Dataflow: Qwik Start - Templates

GSP192

Google Cloud Self-Paced Labs

- This lab teaches creating streaming pipelines using Google's pre-built Dataflow templates, specifically Pub/Sub to BigQuery.

- The template reads JSON messages from a Pub/Sub topic and streams them directly into a BigQuery table.

- Choose either Cloud Shell command line or Cloud Console to create the required BigQuery dataset and table.

- Set up a Cloud Storage bucket to store pipeline artifacts and temporary files.

- Launch the complete streaming pipeline using the Pub/Sub to BigQuery Dataflow template configuration.

Contents

Dataflow: Qwik Start - Templates

```
gcloud auth list
```

4. Click **Authorize**.

Output:

```
ACTIVE: *
ACCOUNT: student-02-77569a450c62@qwiklabs.net

To set the active account, run:
    $ gcloud config set account `ACCOUNT`
```

5. (Optional) You can list the project ID with this command:

```
gcloud config list project
```

01:06:27 ⏱ Time limit

End Lab

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username

student-02-77569a450c62€

Password

1gd108qixhdc

Project ID

qwiklabs-gcp-02-e70062f1

Task 1. Ensure that the Dataflow API is successfully re-enabled

Task 2. Create a BigQuery dataset, BigQuery table, and Cloud Storage bucket using Cloud Shell

Task 3. Create a BigQuery dataset, BigQuery table, and Cloud Storage bucket using the Google Cloud console

Task 4. Run the pipeline

Task 5. Submit a query

Task 6. Test your understanding

0/100

< Previous   Next >

```
Welcome to Cloud Shell! Type "help" to get started, or type "gemini" to try prompting with Gemini CLI.
Your Cloud Platform project in this session is set to qwiklabs-gcp-02-e70062ff8008.
Use `gcloud config set project [PROJECT_ID]` to change to a different project.
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$ gcloud auth list
Credentialed Accounts

ACTIVE: *
ACCOUNT: student-02-77569a450c62@qwiklabs.net

To set the active account, run:
    $ gcloud config set account `ACCOUNT`

student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$ gcloud config list project
[core]
project = qwiklabs-gcp-02-e70062ff8008

Your active configuration is: [cloudshell-31149]
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$
```

Contents

Dataflow: Qwik Start - Templates

Lab setup instructions and requirements

End Lab

01:04:54
⊘ Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username

student-02-77569a450c62(

Password

1gd108qixhdc

Project ID

qwiklabs-gcp-02-e70062f1

Task 1. Ensure that the Dataflow API is successfully re-enabled

To ensure access to the necessary API, restart the connection to the Dataflow API.

- In Cloud Shell, run the following commands to reset the Dataflow API by disabling and re-enabling it for your project.

```
gcloud services disable dataflow.googleapis.com --project
qwiklabs-gcp-02-e70062ff8008 --force
gcloud services enable dataflow.googleapis.com --project
qwiklabs-gcp-02-e70062ff8008
```

Click **Check my progress** to verify the objective.

Disable and re-enable the Dataflow API.

Check my progress

Task 1. Ensure that the Dataflow API is successfully re-enabled

Task 2. Create a BigQuery dataset, BigQuery table, and Cloud Storage bucket using Cloud Shell

Task 3. Create a BigQuery dataset, BigQuery table, and Cloud Storage bucket using the Google Cloud console

Task 4. Run the pipeline

Task 5. Submit a query

Task 6. Test your understanding

Congratulations!

0/100

< Previous

Next >

student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$ gcloud services disable dataflow.googleapis.com --project qwiklabs-gcp-02-e70062ff8008 --force
gcloud services enable dataflow.googleapis.com --project qwiklabs-gcp-02-e70062ff8008
Operation "operations/acat.p17-312191275268-309a3ef5-a66a-45d8-b4a5-5c7707c0c230" finished successfully.
Operation "operations/acf.p2-312191275268-c9a9a01e-89d4-4527-84db-63ba53fcedbd" finished successfully.
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$

Google Skills | Partner

★ 6325   🔥 0   ?   🌐

Contents

Dataflow: Qwik Start - Templates ▾

Lab setup instructions and requirements >

End Lab   01:03:43
? Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username
student-02-77569a450c62(

Password
1gd108qixhdc

Project ID
qwiklabs-gcp-02-e70062f1

Task 2. Create a BigQuery dataset, BigQuery table, and Cloud Storage bucket using Cloud Shell

Let's first create a BigQuery dataset and table.

