

# ML and Data Analytics with Google Cloud Platform

The power of machine learning on any data, any size

# Google

**Google Search**

**I'm Feeling Lucky**

A large, multi-story data center with a complex metal ceiling structure. The floor is filled with rows of server racks, many of which are illuminated with blue and yellow lights. The Google logo is superimposed in the center of the image.

# Google

Google Search

I'm Feeling Lucky



Alex Osterloh  
Solution Engineer, Google  
aosterloh@google.com  
@BigDataWizard





## 1st Wave Colocation

Your kit, someone else's  
building.  
Yours to manage.



## An Evolving Cloud

---

## 2nd Wave Virtualized Data Centers

Standard virtual kit, for rent.  
Still yours to manage.

Google Cloud Platform



?



### 1st Wave Colocation

Your kit, someone else's  
building.  
Yours to manage.




### An Evolving Cloud

---

### 2nd Wave Virtualized Data Centers

Standard virtual kit, for rent.  
Still yours to manage.



### 3rd Wave Automated Services Scalable Data

Focus in insight,  
not infrastructure





### 1st Wave Colocation

Your kit, someone else's  
building.  
Yours to manage.




### An Evolving Cloud

---

### 2nd Wave Virtualized Data Centers

Standard virtual kit, for rent.  
Still yours to manage.



