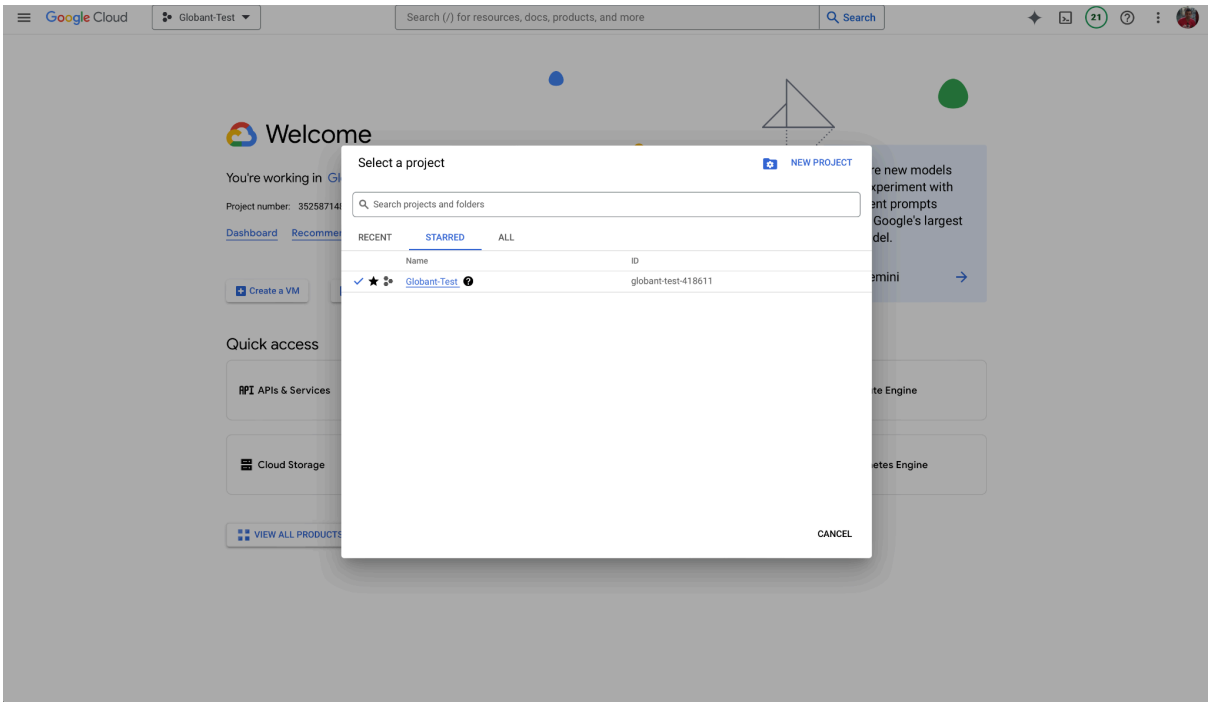
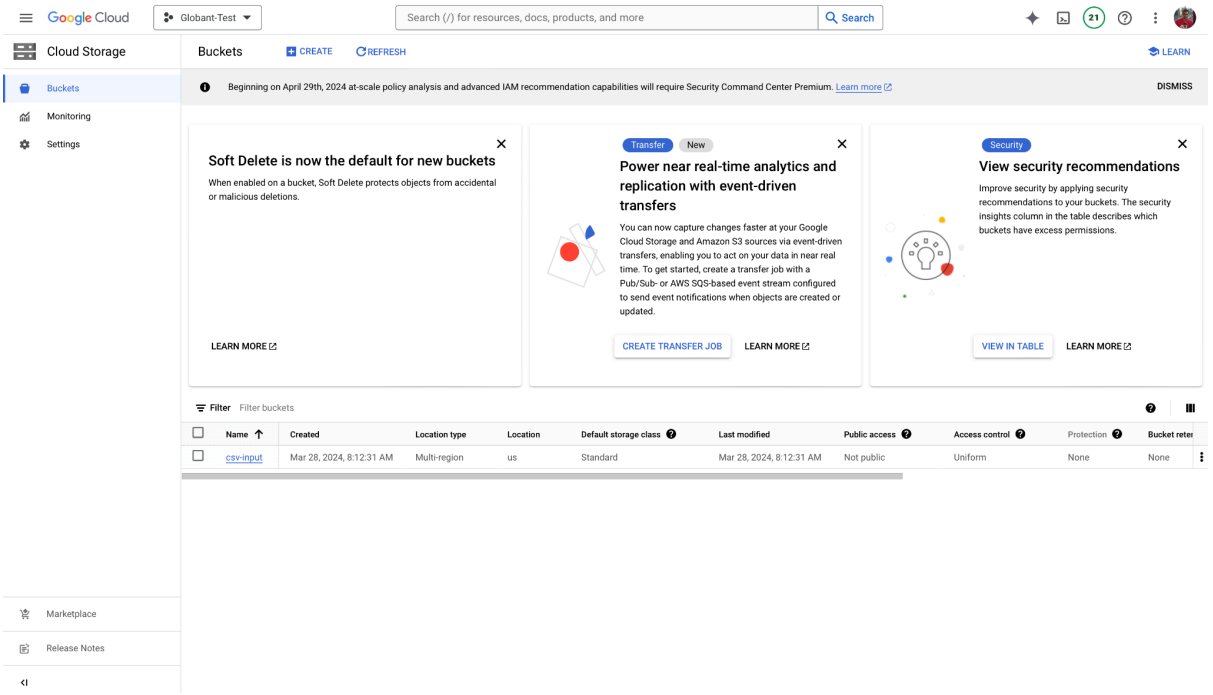
