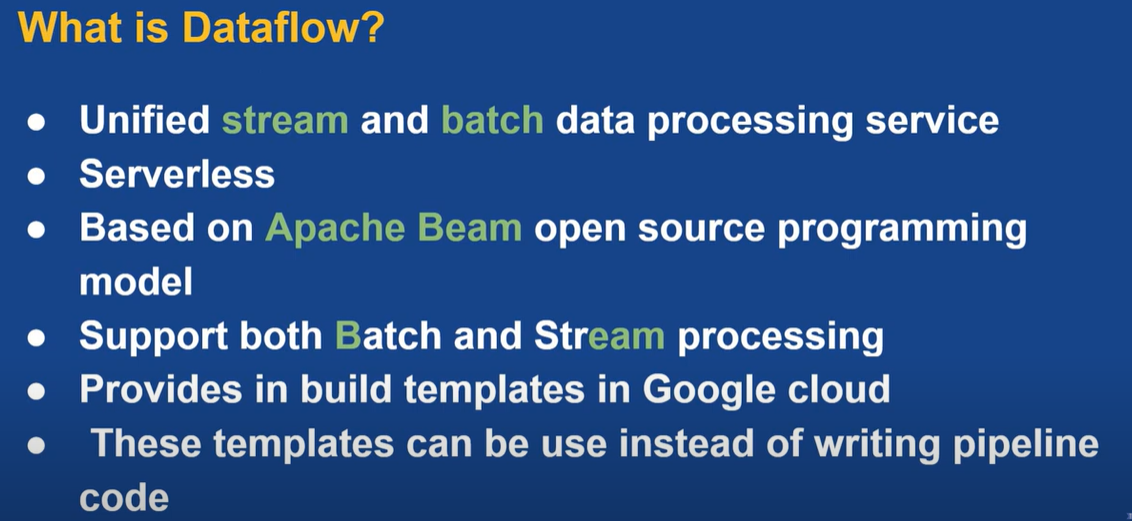
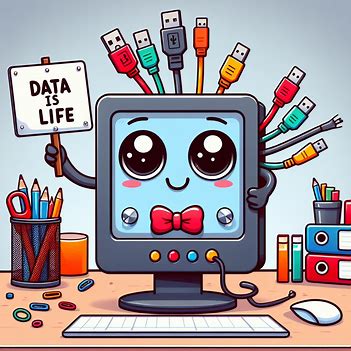
Google Cloud Dataflow

Google Cloud Dataflow is a fully-managed, scalable data processing service for executing batch, stream, and ETL processing patterns. The process of data transfer from Google Cloud Storage to Big Query is known dataflow.









Steps to create dataflow

The steps to create a dataflow to read data source from GCP bucket insert into Big query.

Create GCP bucket

Create Json and [**User-Defined Function**](https://www.techopedia.com/definition/3888/user-defined-function-udf) (UDF) data

