<Start Training>

Welcome to Cloud OnBoard

#GoogleCloudOnBoard

#NowICan

@GoogleCloud_TPE

</Start Training>

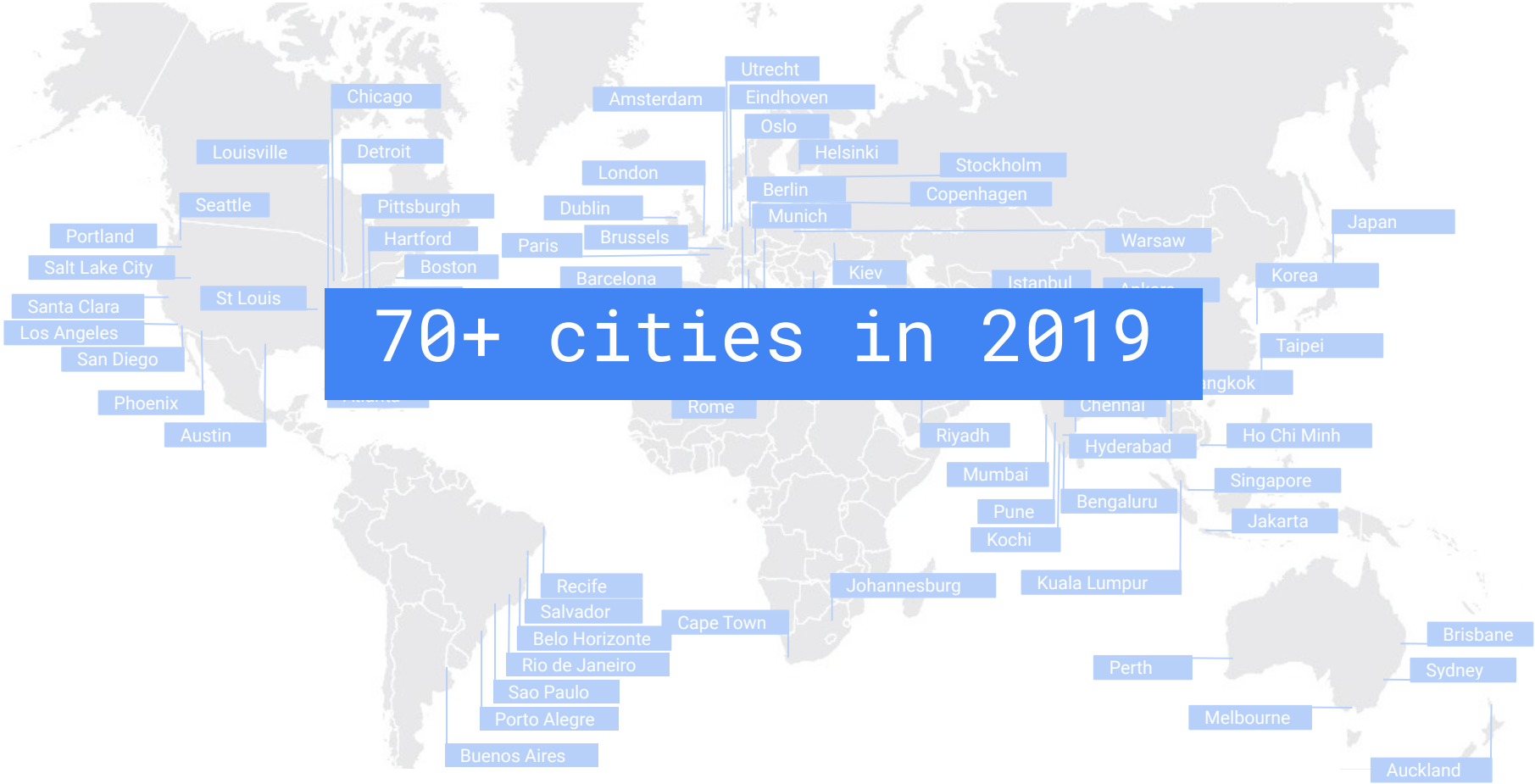
Version #1.1

Welcome

[Ricky Lin]

[Account Executive,Google Cloud]

Google Cloud Cloud OnBoard




Google Cloud
Cloud OnBoard
Chicago
Louisville
Detroit
Seattle
Pittsburgh
Portland
Hartford
Salt Lake City
Boston
St Louis
NYC
Santa Clara
Los Angeles
San Diego
Washington DC
Phoenix
Atlanta
Austin
Utrecht
Amsterdam
Eindhoven
Oslo
Helsinki
London
Stockholm
Berlin
Copenhagen
Dublin
Munich
Paris
Brussels
Warsaw
Korea
Barcelona
Kiev
Istanbul
Ankara
Japan
Madrid
Hamburg
Bucharest
Delhi
Milan
Prague
Dubai
Kolkata
Taipei
Rome
Chennai
Bangkok
Riyadh
Hyderabad
Ho Chi Minh
Mumbai
Singapore
Pune
Bengaluru
Jakarta
Kochi
Recife
Johannesburg
Kuala Lumpur
Salvador
Cape Town
Brisbane
Belo Horizonte
Perth
Sydney
Rio de Janeiro
Sao Paulo
Melbourne
Porto Alegre
Buenos Aires
Auckland

Google Cloud Cloud OnBoard

Getting started on your Google Cloud learning journey

1

Today

Google Cloud Platform Fundamentals: Big Data and Machine Learning

2

Tomorrow

Complete hands-on labs: Baseline: Data, ML, AI quest

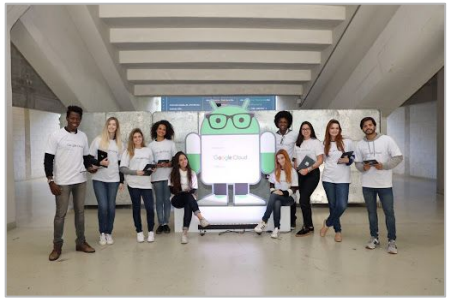
google.qwiklabs.com

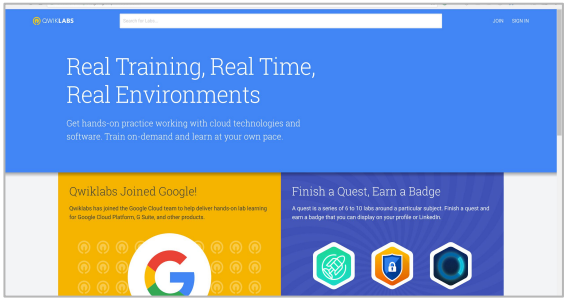
3

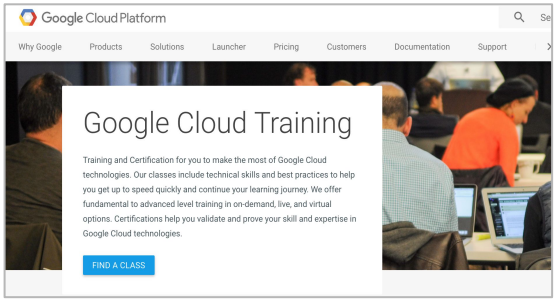
Future

Find more training online

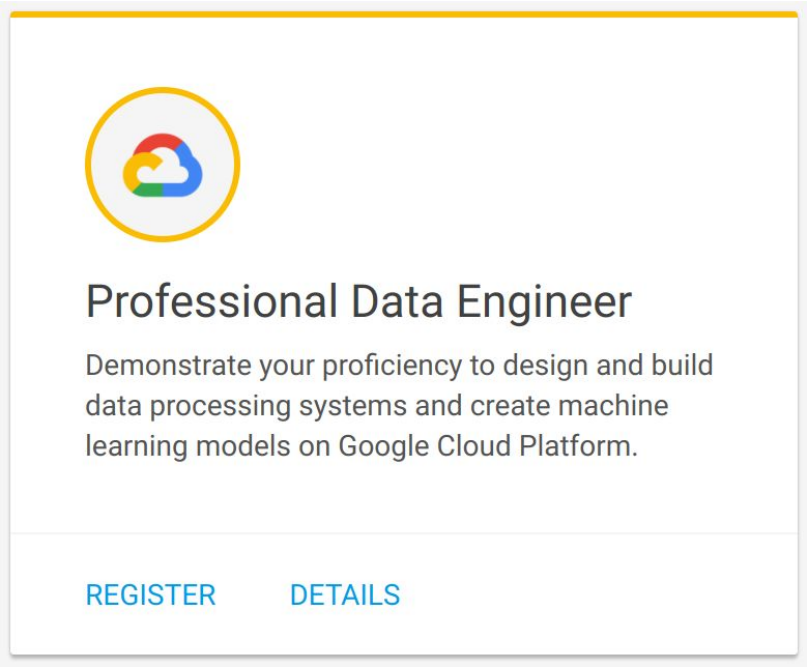
cloud.google.com/training







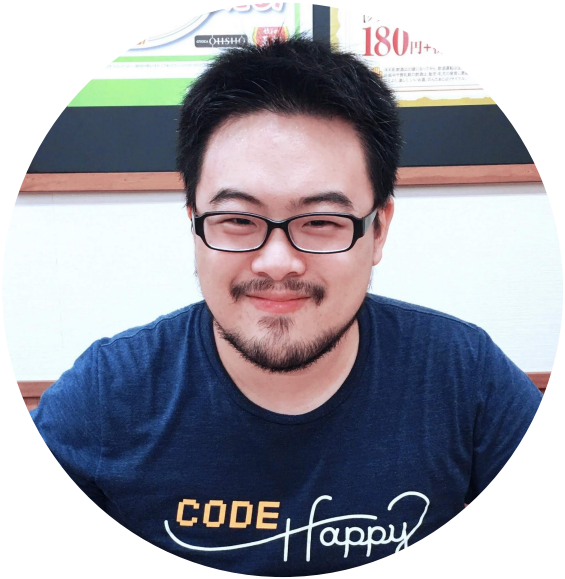
Make Google Cloud certification your goal!



Find study guides, tips, practice exams, and testing sites

cloud.google.com/certification

Instructor



[Eefy Lin]
[Instructor,Google Cloud]

email:[eefylin@google.com]

<Start Training>

#GoogleCloudOnBoard

</Start Training>

Cloud OnBoard <Agenda>

{ Google Cloud Platform Fundamentals: Big Data and Machine Learning

( ) Registration		08:30 AM
	Welcome and opening	10:00 AM
(‘Module 1’)	Introduction to and Fundamentals of GCP	10:15 AM
(‘Module 2’)	Migrate to Google Cloud in 5 mins (Anthos Migrate)	10:45 AM
( ) Lunch		12:00 PM
(‘Module 3’)	Data Analysis on the Cloud	01:00 PM
(‘Module 4’)	Scaling Data Analysis	02:00 PM
( ) Break		03:00 PM
(‘Module 5’)	Machine Learning (w/ Developer’s Sharing)	03:30 PM
(‘Module 6’)	Data Processing Architecture (w/ Customer’s Sharing)	04:45 PM
	Closing Session | Continue learning with Google Cloud	05:15 PM

}

Cloud OnBoard

Introduction
Google Cloud Platform:
Big Data and Machine Learning

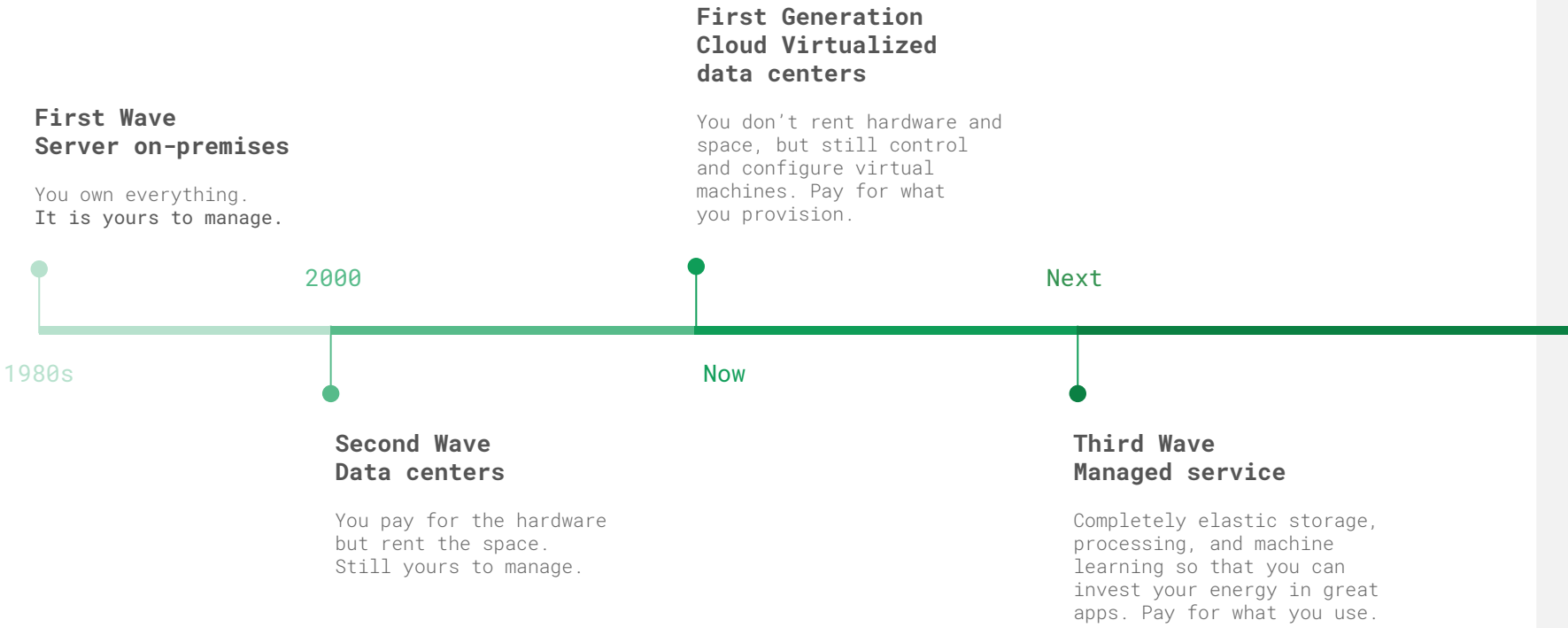
Cloud OnBoard

Agenda

What is Google Cloud Platform

Google Cloud Big Data products

Cloud computing is a continuation of a long-term shift in how computing resources are managed

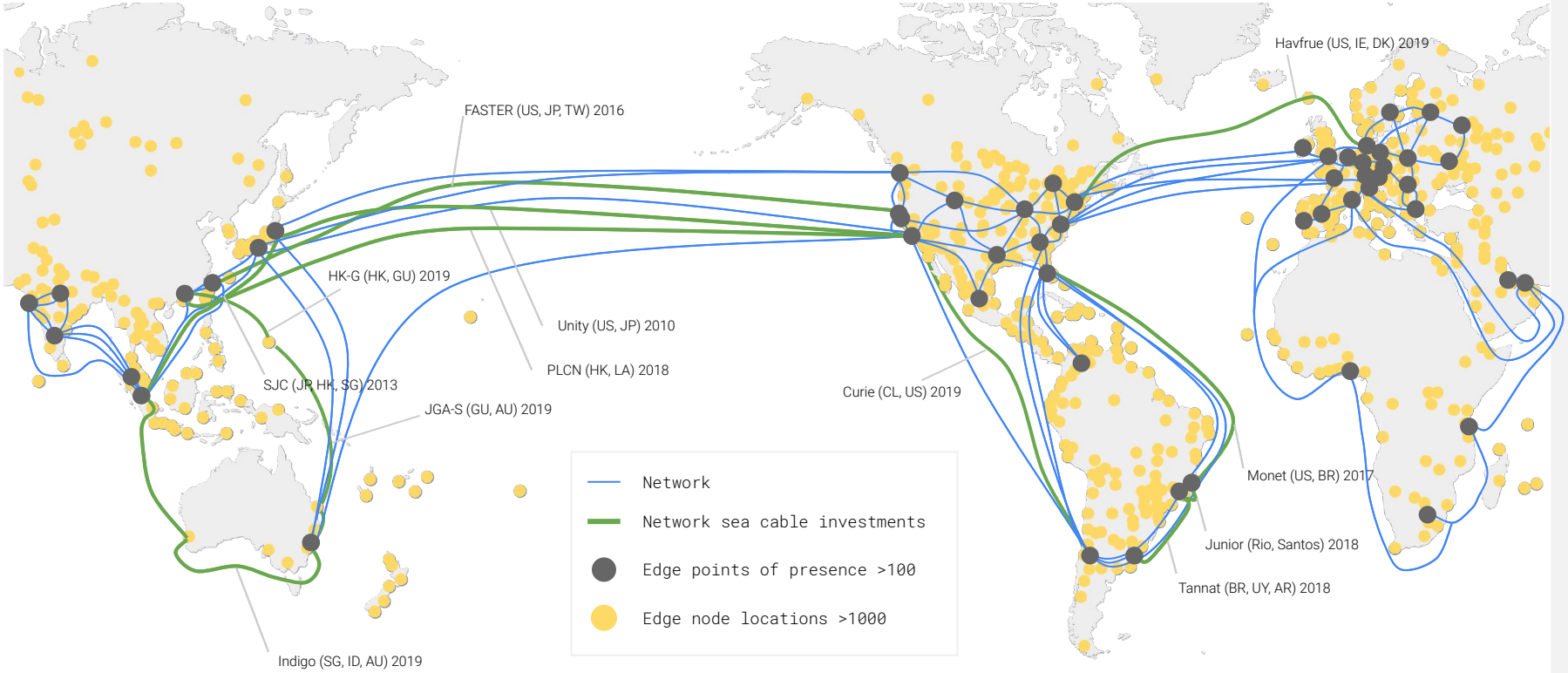



Google’s mission is to organize the world’s information and make it universally accessible and useful

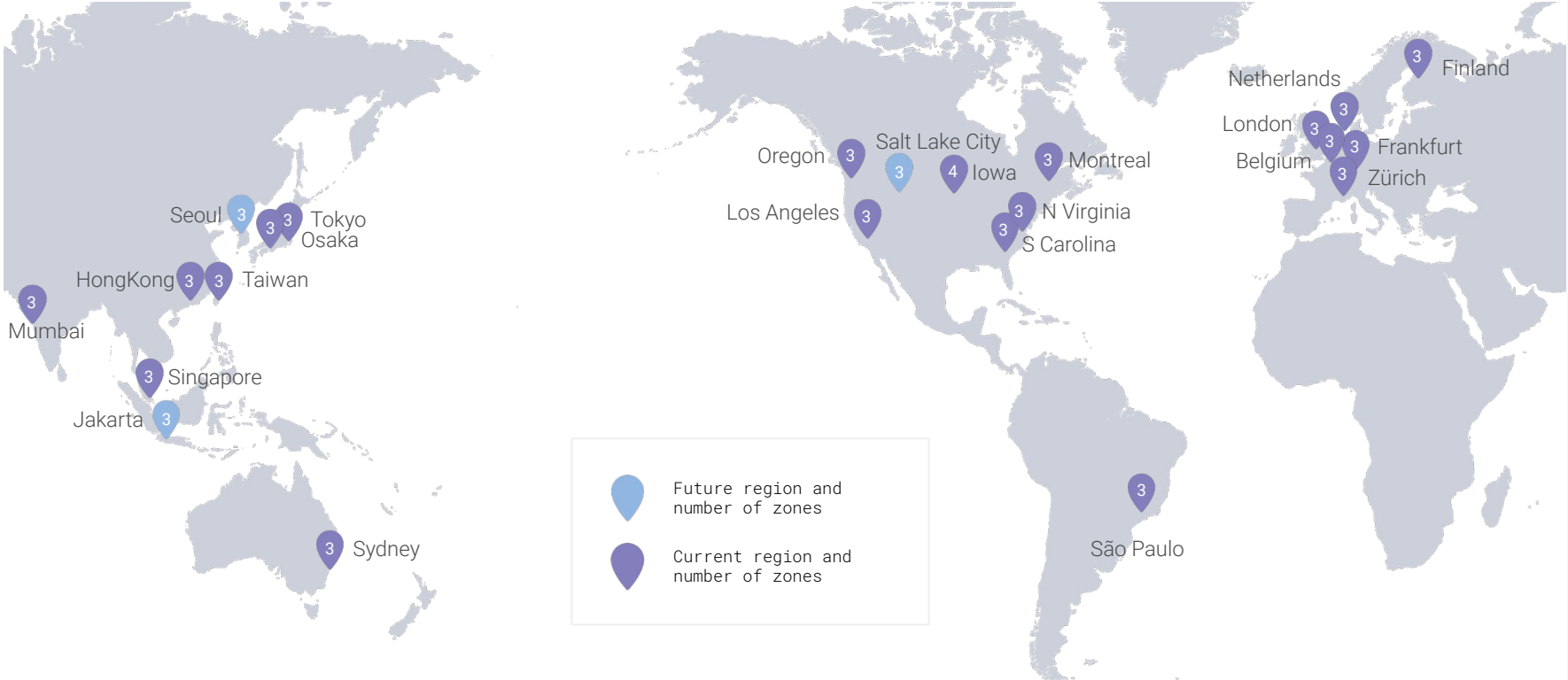

To organize the world's information, Google has been building the most powerful infrastructure on the planet
Google | google.com/datacenters

Google Cloud Cloud OnBoard

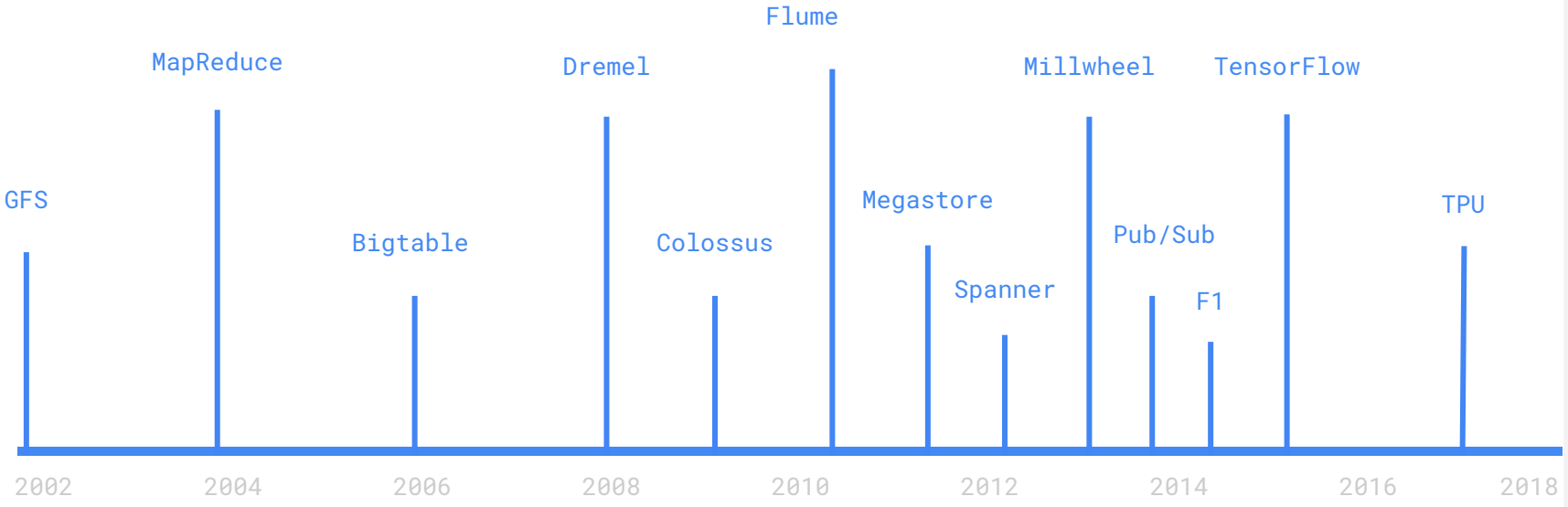
In terms of hardware, Google Cloud has the largest cloud network, with over 100 points of presence, and 100,000s of miles of fiber optic cable.



The network connects 20 regions, 61 zones with 3 more coming

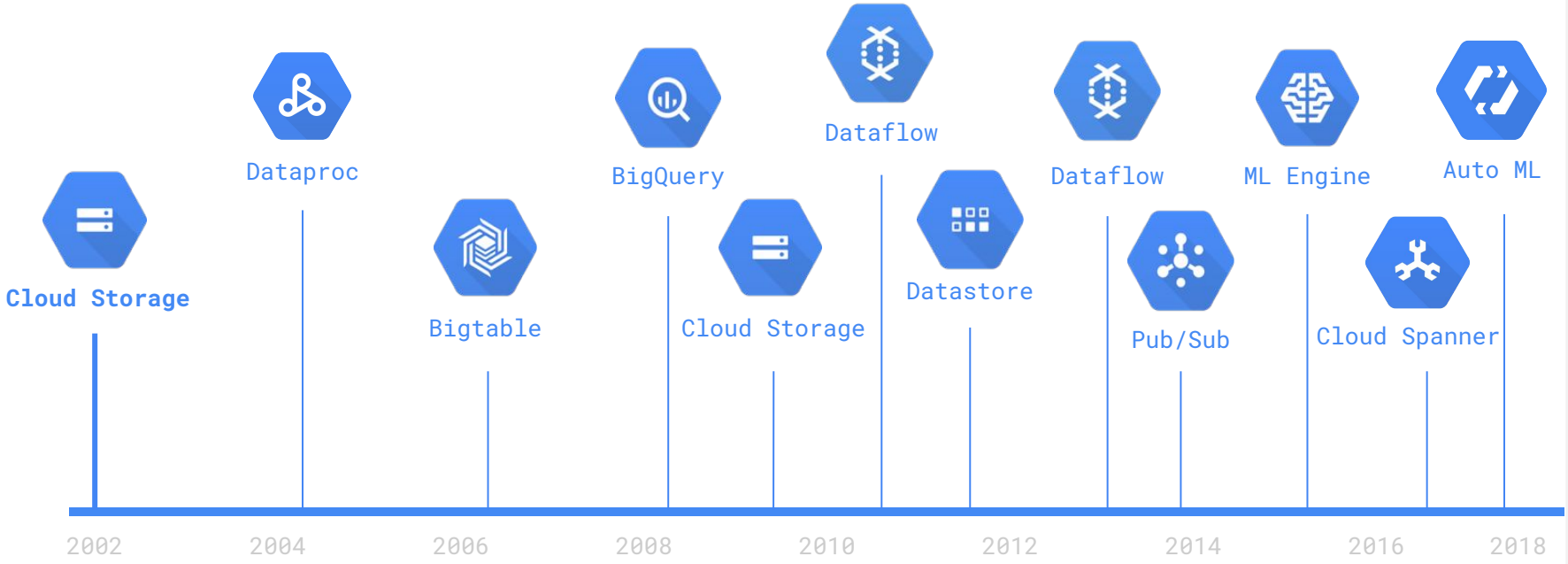


In terms of software, organizing the world's information has meant that Google needed to invent data processing methods



http://research.google.com/pubs/papers.html

Google Cloud opens up that innovation and infrastructure to you



Agenda

What is Google Cloud Platform

Google Cloud Big Data products

A suite of products that can be put together for data processing

Foundation

- Compute Engine
- Cloud Storage

Databases

- Cloud Spanner
- Cloud SQL
- Cloud Bigtable

Analytics and ML

- BigQuery
- Cloud Datalab
- ML APIs

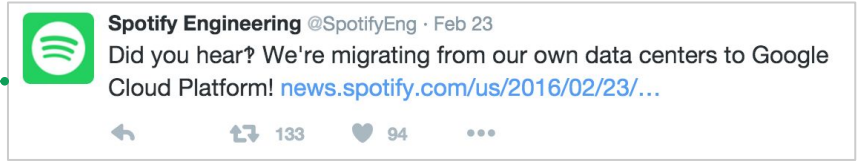
Data-handling frameworks

- Cloud Pub/Sub
- Cloud Dataflow
- Cloud Dataproc

Spotify illustrates the typical journey of companies that come to Google Cloud: From lower costs to increased reliability to business transformation

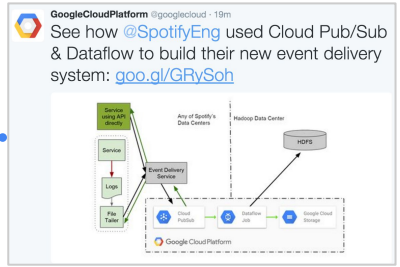
Spend less
No-ops,
Pay for use,
Secure

1



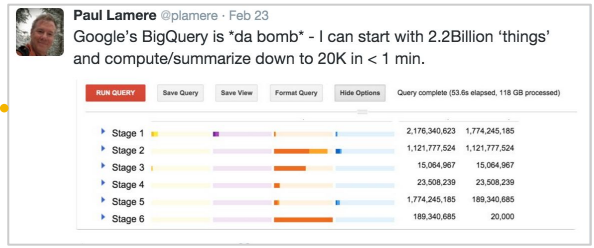
Flexible
Complete

2



Innovative
Powerful

3



A suite of products that can be put together for data processing

Change **where** you compute

Improve scalability and reliability

Change **how** you compute

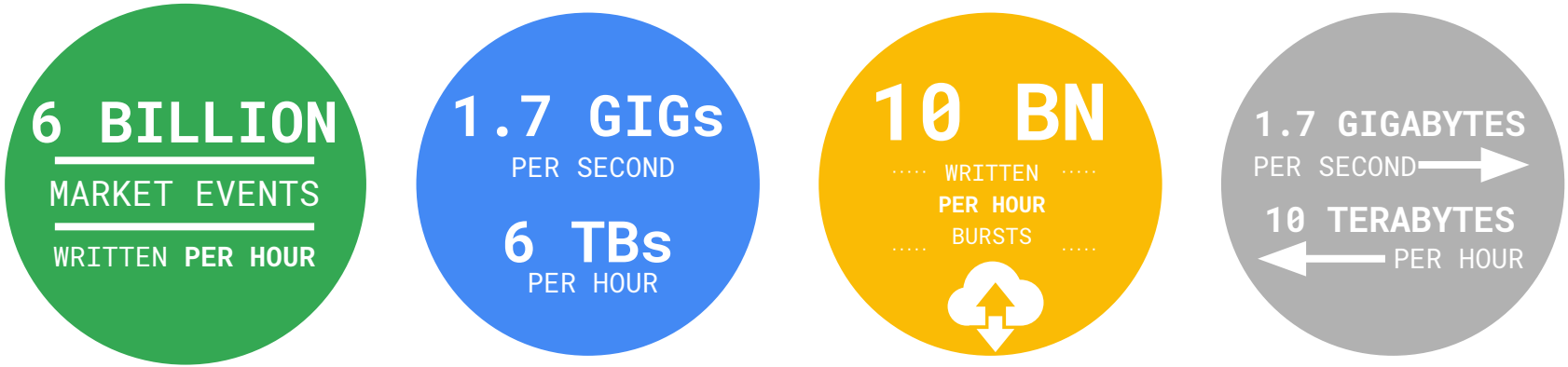
Atomic Fiction lowered their costs with per-minute (now per-second) billing

Change **where** you compute

https://www.youtube.com/watch?v=mBY-RjE15WA

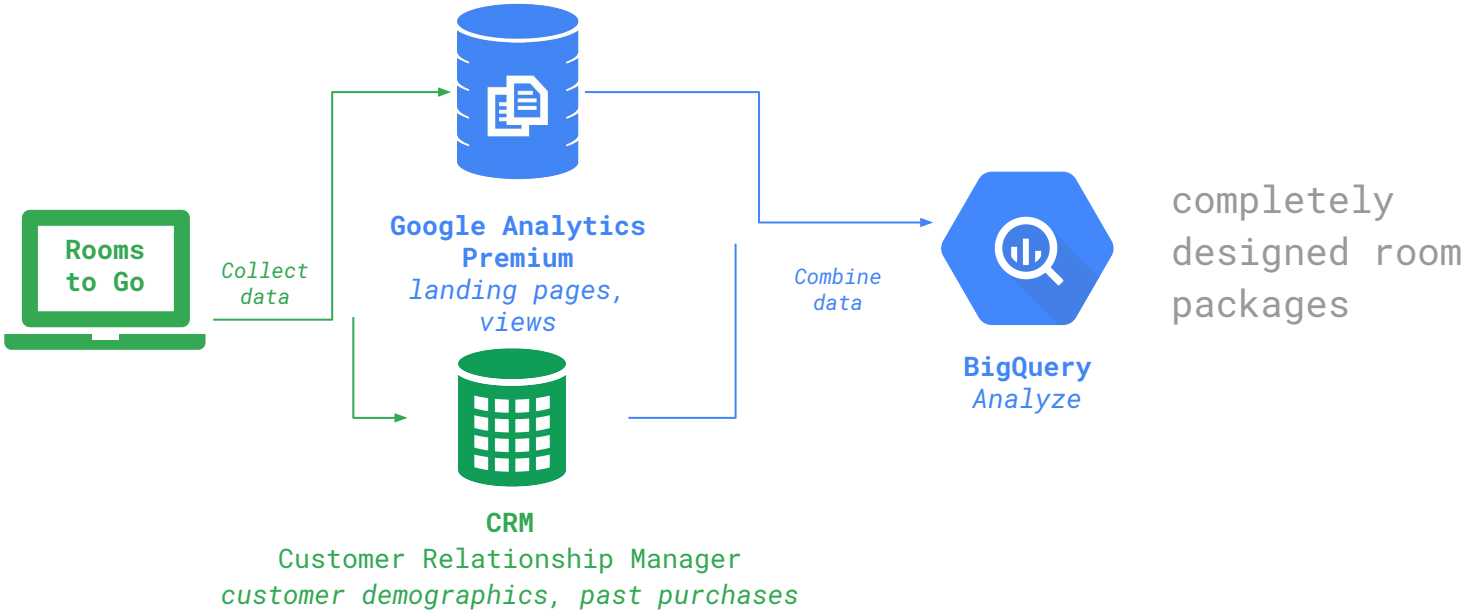
FIS was able to improve reliability and scalability on a massive data-processing challenge

FIS



The Consolidated Audit Trail (CAT) is a data repository of all equities and options orders, quotes, and events; FIS processed the CAT to organize 100 billion market events into an “order lifecycle” in a 4-hour window using Cloud Bigtable.

