Google Cloud

Cloud Computing for Higher Education

April 23, 2017

Introductions

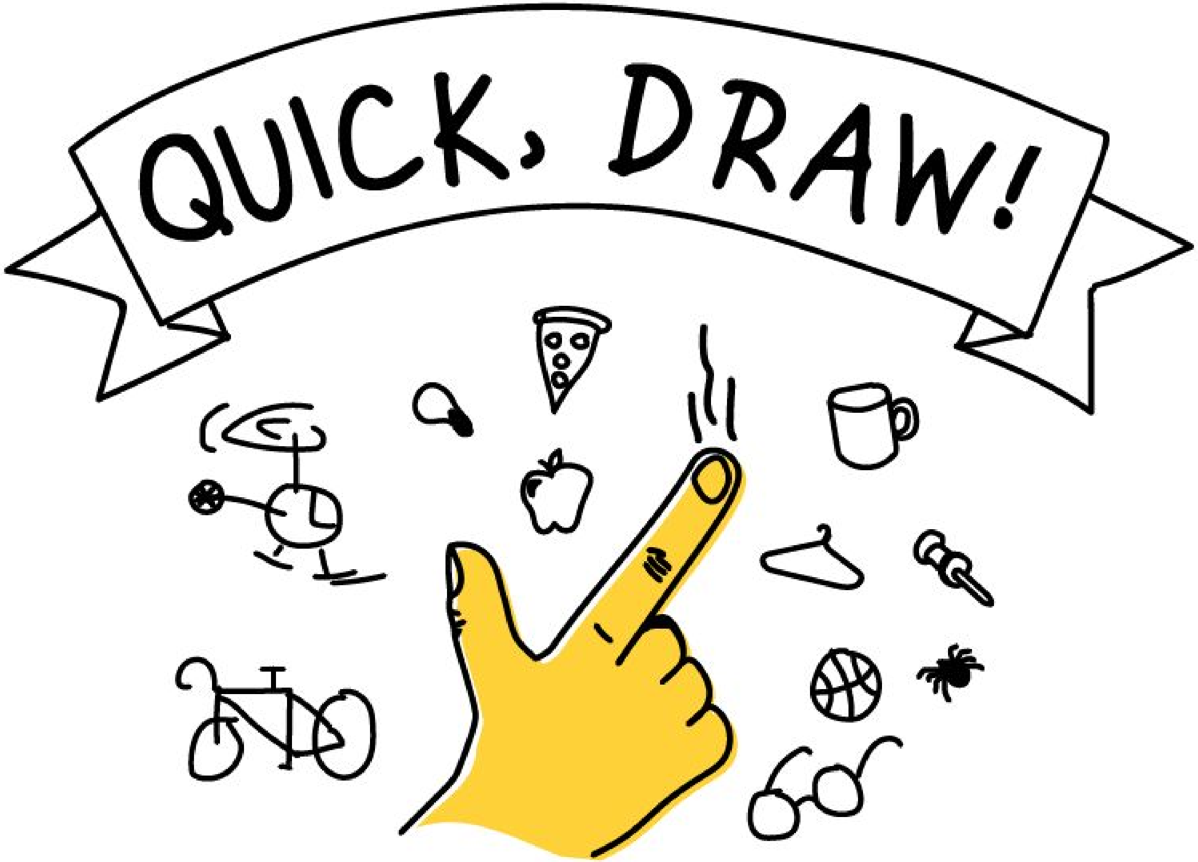
Shel Waggener
Internet2
Senior VP

Bram Bout
Google for Education
Director

Edward Doan
Google for Education
Customer Engineer

Steven Butschi
Google for Education
Higher Ed Manager

Karan Bhatia, PhD
Google Cloud,
Scientific Computing



quickdraw.withgoogle.com

So what?!

Think. Pair. Share.

- What are you hoping to learn in this session?
- What is one question you want to have answered by the time you leave?

Agenda

- Google Cloud Platform for universities
- 4 Ways GCP can help IT support researchers
- Bursting into GCP
    - Video chat with Andrew Sutherland, MIT Mathematics
- Interactive Qwiklabs & Roundtable Discussions
- Wrap Up

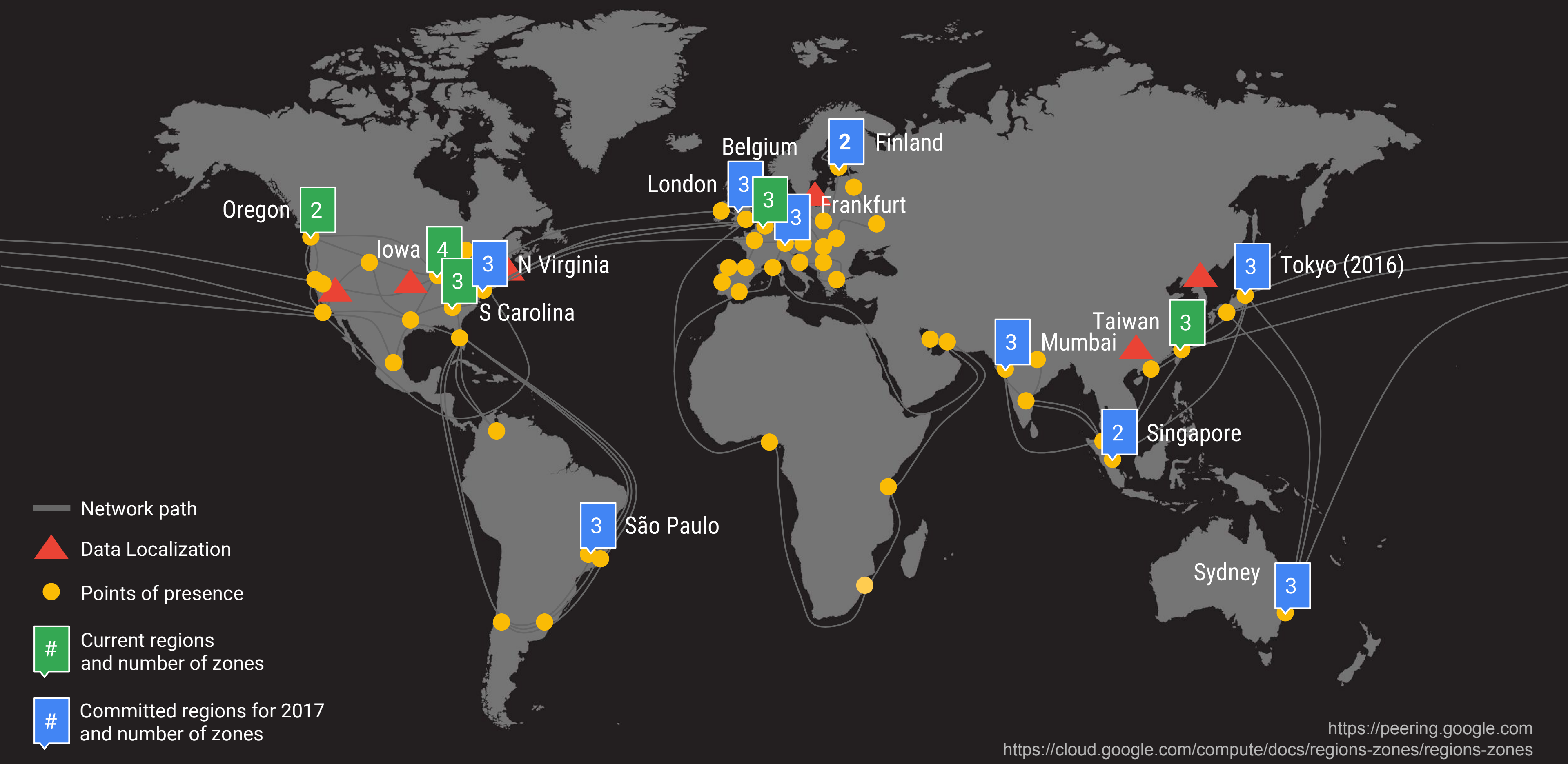
Investing to meet university needs

1 Billion

End users served by GCP customers

$29.4 Billion

Google's trailing 3 Year CAPEX investment



Oregon
2
Iowa
4
3
N Virginia
3
S Carolina
London
3
Belgium
3
3
Frankfurt
2
Finland
3
Tokyo (2016)
Taiwan
3
3
Mumbai
2
Singapore
3
São Paulo
Sydney
3
Network path
Data Localization
Points of presence
#
Current regions
and number of zones
#
Committed regions for 2017
and number of zones
https://peering.google.com
https://cloud.google.com/compute/docs/regions-zones/regions-zones

