Cut Video

Course 1: Exploring and Preparing your Data with BigQuery

Module 1: Introduction

Lesson Title: **Introduction**

Format: Talking head with slides

Video Name: xxx

From Data to Insights with Google Cloud Platform

v1.0

Facilities

Parking

Facilities

Food

Course etiquette

Please silence your phone and take calls outside

Recording this class is prohibited

Ask questions interactively or via chat (online)

Agenda

1- Introduction to Data on the Google Cloud Platform

2 - Big Data Tools

3 - Exploring your Data with SQL in BigQuery

4 - Pricing

Agenda

5 - Cleaning and Transforming Data

6 - Storing and Exporting Data

7 - Ingesting New Datasets

8 - Visualization Basics with Data Studio

Agenda

9 - Joining and Merging Datasets

10 - Advanced Clauses and Functions

11 - Schema Design and Nested Data Structures

12 - Advanced Visualization with Google Data Studio

Agenda

13 - Optimizing for Performance

14 - Advanced Insights with Cloud Datalab

15 - Data Access

Audience and Prerequisites

Target Audiences

1. Data Analysts, Business Analysts, Business Intelligence professionals

2. Data Engineers who will be partnering with Data Analysts to build scalable data solutions on Google Cloud Platform

Prerequisites

1. Basic Knowledge of SQL

Introductions

Your instructor

- Organization
- Background
- Course goals

You

- Name
- Organization
- Job role
- Course goals

Pay special attention to slides with key messages or pitfalls

Module 1

Introduction to Data on the Google Cloud Platform

In this module we will:

- **Highlight Analytics Challenges Faced by Data Analysts**
- Compare Big Data On-Premise vs on the Cloud
- Learn from Real-World Use Cases of Companies Transformed through Analytics on the Cloud
- Navigate Google Cloud Platform Project Basics

Data analysts face query, infrastructure, and storage challenges

"My queries are taking way **too long** to run and is stalling my analysis."

"We're a data department, not an **infrastructure** department. Maintaining and upgrading our own servers is unsustainable."

"We can only **afford to store a subset** of the data our business generates"

"I have no easy way to **combine and query** all the data I've collected"

"My on premise clusters **aren't scaling** with my analysis"

"We don't have a **central data** analytics warehouse or set of tools"

Module 1

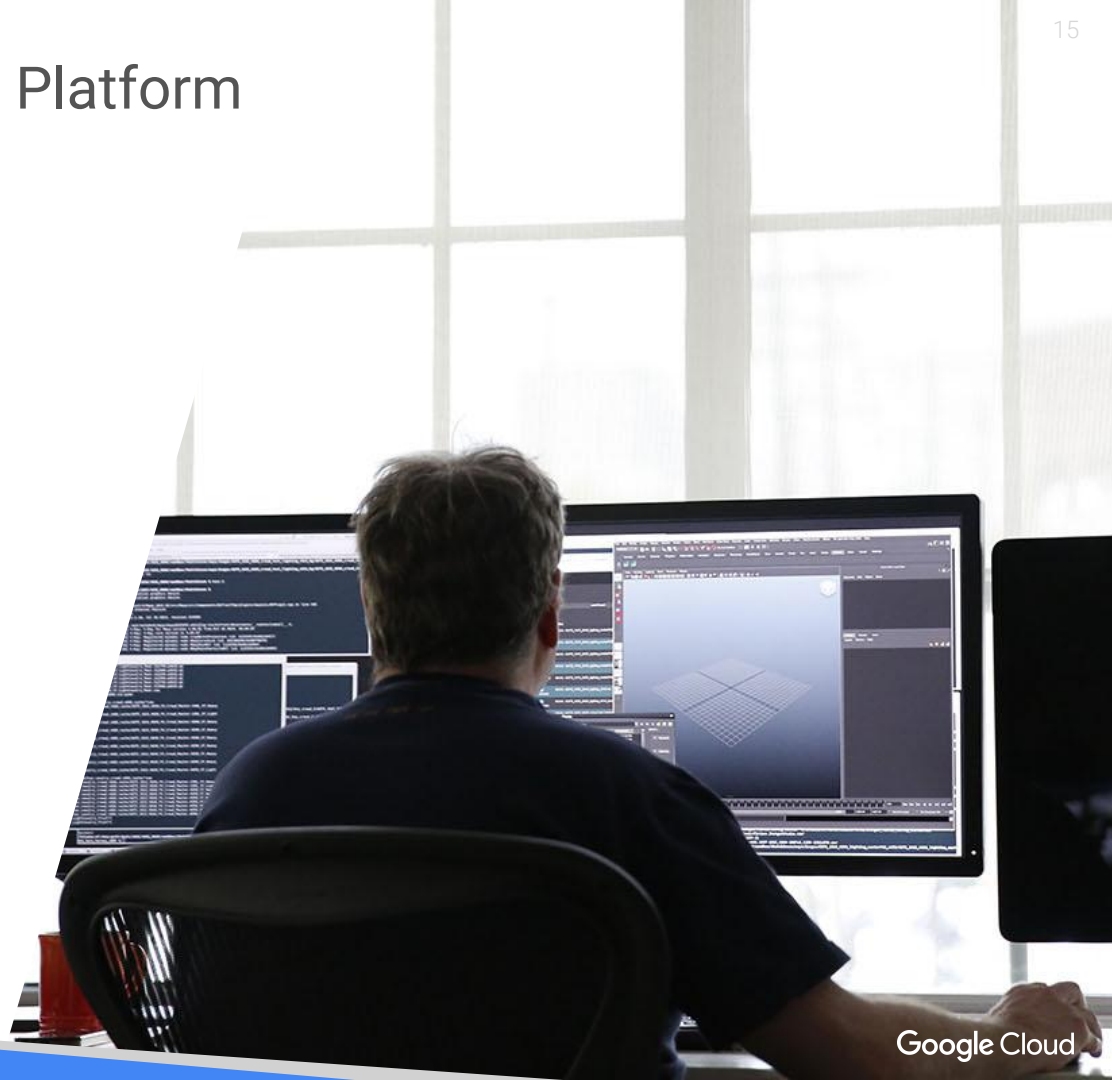
Introduction to Data on the Google Cloud Platform

In this module we will:

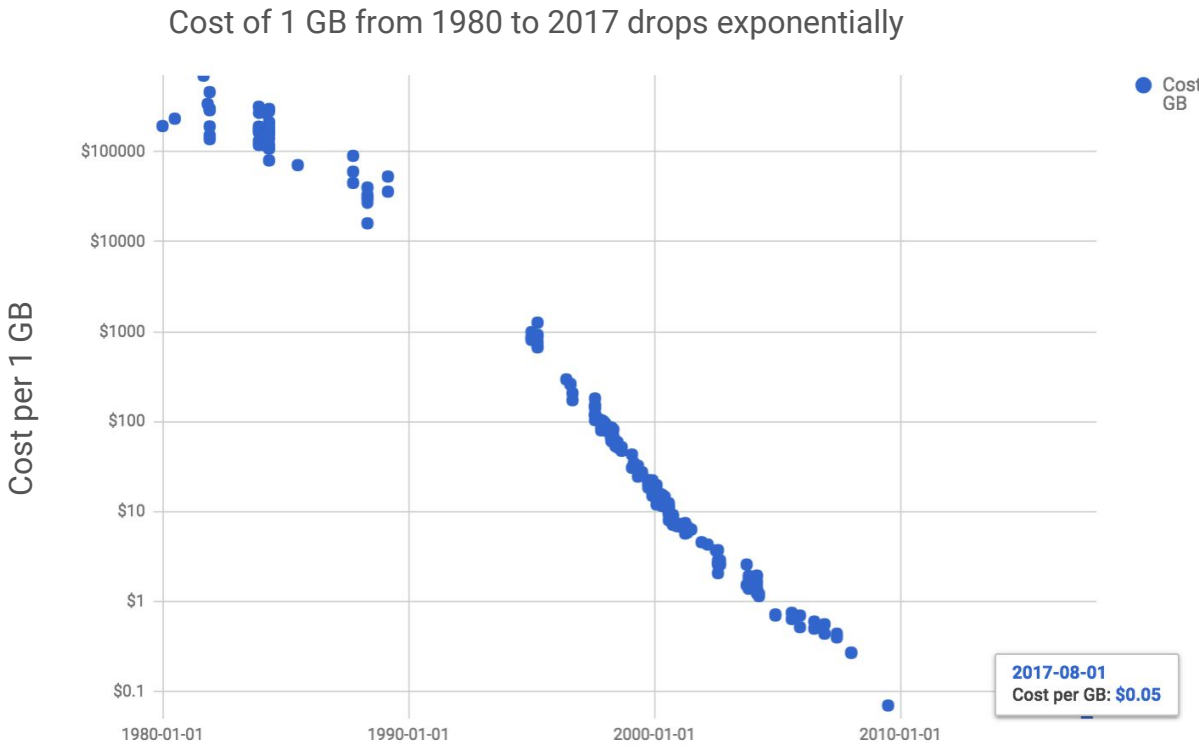
- Highlight Analytics Challenges Faced by Data Analysts
- **Compare Big Data On-Premise vs on the Cloud**
- Learn from Real-World Use Cases of Companies Transformed through Analytics on the Cloud
- Navigate Google Cloud Platform Project Basics

Reasons why Google Cloud Platform is used for Data Analysis

- Storage is Cheap
- Focus on Queries, not Infrastructure
- Massive Scalability

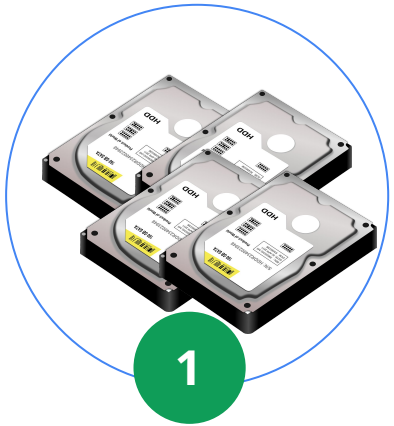


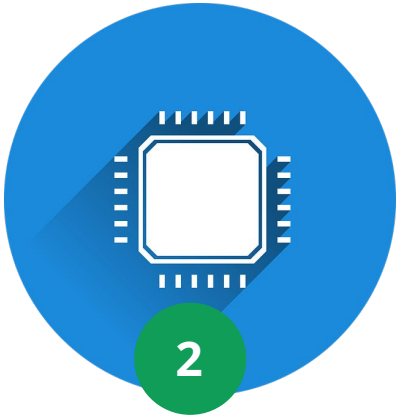
The cost of 1GB of storage has dropped dramatically

