

# STREAMING DATA FROM PUB/SUB TO BIGQUERY USING DATAFLOW

---

## OVERVIEW:

This guide outlines the process of creating a streaming data pipeline that reads JSON-encoded messages from a Pub/Sub topic, processes them using Dataflow, and writes the results to a BigQuery table.

## PREREQUISITES:

- A Google Cloud project with billing enabled.
- Enable the following APIs:
  - Compute Engine API
  - Dataflow API
  - Cloud Logging API
  - Big Query API
  - Pub/Sub API
  - Cloud Storage API
- **IAM ROLES:**
  - Your user account: Service Account User
  - Compute Engine default service account:
    - Dataflow Worker
    - Dataflow Admin
  - Pub/Sub Editor
  - Storage Object Admin
  - BigQuery Data Editor

## STEP 1: CREATE A CLOUD STORAGE BUCKET:

1. Go to Cloud Storage > Buckets.
2. Click 'Create'.
3. Name your bucket uniquely.
4. Choose location type and location.
5. Click 'Create'.

## STEP 2: CREATE A PUB/SUB TOPIC AND SUBSCRIPTION:

1. Navigate to Pub/Sub > Topics.
2. Click 'Create Topic'.

3. Enter a Topic ID.
4. Click 'Create Topic'.

### STEP 3: CREATE A BIGQUERY DATASET AND TABLE:

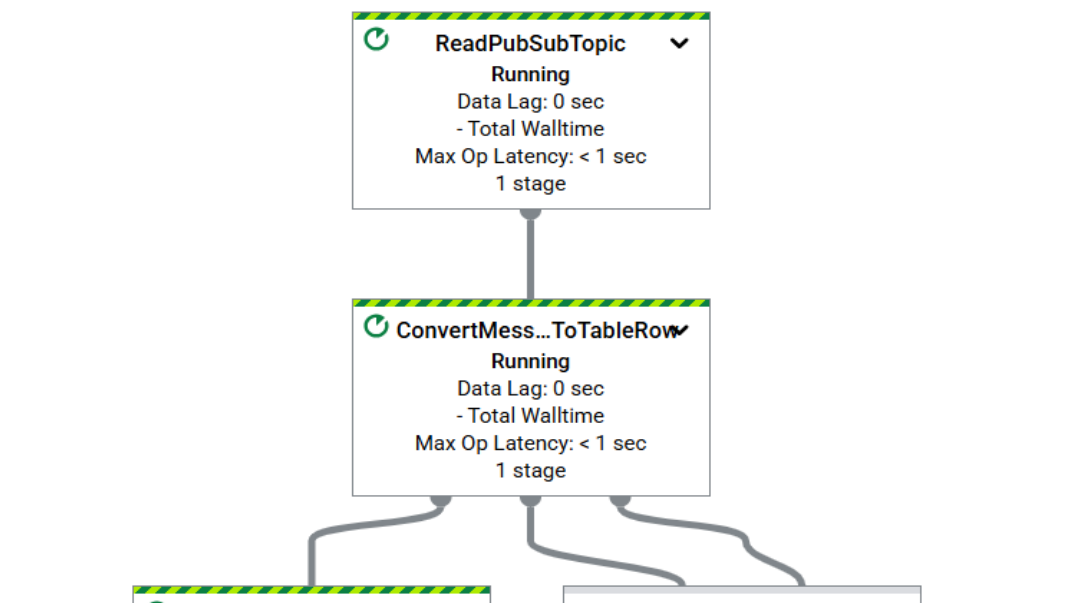
1. Go to BigQuery in the Console.
2. Click your project name in the Explorer.
3. Click 'Create Dataset'.
4. Enter a Dataset ID and click 'Create'.
5. Within the dataset, click 'Create Table'.
6. Choose 'Empty table'.
7. Enter Table name and schema.
8. Click 'Create Table'.

### STEP 4: DEPLOY THE DATAFLOW PIPELINE:


1. Go to Dataflow > Jobs.
2. Click 'Create Job from Template'.
3. Provide a Job name and region.
4. Select template: 'Pub/Sub to BigQuery'.
5. Set the temp location in your Cloud Storage bucket.
6. Click 'Run Job'.