Save Json & UDF data into bucket

Create job from template

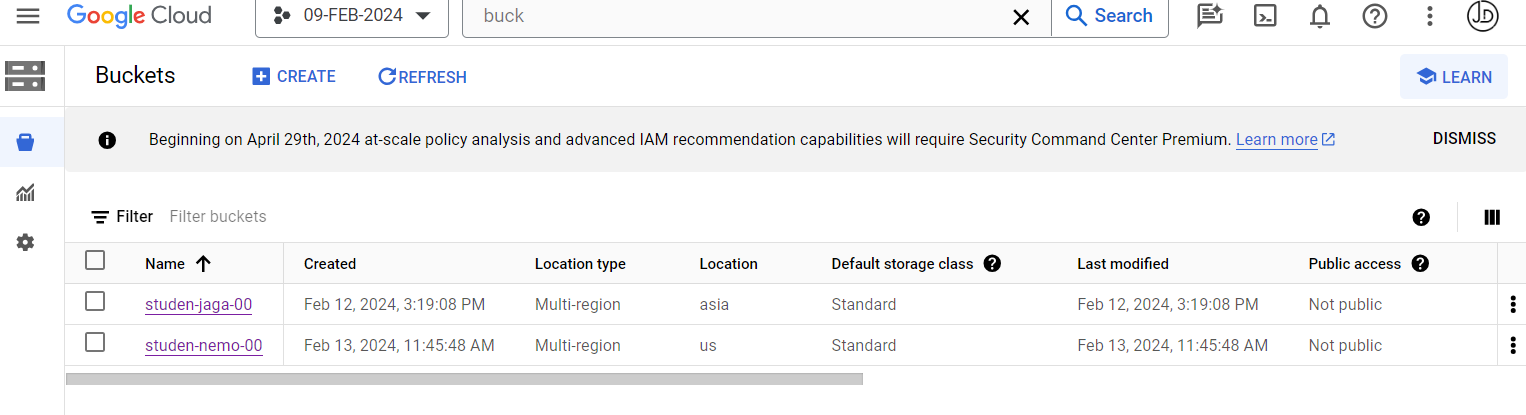
Insert data into bigquery

GCP Bucket

[A GCP (Google Cloud Platform) bucket is a basic container that holds your data in Google Cloud Storage](https://cloud.google.com/storage/docs/buckets). Here are some key points about GCP buckets:

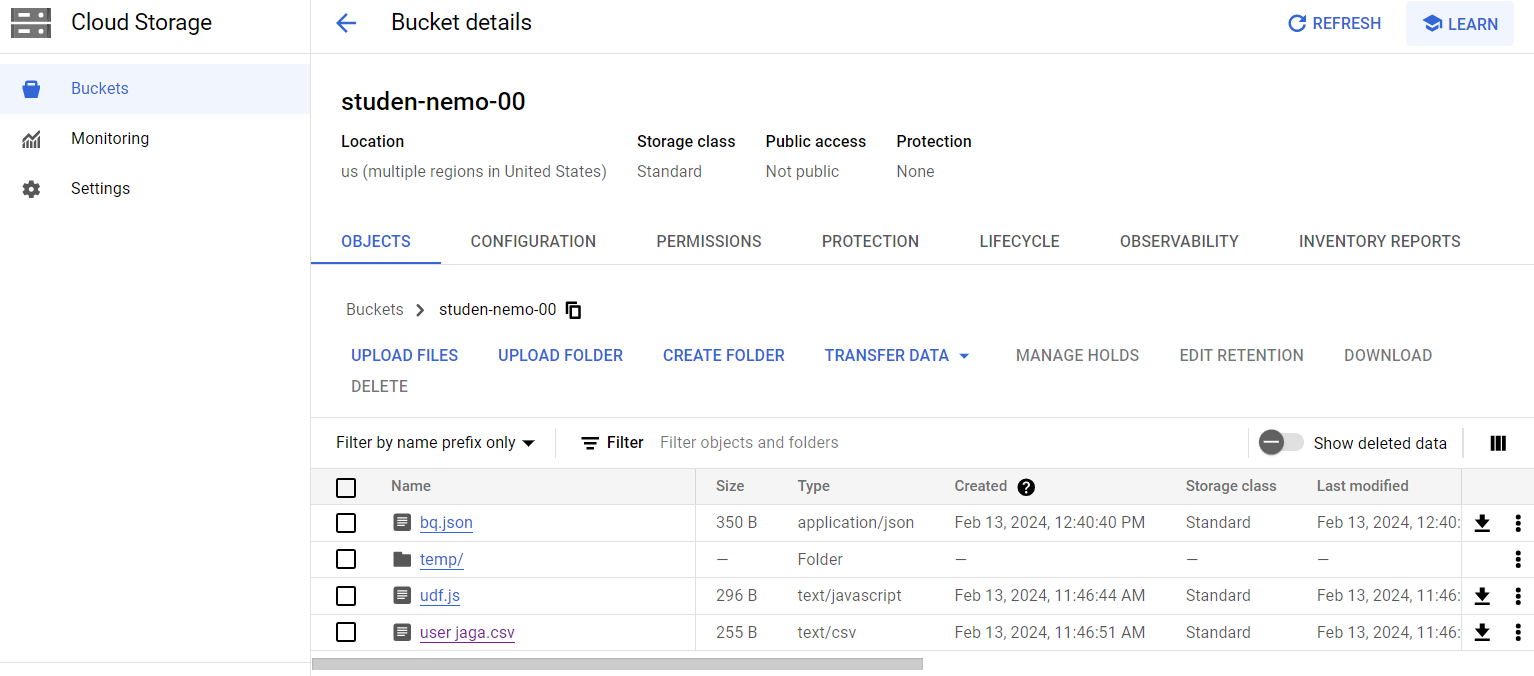
* [Everything that you store in Google Cloud Storage must be contained in a bucket](https://cloud.google.com/storage/docs/buckets).
* [You can use buckets to organize your data and control access to your data](https://cloud.google.com/storage/docs/buckets).
* [Unlike directories and folders, you cannot nest buckets](https://cloud.google.com/storage/docs/buckets)

