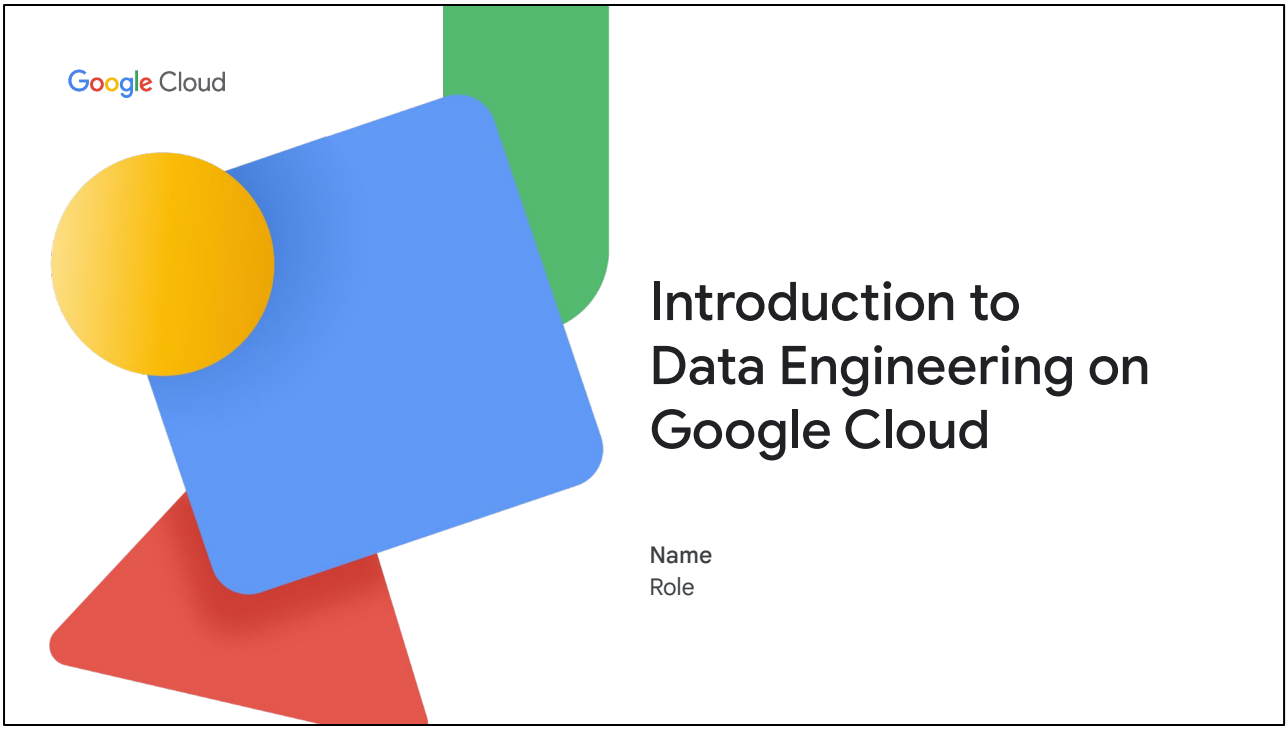
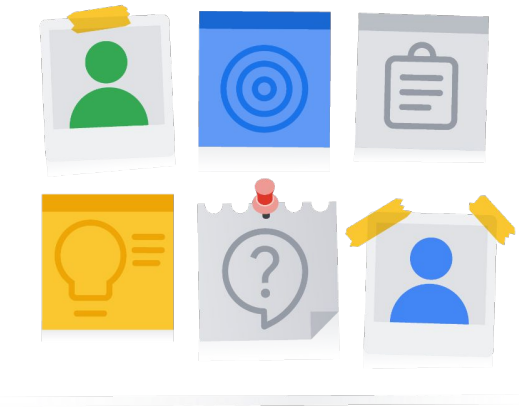


Hello and welcome to Introduction to Data Engineering on Google Cloud. Whether you already work in data engineering and want to learn how to be successful on Google Cloud or you are looking to progress in your career, this course will help you get started. Through a series of lectures, quizzes, and hands-on labs, you learn the fundamentals of data engineering on Google Cloud.