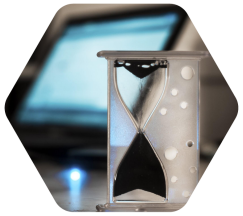
Rooms to Go transformed its business with data and machine learning



https://www.thinkwithgoogle.com/case-studies/rooms-to-go-improves-the-shopper-experience.html

In summary, Google Cloud offers you ways to…

Spend less on ops and administration



Incorporate real-time data into apps and architectures

Apply machine learning broadly and easily

Become a truly data-driven company

Module Review

Module review

Google Cloud Platform is:
*(select **all** of the correct options)*

[A] Operated by Google on the same infrastructure it uses

[B] Most cost-effective if you pre-purchase instances on a yearly basis

[C] A set of modular services from which you can compose cloud-based applications

[D] A platform on which to host scalable and fast distributed applications

Module review

Google Cloud Platform is:
*(select **all** of the correct options)*

[A] Operated by Google on the same infrastructure it uses

[B] Most cost-effective if you pre-purchase instances on a yearly basis

[C] A set of modular services from which you can compose cloud-based applications

[D] A platform on which to host scalable and fast distributed applications

Resources

Google Cloud Platform	https://cloud.google.com/
Data Centers	https://www.google.com/about/datacenters/
Google IT security	https://cloud.google.com/security/overview/whitepaper
Why Google Cloud Platform?	https://cloud.google.com/why-google-cloud/
Pricing Philosophy	https://cloud.google.com/pricing/principles

<Start Training>

#GoogleCloudOnBoard

</Start Training>

Cloud OnBoard

Fundamentals of Google Cloud Platform

Cloud OnBoard

Agenda

CPUs on demand + Demo

A global filesystem + Demo

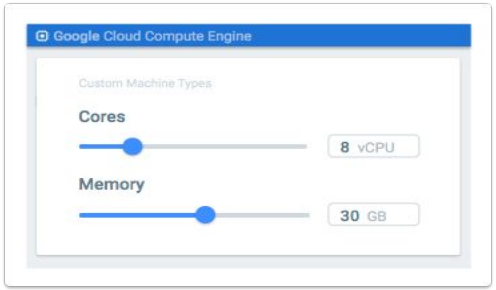
Google Cloud provides an earth-scale computer

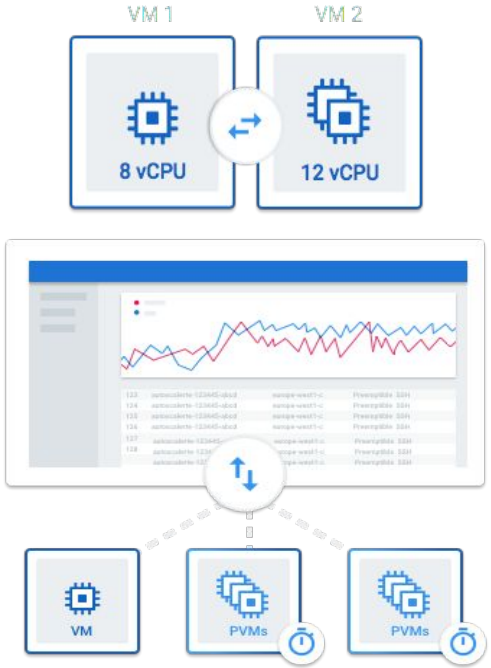
Networking

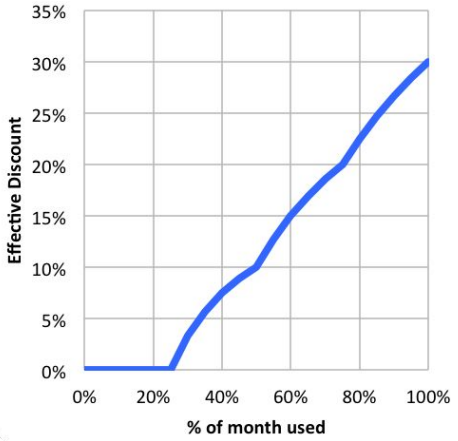
Data storage

Compute power

Custom/changeable machine types, preemptible machines, and automatic discounts lead to simplicity and agility







Demo:

Create a Compute Engine instance

Demo : **Create a Compute Engine Instance**

In this demo, we will :

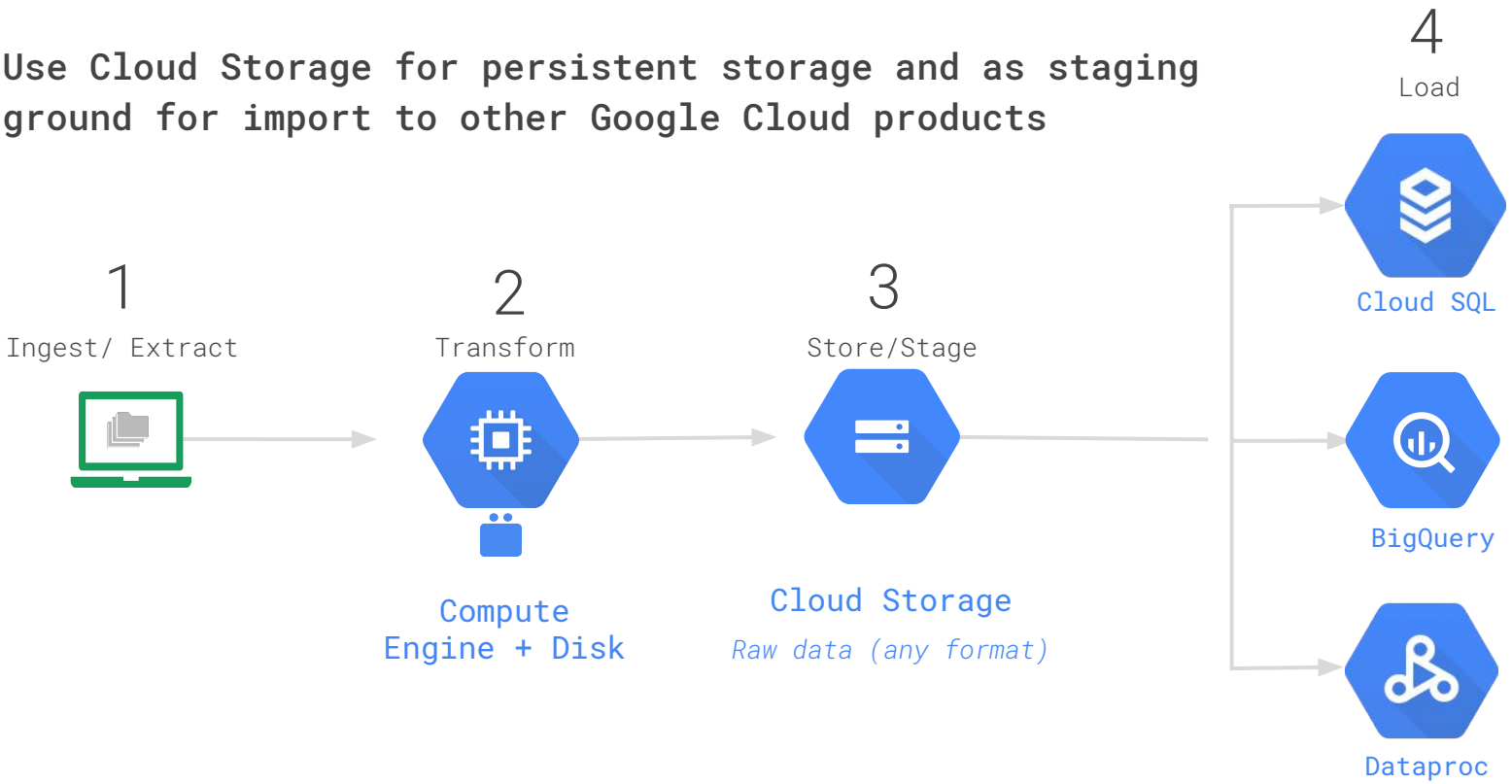
1. Create a Compute Engine instance
2. SSH into the instance
3. Install the software package git (for source code version control)

Agenda

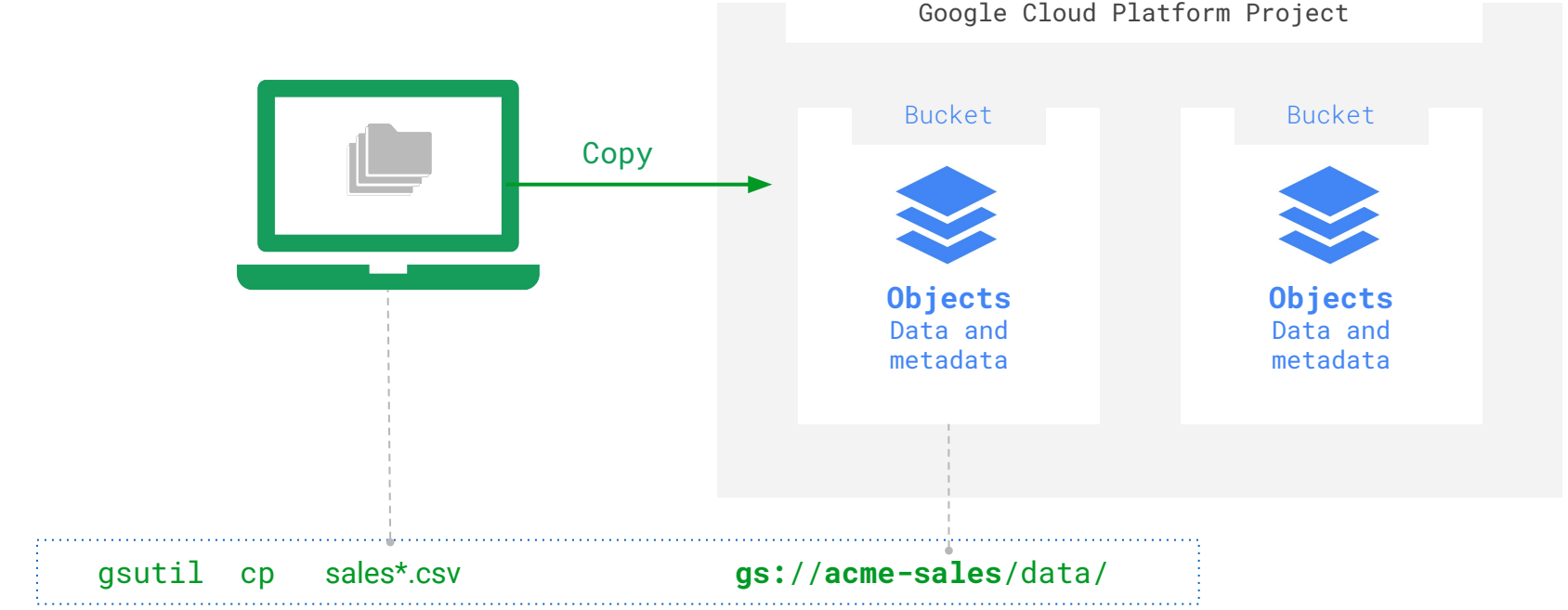
CPUs on demand + Demo

A global filesystem + Demo

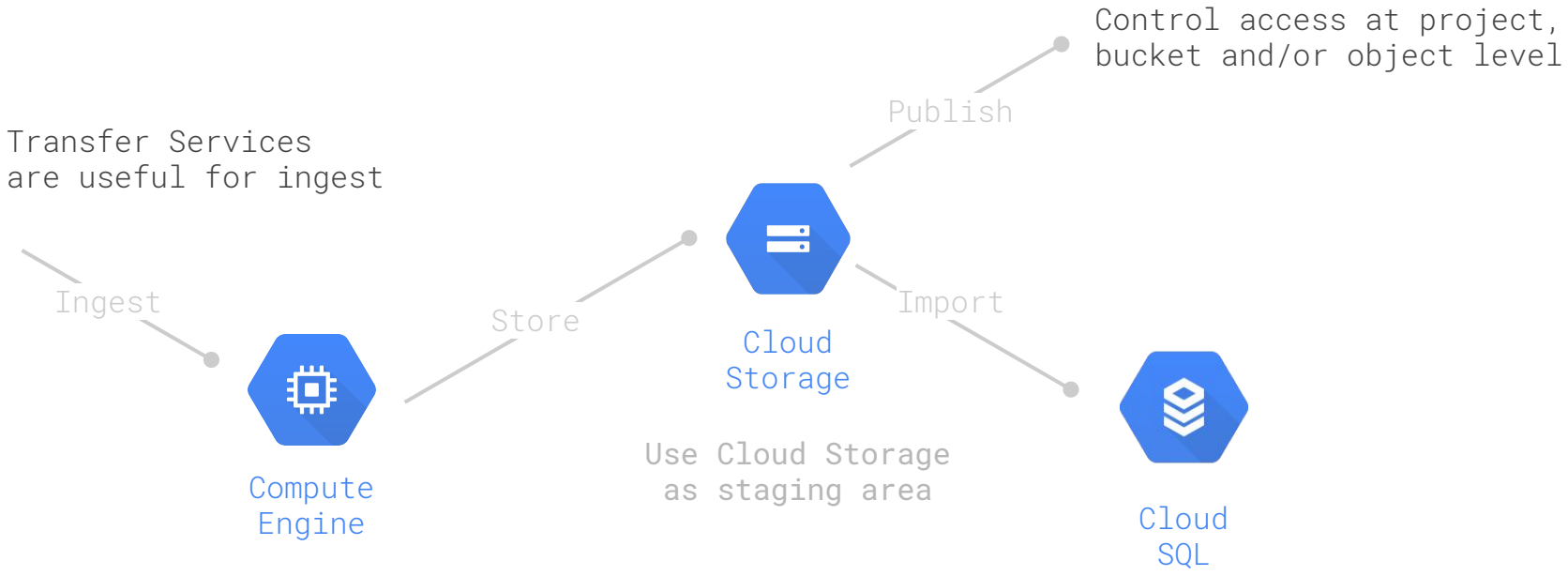
Use Cloud Storage for persistent storage and as staging ground for import to other Google Cloud products



Create a bucket and copy the data over using the Cloud SDK; blobs are referenced through a gs://.../ URL



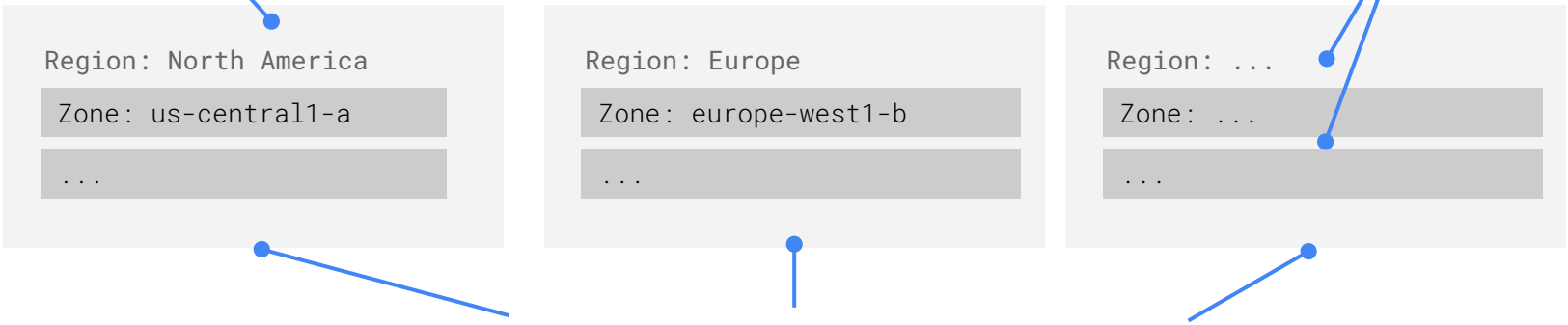
Cloud Storage gives you durability, reliability, and global reach



Control latency and availability with zones and regions

Choose the closest zone/region so as to to reduce latency.

Distribute your apps and data across zones to reduce service disruptions.



Distribute your apps and data across regions for global availability.

Demo:

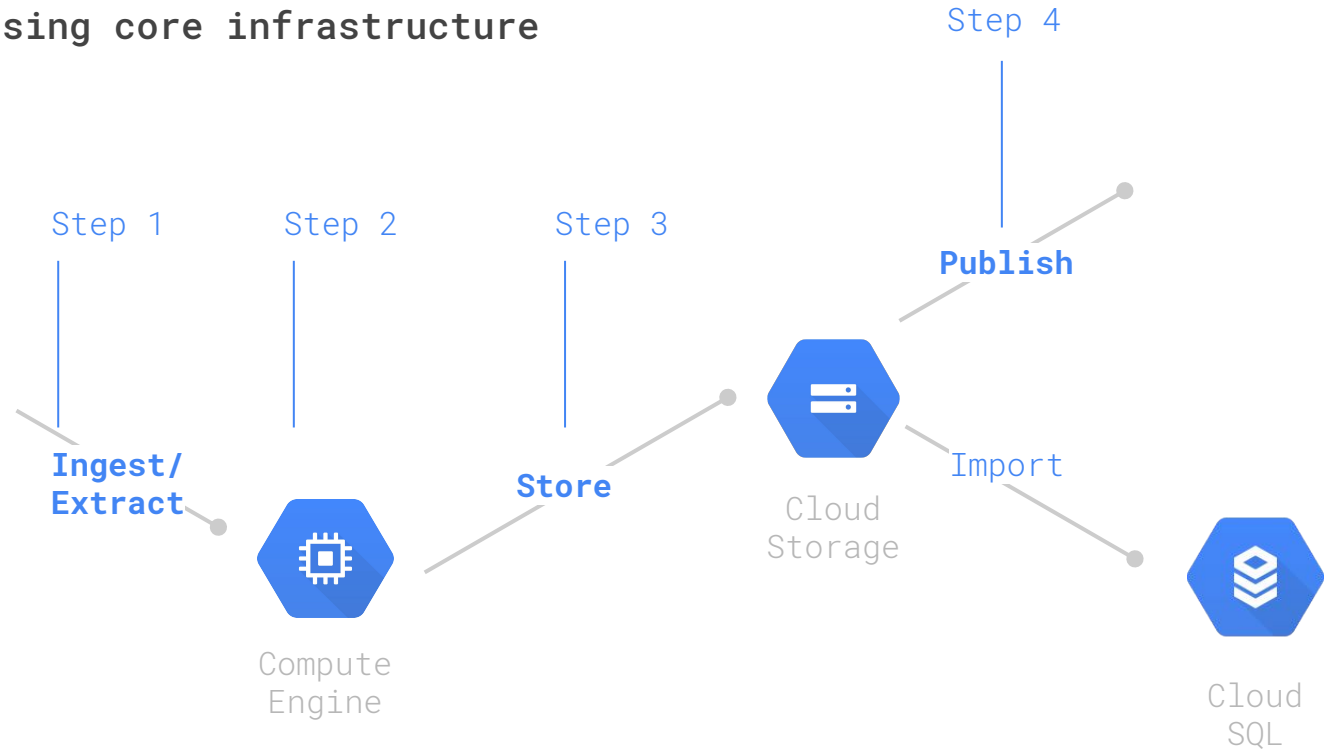
Interact with Cloud Storage

Demo : **Interact with Cloud Storage**

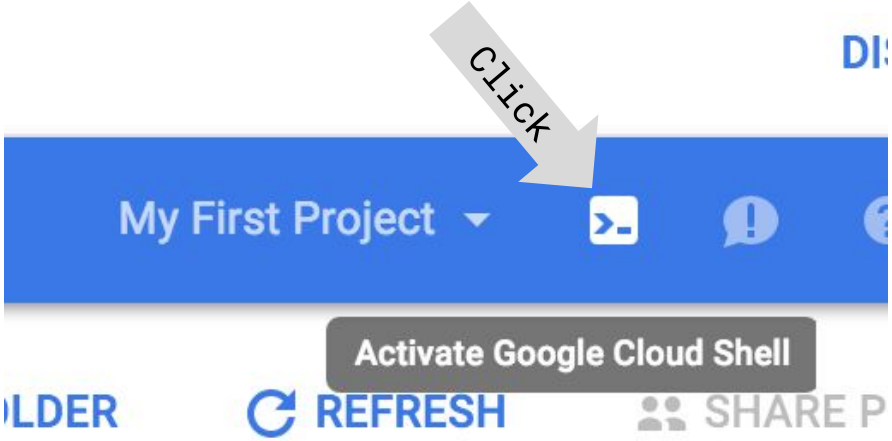
In this demo, we carry out the steps of an ingest-transform-and-publish data pipeline manually

1. Ingest data into a Compute Engine instance
2. Transform data on the Compute Engine instance
3. Store the transformed data on Cloud Storage
4. Publish Cloud Storage data to the web

Ingest-Transform-Publish using core infrastructure



Cloud Shell gives you an easy command-line



Do Now

Cloud Shell comes pre-installed with the tools, libraries, and so on you need to interact with Google Cloud Platform

Google Cloud Customers

Module Review

Module review (1 of 2)

Compute nodes on GCP are:

(select the correct option)

- ❑ Allocated on demand, and you pay for the time that they are up.
- ❑ Expensive to create and teardown
- ❑ Pre-installed with all the software packages you might ever need.
- ❑ One of ~50 choices in terms of CPU and memory

Module review answers (1 of 2)

Compute nodes on GCP are:

(select the correct option)

➔ Allocated on demand, and you pay for the time that they are up.

❑ Expensive to create and teardown

❑ Pre-installed with all the software packages you might ever need.

❑ One of ~50 choices in terms of CPU and memory

Module review (2 of 2)

Google Cloud Storage is a good option for storing data that:

(select all of the correct options)

- ❑ Is ingested in real-time from sensors and other devices
- ❑ Will be frequently read/written from a compute node
- ❑ May be required to be read at some later time
- ❑ May be imported into a cluster for analysis

Module review (2 of 2)

Google Cloud Storage is a good option for storing data that:

(select all of the correct options)

- ❑ Is ingested in real-time from sensors and other devices
- ❑ Will be frequently read/written from a compute node
- ➔ May be required to be read at some later time
- ➔ May be imported into a cluster for analysis

Resources

Google Cloud Platform	https://cloud.google.com/compute/
Datacenters	https://cloud.google.com/storage/
Pricing	https://cloud.google.com/pricing/
Cloud Launcher	https://cloud.google.com/launcher/
Pricing Philosophy	https://cloud.google.com/pricing/principles

Speaker

[Wayne An]

[客戶工程師,Google Cloud]

Cloud OnBoard

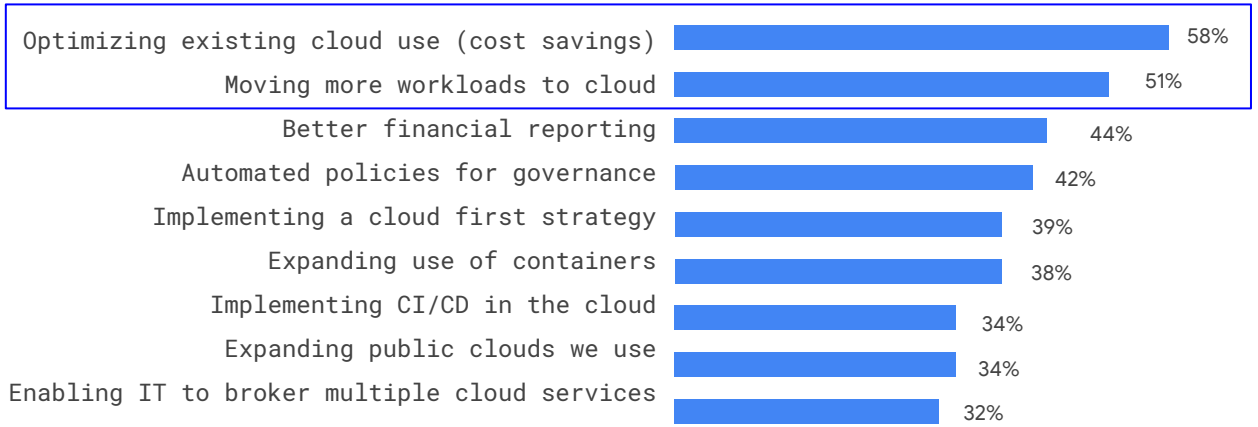
Migrate to Google Cloud in 5 mins

Athos Migrate

Cloud OnBoard

The big migration: enterprises shifting to cloud

Cloud initiatives in 2018



Source: Rightscale 2018 State of the Cloud Report

"The fastest-growing segment of the market is cloud system infrastructure services (infrastructure as a service or IaaS), which is forecast to grow 35.9 percent in 2018 to reach $40.8 billion"

Source: Gartner, https://www.gartner.com/newsroom/id/3871416

Key challenges of enterprise migration

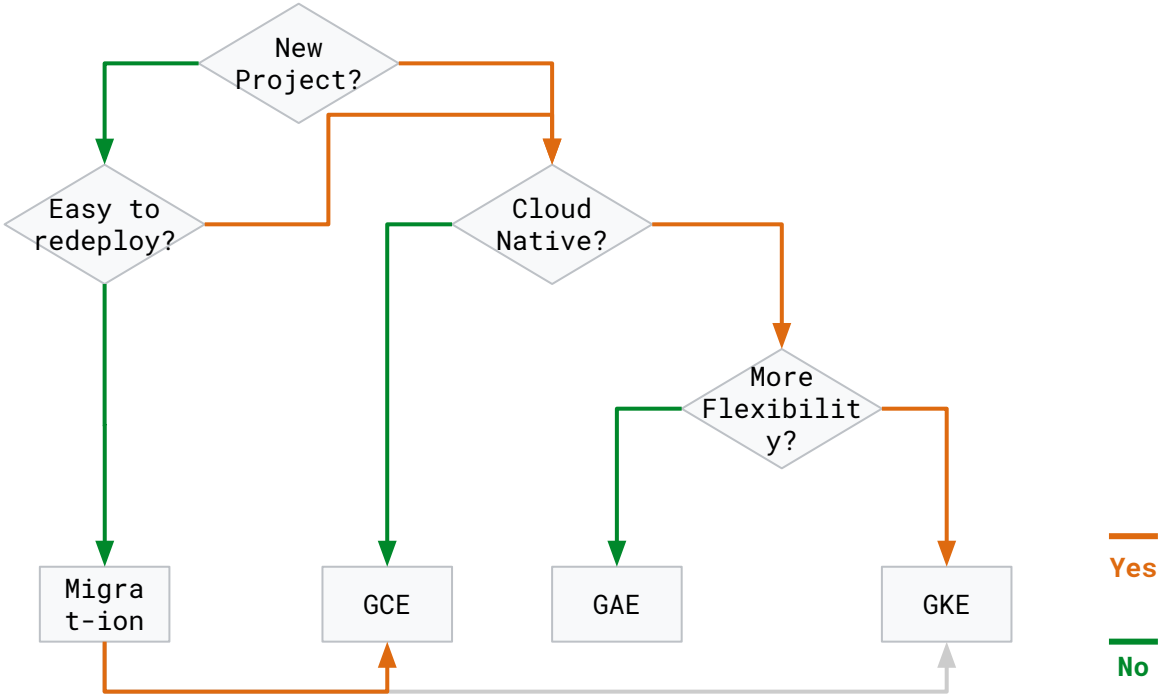
94%

of enterprise migration projects face delays, over budget

2017 Cloud Migration Survey, Dimensional Research

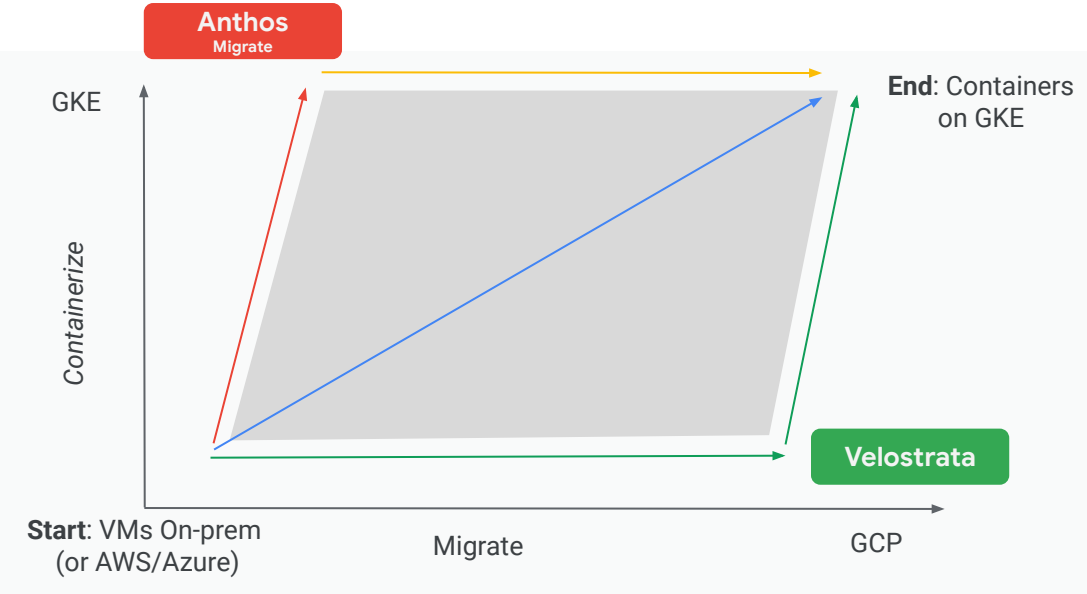
Scale	Data Center Migration: Thousands of workloads
Multi-Source	Migrate on-prem VMs, physical, other clouds
Complexity	Multi-tier apps incl. SAP, w minimized downtime
Risk	Revert to on-prem if needed (cost, performance)

Cloud migration considerations



Vision: Migrate and Modernize VMs!