> **Note:** This task uses the `bq` command-line tool. **Skip down** to Task 3 if you want to complete these steps using the Cloud console.

1. Run the following command to create a dataset called `taxirides`:

```
bq mk taxirides
```

Your output should look similar to:

```
Dataset '<myprojectid:taxirides>' successfully created
</myprojectid:taxirides>
```

Task 1. Ensure that t   20/100
Dataflow API is
successfully re-enabled

Task 2. Create a BigQuery
dataset, BigQuery table,
and Cloud Storage
bucket using Cloud Shell

Task 3. Create a BigQuery
dataset, BigQuery table,
and Cloud Storage
bucket using the Google
Cloud console

Task 4. Run the pipeline

Task 5. Submit a query

Task 6. Test your
understanding

Congratulations!

< Previous   Next >

1. Run the following command to create a dataset called `taxirides`:

```
bq mk taxirides
```

Your output should look similar to:

```
Dataset '<myprojectid:taxirides>' successfully created
</myprojectid:taxirides>
```

Click **Check my progress** to verify the objective.

✅ Create a BigQuery Dataset (name: taxirides).

Check my progress

Assessment Completed!

Now that you have your dataset created, you'll use it in the following step to instantiate a

```
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$ bq mk taxirides
Dataset 'qwiklabs-gcp-02-e70062ff8008:taxirides' successfully created.
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$
```

Contents

Dataflow: Qwik Start - Templates

Lab setup instructions and requirements

End Lab

01:02:24
? Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username
student-02-77569a450c62(

Password
1gd108qixhdc

Project ID
qwiklabs-gcp-02-e70062f1

2. Run the following command to do so:

```
bq mk \
--time_partitioning_field timestamp \
--schema
ride_id:string,point_idx:integer,latitude:float,longitude:float, \
timestamp:timestamp,meter_reading:float,meter_increment:float,ride_
passenger_count:integer -t taxirides.realtime
```

Your output should look similar to:

```
Table 'myprojectid:taxirides.realtime' successfully created
```

Click **Check my progress** to verify the objective.

✓ Create a table in BigQuery Dataset.

Check my progress

Assessment Completed!

< Previous

Next >

Task sidebar

40/100

Task 1. Ensure that t Dataflow API is successfully re-enabled

Task 2. Create a BigQuery dataset, BigQuery table, and Cloud Storage bucket using Cloud Shell

Task 3. Create a BigQuery dataset, BigQuery table, and Cloud Storage bucket using the Google Cloud console

Task 4. Run the pipeline

Task 5. Submit a query

Task 6. Test your understanding

Congratulations!

```
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$ bq mk taxirides
Dataset 'qwiklabs-gcp-02-e70062ff8008:taxirides' successfully created.
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$ bq mk \
--time_partitioning_field timestamp \
--schema ride_id:string,point_idx:integer,latitude:float,longitude:float,\
timestamp:timestamp,meter_reading:float,meter_increment:float,ride_status:string,\
passenger_count:integer -t taxirides.realtime
Table 'qwiklabs-gcp-02-e70062ff8008:taxirides.realtime' successfully created.
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$
```

Contents

Dataflow: Qwik Start - Templates

Lab setup instructions and requirements

End Lab

01:01:45
⏱ Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username
student-02-77569a450c62(

Password
1gd108qixhdc

Project ID
qwiklabs-gcp-02-e70062f1

On its face, the `bq mk` command looks a bit complicated. However, with some assistance from the BigQuery command-line documentation, we can break down what's going on here. For example, the documentation tells us a little bit more about **schema**:

- Either the path to a local JSON schema file or a comma-separated list of column definitions in the form `[FIELD]:[DATA_TYPE]`, `[FIELD]:[DATA_TYPE]`.

In this case, we are using the latter—a comma-separated list.

Create a Cloud Storage bucket using Cloud Shell

Now that we have our table instantiated, let's create a bucket.

