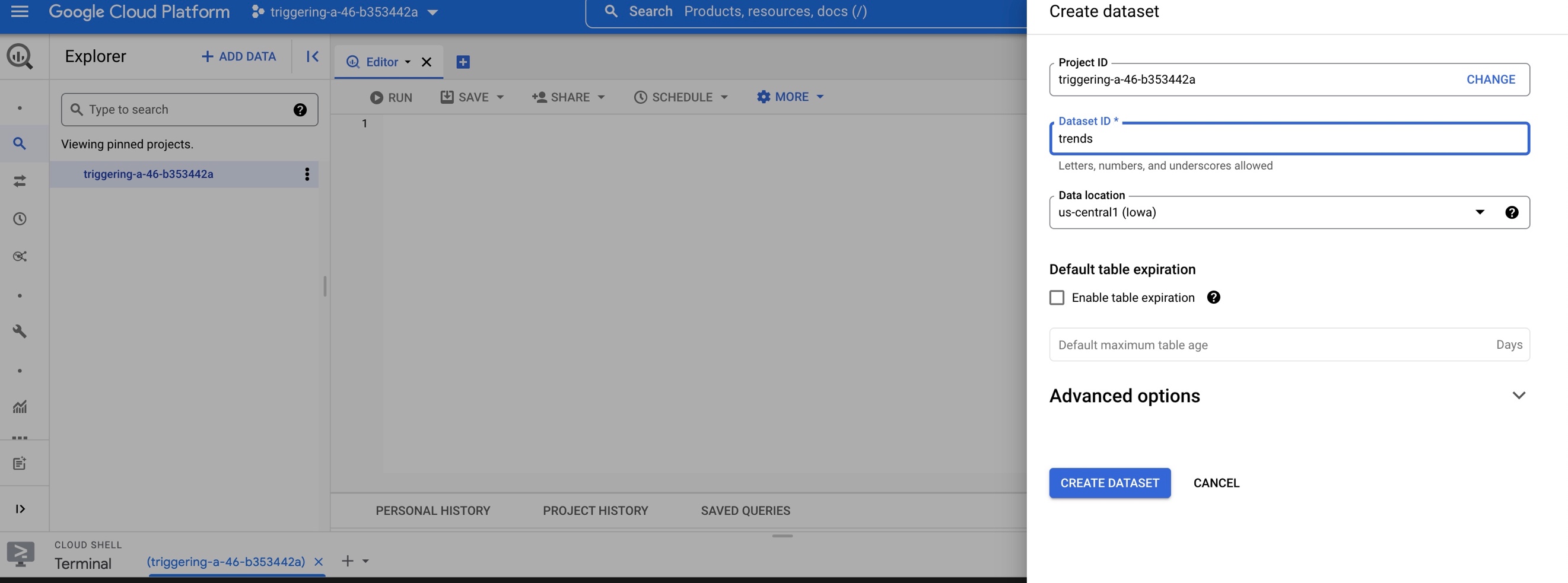
