

# Summary

Google Cloud Platform Fundamentals: Big Data and Machine Learning

---

Version #1.1

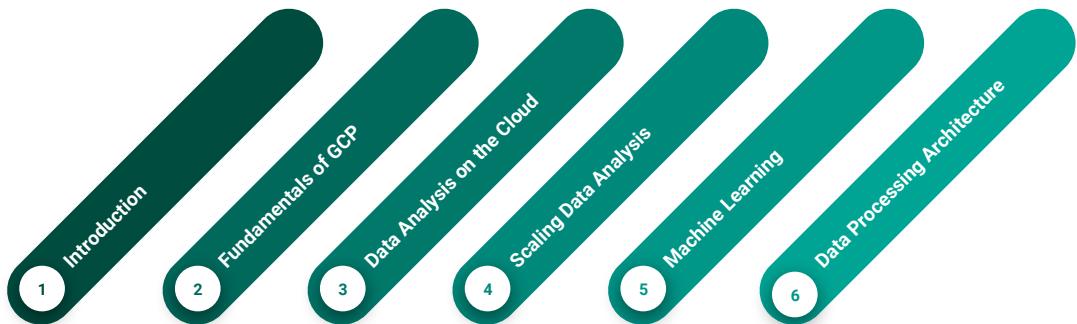


© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

## Notes:

30 minutes

# Agenda



## Notes:

### 1. Introduction

Overview of GCP as a whole, but with emphasis on the data-handling aspects of the platform

- GCP, GCP Big Data
- Usage scenarios
- Create an account on GCP

### 2. Foundation of GCP

Compute and Storage with a focus on their value in data ingest, storage, and federated analysis

- Compute Engine
- Cloud Storage
- Start GCE instance
- Upload data to GCS

### 3. Data analytics on the Cloud

Common use cases that Google manages for you and for which there is an easy migration path to the Cloud

- Cloud SQL
- Dataproc
- Import data into and query Cloud SQL
- Machine Learning with Dataproc

In the morning, we will complete Modules 1 and 2 and get halfway through Module 3.

#### 4a. Scaling data analysis

Change how you compute, not just where you compute with GCP

- Datalab
- Datastore, Big Table
- BigQuery

#### 5. TensorFlow

Change how you compute, not just where you compute with GCP

- TensorFlow
- Datalab instance
- BigQuery
- Demand forecasting with ML

#### 6. Data processing architectures

Scalable, reliable data processing on GCP

- Pub/Sub
- Dataflow

#### 7. Summary

Course summary

- Resources

Please feel free to use the appendixes for self-study.

In the morning, we will get halfway through Module 3.

Please feel free to use the appendixes for self-study.

Please take 5 minutes to give us feedback

[g.co/CloudTrainEval](https://g.co/CloudTrainEval)