Remember, bucket names must be globally unique and are publicly visible. [So, avoid using any personally identifiable information (PII) in bucket names](https://cloud.google.com/storage/docs/buckets). [For more details, you can refer to the official GCP documentation](https://cloud.google.com/storage/docs/buckets).

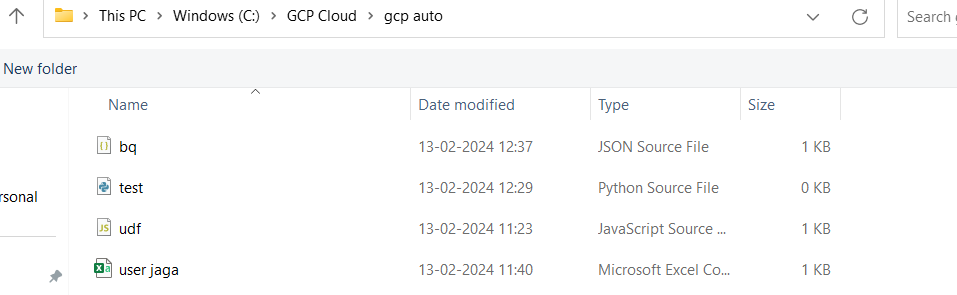


Creating a bucket in Google Cloud Platform (GCP) involves the following steps:

1. [**Log in to Google Cloud Console**](https://www.geeksforgeeks.org/how-to-create-a-google-cloud-storage-bucket/).
2. [**Go to the service section**, search for **Cloud Storage**, and select **Bucket**](https://www.geeksforgeeks.org/how-to-create-a-google-cloud-storage-bucket/).
3. [Click on **Create**](https://www.geeksforgeeks.org/how-to-create-a-google-cloud-storage-bucket/).
4. [**Enter the Name** of your bucket and click on continue](https://www.geeksforgeeks.org/how-to-create-a-google-cloud-storage-bucket/). [Remember, bucket names must be globally unique and are publicly visible](https://cloud.google.com/storage/docs/buckets).
5. [**Select the Location Type** according to your choice](https://www.geeksforgeeks.org/how-to-create-a-google-cloud-storage-bucket/).
6. [**Select Protection Tools** then click on **Create**](https://www.geeksforgeeks.org/how-to-create-a-google-cloud-storage-bucket/).



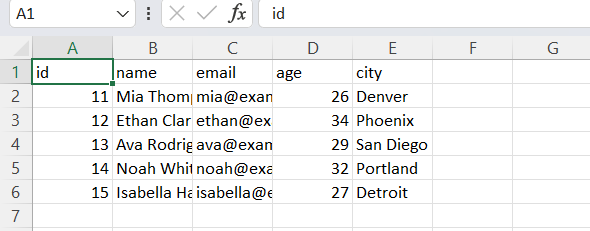
1. [Choose the target GCP bucket where you want to upload your files](https://bing.com/search?q=how+to+upload+data+from+local+storage+to+GCP+bucket).
2. [Locate the files on your local machine and drag and drop them directly into the console](https://bing.com/search?q=how+to+upload+data+from+local+storage+to+GCP+bucket).
3. Save the files on your GCP bucket from local machine.



