ML and Data Analytics with Google Cloud Platform

The power of machine learning on any data, any size
Alex Osterloh
Solution Engineer, Google
aosterloh@google.com
@BigDataWizard
1st Wave
Colocation
Your kit, someone else's building.
Yours to manage.

2nd Wave
Virtualized
Data Centers
Standard virtual kit, for rent.
Still yours to manage.

An Evolving Cloud

Google Cloud Platform
Colocation

Virtualized Data Centers

Automated Services

3rd Wave Scalable Data

An Evolving Cloud

1st Wave Colocation
Your kit, someone else's building. Yours to manage.

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
An Evolving Cloud

1st Wave Colocation
Your kit, someone else’s building. Yours to manage.

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

GFS
MapReduce
BigTable
Dremel
Colossus
Flume
Spanner
Megastore
Millwheel
PubSub


Google Research Publications referenced are available here: [http://research.google.com/pubs/papers.html](http://research.google.com/pubs/papers.html)
We don’t really use MapReduce anymore

Urs Hölzle
SVP Technical Infrastructure
Google
Google Cloud Platform

Big Data

ML

Storage

Compute

Networking

Management

Developer Tools

Mobile

Services
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
BigQuery
Dataflow
Cloud ML

Learn
Enterprise Big Data Architecture on Google

Your Data

PubSub

Cloud Storage

Bigtable

Dataflow

BigQuery

Flink

GCS-Hadoop Connector

Hadoop on Compute Engine

Applications + Reports

PubSub

Fast ETL

Regex

JSON

UDFs

BI Tools

Spreadsheets

Coworkers

Google Cloud Platform
Confidential & Proprietary
Enterprise Big Data Architecture on Google

Your Data
- Cloud Storage
  - GCS-Hadoop Connector
- Hadoop on Compute Engine
- Cloud Dataproc
- PubSub
  - Dataflow
- Bigtable
  - Fast ETL
    - Regex
    - JSON
    - UDFs
- Applications + Reports
  - BI Tools
    - QlikView
  - Spreadsheets
  - Coworkers

Applications + Reports
- BI Tools
- Spreadsheets
- Coworkers

Google Cloud Platform
Confidential & Proprietary
Page 16
Applications that can see, hear & understand
## Examples of applying ML

<table>
<thead>
<tr>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pixels: <img src="image1.png" alt="Image of lion" /></td>
<td>“lion”</td>
</tr>
<tr>
<td>Audio: <img src="image2.png" alt="Audio waveform" /></td>
<td>“see at tuhl res taur aun ts”</td>
</tr>
<tr>
<td><code>&lt;query, doc&gt;</code></td>
<td>$P(\text{click on doc})$</td>
</tr>
<tr>
<td>“Hello, how are you?”</td>
<td>“Bonjour, comment allez-vous?”</td>
</tr>
<tr>
<td>Pixels: <img src="image3.png" alt="Image of child" /></td>
<td>“A close up of a small child holding a stuffed animal”</td>
</tr>
</tbody>
</table>
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

Cloud Machine Learning
The Machine Learning Spectrum

Academic / research

- TensorFlow
- OSS SDK
- Managed Infrastructure

Cloud Machine Learning

- Cloud Datalab Notebook experience

Industry / applications

- Machine Learning APIs
- Translate API
- Vision API
- Speech API
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 3. Check the result: LANDMARK_DETECTION

```json
{
  "responses": [
    {
      "landmarkAnnotations": [
        {
          "description": "Geneva",
          "score": 0.34884268,
          "locations": [
            {
              "latLng": {
                "latitude": 46.201215,
                "longitude": 6.14333300000001
              }
            }
          ]
        }
      ],
      "boundingPoly": {
        "vertices": [
          {
            "y": 318,
            "x": 218
          },
          {
            "y": 318,
            "x": 567
          },
          {
            "y": 365,
            "x": 567
          },
          {
            "y": 365,
            "x": 218
          }
        ]
      }
    }
  ]
}```
Step 3. Check the result: TEXT_DETECTION
{
 "responses": [
   {
     "textAnnotations": [
       {
         "locale": "en",
         "description": "MobilB",
         "boundingPoly": {
           "vertices": [
             {
               "y": 616,
               "x": 233
             },
             {
               "y": 616,
               "x": 261
             },
             {
               "y": 629,
               "x": 261
             },
             {
               "y": 629,
               "x": 233
             }
           ]
         }
       },
       {
         "description": "MobilB",
         "boundingPoly": {
           "vertices": [
             {
               "y": 616,
               "x": 233
             },
             {
               "y": 616,
               "x": 261
             },
             {
               "y": 629,
               "x": 261
             },
             {
               "y": 629,
               "x": 233
             }
           ]
         }
       }
     ]
   }
]
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

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

Click for Demo
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

TensorFlow
Cloud Machine Learning
Machine Learning APIs

Industry / applications

Academic / research

OSS SDK
Managed Infrastructure
Cloud Datalab Notebook experience

Translate API
Vision API
Speech API
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: \textit{n-dimensional arrays}

Using a flow graph: \textit{data flow computation framework}

- Train on CPUs, GPUs
- Run wherever you like (local, cloud, mobile)
The Machine Learning Spectrum

- **TensorFlow**
  - OSS SDK
- **Cloud Machine Learning**
  - Managed Infrastructure
  - Cloud Datalab Notebook experience
- **Machine Learning APIs**
  - Translate API
  - Vision API
  - Speech API

**Academic / research**

**Industry / applications**
What Cloud Machine Learning Can Do

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

Alex Osterloh
aosterloh@google.com