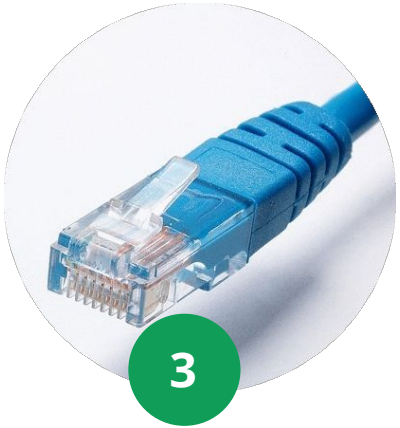


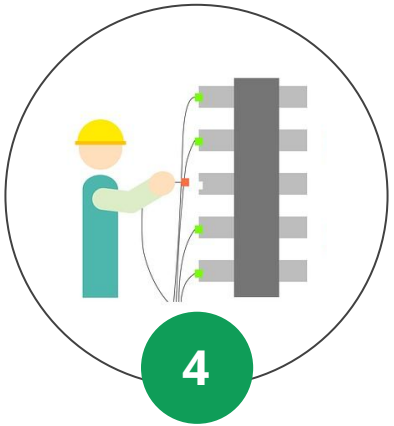
Google Cloud

Traditional big data platforms require an investment in infrastructure



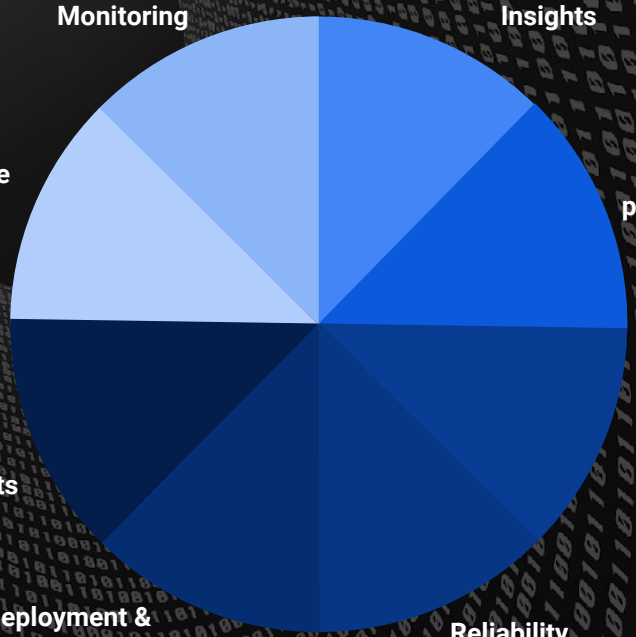






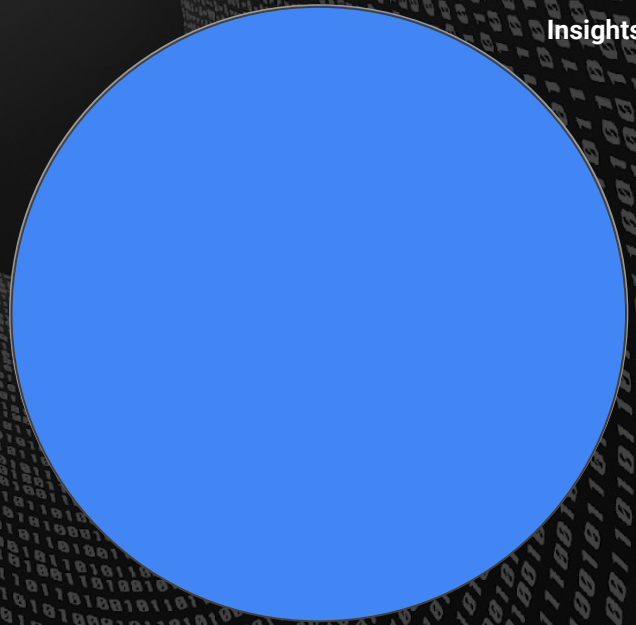
Time to Understanding

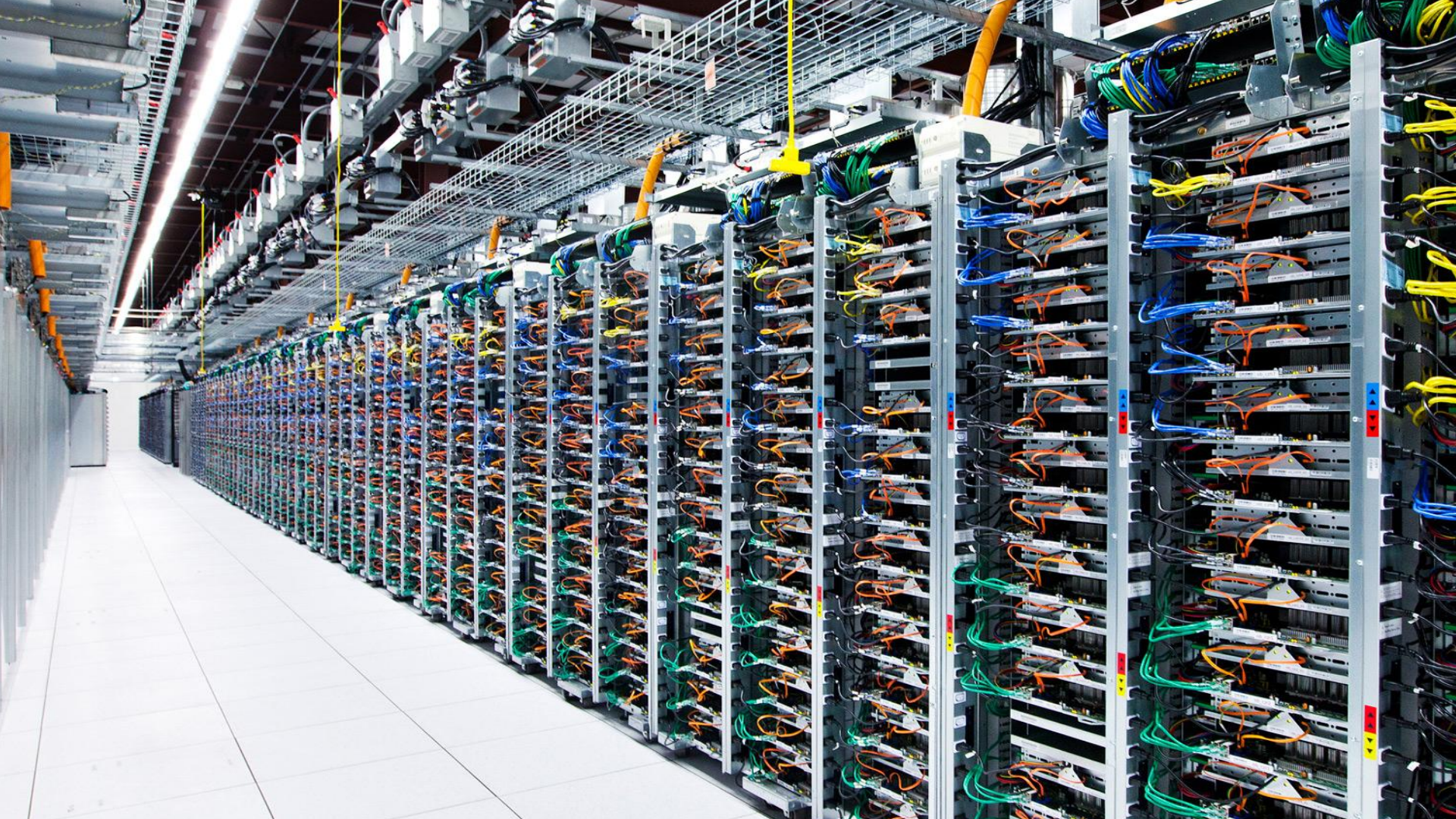
Typical Big Data Processing



Time to Understanding

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

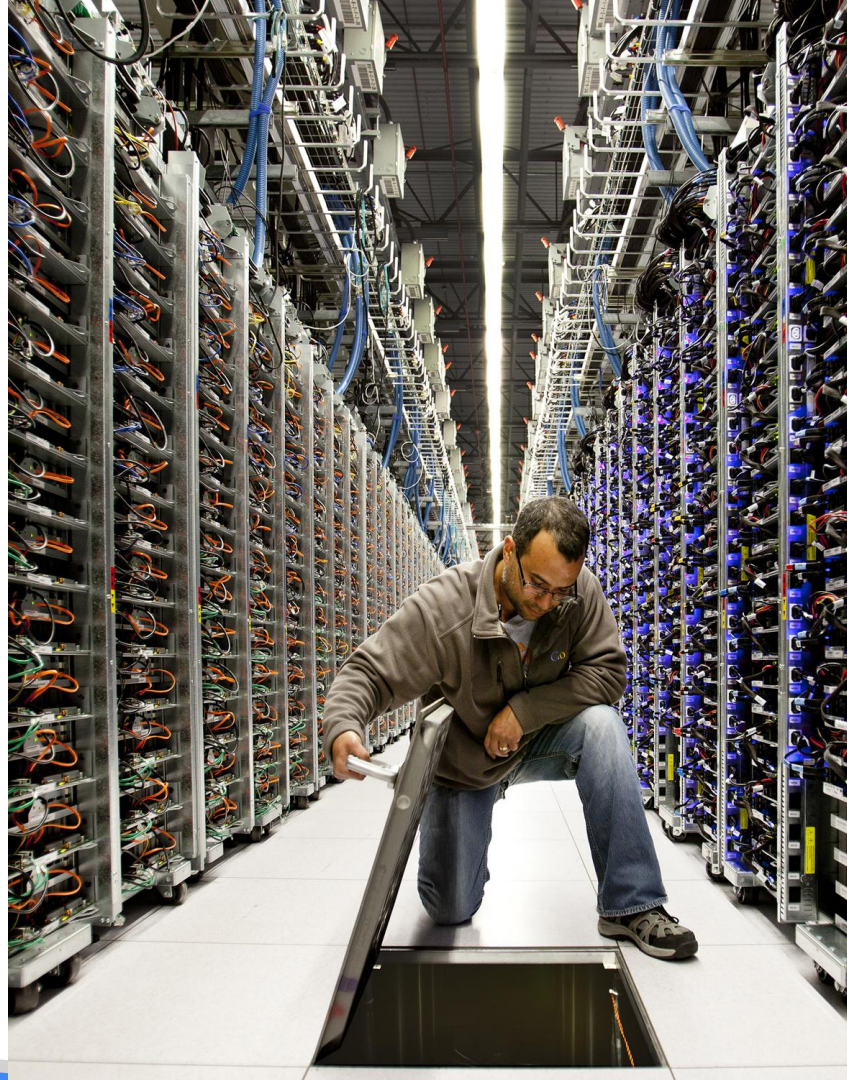




“[Google's] ability to build, organize, and operate a huge network of servers and fiber-optic cables with an efficiency and speed that rocks physics on its heels.

