Architecting the Future

Google Cloud Platform

Edward Doan
edoan@google.com | @edwardd
Agenda

01 How is Google’s Cloud Different?
02 What can you do with GCP?
03 Getting Started
04 Q&A
Why are we here today?
Hitting the limits early on

The Anatomy of a Large-Scale Hypertextual Web Search Engine

1996, Sergey Brin and Lawrence Page
Computer Science Department, Stanford University, Stanford, CA 94305
Google Cloud brings proven tech from Google products

- **Search**
  - Search ranking
  - Speech recognition

- **Gmail**
  - Smart Reply
  - Spam classification

- **Photos**
  - Photos search

- **Translate**
  - Text, graphic, and speech translations

- **Android**
  - Keyboard & speech input

- **Drive**
  - Intelligence in Apps

- **Play**
  - App recommendations
  - Game developer experience

- **Chrome**
  - Search by Image

- **Maps**
  - Street View image
  - Parsing Local Search

- **YouTube**
  - Video recommendations
  - Better thumbnails

- **Cardboard**
  - Smart stitching

- **Ads**
  - Richer Text Ads
  - Automated Bidding
Does your server room / data center look like this?
580,000 cores on preemptible VMs

“$10^{17}$ hyperelliptic curves of genus 3 in an effort to find curves whose L-functions can be easily computed, and which have potentially interesting Sato-Tate distributions”

Products used: Google Compute Engine, Cloud Storage, DataStore
Spotify

75M+ Users
2B+ Playlists
30M+ Songs

Products used: App Engine, BigQuery, Dataflow, Dataproc and Pub/Sub
Snapchat

700 million photos and videos per day

Google App Engine scaled seamlessly during growth

Small team is able to innovate quickly, globally
Google Network

The largest cloud network, comprised of 100,000s of miles of fiber cable and 8 subsea cables.
Advancing the state of the art.

MapReduce: Simplified Data Processing on Large Clusters
Jeffrey Dean and Sanjay Ghemawat
jdean@google.com, sghemawat@google.com
Google, Inc.

Abstract

MapReduce is a programming model and an associated implementation for processing and generating large data sets. The model was designed specifically to process geographical data, where a job consists of a set of tasks that can be run in parallel. The MapReduce model makes it easy to process large data sets using simple programming model and an associated implementation for processing and generating large data sets. The model was designed specifically to process geographical data, where a job consists of a set of tasks that can be run in parallel. The MapReduce model makes it easy to process large data sets using simple programming model and an associated implementation.

Bigtable: A Distributed Storage System for Structured Data
Fuy Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Heik, Deborah A. Wallace, Mike Armstrong, Tudor Cristian Cotrupe, Sanjay Kumar, Robert E. Gruber
jdean@google.com, sghemawat@google.com
Google, Inc.

Abstract

Bigtable is a distributed storage system for managing structured data that is designed to scale to a very large number of data sources across thousands of commodity servers. Bigtable provides an abstraction that simplifies the task of scaling by providing a consistent interface to large collections of data across a network. Bigtable's design is based on a layered approach to data storage and retrieval. Bigtable's key-value interface provides an efficient way to store and retrieve data. Bigtable is designed to be scalable and fault-tolerant, and supports a wide range of data storage applications. Bigtable is designed to be scalable and fault-tolerant, and supports a wide range of data storage applications.

Distributed: Interactive Analysis of Web-Scale Datasets
Sergey Melnik, Andrey Gubarev, Jingjing Long, Geoffrey Stormer, Shane Silverman, Xuan V. Tran
melnik@alum.mit.edu

ABSTRACT

Distributed is a scalable, interactive analytic query system for analyzing and storing web-scale data. Distributed is the first system to provide interactive access to trillions of data points. Distributed provides a scalable and fault-tolerant architecture that enables efficient, interactive analysis of web-scale datasets. Distributed is designed to be scalable and fault-tolerant, and supports a wide range of data storage applications. Distributed is designed to be scalable and fault-tolerant, and supports a wide range of data storage applications.