### 3rd Wave Automated Services Scalable Data

Focus in insight,  
not infrastructure



Google Cloud Platform



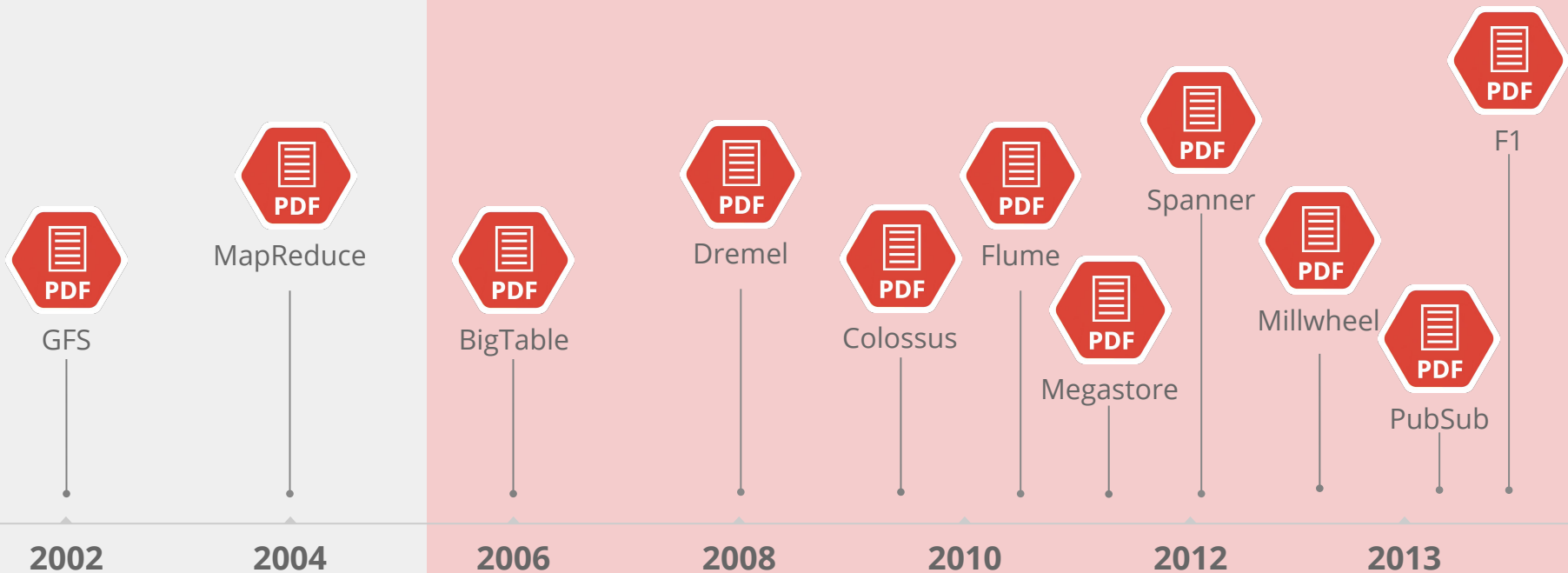


***“Google is living a few years in the future and sending the rest of us messages”***

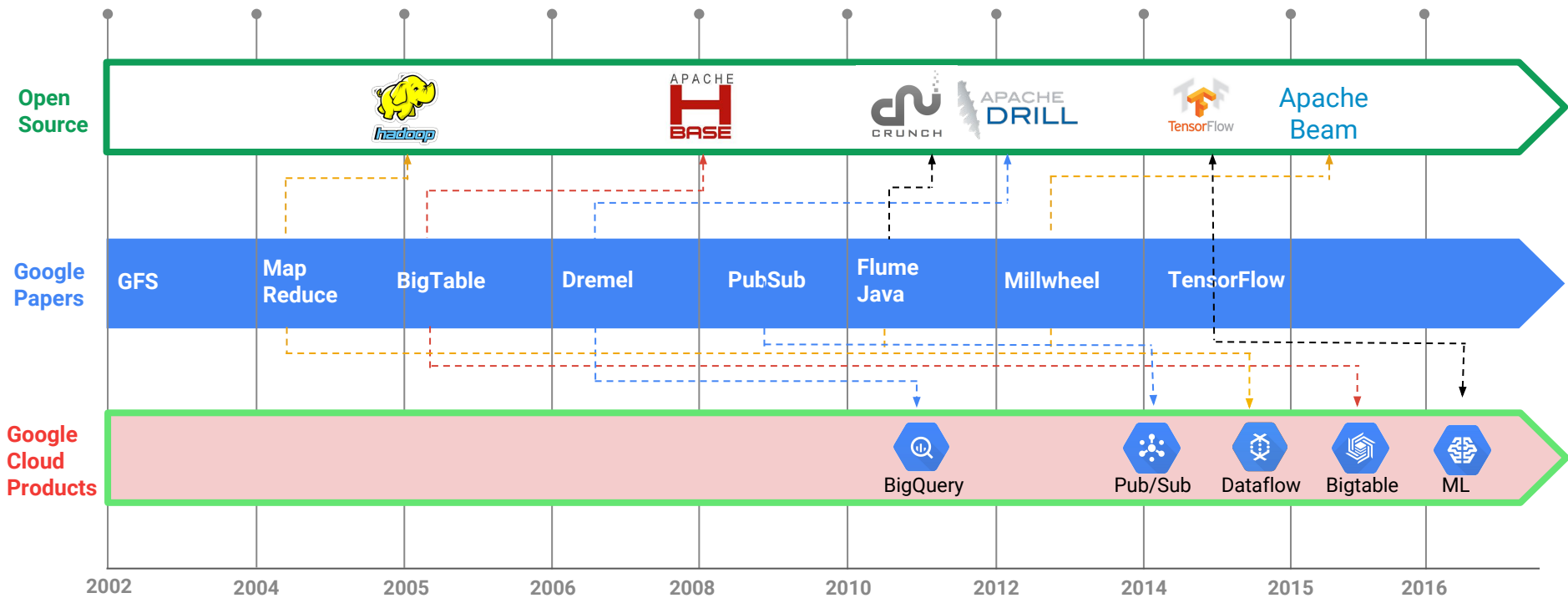
Doug Cutting  
Chief Architect Cloudera



# Google Research in Data Technologies



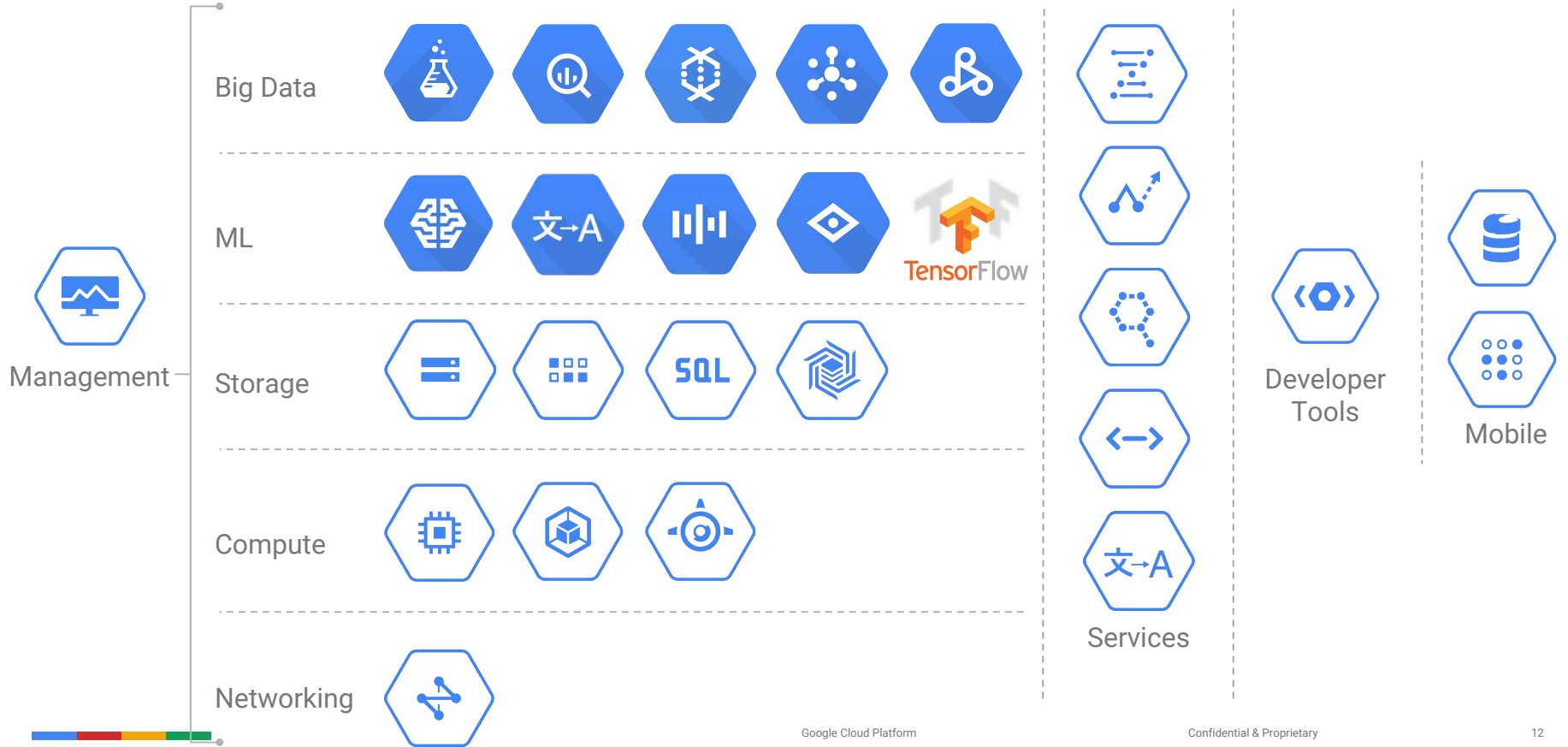
# 10+ Years of Tackling Data Problems



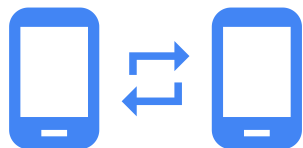


“ We don't really use MapReduce anymore ”