Json and User-Defined Functio(UDF)

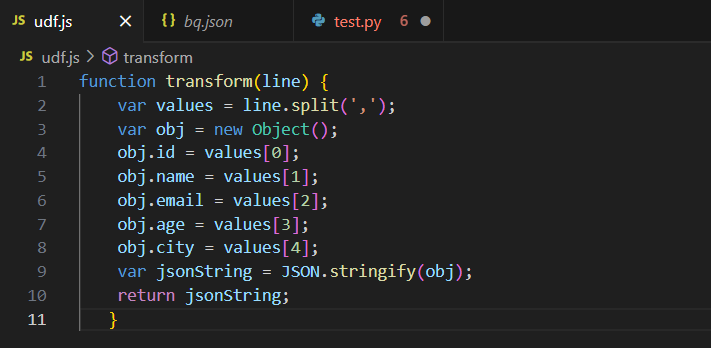
[In Google Cloud Platform (GCP), JSON and User-Defined Functions (UDFs) can be used in conjunction with services like Dataflow and Cloud Storage](https://cloud.google.com/blog/topics/developers-practitioners/extend-your-dataflow-template-with-udfs)

[when you’re working with User-Defined Functions (UDFs), the logic you apply to CSV data can often be similar to the logic you apply to JSON data](https://docs.snowflake.com/en/user-guide/unstructured-data-java).

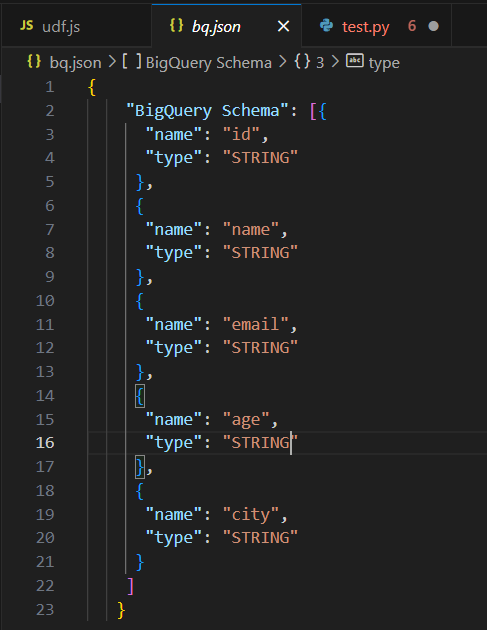


CSV

 If you’re writing a UDF to process data from a CSV file, you might write a function that takes a row of data as input, performs some transformation on the data, and returns the transformed data. [Similarly, if you’re writing a UDF to process data from a JSON file, you might write a function that takes a JSON object as input, performs some transformation on the data, and returns the transformed data](https://docs.snowflake.com/en/user-guide/unstructured-data-java).



[User-Defined Functions (UDFs)](https://cloud.google.com/blog/topics/developers-practitioners/extend-your-dataflow-template-with-udfs)