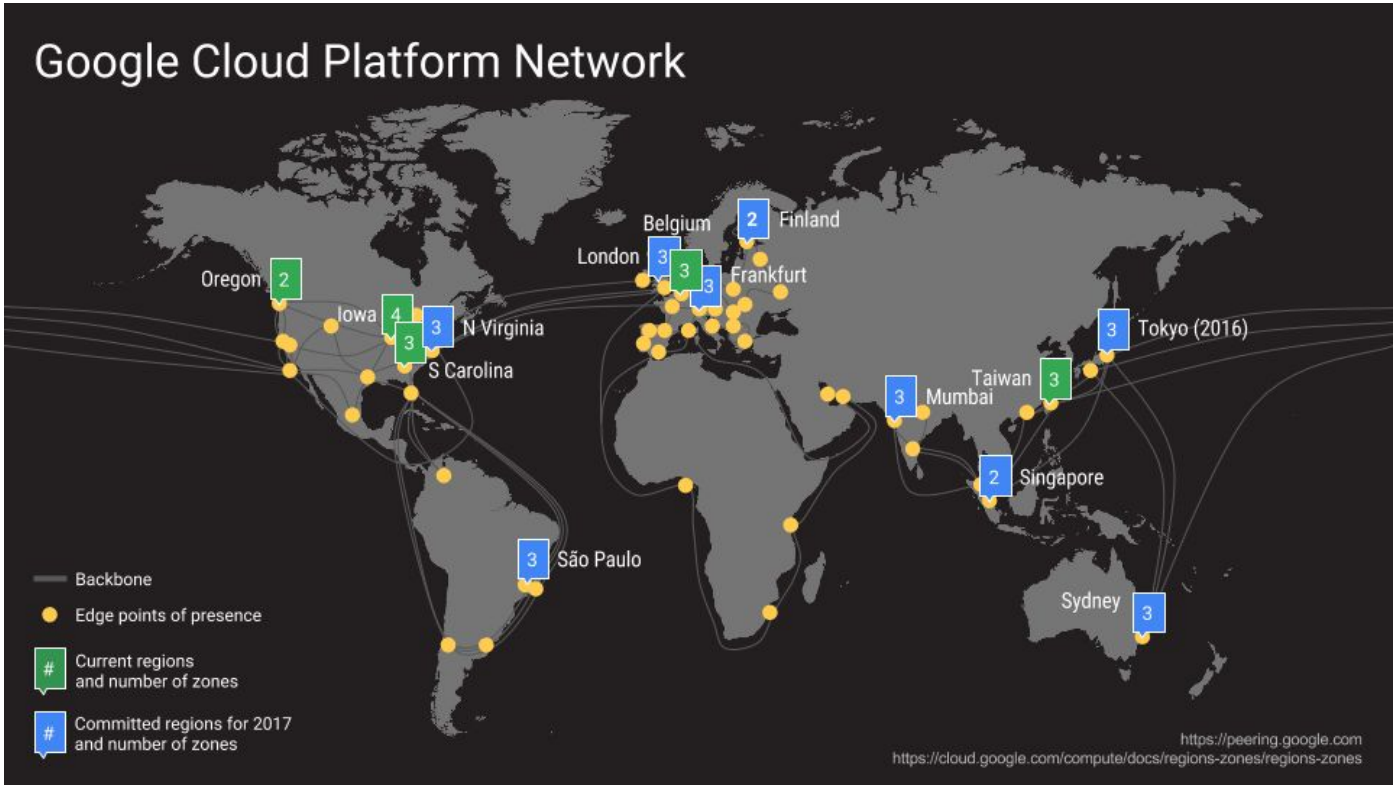
This is what makes Google Google: its physical network, its thousands of fiber miles, and those many thousands of servers that, in aggregate, add up to the **mother of all clouds**.”

- Wired



Google Cloud Platform opens Google-scale big data analysis

1. **Global data centers**
2. **Global network**
3. **Edge locations in 30+ countries**
4. **Software-defined networking**
   (why this matters)



Google Cloud

Before

Assembly required

On Premise

You Manage Hardware

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

Storage

Processing

Memory

Network

After

True On-Demand Cloud

In the Cloud

An actual, global elastic cloud

Invest your time in query writing, not infrastructure

Ad Hoc Querying and Scalable Storage

Google Cloud Platform enables on-demand scalability

On-Premise

Underprovisioned
(demand > capacity)

Query Processing Time

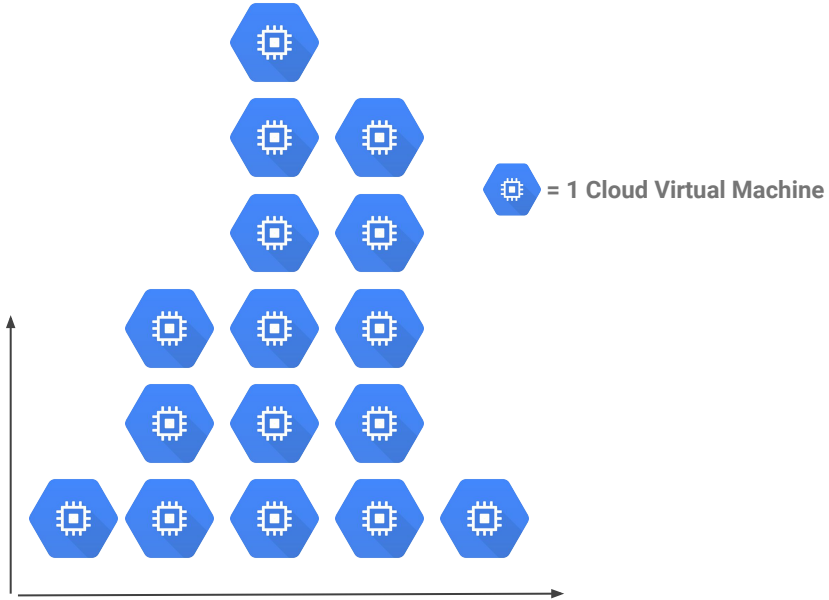
(120s, 1 Machine)

Overprovisioned
(demand < capacity)

Query Processing Time

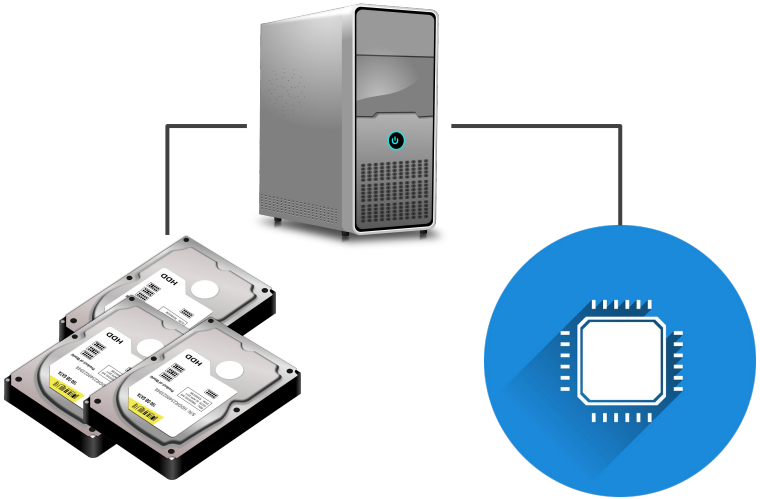
(10s, 120 Machines)