Use the Project ID as the bucket name to ensure a globally unique name: `qwiklabs-gcp-02-e70062ff8008`

- Run the following commands to do so:

```
export BUCKET_NAME=qwiklabs-gcp-02-e70062ff8008
```

Previous

Next

Contents

Dataflow: Qwik Start - Templates

Lab setup instructions and requirements

End Lab 01:01:12 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username
student-02-77569a450c62

Password
1gd108qixhdc

Project ID
qwiklabs-gcp-02-e70062f1

Create a Cloud Storage bucket using Cloud Shell

Now that we have our table instantiated, let's create a bucket.

Use the Project ID as the bucket name to ensure a globally unique name: `qwiklabs-gcp-02-e70062ff8008`

- Run the following commands to do so:

```
export BUCKET_NAME=qwiklabs-gcp-02-e70062ff8008
```

```
gsutil mb gs://$BUCKET_NAME/
```

Click **Check my progress** to verify the objective.

✓ Create a Cloud Storage bucket.

Check my progress

Assessment Completed!

60/100

Task 1. Ensure that t Dataflow API is successfully re-enabled

Task 2. Create a BigQuery dataset, BigQuery table, and Cloud Storage bucket using Cloud Shell

Task 3. Create a BigQuery dataset, BigQuery table, and Cloud Storage bucket using the Google Cloud console

Task 4. Run the pipeline

Task 5. Submit a query

Task 6. Test your understanding

Congratulations!

< Previous

Next >

https://console.cloud.google.com/home/dashboard?project=qwiklabs-gcp-02-e70062ff8008&pli=1&cloudshell=true

Free AI Paraphrasin... | Transcript of Case St... | GCP-LAB | Hydra | datawarehouse | Products

Google Cloud | qwiklabs-gcp-02-e70062ff8008 | Search (/) for resources, docs, products, and more | Search

CLOUD SHELL
Terminal | (qwiklabs-gcp-02-e70062ff8008) ✕ | + ▾ | ✏️ Open Editor

ⓘ Gemini CLI is available in Cloud Shell terminal! Type gemini to try it. Learn more | Don't show again | Dismiss

```
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$ bq mk taxirides
Dataset 'qwiklabs-gcp-02-e70062ff8008:taxirides' successfully created.
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$ bq mk \
--time_partitioning_field timestamp \
--schema ride_id:string,point_idx:integer,latitude:float,longitude:float,\
timestamp:timestamp,meter_reading:float,meter_increment:float,ride_status:string,\
passenger_count:integer -t taxirides.realtime
Table 'qwiklabs-gcp-02-e70062ff8008:taxirides.realtime' successfully created.
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$ export BUCKET_NAME=qwiklabs-gcp-02-e70062ff8008
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$ gsutil mb gs://$BUCKET_NAME/
Creating gs://qwiklabs-gcp-02-e70062ff8008/...
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$
```

Contents

Dataflow: Qwik Start - Templates

Task 4. Run the pipeline

Deploy the Dataflow Template:

```
gcloud dataflow jobs run iotflow \
    --gcs-location gs://dataflow-templates-us-west1/latest/PubSub_t
    --region us-west1 \
    --worker-machine-type e2-medium \
    --staging-location gs://qwiklabs-gcp-02-e70062ff8008/temp \
    --parameters inputTopic=projects/pubsub-public-data/topics/taxi
```

In the **Navigation menu**, click **View All Products** > **Analytics** > **Dataflow** > **Jobs**. You will see your dataflow job.

Please refer the document for more information.

> **Note:** You may need to wait a minute for the activity tracking to complete.

Click **Check my progress** to verify the objective.

```
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$ gcloud dataflow jobs run iotflow \
    --gcs-location gs://dataflow-templates-us-west1/latest/PubSub_to_BigQuery \
    --region us-west1 \
    --worker-machine-type e2-medium \
    --staging-location gs://qwiklabs-gcp-02-e70062ff8008/temp \
    --parameters inputTopic=projects/pubsub-public-data/topics/taxirides-realtime,outputTableSpec=qwiklabs-gcp-02-e70062ff8008:taxirides.realtime
createTime: '2025-12-27T10:15:14.792462Z'
currentStateTime: '1970-01-01T00:00:00Z'
id: 2025-12-27_02_15_13-8394203768905264500
location: us-west1
name: iotflow
projectId: qwiklabs-gcp-02-e70062ff8008
startTime: '2025-12-27T10:15:14.792462Z'
type: JOB_TYPE_STREAMING
student_02_77569a450c62@cloudshell:~ (qwiklabs-gcp-02-e70062ff8008)$
```

Contents

Dataflow: Qwik Start - Templates

Lab setup instructions and requirements

End Lab 00:57:16 ⏱ Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username

student-02-77569a450c62(

Password

1gd108qixhdc

Project ID

qwiklabs-gcp-02-e70062f1

Task 5. Submit a query

You can submit queries using standard SQL.