Introduction

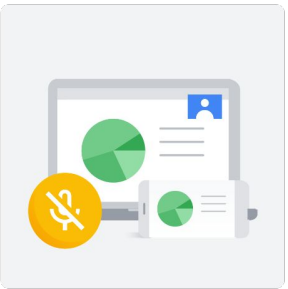
Your instructor and you

- Background
- Position
- Organization



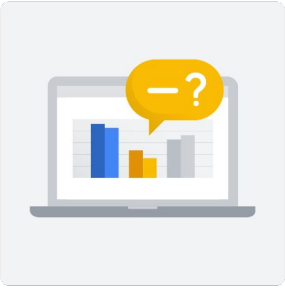
ILT INTRODUCTION

Etiquette



Mute microphone

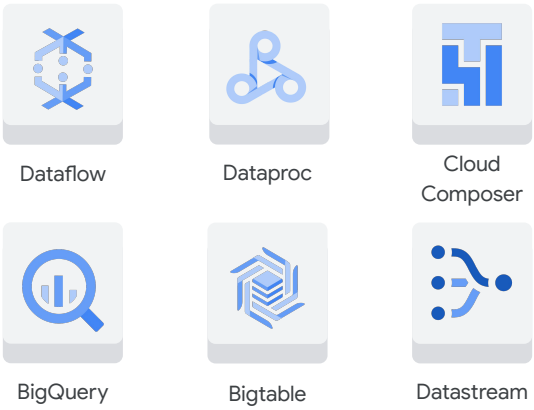
No recording



Ask questions

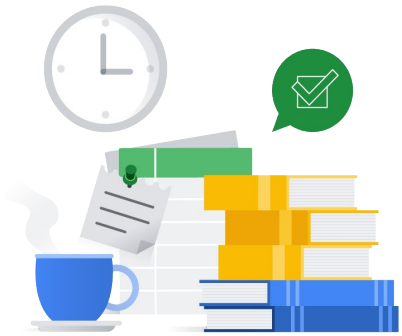
Target audience

- Data Engineer
- Database Admin
- System Admin



This course was designed for data engineers or anyone interested in preparing and storing data assets for further usage in their organization. This involves using tools such as, but not limited to Dataflow, Dataproc, Cloud Composer, BigQuery, Bigtable, and Datastream.

