
microsec latency

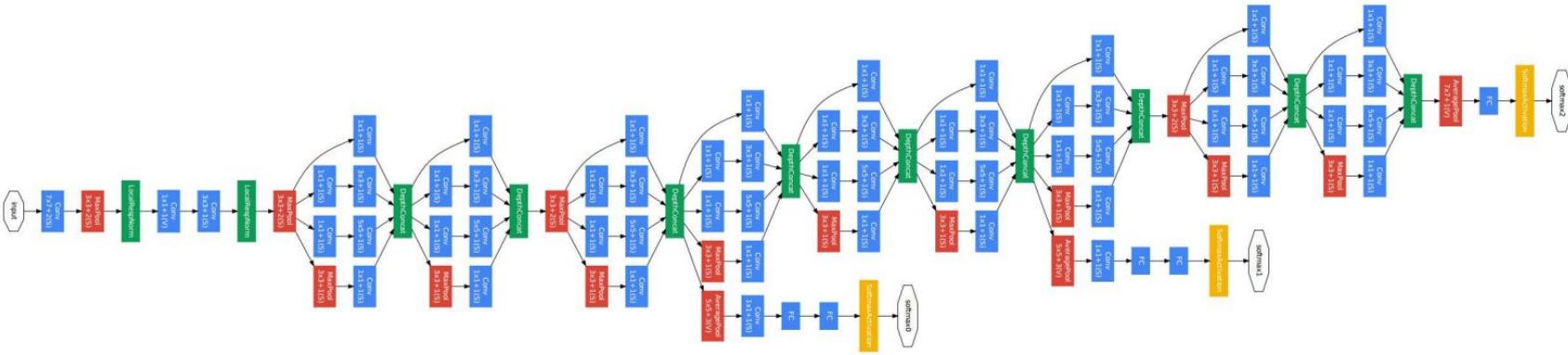
Borg

No VMs, pure containers
10K - 20K nodes per Cell
DC-scale job scheduling
CPUs, mem, disks and IO



Google Cloud +
Neural Network =
Google Brain



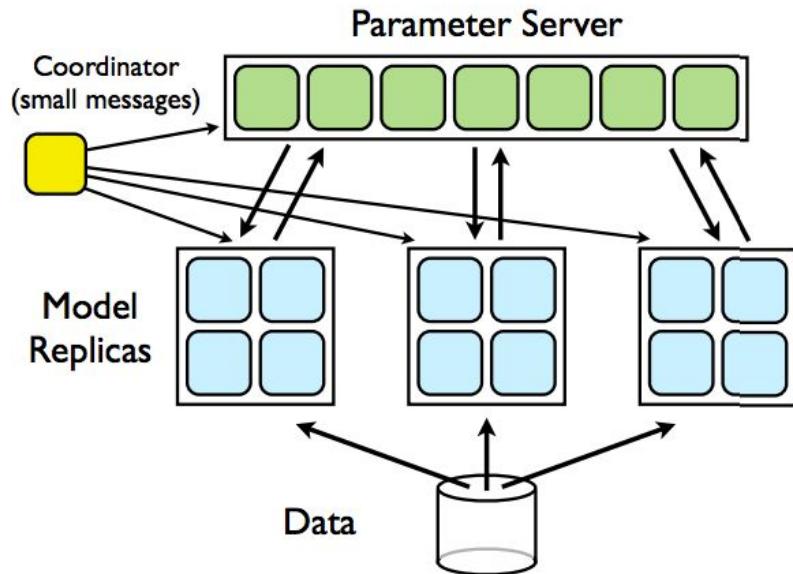


The Inception model (GoogLeNet, 2015)

What's the scalability of Google Brain?