1. In the BigQuery **Editor**, add the following to query the data in your project:

```
SELECT * FROM `qwiklabs-gcp-02-e70062ff8008.taxirides.realtime` LIM
```

2. Now click **RUN**.

If you run into any issues or errors, run the query again (the pipeline takes a minute to start up.)

3. When the query runs successfully, you'll see the output in the **Query Results** panel as shown below:

Query results   ⬆ SAVE AS ▾   📊 EXPLORE IN DATA STUDIO

Query complete (2.116 sec elapsed, 0 B processed)

Job information   Results   JSON   Execution details

Row	ride_id	point_idx	latitude	longitude	timestamp
1	b0810fbd-78a8-4159-b9ff-963695e2a23d	225	40.753550000000004	-73.98504000000001	2018-07-25 23:28:20.870530 UTC

100/100

< Previous   Next >

SELECT * FROM `qwiklabs-gcp-02-e70062ff8008.taxirides.realtime` LIMIT 1000

This query will process 0 B when run.

Query results

Row	ride_id	point_idx	latitude	longitude	timestamp
1	8814a53e-6a17-462a-b658-8cd...	618	40.68679	-73.9756400000...	2025-12-27 10:15:32.732
2	df84cdbf-7909-4c5f-8747-d915...	204	40.64545000000...	-73.79948	2025-12-27 10:15:32.754
3	6d380912-0a34-4606-9afd-484...	232	40.76911000000...	-73.88187	2025-12-27 10:15:37.877
4	14493bc6-9d0a-4e00-bc97-1c6...	178	40.72806000000...	-74.00203	2025-12-27 10:15:32.671
5	ddf0f51d-6c57-4e0e-8fb4-a9eb...	2534	40.74160000000...	-73.95552	2025-12-27 10:15:37.868
6	a3193e2e-ed6c-4730-88e4-573...	151	40.75144	-73.9903200000...	2025-12-27 10:15:32.616

Results per page: 50   1 – 50 of 1000

Job history   Show

Contents

Dataflow: Qwik Start - Templates

Lab setup instructions and requirements

End Lab 00:56:03 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username
student-02-77569a450c62(

Password
1gd108qixhdc

Project ID
qwiklabs-gcp-02-e70062f1

2. Now click **RUN**.

If you run into any issues or errors, run the query again (the pipeline takes a minute to start up.)

3. When the query runs successfully, you'll see the output in the **Query Results** panel as shown below:

Query results	⬇ SAVE AS ▼	📊 EXPLORE IN DATA STUDIO

Query complete (2.116 sec elapsed, 0 B processed)

Job information	**Results**	JSON	Execution details

Row	ride_id	point_idx	latitude	longitude	timestamp
1	b0810fbd-78a8-4159-b9ff-963695e2a23d	225	40.753550000000004	-73.98504000000001	2018-07-25 23:28:20.870530 UTC
2	1a10dc8b-3623-41bf-938a-9fca26c2ae10	311	40.752930000000006	-73.96584	2018-07-25 23:24:10.608380 UTC
3	5253c100-1a30-4a3e-89ee-6c0c861cf44f	224	40.74331	-73.99172	2018-07-25 23:26:34.636480 UTC
4	3efa96c2-4695-4c0b-96b6-da33a4b74ccf	8	40.7533	-73.97832000000001	2018-07-25 23:24:06.823150 UTC
5	d6d37615-ccba-4416-9932-e956e0f0ba65	747	40.682140000000004	-74.00594000000001	2018-07-25 23:24:10.103770 UTC

Great work! You just pulled 1000 taxi rides from a Pub/Sub topic and pushed them to a BigQuery table. As you saw firsthand, templates are a practical, easy-to-use way to run Dataflow jobs. Be sure to check out, in the Dataflow Documentation, some other Google Templates in the Get started with Google-provided templates Guide.

100/100

< Previous   Next >