

Activity Overview

As you have been learning, a data pipeline is a series of processes that transport data from different sources to their final destination for storage and analysis. Data pipelines automate the processes involved in extracting, transforming, combining, validating, and loading data for further analysis and visualization. Effective data pipelines also help eliminate errors and combat system latency. In this activity, you will complete a tutorial in which you will create a streaming pipeline using a Dataflow template. Please note that this activity is optional and will not prevent you from completing this course or receiving your certificate. Google Dataflow and BigQuery are Google Cloud products that you can try at no cost. You will need to fill out your credit card information to begin your trial, but you will not be charged if you follow the free trial's terms. For more information, review this article about free cloud features and trial offer. If you prefer not to open a billing account, you may skip to the next course item. Step-By-Step Instructions

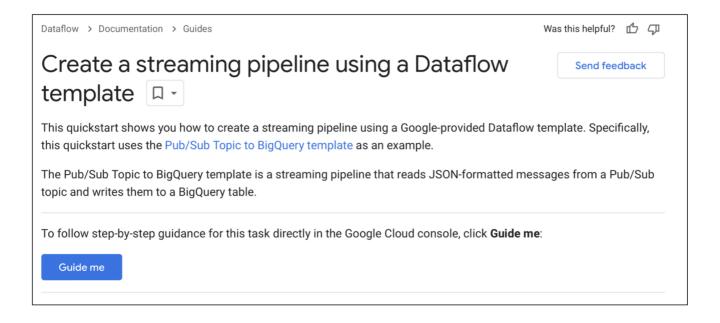
Follow the instructions to complete each step of the activity. Then, answer the questions at the end of the activity before going to the next course item.

Step 1: Log into Google Cloud

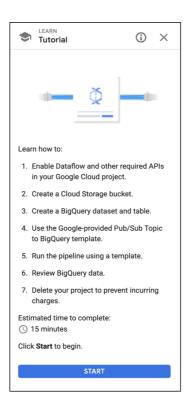
To begin the quickstart guide, log into your Google Cloud account. If you have not created an account, be sure to complete the activity that involved <u>creating a Google Cloud account</u> \Box .

Step 2: Begin the tutorial

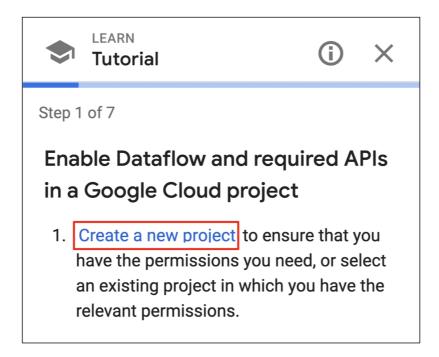
Once you are logged in, open the <u>streaming pipeline template</u> quickstart guide. Then click Guide me.



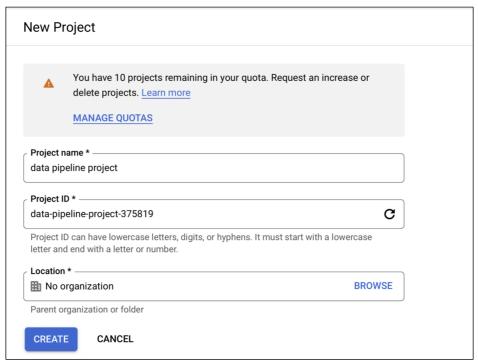
To begin the tutorial, click START.



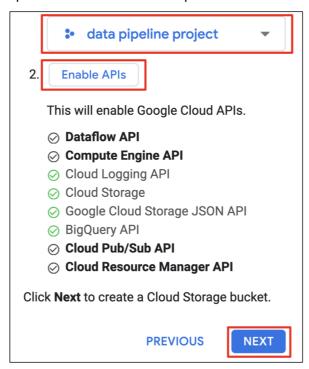
> Step 3: Create a new project and enable APIs For Step 1 of the tutorial, click create a new project.



Give your project a name and project ID. Then click CREATE.



Select your project from the dropdown menu in the tutorial pane and click ENABLE APIs. Then click NEXT.



Step 4: Complete the tutorial

Proceed with the steps of the tutorial, which is estimated to take about 15 minutes. The remaining steps of he tutorial are as follows:

- n Step 2, you will set up a Cloud Storage bucket. Buckets are the basic containers that hold your data.
- n Step 3, you will create a BigQuery dataset and BigQuery table within your project.
- n Steps 4 and 5, you will be guided through running the pipeline. You'll use an existing template that runs a pipeline from Pub/Sub to BigQuery. Pub/Sub is a cloud service that opens a communication channel for a publishing provider and subscriber for a data source. It is used to ingest data before moving it to BigQuery with a data pipeline.
- n Step 6, you will run a query on the table that now holds the data you brought in from Pub/Sub using the Dataflow template.
- n Step 7, you will delete your project. Make sure to complete this step in order to avoid incurring charges on our Google Cloud account.

1.	Did you complete this activity?	1 / 1 point
	YesNo	
	Correct Thank you for completing this activity! Creating a pipeline in Dataflow is a great example of how you'll transform and move data from a source to a table. Please complete the following quiz questions and review the feedback.	
2.	What are some benefits of data pipelines? Select all that apply.	1 / 1 point
	Automate data analysis processes	
	Correct Data pipelines automate the processes involved in extracting, transforming, combining, validating, and loading data for further analysis and visualization. Effective data pipelines also help eliminate errors and combat system latency.	
	Prevent system latency	
	Correct Data pipelines automate the processes involved in extracting, transforming, combining, validating, and loading data for further analysis and visualization. Effective data pipelines also help eliminate errors and combat system latency.	
	Assign roles cross-team	
	Help eliminate errors	
	Ocrrect Data pipelines automate the processes involved in extracting, transforming, combining, validating, and loading data for further analysis and visualization. Effective data pipelines also help eliminate errors and combat system latency.	
3.	While working for a client, you need to create a streaming pipeline in Dataflow. What template could you use to quickly move Pub/Sub data to BigQuery?	1 / 1 point
	The Pub/Sub Topic to BigQuery template	
	The Dataflow to BigQuery template	
	The Cloud Storage to BigQuery template	
	The Compute Engine to BigQuery template	
	Correct The Pub/Sub Topic to BigQuery template is a streaming pipeline that reads JSON-formatted messages from a Pub/Sub topic and writes them to a BigQuery table. You can use this template as a quick solution to move Pub/Sub data to BigQuery.	