**Urs Hölzle**  
SVP Technical  
Infrastructure Google



# The Big Data Lifecycle



---

Capture

Pub/Sub



---

Store

Storage

SQL

Datastore

BigTable



---

Process

Dataflow



---

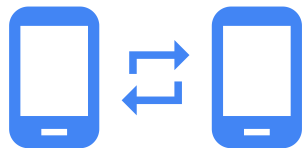
Analyze

BigQuery

Dataflow

Cloud ML

# The Big Data Lifecycle



---

Capture

Pub/Sub



---

Store

Storage

SQL

Datastore

BigTable



---

Process

Dataflow



---

~~Analyze~~

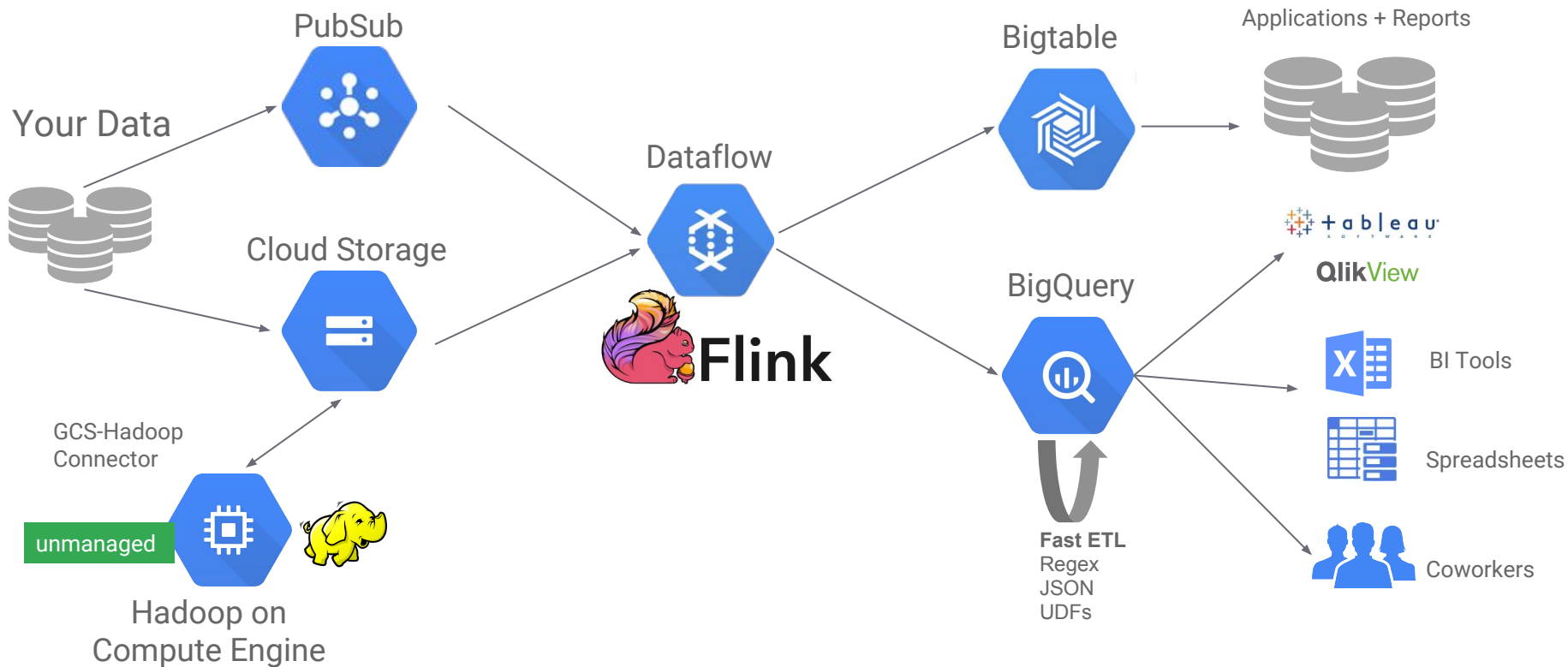
Learn

BigQuery

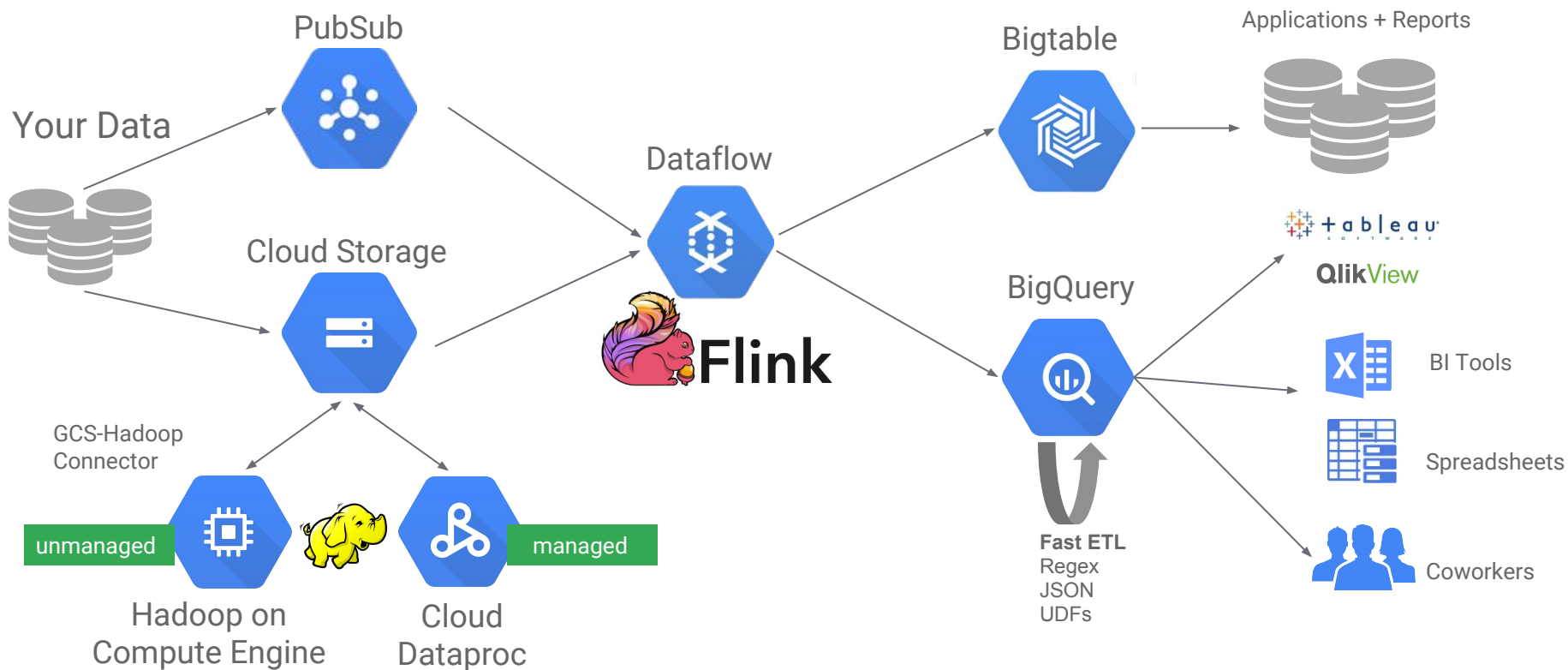
Dataflow

Cloud ML

# Enterprise Big Data Architecture on Google



# Enterprise Big Data Architecture on Google







### networking

### Storage

### Big Data

BigQuery

Cloud Dataflow

Cloud Dataproc

Clusters

Jobs

Genomics

Pub/Sub

New instance

New instance group

Reset

Start

Stop

Delete

### VM instances

CPU utilization

1 hour

6 hours

12

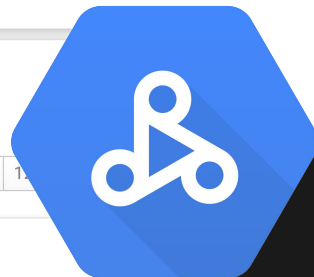
### CPU

% CPU

14

12

10



datapro-fun x +

```

Downloading file:///hello-world.py: 147 B/147 B
polleyg@datapro-fun:/var/tmp$ cat hello-world.py
#!/usr/bin/python
import pyspark
sc = pyspark.SparkContext()
rdd = sc.parallelize(['Hello', 'world!'])
words = sorted(rdd.collect())
print words

polleyg@datapro-fun:/var/tmp$ gcloud beta dataproc jobs submit pyspark --cluster datapro-train-challenge hello-world.py
Copying file:///hello-world.py [Content-Type=text/x-python]...
Uploading ...501ec-c1bc-44d6-ac61-03c20cf16a3b/hello-world.py: 147 B/147 B
Job [d2afd530-8bcd-45b8-b2f6-4a73497a4e66] submitted.
Waiting for job output...
15/10/12 06:34:02 INFO akka.event.slf4j.Slf4jLogger: Slf4jLogger started