"Large Scale Distributed Systems for Training Neural Networks", NIPS 2015

- Inception / ImageNet: **40x with 50 GPUs**
- RankBrain: **300x with 500 nodes**



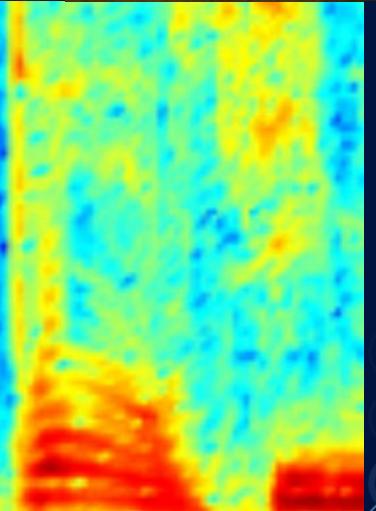
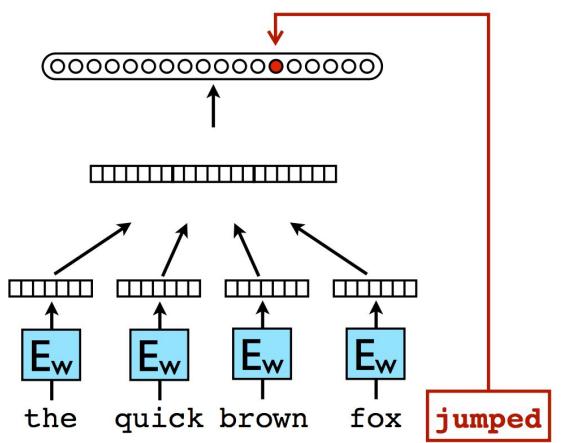
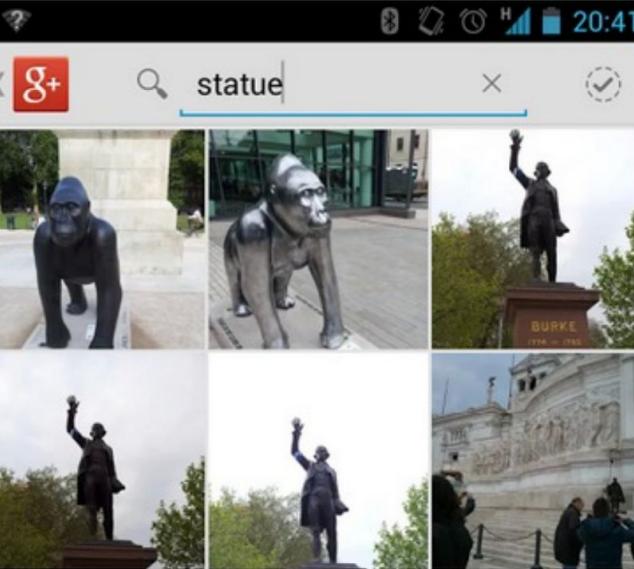
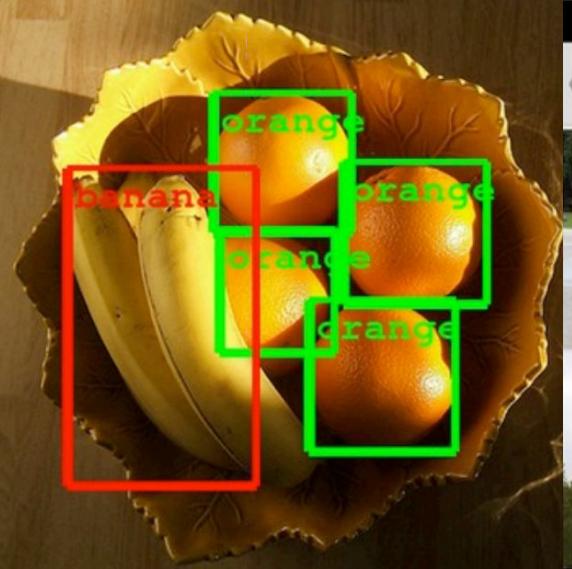
Data Parallelism