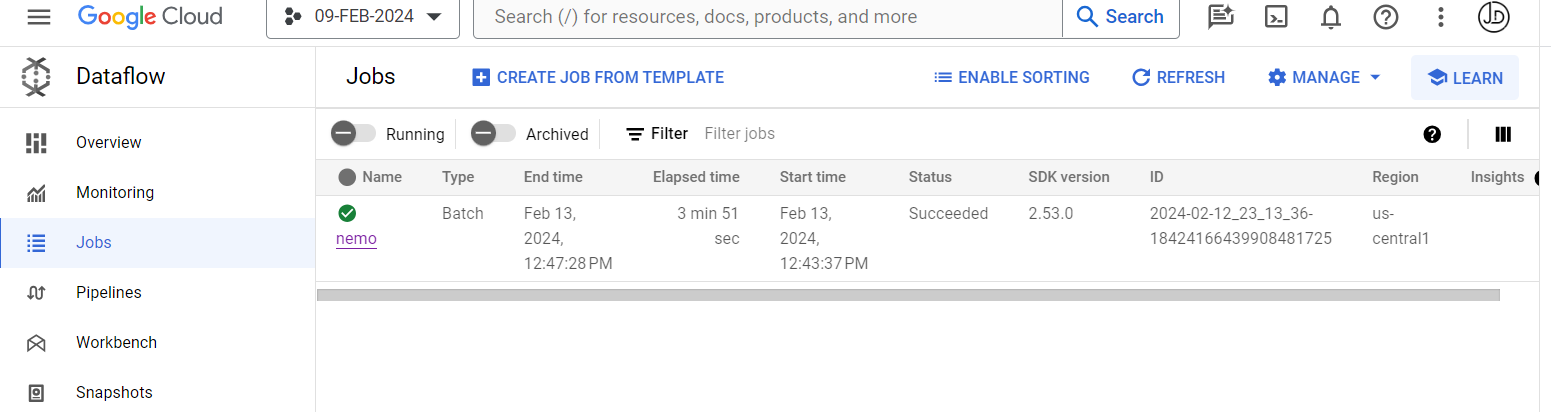


[JSON](https://cloud.google.com/blog/topics/developers-practitioners/extend-your-dataflow-template-with-udfs)

Create job from template

[In Google Cloud Platform (GCP), a job template is a predefined configuration for a job that you can use when creating new job](https://cloud.google.com/run/docs/create-jobs).

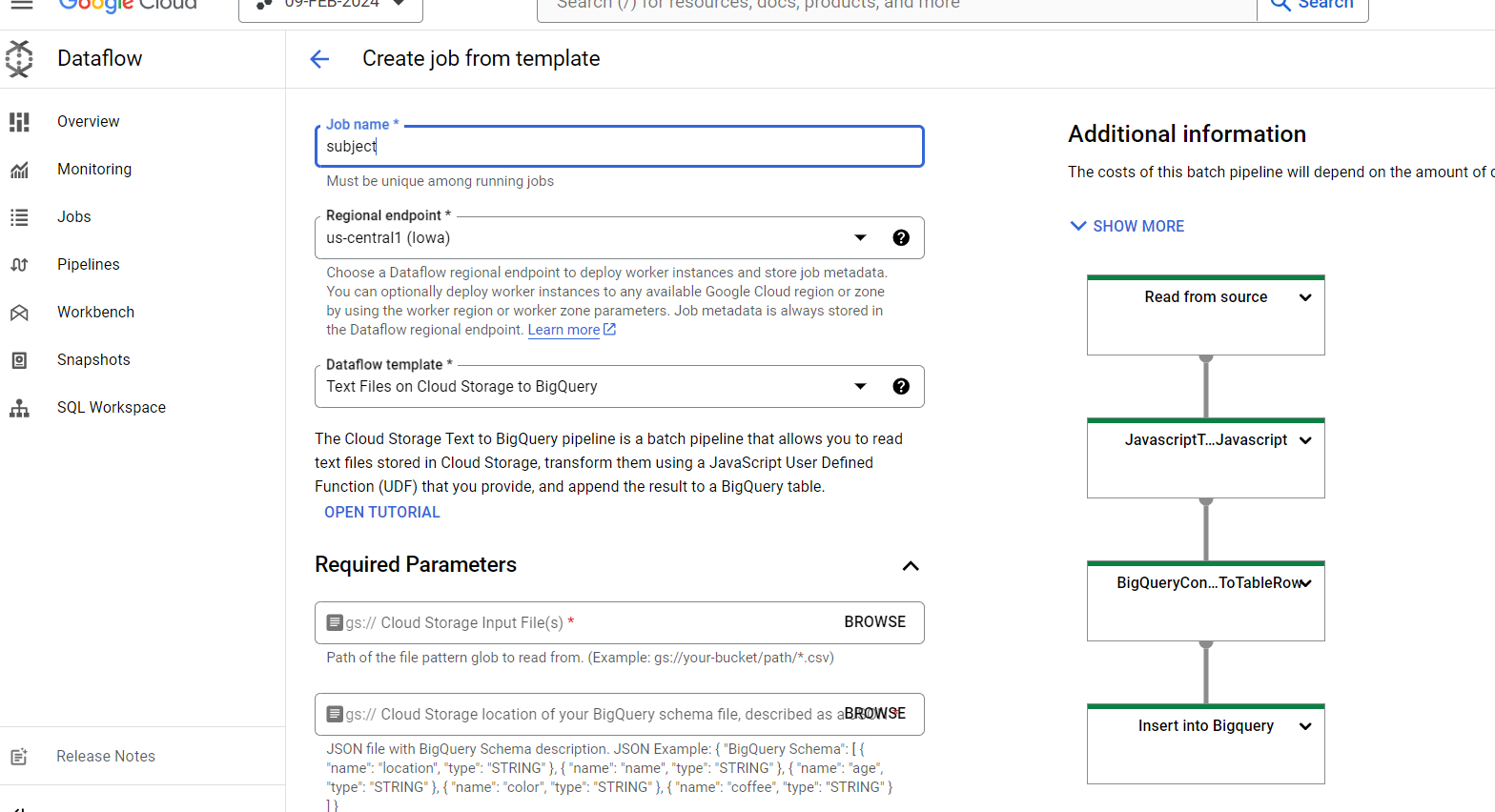
[The exact usage of job templates can vary depending on the specific GCP service](https://cloud.google.com/run/docs/create-jobs).