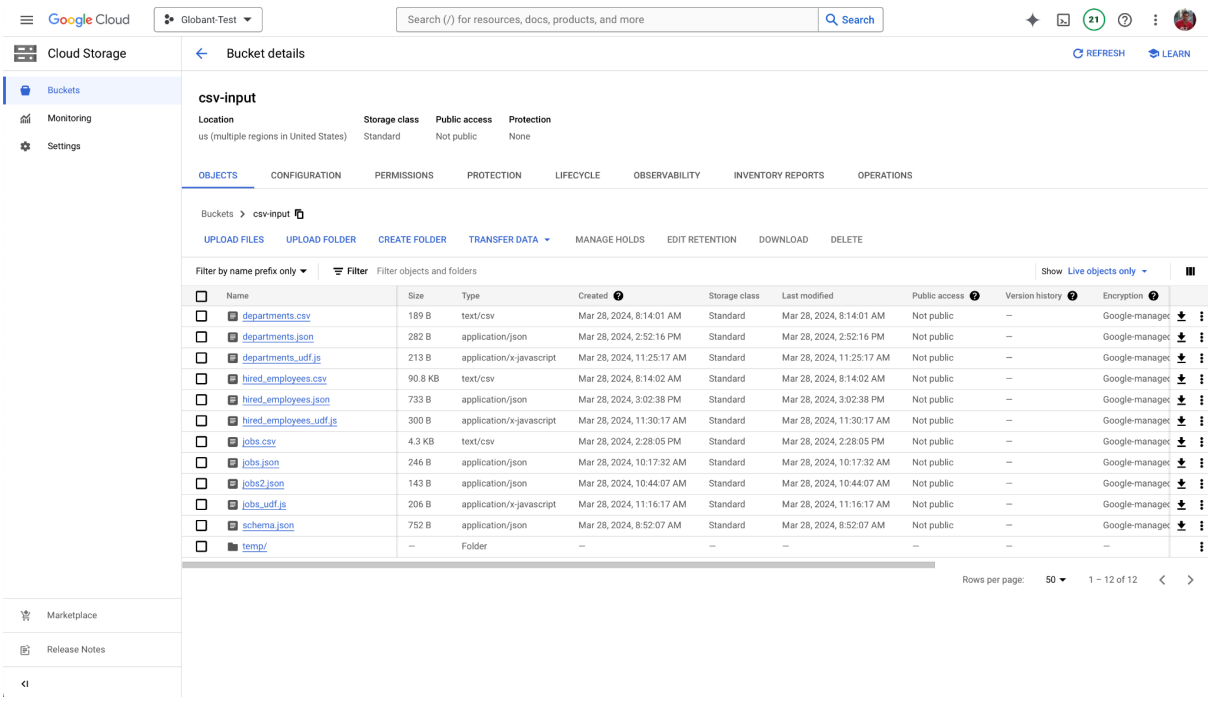
Github: https://github.com/martinmvelez/Globant