Helpful knowledge



- Google Cloud basics

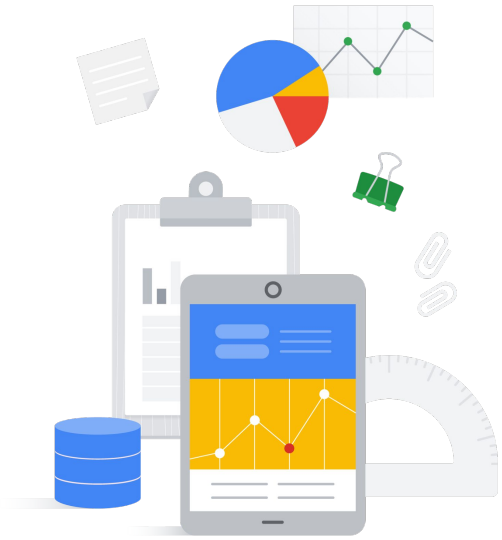
Helpful to **be familiar with**

- Querying data using SQL
- Python programming
- Data modelling and ETL

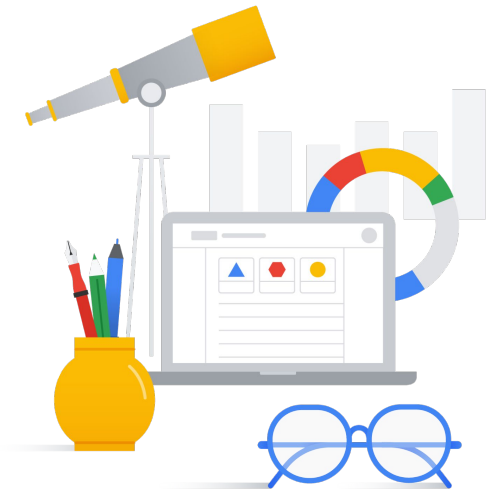
- Prior Google Cloud experience at the fundamental level using Cloud Shell and accessing products from the Google Cloud console.
- Basic proficiency with a common query language such as SQL.
- Experience with data modeling and ETL (extract, transform, load) activities.
- Experience developing applications using a common programming language such as Python.

Learning objective #01

Understand the role of a data engineer.



In this course, you learn about the duties and responsibilities of a data engineer.



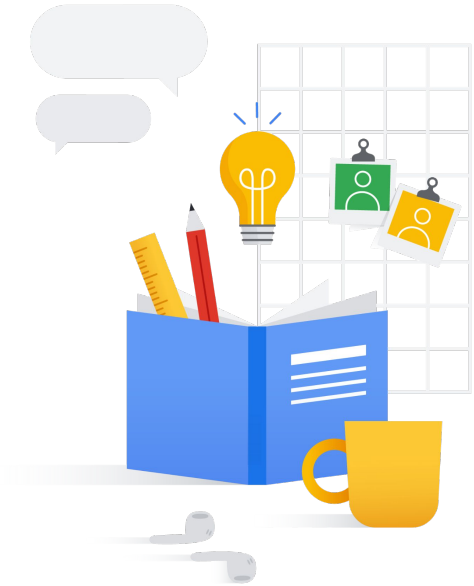
Learning objective #02

Identify data engineering tasks and core components used on Google Cloud.

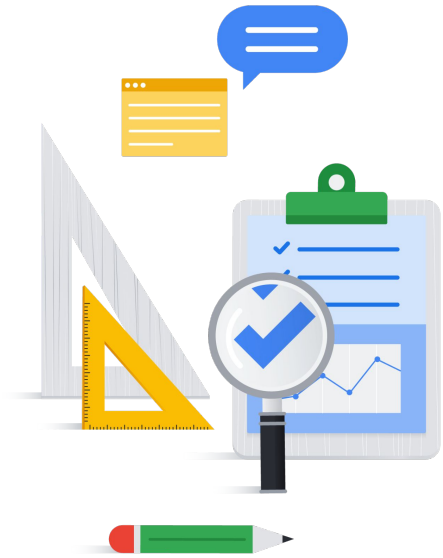
Identify data engineering tasks and core components used on Google Cloud to accomplish those tasks.

Learning objective #03

Understand how to create and deploy data pipelines of varying patterns on Google Cloud.



Understand how to create and deploy data pipelines of varying patterns on Google Cloud.



Learning objective #04

Identify and utilize various automation techniques on Google Cloud.

And identify and utilize various automation techniques on Google Cloud to complete data engineering tasks.

Agenda

01 Data Engineering Tasks and Components

02 Data Replication and Migration

03 The Extract and Load Data Pipeline Pattern

04 The Extract, Load, and Transform Data Pipeline Pattern

05 The Extract, Transform, and Load Data Pipeline Pattern

06 Automation Techniques

The course is divided into six modules designed to address the learning objectives.

First, we look at data engineering tasks and components on Google Cloud.

Next, we explore data replication and migration.

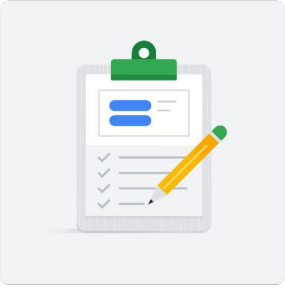
Then, we explore each of the three main data pipeline patterns: extract and load; extract, load, and transform; and
extract, transform, and load.