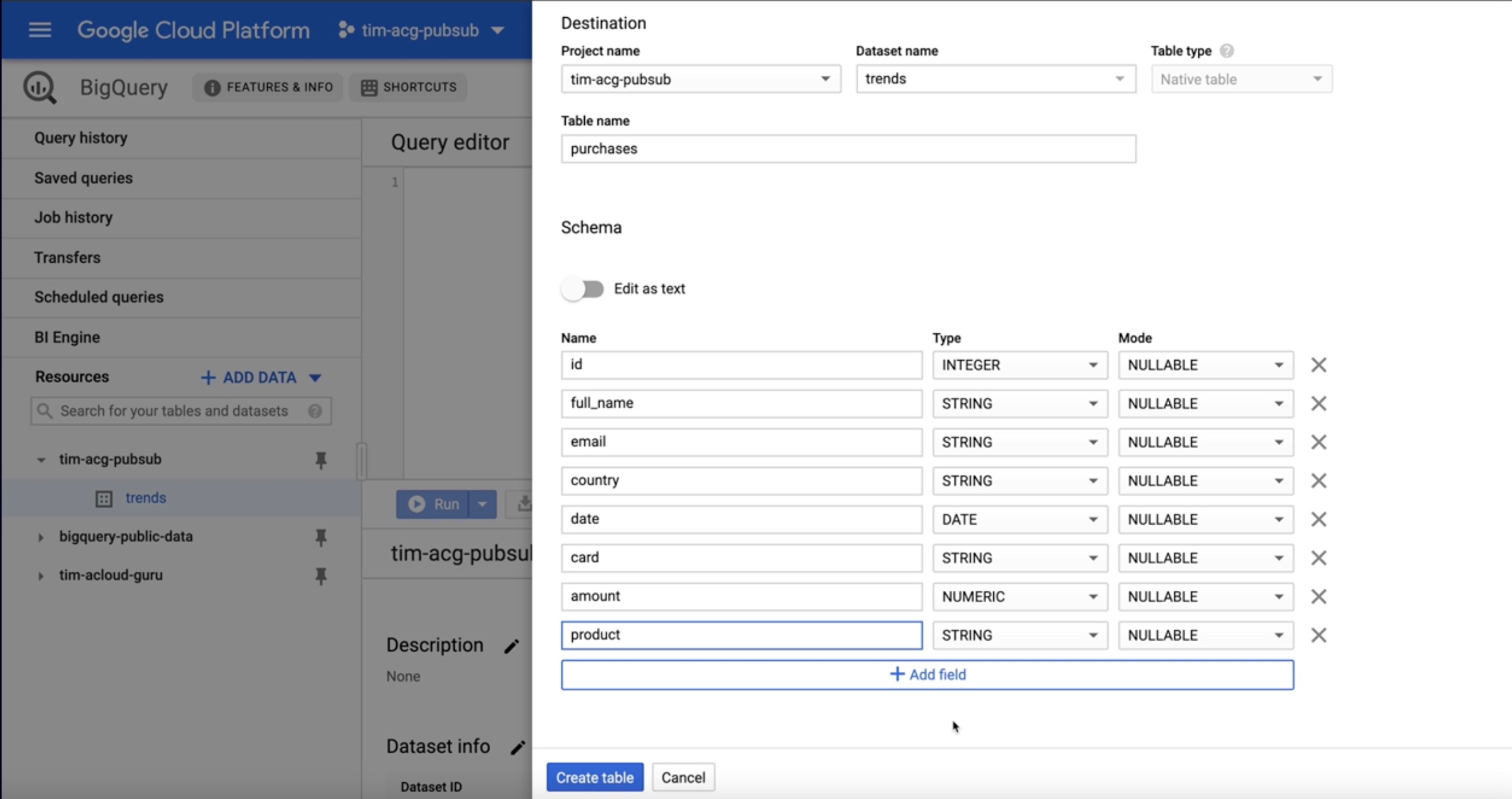
**Pub/Sub -> DataFlow -> BigQuery**