With the purpose of developing this exercise, i created Globant-Test project in GCP.

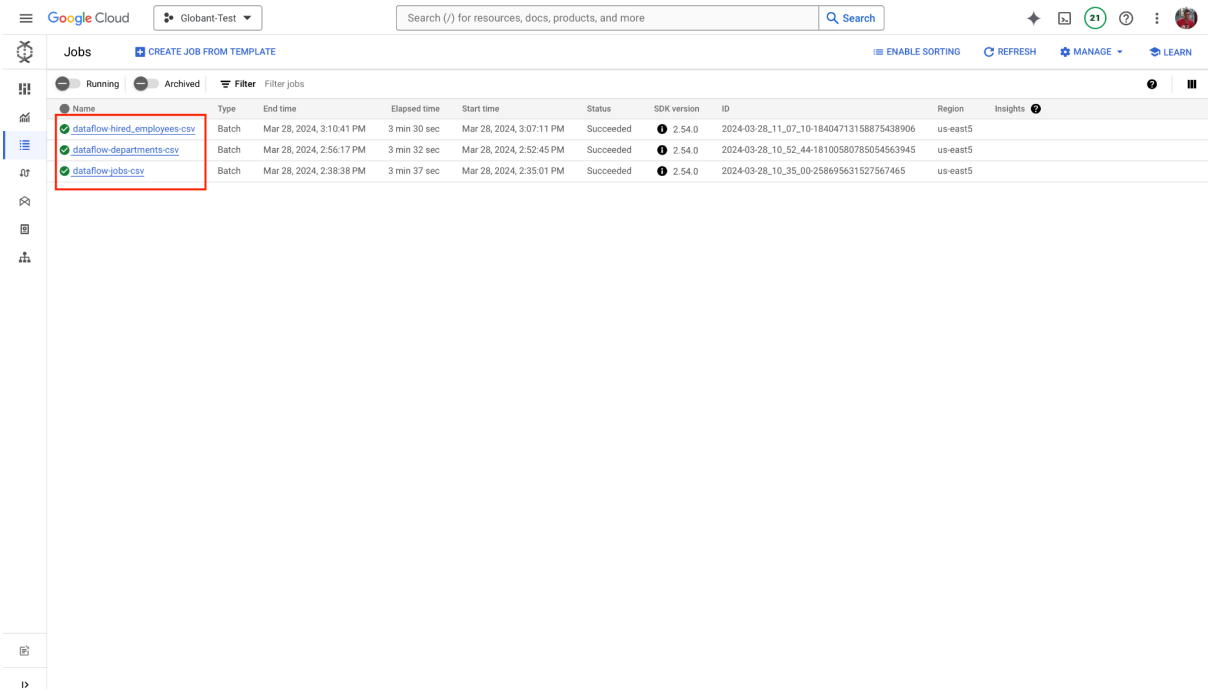


Buckets
To upload the files departments.csv, jobs.csv, hired_employees.csv i created the bucket csv-input