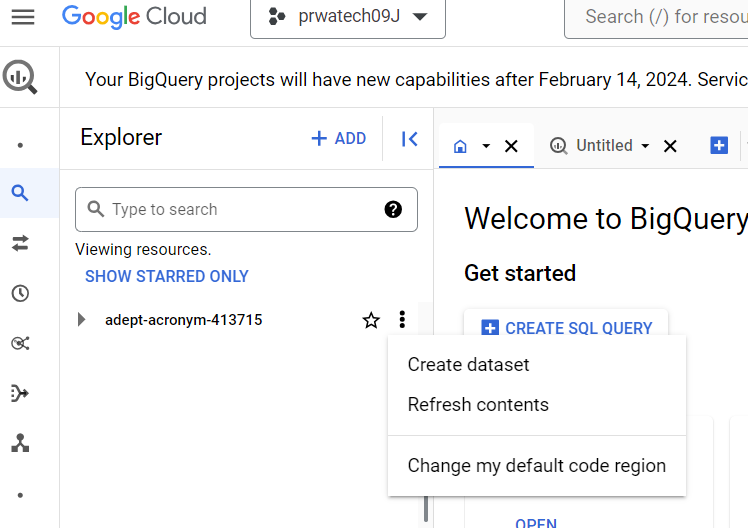
1. Go to the google cloud console
2. Select the Dataflow
3. Click on create job from template
4. Specify the necessary parameters