We conclude the course by examining automation techniques important to a data engineer.

Hands-on labs

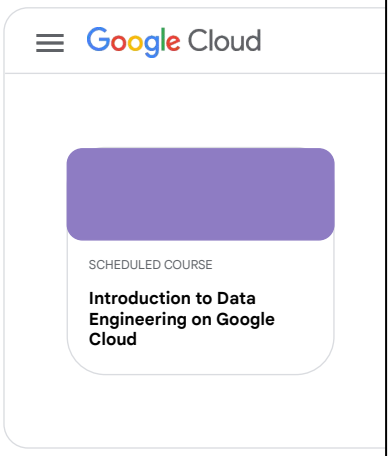
For each lab, Google Cloud Skills Boost offers:

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



Open Cloud Skills Boost

1. **Open an incognito window** (or private/anonymous window).
2. **Go** to the URL: **cloudskillsboost.google**
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 the **Welcome page**.

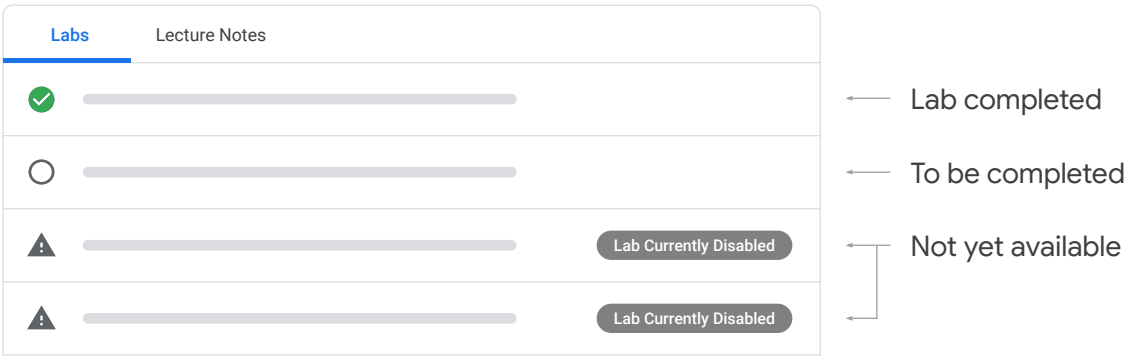


Access issues

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

View your labs

Do NOT launch a lab until instructed to do so!



View lecture notes

Labs | Lecture Notes

01

02

03

04

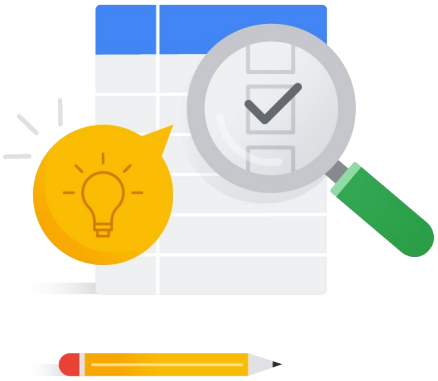
You can download these as PDF files

Lab: Loading Data into BigQuery

30 min

Learning objectives

- Load data into BigQuery from various sources.
- Load data into BigQuery using the CLI and the Google Cloud console.
- Use DDL to create tables.



In this lab, you practice loading data into BigQuery. The primary objective of this lab is to load data into BigQuery using both the command-line interface and the Google Cloud console. You also experience loading several datasets into BigQuery and using the Data Description Language or DDL.