Google Cloud Platform



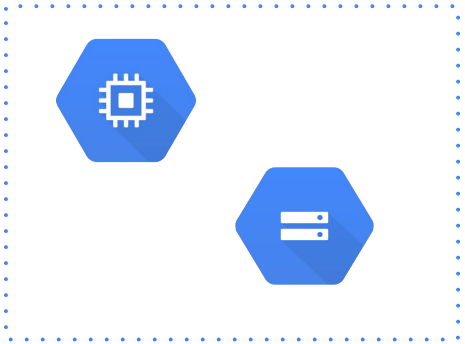
Separation of storage and computing power enables efficient resource allocation

On-Premise



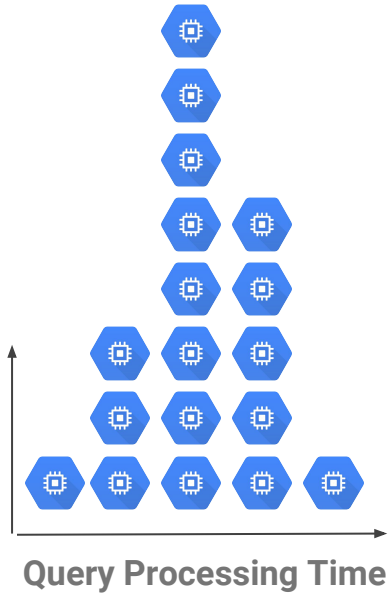
Pay for ability to use processing power even when no queries running

Google Cloud Platform



Pay for only the resources you are using and no more

BigQuery scales automatically and you only pay for what you use



Fully-Managed Infrastructure Scales to Process Faster

$

.. and you only pay for bytes processed + storage

Module 1

Introduction to Data on the Google Cloud Platform

In this module we will:

- Highlight Analytics Challenges Faced by Data Analysts
- Compare Big Data On-Premise vs on the Cloud
- **Learn from Real-World Use Cases of Companies Transformed through Analytics on the Cloud**
- Navigate Google Cloud Platform Project Basics

Store Petabytes of Data

*"[Our mission is] to **make our data so intelligent** it has the answer before the question is even asked. It was a stretch goal but essentially one that means we have to **capture all the data** we produce - both now and in the future."*

Dan Nelson - Head of Data
Ocado

ocado

Google Cloud

Focus on your Business, not Hardware

"The less time that we can spend solving problems that are already solved, like scaling,... the ***more time and energy we can spend on turning our data into value****"*

Nicholas Harteau - VP Infrastructure
Spotify

Module 1

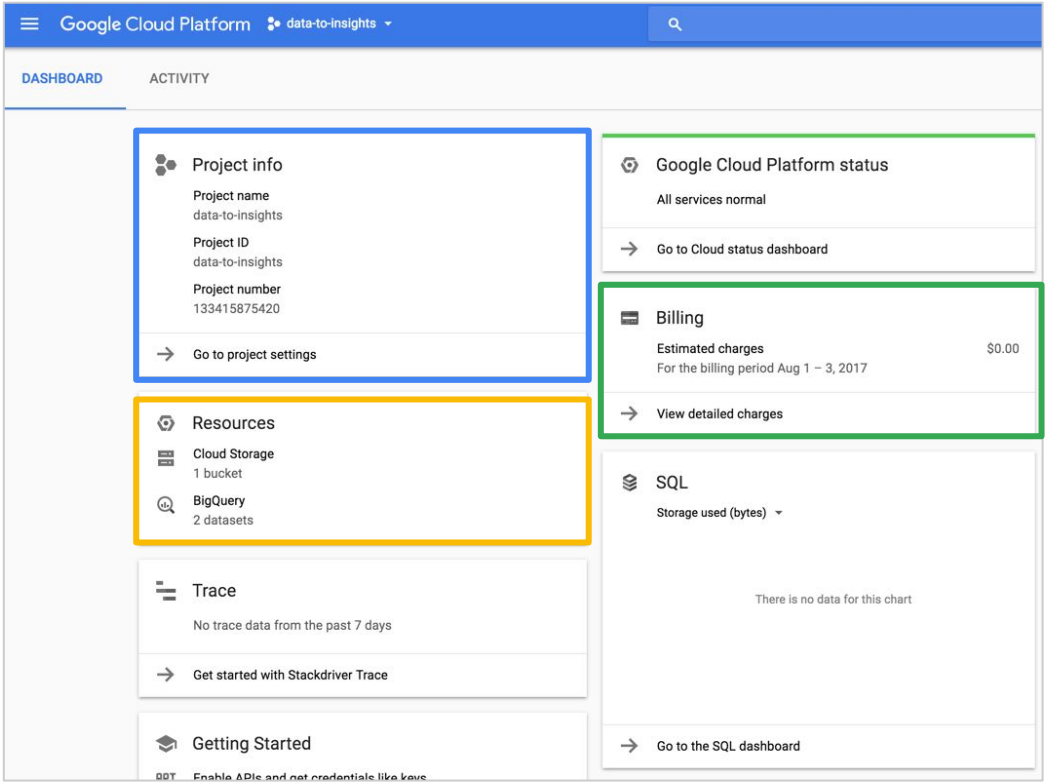
Introduction to Data on the Google Cloud Platform

In this module we will:

- Highlight Analytics Challenges Faced by Data Analysts
- Compare Big Data On-Premise vs on the Cloud
- Learn from Real-World Use Cases of Companies Transformed through Analytics on the Cloud
- **Navigate Google Cloud Platform Project Basics**

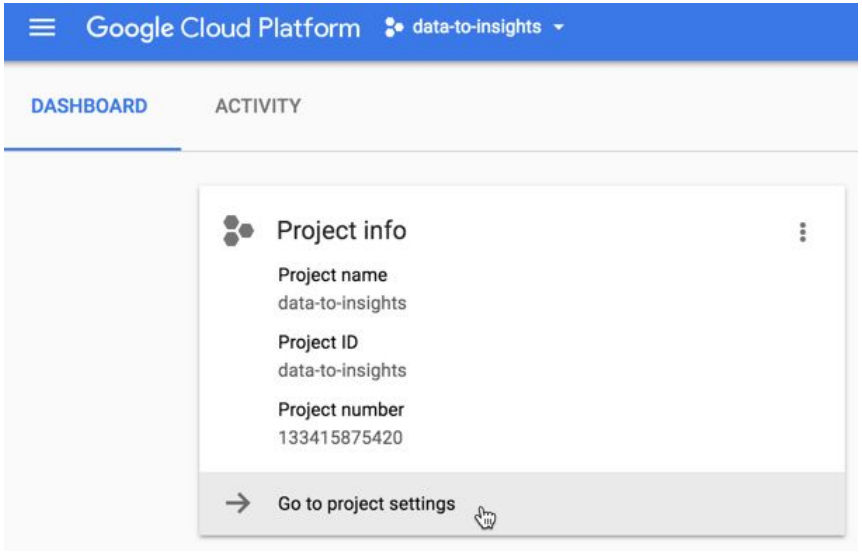
Navigate the Google Cloud Platform using the dashboard