= split data, share model

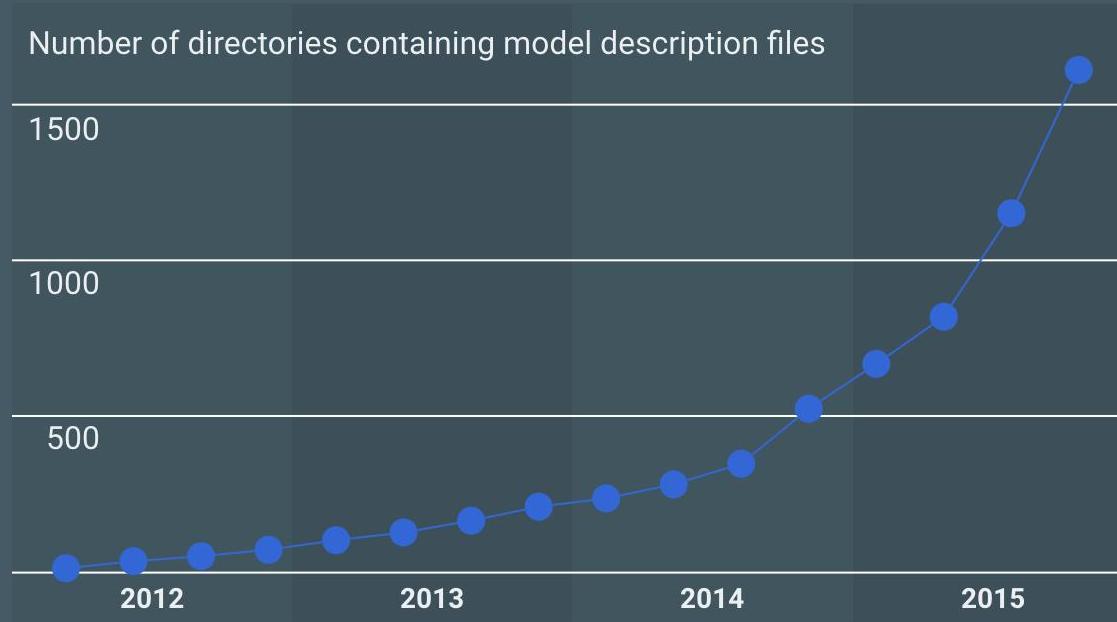
(but ordinary network is
1,000x slower than GPU and
doesn't scale)



Large-scale neural network
for everyone



Growing use of deep learning at Google



Across many areas

- AlphaGo
- Apps
- Maps
- Photos
- Gmail
- Speech
- Android
- YouTube
- Translation
- Robotics Research
- Image Understanding
- Natural Language Understanding
- Drug Discovery



Cloud Vision API

Pre-trained models. No ML skill required

REST API: receives images and returns a JSON

\$2.5 or \$5 / 1,000 units (free** to try)**

Public Beta - cloud.google.com/vision





Label
Detection



Demo

Cloud Speech API

Pre-trained models. No ML skill required

REST API: receives audio and returns texts

Supports **80+** languages

Streaming or non-streaming

Limited Preview - cloud.google.com/speech



Demo Video

(...what if we want to find
classifications the Vision API
doesn't know about?)



TensorFlow and Google Cloud Machine Learning

The Machine Learning Spectrum



A brief look at TensorFlow

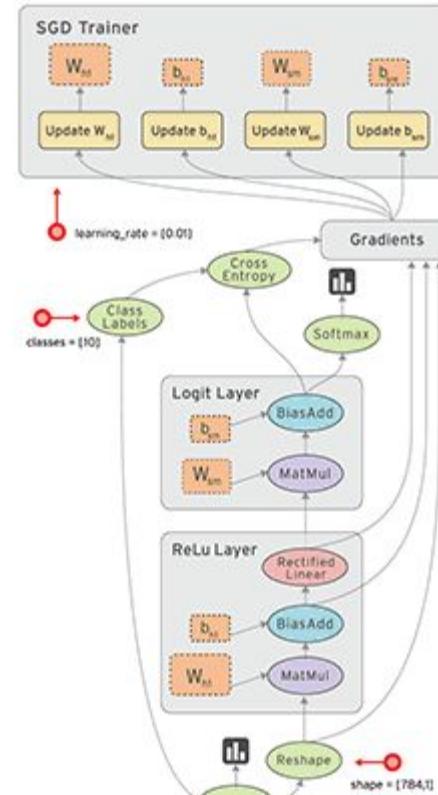
<http://tensorflow.org>

OSS library for numerical computation using data flow graphs.