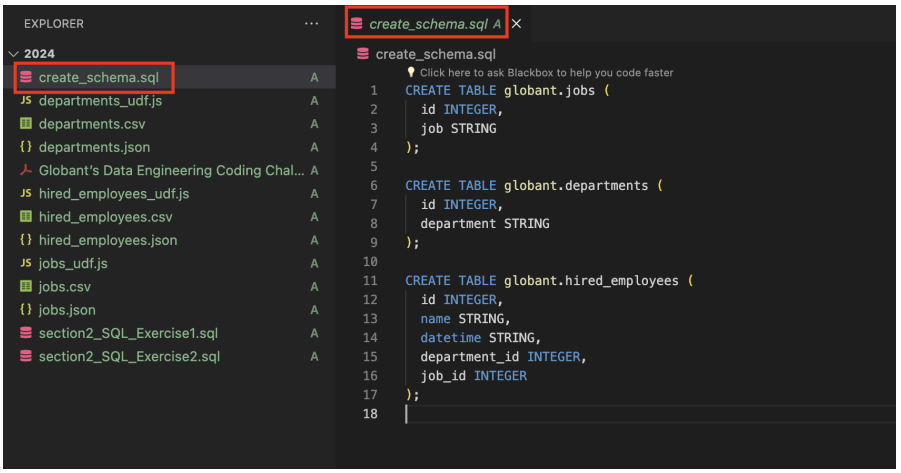


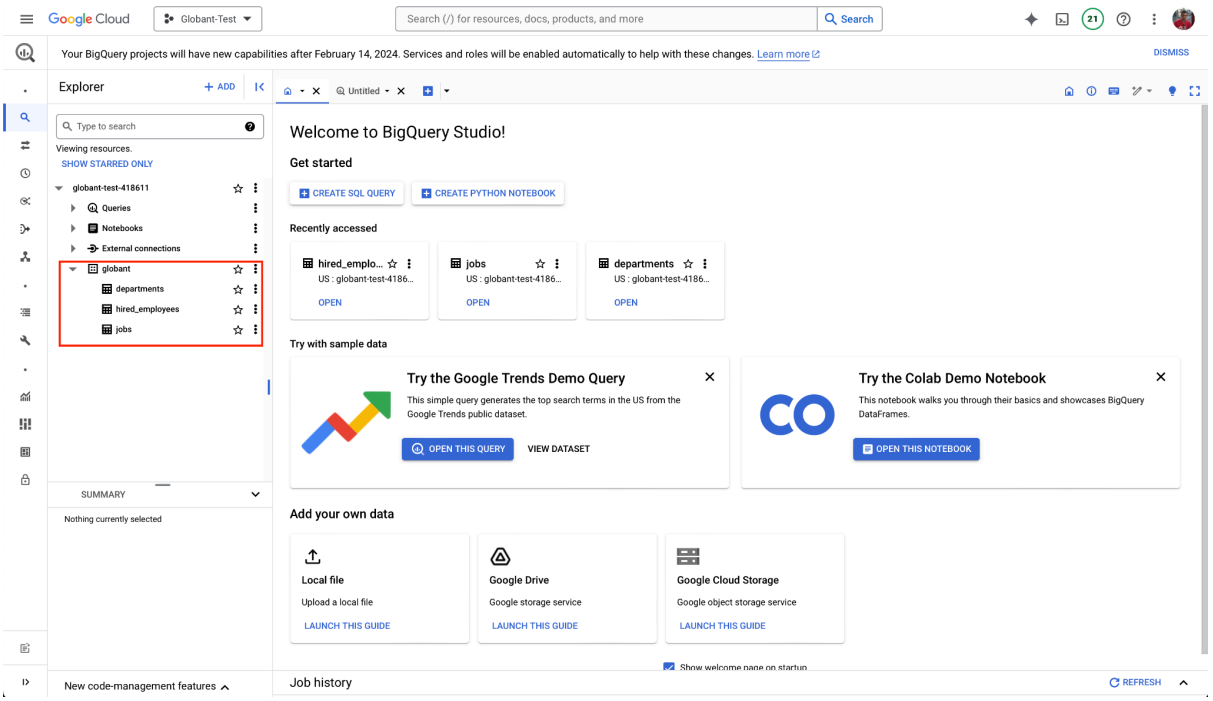

Google Cloud
Globant-Test
Search (/) for resources, docs, products, and more
Search
Cloud Storage
Buckets
Monitoring
Settings
Bucket details
REFRESH
LEARN
csv-input
Location
us (multiple regions in United States)
Storage class
Standard
Public access
Not public
Protection
None
OBJECTS
CONFIGURATION
PERMISSIONS
PROTECTION
LIFECYCLE
OBSERVABILITY
INVENTORY REPORTS
OPERATIONS
Buckets > csv-input
UPLOAD FILES
UPLOAD FOLDER
CREATE FOLDER
TRANSFER DATA
MANAGE HOLDS
EDIT RETENTION
DOWNLOAD
DELETE
Filter by name prefix only
Filter
Filter objects and folders
Show Live objects only
Name
Size
Type
Created
Storage class
Last modified
Public access
Version history
Encryption
departments.csv 189 B text/csv Mar 28, 2024, 8:14:01 AM Standard Mar 28, 2024, 8:14:01 AM Not public
departments.json 282 B application/json Mar 28, 2024, 2:52:16 PM Standard Mar 28, 2024, 2:52:16 PM Not public
departments_udf.js 213 B application/x-javascript Mar 28, 2024, 11:25:17 AM Standard Mar 28, 2024, 11:25:17 AM Not public
hired_employees.csv 90.8 KB text/csv Mar 28, 2024, 8:14:02 AM Standard Mar 28, 2024, 8:14:02 AM Not public
hired_employees.json 733 B application/json Mar 28, 2024, 3:02:38 PM Standard Mar 28, 2024, 3:02:38 PM Not public
hired_employees_udf.js 300 B application/x-javascript Mar 28, 2024, 11:30:17 AM Standard Mar 28, 2024, 11:30:17 AM Not public
jobs.csv 4.3 KB text/csv Mar 28, 2024, 2:28:05 PM Standard Mar 28, 2024, 2:28:05 PM Not public