INTERNET 2

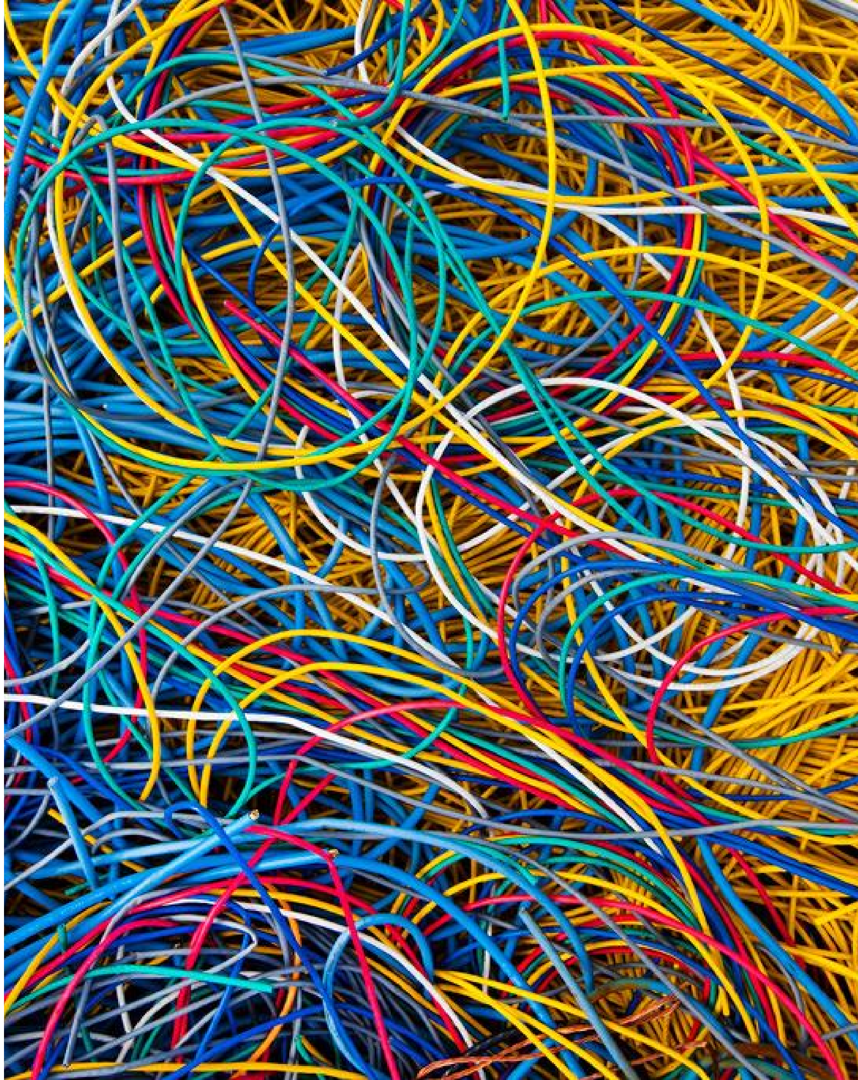
Google Cloud Platform

Education Egress Waiver Program for Internet2 Members

INTERNET2®

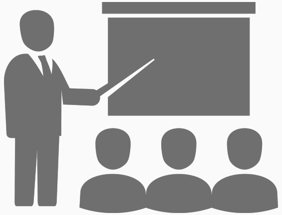
Fill out this form to receive updates:

goo.gl/NE7YzL

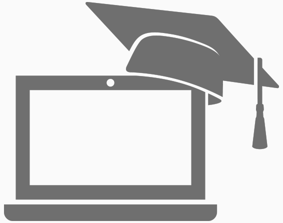


Three ways cloud computing can help universities transform

Google Cloud



Teaching



Administration

Research


Help students
build what’s next
Google
google.com/datacenters

Teaching

Faculty in select countries

+

Teaching university courses

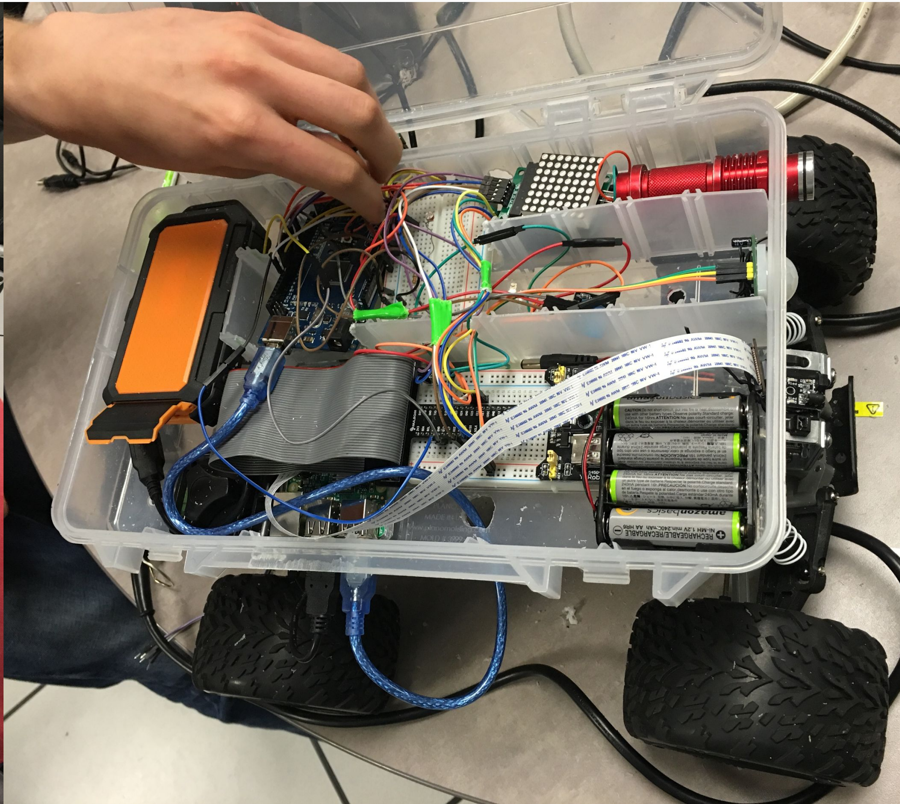
+

In computer science or related fields

Eligible countries

- Austria
- Belgium
- Bulgaria
- Canada
- Czech Republic
- Cyprus
- Denmark
- Estonia
- Finland
- France
- Germany
- Greece
- Hungary
- Ireland
- Israel
- Italy
- Latvia
- Lithuania
- Luxembourg
- Malta
- Netherlands
- Poland
- Portugal
- Romania
- Slovakia
- Spain
- Sweden
- Switzerland
- United Kingdom
- United States

California State University, San Bernardino



Be more efficient with a cloud you can trust

Administration

Efficient IT

Affordable, powerful, dependable infrastructure

- Storage
- Disaster Recovery
- High Performance Computing

The next frontier

Create insights from smart, Big Data

- Recruitment ROI
- Student success
- Alumni donations

Administration

Efficient IT

- Pre-minute billing
- Orbitera
- Customizable VMs
- Sustainable use discount
- Future discounts

The next frontier

- BigQuery
- TensorFlow
- Structured & unstructured data
- Machine Learning Models & APIs