Operates over **tensors**: *n-dimensional arrays*

Often used for deep learning, but more broadly applicable

- Train on CPUs, GPUs
- Run wherever you like (local, cloud, mobile)



The TensorFlow Inception-v3 model

- Inception-v3 is trained for the ImageNet Large Visual Recognition Challenge using the data from 2012.
- Knows 1000 image classes
- <https://www.tensorflow.org> → Image Recognition Tutorial



The Machine Learning Spectrum





Google Cloud Machine Learning

(limited preview)

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



What Cloud Machine Learning Can Do

- Fully managed service and **SDK**
- Train custom **TensorFlow** graphs at scale
- Regression and classification
- Batch and online predictions at scale
- Integrated **Datalab** environment

<http://cloud.google.com/ml>



Use your own data to train models



Use your own data to train models





How can we learn new image classes?

Demo: Can I Hug That?





<https://flic.kr/p/BTSBc1>

Your recommendation

Huggable. With a score of **0.990581512451**, we're pretty sure.

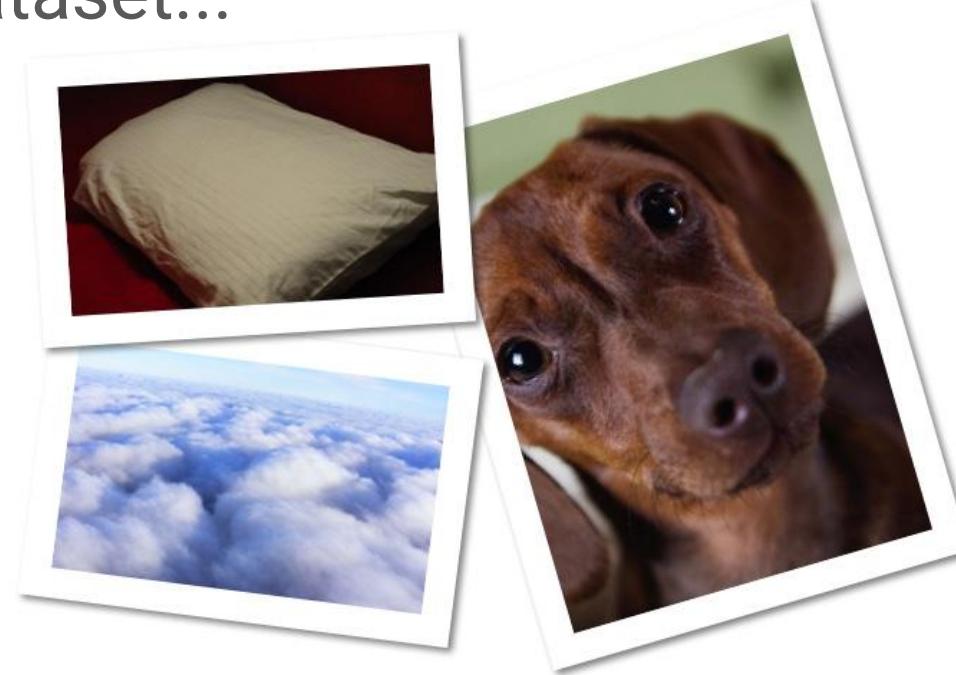


<https://flic.kr/p/BTSBc1>



Behind the Scenes: Training the Classifier (Transfer Learning)

Image Dataset...