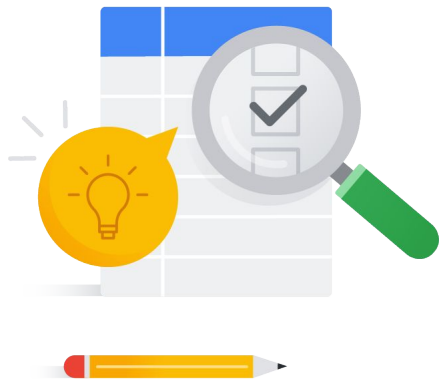
Lab URL: https://www.cloudskillsboost.google/catalog_lab/2119

Lab: Datastream: PostgreSQL Replication to BigQuery

45 min

Learning objectives

- Prepare a Cloud SQL for PostgreSQL instance using the Google Cloud console.
- Import data into the Cloud SQL instance.
- Create a Datastream connection profile for the PostgreSQL database.
- Create a Datastream connection profile for the BigQuery destination.
- Create a Datastream stream and start replication.
- Validate that the existing data and changes are replicated correctly into BigQuery.



In this lab, you use Datastream to replicate data from PostgreSQL to BigQuery. You prepare and load a Cloud SQL for PostgreSQL instance. You create Datastream connection profiles for the source and target. You then create a Datastream processing stream and start replication. Finally, you validate the replication in BigQuery.

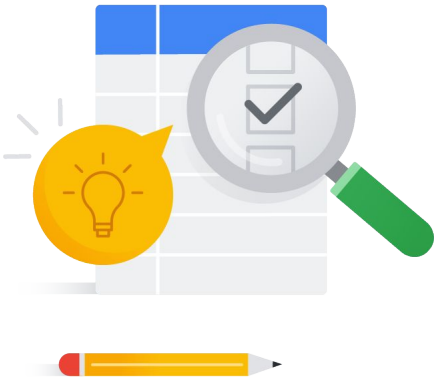
Lab URL: https://www.cloudskillsboost.google/catalog_lab/5777

Lab: BigLake: Qwik Start

45 min

Learning objectives

- Create and view a connection resource.
- Set up access to a Cloud Storage data lake.
- Create a BigLake table.
- Query a BigLake table through BigQuery.
- Set up access control policies.
- Upgrade an external table to be a BigLake table.



In this lab, you use BigLake to connect to various external data sources.

You configure a connection resource and set up access to a Cloud Storage data lake.

You create and query a BigLake table and set up access control policies.

Finally, you upgrade an existing external table to be a BigLake table.

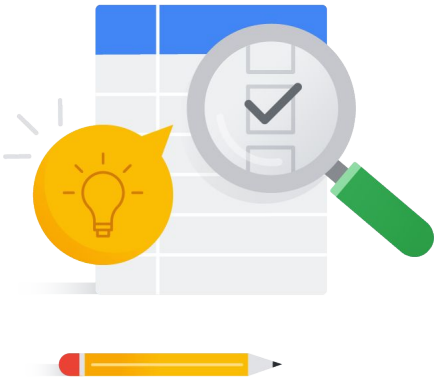
Lab URL: https://www.cloudskillsboost.google/catalog_lab/4896

Lab: Create and Execute a SQL Workflow in Dataform

30 min

Learning objectives

- Create a Dataform repository.
- Create and initialize a Dataform development workspace.
- Create and execute a SQL workflow.
- View execution logs in Dataform.



In this lab, you use Dataform to create and execute a SQL workflow. First, you create a Dataform repository. Second, you create and initialize a Dataform development workspace. Then, you create and execute a SQL workflow. Finally, you view execution logs in Dataform to confirm completion.

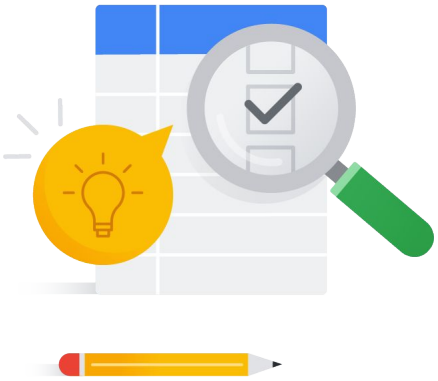
Lab URL: https://www.cloudskillsboost.google/catalog_lab/20934

Lab: Use Dataproc Serverless for Spark to load BigQuery

30 min

Learning objectives

- Configure the environment.
- Download lab assets.
- Configure and execute the Spark code.
- View data in BigQuery.