jobs.json 246 B application/json Mar 28, 2024, 10:17:32 AM Standard Mar 28, 2024, 10:17:32 AM Not public
jobs2.json 143 B application/json Mar 28, 2024, 10:44:07 AM Standard Mar 28, 2024, 10:44:07 AM Not public
jobs_udf.js 206 B application/x-javascript Mar 28, 2024, 11:16:17 AM Standard Mar 28, 2024, 11:16:17 AM Not public
schema.json 752 B application/json Mar 28, 2024, 8:52:07 AM Standard Mar 28, 2024, 8:52:07 AM Not public
temp/ – Folder
Rows per page: 50
1 – 12 of 12
Marketplace
Release Notes

Dataflow


Google Cloud
Globant-Test
Search (/) for resources, docs, products, and more
Search
Jobs
CREATE JOB FROM TEMPLATE
ENABLE SORTING
REFRESH
MANAGE
LEARN
Running
Archived
Filter
Filter jobs
Name
Type
End time
Elapsed time
Start time
Status
SDK version
ID
Region
Insights
dataflow-hired_employees-csv Batch Mar 28, 2024, 3:10:41 PM 3 min 30 sec Mar 28, 2024, 3:07:11 PM Succeeded 2.54.0 2024-03-28_11_07_10-18404713158875438906 us-east5
dataflow-departments-csv Batch Mar 28, 2024, 2:56:17 PM 3 min 32 sec Mar 28, 2024, 2:52:45 PM Succeeded 2.54.0 2024-03-28_10_52_44-18100580785054563945 us-east5
dataflow-jobs-csv Batch Mar 28, 2024, 2:38:38 PM 3 min 37 sec Mar 28, 2024, 2:35:01 PM Succeeded 2.54.0 2024-03-28_10_35_00-2586956315275674 65 us-east5

Bigquery
I created the dataset globant and the tables
Note: code of table creation create_schema.sql



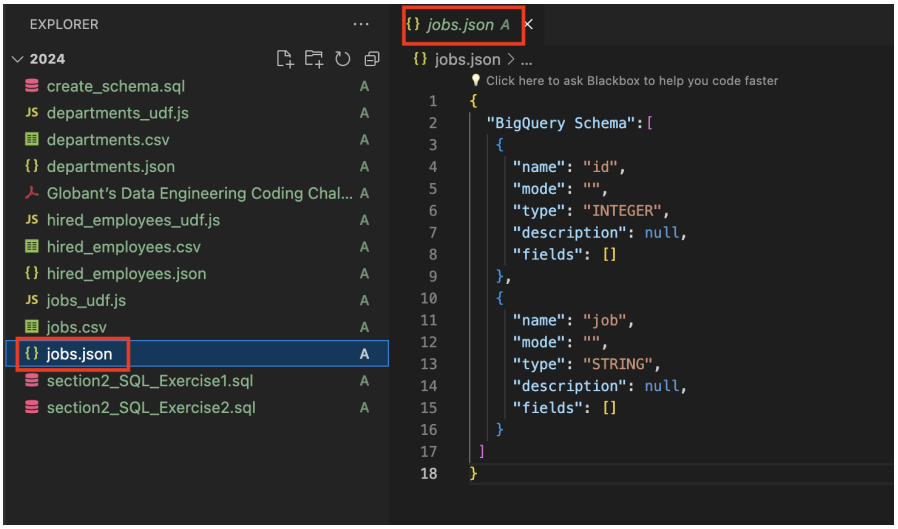


APIs
In the following i will describe the ingestion process for each one.