15/10/12 06:34:02 INFO Remoting: Starting remoting
15/10/12 06:34:02 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://sparkDriver@10.240.0.4:53655]
15/10/12 06:34:02 INFO org.spark-project.jetty.server.Server: jetty-8.y.z-SNAPSHOT
15/10/12 06:34:02 INFO org.spark-project.jetty.server.AbstractConnector: Started SocketConnector@0.0.0.0:37047
15/10/12 06:34:02 INFO org.spark-project.jetty.server.Server: jetty-8.y.z-SNAPSHOT
15/10/12 06:34:03 INFO org.spark-project.jetty.server.AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040
15/10/12 06:34:03 WARN org.apache.spark.metrics.MetricsSystem: Using default name DAGScheduler for source because spark.ap
15/10/12 06:34:03 INFO org.apache.hadoop.yarn.client.RMPProxy: Connecting to ResourceManager at datapro-train-challenge-m/
15/10/12 06:34:05 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl: Submitted application application_1444631143

```

My work c

My commute to

uneventful affa

mobile devices,

people who like

have been up to

mplementation class:org  
d  
own - will be ignored  
e ignored



# Applications that can see, hear & understand



Google machine learning

All News Books Videos Images More Search tools

About 124,000,000 results (0.42 seconds)

**Machine Learning - 5 Courses & a Capstone Project**  
www.coursera.org/machine-learning  
Earn a U of Washington Certificate!  
Learn 24/7 On Any Device - Flexible Schedule - Top Instructors - Top Universities  
Highlights: Case-Based Introduction, Real-World Problem...  
Sign Up for Free - How Coursera Works