The power to
change our future

Research

Do more with less

Pricing that allows you to do more on the Cloud

Innovative Tools

Machine learning and advanced APIs

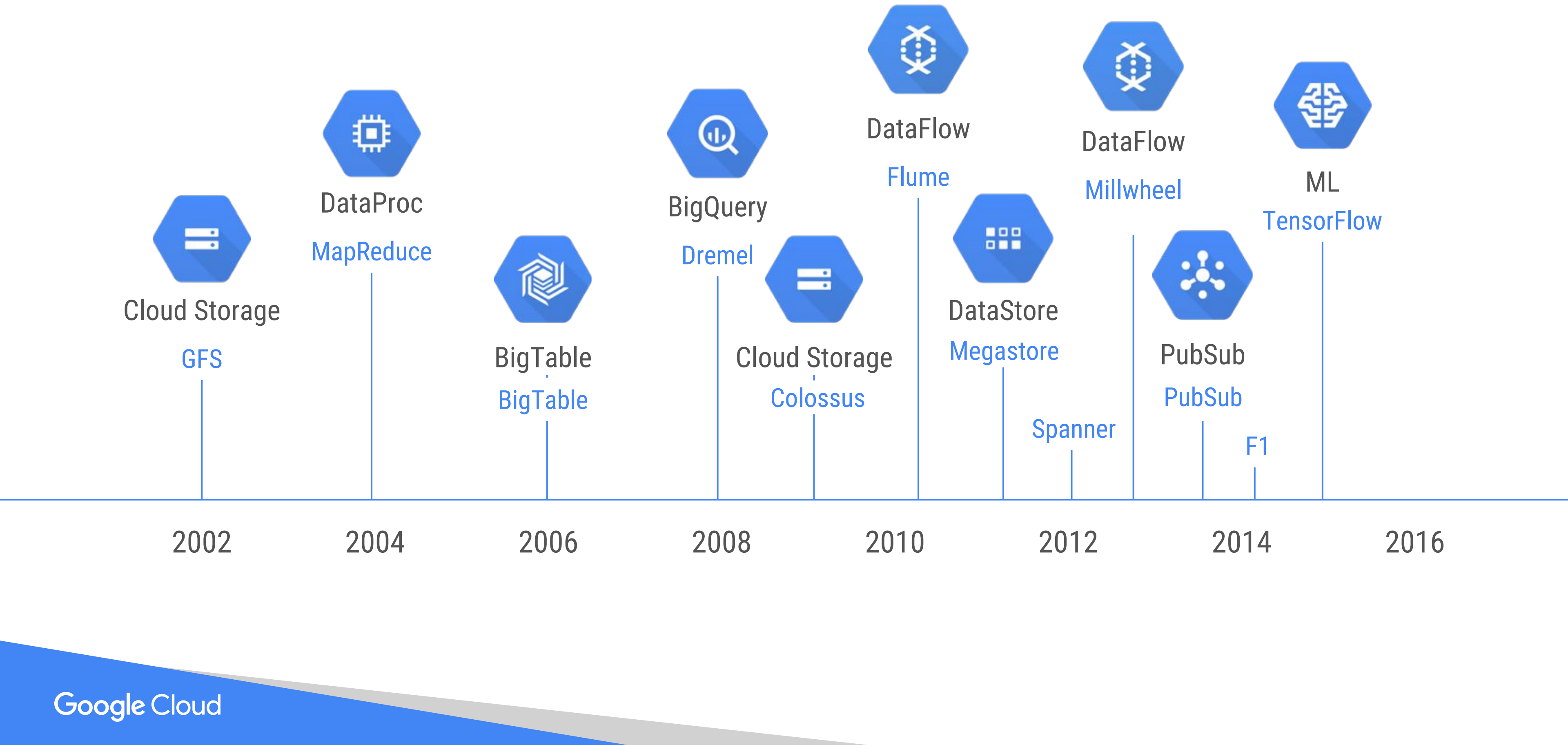
Start where you are

Hybrid cloud opportunities

Data & Analytics

Application Development

Infrastructure & Operations


Cloud Storage
GFS
DataProc
MapReduce
BigTable
BigTable
BigQuery
Dremel
Cloud Storage
Colossus
DataFlow
Flume
DataStore
Megastore
Spanner
DataFlow
Millwheel
PubSub
PubSub
F1
ML
TensorFlow
2002
2004
2006
2008
2010
2012
2014
2016
Google Cloud

Leading open source communities

Kubernetes

#1 Highest Engagement on Github

Tensorflow

#2 Highest Engagement on Github

Source: Analyzing GitHub issues and comments with BigQuery

Funding Agency Partnerships

- National Science Foundation
  - BIGDATA
- National Institute of Health

Google Cloud

4 ways GCP can help you help researchers

1.

Machine Learning
- Vision API
- Tensorflow

2.

Simplify billing with Orbitera

3.

Analyze lots of data quickly with BigQuery

4.

