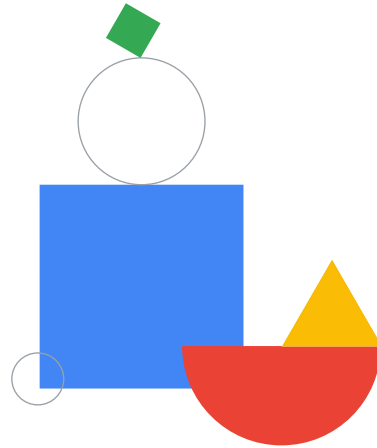# Data Engineering
on Google Cloud

Hello and welcome to **Data Engineering on Google Cloud**.

In this 4-day course, I look forward to showing you how to design data processing systems, build end-to-end data pipelines, analyze data, and implement machine learning.

In addition to the lectures, you will also complete a series of hands-on labs.

Introductions

Your instructor
- Organization
- Background
- Course goals

You
- Name
- Organization
- Job role
- Course goals

Google Cloud

Introductions:
Your instructor + You
Background
Position
Organization

Facilities

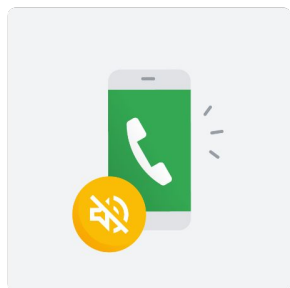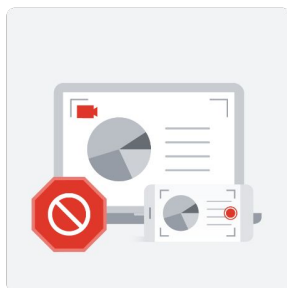| Parking	Facilities	Food

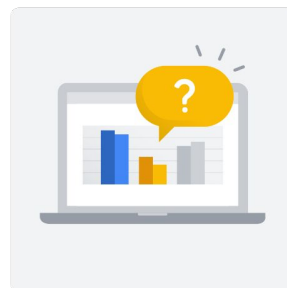Facilities:
- Parking
- Facilities
- Food

Etiquette

No calls

No recording

Ask questions

Google Cloud

Course etiquette:
- Please silence your phone and take calls outside.
- Recording this class is prohibited.
- Ask questions interactively or via chat (online).

Data Engineering on Google Cloud learning path

1	Modernizing Data Lakes and Data Warehouses with Google Cloud
2	Building Batch Data Pipelines on Google Cloud
3	Building Resilient Streaming Analytics Systems on Google Cloud
4	Smart Analytics, Machine Learning and AI on Google Cloud
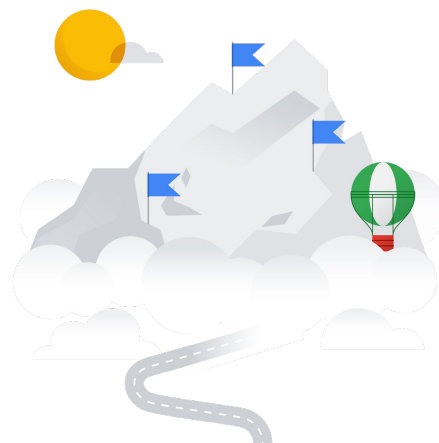
Google Cloud

Each of the four days that make up this course focuses on a specific aspect of data engineering.

We will first discuss the differences between data lakes and data warehouses, the two key components of any data pipeline. We will highlight use-cases for each type of storage and dive into the available data lake and warehouse solutions on Google Cloud in technical detail. Also, we will describe the role of a data engineer, the benefits of a successful data pipeline to business operations, and examine why data engineering should be done in a cloud environment.

Data pipelines typically fall under one of the Extract-Load, Extract-Load-Transform or Extract-Transform-Load paradigms. So during day 2, Building Batch Data Pipelines, describes which paradigm should be used and when for batch data. Furthermore, we will cover several technologies on Google Cloud for data transformation including BigQuery, executing Spark on Dataproc, pipeline graphs in Data Fusion and serverless data processing with Dataflow.