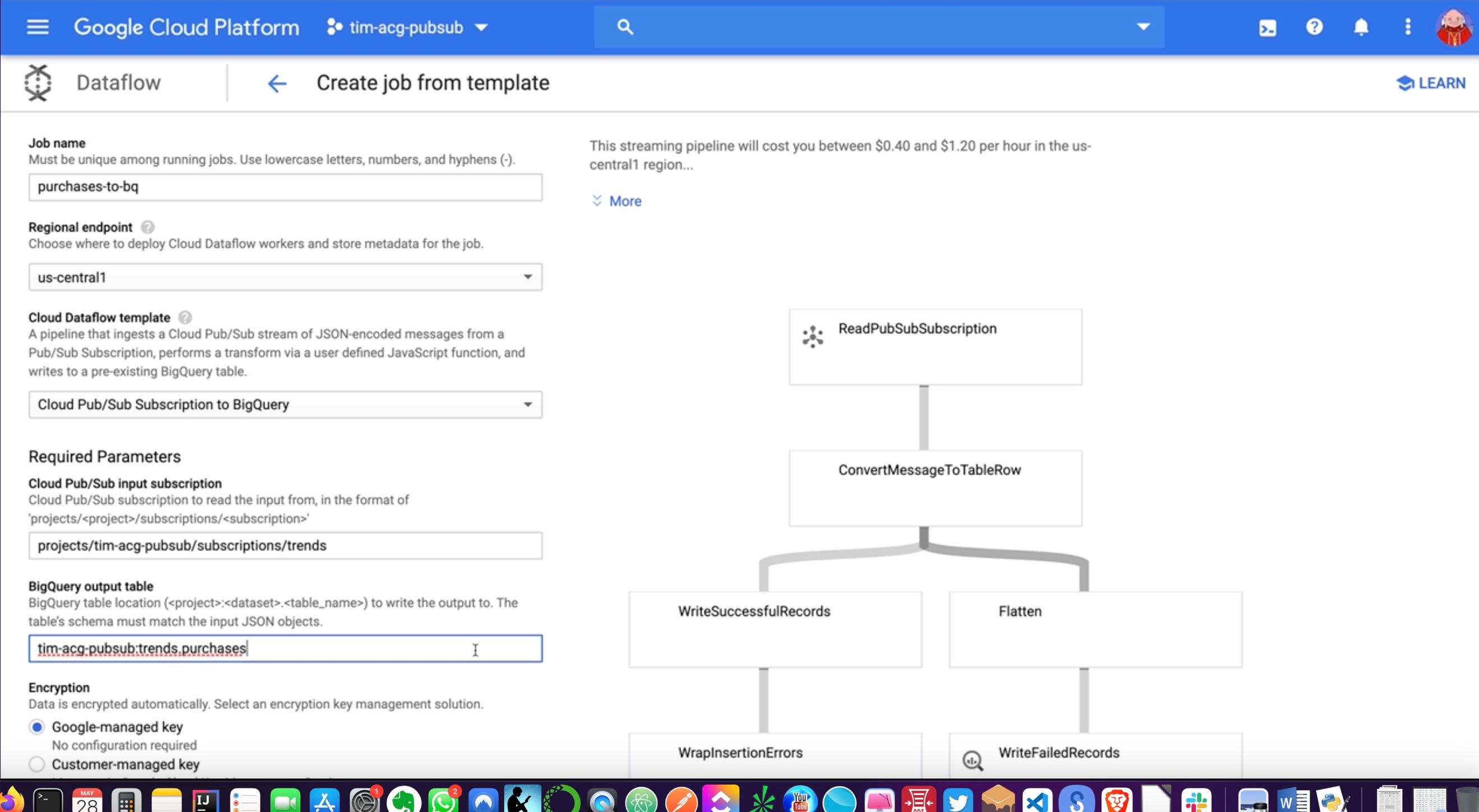
1. Create a bucket – **PROJECT\_NAME-temp**
2. Create a Pub/Sub Topic name – **purchases** (projects/triggering-a-46-b353442a/topics/purchases)
3. Create a Pub/Sub Subscription for the topic – **trends** (projects/triggering-a-46-b353442a/subscriptions/trends)
4. BigQuery -> Create dataset – **trends**

