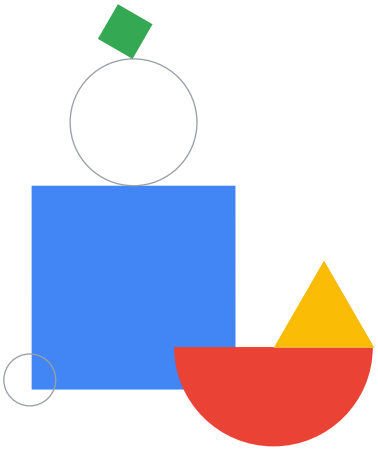
Google Cloud

Data Engineering on Google Cloud



Hello and welcome to Data Engineering on Google Cloud.

Introductions

Your instructor

- Organization
- Background
- Course goals

You

- Name
- Organization
- Job role
- Course goals

Introductions:
Your instructor + You
Background
Position
Organization

Facilities

Parking

Facilities

Food

Facilities:

- Parking
- Facilities
- Food

Course etiquette

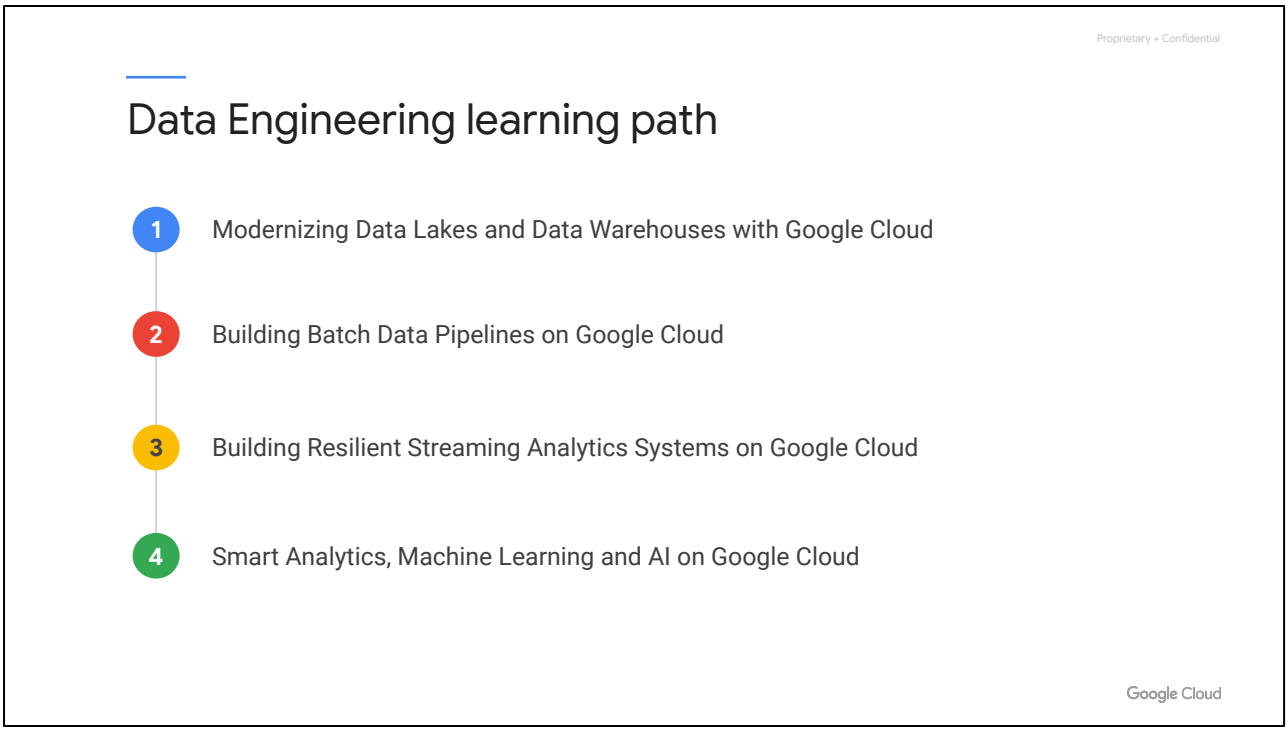
Please silence your phone and take calls outside.

Recording this class is prohibited.

Ask questions interactively or via chat (online).

Course etiquette:

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



As part of the Data Engineering learning path, we will first discuss the differences between data lakes and data warehouses, the two key components of any data pipeline. This course highlights use-cases for each type of storage and dives into the available data lake and warehouse solutions on Google Cloud in technical detail. Also, this course describes the role of a data engineer, the benefits of a successful data pipeline to business operations, and examines why data engineering should be done in a cloud environment.