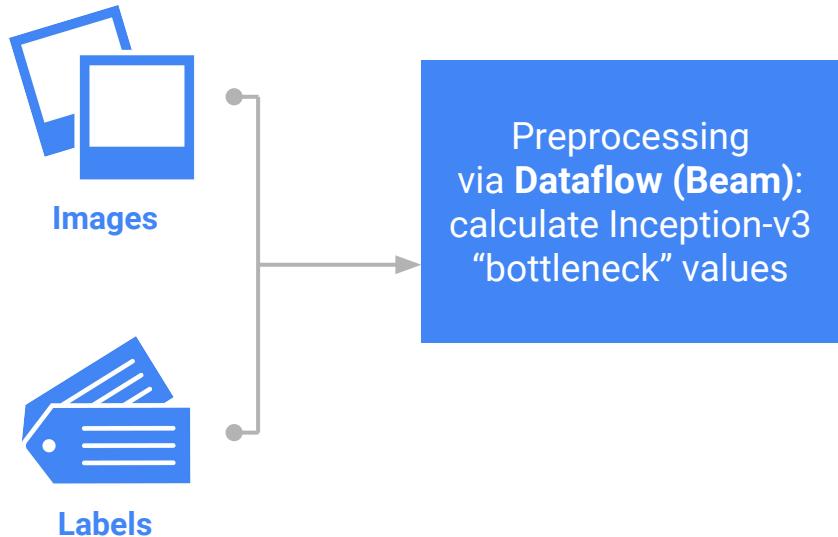
<https://flic.kr/p/c3g4pj>, <https://flic.kr/p/71HVKs>, <https://flic.kr/p/bQhDfg>

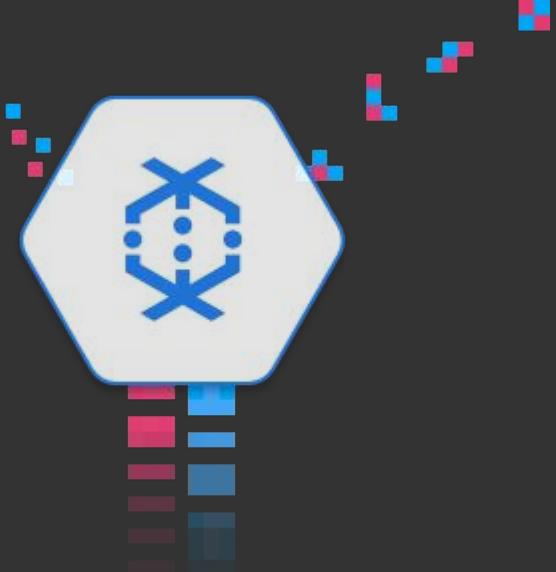
Image Dataset...



<https://flic.kr/p/pJYUZM>, <https://flic.kr/p/4xzkwu>, <https://flic.kr/p/gB7747>

Learning new image classifications

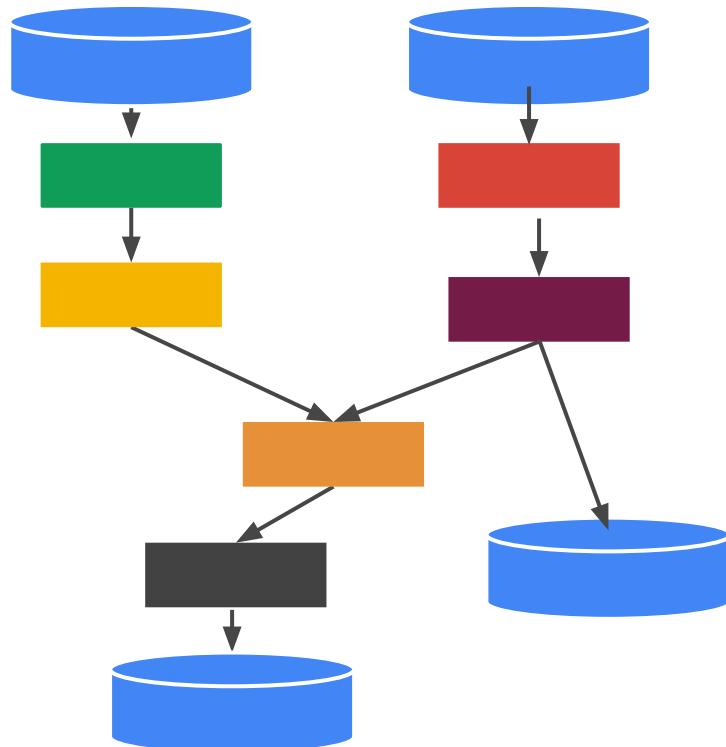




Cloud Dataflow (Apache Beam)