1. INTRODUCTION

Large scale analytic data processing has become ubiquitous in web applications and services, and has led to a focus on large scale analytic data processing. Distributed is designed to be scalable and fault-tolerant, and supports a wide range of data storage applications. Distributed is designed to be scalable and fault-tolerant, and supports a wide range of data storage applications.
Bigtable: A Distributed Storage System for Structured Data

Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallace, Mike Burrows, Tushar Chandra, Andrew Colk, Robert E. Gruber

Google, Inc.

Abstract

Bigtable is a distributed storage system for managing structured data that is designed to scale on very large clusters of commodity servers. It is built on top of the Google File System (GFS) and the Google cluster management system (HBase). Bigtable shares many design and implementation principles with GFS and HBase.

1 Introduction

Bigtable is a distributed data management system that is designed to store and retrieve large amounts of structured data. It is built on top of the Google File System (GFS) and the Google cluster management system (HBase). Bigtable shares many design and implementation principles with GFS and HBase.

2 Data Model

A Bigtable is a sparse, distributed, persistent, multi-dimensional, and multi-user data store. The data is organized into a set of tables, each of which is composed of multiple columns. Each column is associated with a set of values, called a cell. The values in each cell are stored as a set of key-value pairs.

...and creating a revolution
Easy to deploy Google Innovation

- DataProc: MapReduce (2004)
- BigTable (2006)
- BigQuery (2008)
- Dremel (2008)
- Cloud Storage: Colossus (2008)
- DataFlow: Flume (2010)
- DataStore: Megastore (2010)
- Spanner (2012)
- PubSub: PubSub (2014)
- ML: TensorFlow (2016)
02 What Can You Do with GCP?
Google Cloud Platform

Compute
- App Engine (PaaS)
- Compute Engine (IaaS)
- Container Engine (powered by Kubernetes)

Storage
- Cloud Storage
- Cloud SQL
- Cloud Datastore
- Big Table

Big Data
- BigQuery
- Pub/Sub
- Dataflow
- Dataproc
- Datalab

Machine Learning
- Cloud ML
- Cloud Vision
- Translate API
- Speech API
Compute: A Continuum

- **Compute Engine**: Virtualized hardware at Google speed.
- **Container Engine**: Manages your container cluster and actively schedules your containers.
- **App Engine**: Build your scalable app faster.

**IaaS**

**PaaS**
High-performance Virtual Machines

(Really) Pay for what you use

Fast, Easy Provisioning

High-performance Virtual Machines

Batch Computing

Compliance & Security
Hardware accelerated with GPUs

- AMD FirePro S9300 x2
  Coming Soon
- NVIDIA® Tesla® P100s
  Coming Soon
- NVIDIA® Tesla® K80s

$0.77 ON DEMAND PRICE GPU / HOUR for K80
Storage and Databases

Blob
- Cloud Storage
- Good for: Structured and unstructured binary or object data
  - Such as: Images, large media files, backups

NoSQL
- Cloud Datastore
- Good for: Getting started, App Engine, serve use cases
  - Such as: User profiles, product catalog

- Cloud Bigtable
- Good for: Heavy read + write, events, and analytical data
  - Such as: AdTech, Financial and IoT data

SQL
- Cloud SQL
- Good for: Web frameworks, existing applications
  - Such as: User credentials, customer orders
But what does it all mean?

- Do we have any actionable analytics from our big data in the cloud?
- Yes, the data shows that my productivity plunges whenever you learn new jargon.
- Maybe in-memory computing will accelerate your applications.
- Plunge, plunge, plunge.
Big Data and Analytics
BigQuery

- Fully Managed, **No-Ops** Data Warehouse
- Petabyte-Scale and Fast
- Convenience of SQL
- Externalization of Google Dremel
Fun BigQuery Stats

- Largest query by rows: 10.5 Trillion rows
- Largest query by data size: 2.1 Petabytes
- Largest storage customer: 62 Petabytes
- Streaming ingest at peak: 4.5 Million rows per second
BigQuery - explained

1. We just rented ~2000 cores from Google for ~12 seconds
2. We only paid ~$0.13
3. Most importantly, complexity is hidden from end user
4. Users do not think about cores
Machine Learning
Machine Learning is...