In this lab, you use Dataproc Serverless for Spark to load BigQuery.

First, you configure the environment.

Next you download lab assets.

You then configure and execute the Spark code.

Finally, you view the data in BigQuery.

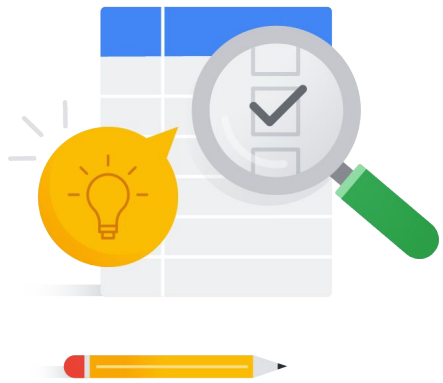
reference: https://www.cloudskillsboost.google/authoring/labs/31670

Lab: Creating a Streaming Data Pipeline for a Real-Time Dashboard with Dataflow

45 min

Learning objectives

- Create a Dataflow job from a template.
- Stream data via Dataflow pipeline into BigQuery.
- Monitor a Dataflow pipeline in BigQuery.
- Analyze results with SQL.
- Visualize key metrics in Looker Studio.



In this lab, you create a streaming data pipeline for a real-time dashboard with Dataflow.

You create a Dataflow job from a template.

You then monitor a pipeline loading data into BigQuery.

After that, you examine the data loaded using SQL.

Finally, you visualize key metrics using Looker Studio.

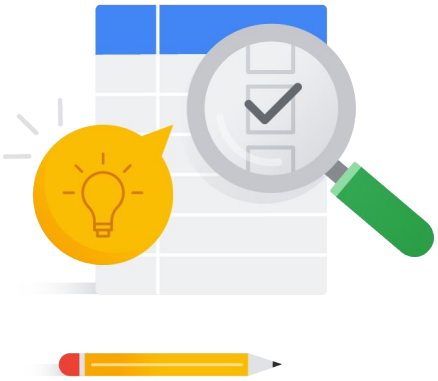
reference: https://www.cloudskillsboost.google/catalog_lab/1796

Lab: Use Cloud Run Functions to Load BigQuery

30 min

Learning objectives

- Create a Cloud Run function.
- Deploy and test the Cloud Run function.
- View data in BigQuery and review Cloud Run function logs.



In this lab, you create a Cloud Run function to Load BigQuery.

You create a Cloud Run function using the Cloud SDK.

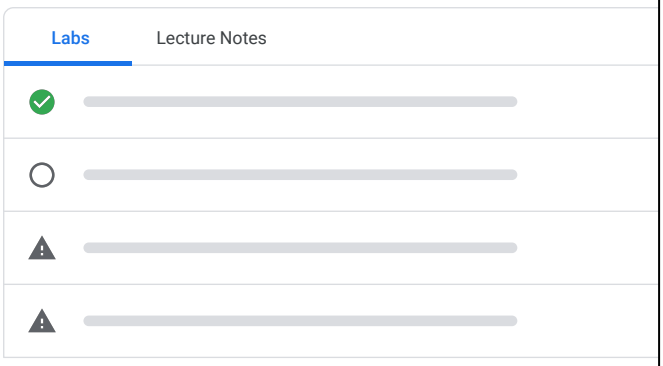
You then deploy and test the Cloud Run function.

Finally, you view data in BigQuery and review Cloud Run function logs.

reference: https://www.cloudskillsboost.google/authoring/labs/31673

Thank you for attending this training!

We love your feedback! Please take a minute to complete the survey and help us improve our courses.



Complete Survey: Google Cloud Learning | Evaluation

Google Cloud

Let’s get started!