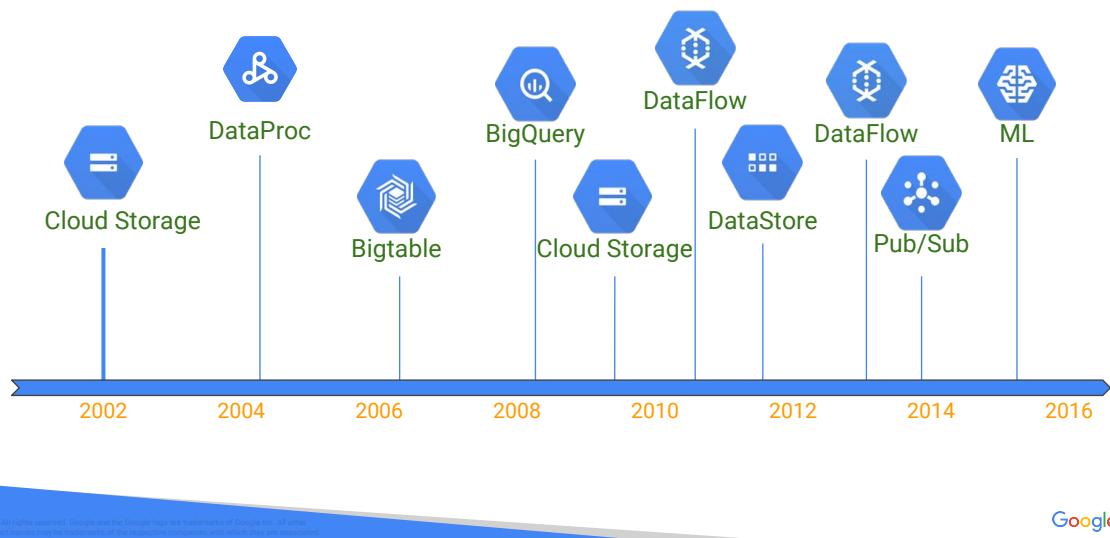
This URL is case-insensitive

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.



The class code is always a tag in the QL class. Instructor can always find the class code (if they don't have it already) when they are in the class in QL, click Edit Class and the code is in the tags. Not ideal - but we'll figure out how to make it more prominent

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



### Notes:

Google Research Publications referenced are available here:

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

The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines, 2009

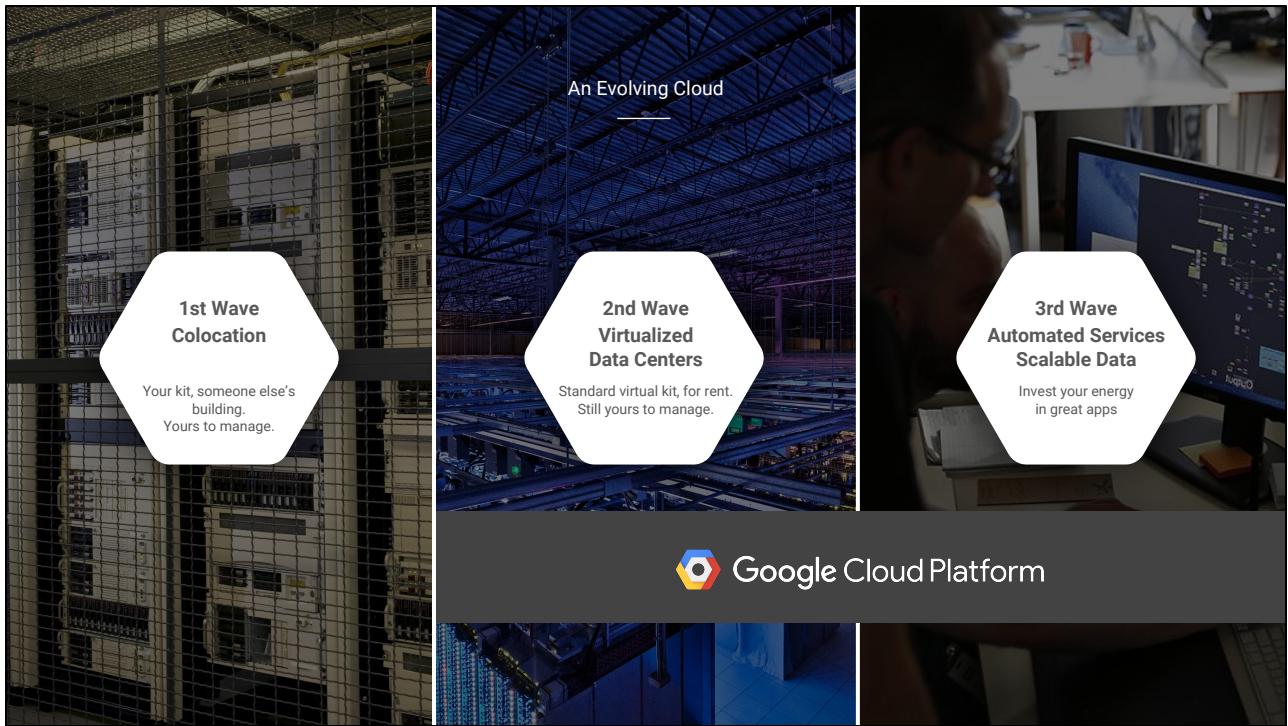
<http://research.google.com/pubs/pub35290.html>

So, the mode is now to provide the exact implementations that Google uses, and give you a way to use them directly. The APIs are open-sourced, but not Google's implementations (the Apache Beam/DataFlow model). Starting with Bigtable, there are no exact equivalents any more. (Bigtable != HBase/MongoDB and BigQuery != Amazon RedShift).