Access & add to public data sets

Machine Learning

Google Cloud brings proven tech from Google products

Search
Search ranking
Speech recognition

Android
Keyboard & speech input

Play
App recommendations
Game developer experience

Gmail
Smart Reply
Spam classification

Drive
Intelligence in Apps

Chrome
Search by Image

Photos
Photos search

YouTube
Video recommendations
Better thumbnails

Maps
Street View image
Parsing Local Search

Translate
text, graphic, and
speech translations

Cardboard
Smart stitching

Ads
Richer Text Ads
Automated Bidding

Two flavors of machine learning

API

Vision | Translation | Speech BETA | Natural Language

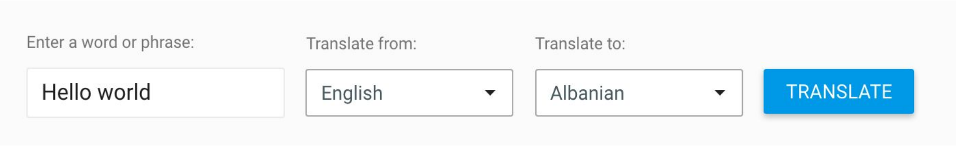
Pre-Trained Models

Storage | BigQuery | Datalab | Tensor Flow | Pipelines | Model Management

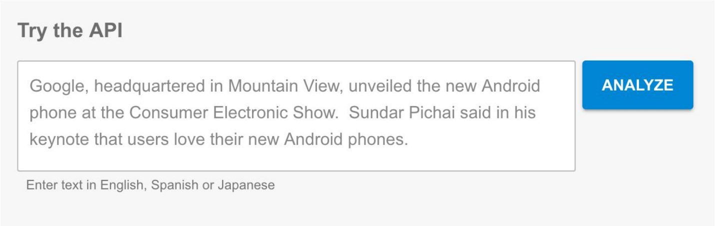
Build Your Own Model

Google Cloud Machine Learning Services

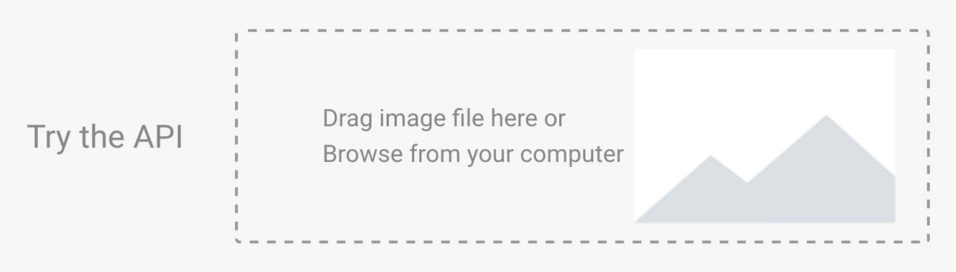
cloud.google.com/translate/



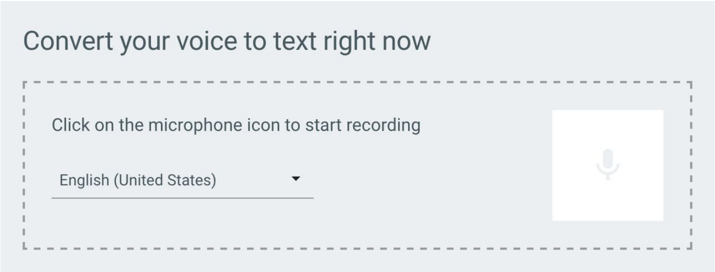
cloud.google.com/natural-language/

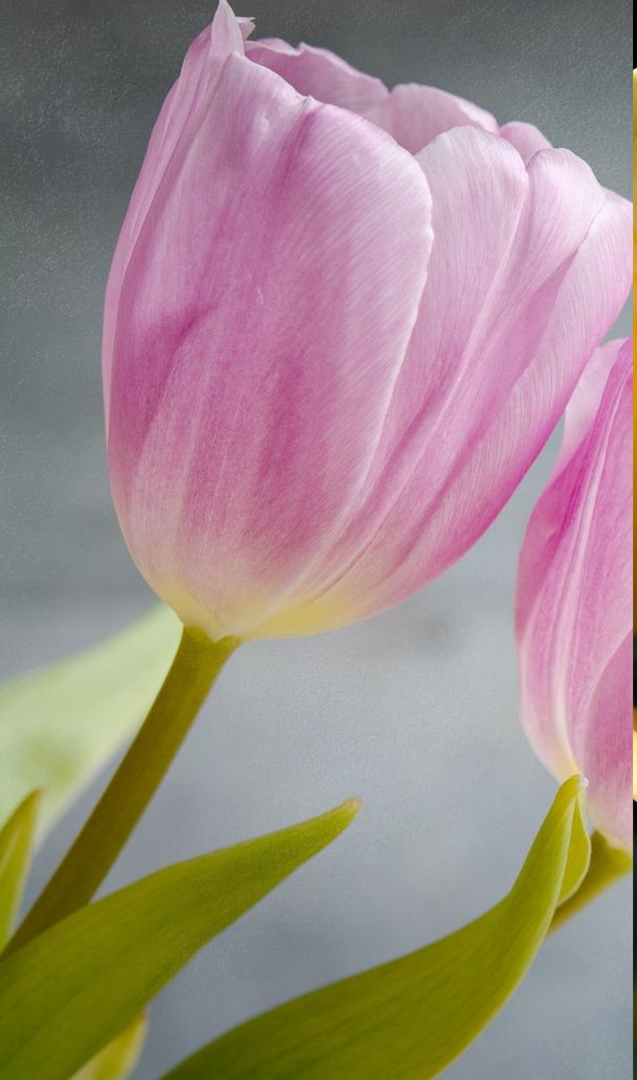


cloud.google.com/vision/



cloud.google.com/speech/





Two flavors of machine learning

API

Vision

Translation

Speech BETA

Natural Language

Pre-Trained Models

Storage

BigQuery

Datalab

Tensor Flow

Pipelines

Model Management

Build Your Own Model

Cloud ML Engine

- **PaaS** for Tensorflow
- **Scale** your training up to 100 workers
- Automatic **monitoring** and **logging**
- Easy transition from training to **prediction**
- Built in model **version management**
- **No lock-in.** Option to download your trained models for on-premise or mobile deployment

Hardware Accelerated

- Available Today: **NVIDIA K80 GPU**
- Coming Soon: **Tensor Processing Unit (TPU)**
- Custom ASIC built and optimized for TensorFlow
- Used in production at Google for over 16 months
- 7 years ahead of GPU performance per watt

CloudML is part of a bigger picture

Capture
Pub/Sub

Store
Cloud Storage

BigQuery

Cloud SQL

Datastore

BigTable

Process
Dataflow
Dataproc

Analyze
BigQuery

Dataflow

Datalab

Insight
Cloud ML Engine

DataLab

A better developer experience

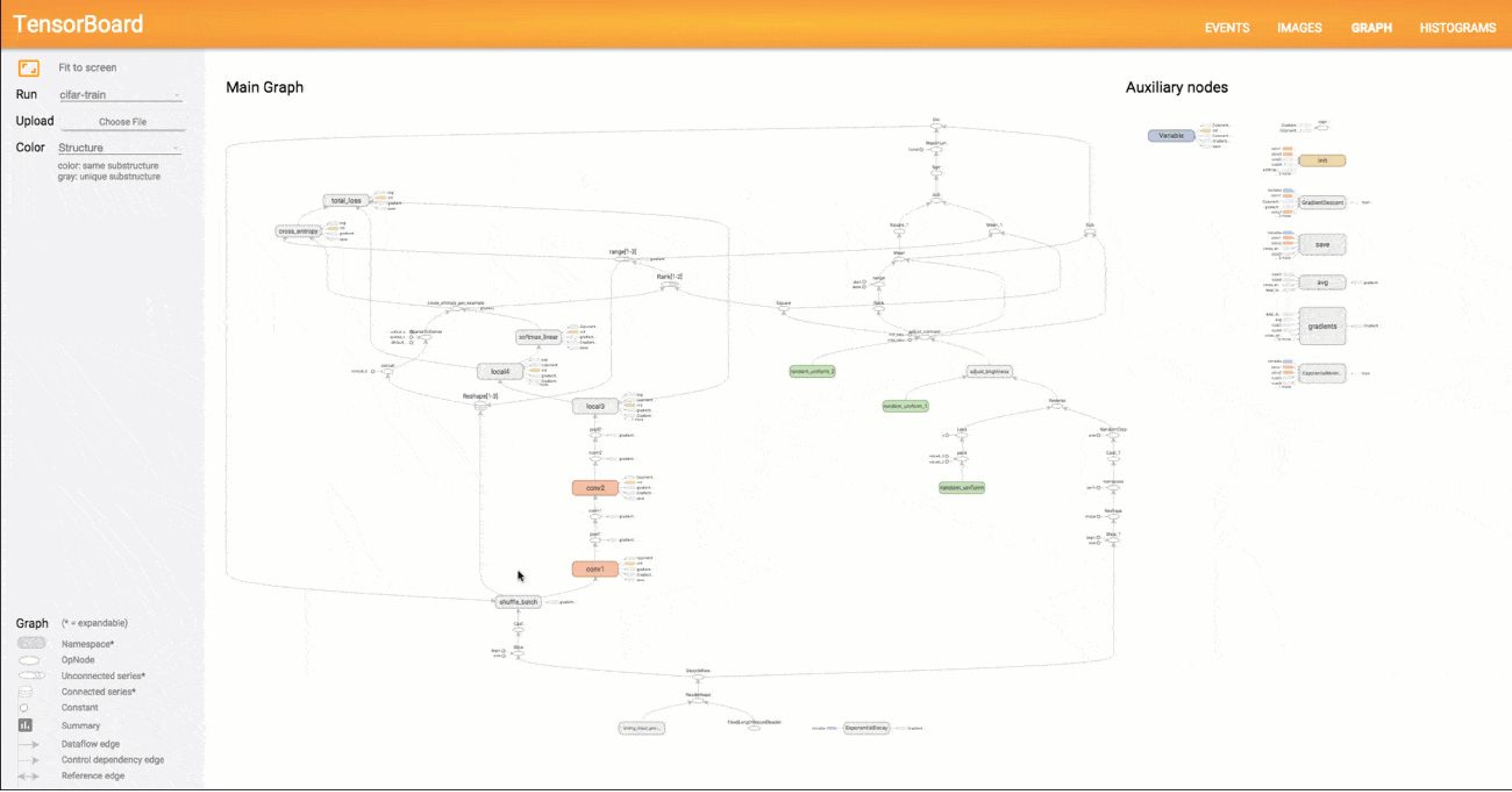
TensorFlow

- World's most popular ML framework
- Developer friendly yet performance optimized
- **Powers over 100 Google services**
- Managed infrastructure with Cloud ML
- Tutorials at https://www.tensorflow.org