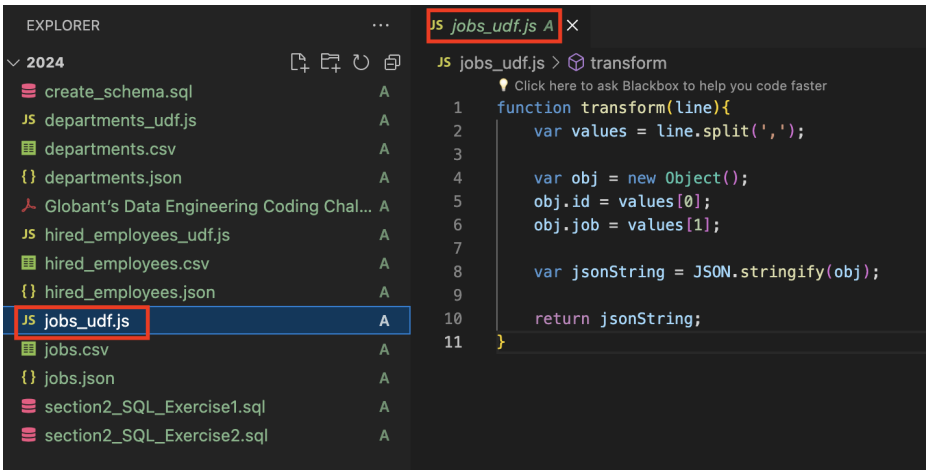
Jobs.csv

I created jobs.json, jobs_udf.js

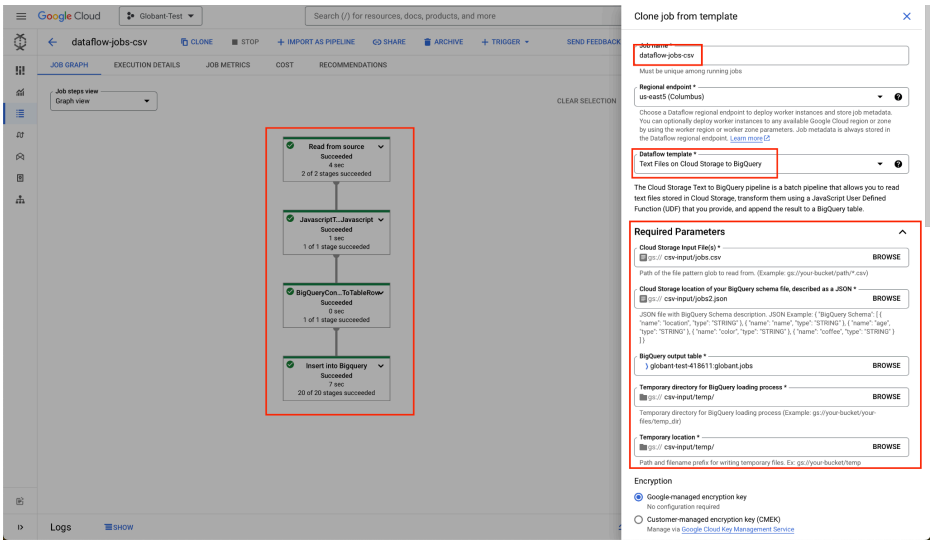
jobs.json: is the schema of the table



jobs_udf.js: Script for the csv ingestion



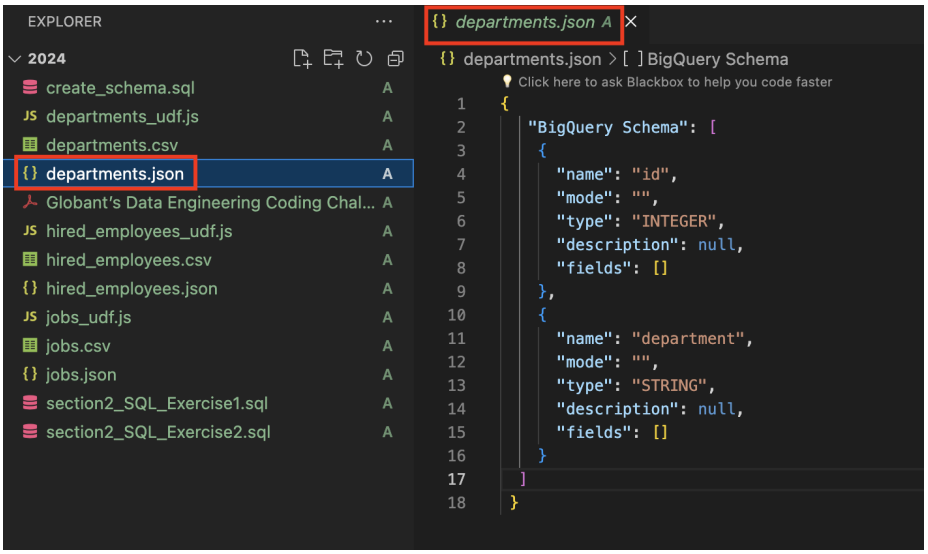
Dataflow Job



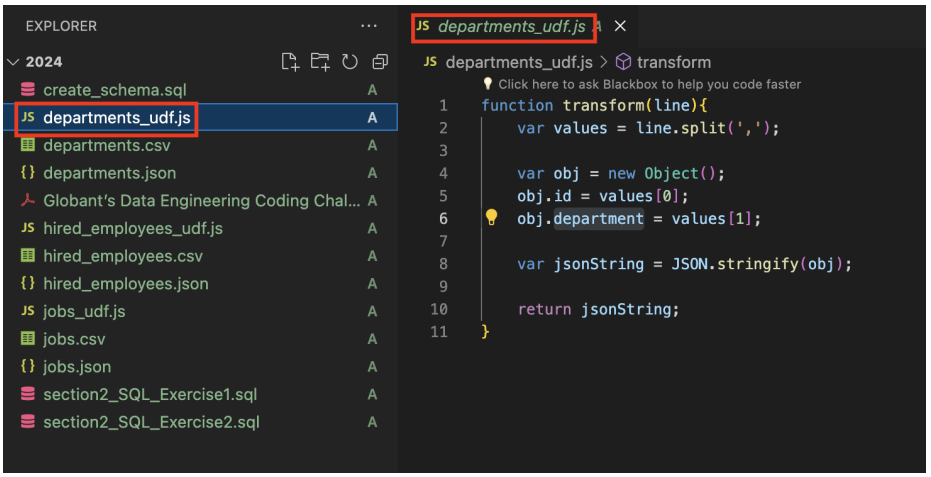
Departments.csv

I created departments.json, departments_udf.js