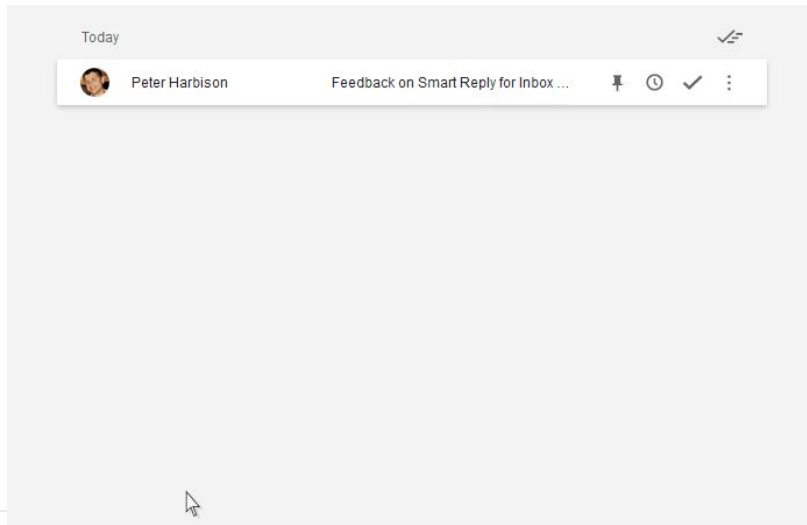
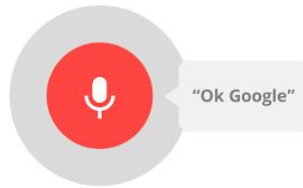
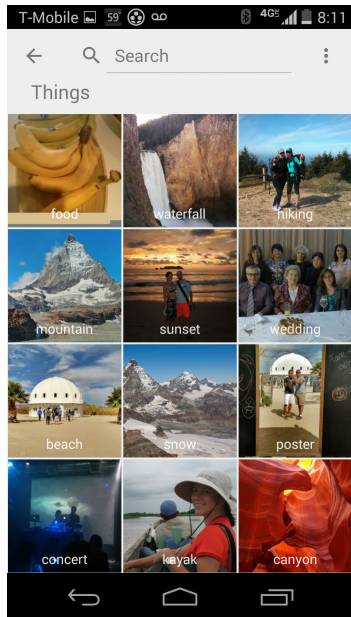
**MIT ProEd Class Jun 18-22 - mit.edu**  
prof.mit.edu/MachineLearning  
Big Data Machine Learning at MIT. Learn, Implement, Predict. Register

**Machine learning** is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. In 1959, Arthur Samuel defined **machine learning** as a "Field of study that gives computers the ability to learn without being explicitly programmed".

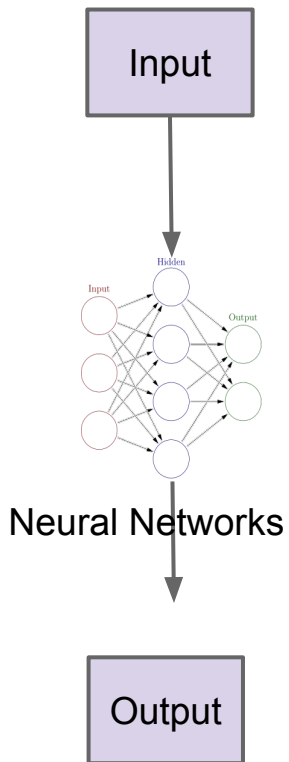
Machine learning - Wikipedia, the free encyclopedia  
https://en.wikipedia.org/wiki/Machine\_learning


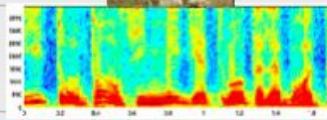

More about Machine learning

About this result • Feedback



# Examples of applying ML



	Input	Output
Pixels:		"lion"
Audio:		"see at tuhl res taur aun ts"
	<query, doc>	P(click on doc)
	"Hello, how are you?"	"Bonjour, comment allez-vous?"
Pixels:		"A close up of a small child holding a stuffed animal"

# Machine Learning Use Cases

## Structured Data

### *Classification/ Regression*

- Customer Churn Analysis
- Product Diagnostics
- Forecasting

### *Recommendation*

- Content Personalization
- Product X-Sells/Up-sells

### *Anomaly Detection*

- Fraud Detection
- Asset Sensor Diagnostics
- Log Metric Anomalies

## Unstructured Data

### *Image Analytics*

- Identify damaged shipments
- Explicit Content Classification
- Identify “styles” in images

### *Text Analytics*

- Call Center log analysis
- Language Identification
- Topic Classification
- Sentiment Analysis