Offers the option to choose when/if to modernize



"Container adoption **likely to grow** the highest in pubic cloud. Recommendation for **container orchestration**: Make it extremely easy to **accelerate migration to cloud**"

McKinsey container adoption survey 2018

Google Cloud

Solution: Velostrata enterprise cloud migration

Purpose-built, Enterprise-grade, field proven:

- Agentless Fast Switchover
- Cross-Cloud Migration
- Testing to Migration Automation

Free for migrations to GCP

Google Cloud

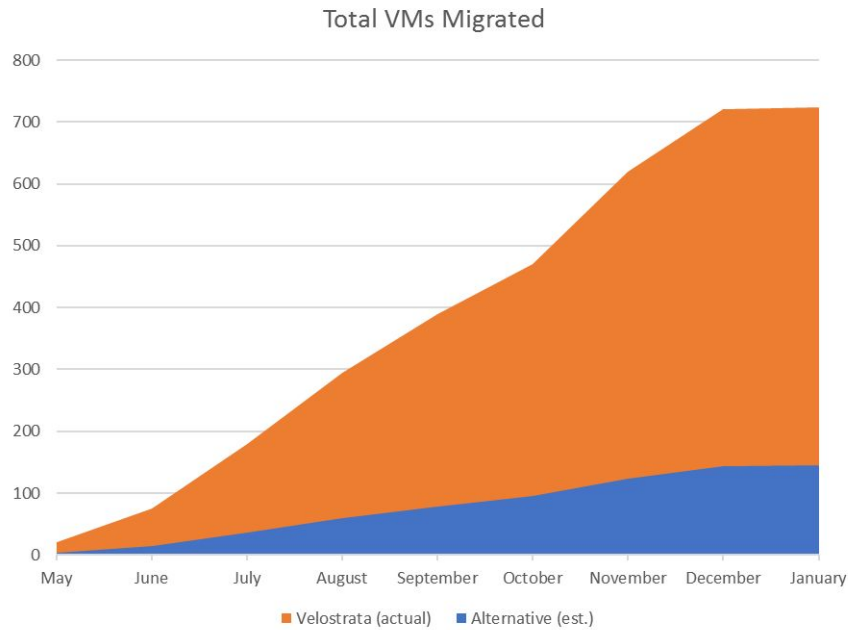
The faster we migrate, the faster everyone benefits

Example: Large energy company

- 700 VMs -- Enterprise apps
  - 140 TB of data
  - 100-200 Mbps WAN link
- Completed migration in 6 mo.
- 5X faster than projections w/ other alternatives
- Cost Savings, over $1M est.

Faster migration =

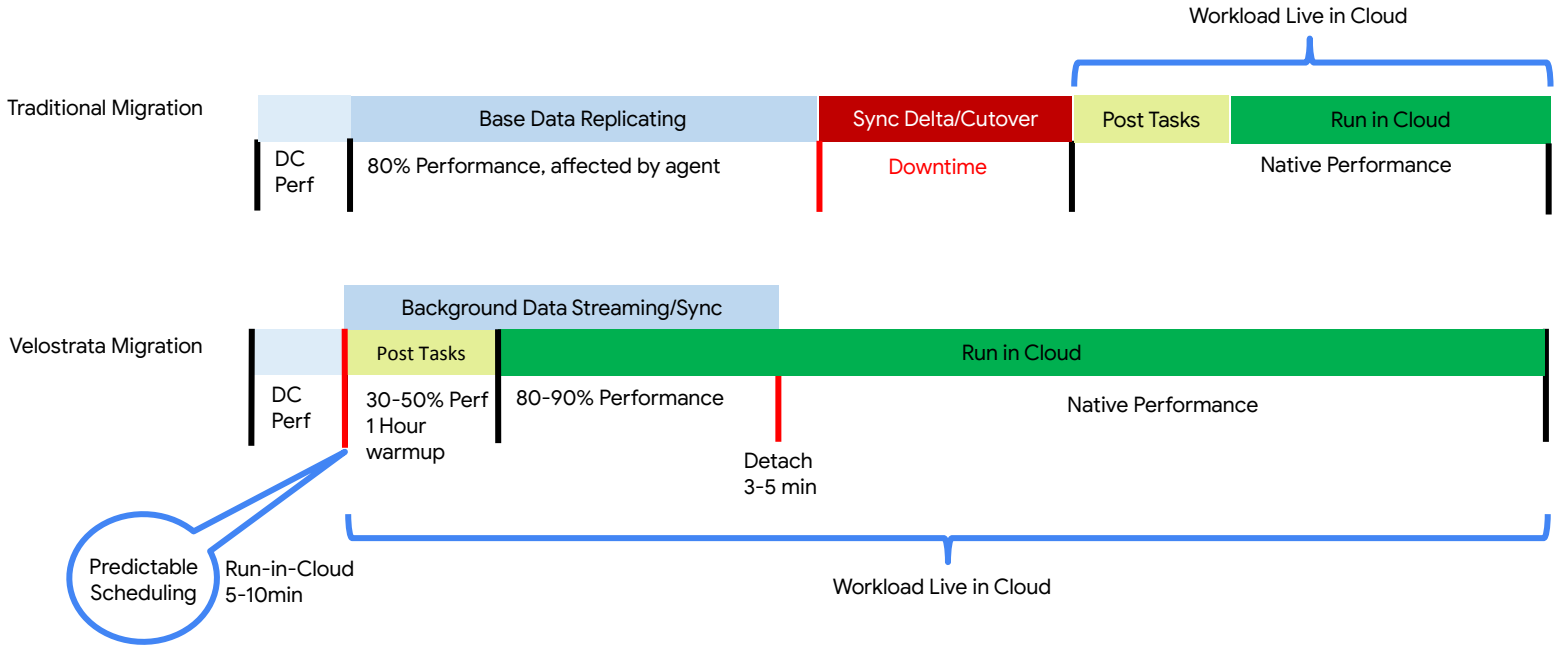
Faster agility, lower cost of migration.



Google Cloud

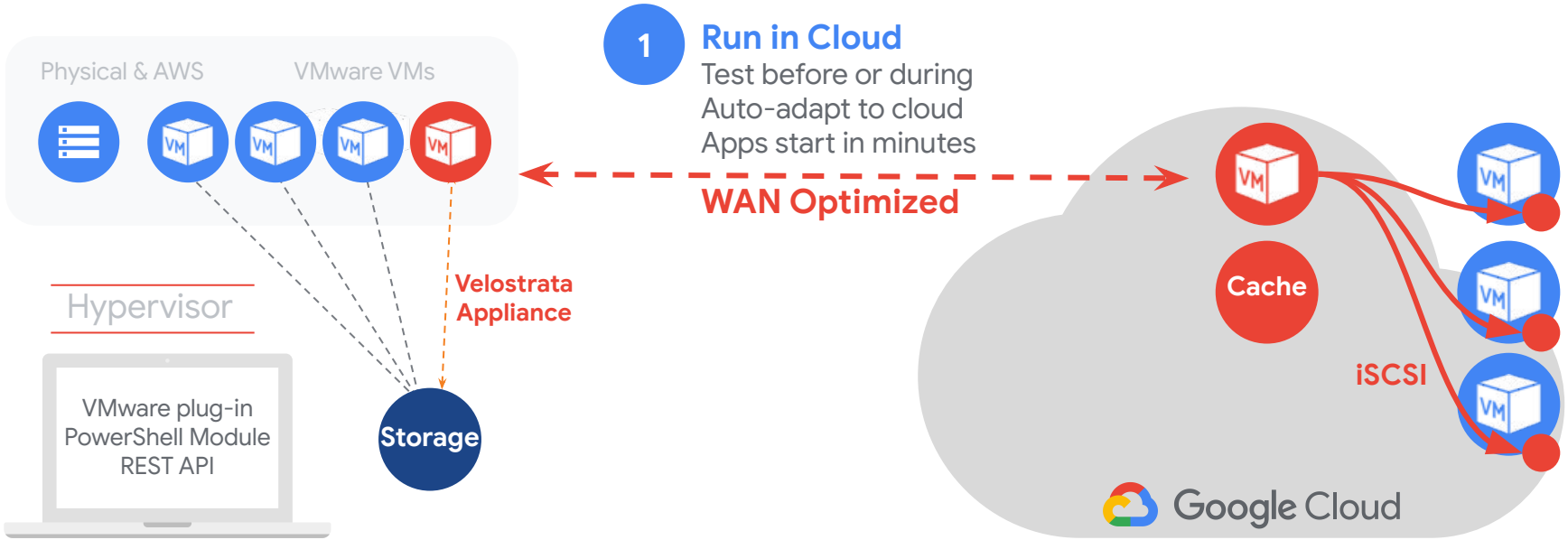
Architecture

Velostrata: Migrate VM to Cloud within Mins (Patented Technology)

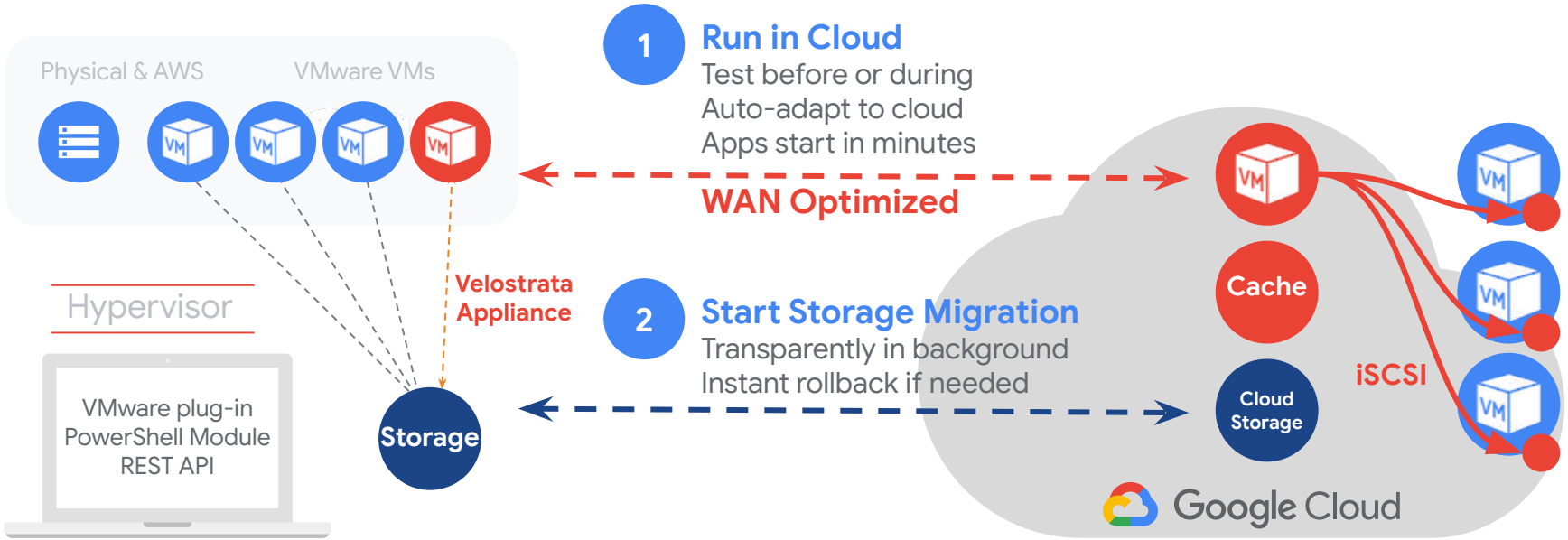


Google Cloud

Real-time agentless streaming technology

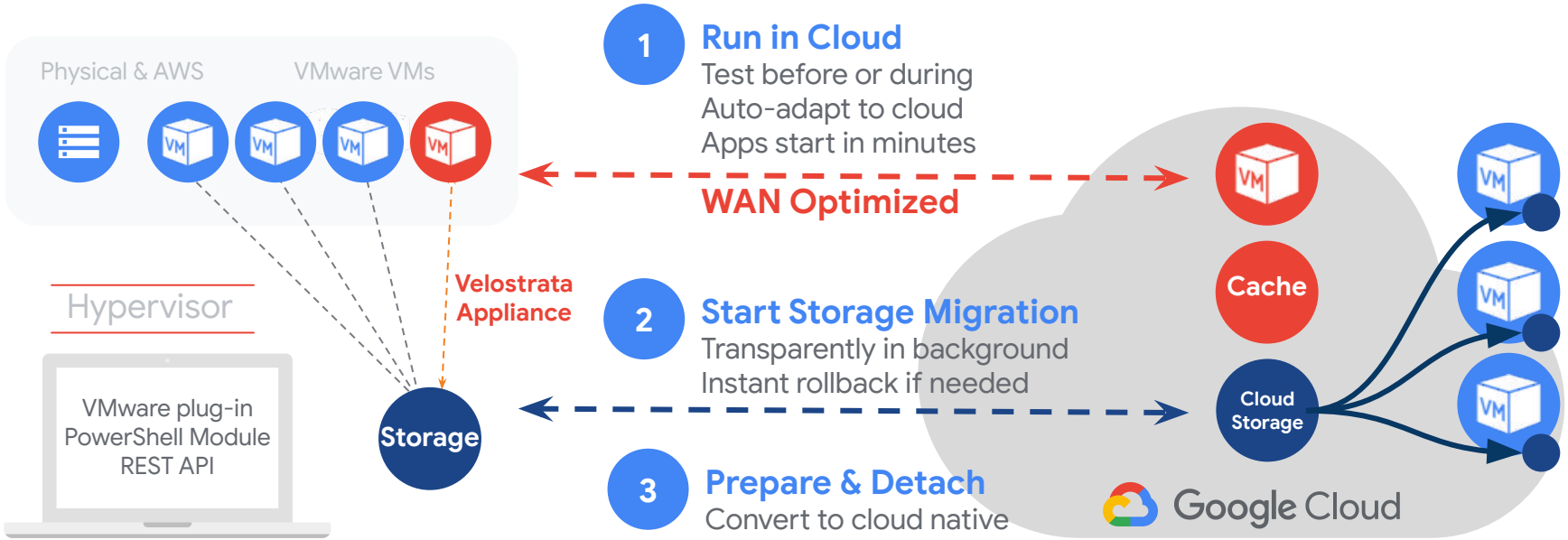


Real-time agentless streaming technology



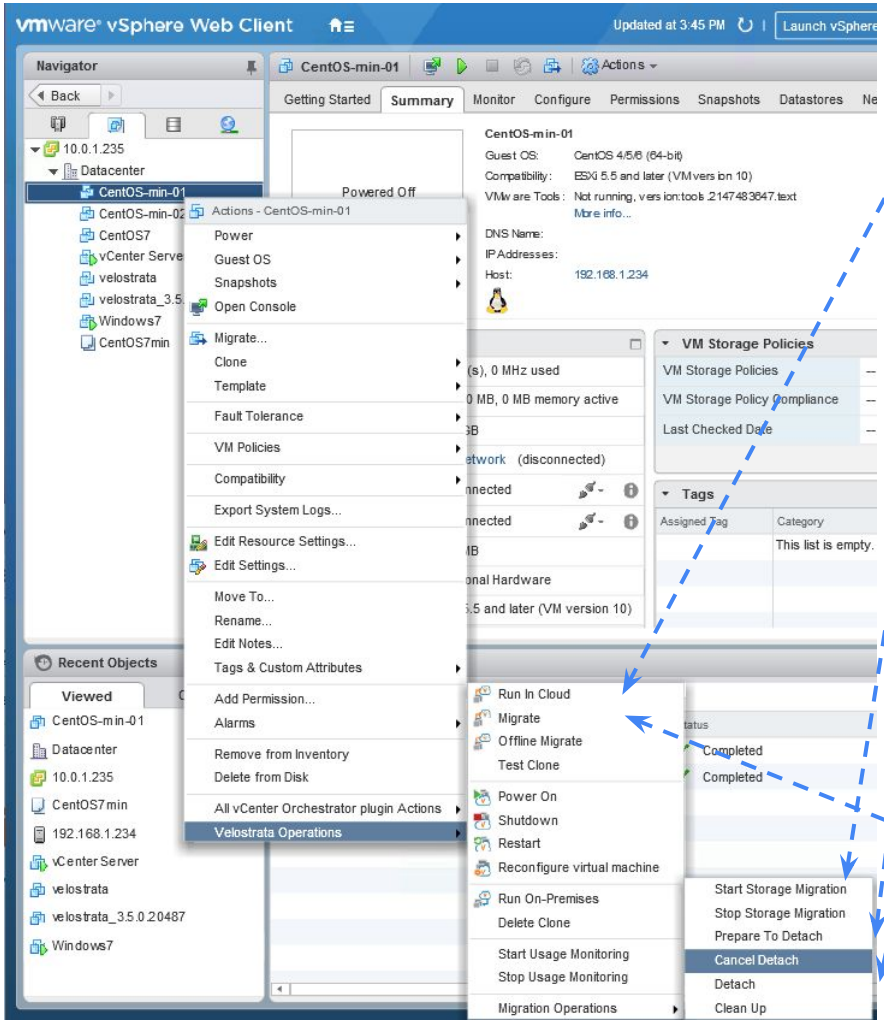
Google Cloud

Real-time agentless streaming technology



Google Cloud

Features

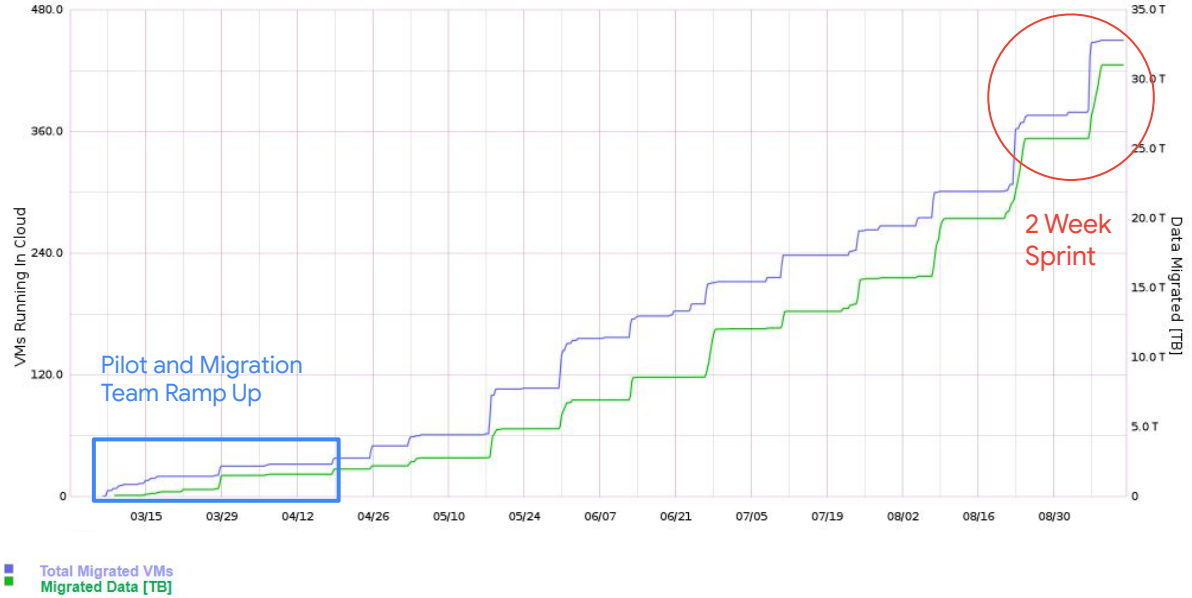


- Run in Cloud
  - Power off on-prem VM
  - Create VM on Cloud
  - Power on VM on Cloud and use iSCSI to access streaming data from on-prem
- Start Storage Migration
  - Copy data from on-prem to GCS
- Preparing to Detach
  - exporter VM/instance is created
  - Native cloud drives are created as the source VM
  - The exporter reads all the VM data from the Cloud Storage and writes to PD, continues to read and write the changes committed until the detach operation
- Detach
  - VM will shut down
  - exporter perform last synchronization
  - adjusting VM size
  - attach the native disks to the instance
  - re-start the instance
- The Migrate operation performs three subtasks:
  - Run in the cloud
  - Storage Migration
  - Prepare to detach

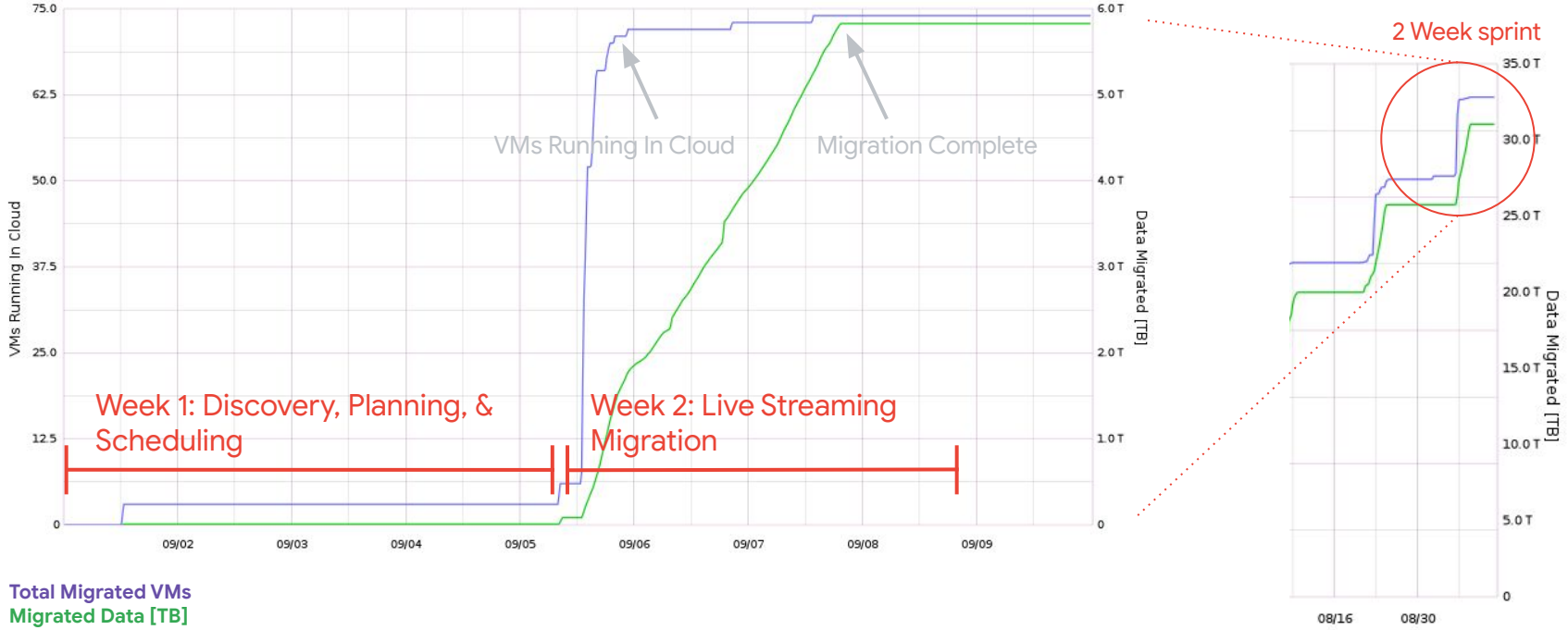
Real World Example

Scale validation - real world example

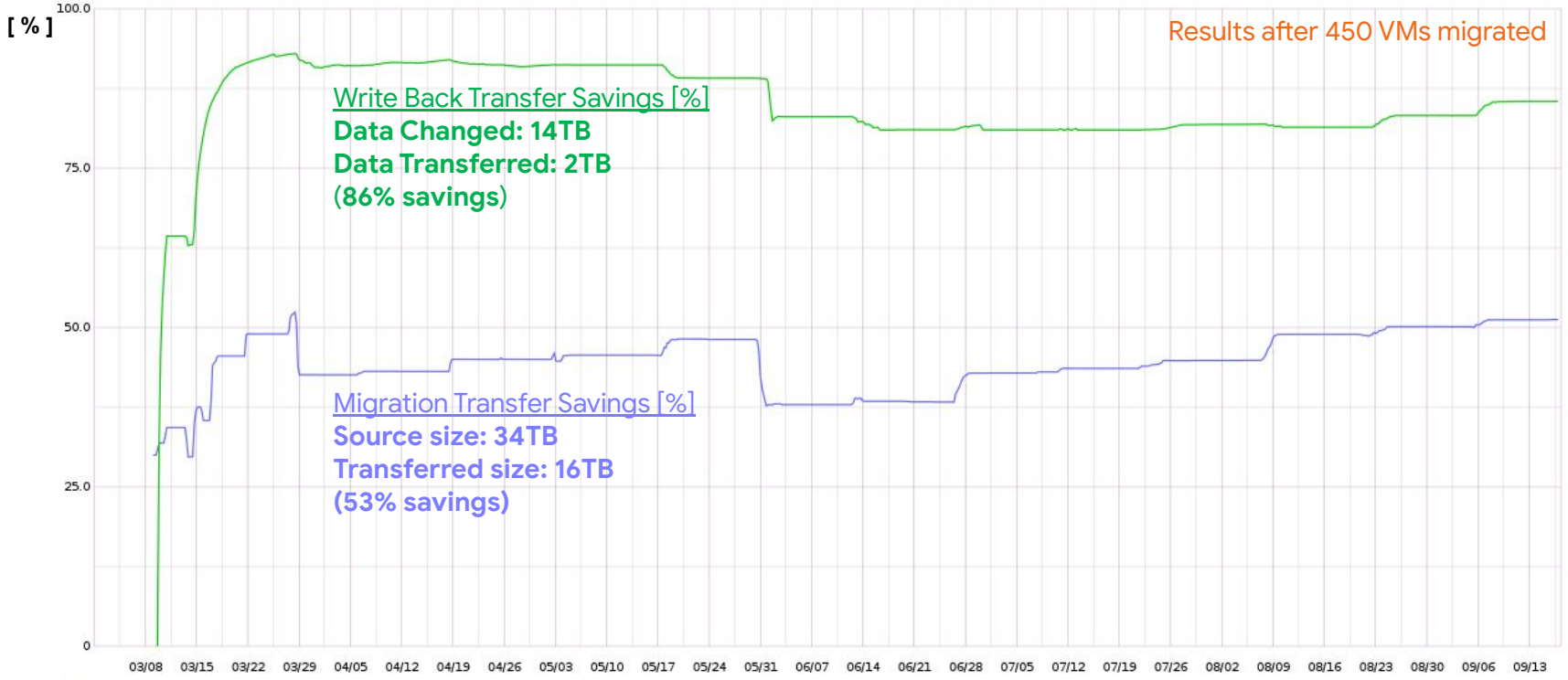
- 2-week sprint size grows to 40-100 VMs
- Up from 6-12 during PoC
- 8-10x improvement = 8-10x faster completion



Scale validation - real world example



WAN Optimization - real world example



Demo

Anthos

Google's Velostrata is now Anthos Migrate

Introducing Anthos, the latest innovation in hybrid and multi-cloud computing that allows you to modernize applications, automate policy, provide security at scale, and can run on premises, or in the cloud.

Anthos

Anthos

Transform your IT and build apps for today and the future

Google Kubernetes Engine

Eliminate the need to install, manage, and operate your own Kubernetes clusters in the cloud.

LEARN MORE →

GKE On-Prem

Run Kubernetes clusters reliably, efficiently, and securely — anywhere.

LEARN MORE →

Istio on GKE

Enhance your application observability, security, and control with Istio-powered service mesh with single-click deployment from Google Cloud Console.

LEARN MORE →

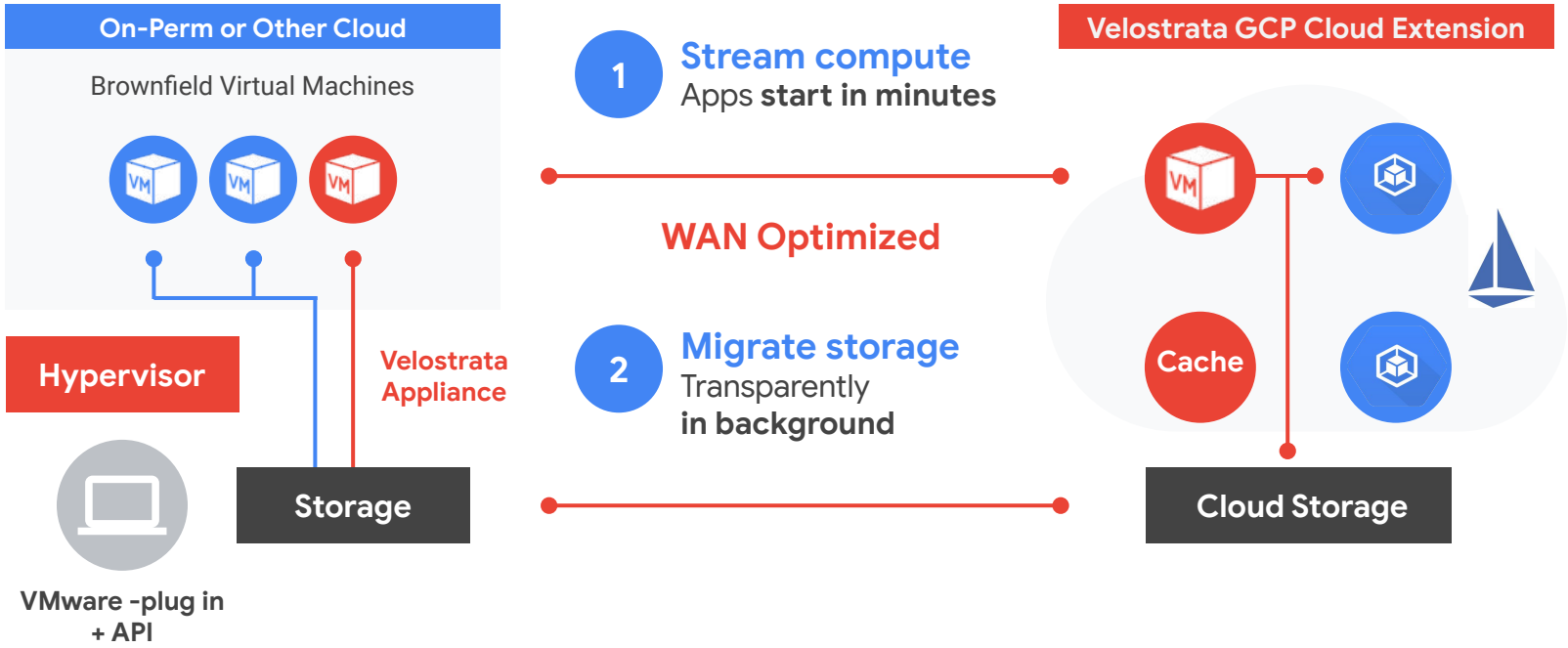
Anthos Config Management

Automate policy and security at scale for your hybrid Kubernetes deployments.

