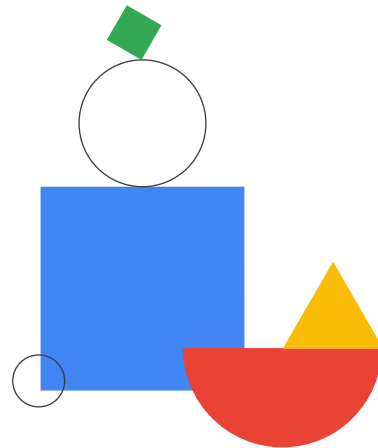


# From Data to Insights with Google Cloud



Course Introduction

Interested in querying billions of records from Wikipedia in seconds? What about how Google BigQuery effectively scales to handle petabytes of data? Well... Welcome to our course: From Data to Insights with Google Cloud .

This class teaches you, the data analyst, how to derive insights through large-scale data analysis and visualization. You'll dive deep into data insights and learn how to use the big data tools on the Google Cloud with interactive labs.

With this class, you'll explore, clean, load, visualize, and extract insights from diverse datasets and you'll pick up some advanced concepts like effective schema design, data cleansing through a powerful new tool, optimizing for query performance, and and more. Let's get started.

# Introductions

## Your instructor

- Organization
- Background
- Course goals

## You

- Name
- Organization
- Job role
- Course goals



# 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

## Prerequisites

1. Basic knowledge of SQL

This class is targeted primarily at data analysts who query their business datasets using SQL and create insightful reports and dashboards.

### Class Questions

How many of you are currently working on business intelligence or as a data or business analyst?

How many of you are more on the data engineering side? Working with infrastructure? Cloud computing to support data analysts or data scientists?

How many are familiar with SQL?

## 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).



## Day 1 Agenda: Exploring, Transforming, and Visualizing Data



- 01 Introduction to Data on Google Cloud
- 02 Analyzing Large Datasets with BigQuery  
Lab: Exploring a BigQuery Public Dataset
- 03 Exploring your Public Dataset with SQL  
Lab: Explore your Ecommerce Dataset with SQL in BigQuery
- 04 Cleaning and Transforming your Data with Dataprep  
Lab: Creating a Data Transformation Pipeline with Dataprep
- 05 Visualizing Insights and Creating Scheduled Queries  
Lab: How to Build a BI Dashboard Using Google Data Studio and BigQuery

This course is broken into three parts that are roughly distributed over the three days of this course.

**“Exploring, Transforming, and Visualizing Data”** is the first part of the course.


Let’s look at the agenda of topics we will cover. We are going to start with the basics of the Google Cloud and why letting the cloud handle your compute and storage needs enable massive scalability.

After the fundamentals of cloud, we will go into the Big Data tools available to you as an analyst (with a focus on BigQuery, Data Studio, and Dataprep to start)


The third module is where we’ll start coding in SQL with BigQuery with interactive examples. Here we’ll introduce our course financial dataset on over 130 million U.S. charities.

In the fourth module we will explore how you can clean and explore with a UI tool called Dataprep.

And in the last module of the day, we will look at how you can visualize your insights and setup a reporting pipeline with scheduled queries.



## Day 2 Agenda: Creating and Optimizing your Data Warehouse



- 06 **Storing and Ingesting New Datasets**  
Lab: Ingesting New Datasets into BigQuery
- 07 **Enriching your Data Warehouse with JOINS**  
Lab: Troubleshooting and Solving Data Join Pitfalls
- 08 **Partitioning your Queries and Tables for Advanced Insights**  
Lab: Creating Date-Partitioned Tables in BigQuery
- 09 **Designing Schemas that Scale: Arrays and Structs in BigQuery**  
Lab: Schema Design for Performance: Arrays and Structs in BigQuery
- 10 **Optimizing Queries for Performance**
- 11 **Controlling Access with Data Security Best Practices**

**“Creating and Optimizing your Data Warehouse”** is the second part of the course.

In this part, we start by looking at how you can bring your own datasets to the cloud.


Followed by how you can enrich your datasets by combining data across multiple tables with SQL joins and unions.

In module 8 we discuss how you can glean advanced insights and performance with partitioned queries and partitioned tables.