Linear Regression VS Neural Network

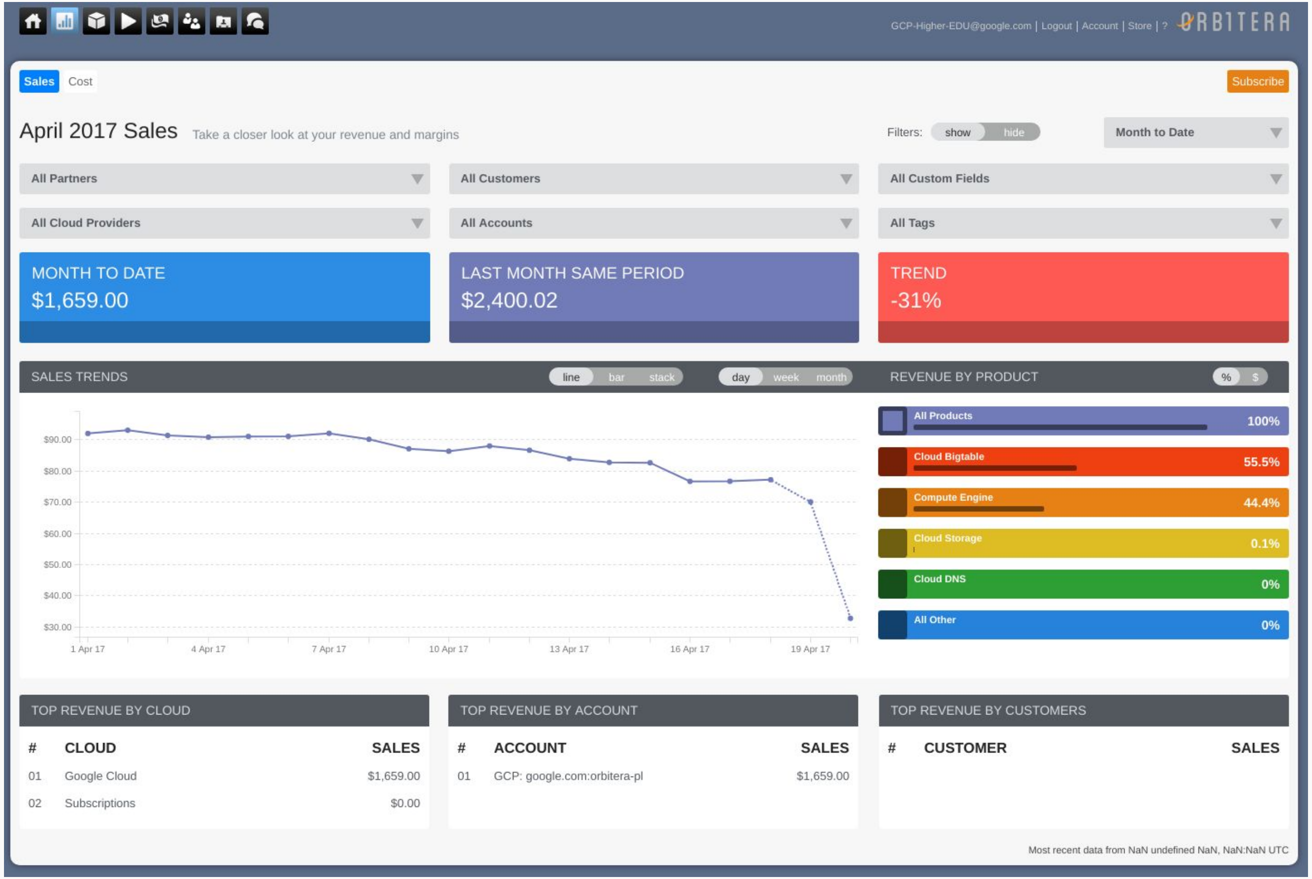
```
import tensorflow as tf

#Define input feature columns
sq_footage = tf.contrib.layers.real_valued_column("sq_footage")
feature_columns = [sq_footage]

#Define input function
def input_fn(feature_data,label_data=None):
  return {"sq_footage":feature_data}, label_data

#Instantiate Linear Regression Model
estimator = tf.contrib.learn.LinearRegressor(
  feature_columns=feature_columns,
  optimizer=tf.train.FtrlOptimizer(learning_rate=100))

#Train
estimator.fit(
  input_fn=lambda:input_fn(tf.constant([1000,2000]),
                           tf.constant([100000,200000])),
  steps=100)

#Predict
estimator.predict(input_fn=lambda: input_fn(tf.constant([3000])))
```

```
import tensorflow as tf

#Define input feature columns
sq_footage = tf.contrib.layers.real_valued_column("sq_footage")
feature_columns = [sq_footage]

#Define input function
def input_fn(feature_data,label_data=None):
  return {"sq_footage":feature_data}, label_data

#Instantiate Neural Network Model
estimator = tf.contrib.learn.DNNRegressor(
  feature_columns=feature_columns, hidden_units=[10,10])


#Train
estimator.fit(
  input_fn=lambda:input_fn(tf.constant([1000,2000]),
                           tf.constant([100000,200000])),
  steps=100)

#Predict
estimator.predict(input_fn=lambda: input_fn(tf.constant([3000])))
```


TensorBoard
EVENTS
IMAGES
GRAPH
HISTOGRAMS
Fit to screen
Run
cifar-train
Upload
Choose File
Color
Structure
color: same substructure
gray: unique substructure
Main Graph
Auxiliary nodes
total_loss
cross_entropy
softmax_linear
local4
local3
conv2
conv1
shuffle_batch
Variable
init
GradientDescent
save
avg
gradients
Graph
(* = expandable)
Namespace*
OpNode
Unconnected series*
Connected series*
Constant
Summary
Dataflow edge
Control dependency edge
Reference edge

Billing with Orbitera

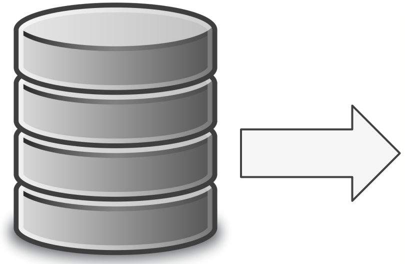
- Manage and govern how users are consuming cloud services that are bought from one or more vendors
- Set up your own price books and generate internal invoices for departments
- **Free** for GCP usage



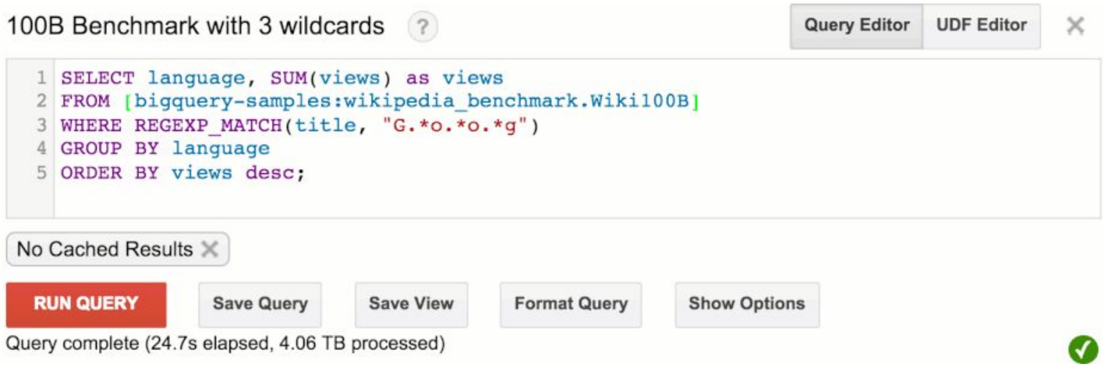
BigQuery: Analyze 100B rows in Wikipedia

- Let's check out the publically-available bigquery-samples:wikipedia_benchmark, specifically the Wiki100B table. This table contains 100 billion rows and is about 7 Terabytes in size.