LEARN MORE →

Google Cloud

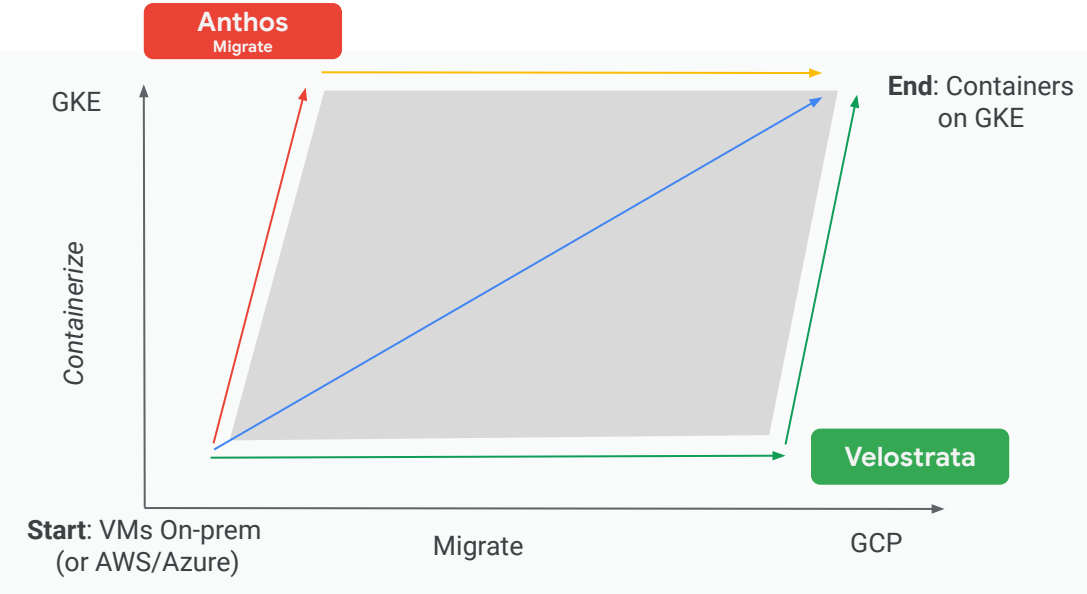
Migration to GKE enabled by streaming!



Google Cloud

Vision: Migrate and Modernize VMs!

Offers the option to choose when/if to modernize



"Container adoption **likely to grow** the highest in pubic cloud. Recommendation for **container orchestration**: Make it extremely easy to **accelerate migration to cloud**"

McKinsey container adoption survey 2018

Google Cloud

<Start Training>

#GoogleCloudOnBoard

</Start Training>

Cloud OnBoard

Data Analysis on the Cloud

Cloud OnBoard

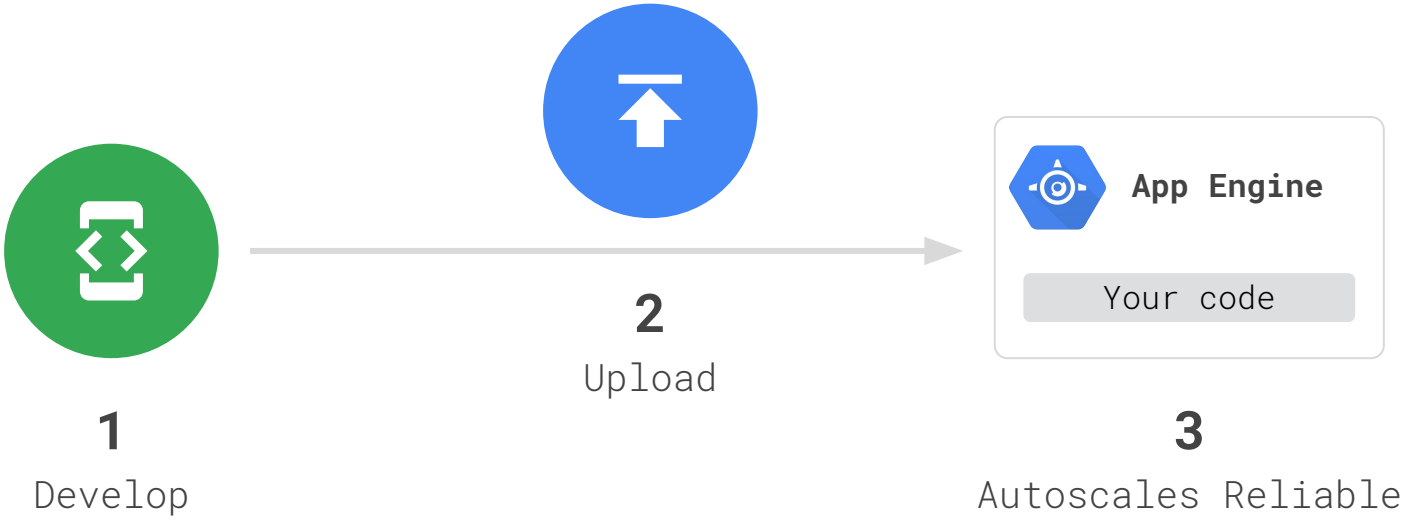
Agenda

Stepping stones to transformation

Your SQL database in the cloud + Demo

Managed Hadoop in the cloud + Demo

Google Cloud Platform began in 2008, with App Engine, a serverless way to run web applications

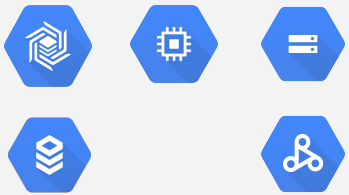


http://googleappengine.blogspot.com/2008/04/introducing-google-app-engine-our-new.html
http://googleappengine.blogspot.com/2013/05/the-google-app-engine-blog-is-moving.html


App Engine
App Engine Flex
Container Engine
Compute Engine
There [was] something fundamentally wrong with what we were doing in 2008 … We didn't get the right stepping stones into the cloud …
-- Eric Schmidt, Executive Chairman, Google

GCP now consists of a suite of products that together provide these stepping stones in a business' transformative journey

Change **where** you compute



Cost effective virtual machines, storage, Hadoop, and MySQL to migrate your current workloads to the public cloud.

Flexibility, scalability and reliability

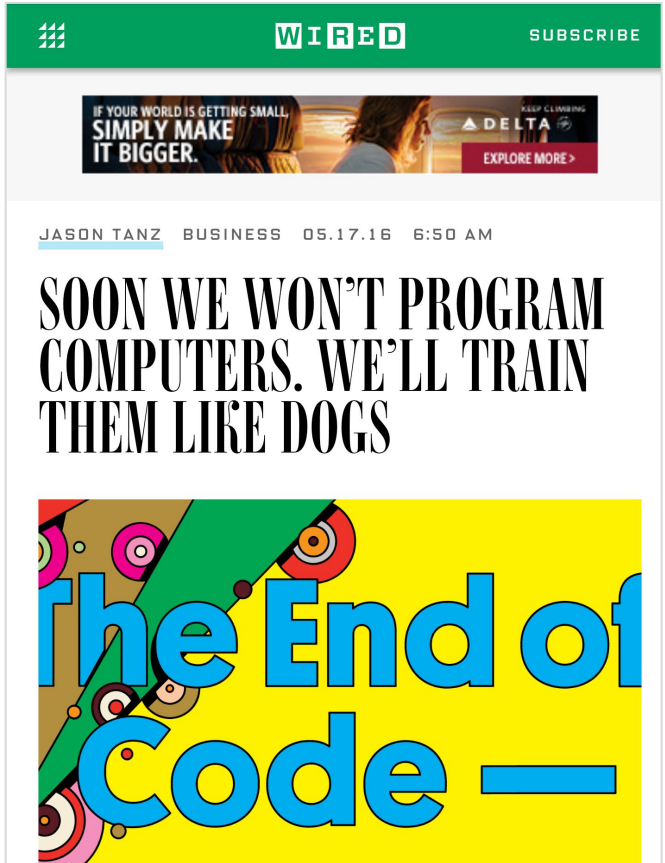
Reliable, autoscaling messaging, data processing, and storage.

Change **how** you compute

Fully managed products for data warehousing, data analysis, streaming, and machine learning.

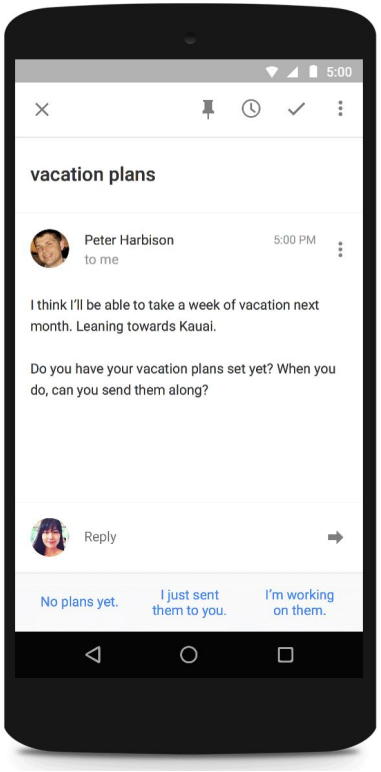
Machine learning. This is the next transformation … the programming paradigm is changing. Instead of programming a computer, you teach a computer to learn something and it does what you want.

Eric Schmidt,
Executive Chairman,
Google



WIRED's headline

"If you want to teach a neural network to recognize a cat, for instance, you don't tell it to look for whiskers, ears, fur, and eyes. You simply show it thousands and thousands of photos of cats, and eventually it works things out."



Machine Learning is not new, but it is now mainstream

Search
People who bought ...
Spam filtering
Suggest next video
Route planning
Smart Reply

?

What's common to all of these use cases of Machine Learning?

There are three components in a recommendation system

Rating	Training	Recommending
Users rate a few houses explicitly or implicitly	A machine learning model is created to predict a user's rating of a house	For each user, the model is applied to every unrated house and the top 5 houses for that user are saved.

? What else is needed?

The ML algorithm essentially clusters users and items

1

Who is like this user?

2

Is this a good house?

3

Predict rating

Is this house similar to houses that people similar to this user like?

Predicted rating = user-preference * item-quality

?

How often do you need to compute the predicted ratings?

Where would you save them?

In addition to the ML algorithm, you also need sophisticated data management

Data Collection	Scalable front end to collect customer actions
Data Analysis	Data that is accessible and not silo-ed
Machine Learning	(Re-)training and experimentation
Serving	Scalable, real-time system to serve recommendations

Agenda

Stepping stones to transformation

Your SQL database in the cloud + Demo

Managed Hadoop in the cloud + Demo

Choose your storage solution based on your access pattern

	Cloud Storage	Cloud SQL	Datastore	Bigtable	BigQuery
Capacity	Petabytes +	Gigabytes	Terabytes	Petabytes	Petabytes
Access metaphor	Like files in a file system	Relational database	Persistent Hashmap	Key-value(s), HBase API	Relational
Read	Have to copy to local disk	SELECT rows	Filter objects on property	scan rows	SELECT rows
Write	One file	INSERT row	put object	put row	Batch/stream
Update granularity	An object (a "file")	Field	Attribute	Row	Field
Usage	Store blobs	No-ops SQL database on the cloud	Structured data from AppEngine apps	No-ops, high throughput, scalable, flattened data	Interactive SQL* querying fully managed warehouse

Cloud SQL is a fully managed database service

Cloud SQL
Google-managed
MySQL or Postgres

- Flexible pricing
- Familiar
- Managed backups
- Automatic replication
- Fast connection from GCE & GAE
- Connect from anywhere
- Google Security

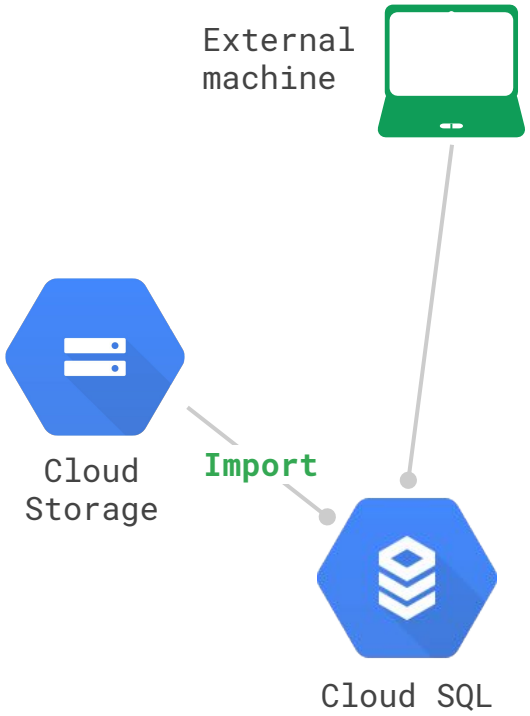
Demo:

Set up rentals data in Cloud SQL

Demo: Setup rentals data in Cloud SQL

In this demo, we populate rentals data in Cloud SQL for the recommendation engine to use:

1. Create Cloud SQL instance
2. Create database tables by importing .sql files from Cloud Storage
3. Populate the tables by importing .csv files from Cloud Storage
4. Allow access to Cloud SQL
5. Explore the rentals data using SQL statements from Cloud Shell



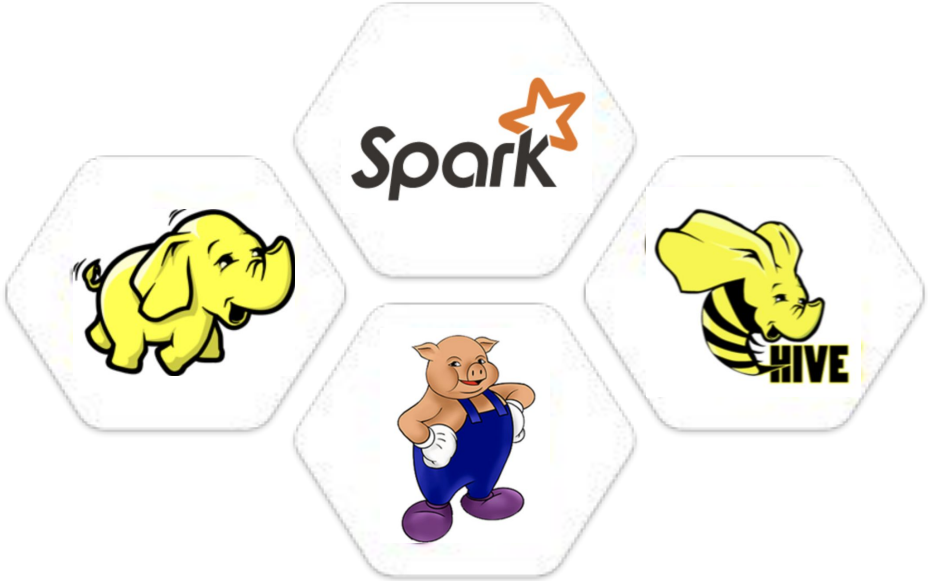
Agenda

Stepping stones to transformation

Your SQL database in the cloud + Demo

Managed Hadoop in the cloud + Demo

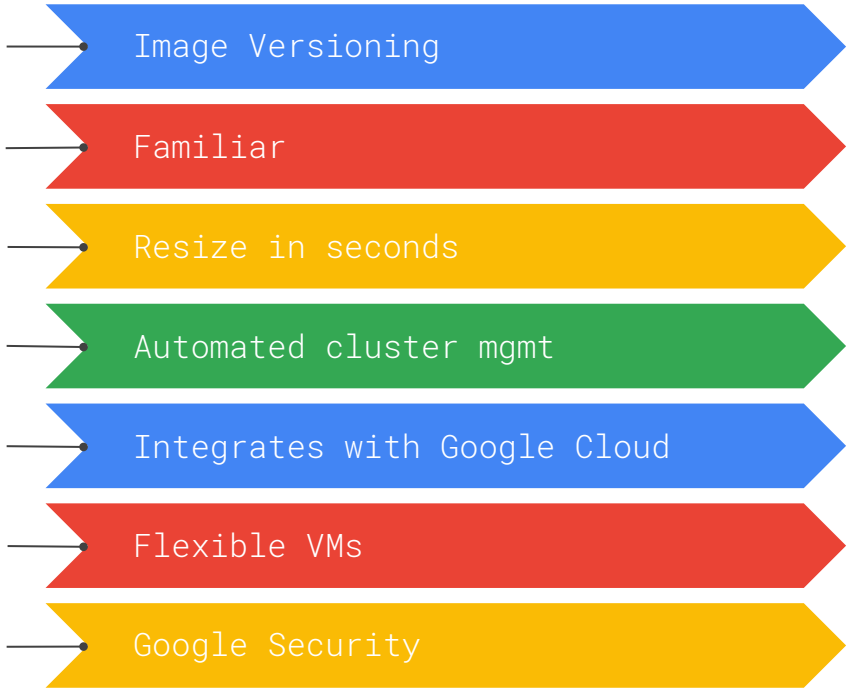
There is a rich open-source ecosystem for big data



http://hadoop.apache.org/
http://pig.apache.org/
http://hive.apache.org/
http://spark.apache.org/

Dataproc reduces the cost and complexity associated with Spark and Hadoop clusters

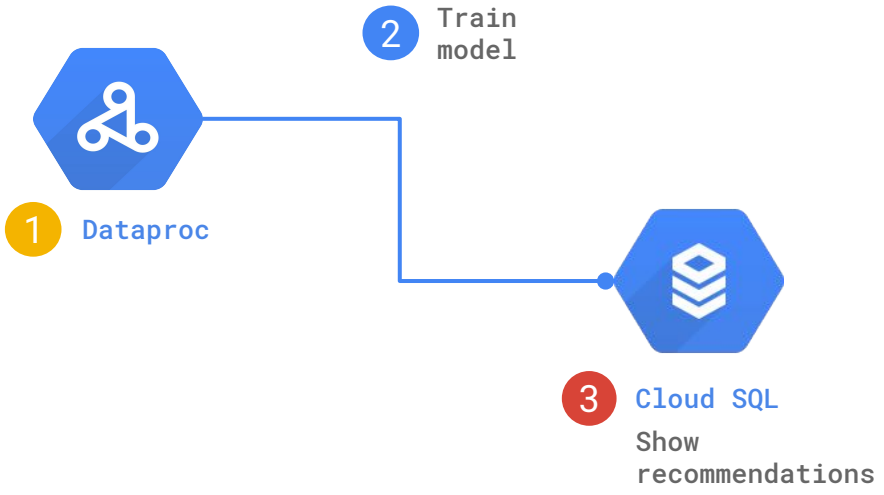
Dataproc
Google-managed:
Hadoop
Pig
Hive
Spark



Demo:

Recommendations ML with Dataproc

Demo: Recommendations ML with Cloud Dataproc



In this demo, we implement machine learning recommendations using Cloud Dataproc:

1. Launch Dataproc
2. Train and apply ML model written in PySpark to create product recommendations
3. Explore inserted rows in Cloud SQL

Module Review

Module review (1 of 2)

Relational databases are a good choice when you need:

(select all of the correct options)

- ❑ Streaming, high-throughput writes
- ❑ Fast queries on terabytes of data
- ❑ Aggregations on unstructured data
- ❑ Transactional updates on relatively small datasets

Module review (1 of 2)

Relational databases are a good choice when you need:

(select all of the correct options)

- ❑ Streaming, high-throughput writes
- ❑ Fast queries on terabytes of data
- ❑ Aggregations on unstructured data
- ✓ Transactional updates on relatively small datasets

Module review (2 of 2)

Cloud SQL and Cloud Dataproc offer familiar tools (MySQL and Hadoop/Pig/Hive/Spark). What is the value-add provided by Google Cloud Platform?

(select all of the correct options)

- ❑ It's the same API, but Google implements it better
- ❑ Google-proprietary extensions and bug fixes to MySQL, Hadoop, and so on
- ❑ Fully-managed versions of the software offer no-ops
- ❑ Running it on Google infrastructure offers reliability and cost savings

Module review (2 of 2)

Cloud SQL and Cloud Dataproc offer familiar tools (MySQL and Hadoop/Pig/Hive/Spark). What is the value-add provided by Google Cloud Platform?

(select all of the correct options)

- ❑ It's the same API, but Google implements it better
- ❑ Google-proprietary extensions and bug fixes to MySQL, Hadoop, and so on
- ✓ Fully-managed versions of the software offer no-ops
- ✓ Running it on Google infrastructure offers reliability and cost savings

Resources

Cloud SQL	https://cloud.google.com/sql/
Cloud Dataproc	https://cloud.google.com/dataproc/
Cloud Solutions	https://cloud.google.com/solutions/

<Start Training>

#GoogleCloudOnBoard

</Start Training>

Cloud OnBoard

Scaling Data Analysis

Cloud OnBoard

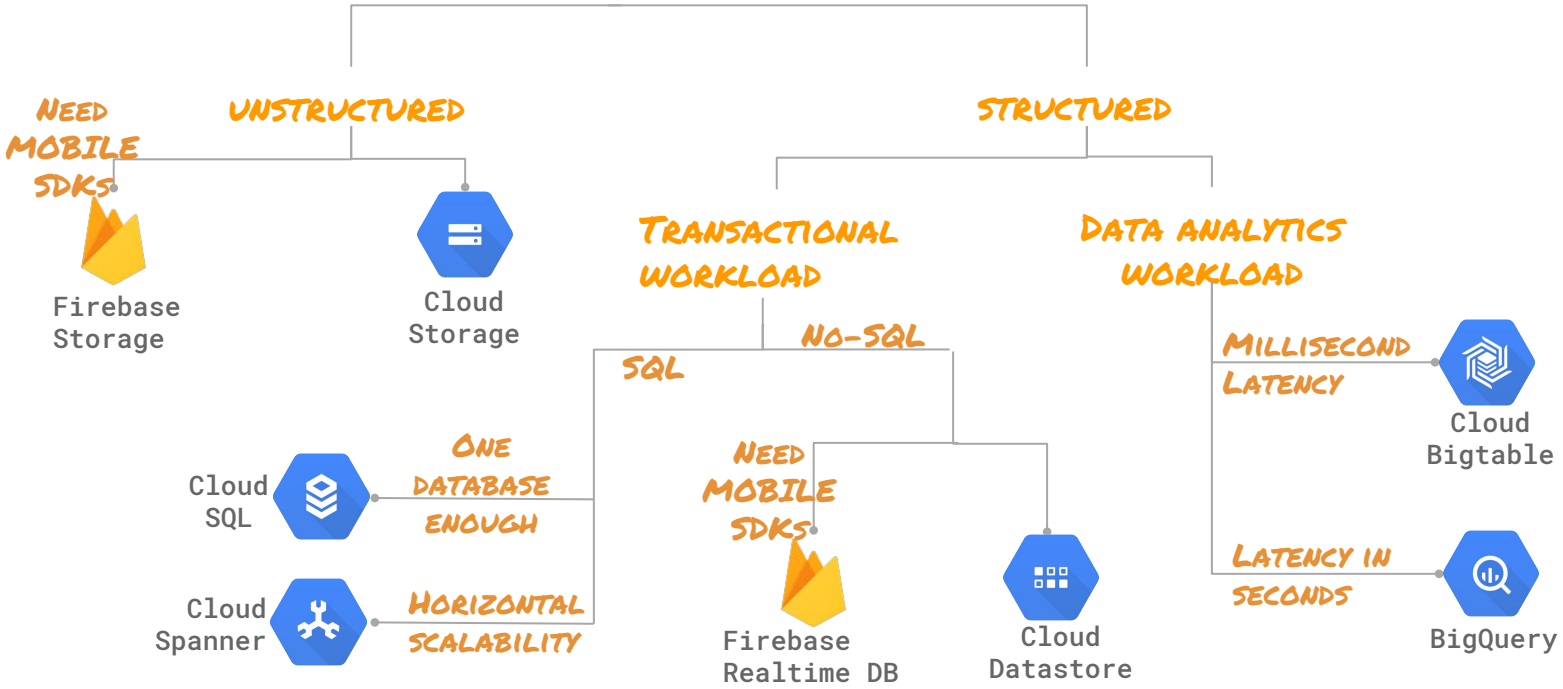
Agenda

Fast random access

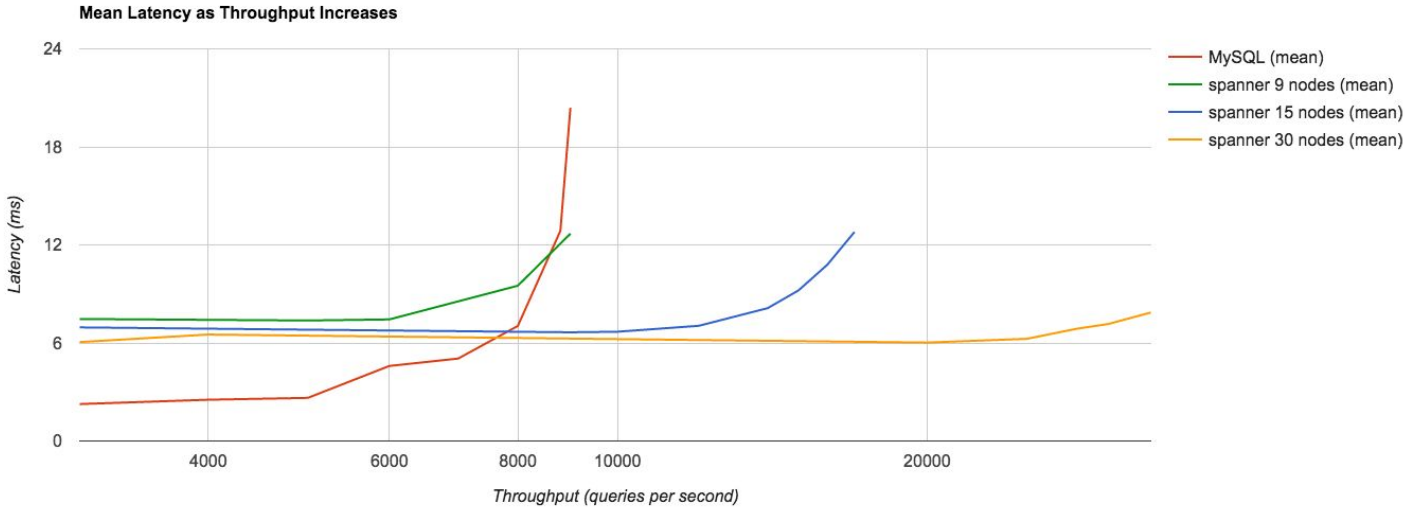
Warehouse and interactively query petabytes

Interactive, iterative development + Demo

Choosing where to store data on GCP



Use cloud spanner if you need globally consistent data or more than one Cloud SQL instance



Source:
https://quizlet.com/blog/quizlet-cloud-spanner

Comparing storage options: technical details

	Cloud Datastore	Bigtable	Cloud Storage	Cloud SQL	Cloud Spanner	BigQuery
Type	NoSQL document	NoSQL wide column	Blobstore	Relational SQL for OLTP	Relational SQL for OLTP	Relational SQL for OLAP
Transactions	Yes	Single-row	No	Yes	Yes	No
Complex queries	No	No	No	Yes	Yes	Yes
Capacity	Terabytes+	Petabytes+	Petabytes+	500 GB	Petabytes	Petabytes+
Unit size	1 MB/entity	~10 MB/cell ~100 MB/row	5 TB/object	Determined by DB engine	10,240 MiB/ row	10 MB/row

Comparing storage options: use cases

	Cloud Datastore	Bigtable	Cloud Storage	Cloud SQL	Cloud Spanner	BigQuery
Type	NoSQL document	NoSQL wide column	Blobstore	Relational SQL for OLTP	Relational SQL for OLTP	Relational SQL for OLAP
Best for	Getting started, App Engine applications	“Flat” data, Heavy read/write, events, analytical data	Structured and unstructured binary or object data	Web frameworks, existing applications	Large-scale database applications (> ~2 TB)	Interactive querying, offline analytics
Use cases	Getting started, App Engine applications	AdTech, Financial and IoT data	Images, large media files, backups	User credentials, customer orders	Whenever high I/O, global consistency is needed	Data warehousing

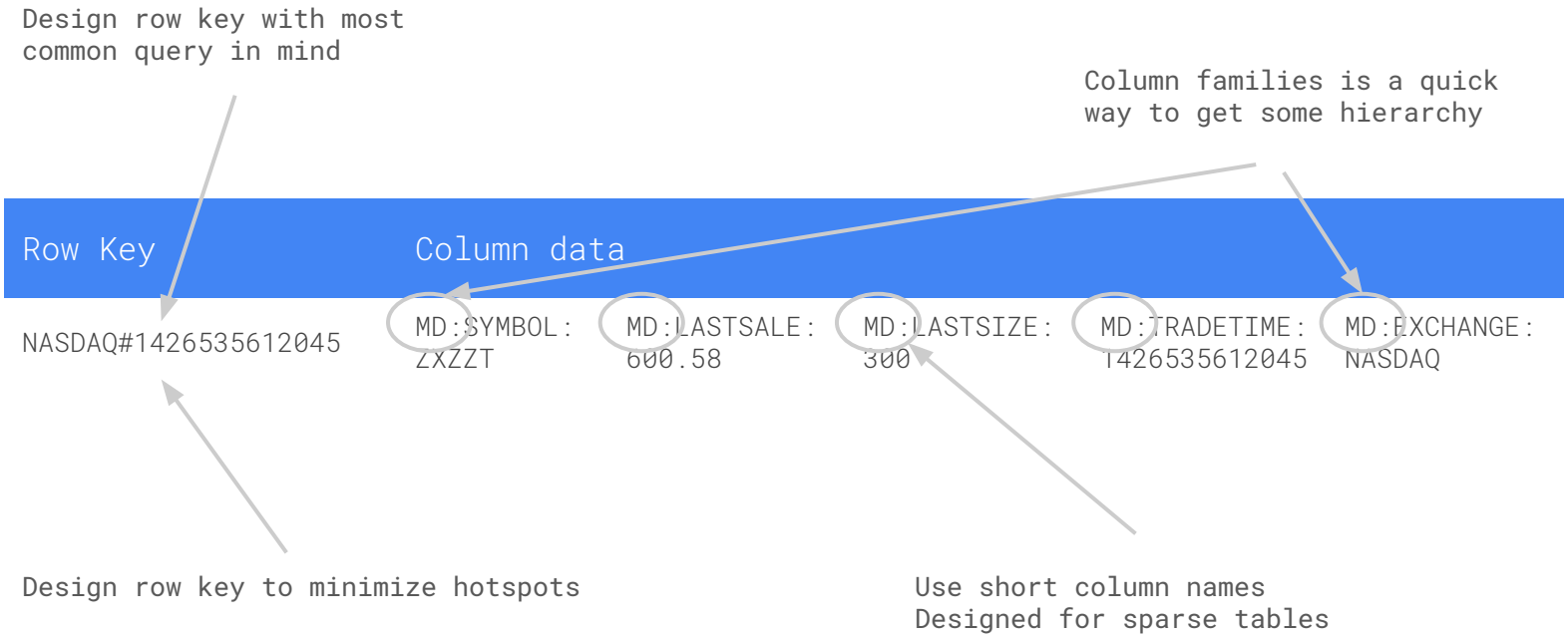
Bigtable is meant for high throughput data where access is primarily for a range of Row Key prefixes