Necessary Parameters

* Name the job
* Select Datafloe template
* Browse required parameters

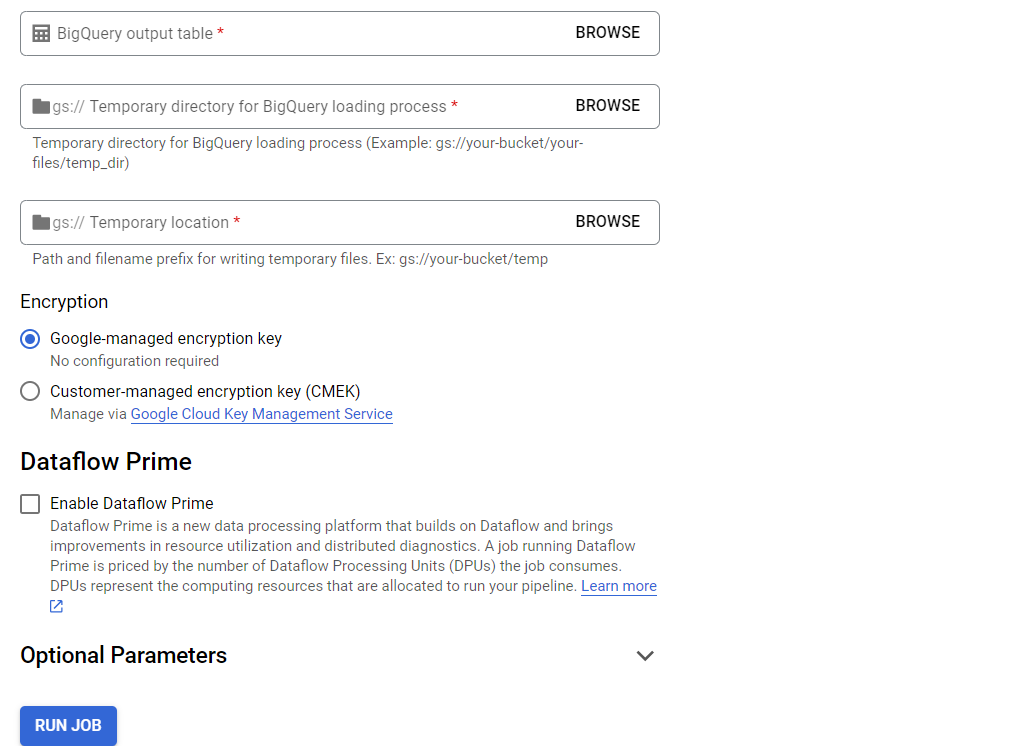


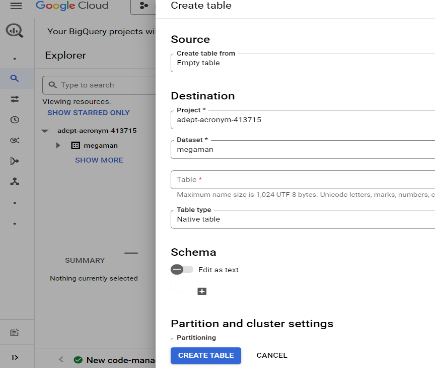
In the required parameters select the CSV file , JSON and brows the created bigquery output table also give temporary direction and location.