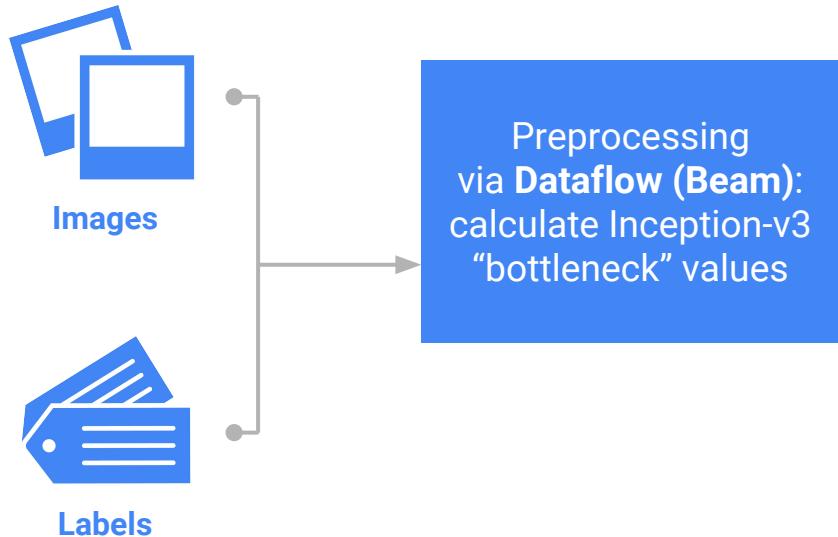
<http://cloud.google.com/dataflow>

What is Apache Beam/Cloud Dataflow?

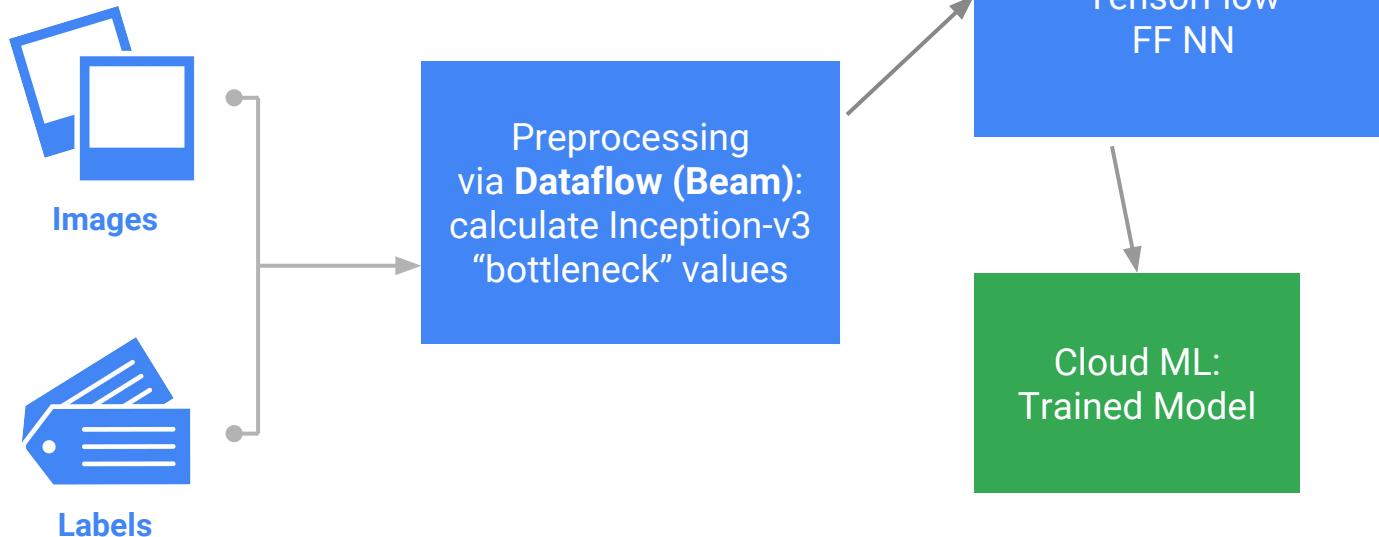


- Apache Beam is a **unified model** for building data processing pipelines that handle **bounded** and **unbounded** data
- Apache Beam is a **collection of SDKs** for building parallelized data processing pipelines
- Google Cloud Dataflow is a **managed service** for executing parallelized data processing **pipelines** written using Apache Beam