Row Key	Column data				
NASDAQ#1426535612045	MD:SYMBOL: ZXZZT	MD:LASTSALE: 600.58	MD:LASTSIZE: 300	MD:TRADETIME: 1426535612045	MD:EXCHANGE: NASDAQ
…	…	…	…	…	…

Tables should be tall and narrow
Store changes as new rows

Bigtable will automatically compact the table

Short meaningful column names reduce storage and RPC overhead



Can work with Bigtable using the HBase API

```
import org.apache.hadoop.hbase.*;
import org.apache.hadoop.hbase.client.*;
import org.apache.hadoop.hbase.util.*;

byte[] CF = Bytes.toBytes("MD");  // column family
Connection connection = ConnectionFactory.createConnection(...)
Table table = null;
try {
      table = connection.getTable(TABLE_NAME);
      Put p = new Put(Bytes.toBytes("NASDAQ#GOOG  #1234561234561"));
      p.addColumn(CF, Bytes.toBytes("SYMBOL"), Bytes.toBytes("GOOG"));
      p.addColumn(CF, Bytes.toBytes("LASTSALE"), Bytes.toBytes(742.03d));
      ...
      table.put(p);
} finally {
      if (table != null) table.close();
}
```

Comparing storage options: technical details

	Cloud Datastore	Bigtable	Cloud Storage	Cloud SQL	Cloud Spanner	BigQuery
Type	NoSQL document	NoSQL wide column	Blobstore	Relational SQL for OLTP	Relational SQL for OLTP	Relational SQL for OLAP
Transactions	Yes	Single-row	No	Yes	Yes	No
Complex queries	No	No	No	Yes	Yes	Yes
Capacity	Terabytes+	Petabytes+	Petabytes+	500 GB	Petabytes	Petabytes+
Unit size	1 MB/entity	~10 MB/cell ~100 MB/row	5 TB/object	Determined by DB engine	10,240 MiB/ row	10 MB/row

Comparing storage options: use cases

	Cloud Datastore	Bigtable	Cloud Storage	Cloud SQL	Cloud Spanner	BigQuery
Type	NoSQL document	NoSQL wide column	Blobstore	Relational SQL for OLTP	Relational SQL for OLTP	Relational SQL for OLAP
Best for	Getting started, App Engine applications	“Flat” data, Heavy read/write, events, analytical data	Structured and unstructured binary or object data	Web frameworks, existing applications	Large-scale database applications (> ~2 TB)	Interactive querying, offline analytics
Use cases	Getting started, App Engine applications	AdTech, Financial and IoT data	Images, large media files, backups	User credentials, customer orders	Whenever high I/O, global consistency is needed	Data warehousing

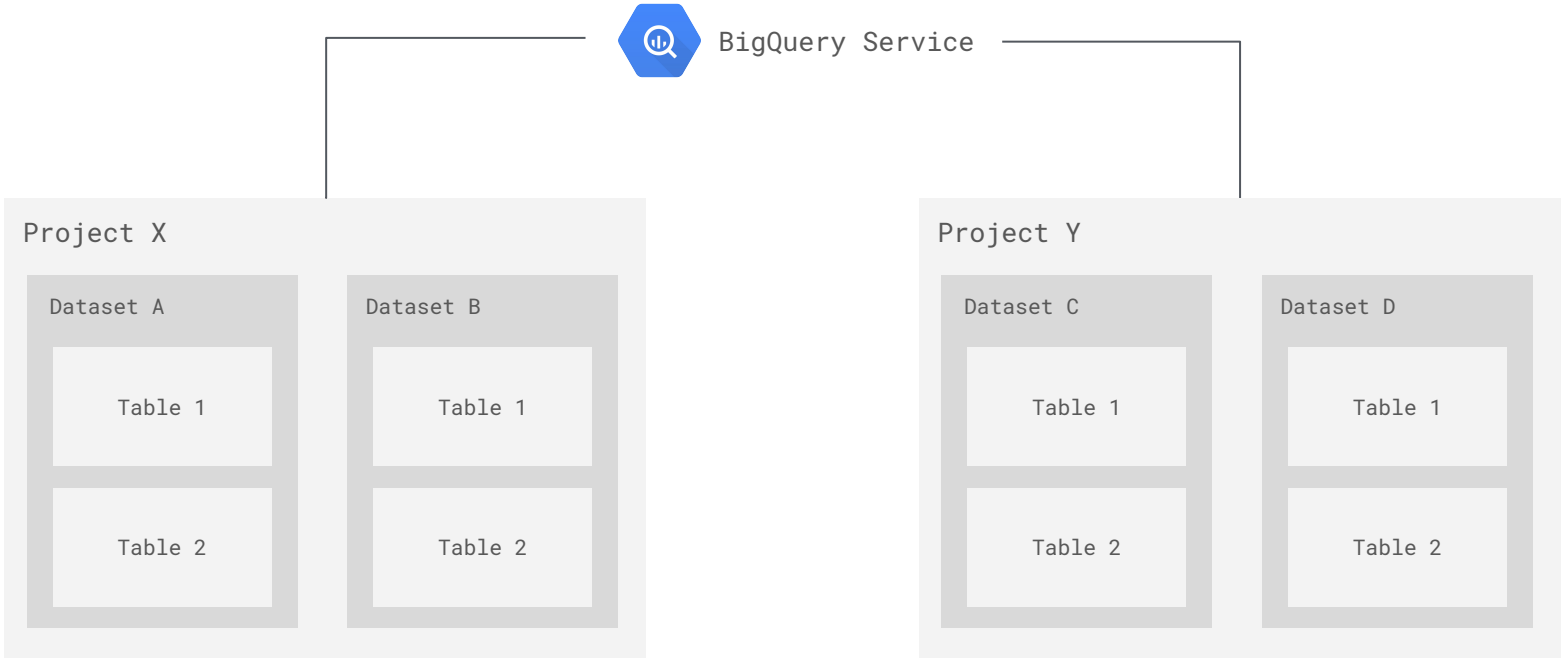
Agenda

Fast random access

Warehouse and interactively query petabytes

Interactive, iterative development + Demo

BigQuery is a fully managed data warehouse that lets you do ad-hoc SQL queries on massive volumes of data



A demo of BigQuery on a 10 billion-row dataset shows what it is and what it can do

```
#standardsql
SELECT
    language, SUM(views) as views
FROM `bigquery-samples.wikipedia_benchmark.Wiki10B`
WHERE
    title like "%google%"
GROUP by language
ORDER by views DESC
```

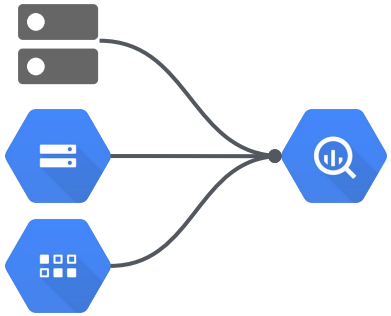
Familiar, SQL 2011 query language

Interactive ad-hoc analysis of petabyte-scale databases

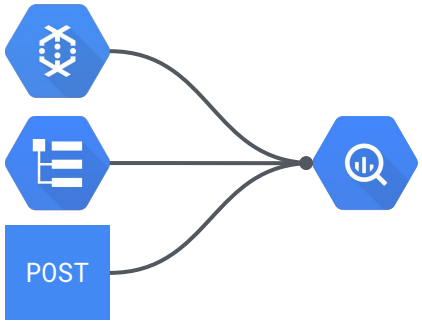
No need to provision clusters

Three ways of loading data into BigQuery

Files on disk or Cloud Storage



Stream Data



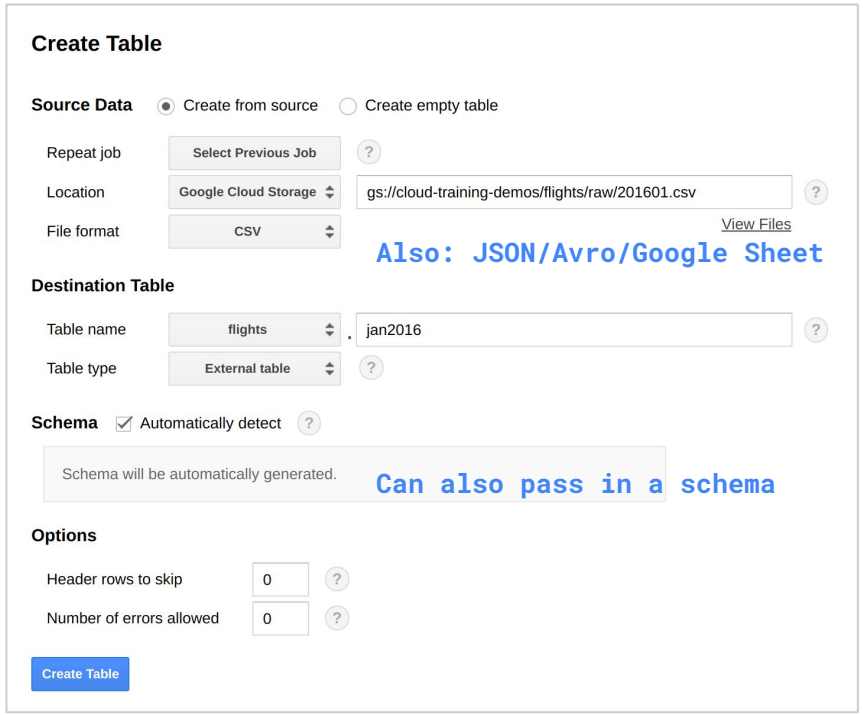
Federated data source

CSV
JSON
AVRO
Google Sheets

Serverless ETL

With Federated data sources, you can directly query files on Cloud Storage, without having to ingest them into BigQuery

Also: Google Drive, Bigtable



Comparing storage options: technical details

	Cloud Datastore	Bigtable	Cloud Storage	Cloud SQL	Cloud Spanner	BigQuery
Type	NoSQL document	NoSQL wide column	Blobstore	Relational SQL for OLTP	Relational SQL for OLTP	Relational SQL for OLAP
Transactions	Yes	Single-row	No	Yes	Yes	No
Complex queries	No	No	No	Yes	Yes	Yes
Capacity	Terabytes+	Petabytes+	Petabytes+	500 GB	Petabytes	Petabytes+
Unit size	1 MB/entity	~10 MB/cell ~100 MB/row	5 TB/object	Determined by DB engine	10,240 MiB/ row	10 MB/row

Comparing storage options: use cases

	Cloud Datastore	Bigtable	Cloud Storage	Cloud SQL	Cloud Spanner	BigQuery
Type	NoSQL document	NoSQL wide column	Blobstore	Relational SQL for OLTP	Relational SQL for OLTP	Relational SQL for OLAP
Best for	Getting started, App Engine applications	“Flat” data, Heavy read/write, events, analytical data	Structured and unstructured binary or object data	Web frameworks, existing applications	Large-scale database applications (> ~2 TB)	Interactive querying, offline analytics
Use cases	Getting started, App Engine applications	AdTech, Financial and IoT data	Images, large media files, backups	User credentials, customer orders	Whenever high I/O, global consistency is needed	Data warehousing

Agenda

Fast random access

Warehouse and interactively query petabytes

Interactive, iterative development + Demo

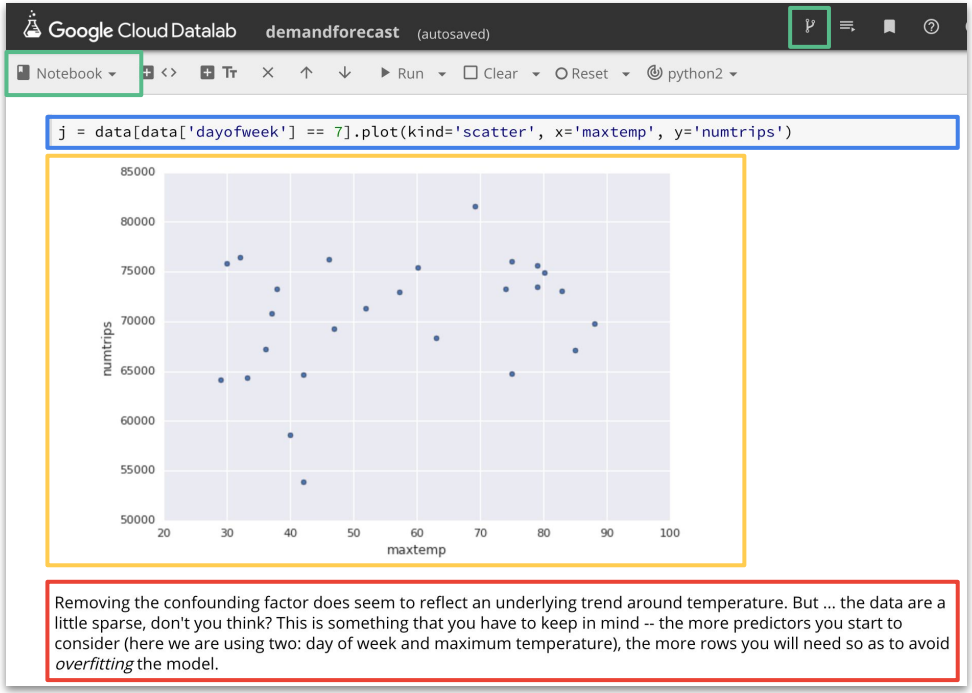
Increasingly, data analysis and machine learning are carried out in self-descriptive, shareable, executable notebooks

Share

Code

Output

Markup



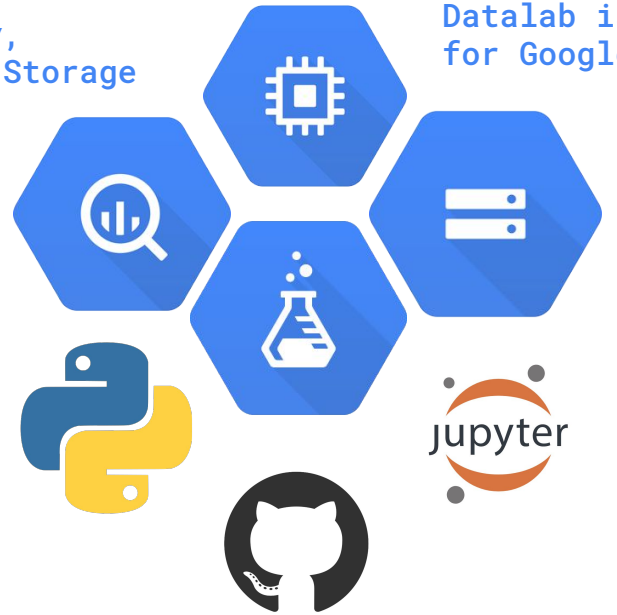
A typical notebook contains code, charts, and explanations

Image Source: Git Logo from Wikipedia

Datalab is an open-source notebook built on Jupyter (IPython)

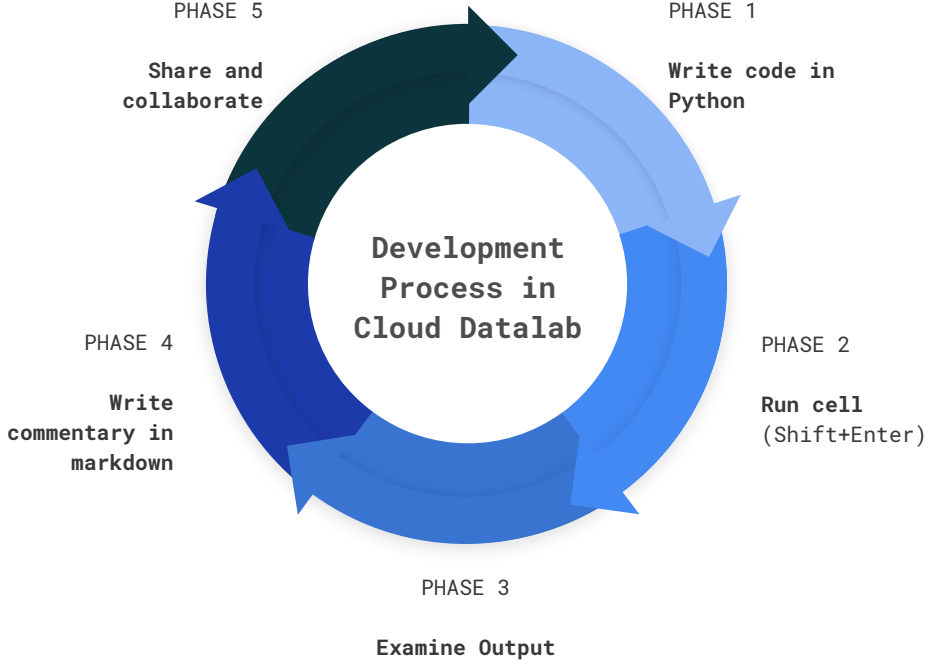
Analyze data in BigQuery,
Compute Engine or Cloud Storage

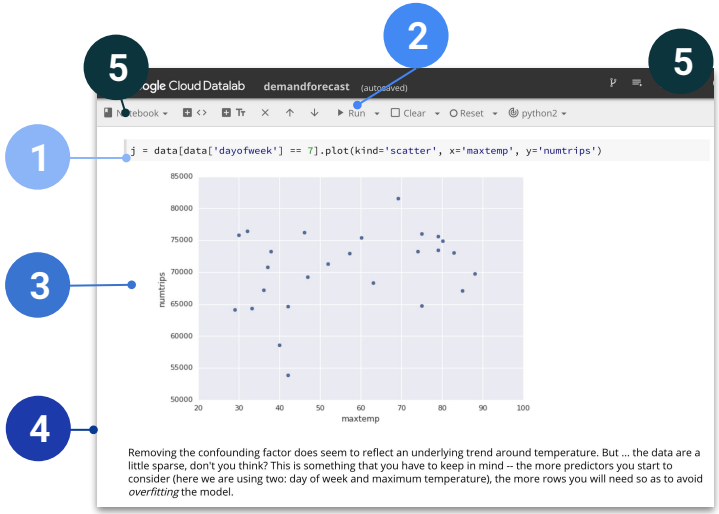
Datalab is free—just pay
for Google Cloud resources



Use existing
Python packages

Datalab notebooks are developed in an iterative, collaborative process





Datalab supports BigQuery

%%sql

```
%bq query -n wxquery
SELECT EXTRACT (DAYOFYEAR FROM CAST(CONCAT(@YEAR,'-',mo,'-',da) AS TIMESTAMP)
       MIN(EXTRACT (DAYOFWEEK FROM CAST(CONCAT(@YEAR,'-',mo,'-',da) AS TIMEST
       MIN(min) mintemp, MAX(max) maxtemp, MAX(IF(prcp=99.99,0,prcp)) rain
FROM `bigquery-public-data.noaa_gsod.gsod*`
WHERE stn='725030' AND _TABLE_SUFFIX = @YEAR
GROUP BY 1 ORDER BY daynumber DESC
```

	daynumber	dayofweek	mintemp	maxtemp	rain
0	365	5	46.0	48.2	0.17
1	364	4	34.0	48.0	0.13
2	363	3	33.8	46.9	0.37
3	362	2	39.0	62.1	0.02
4	361	1	46.0	62.6	0.14

```
query_parameters = [
  {
    'name': 'YEAR',
    'parameterType': {'type': 'STRING'},
    'parameterValue': {'value': 2015}
  }
]
weather = wxquery.execute(query_params=query_parameters).result().to_dataframe()
weather[:5]
```

To Pandas

Demo:

Create ML dataset with BigQuery

Demo: Create ML dataset with BigQuery

In this demo, we use BigQuery to create a dataset that we later use to build a taxi demand forecast system using Machine Learning.

- What kinds of things affect taxi demand?
- What are some ways to measure “demand”?

Demo: Create ML dataset with BigQuery

In this demo, we use BigQuery to create a dataset that we later use to build a taxi demand forecast system using Machine Learning.

1. Use BigQuery and Datalab to explore and visualize data
2. Build a Pandas dataframe that will be used as the training dataset for machine learning using TensorFlow

Module Review

Module review

Match the use case on the left with the product on the right

Global consistency needed	1. Datalab
High-throughput writes of wide-column data	2. BigTable
Warehousing structured data	3. BigQuery
Develop Big Data algorithms interactively in Python	4. Spanner

Module review

Match the use case on the left with the product on the right

Global consistency needed **(4)**	1. Datalab
High-throughput writes of wide-column data **(2)**	2. BigTable
Warehousing structured data **(3)**	3. BigQuery
Develop Big Data algorithms interactively in Python **(1)**	4. Spanner

<Start Training>

#GoogleCloudOnBoard

</Start Training>

Cloud OnBoard

Machine Learning

Cloud OnBoard

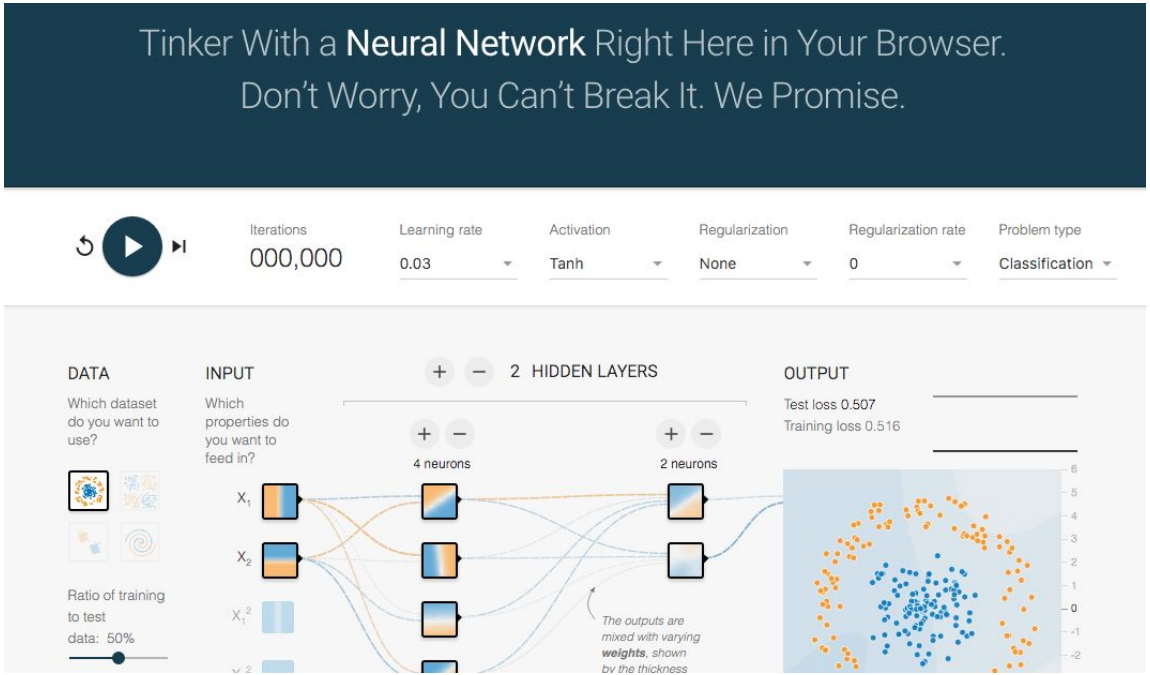
Agenda

Machine learning with TensorFlow + Demo

Pre-built machine learning models + Demo

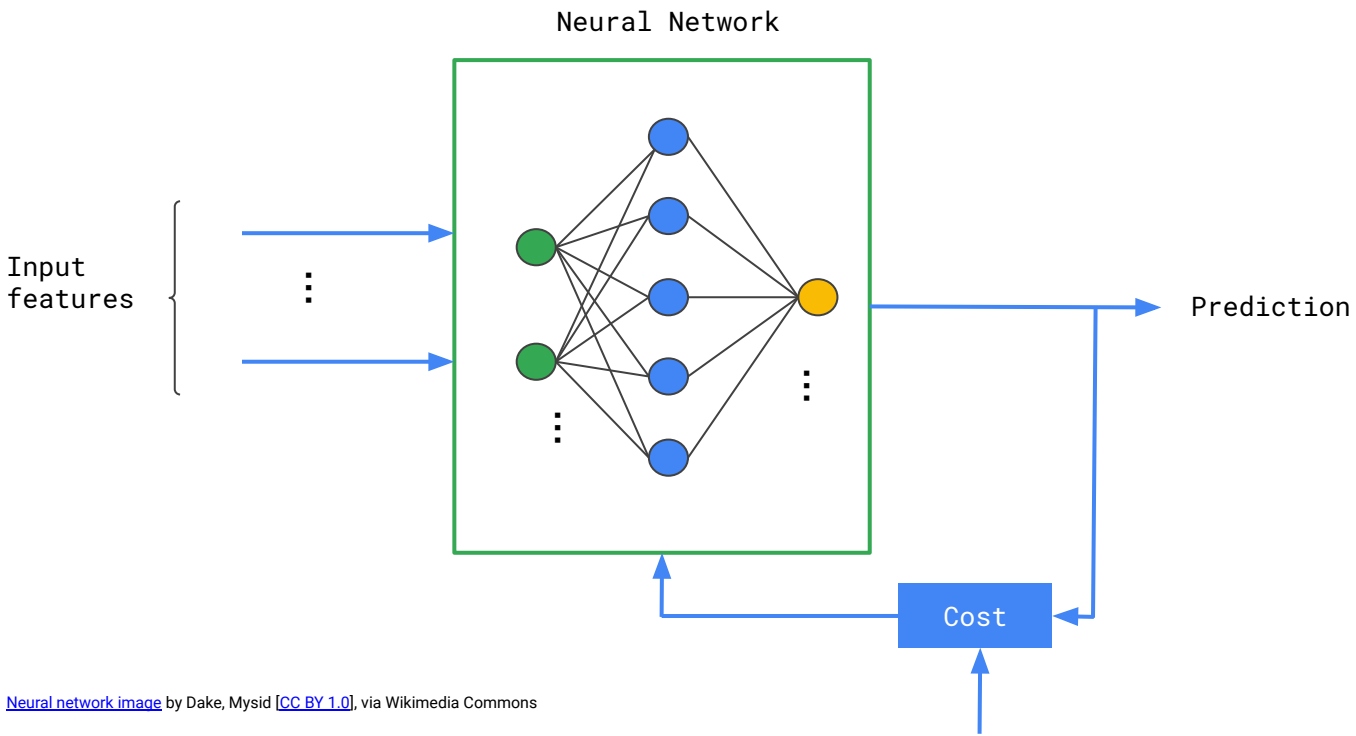
TensorFlow is an open source library that underlies many Google products

Demo: Playing with neural networks to learn what they are



http://playground.tensorflow.org/

Supervised machine learning requires features and labels



Neural network image by Dake, Mysid [CC BY 1.0], via Wikimedia Commons

Machine Learning with TensorFlow involves four steps:

Gather training data (input features and labels)

Create model

Train the model based on input data

Use the model on new data

Gather training data and select input features

1 Gather Data

Input features

	daynumber	dayofweek	mintemp	maxtemp	rain	numtrips
104	77	4	28.9	37.9	0.01	51635
9	356	2	32.0	43.0	0.00	46781
114	67	1	35.1	48.0	0.00	57377
11	354	7	30.0	37.9	0.00	74101
316	49	3	19.0	39.9	0.05	28463

discard

target

All input features need to be numeric

1 Gather Data

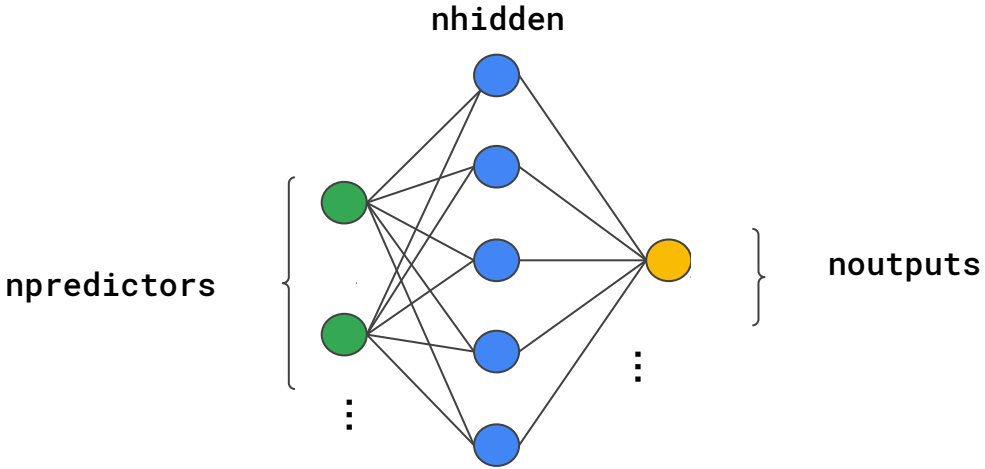
Use as-is

One-hot encoding

	mintemp	maxtemp	rain	day_1	day_2	day_3	day_4	day_5	day_6	day_7
104	28.9	37.9	0.01	0	0	0	1	0	0	0
9	32.0	43.0	0.00	0	1	0	0	0	0	0
114	35.1	48.0	0.00	1	0	0	0	0	0	0
11	30.0	37.9	0.00	0	0	0	0	0	0	1
316	19.0	39.9	0.05	0	0	1	0	0	0	0