departments.json: is the schema of the table



departments.js: Script for the csv ingestion

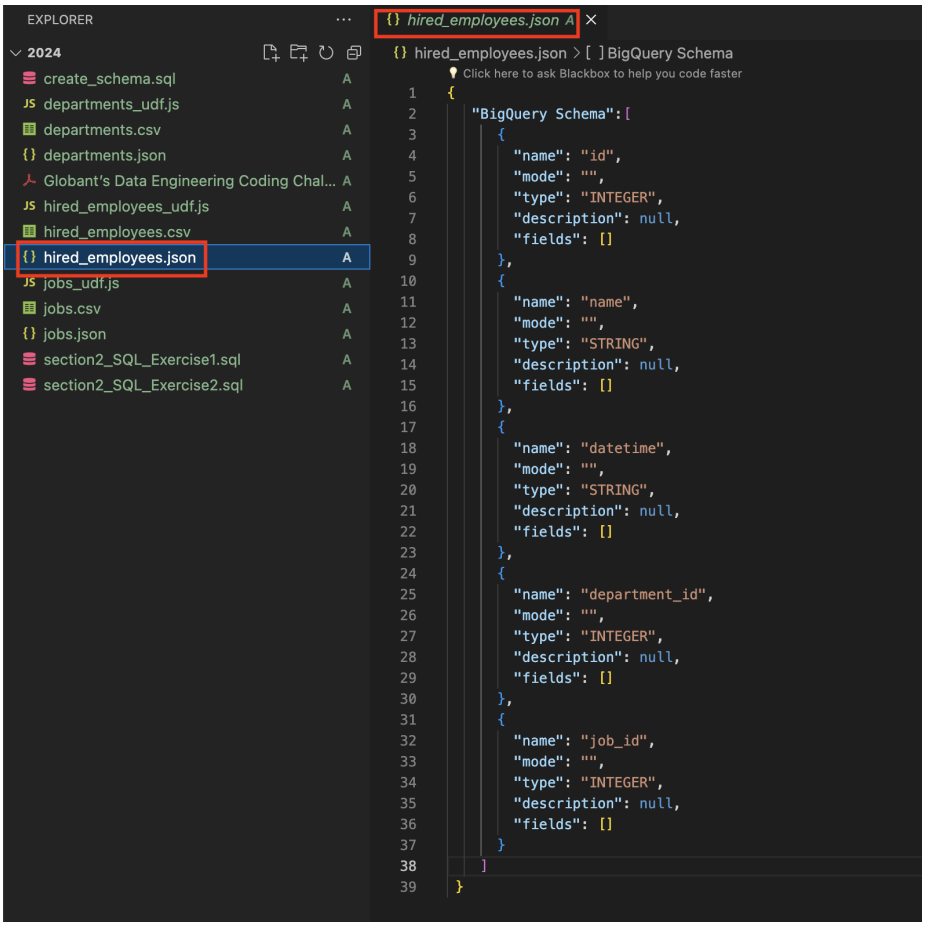


Dataflow Job

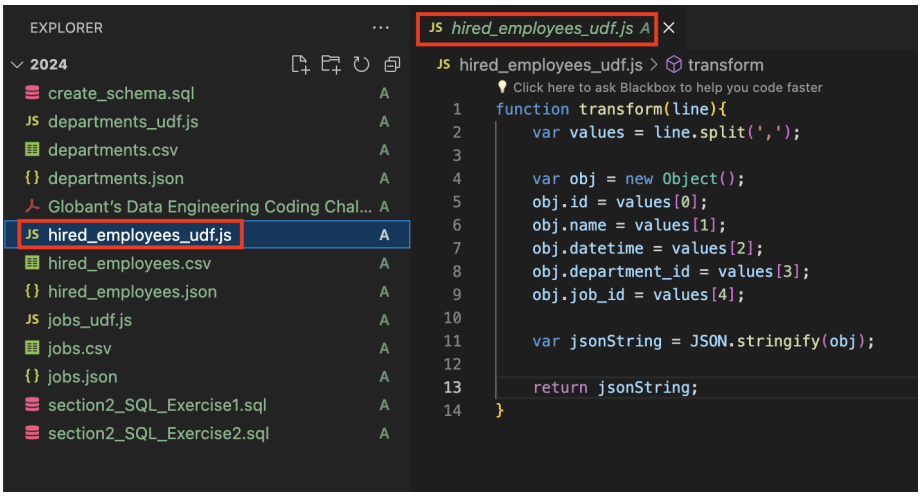
hired_employees.csv

I created hired_employees.json, hired_employees_udf.js

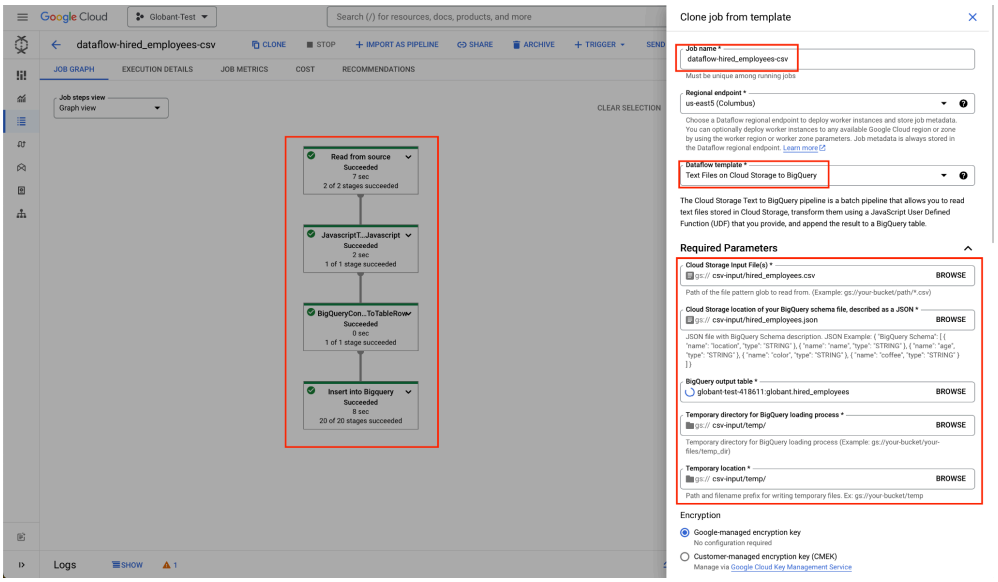
hired_employees.json: is the schema of the table