- BigQuery has to:
    - Read about 1TB of data, then uncompress it to 4TB (assuming ~4:1 compression)
    - Execute 100 billion regular expressions with 3 wildcards each
    - Distribute 1.25TB of data across the network (1TB compressed for initial read, and 0.25TB for the aggregation)



Wikipedia: 100Billion rows, 7 TeraBytes



Source: https://cloud.google.com/blog/big-data/2016/01/anatomy-of-a-bigquery-query

Anatomy of a BigQuery Query (part 2)

To run in 30s:

- Read 4TB of compressed data (1TB raw)
- Execute 100B regular expressions
- Distribute 1.25 TB of data across the network

Means:

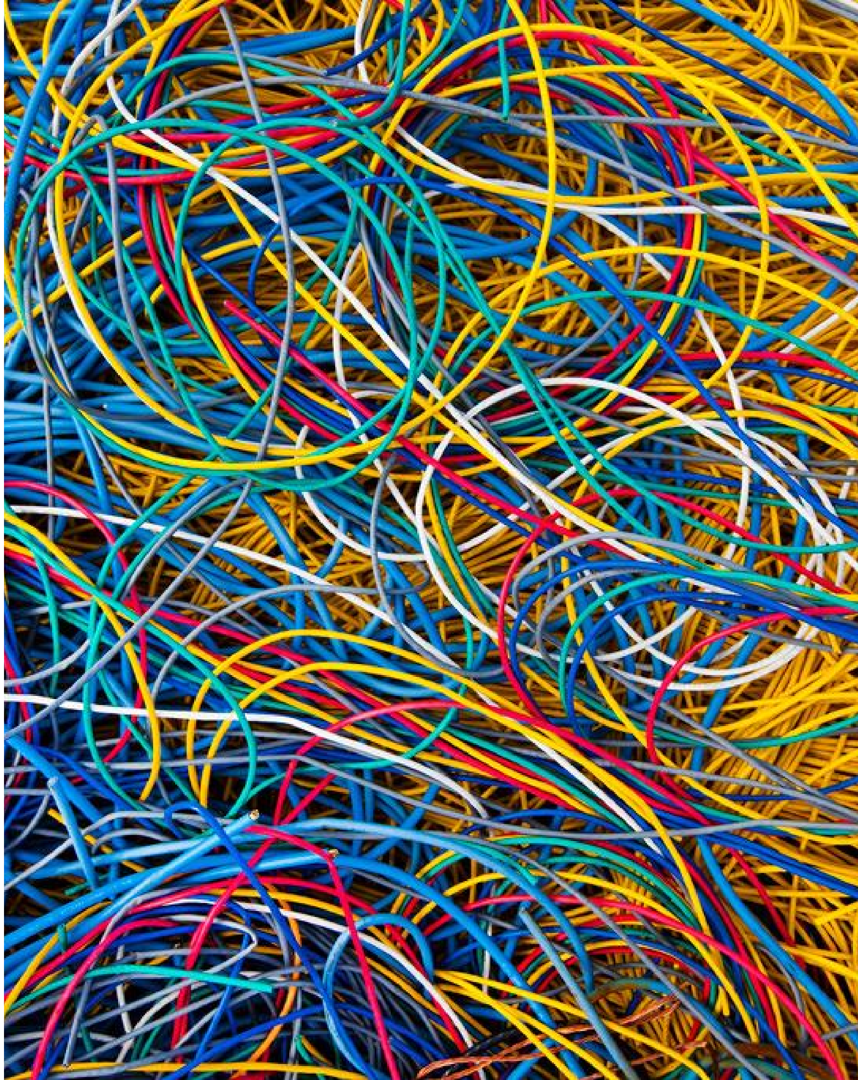
- About 330 100MB/sec dedicated hard-drives to read 1TB of data
- A 330 Gigabit network to shuffle the 1.25 TB of data
- 3,300 cores to uncompress 1TB of data and process 100 billion regular expressions at 1 μsec per

Education Egress Waiver Program for Internet2 Members

INTERNET2®

Fill out this form to receive updates:

goo.gl/NE7YzL



Google Cloud Public Datasets Program

Mission:
Facilitate the onboarding of datasets into Google Cloud products

Google Cloud

Next

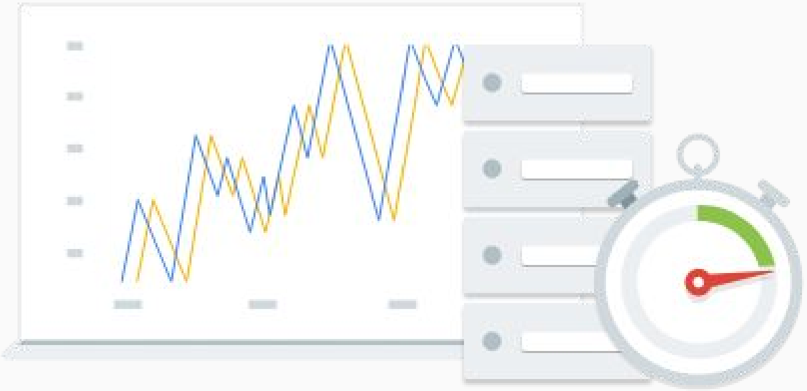
42+ datasets

NOAA
Bay Area BikeShare
NASA
NYC OpenData
citi bike
esa
DATA
GDELT
looker
HHS
Federal Election Commission United States of America
stackoverflow
GitHub
Hacker News
NYC Taxi & Limousine Commission
IRS

You can contribute too!