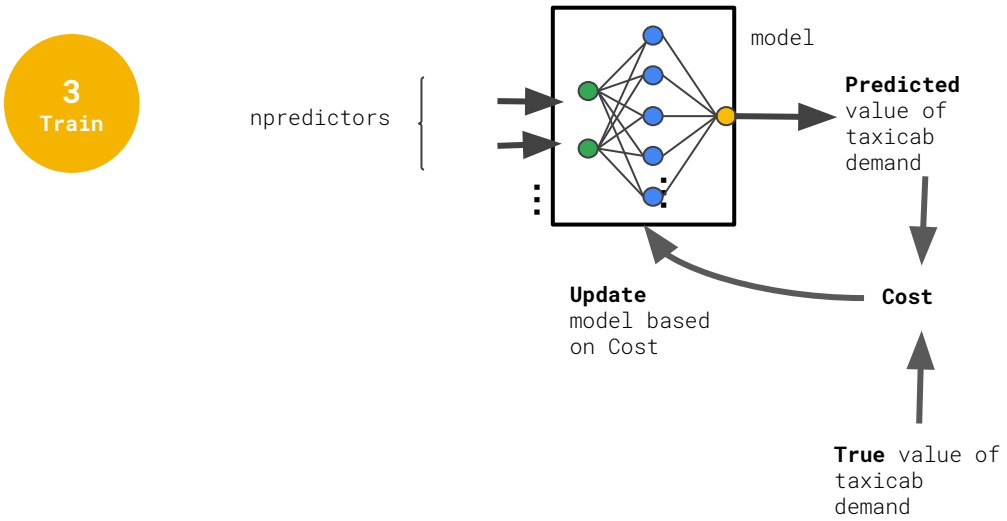
Create a neural network model, defining the number of feature columns and hidden units

2
Create



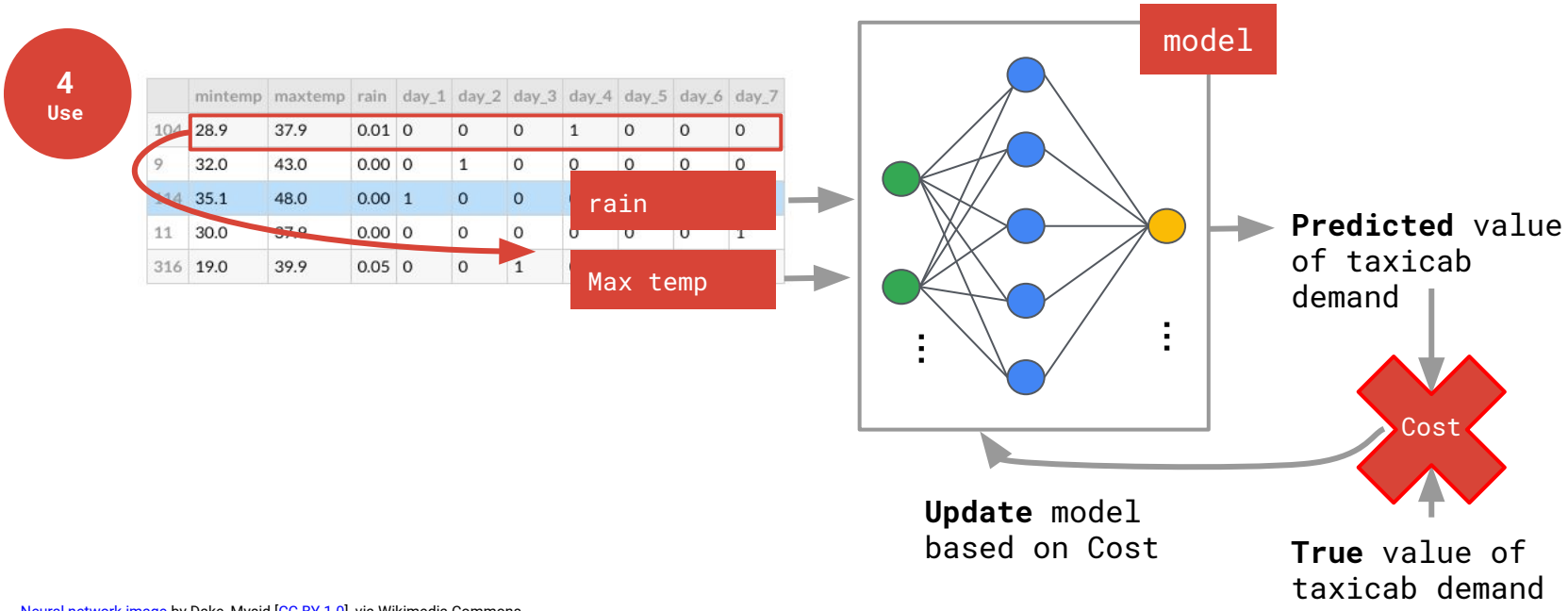

```
estimator = DNNRegressor(hidden_units=[5], feature_columns=[...])
```

Train the model on the collected data




```
estimator.fit(predictors, targets, steps=1000)
```

Train the model on the collected data



Train the model on the collected data

4
Use

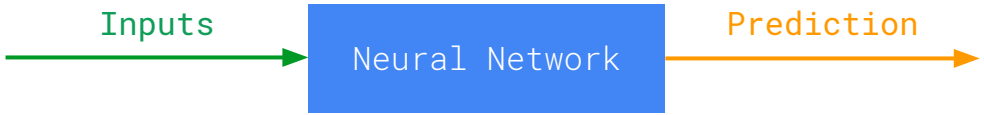
```
input = pd.DataFrame.from_dict(data =
                                {'dayofweek' : [4, 5, 6],
                                 'mintemp' : [60, 15, 60],
                                 'maxtemp' : [80, 80, 65],
                                 'rain' : [0, 0.8, 0]})


# read trained model from /tmp/trained_model
estimator = DNNRegressor(model_dir='/tmp/trained_model',
                                           hidden_units=[5])

pred = estimator.predict(input.values)
print pred
```

Demo 2 Part 2:

Carry out ML with TensorFlow

Demo 2, Part 2: Carry out ML with TensorFlow

In this demo, we build a neural network to predict taxicab demand on a day-by-day basis using TensorFlow.

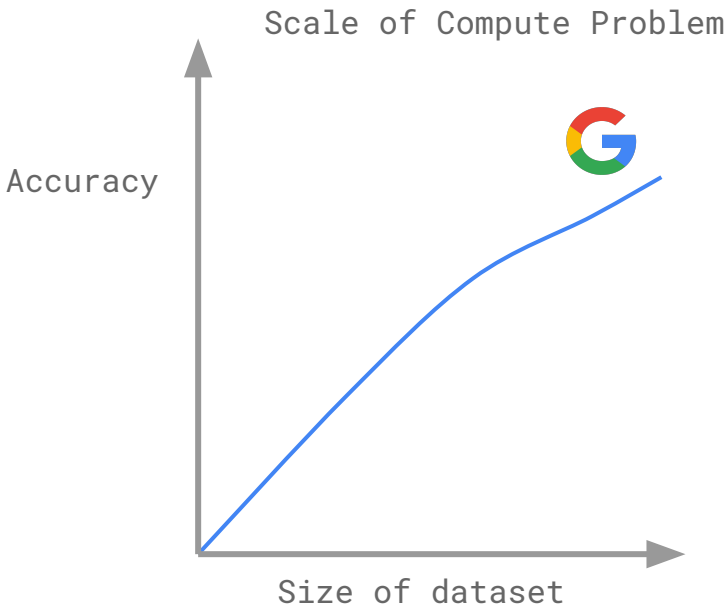


Agenda

Machine learning with TensorFlow + Demo

Pre-built machine learning models + Demo

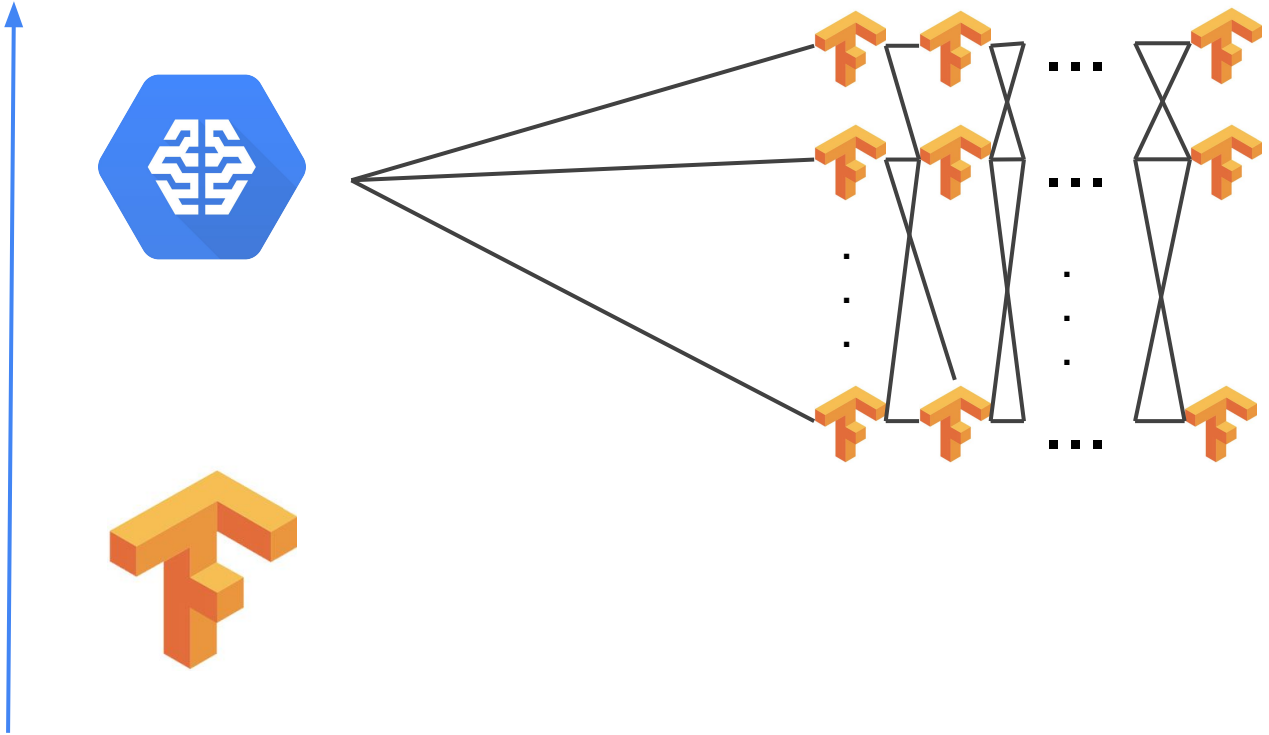
The accuracy of a ML problem is driven largely by the size and quality of the dataset; this is why ML requires massive compute



https://cloudplatform.googleblog.com/2016/05/Google-supercharges-machine-learning-tasks-with-custom-chip.html

CloudML Engine simplifies the use of Distributed TensorFlow

Size of dataset



ML APIs are pre-trained ML models (trained off Google's data) for common tasks; they are accessible through REST APIs

Use your own data to train models

TensorFlow

Cloud Machine Learning Engine

Machine Learning as an API

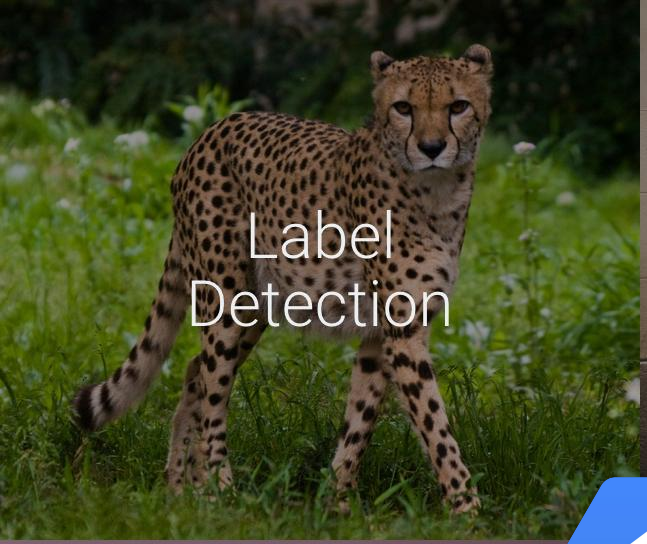
Cloud Vision API

Cloud Speech API

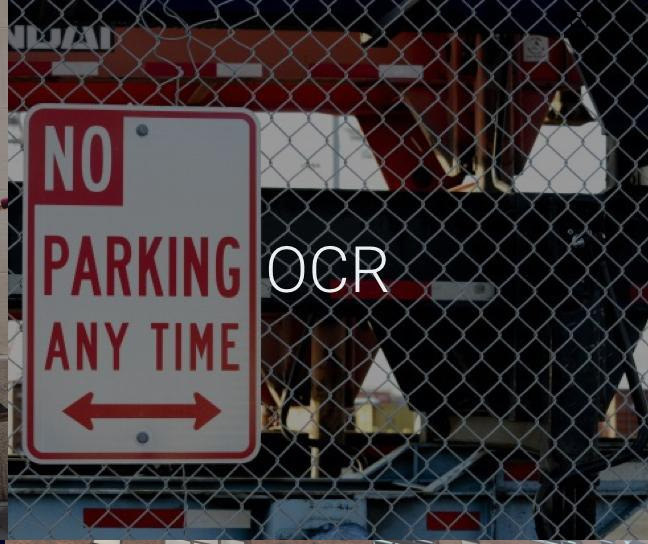
Cloud Natural Language API

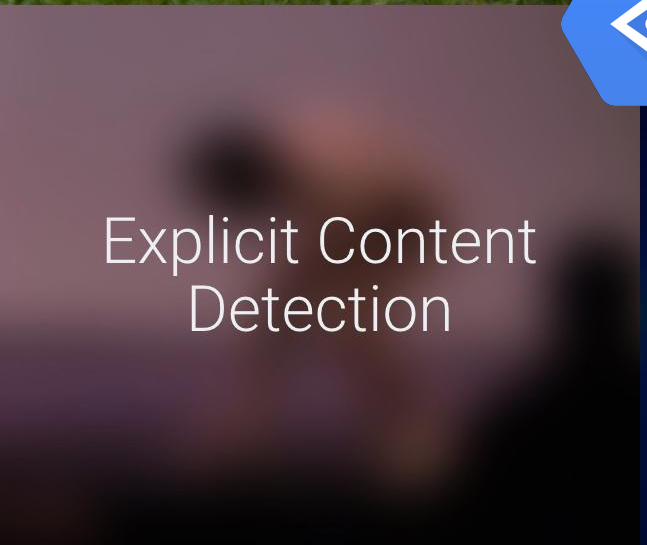
Cloud Translation API

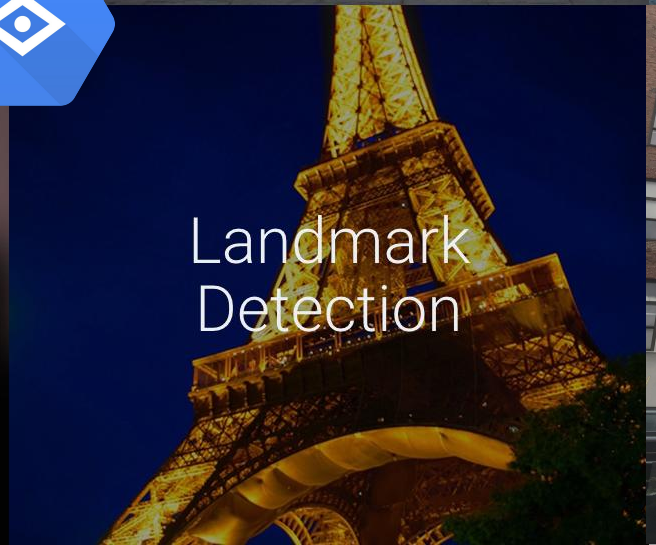
Cloud Video Intelligence


Label
Detection


Face
Detection

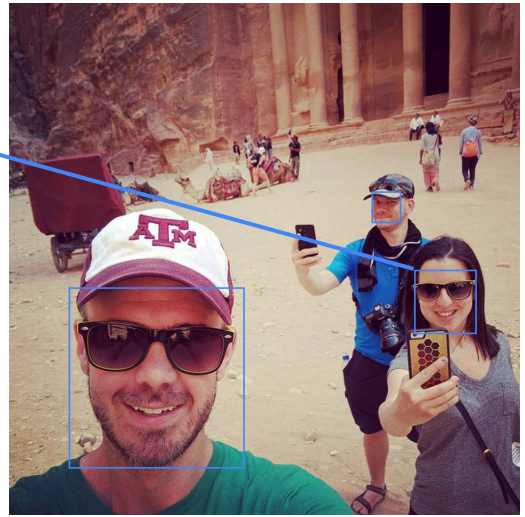

NO
PARKING
ANY TIME
OCR


Explicit Content
Detection


Landmark
Detection


Logo Detection
Google
Google Cloud

Face detection



```
"faceAnnotations" : [
  {
    "headwearLikelihood" : "VERY_UNLIKELY",
    "surpriseLikelihood" : "VERY_UNLIKELY",
    rollAngle" : -4.6490049,
    "angerLikelihood" : "VERY_UNLIKELY",
    "landmarks" : [
      {
        "type" : "LEFT_EYE",
        "position" : {
          "x" : 691.97974,
          "y" : 373.11096,
          "z" : 0.000037421443
        }
      },
      ...
    ],
    "boundingPoly" : {
      "vertices" : [
        {
          "x" : 743,
          "y" : 449
        },
        ...
```

```
"detectionConfidence" : 0.93568963,
"joyLikelihood" : "VERY_LIKELY",
"panAngle" : 4.150538,
"sorrowLikelihood" : "VERY_UNLIKELY",
"tiltAngle" : -19.377356,
"underExposedLikelihood" : "VERY_UNLIKELY",
"blurredLikelihood" : "VERY_UNLIKELY"
```

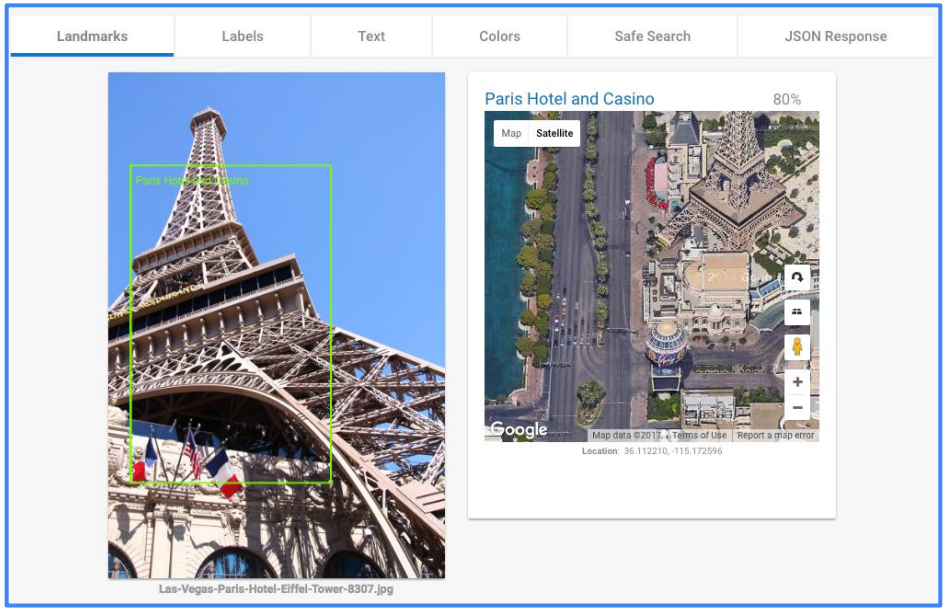
Web annotations


```
{
  "entityId": "/m/0gff2yr",
  "score": 5.92256,
  "description": "ArtScience Museum"
}
```

```
{
  "entityId": "/m/016ms7",
  "score": 1.44038,
  "description": "Ford Anglia"
}
```

```
{
  "entityId": "/m/0h898pd",
  "score": 7.4162,
  "description": "Harry Potter (Literary Series)"
}
```


Try it in the browser with your own images

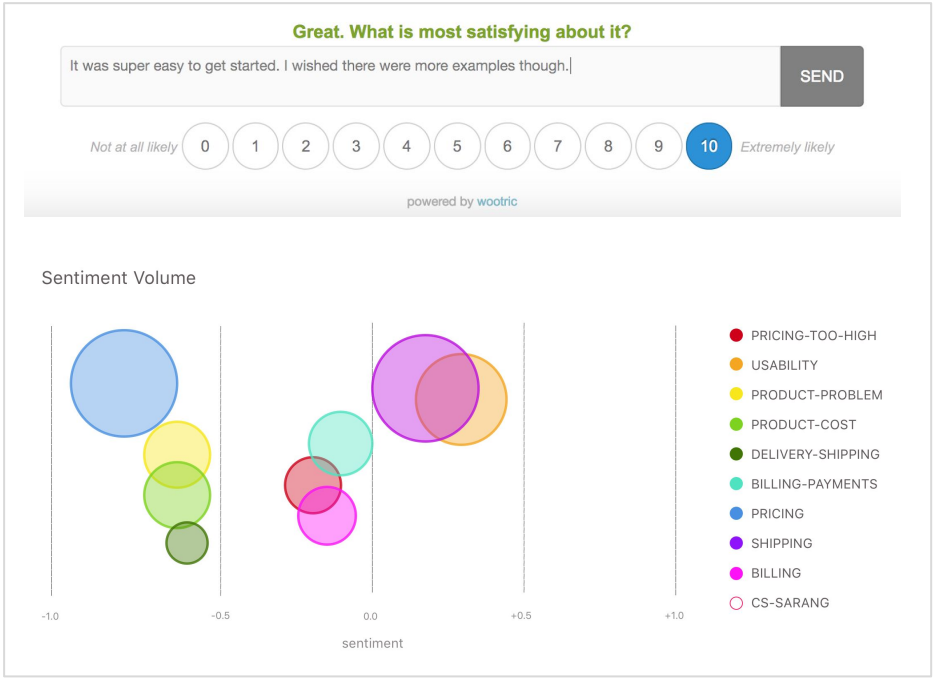


cloud.google.com/vision

The Translation API supports 100+ languages

https://cloud.google.com/translate/

Wootric uses the Cloud Natural Language API (entity and sentiment) to make sense of qualitative customer feedback



WOOTRIC
WIN CUSTOMERS FOR LIFE

Extracted entities are tied into a knowledge graph


```
{
  "name": "Joanne 'Jo' Rowling",
  "type": "PERSON",
  "metadata": {
    "mid": "/m/042xh",
    "wikipedia_url": "http://en.wikipedia.org/wiki/J._K._Rowling"
}
```


Joanne "Jo" Rowling, *pen names* ***J. K. Rowling*** *and* ***Robert Galbraith***, *is a* ***British*** *novelist, screenwriter and film producer best known as the author of the* ***Harry Potter*** *fantasy series*


```
{
  "name": "British",
  "type": "LOCATION",
  "metadata": {
    "mid": "/m/07ssc",
    "wikipedia_url": "http://en.wikipedia.org/wiki/United_Kingdom"
}
```



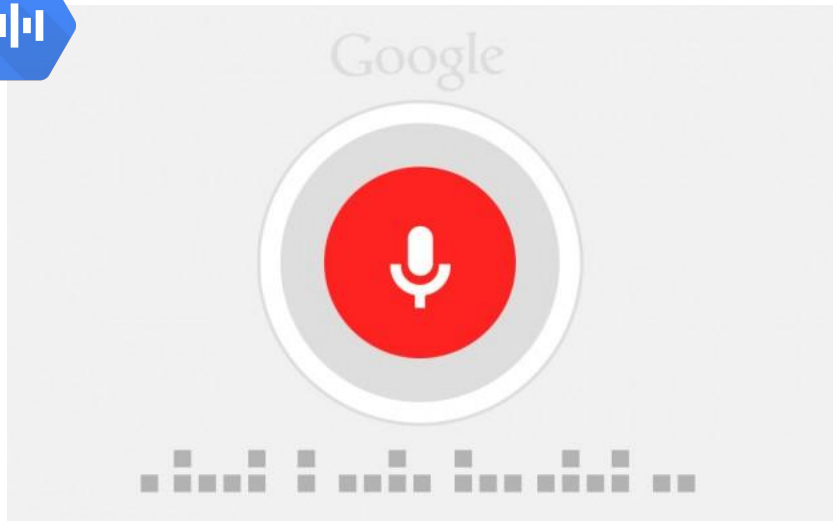
```
{
  "name": "Harry Potter",
  "type": "PERSON",
  "metadata": {
    "mid": "/m/078ffw",
    "wikipedia_url":
"http://en.wikipedia.org/wiki/Harry_Potter"
}
```

When you analyze sentiment, you get a score (positive/negative) as well as a magnitude (how intense?)

The food was excellent, I would definitely go back!


```
{
  "documentSentiment": {
    "score": 0.8,
    "magnitude": 0.8
  }
}
```

The Cloud Speech API can be used to transcribe audio to text



http://cloud.google.com/speech

Like the Vision API, the Video Intelligence API can identify labels in a video, along with a timestamp


```
{
  "description": "Bird's-eye view",
  "language_code": "en-us",
  "locations": {
    "segment": {
      "start_time_offset": 71905212,
      "end_time_offset": 73740392
    },
    "confidence": 0.96653205
  }
}
```


https://cloud.google.com/video-intelligence/

Demo 2 Part 3:

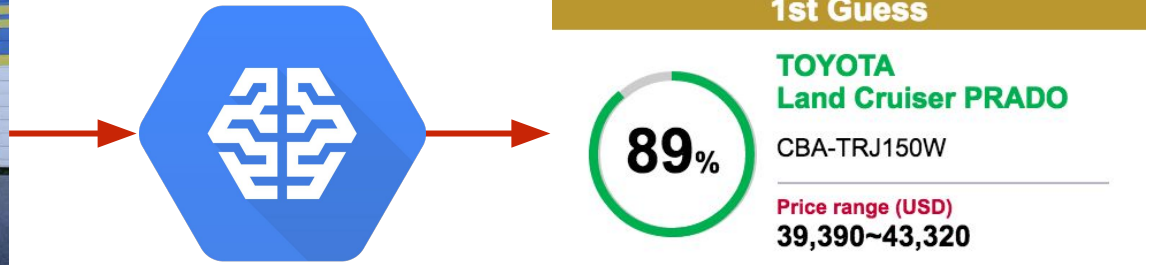
Machine Learning APIs

Demo 2, Part 3: Machine Learning APIs

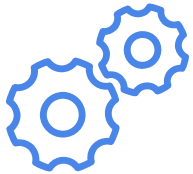
Use several of the Machine Learning APIs (Vision, Translate, Natural Language Processing, Speech) from Python

“How much is this car worth?”

AUCNET



"Thanks to the Google Cloud Platform, Ocado was able to use the power of cloud computing and train our models in parallel."



Improves natural language processing of customer service claims

"Hi Ocado,
I love your website. I have children so it's easier for me to do the shopping online.
Many thanks for saving my time!
Regards"

Feedback

Customer is happy

50%

of enterprises will be spending more per annum on bots and chatbot creation than traditional mobile app development by 2021 — *Gartner*

Custom image model to price cars

Build off NLP API to route customer emails

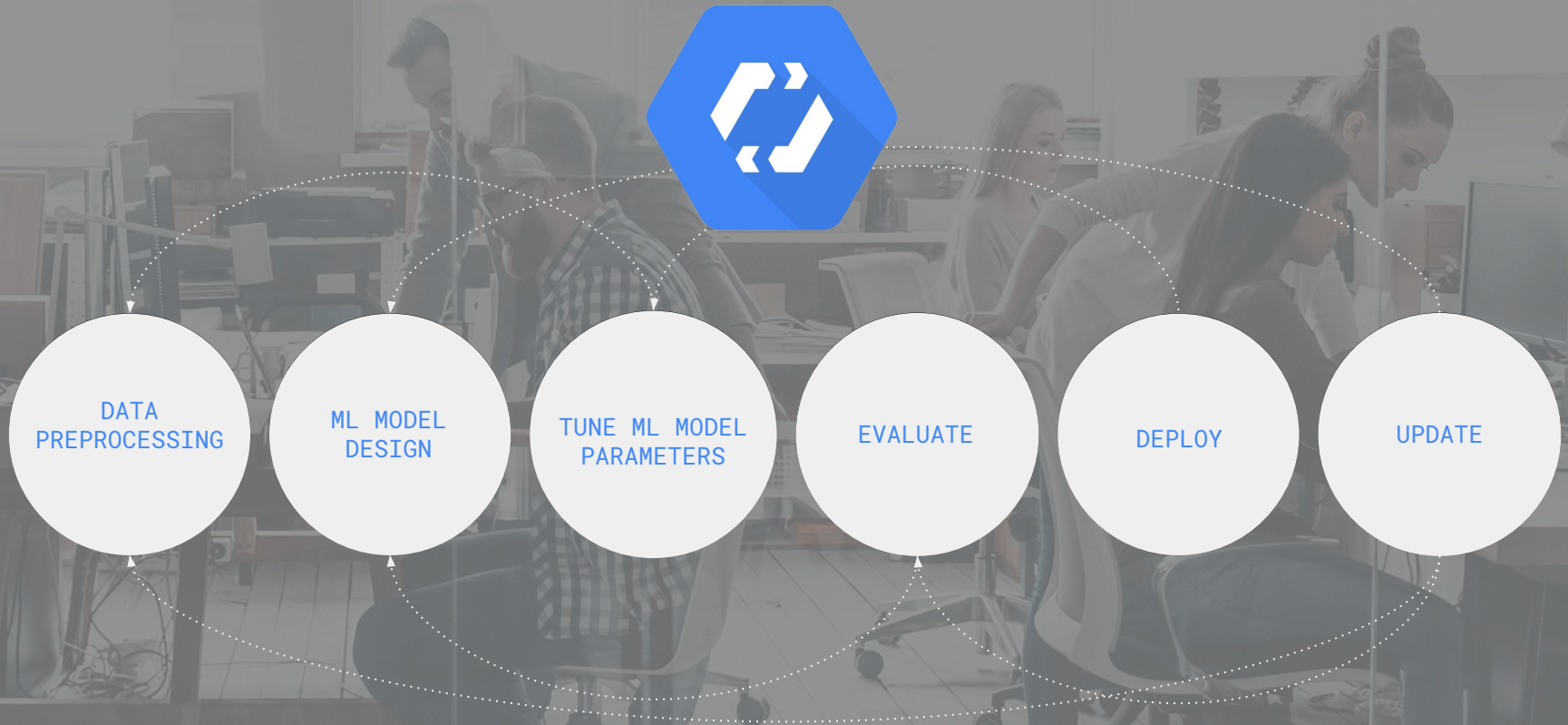
Use Vision API as-is to find text in memes

Use Dialogflow to create a new shopping experience

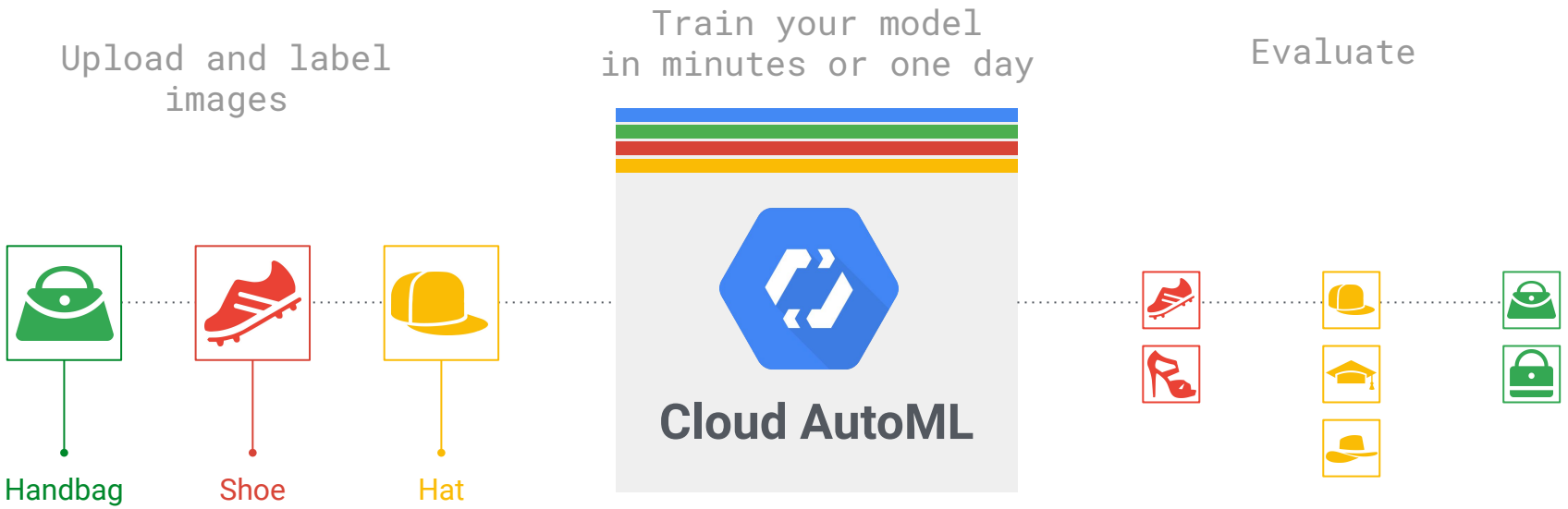

Photo Dataset
Cloud AutoML Vision
Generate predictions with a REST API
Rest API
Train
Deploy
Serve

Introducing Cloud AutoML

A technology that can automatically create a Machine Learning Model



Cloud AutoML Vision



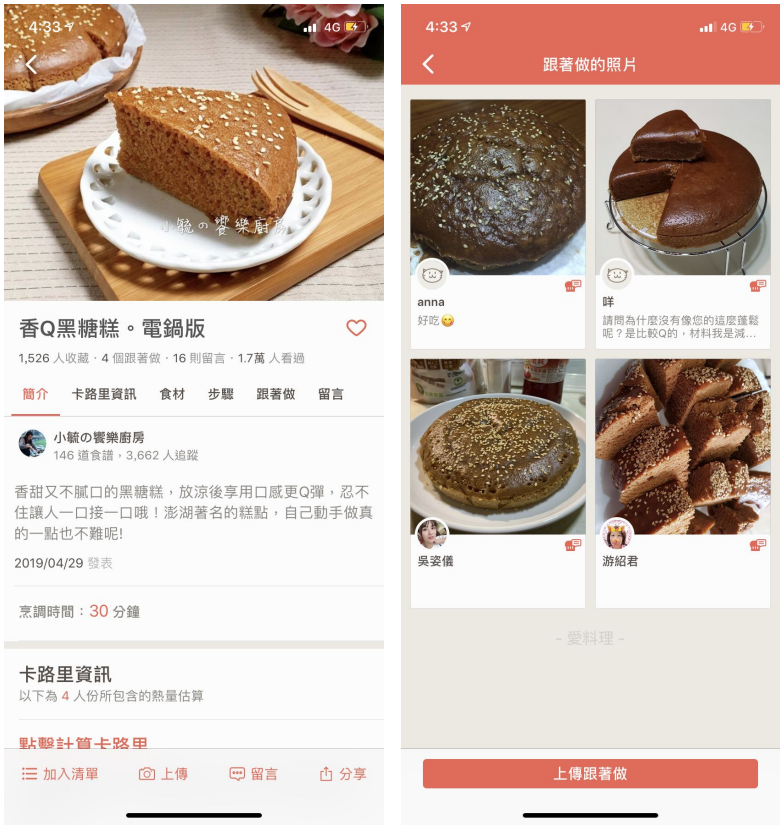
Model is now trained and ready to make prediction.
This model can scale as needed to adapt to customer demands.