One branch of the field of Artificial Intelligence

A way of solving problems without explicitly codifying the solution

A way of building systems that improve themselves over time

Source: Neota Logic
The Machine Learning Spectrum

Use/extend OSS SDK  Build custom models  Use pre-built models

ML researcher  Data Scientist  App Developer

CloudML

Scale, No-ops Infrastructure
TPU - 7 years ahead of GPU in terms of price/performance

Google Trained Models

Cloud Translate  Cloud Vision  Cloud Natural Language  Cloud Speech

Google Trained Models: Natural Language, Speech, Vision

TensorFlow

Build custom models

CloudML

Scale, No-ops Infrastructure
TPU - 7 years ahead of GPU in terms of price/performance

Google Trained Models: Natural Language, Speech, Vision

Google Cloud

Use/extend OSS SDK  Build custom models  Use pre-built models

ML researcher  Data Scientist  App Developer
Democratising ML with Pre-Trained Models

Fully trained ML models from Google Cloud that allow a general developer to take advantage of rich machine learning capabilities with simple REST based services.

Cloud Translate  Cloud Vision  Cloud Natural Language  Cloud Speech  Cloud Video Intelligence  Cloud Jobs API  Stay tuned...
What do you think that cloud looks like?

Google → Search by image

[UPLOADING...]

Keep trying, Google.

Best guess for this image: CLOUD
Vision API Demo
Cloud Natural Language API

- Call API from anywhere, with support for embedded Text, and Google Cloud storage
- Support for English, Spanish and Japanese
Cloud Translate API

Dynamically translate between thousands of available language pairs
Built on Open Source

Created by Google Brain team

Most popular ML project on Github
- Over 480 contributors
- 10,000 commits in 12 months

Multiple deployment options:
- Mobile, Desktop, Server, Cloud
- CPU, GPU
03 Getting Started
GCP Computer Science credits

Teaching

Faculty in select countries + Teaching university courses + In computer science or related fields

@bagibson
Redeem the coupon by going to https://console.cloud.google.com/education

Education grants
Please enter the coupon code provided to you via the Google Cloud Platform Education Grants program to receive credit for Google Cloud Platform. Get what you need to build and run your apps, websites and services.

Coupon code
2REJFPEN-BU7V-7EVL

Credit amount | Expiration date | Course
$1.00         | Jun 1, 2018     |

Please email me updates regarding feature announcements, performance suggestions, feedback surveys and special offers.

☐ Yes  ☐ No

Google Cloud Platform education grants credits terms and conditions
By clicking “Accept and continue” below, you, on behalf of yourself and the organization you represent (“You”) agree to these terms and conditions:
Upon redeeming, you’ll be taken to the GCP console and can get started:

<table>
<thead>
<tr>
<th>DASHBOARD</th>
<th>ACTIVITY</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project info</strong></td>
<td><strong>Compute Engine</strong></td>
</tr>
<tr>
<td>Project name</td>
<td>CPU (%)</td>
</tr>
<tr>
<td>TEST</td>
<td></td>
</tr>
<tr>
<td>Project ID</td>
<td>extended-aroon-142620</td>
</tr>
</tbody>
</table>

![Google Cloud Platform Console](image-url)
Create a Project

1. Go to Create a project
2. Give your project a name
3. Select the billing account that was created in the previous step
Create a Linux VM

1. In the Cloud Platform Console, go to the VM Instances page. (Compute Engine -> VM Instances)
2. Click the Create instance button.
3. In the Boot disk section, click Change to begin configuring your boot disk.
4. In the OS images tab, choose the Debian 8 image.
5. Click Select.
6. In the Firewall section, select Allow HTTP traffic.
7. Click the Create button to create the instance.
Connect to your instance

1. In the Cloud Platform Console, go to the **VM Instances** page.
2. In the list of virtual machine instances, click **SSH** in the row of the instance that you want to connect to.

You now have a terminal window for interacting with your Linux instance.

Don’t forget to suspend or delete your instance to avoid incurring charges!
Codelab: Deploy a Bookshelf application to App Engine