1. **Projects**
2. **Resources**
3. **Billing**



1. Projects organize and govern your activities in the cloud

- **Navigate and launch cloud tools** for your project by exploring the Products and Services menu
- **Work collaboratively** by adding project users through IAM (Identity and Access Management)
- **Authorize Tools and Apps** through the API manager

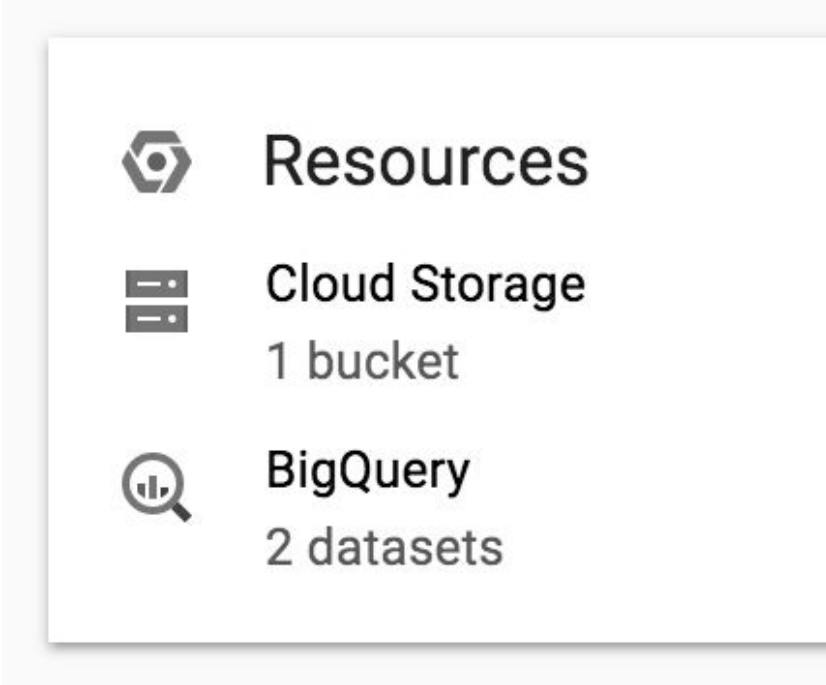


2. Resources are what you are using in the cloud

Commonly used by data analysts:

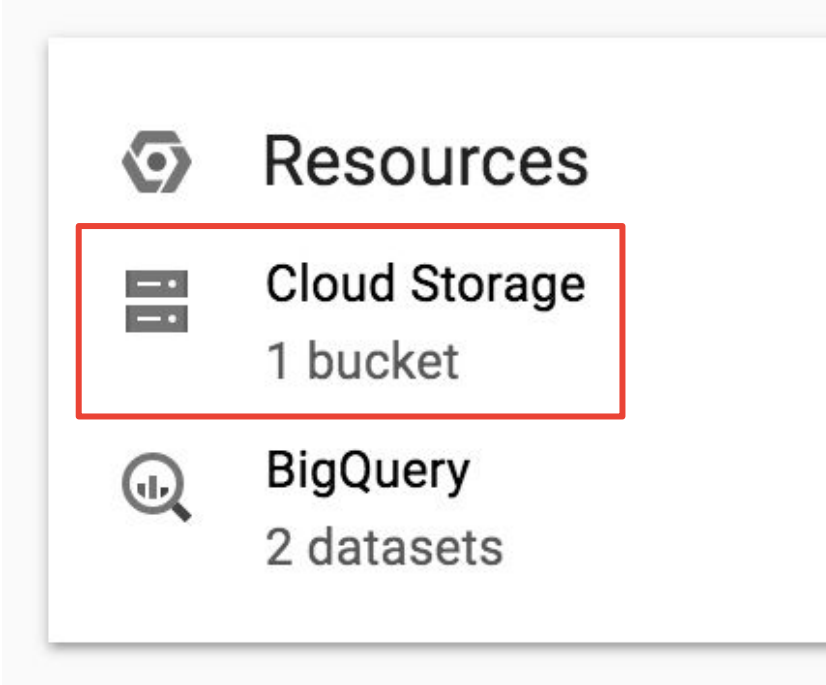
- **Storage** in Google Cloud Storage
  - Example: You use a Bucket for uploading large CSV files to ingest later for analysis

- **Datasets** in Google BigQuery
  - Example: You perform analysis on raw data and create a brand new dataset



2. The Cloud Storage Bucket is your goto for scalable storage

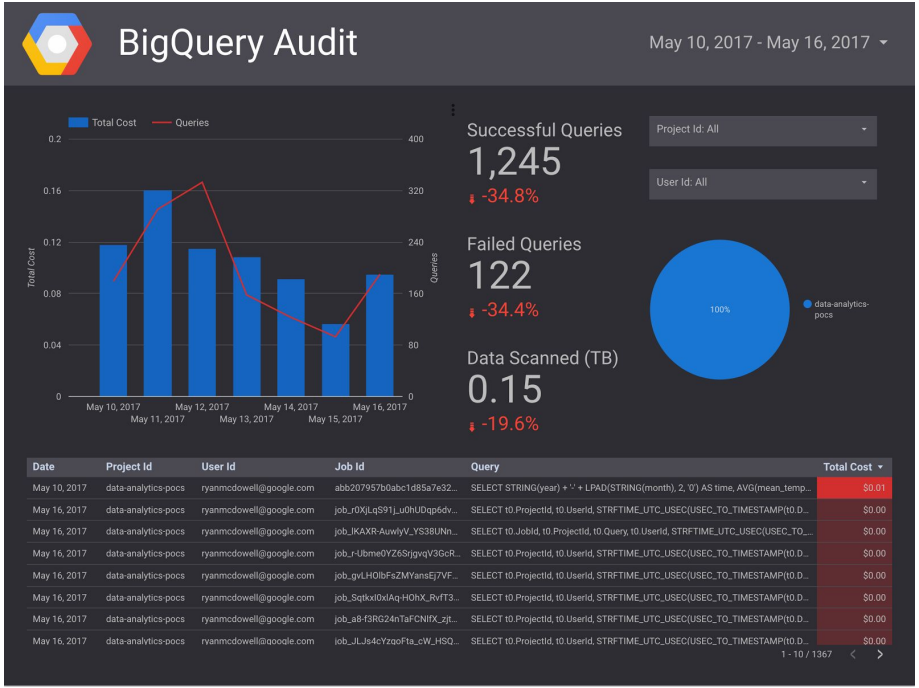
- Buckets are scalable containers that hold your data.
- You can create and upload files to your buckets within your Cloud Console



3. You are billed for the resources you use

Commonly used by data analysts:

- **Storage** in Google Cloud Storage
  - Billed for Bucket Storage
- **Datasets** in Google BigQuery
  - Billed for Query processing
  - Billed for Table Storage