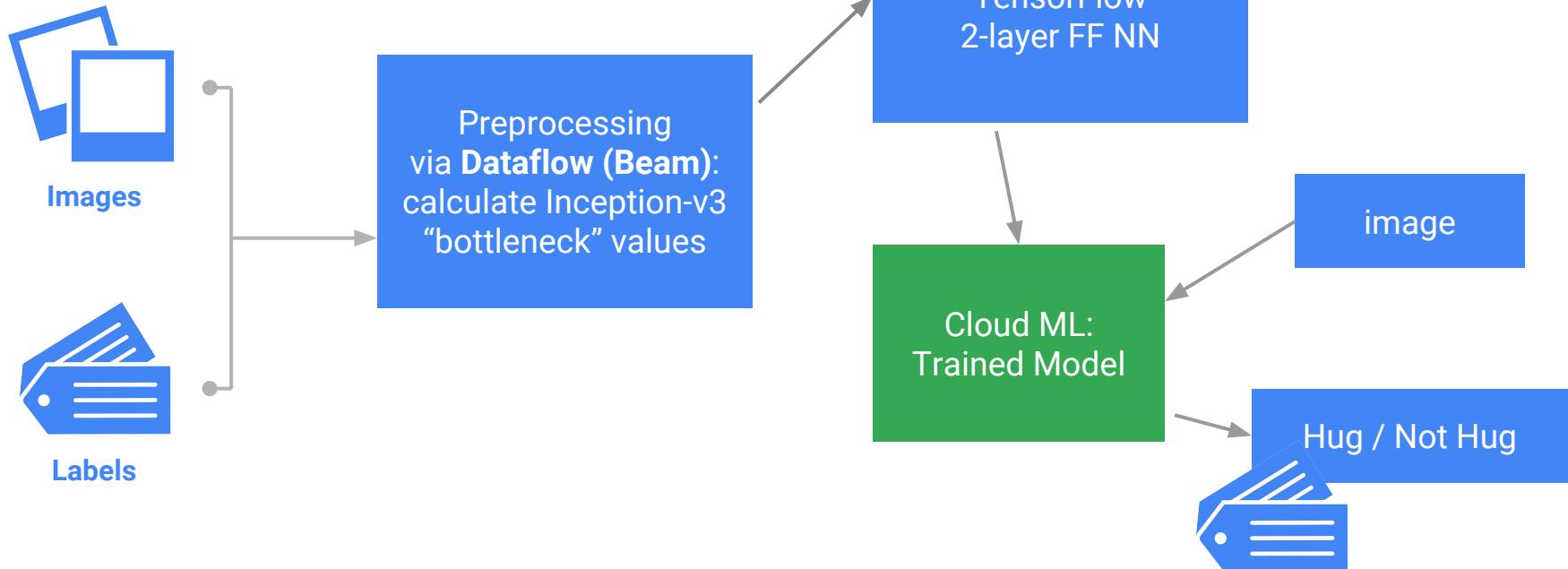
Learning new image classifications



Learning new classifications



Learning new classifications



Cloud Datalab makes it easy to
run preprocessing, training, and
prediction tasks

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





(Brief) Demo

Notebook ▾

[+ Add Code](#) [+ Add Markdown](#) [Delete](#) [▲ Move Up](#) [▼ Move Down](#) [▶ Run ▾](#) [Clear ▾](#) [Reset Session](#)