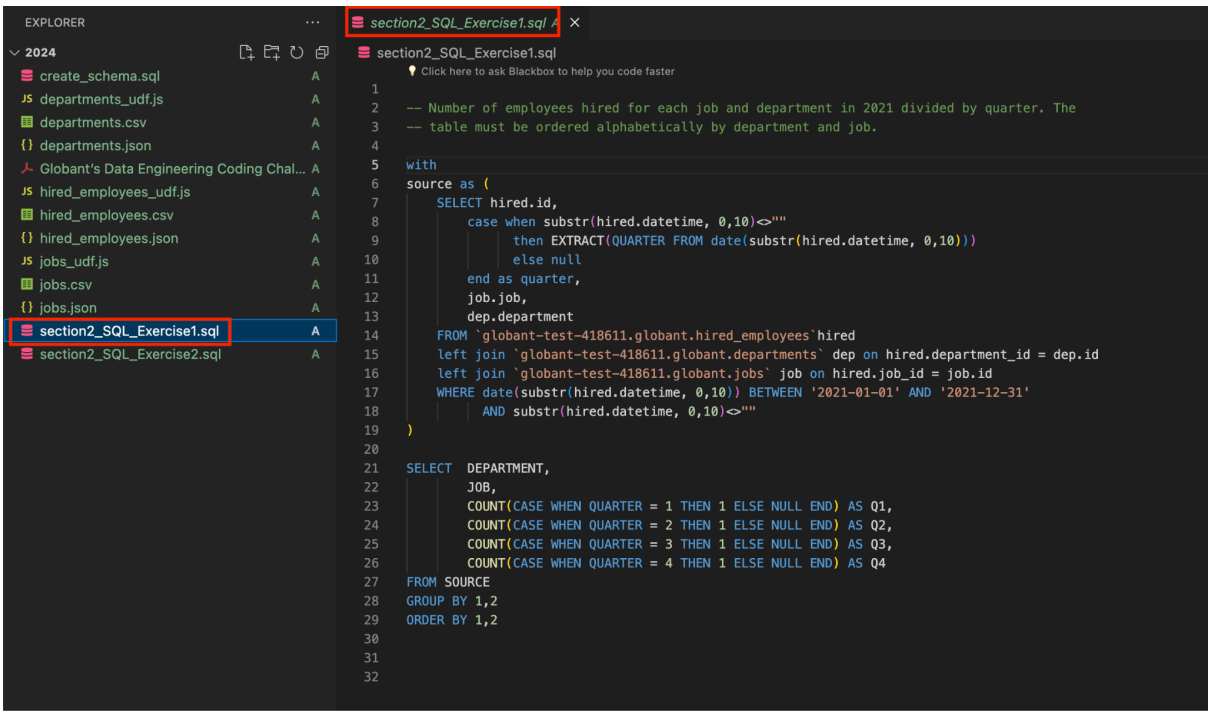
hired_employees.js: Script for the csv ingestion



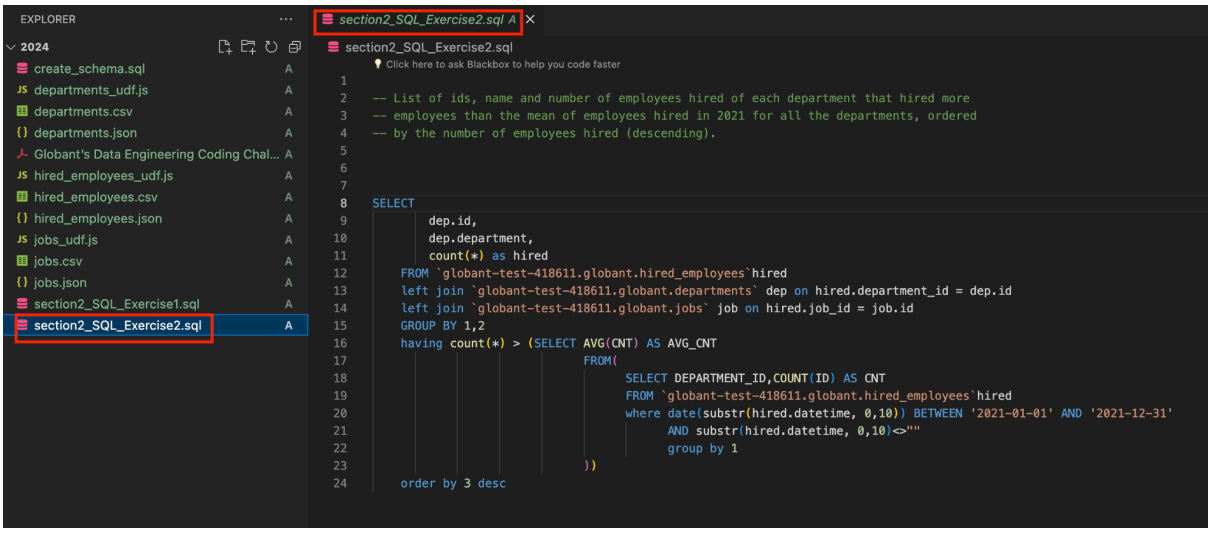
Dataflow Job



Exercise 1



Exercise 2


EXPLORER
2024
create_schema.sql
departments_udf.js
departments.csv
departments.json
Globant's Data Engineering Coding Chal...
hired_employees_udf.js
hired_employees.csv
hired_employees.json
jobs_udf.js
jobs.csv
jobs.json
section2_SQL_Exercise1.sql
section2_SQL_Exercise2.sql
section2_SQL_Exercise2.sql A
section2_SQL_Exercise2.sql
Click here to ask Blackbox to help you code faster
-- List of ids, name and number of employees hired of each department that hired more
-- employees than the mean of employees hired in 2021 for all the departments, ordered
-- by the number of employees hired (descending).
SELECT
dep.id,
dep.department,
count(*) as hired
FROM `globant-test-418611.globant.hired_employees`hired
left join `globant-test-418611.globant.departments` dep on hired.department_id = dep.id
left join `globant-test-418611.globant.jobs` job on hired.job_id = job.id
GROUP BY 1,2
having count(*) > (SELECT AVG(CNT) AS AVG_CNT
FROM(
SELECT DEPARTMENT_ID,COUNT(ID) AS CNT
FROM `globant-test-418611.globant.hired_employees`hired
where date(substr(hired.datetime, 0,10)) BETWEEN '2021-01-01' AND '2021-12-31'
AND substr(hired.datetime, 0,10)<>""
group by 1
))
order by 3 desc