Dataset and Table

* Go to the Bigquery
* Create dataset
* Create table in your dataset



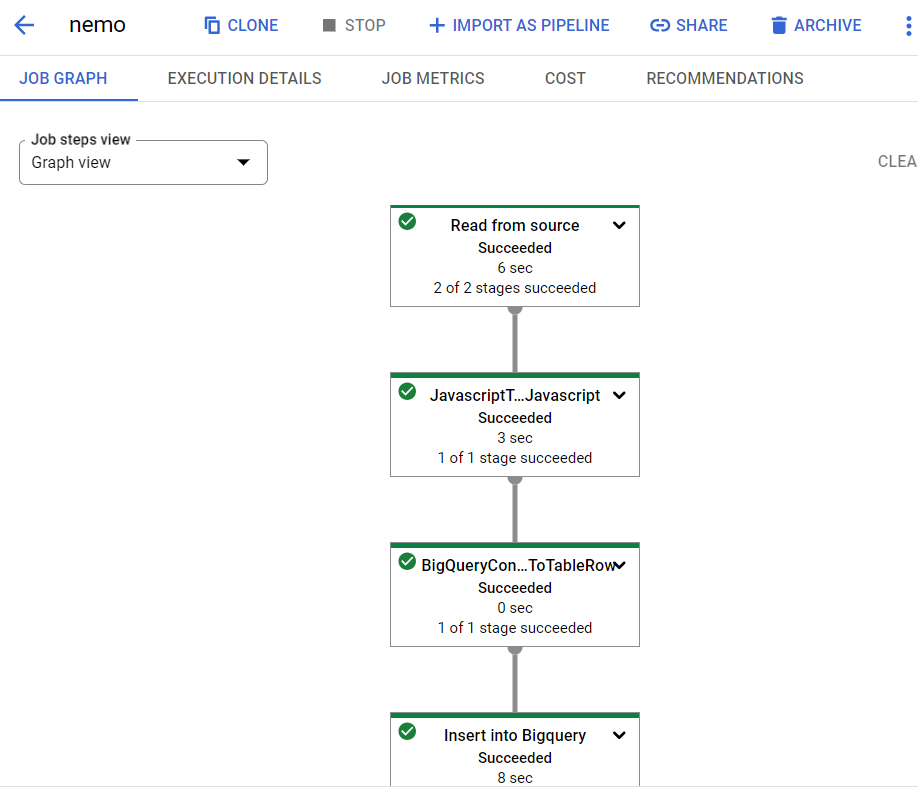


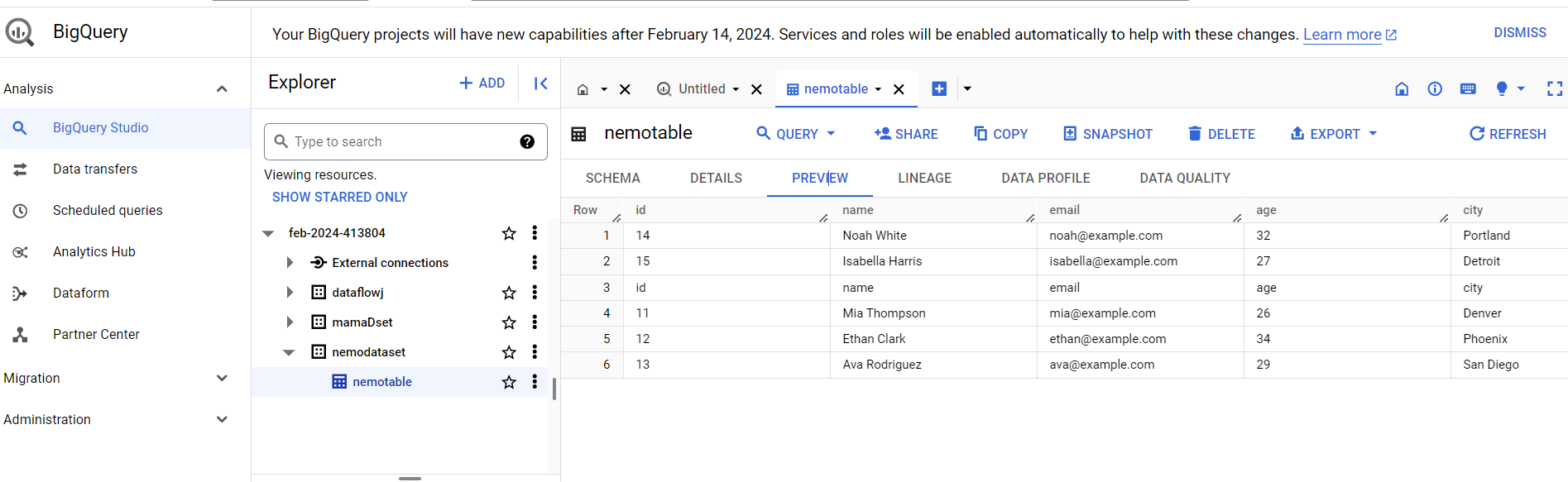
After browse every data from bucket run job .

Insert data into bigquery

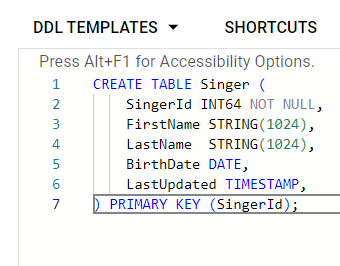
[After successfully creating a job template in Google Cloud Platform (GCP), you can use it to create and execute jobs](https://cloud.google.com/dataproc/docs/concepts/workflows/overview).

* [**Instantiating a Workflow Template**: Instantiating a Workflow Template launches a Workflow](https://cloud.google.com/dataproc/docs/concepts/workflows/overview). [A Workflow is an operation that runs a Directed Acyclic Graph (DAG) of jobs on a cluster](https://cloud.google.com/dataproc/docs/concepts/workflows/overview).





1. [When you load CSV data from Cloud Storage, you can load the data into a new table or partition, or you can append to or overwrite an existing table or partition](https://cloud.google.com/bigquery/docs/loading-data-cloud-storage-csv).
2. [The dataset that contains the table must be in the same regional or multi-regional location as the Cloud Storage bucket](https://cloud.google.com/bigquery/docs/loading-data-cloud-storage-csv).
3. [**Use Dataflow to read CSV file from Storage and write into BigQuery**: You can use Google Cloud Dataflow to read a CSV file from Google Cloud Storage and write it into a BigQuery table](https://cloud.google.com/bigquery/docs/loading-data-cloud-storage-csv).



INSERT INTO Singer (SingerId, FirstName, LastName)

VALUES(1, 'Marc', 'Richards'),

(2, 'Catalina', 'Smith'),

(3, 'Alice', 'Trentor');