# The Spectrum of Machine Learning

# Use pretrained models



Cloud  
Translate API



Cloud  
Vision API



Cloud  
Speech API

Or use your own data to train models



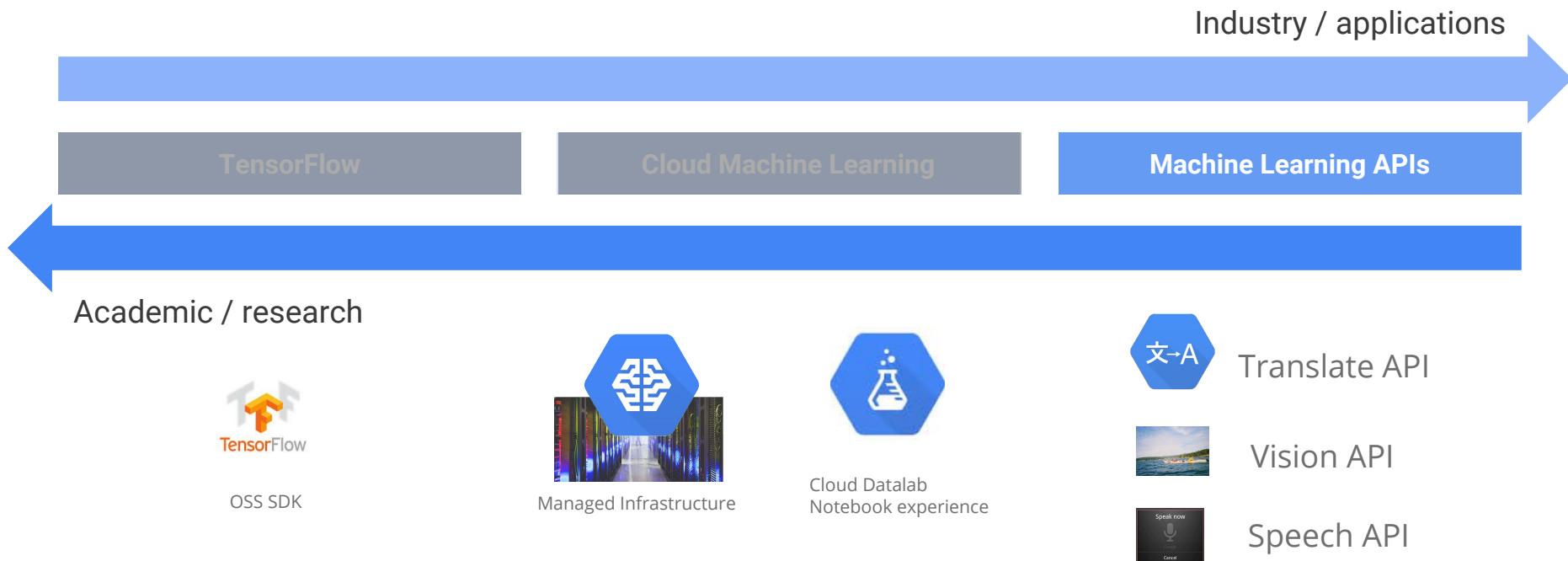
TensorFlow

[www.tensorflow.org](http://www.tensorflow.org)



Cloud Machine Learning

# The Machine Learning Spectrum





# Google Cloud Vision API

- Detect faces, landmarks, logos, text, and more
- Perform sentiment analysis
- Straightforward REST API
- Works on a base64-encoded image
- Connects to Google Cloud Storage
- Returns label, score pair



### Step 1. Upload Image File

Choose File No file chosen



### Step 2. Choose Detection Type

Label Detection

Text Detection

Face Detection

Landmark Detection

Logo Detection

Safe Search Detection

Image Properties

### Step 3. Check the result: LANDMARK\_DETECTION

```
{
  "responses": [
    {
      "landmarkAnnotations": [
        {
          "description": "Geneva",
          "score": 0.34884268,
          "locations": [
            {
              "latLng": {
                "latitude": 46.201215,
                "longitude": 6.143333000000001
              }
            }
          ]
        }
      ],
      "boundingPoly": {
        "vertices": [
          {
            "y": 318,
            "x": 218
          },
          {
            "y": 318,
            "x": 567
          },
          {
            "y": 365,
            "x": 567
          },
          {
            "y": 365,
            "x": 218
          }
        ]
      }
    }
  ]
}
```

### Step 1. Upload Image File

Choose File No file chosen



### Step 2. Choose Detection Type

Label Detection
<b>Text Detection</b>
Face Detection
Landmark Detection
Logo Detection
Safe Search Detection
Image Properties

### Step 3. Check the result: TEXT\_DETECTION

```
{
  "responses": [
    {
      "textAnnotations": [
        {
          "locale": "no",
          "description": "MobilB\n",
          "boundingPoly": {
            "vertices": [
              {
                "y": 616,
                "x": 233
              },
              {
                "y": 616,
                "x": 261
              },
              {
                "y": 629,
                "x": 261
              },
              {
                "y": 629,
                "x": 233
              }
            ]
          }
        },
        {
          "description": "MobilB",
          "boundingPoly": {
            "vertices": [
              {
                "y": 616,
```

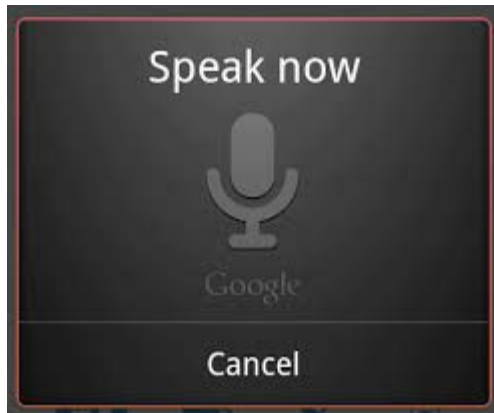


# Google Cloud Speech API

- Pass raw audio data and language
- Returns a transcript of the audio data
- Works across >80 languages
- Receive response in streaming **or** non-streaming



# Speech API



[Click for Demo](#)

- Enable voice interface to devices and applications
- Transcribe audio from stored media
- Multiple language support
- Access from mobile devices