After this course, try exporting BigQuery logs using this tutorial to recreate the above Data Studio billing dashboard

Google Cloud

Module Summary: Scale with the Google Cloud Platform

Overcome query speed, infrastructure, and cost challenges

Efficiently scale your compute and storage needs

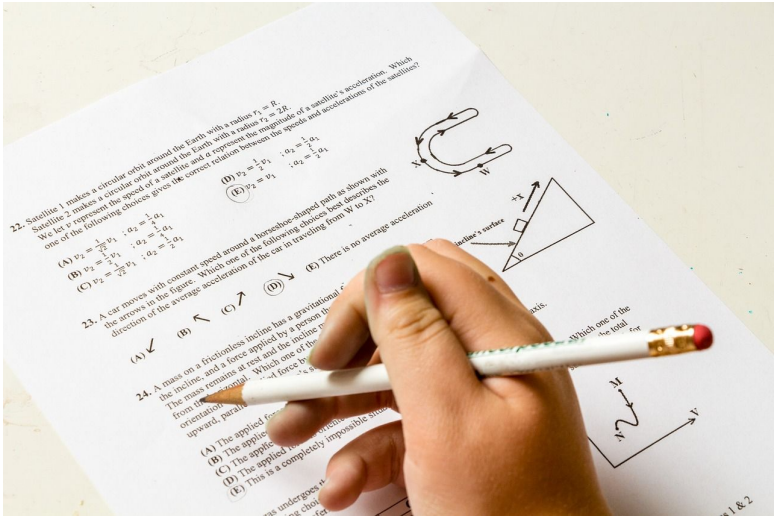
Manage and monitor your project resources in one place

Evangelize data analysis in your organization

Lab 0
Getting Started with Google Cloud Platform and Qwiklabs

Getting started with Google Cloud Platform and Qwiklabs

- Open an incognito window
- Navigate to: googlecloud.qwiklabs.com
- Create a new account with the email address you used when you registered for this course



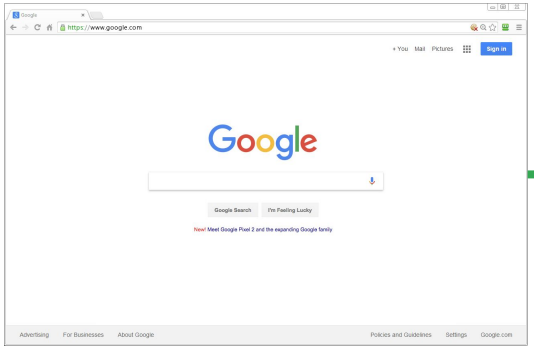
Google Cloud

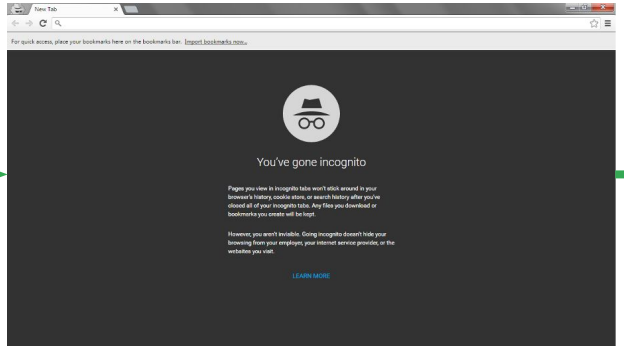
What you get

For each lab, Qwiklabs offers:

- A free set of resources for a fixed amount of time
- A clean environment with permissions

Qwiklabs sign-in process





Open an incognito browser

From the incognito browser, sign in to Qwiklabs

QWIKLABS

Select the lab and click Start Lab

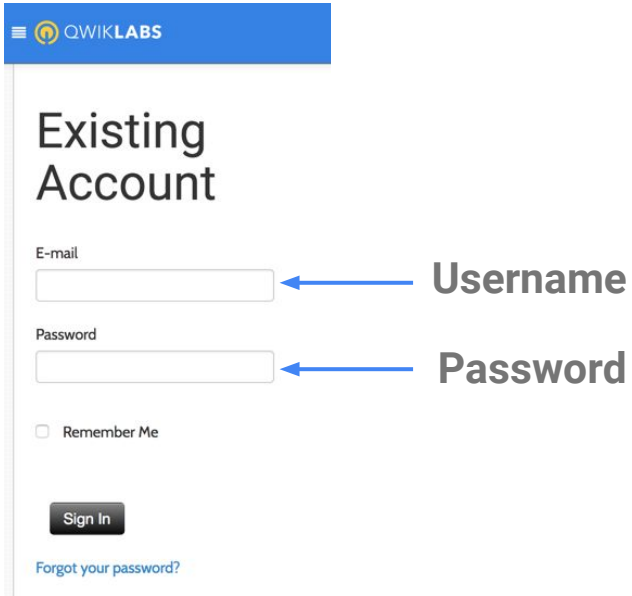
Sign in to the GCP console using the provided credentials

Open Qwiklabs

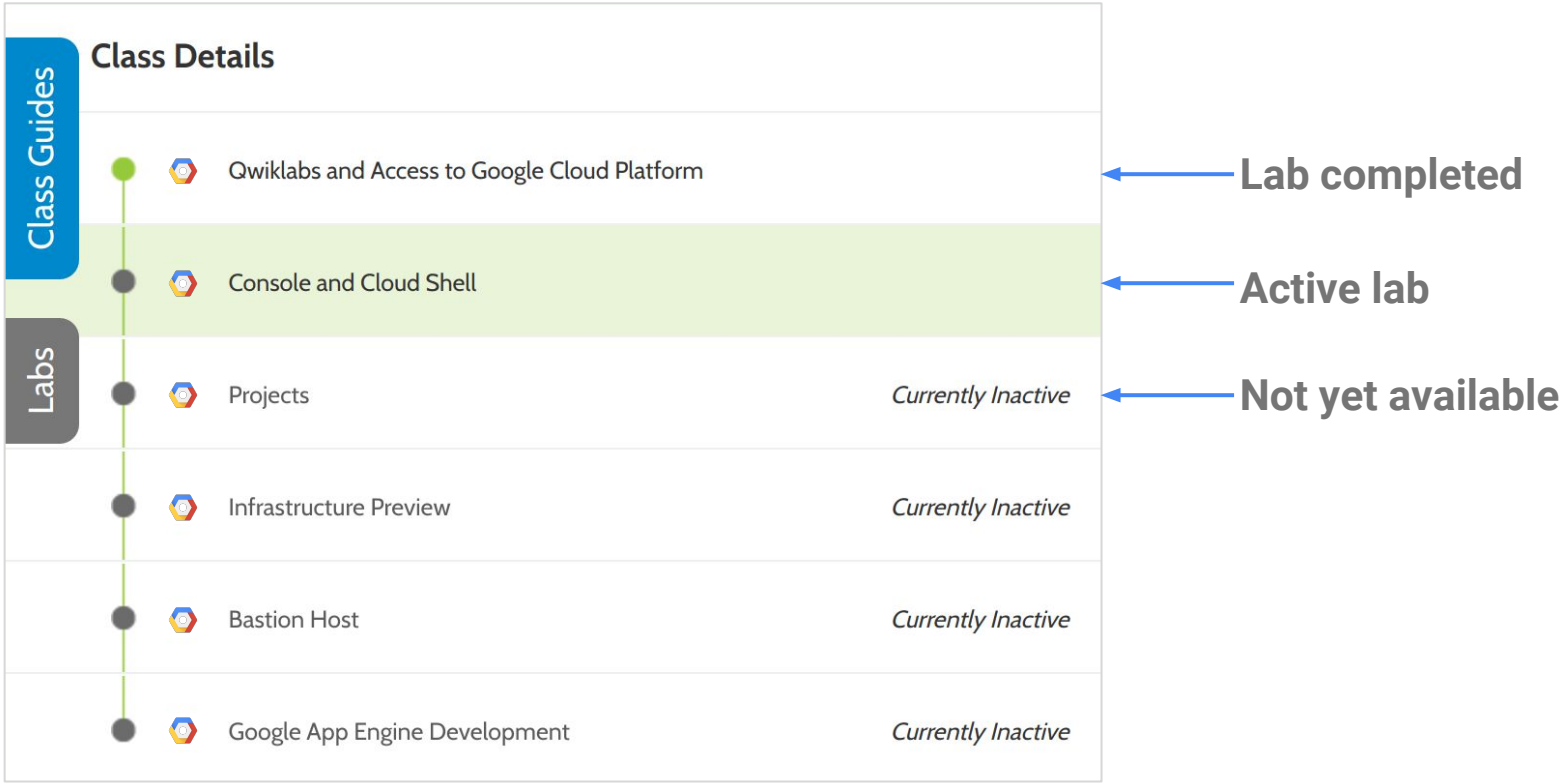
1. **Open an incognito window** (or private/anonymous window).