Data pipelines typically fall under one of the Extra-Load, Extract-Load-Transform or Extract-Transform-Load paradigms. So the next course, Building batch data pipelines, describes which paradigm should be used and when for batch data. Furthermore, it covers several technologies on Google Cloud for data transformation including BigQuery, executing Spark on Dataproc, pipeline graphs in Cloud Data Fusion and serverless data processing with Dataflow.

Processing streaming data is becoming increasingly popular as streaming enables businesses to get real-time metrics on business operations. So the third course covers how to build streaming data pipelines on Google Cloud. Pub/Sub is described for handling incoming streaming data. The course also covers how to apply aggregations and transformations to streaming data using Dataflow, and how to store processed records to BigQuery or Cloud Bigtable for analysis.

Incorporating machine learning into data pipelines increases the ability of businesses

to extract insights from their data. The final course covers several ways machine learning can be included in data pipelines on Google Cloud depending on the level of customization required. For little to no customization, this course covers AutoML. For more tailored machine learning capabilities, this course introduces AI Platform Notebooks and BigQuery Machine Learning. Also, this course covers how to productionalize machine learning solutions using Kubeflow.

Lab environment

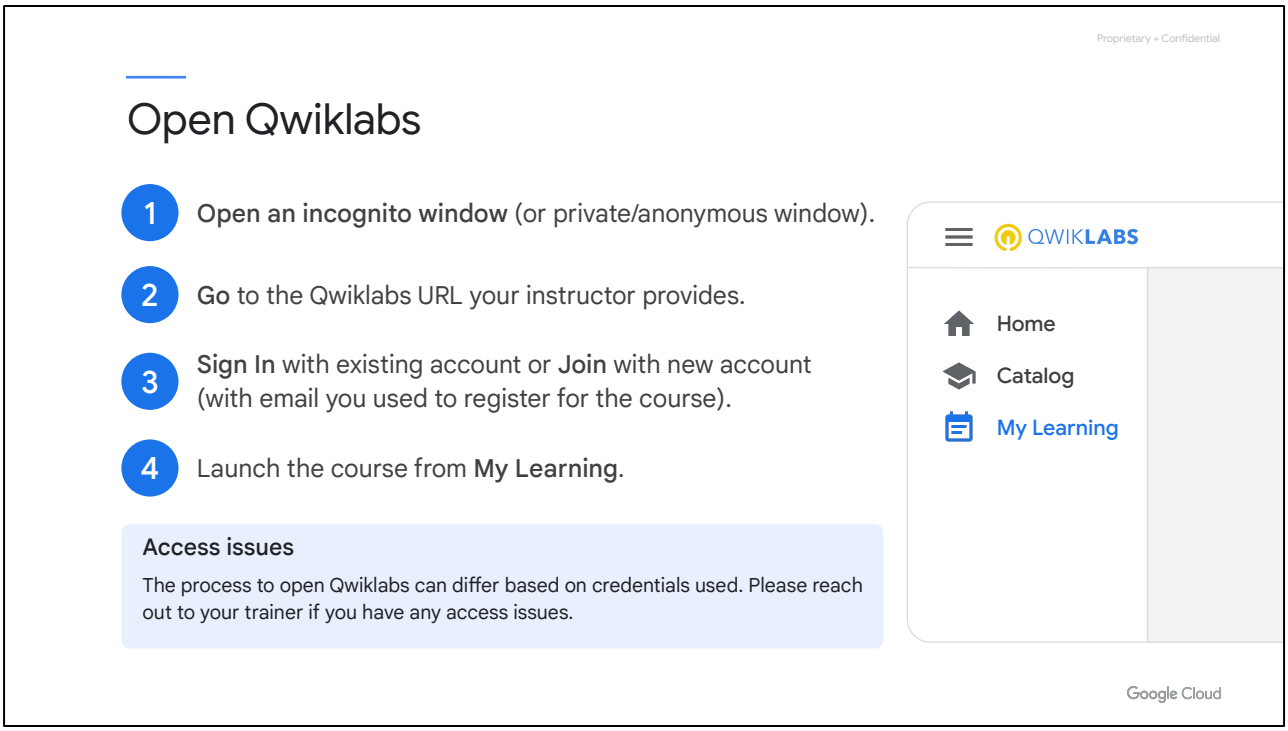
For each lab, Qwiklabs offers:

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

This course includes interactive hands-on labs through the Qwiklabs platform.