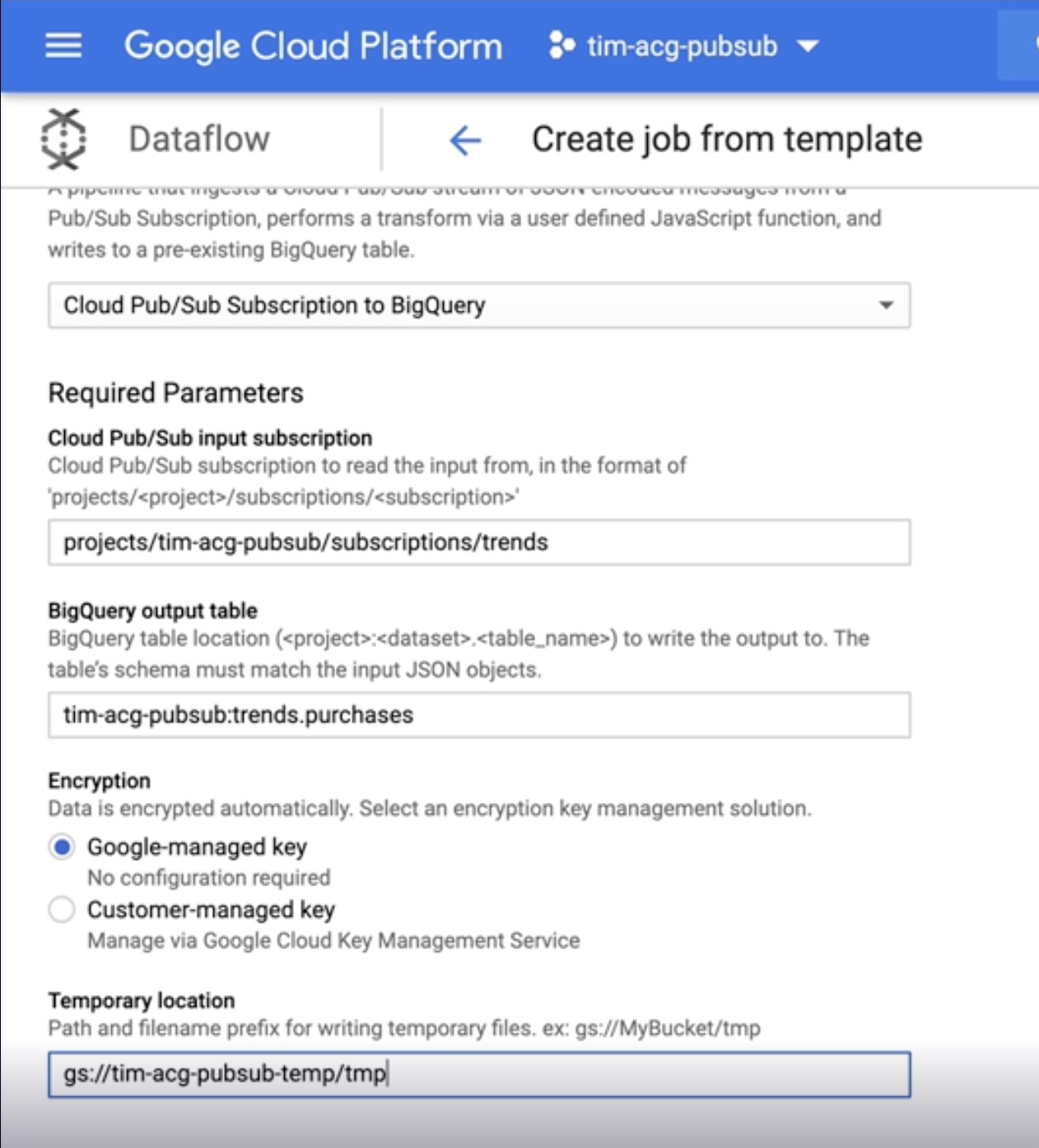


1. Create a table – **purchases** in the dataset with below fields



1. Once the table is created, go to the table-> details -> note the Table-ID. Table is ready to receive the data.
2. DataFlow -> Create job from template
   1. Job name : purchases-to-bq
   2. Cloud Dataflow template : Cloud Pub/Sub Subscription to BigQuery
   3. Cloud Pub/Sub input Subscription : projects/triggering-a-46-b353442a/subscriptions/trends
   4. BigQuery output table : project\_name:trends-purchases (Table-ID)
   5. Temporary location : PROJECT\_NAME-temp/tmp/ (easy to manage)





1. Run Job
2. Clone the repo

git clone <https://github.com/manisekharan/gcp_pubsub_dataflow_bigquery.git>

1. Update project name, topic name and file location (same level as transactions.py)
2. Install google-cloud-pubsub library

sudo pip3 install google-cloud-pubsub

1. Run the script to send the data to pubsub

python3 transactions.py

1. Data is loaded to BigQuery table for analysis