# Speech API Demo

*“What are you sinking about ? ”*



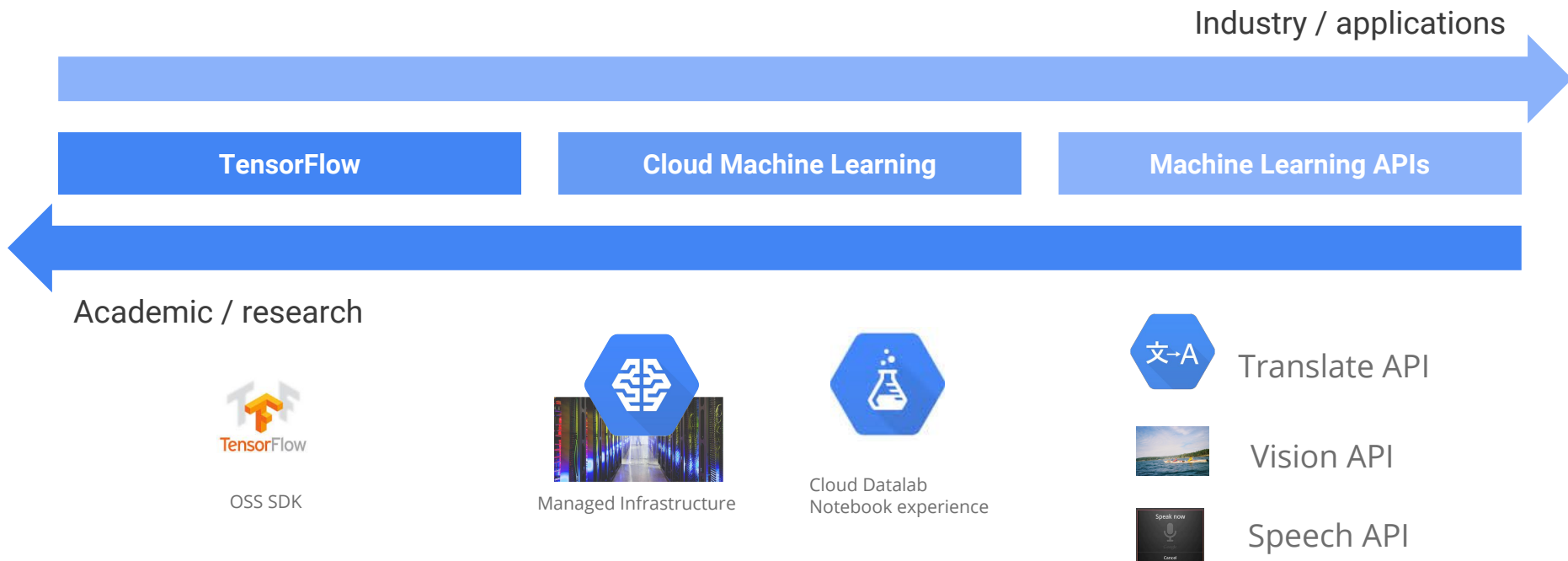
Click for Demo

# Google Cloud Translate API

- translate text between thousands of language pairs.
- let's websites and programs integrate with Google Translate programmatically