Speaker

[Richard Lee]

[CTO/Co-founder, 愛料理]


愛料理



愛料理 is the leading recipe platform @ Taiwan / Hong Kong

- 170k recipes by users & counting
- Apps with 5 millions downloads
- First **Actions on Google** TW partner

愛料理

Google Cloud

We love & benefits from Google Cloud for years

Recipe Q&A Interactive Chatbot

Cooking is usually for families, and we bring the content to where they chat.

We build with GCP technologies:

- Dialogflow
- Cloud Firestore
- Cloud Functions

愛料理 Actions on Google

Hey Google, where's my favorite recipes?

We build with GCP technologies:

- Dialogflow
- Cloud Functions

食譜社群平台
愛料理

愛料理 生活誌 市集 影音 試用推薦 VIP 一週菜單 | RL

找食譜 | 搜尋食譜名 | 搜尋食材，以空格分開 | 食譜收藏 | 寫食譜

旋轉咖啡機63折 第一款琺瑯保溫杯 凱蒂貓折疊餐盒 無油健康氣炸鍋

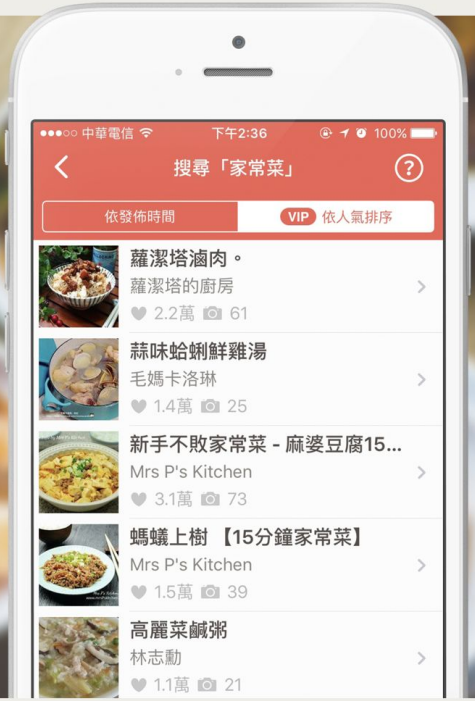
快速搜尋零失敗的人氣食譜

成為愛料理 VIP 獨享更多服務

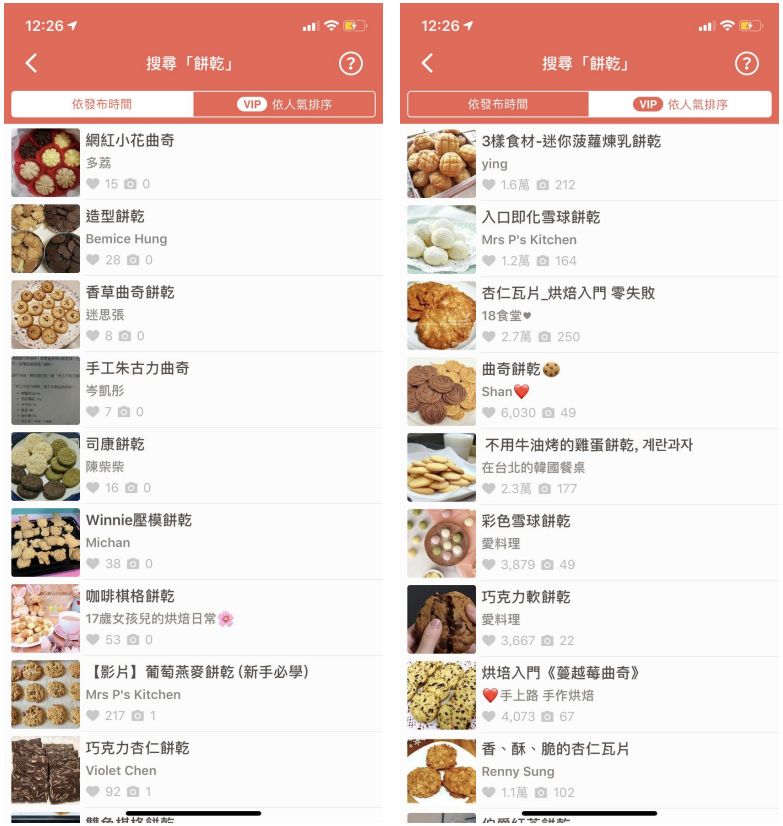
快速找到零失敗的人氣食譜，挑戰從未做過的料理，從此輕鬆變化每日菜色。

查看訂閱方案

你可隨時取消訂閱，請參考 VIP 同意條款。



Google Cloud



VIP feature - Popularity sorting

We consider several factors to rank content for VIP users to get best recipe.

We build with GCP technologies:

- BigQuery
- Cloud SQL
- Composer
- AutoML Natural Languages
- AutoML Vision (Testing)

Social listening on Instagram

As we publish content on most major social networks, it's important to gather user comments & respond.

We build with GCP technologies:

- BigQuery
- Cloud Functions
- AutoML Natural Languages

Product recommendation on LINE

To profit from chatbot, we recommend products based on user historical behaviors.

We build with GCP technologies:

- BigQuery
- Dialogflow
- Dataflow
- Cloud Pub/Sub
- Cloud Functions

We love Google Cloud because:

Build next generation UI/UX upon GCP

Form a data-driven environment for better business decisions

Using state-of-art ML models to bring intelligence to content

Google Cloud Why Google Solutions Products Pricing Getting started Contact sales Docs Support Language Console

Polydice: Delivering mobile, personalized recipe-sharing

Using Google Cloud Platform, Polydice has achieved greater stability and efficiency, and lower costs, while establishing a strong foundation to become an AI-led company.

Polydice

About Polydice

Founded in 2010, Polydice operates the iCook social recipe website and application. iCook is one of the most popular recipe services in Taiwan with approximately 5 million visitors per month and approximately 4.5 million downloads.

Industries: Gaming, Media & Entertainment

Thank you & enjoy cooking!

Demo:

Module Review

Module review

Match the use case on the left with the product on the right

Create, test new machine learning methods
No-ops, custom machine learning applications at scale
Automatically reject inappropriate image content
Build application to monitor Spanish twitter feed
Transcribe customer support calls

1. Vision API
2. TensorFlow
3. Speech API
4. Cloud ML
5. Translation API

Module review

Match the use case on the left with the product on the right

Create, test new machine learning methods **(2)**
No-ops, custom machine learning applications at scale **(4)**
Automatically reject inappropriate image content **(1)**
Build application to monitor Spanish twitter feed **(5)**
Transcribe customer support calls **(3)**

1. Vision API
2. TensorFlow
3. Speech API
4. Cloud ML
5. Translation API

Resources (1 of 2)

Cloud Spanner	https://cloud.google.com/spanner/
Cloud Bigtable	https://cloud.google.com/bigtable/
Google BigQuery	https://cloud.google.com/bigquery/
Cloud Datalab	https://cloud.google.com/datalab/
TensorFlow	https://www.tensorflow.org/

Resources (2 of 2)

Cloud Machine Learning	https://cloud.google.com/ml/
Vision API	https://cloud.google.com/vision/
Translation API	https://cloud.google.com/translate/
Speech API	https://cloud.google.com/speech/
Video Intelligence API	https://cloud.google.com/video-intelligence

<Start Training>

#GoogleCloudOnBoard

</Start Training>

Cloud OnBoard

Data Processing Architecture

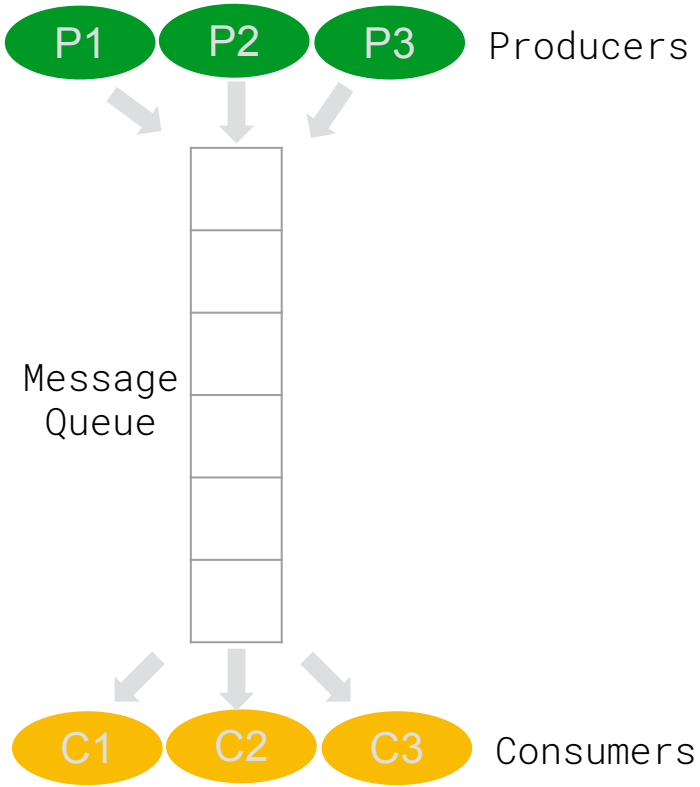
Cloud OnBoard

Agenda

Message-oriented architectures

Serverless data pipelines

GCP Reference Architecture

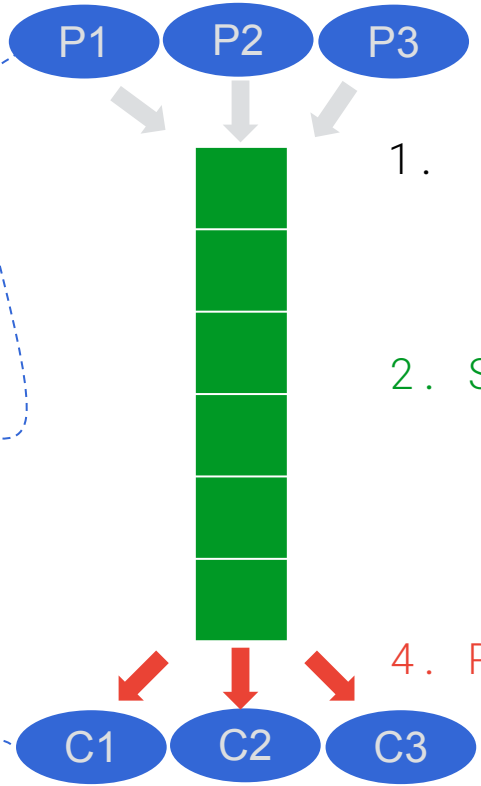


Asynchronous processing is useful for long-lived tasks or to have loose coupling between two systems

Potential use cases:

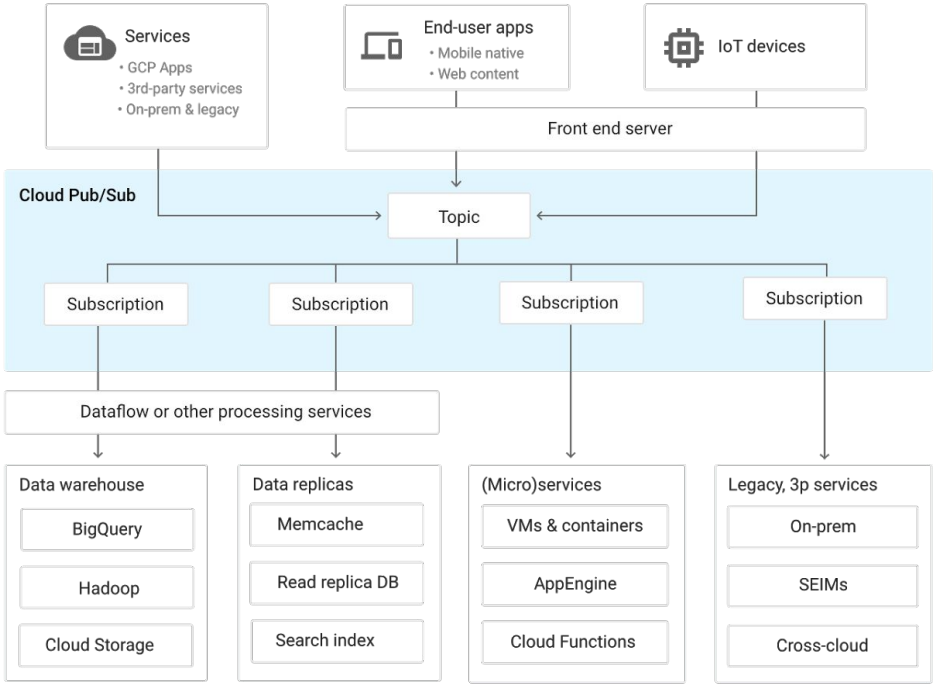
1. Send an SMS
2. Train ML model
3. Process data from multiple sources
4. Weekly reports …

For robust asynchronous processing, you need:



1. A global, highly available queue
2. Scale without over-provisioning
3. Queue must be interoperable
4. Reliable delivery of messages

Pub/Sub provides a no-ops, serverless global message queue



Agenda

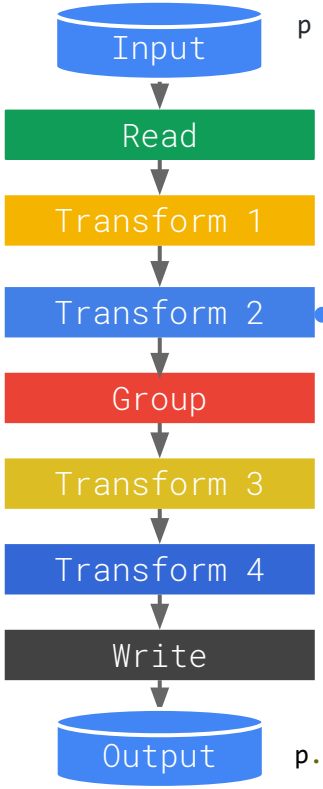
Message-oriented architectures

Serverless data pipelines

GCP Reference Architecture

Google Cloud Cloud OnBoard

Dataflow offers NoOps data pipelines in Java and Python




```
p = beam.Pipeline(options=options)

lines = p | beam.io.ReadFromText('gs://...')

traffic = lines | beam.Map(parse_data).with_output_types(unicode)

                | beam.Map(get_speedbysensor) # (sensor, speed)

                | beam.GroupByKey() # (sensor, [speed])

                | beam.Map(avg_speed) # (sensor, avgspeed)

                | beam.Map(lambda tup: '%s: %d' % tup))

output = traffic | beam.io.WriteToText('gs://...]')

p.run();
```

Open-source API (Apache Beam) can be executed on Flink, Spark, etc. also

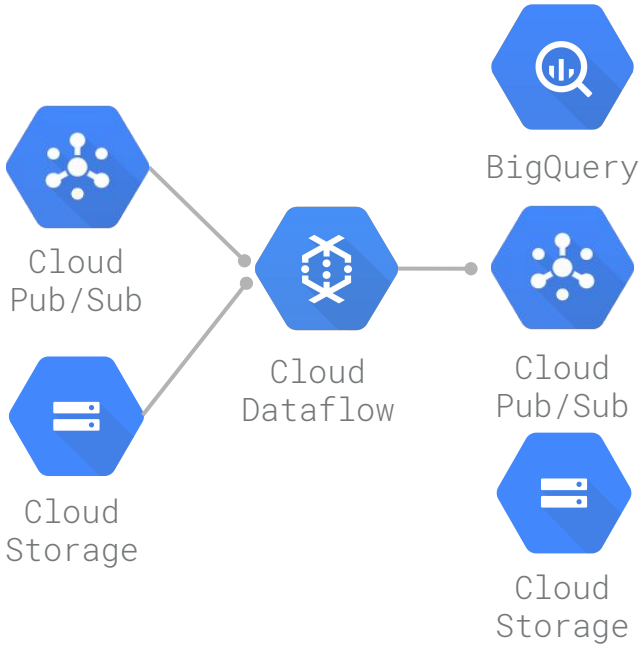
Map

Group-By

Reduce

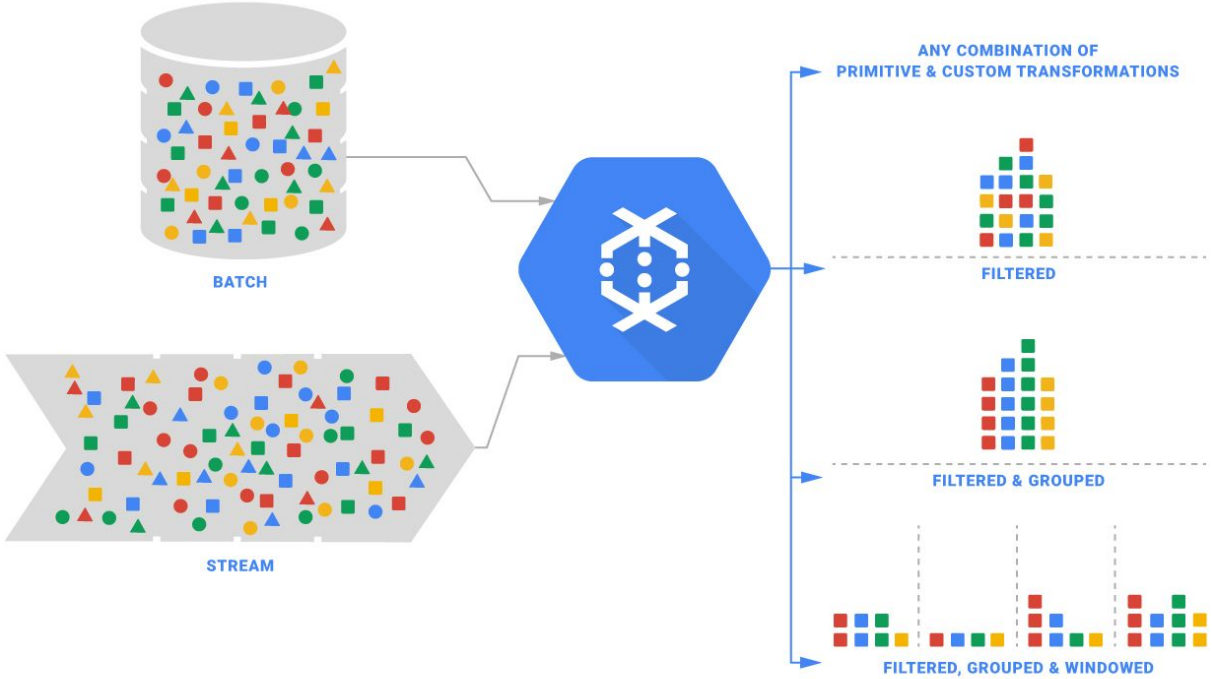
Each of these steps is run in parallel and autoscaled by execution framework

Same code does real-time and batch




```
options = PipelineOptions(pipeline_args)
options.view_as(StandardOptions).streaming = True
p = beam.Pipeline(options=options)
   lines = p | beam.io.ReadStringsFromPubSub(input_topic)
   traffic = (lines
      | beam.Map(parse_data).with_output_types(unicode)
      | beam.Map(get_speedbysensor) # (sensor, speed)
      | beam.WindowInto(window.FixedWindows(15, 0))
      | beam.GroupByKey() # (sensor, [speed])
      | beam.Map(avg_speed) # (sensor, avgspeed)
      | beam.Map(lambda tup: '%s: %d' % tup))
   traffic | beam.io.WriteStringsToPubSub(output_topic)

p.run()
```

Dataflow does ingest, transform, and load; consider using it instead of Spark



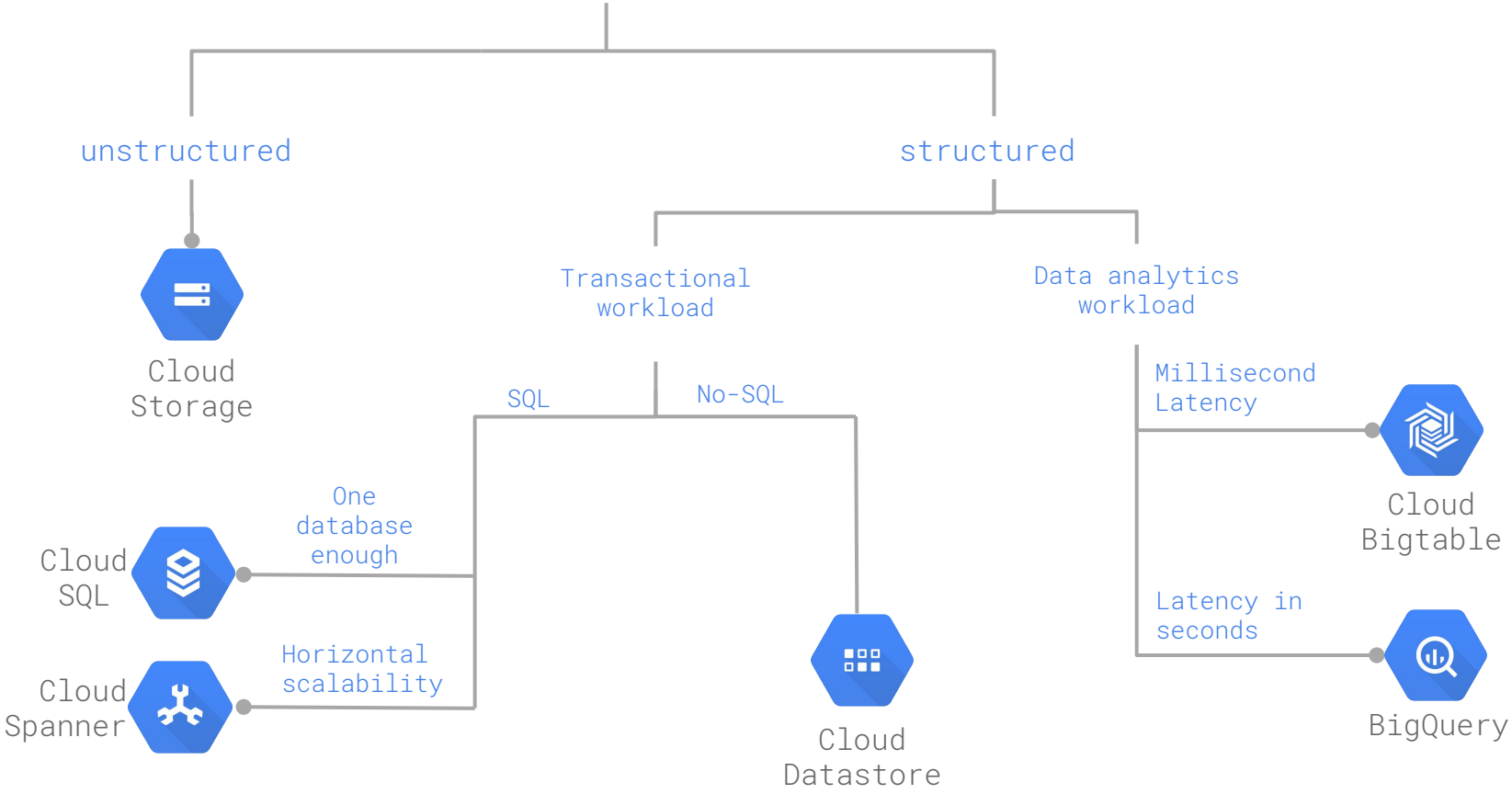
Agenda

Message-oriented architectures

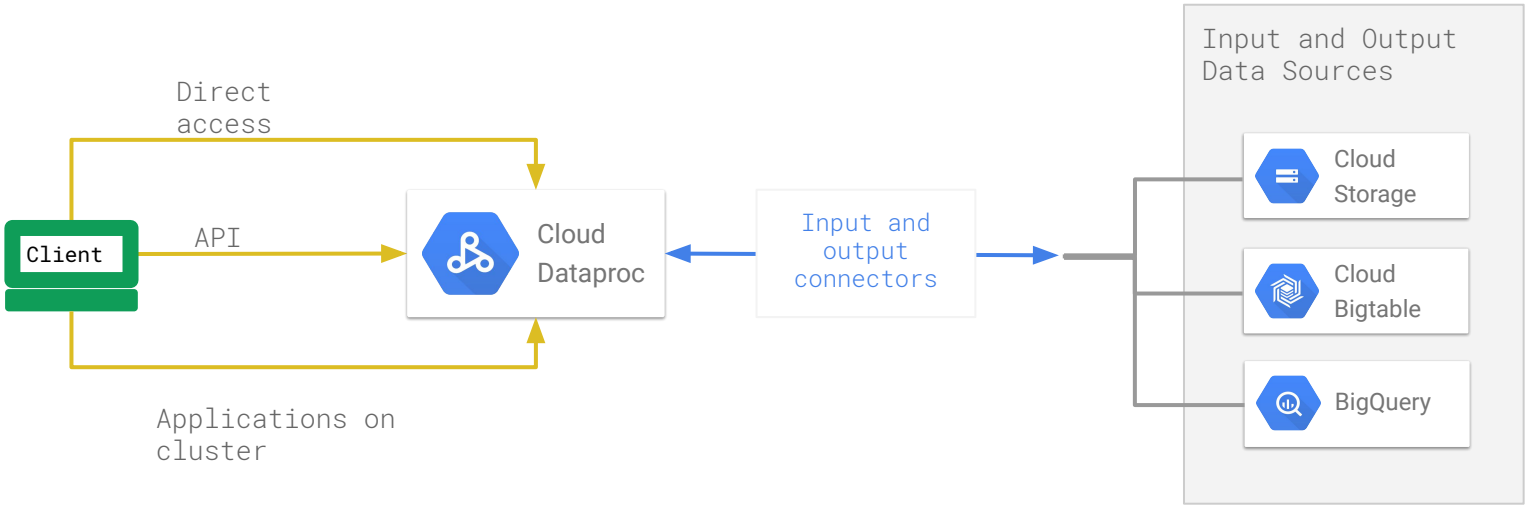
Serverless data pipelines

GCP Reference Architecture

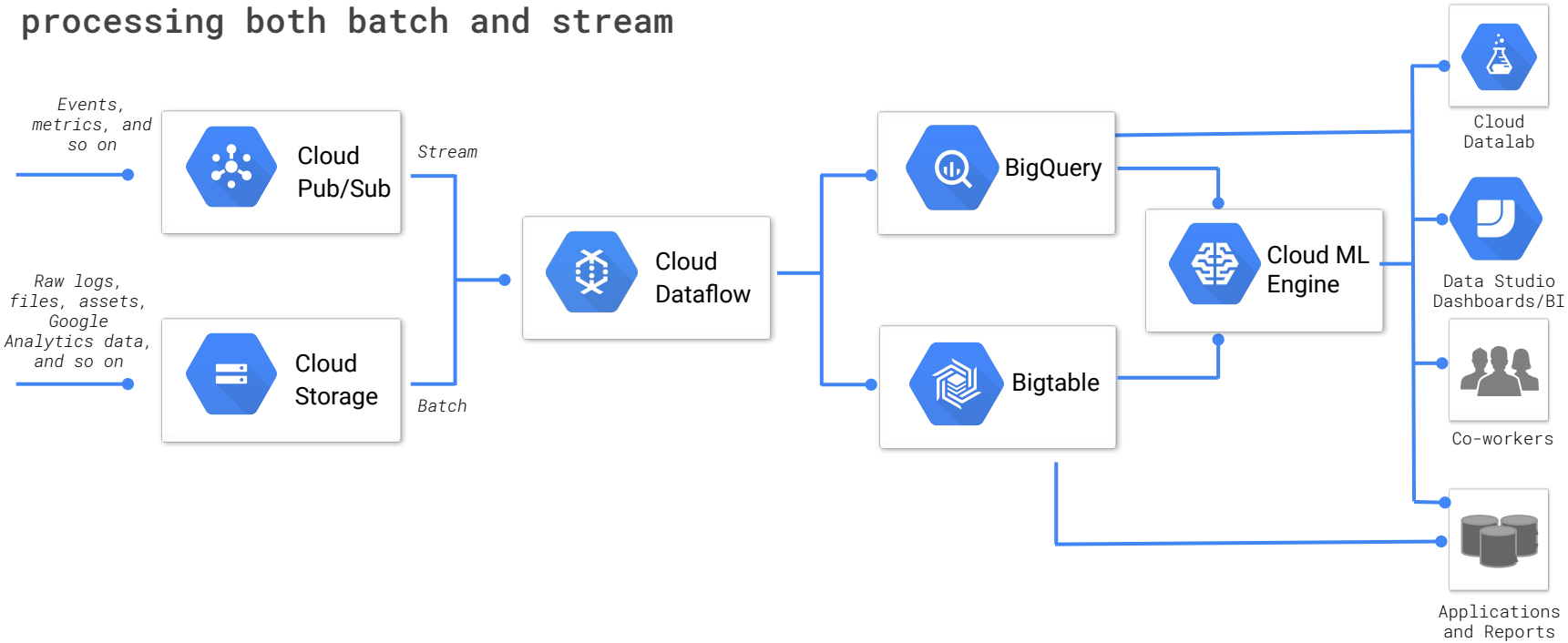
Choosing where to store data on GCP



Run Spark/Hadoop jobs on Cloud Dataproc



On GCP, you can have the same data processing pipeline for processing both batch and stream