Processing streaming data is becoming increasingly popular as streaming enables organizations to get real-time metrics on operations. So day 3 covers how to build streaming data pipelines on Google Cloud. Pub/Sub is the primary product for handling incoming streaming data. We will also cover how to apply aggregations and transformations to streaming data using Dataflow, and how to store processed records in BigQuery or Bigtable for analysis.

Incorporating machine learning into data pipelines increases the ability of organizations to extract insights from their data. The final day of the course covers several ways for machine learning to be included in data pipelines on Google Cloud depending on the level of customization required. For little to no customization, we'll cover AutoML. For more tailored machine learning capabilities, you'll be introduced to Notebooks and BigQuery Machine Learning. Also, we'll cover how to productionize machine learning solutions using Kubeflow.

Lab environment

For each lab, Qwiklabs offers:

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

Google Cloud

Open Qwiklabs

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

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

3 **Sign In** with existing account or **Join** with new account (with email you used to register for the course).

4 Launch the course from **Profile**.

Access issues
The process to open Qwiklabs can differ based on credentials used. Please reach out to your trainer if you have any access issues.

☰   🔵 QWIK**LABS**

🏠 Home

🎓 Catalog

📅 Profile

Google Cloud
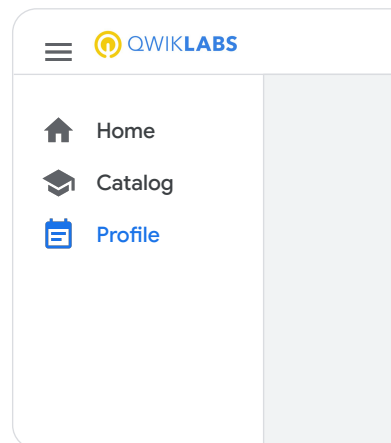
View your labs

Do **NOT** launch a lab until instructed to do so!

Labs	Lecture Notes

✅ ——————————————————————————— ← Lab completed

◯ ——————————————————————————— ← To be completed

⚠ ————————————————— `Lab Currently Disabled` ← Not yet available

⚠ ————————————————— `Lab Currently Disabled` ←

Google Cloud

View lecture notes

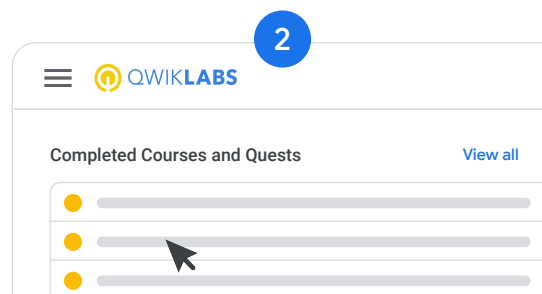| Labs	**Lecture Notes**
01	⬇
02	⬇
03	⬇
04	⬇

You can download these as PDF files

End of class - Materials

Materials are available for 2 years
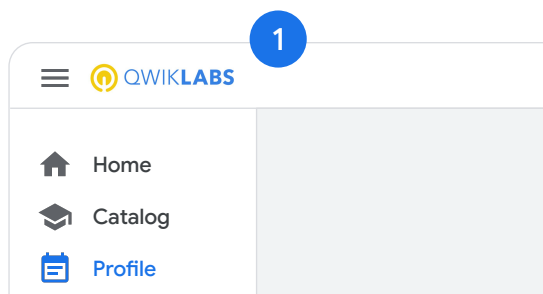
Click on **Profile** in the left-hand navigation bar

Select the class from the Completed Courses list

You can view the course materials within Qwiklabs as follows:

1. Click on **Profile** in the left-hand navigation bar.
2. Scroll down to the **Completed Courses** section.
3. Select the class from the **Completed Courses** list.

Materials are available for 2 years following the completion of a course.