And in module 9, how you can design your schema for scale with arrays and structs.

Module 10 provides query performance tips -- and pitfalls to avoid.

And we end this part of the course with data security and permissions for access.



## Day 3 Agenda: Machine Learning for Structured and Unstructured Datasets



**12** Predicting Visitor Return Purchases with  
BigQuery ML  
Lab: Predicting Visitor Purchases with BigQuery ML

**13** Deriving Insights From Unstructured Data Using  
Machine Learning  
Lab: Extract, Analyze, and Translate Text from Images  
with the Cloud ML APIs  
Lab: Classify Images of Clouds in the Cloud with  
AutoML Vision

**“Machine Learning for Structured and Unstructured Datasets”** is the last part of the course.

In this part, we look at how you can create machine learning models using just SQL in BigQuery.

Before exploring the world of unstructured data, like images and free text, while we leverage pre-built ML models like the Vision API and AutoML.



## Lab environment

For each lab, Qwiklabs offers:

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



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

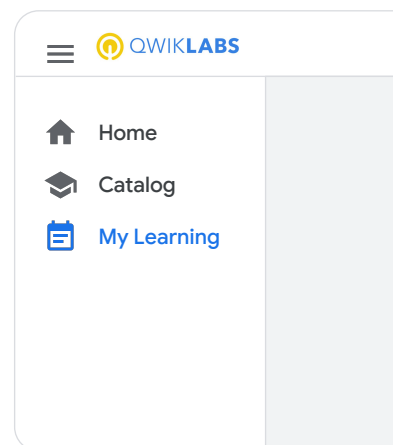
- A free set of resources for a fixed amount of time
- A clean environment with permissions

# 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 **My Learning**.

## 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.

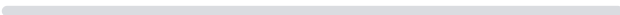









Google Cloud

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**.

## View lecture notes

Labs	Lecture Notes
01	 
02	 
03	 
04	 

You can download  
these as PDF files

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

## View your labs

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

The screenshot shows a user interface with two tabs: 'Labs' (selected) and 'Lecture Notes'. Below the tabs is a list of four lab entries. Each entry consists of a status icon, a progress bar, and a 'Lab Currently Disabled' button. The status icons are: a green checkmark, an empty circle, a triangle with an exclamation mark, and another triangle with an exclamation mark. To the right of the interface, three arrows point to the corresponding lab entries with labels: 'Lab completed' points to the first entry, 'To be completed' points to the second entry, and 'Not yet available' points to the third and fourth entries.

Status	Progress	Notes
Completed	Full	
To be completed	Empty	
Not yet available	Empty	Lab Currently Disabled
Not yet available	Empty	Lab Currently Disabled

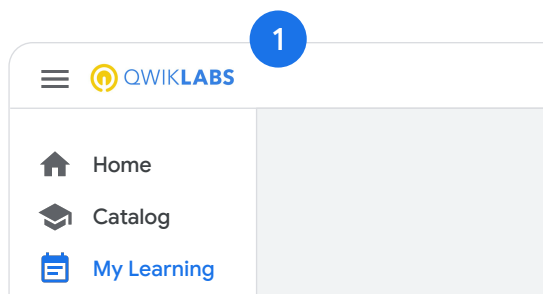
After you launch the course, you can view your labs. The lab list will indicate whether a lab is:

- Completed (by you)
- Active
- Not yet available

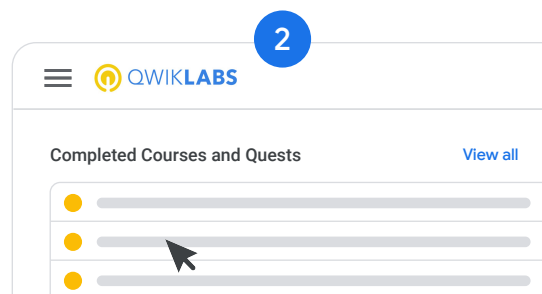
Your instructor will let you know when it's time to launch a lab. Once you start a lab, you won't be able to pause and restart it, so you'll need a continuous block of time to complete the work.

## End of class - Materials

Materials are available for 2 years



Click on **My Learning** 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 **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.