Navigation

Run data pre-processing

```
INPUT_DIR='gs://cloud-ml-data/img/flower_photos'  
OUTPUT_DIR='gs://[REDACTED]/img/flower_photos'  
CLOUD_PROJECT='[REDACTED]'  
NUMBER_OF_WORKERS=5
```

```
%%bash -s "$INPUT_DIR" "$OUTPUT_DIR" "$CLOUD_PROJECT" "$NUMBER_OF_WORKERS"  
../preprocess.sh $1 $2 $3 $4
```

```
Found 5 labels.  
dandelion  
tulips  
sunflowers  
roses  
daisy
```

```
+ prepare_tmp  
+ rm -Rf /tmp/image_classification  
+ mkdir -p /tmp/image_classification  
+ set +e  
+ prepare_csv_data gs://cloud-ml-data/img/flower_photos gs://cm  
l/img/flower_photos
```



Notebook Outline

- [Building custom image models](#)
- [Install pre-requisites](#)
- [Run data pre-processing](#)
- [Load dictionary](#)
- [Train the model](#)
 - [Train locally](#)
 - [Train on the cloud](#)
- [Do online predictions](#)
 - [Copy data for making the predictions](#)
 - [Make the predictions](#)
 - [Visualize prediction](#)
 - [Do predictions using python](#)
- [Do batch predictions](#)
 - [Copy files and display results](#)

[Notebook](#) ▾[+ Add Code](#) [+ Add Markdown](#) [Delete](#) [▲ Move Up](#) [▼ Move Down](#) [▶ Run](#) ▾ [Clear](#) ▾ [O Reset Session](#)[Navigation](#)

Make the predictions

```
import preprocess
import re
from google.protobuf import json_format
preprocessor = preprocess.LocalImageProcessor(image_graph_uri='./classify_image'

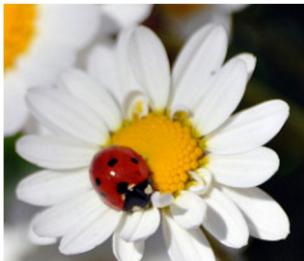
with preprocess._open('./img.jpg') as f:
    example = preprocessor.process([f.read()], ['./img.jpg'])[0]
    json_example = re.sub(r'\s+', ' ', json_format.MessageToJson(example))

%%ml predict -m image.v1 --cloud -p -d json_example
```

	label	score
0	4	0.972606

Visualize prediction

```
from IPython.core.display import HTML
HTML('<div style="padding: 16px"><span>%s (%.2f)</span></div>' :
```



Notebook Outline

[Building custom image models](#)[Install pre-prerequisites](#)[Run data pre-processing](#)[Load dictionary](#)[Train the model](#)[Train locally](#)[Train on the cloud](#)[Do online predictions](#)[Copy data for making the predictions](#)[Make the predictions](#)[Visualize prediction](#)[Do predictions using python](#)[Do batch predictions](#)[Copy files and display results](#)

Ready to use Machine Learning models



Cloud Vision API

Beta



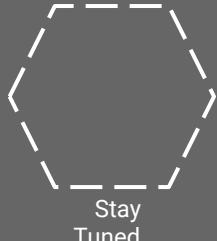
Cloud Speech API

Alpha



Cloud Translate API

GA



Use your own data to train models



Cloud Dataflow
(Apache Beam)



Cloud Machine Learning

Alpha



Cloud Storage



Google BigQuery

GA

GA



Cloud Datalab

Beta



Develop - Model - Test

NEW

Links & Resources

Google Cloud Machine Learning: <http://bit.ly/gcp-ml> , <https://cloud.google.com/ml/>

Google Cloud Speech API: <http://bit.ly/gcp-speech>

Google Cloud Vision API: <http://bit.ly/gcp-vision>

Dataflow, Apache Beam: <https://cloud.google.com/dataflow/> , <http://incubator.apache.org/projects/beam.html>

TensorFlow: <http://tensorflow.org> ,
https://www.tensorflow.org/versions/master/how_tos/image_retraining/index.html

'TensorFlow for Poets', by Pete Warden: <http://bit.ly/tensorflow-for-poets>

Images we used today: <http://bit.ly/hug-images>

[Large Scale Distributed Systems for Training Neural Networks](#), Jeff Dean and Oriol Vinals



Thank you!

Questions?