2. **Go** to the Qwiklabs URL your instructor provides.

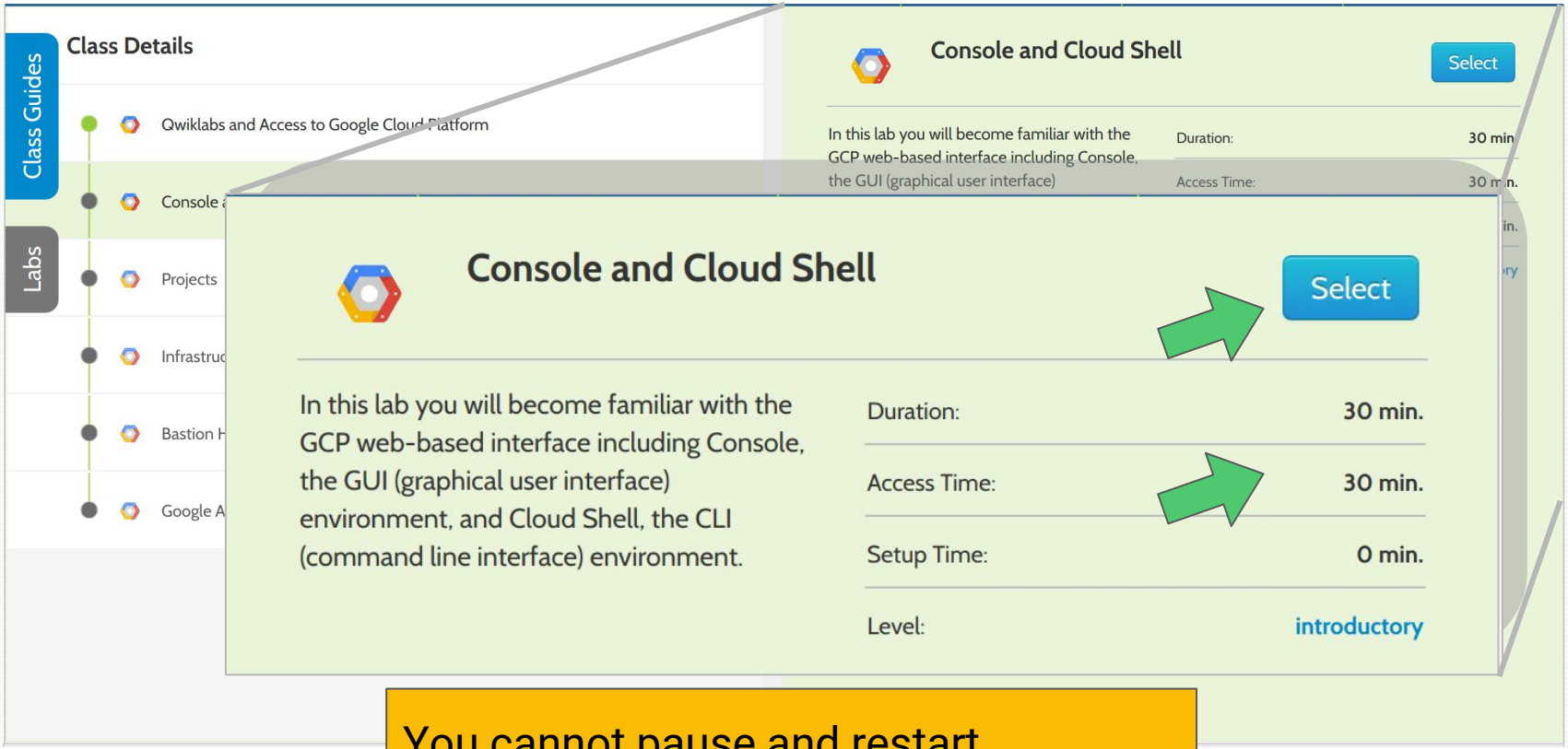
3. **Sign in** and launch the course (with credentials you used to register for the course).



View your labs



Select a lab



You cannot pause and restart

Run a lab

1. Click 

2. Note the following

3. Click 

and sign in

4. Accept terms and note the project 

End a lab

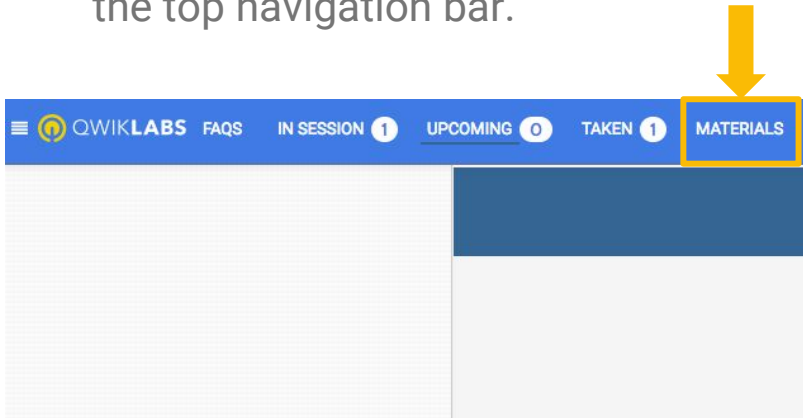
- When done, click 

to free your resources.

Some labs may require you to NOT end the lab; the instructor will inform you.

Course materials: End of class

1. Click **Materials** on the top navigation bar.



2. Select the class from the Course Materials list.