CLE







 **WriteSuccessfulRecords** ▼

**Running**  
Data Lag: 0 sec  
- Total Walltime  
Max Op Latency: < 1 sec  
2 stages

**Flatten**  
Data Lag: 0 sec  
- Total Walltime  
Max Op Latency: < 1 sec  
0 stages


 **WrapInsertionErrors** ▼

**Running**  
Data Lag: 0 sec  
- Total Walltime  
Max Op Latency: < 1 sec  
1 stage


 **WriteFailedRecords** ▼

**Running**  
Data Lag: 0 sec  
- Total Walltime  
Max Op Latency: < 1 sec  
2 stages




 **WrapInsertionErrors** ▼

**Running**  
Data Lag: 0 sec  
- Total Walltime  
Max Op Latency: < 1 sec  
1 stage

 **WriteFailedRecords** ▼

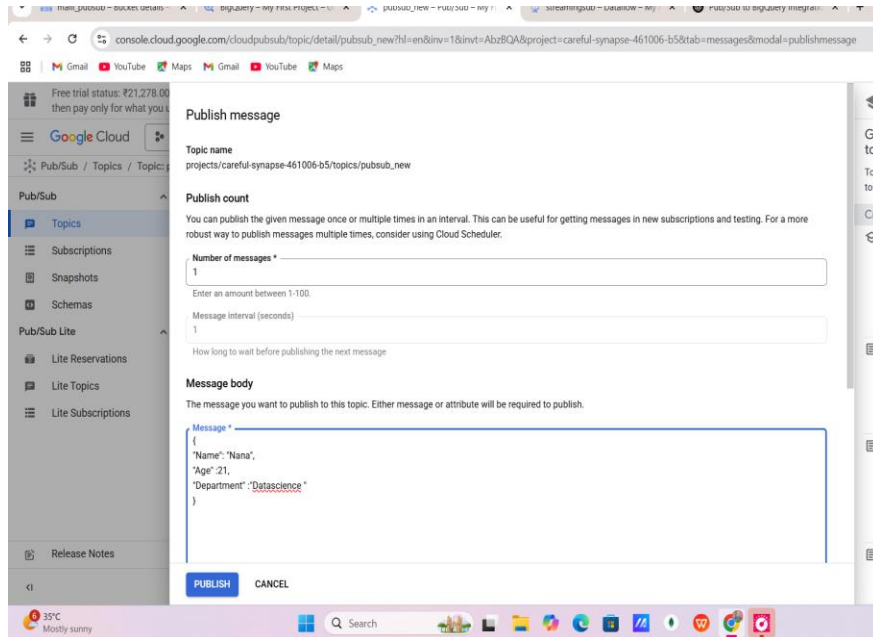
**Running**  
Data Lag: 0 sec  
- Total Walltime  
Max Op Latency: < 1 sec  
2 stages

 **WriteFailedRecords2** ▼

**Running**  
Data Lag: 0 sec  
- Total Walltime  
Max Op Latency: < 1 sec  
2 stages

## STEP 5: PUBLISH MESSAGES TO PUB/SUB:

1. Go to Pub/Sub > Topics.
2. Click your topic.
3. Click 'Publish Message'.
4. Enter a JSON message.
5. Click 'Publish'.



## STEP 6: VERIFY DATA IN BIGQUERY:

1. Go to BigQuery in the Console.
2. Expand your dataset in the Explorer.
3. Click your table.
4. Click 'Preview' to view the ingested data.

Free trial status: ₹21,278.00 credit and 84 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

Google Cloud My First Project Search (/) for resources, docs, products, and more Search

Explorer + Add data

Search BigQuery resources

careful-synapse-461006-b5

newpubsub

pubsubnew

Show more

Untitled query

```
1 select * from 'careful-synapse-461006-b5.newpubsub.pubsubnew'
```

Query completed

Query results

Save results Open in

Job information Results Chart JSON Execution details Execution graph

Row	Name	Age	Department
1	Nana	21	Datascience
2	Sona	21	Computerscience
3	Sona	21	Computerscience
4	Sona	21	Computerscience
5	Arjun	22	Engineering

Results per page: 50 1 - 5 of 5

Repository Preview Job history Refresh

33°C Mostly sunny

Search

ENG IN 13:01 02-06-2025