Visit: https://cloud.google.com/public-datasets/

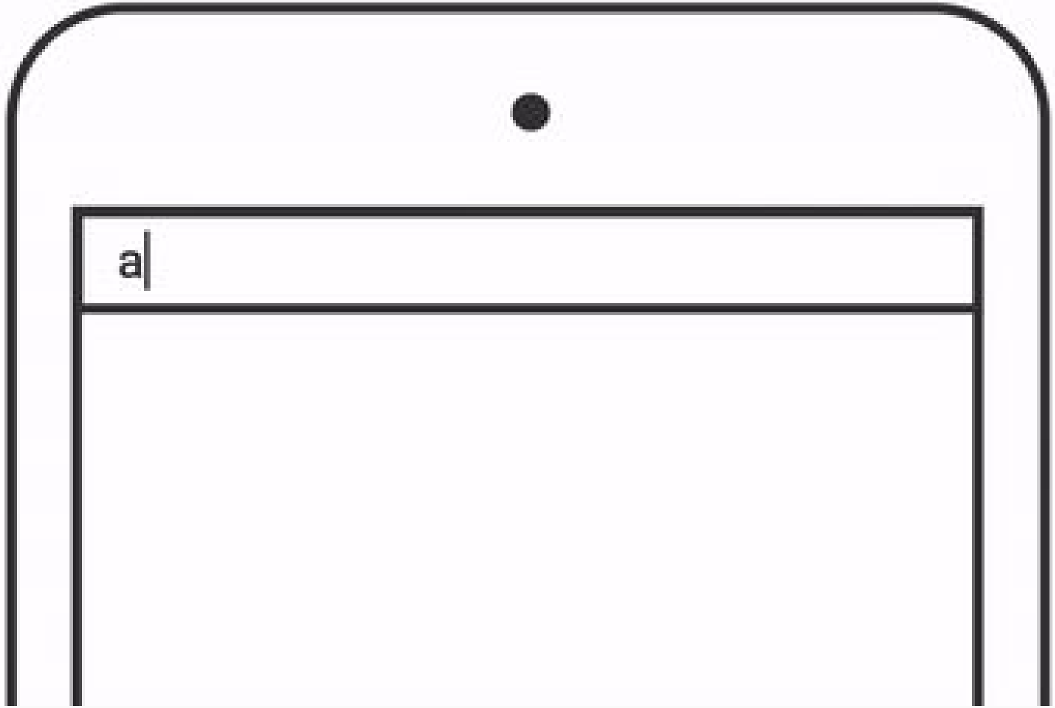
Email: bq-public-data@google.com



Break

& AutoDraw Activity

Autodraw.com *(try this on your phone!)*



Deep dive: Research

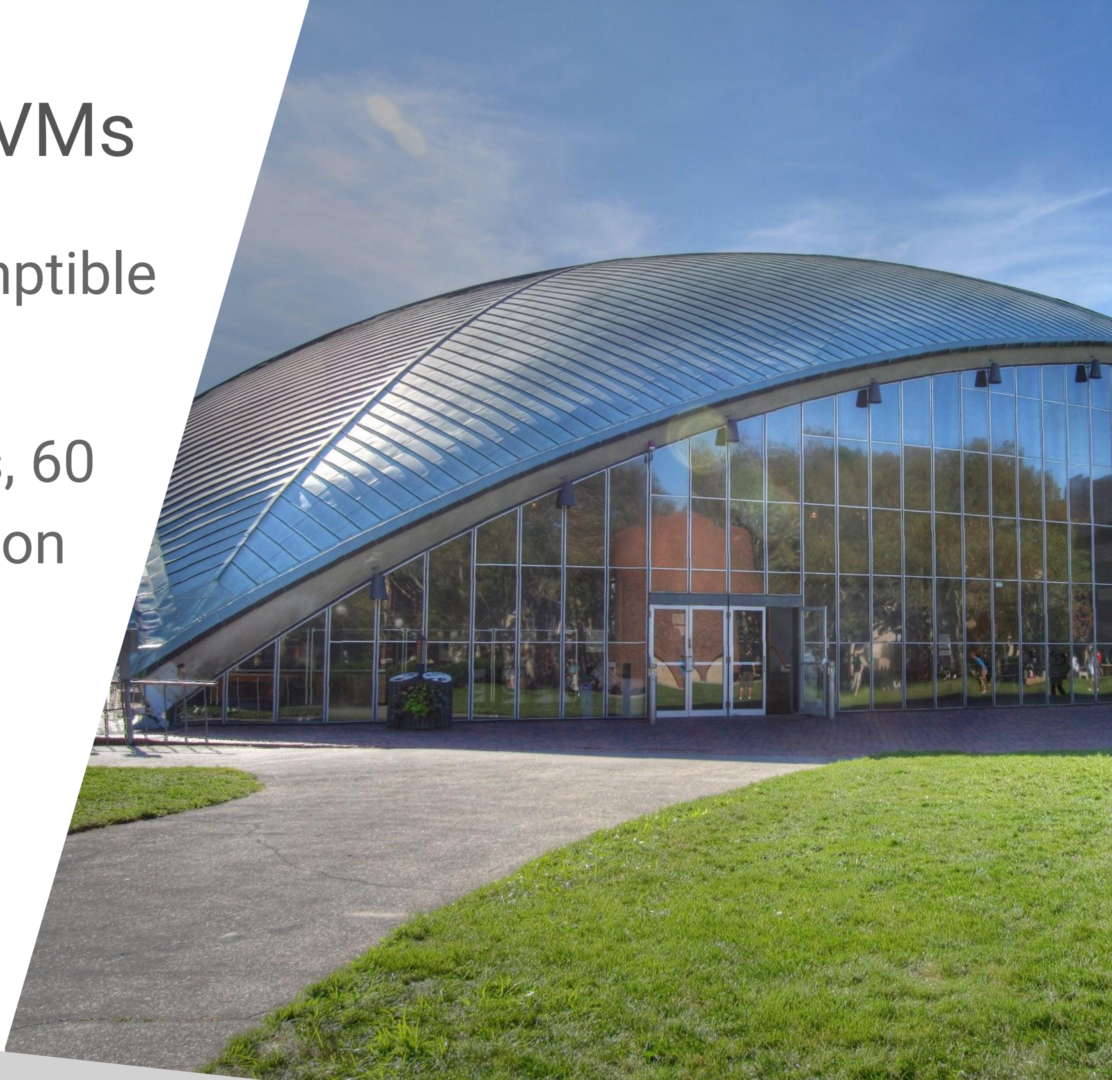
MIT Research w/ VMs

220,000 cores on preemptible VMs

2,250 32-core instances, 60 CPU-years of computation in a single afternoon

Answers in hours v. months

Products used: Google Compute Engine, Cloud Storage, DataStore



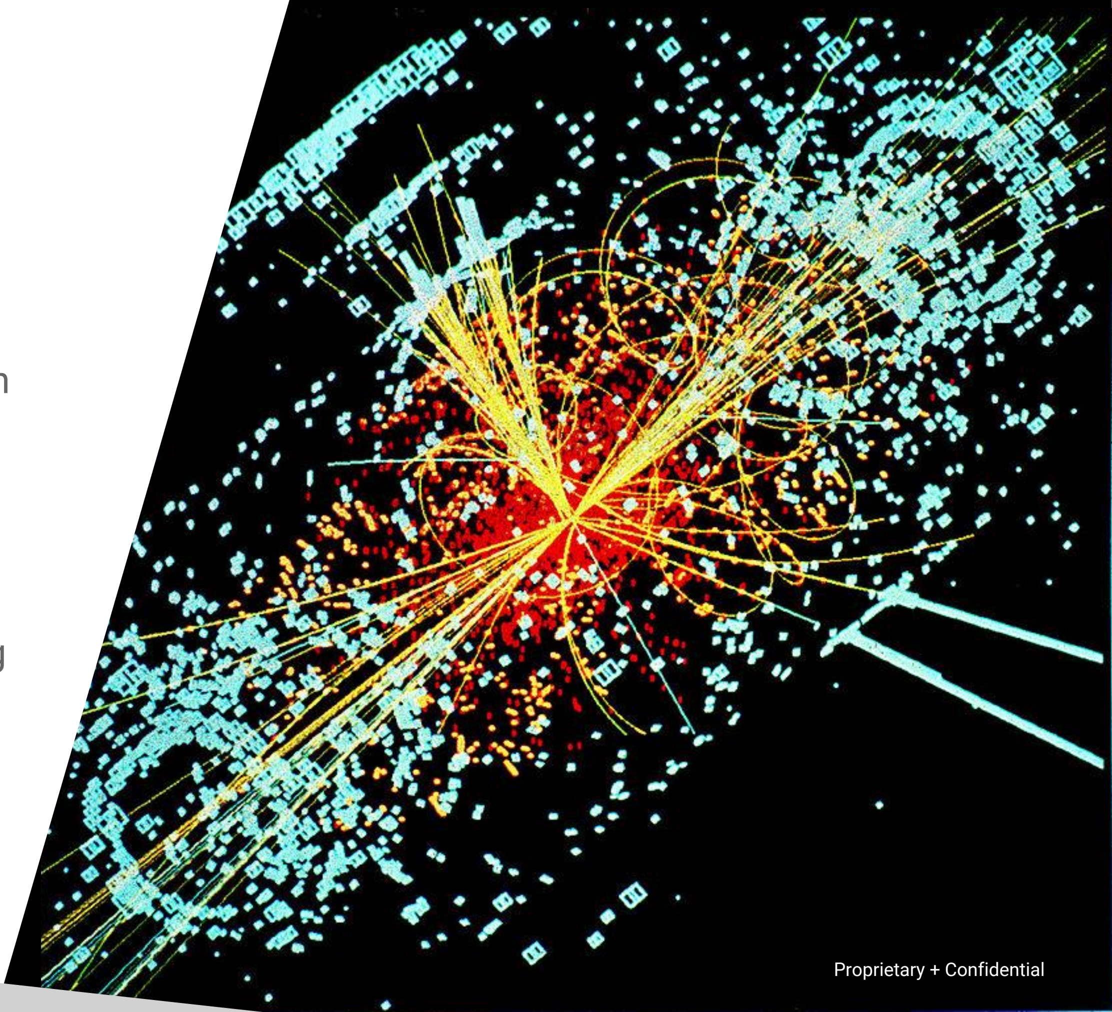
SC16 CMS Demonstrator

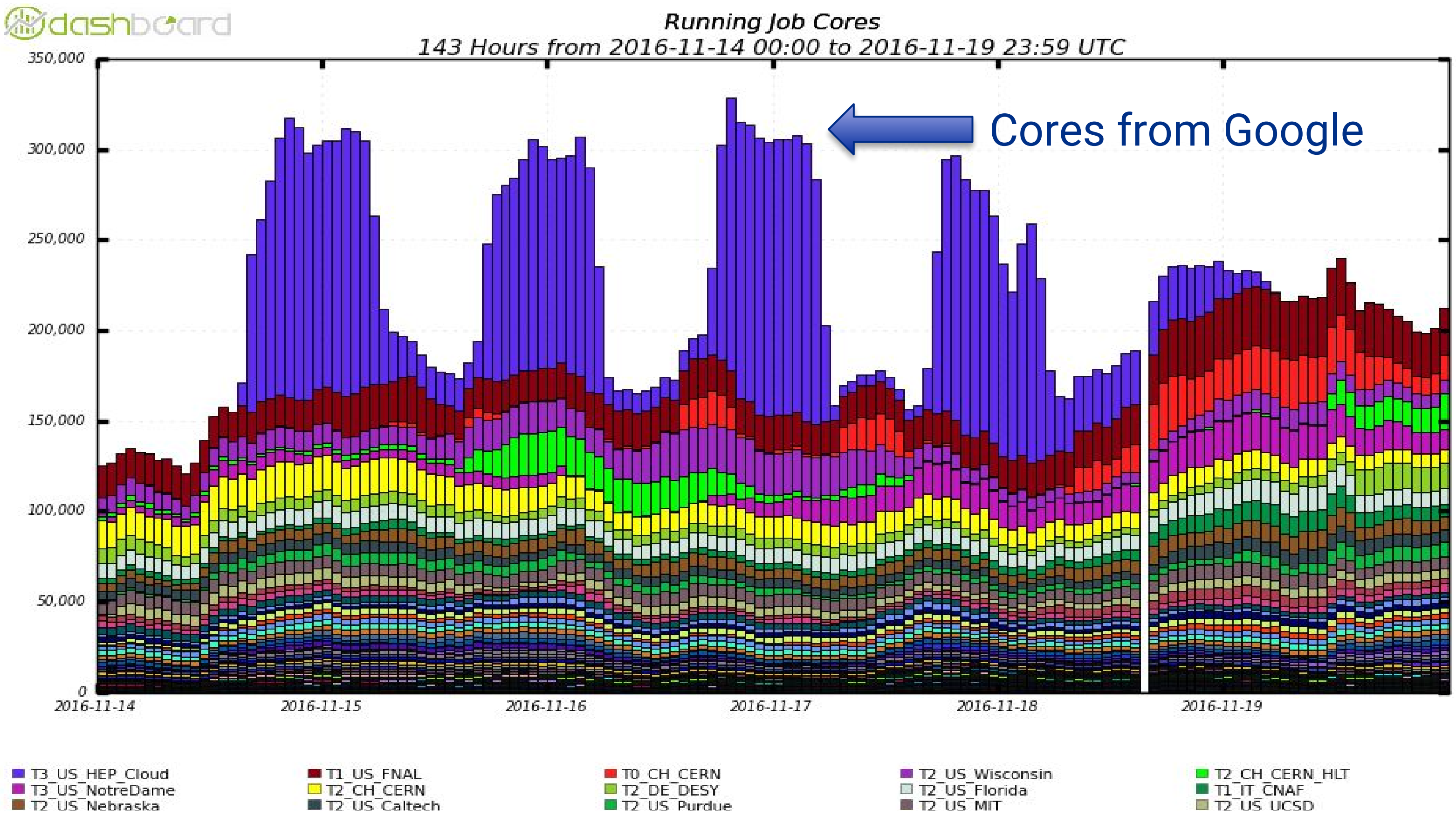
Target: generate 1 Billion events in 48 hours during Supercomputing 2016 on Google Cloud via HEPCloud

35% filter efficiency = stage out 380 million events → 150 TB output

Double the size of global CMS computing resources

CMS Higgs Event - credit: CERN
https://commons.wikimedia.org/wiki/File:CMS_Higgs-event.jpg



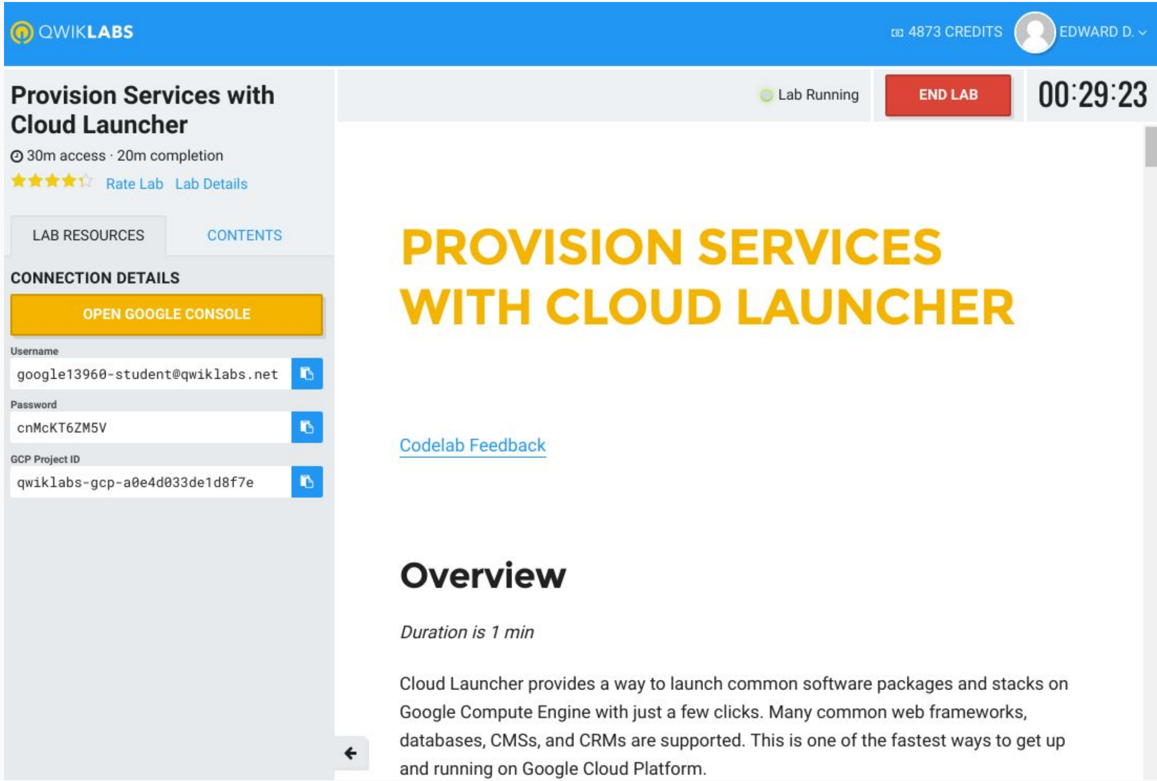

dashboard
Running Job Cores
143 Hours from 2016-11-14 00:00 to 2016-11-19 23:59 UTC
350,000
300,000
250,000
200,000
150,000
100,000
50,000
0
2016-11-14
2016-11-15
2016-11-16
2016-11-17
2016-11-18
2016-11-19
Cores from Google
T3_US_HEP_Cloud
T1_US_FNAL
T0_CH_CERN
T2_US_Wisconsin
T2_CH_CERN_HLT
T3_US_NotreDame
T2_CH_CERN
T2_DE_DESY
T2_US_Florida
T1_IT_CNAF
T2_US_Nebraska
T2_US_Caltech
T2_US_Purdue
T2_US_MIT
T2_US_UCSD

Interactive demos

Your turn to play!

https://google.qwiklabs.com/focuses/3241

Get a token from Edward



Roundtable discussions

Questions

1. What is one thing you learned today that excites you?
2. How could you see Google helping your university?
3. Any advice or feedback for us?

What's next?

Talk to us here

West: Lauren, Ed, Angela

East: Alicia

Get grants

Apply for CS class grants at cloud.google.com/edu

Train up

Access free training at coursera.org/googlecloud

Learn more

We'll email you future opportunities like webinars

Thank you!

Google Cloud