<http://db-engines.com/en/system/Google+Cloud+Bigtable%3BHBase%3BMongoDB>: The main difference is that Bigtable is no-ops (hosted). It is also more performant for very, very large databases.

<https://www.quora.com/How-good-is-Google's-BigQuery-as-compared-to-Amazons-Redshift>: The differences here are similar. BigQuery is no-ops where Amazon Redshift requires provisioning. The quora answer by Peter Mueller says what the bloodless word "provisioning" means in practice -- They move data from Amazon S3 to Google Cloud Platform just so they don't have to

worry about determining how much hardware they need.



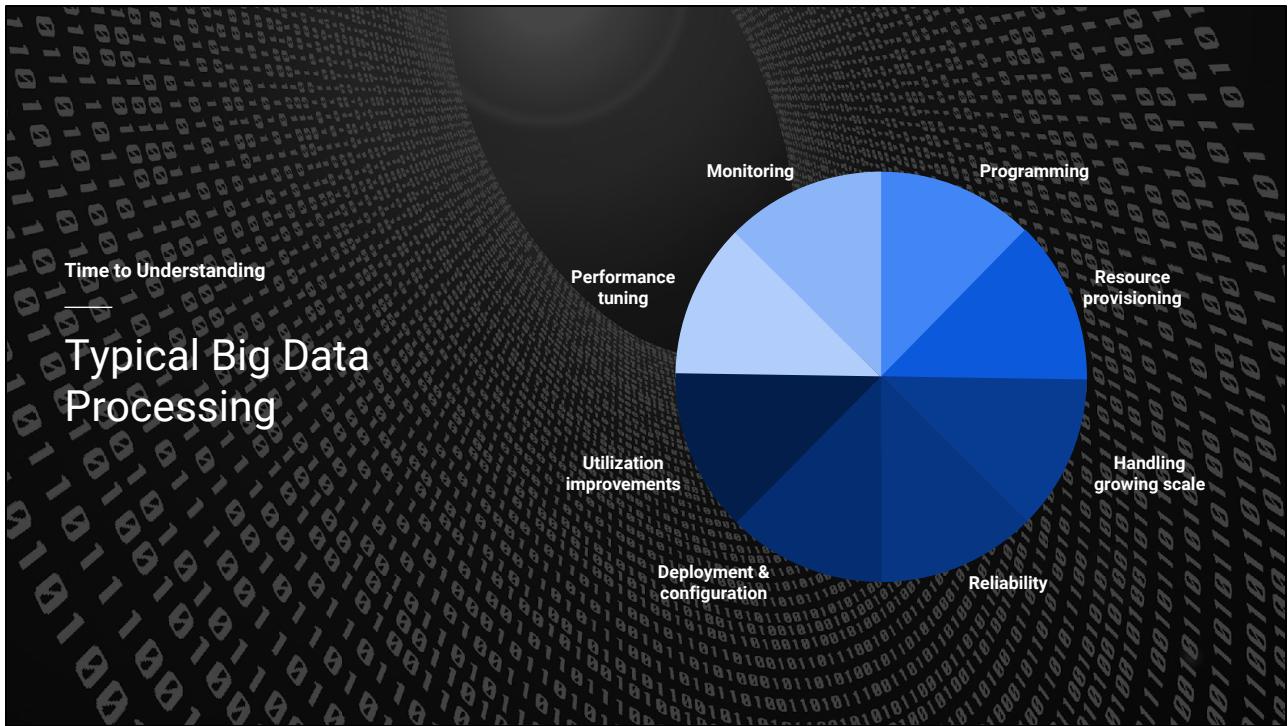
## Notes:

**Google Cloud Platform bridges the 2nd wave of cloud with the 3rd wave.**

**Most organizations are still working in the virtualized server world of the 2nd wave**

Google Cloud Platform lets you **build out from existing investments to take advantage of new technologies like Containers, Platform-as-a-service, and horizontally scalable data**, all while taking advantage of Google's best-in-class cloud infrastructure.