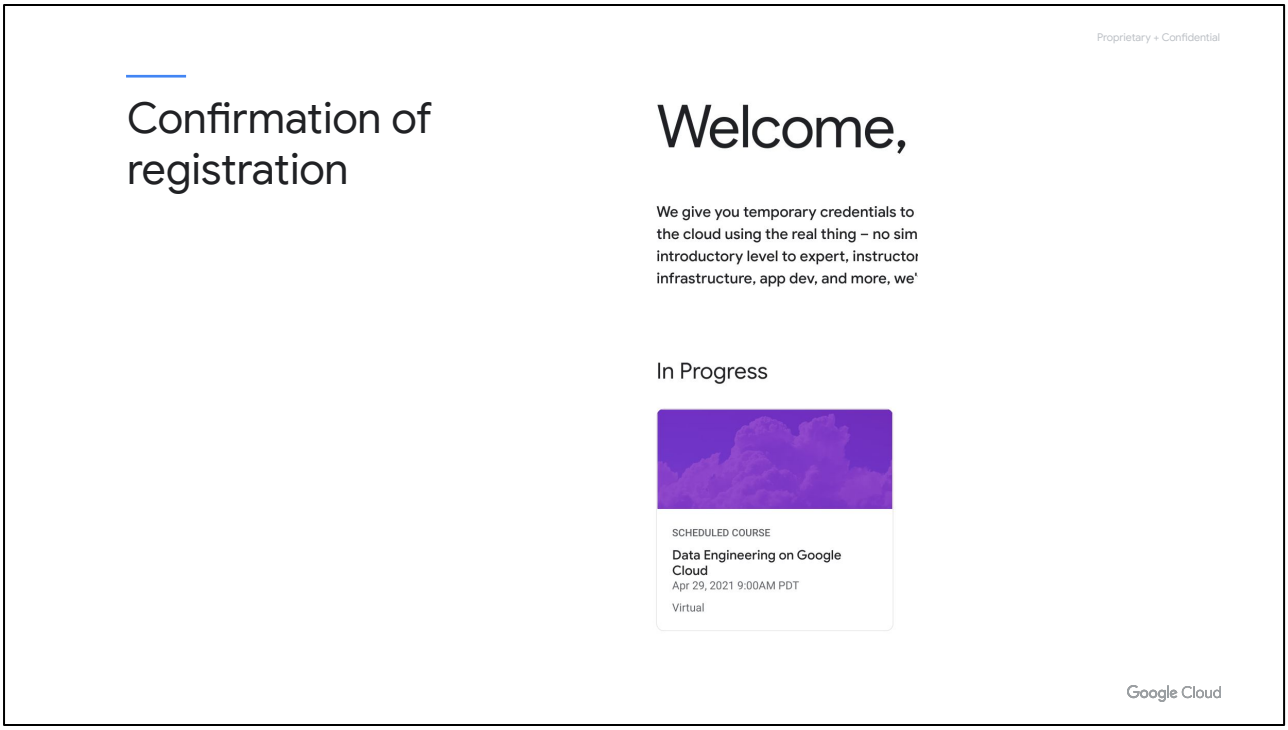
Qwiklabs provisions you with Google account credentials, so you can access the Cloud Console for each lab at no cost. Specifically, for each lab, Qwiklabs offers:

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

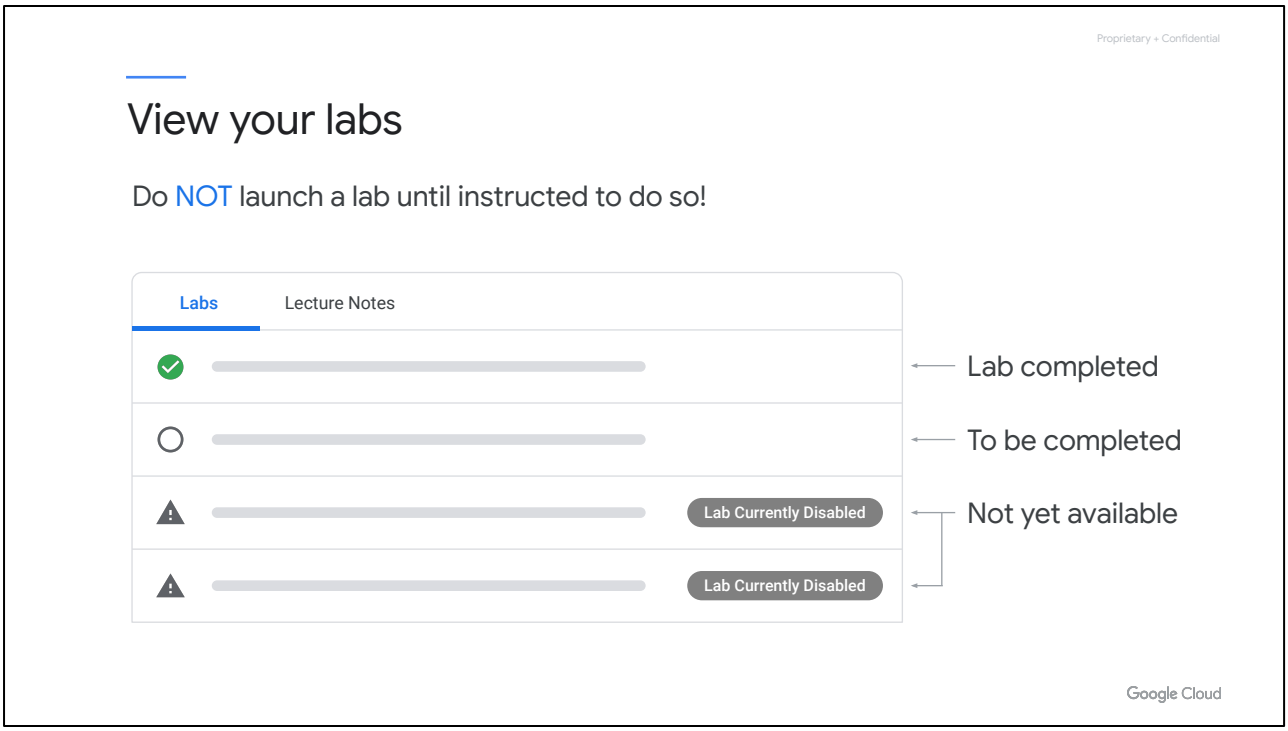


Let's go ahead and open Qwiklabs:

1. **Open an incognito window** (or private/anonymous window). Use of an incognito browser window reduces the risk that you will accidentally do the labs using your own Google Cloud account instead of Qwiklabs.
2. **Go** to the Qwiklabs URL your instructor provides.
3. **Sign** in with an existing account or **Join** with a new account (with email you used to register for the course).
4. Launch the course from **My Learning**.



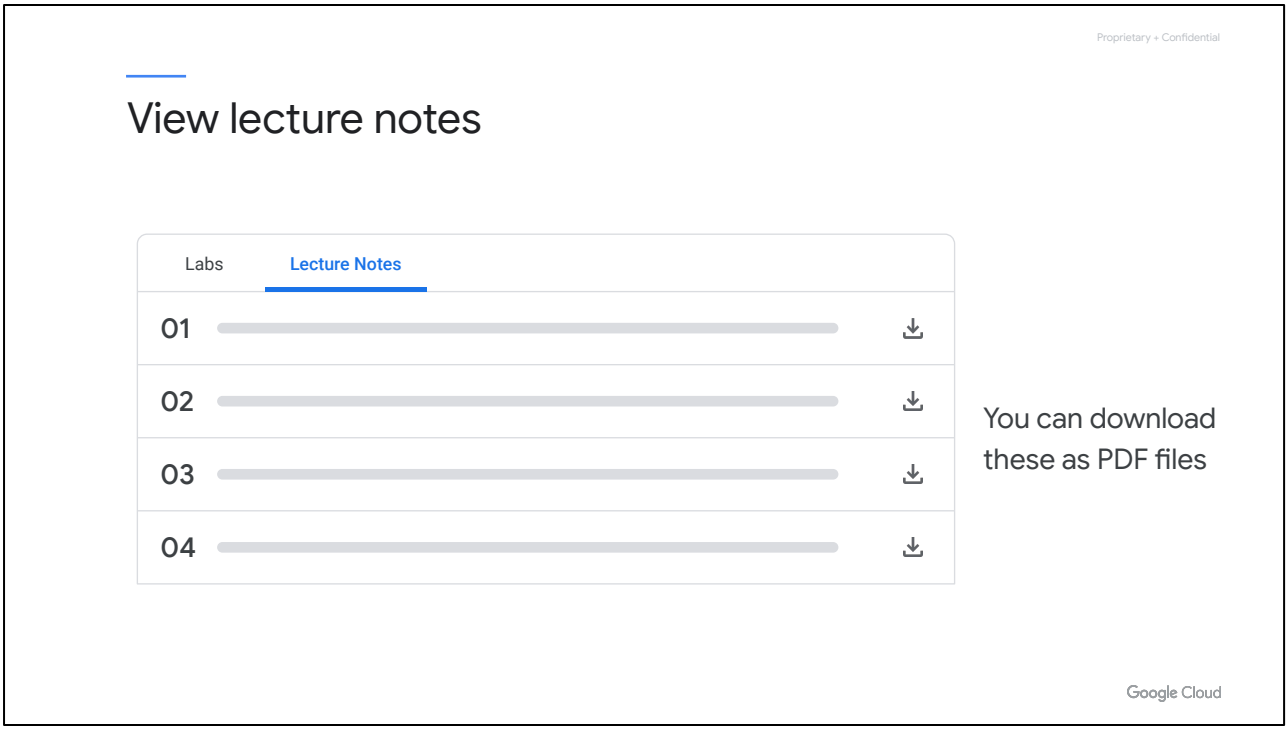
You should see a screen similar to this one once you have registered successfully.