Speaker

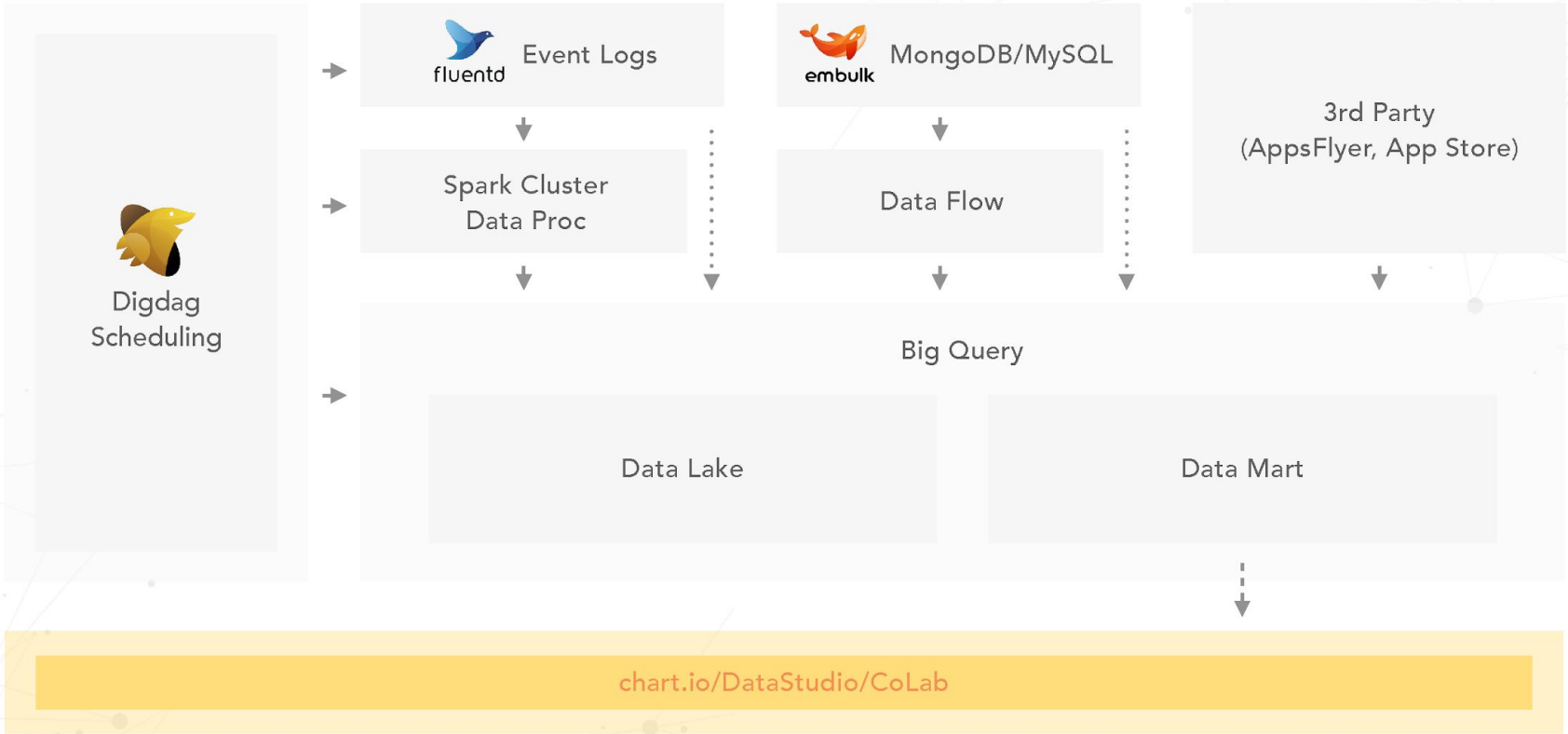
[Randy Huang]

[Director of Data,17 Media]

17 Media

17 ENTERTAINMENT | Taktor Goodnight HandsUP UN 由你娛樂 UNICORN ENTERTAINMENT

Data Architecture



Big Query

Flexible schema changing

~~Startup~~ business requirements do change on the fly

Query Offloading (MySQL, MongoDB, Firebase)

Streaming Insert for both Client/Backend Event Logs

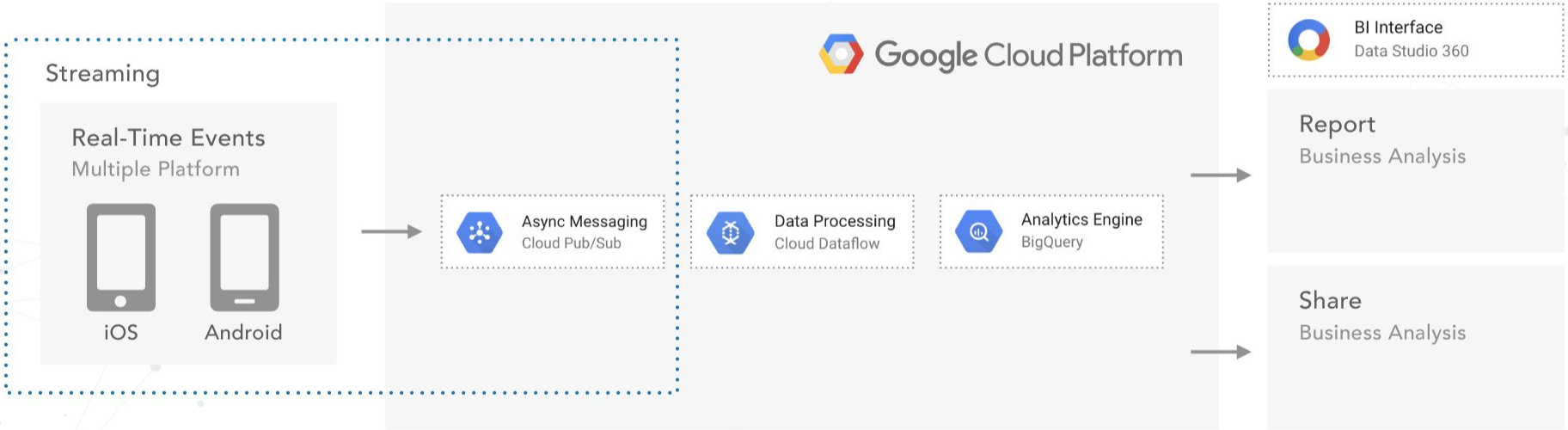
Day Partition to reduce cost

Dataset permission separately by BU

Focus on finding Business Insight instead of maintaining database

Architecture

LiveStream > Mobile Performance Monitoring



Case 1

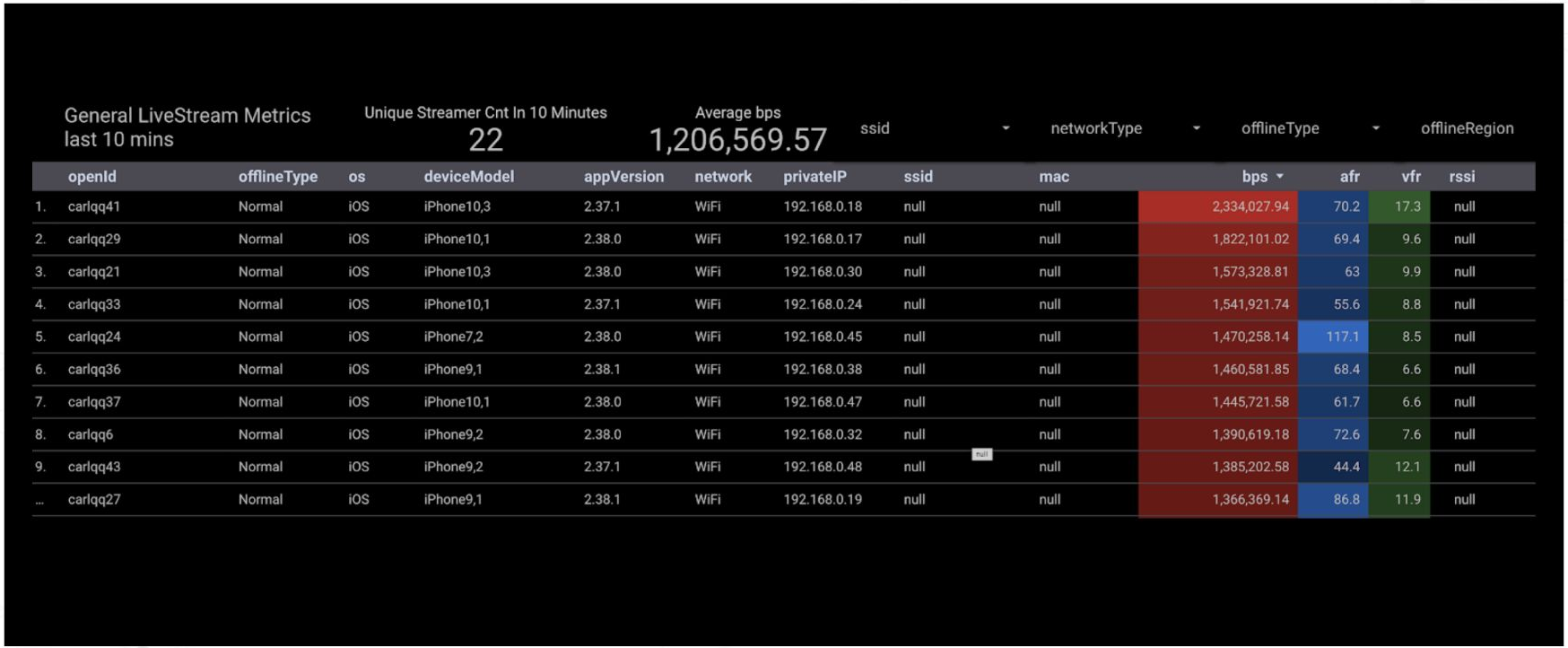
Real-Time Monitor
Streaming Performance

Case 1

Real-Time Monitor
Streaming Performance

Dashboard Screenshot

Streamer General Board



General LiveStream Metrics
last 10 mins

Unique Streamer Cnt In 10 Minutes
22

Average bps
1,206,569.57

ssid | networkType | offlineType | offlineRegion

	openId	offlineType	os	deviceModel	appVersion	network	privateIP	ssid	mac	bps	afr	vfr	rssi
1.	carlqq41	Normal	iOS	iPhone10,3	2.37.1	WiFi	192.168.0.18	null	null	2,334,027.94	70.2	17.3	null
2.	carlqq29	Normal	iOS	iPhone10,1	2.38.0	WiFi	192.168.0.17	null	null	1,822,101.02	69.4	9.6	null
3.	carlqq21	Normal	iOS	iPhone10,3	2.38.0	WiFi	192.168.0.30	null	null	1,573,328.81	63	9.9	null
4.	carlqq33	Normal	iOS	iPhone10,1	2.37.1	WiFi	192.168.0.24	null	null	1,541,921.74	55.6	8.8	null
5.	carlqq24	Normal	iOS	iPhone7,2	2.38.0	WiFi	192.168.0.45	null	null	1,470,258.14	117.1	8.5	null
6.	carlqq36	Normal	iOS	iPhone9,1	2.38.1	WiFi	192.168.0.38	null	null	1,460,581.85	68.4	6.6	null
7.	carlqq37	Normal	iOS	iPhone10,1	2.38.0	WiFi	192.168.0.47	null	null	1,445,721.58	61.7	6.6	null
8.	carlqq6	Normal	iOS	iPhone9,2	2.38.0	WiFi	192.168.0.32	null	null	1,390,619.18	72.6	7.6	null
9.	carlqq43	Normal	iOS	iPhone9,2	2.37.1	WiFi	192.168.0.48	null	null	1,385,202.58	44.4	12.1	null
...	carlqq27	Normal	iOS	iPhone9,1	2.38.1	WiFi	192.168.0.19	null	null	1,366,369.14	86.8	11.9	null

Case 2

User Behavior Events

Aggregation Table → Dataflow → Event Tables

Table Details: backendEvent

Schema | Details | Preview

timestamp	TIMESTAMP	REQUIRED	Describe this field...
eventName	STRING	NULLABLE	Describe this field...
eventDim	RECORD	REPEATED	Describe this field...
eventDim.**key**	STRING	NULLABLE	Describe this field...
eventDim.**value**	RECORD	NULLABLE	Describe this field...
eventDim.value.**string**	STRING	NULLABLE	Describe this field...
eventDim.value.**integer**	INTEGER	NULLABLE	Describe this field...
eventDim.value.**float**	FLOAT	NULLABLE	Describe this field...
eventDim.value.**boolean**	BOOLEAN	NULLABLE	Describe this field...
userID	STRING	NULLABLE	Describe this field...
isSandbox	BOOLEAN	NULLABLE	Describe this field...
insertTs	TIMESTAMP	NULLABLE	insert timestamp from fluentd

Add New Fields

- AutoFollow
- ClientPushNotification
- DailyQuestCheckin
- DailyUserExperience
- GetMission
- GetMissionReward
- GloryRoad
- LiveStreamActivity
- LiveStreamEventInfo
- LiveStreamInfo
- LoyaltyTitleUse
- NewbieQuestComplete
- NotifBatchActiveFollowerCount
- NotifBatchExcludeUnsubscribeUser
- NotifBatchFollowerCount
- NotifBeforeOneSignalAPI
- NotifTotalFollowerCount
- OneSignalCreateNotification
- PokeChallengeInfo
- PokeChallengeRoundInfo
- PokeChallengeSendGift
- PrivateMessageEnabled
- PrivateMessageSent
- ReferralVerify
- SendGift
- StampUse
- StreamComment
- StreamPushLogWithWifiInfo
- TriviaGetExtraLife
- TriviaQuestionInfo
- TriviaUserSubmit

Thank you

Demo:

Module Review

Module review

Match the use case on the left with the product on the right

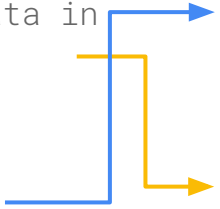
A.	Decoupling producers and consumers of data in large organizations and complex systems	1.	Cloud Dataflow
B.	Scalable, fault-tolerant multi-step processing of data	2.	Cloud Pub/Sub

Module review

Match the use case on the left with the product on the right

A. Decoupling producers and consumers of data in large organizations and complex systems

B. Scalable, fault-tolerant multi-step processing of data



1. Cloud Dataflow

2. Cloud Pub/Sub

Resources (1 of 2)

Cloud Pub/Sub	https://cloud.google.com/pubsub/
Cloud Dataflow	https://cloud.google.com/dataflow/
Processing media using Cloud Pub/Sub and Compute Engine	https://cloud.google.com/solutions/media-processing-pub-sub-compute-engine

Resources (2 of 2)

Reverse Geocoding of Geolocation Telemetry in the Cloud Using the Maps API	https://cloud.google.com/solutions/reverse-geocoding-geolocation-telemetry-cloud-maps-ap
Using Cloud Pub/Sub for Long-running Tasks	https://cloud.google.com/solutions/using-cloud-pub-sub-long-running-tasks

<Start Training>

#GoogleCloudOnBoard

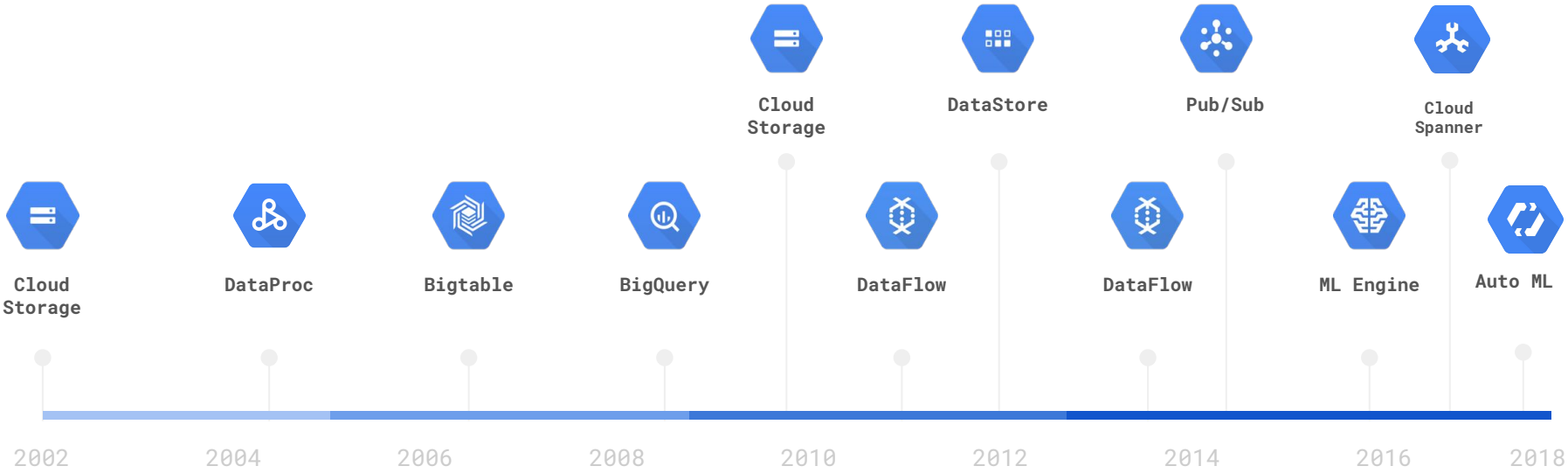
</Start Training>

Cloud OnBoard

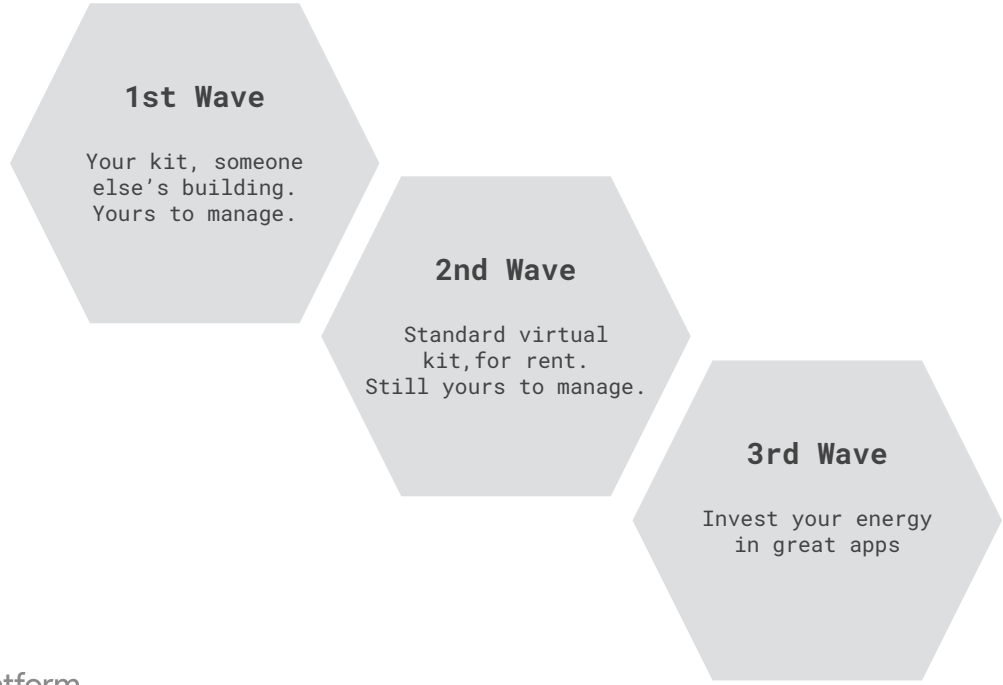
Summary

Cloud OnBoard

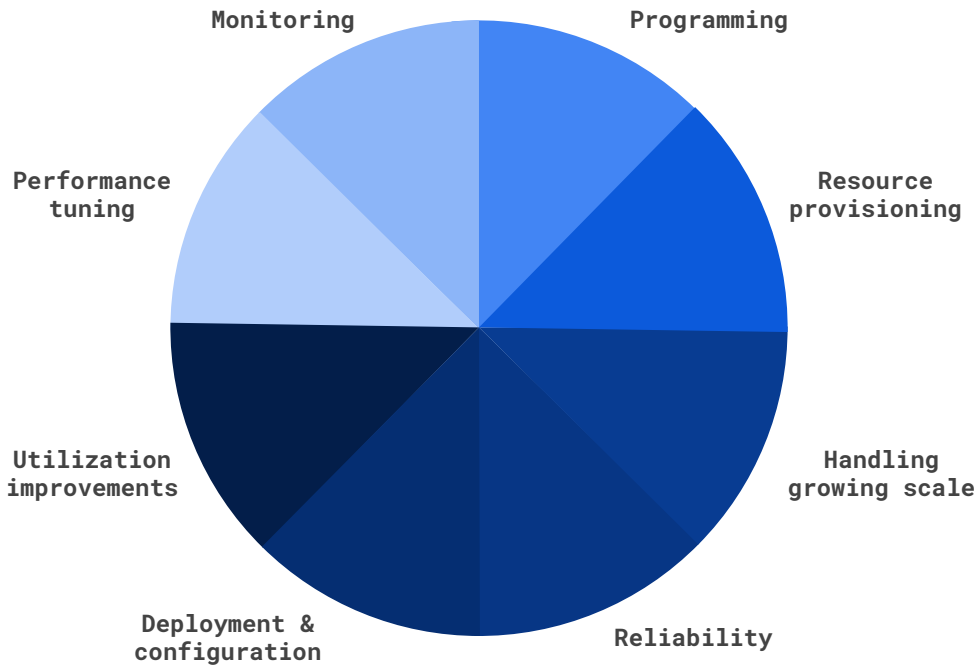
Google Cloud provides a way to take advantage of Google's investments in infrastructure and data processing innovation



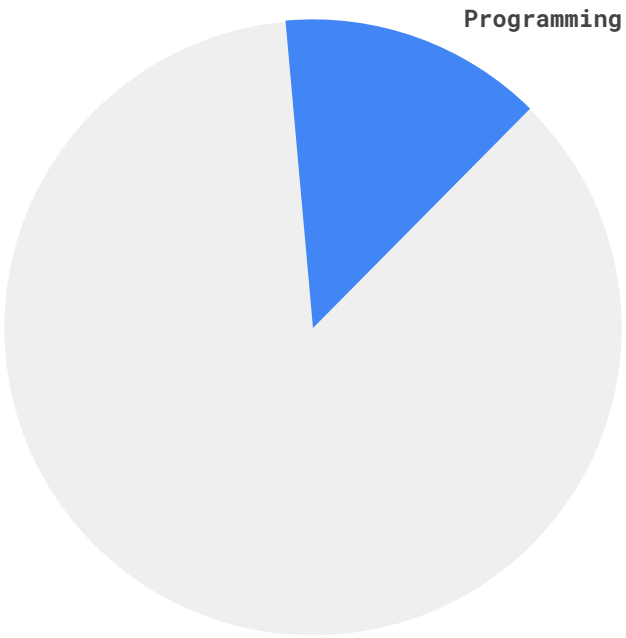
An Evolving Cloud



Typical Big Data Processing



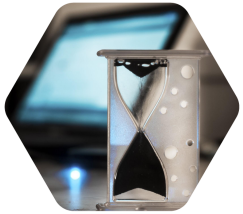
Big Data with Google: Focus on insight, not infrastructure.



In summary, GCP offers you ways to...

Spend less on ops and administration

We've "automated out" the complexity of building and maintaining data and analytics systems.

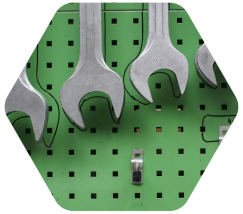


Incorporate real-time data into apps and architectures

To get the most out of data and secure competitive advantage.

Apply machine learning broadly and easily

We make it simple and practical to incorporate machine learning models within custom applications.



Create citizen data scientists

Transform your organization into a truly data driven company. Putting tools into hands of domain experts.

Resources

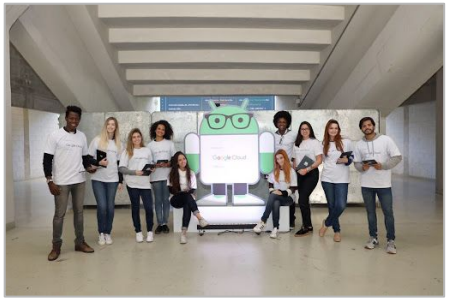
Big data and machine learning blog	https://cloud.google.com/blog/big-data/
Google Cloud Platform blog	https://cloudplatform.googleblog.com/
Google Cloud Platform curated articles	https://medium.com/google-cloud

Next Steps on your Google Cloud learning journey

1

Today

Google Cloud Platform Fundamentals: Big Data and Machine Learning

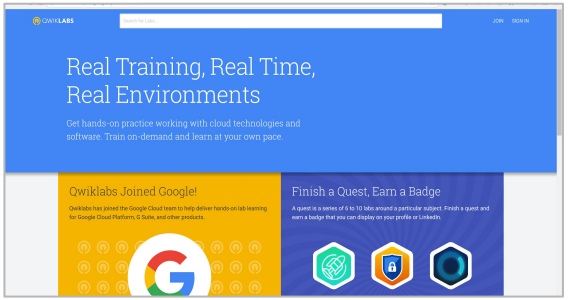


2

Tomorrow

Complete hands-on labs: Baseline: Data, ML, AI quest

google.qwiklabs.com



3

Future

Find more training online

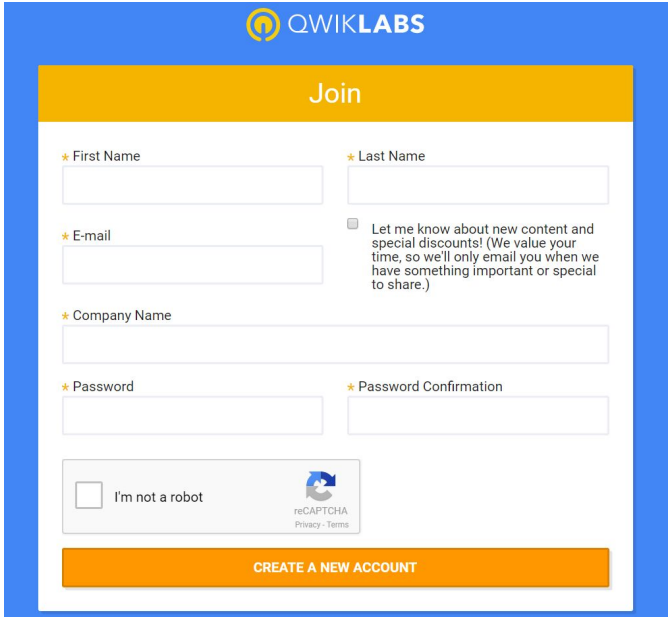
cloud.google.com/training

Google Cloud Cloud OnBoard

1 month of free access to Qwiklabs.

Complete the Baseline: Data, ML, AI Quest within your first month, and earn a **second free month of unlimited access**.

1. Receive a follow up email 1 day after event
2. Create Qwiklabs account with the email you used to register for Cloud OnBoard
3. Open your email and confirm account
4. Return to Qwiklabs and log in
5. Enroll in the Baseline: Data, ML, AI quest and take your first lab!



1 month free on Google Cloud Specializations on Coursera

How to claim your offer

1. Go to https://bit.ly/2IWgEPV
2. Click on one of the Google Cloud Specializations displayed on the Coursera promo landing page
3. Complete checkout. A credit card is required for checkout, but will not be charged for the first month and the subscription can be canceled any time.

Google Cloud

* Terms of promotion: 100% off your first month on select Google Cloud Specializations on Coursera. Limit one use per user. Valid until December 31, 2019. While supplies last.

The promotion can be used on the following Google Cloud Specializations

Architecting with Google Cloud Platform	SDP
Advanced Machine Learning with TensorFlow on Google Cloud Platform	SDP
Data Engineering on Google Cloud Platform	SDP
Developing Applications with Google Cloud Platform	SDP
From Data to Insights with Google Cloud Platform	SDP
Machine Learning with TensorFlow on Google Cloud Platform	SDP
Networking in Google Cloud Platform Specialization	SDP
Security in Google Cloud Platform	SDP

Become a Google Certified Professional Data Engineer to demonstrate your ability to enable data-driven decision making

Take training courses	Practice with Qwiklabs	Use GCP in your work	Schedule your exam	Get certified!
Get up to speed and learn best practices. Classes focus on technical skills covered in the exam.	Use Qwiklabs to get hands on experience with Google Cloud technologies.	Use GCP in your work and draw on that experience to prepare for the exam.	Register and find a location near you.	Demonstrate your ability to design, build, maintain, and troubleshoot data processing systems.

https://cloud.google.com/certification/data-engineer

<Start Training>

#GoogleCloudOnBoard

</Start Training>

<Thank_You!>

See you next time

</Thank_You!>