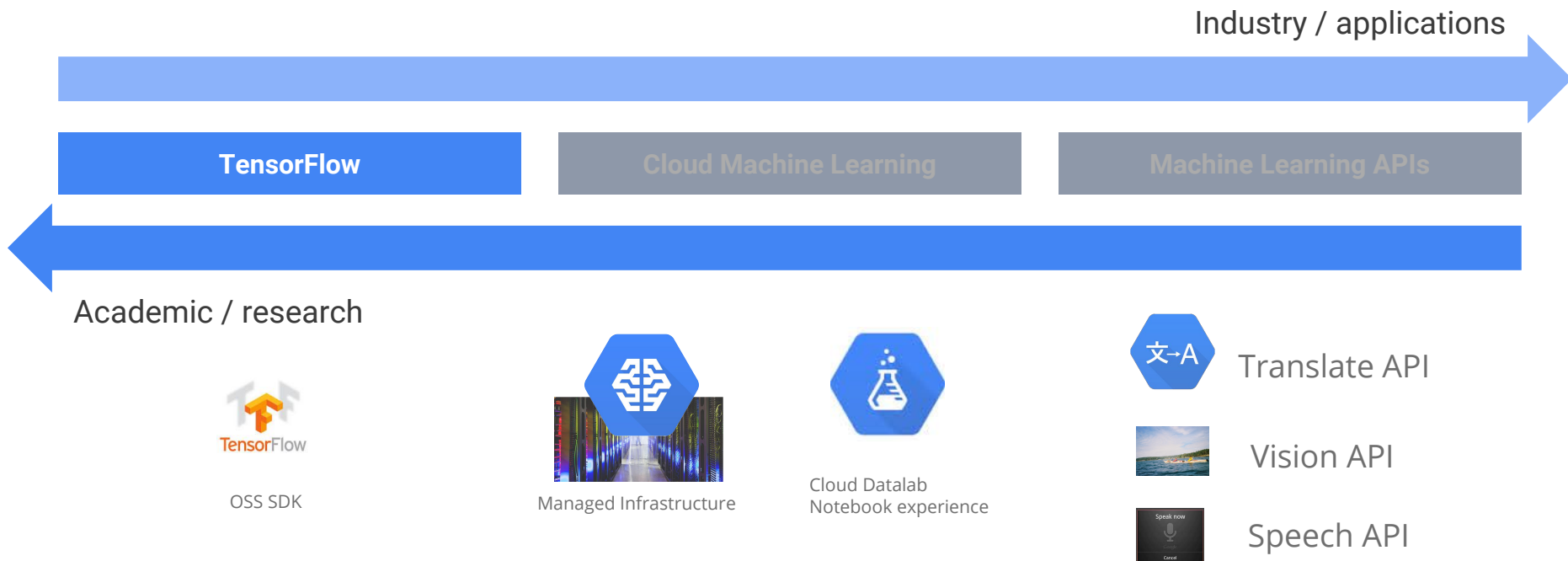


# The Machine Learning Spectrum





# The Machine Learning Spectrum



# A brief look at TensorFlow

Largest Machine Learning repository on GitHub

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

Using a **flow graph**: *data flow computation framework*

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



# A brief look at TensorFlow

Largest Machine Learning repository on GitHub

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

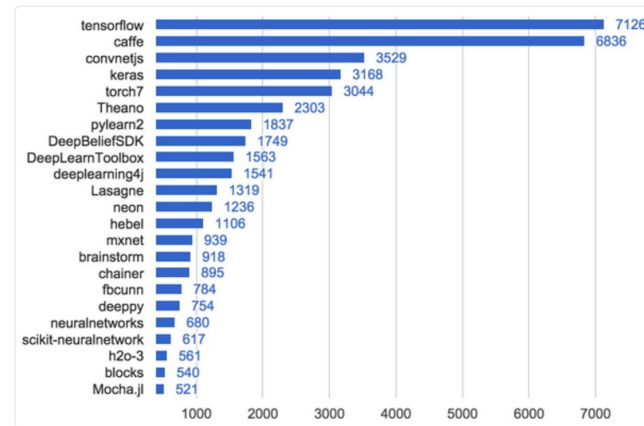
Using a **flow graph**: *data flow computation framework*

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

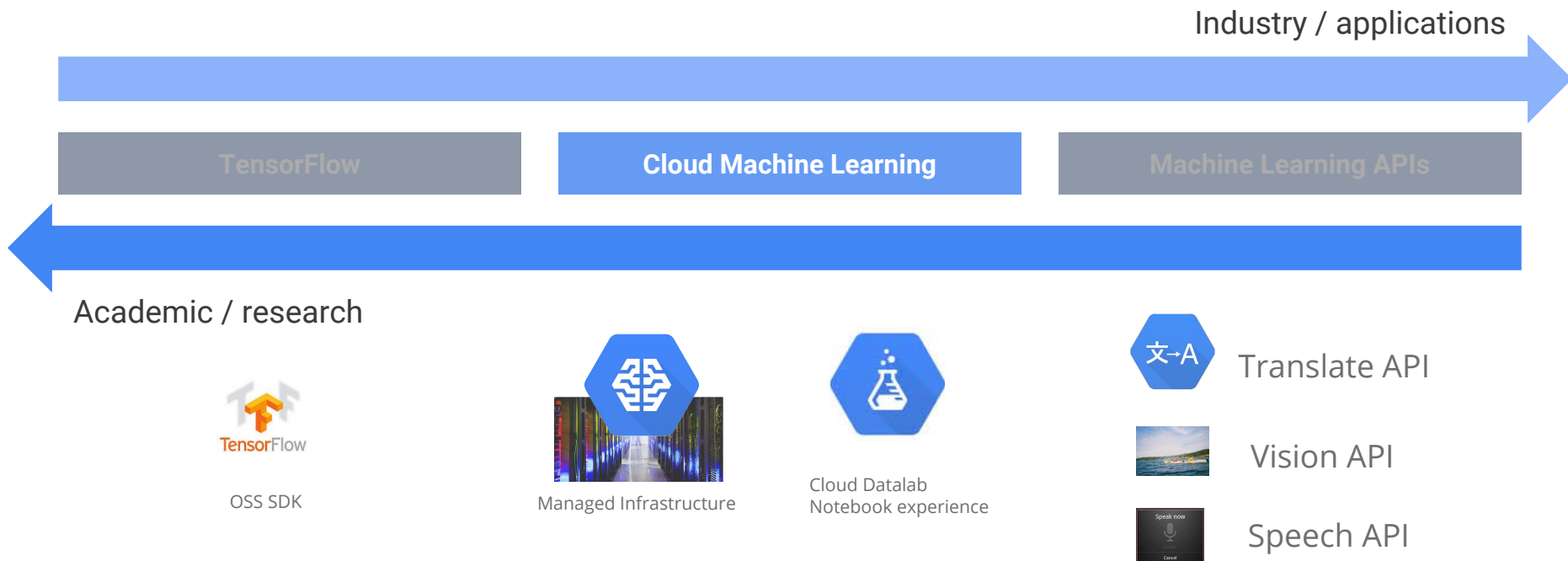
Kyle McDonald  
@kcmcd



2010-2014: a new deep learning toolkit is released every 47 days. 2015: every 22 days.  
tensorflow & caffe top github



# The Machine Learning Spectrum



# What Cloud Machine Learning Can Do

- Fully managed service
- Train using a custom Tensor Flow graph
- Batch and online predictions, at scale
- Integrated Datalab experience
- Regression and classification tasks



Want more ? → <http://bit.ly/gcp16data>



A large industrial facility, possibly a data center or manufacturing plant, is shown at night. The building has a complex structure with many windows and is illuminated from within, creating a warm glow. The sky is dark with some clouds, and there are some lights visible in the distance. The text "Thank You" is overlaid in the center in a large, white, sans-serif font.

# Thank You

Alex Osterloh  
[aosterloh@google.com](mailto:aosterloh@google.com)