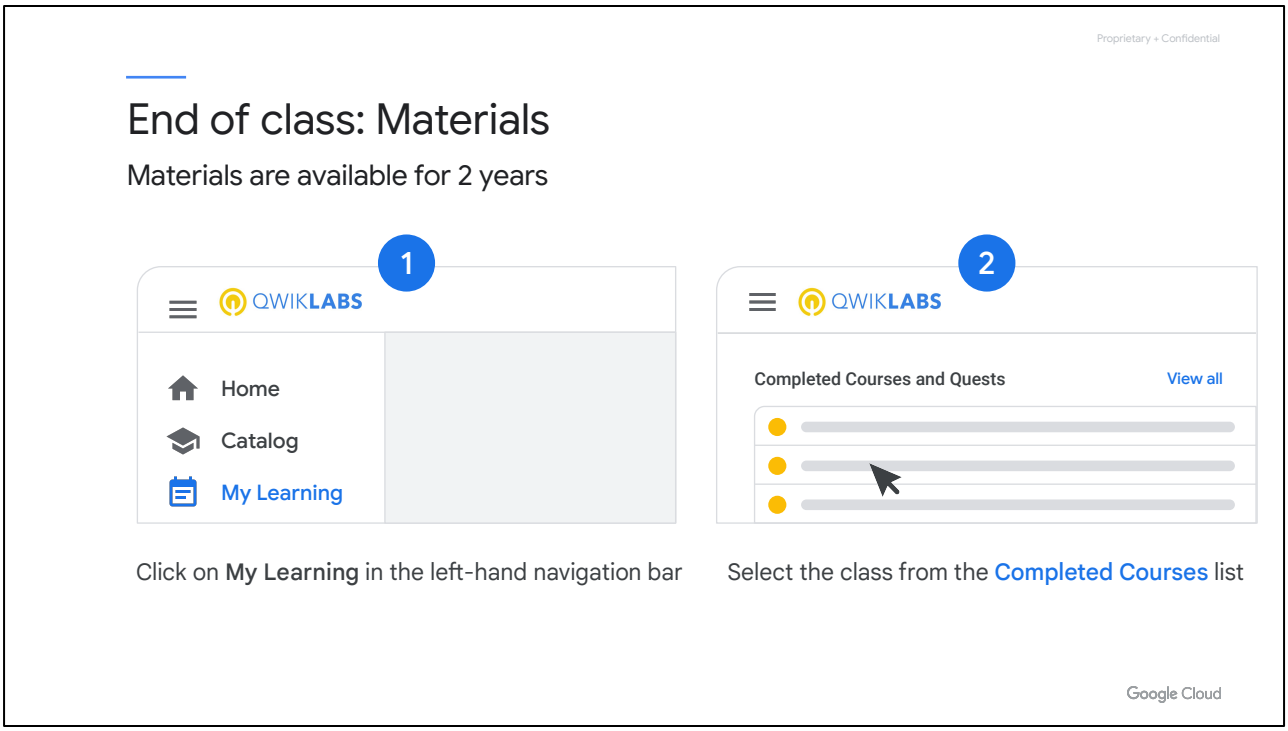
After you launch the course, you can view your labs. The lab list will indicate whether a lab is:

- Completed (by you)
- To be completed
- Not yet available

Your instructor will let you know when it's time to launch a lab. Once you start a lab, you won't be able to pause and restart it, so you'll need a continuous block of time to complete the work.



Within the course, you can also view the lecture notes. You can download these as PDF files.



You can view the course materials within Qwiklabs as follows:

1. Click on ***My Learning*** in the left-hand navigation bar.
2. Select the class from the ***Completed Courses*** list.

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