Using this approach, **organizations can achieve unprecedented software developer productivity**, while improving performance and reducing cost of running existing applications

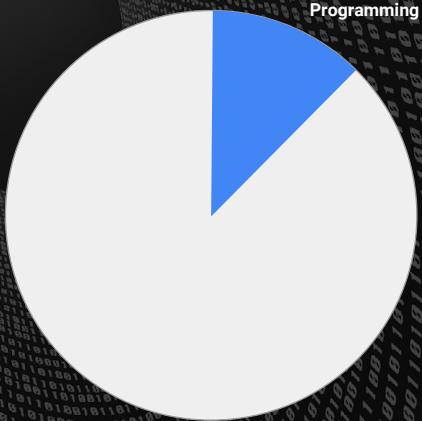


### Notes:

- Then what? How do you find the needle in the haystack of data? Traditional DB not up to the challenge of data on this scale.
- Most BigData today is Map Reduce - a research paper originally published by Google back in 2004
- But if you look at most big data projects and really look and where time and money is being spent, you see that most of the time isn't spent getting insights from the data
- it's spent on the care and feeding of the machinery - managing infrastructure, manipulating data, monitoring and performance

Time to Understanding

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



### Notes:

- We realized this ourselves at Google several years ago, so we started developing systems that let us:
- 1) **scale with your data growth** even as it explodes
- 2) are **managed** so that you aren't wasting time on dealing with all of the underlying complexities
- 3) are just generally **magically awesome** so you can get back to **data insights, and not data mgmt**
- Let's take a look at 2 specific services

## 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.

## We offer courses and labs to meet your ongoing learning needs

Instructor-led training courses	Self-paced labs
<p><b>From Data to Insights with Google Cloud Platform*</b> Learn how to derive insights through data analysis and visualization</p> <ul style="list-style-type: none"> <li>1. Exploring and Preparing your Data</li> <li>2. Visualizing Insights</li> <li>3. Advanced Insights with BigQuery</li> <li>4. Applying ML to your data</li> </ul> <p><b>For data analysts</b></p> <p><a href="https://cloud.google.com/training/data-ml">https://cloud.google.com/training/data-ml</a></p> <p>*Also available on-demand at coursera.org</p>	<p><b>Data Engineering on Google Cloud Platform*</b> Learn how to use the big data and machine learning capabilities of GCP.</p> <ul style="list-style-type: none"> <li>1. Leveraging unstructured data</li> <li>2. Data Analysis (BigQuery/Dataflow)</li> <li>3. Serverless ML with Tensorflow</li> <li>4. Resilient Streaming Systems</li> </ul> <p><b>For data engineers</b></p>

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.



### Instructor led courses:

**Data Analyst Track: From Data to Insights with Google Cloud Platform:** Learn how to derive insights through data analysis and visualization using the Google Cloud Platform.

**Data Engineer Track: Google Cloud Platform Fundamentals: Big Data & Machine Learning:** This course introduces participants to the big data capabilities of Google Cloud Platform.

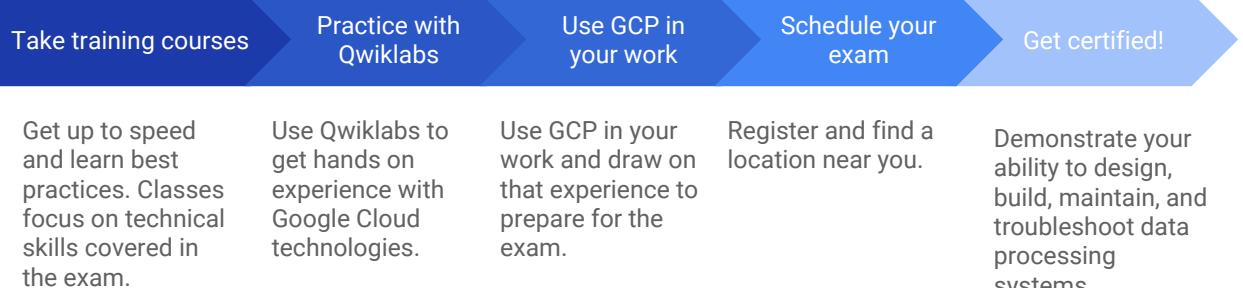
### Self paced labs:

**Data engineering quest on Qwiklabs.** This is a series of 9 labs that allows hands on experience with Data Analysis and ML on Google Cloud

Online Courses:

**Data Engineering on Google Cloud Platform Specialization:** A series of courses offered online by Coursera.

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



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

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.



# 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>

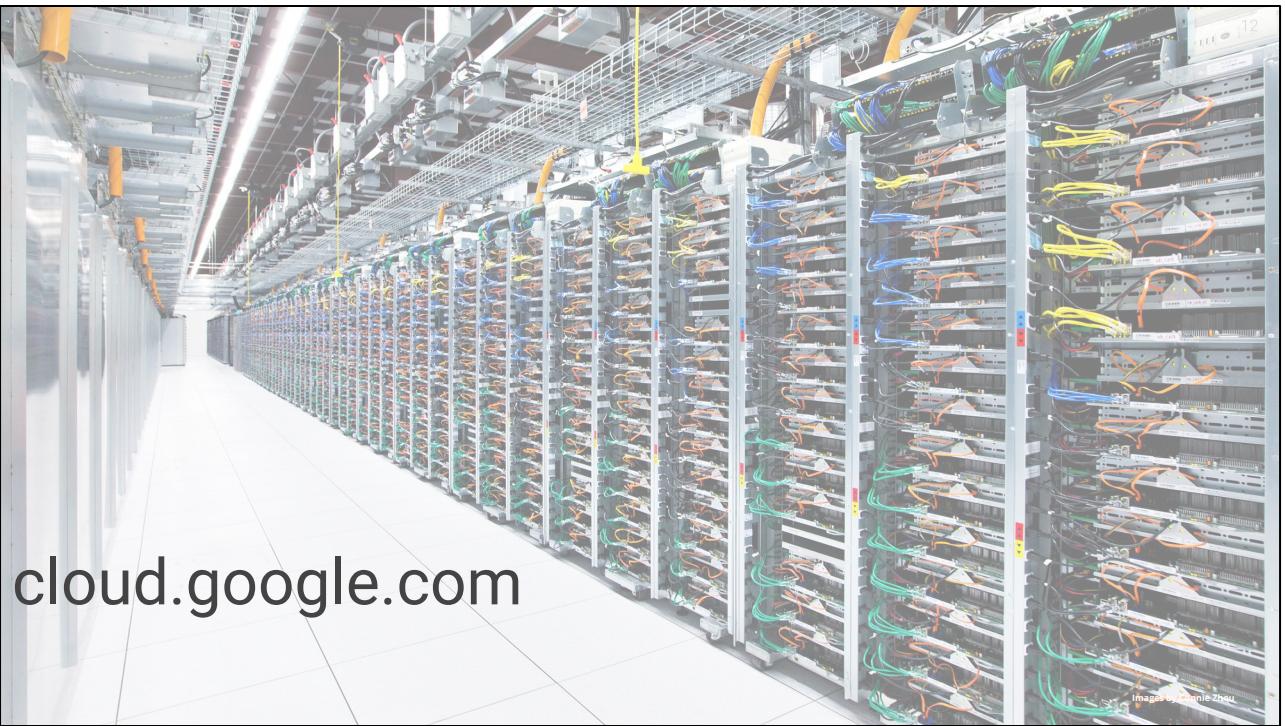
---

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.



## Notes:

Point out that the third is a community; they can write articles in medium and submit their article to google-cloud



cloud.google.com