COVID-19 DATA ANALYSIS

OBJECTIVE

- To perform data analysis and visualization on Covid-19 data using ETL tools.

- To perform data analysis on Covid_19 data and extract the meaningful information from the dataset which will help in taking quick and informed decision.

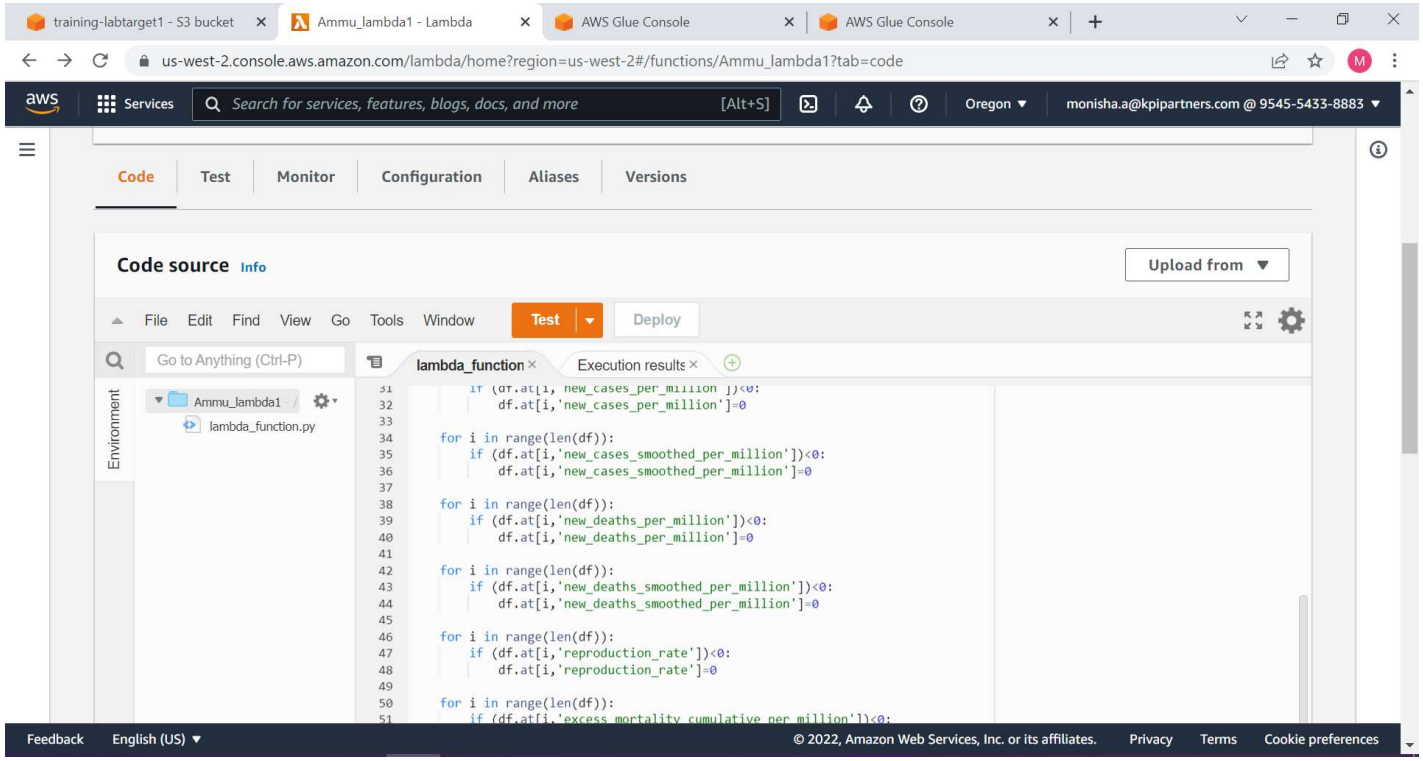
TECHNOLOGIES USED

- AWS Glue
- SparkSQL
- Redshift
- AWS S3
- SparkSQL
- Crawler
- AWS Lambda
- PySpark

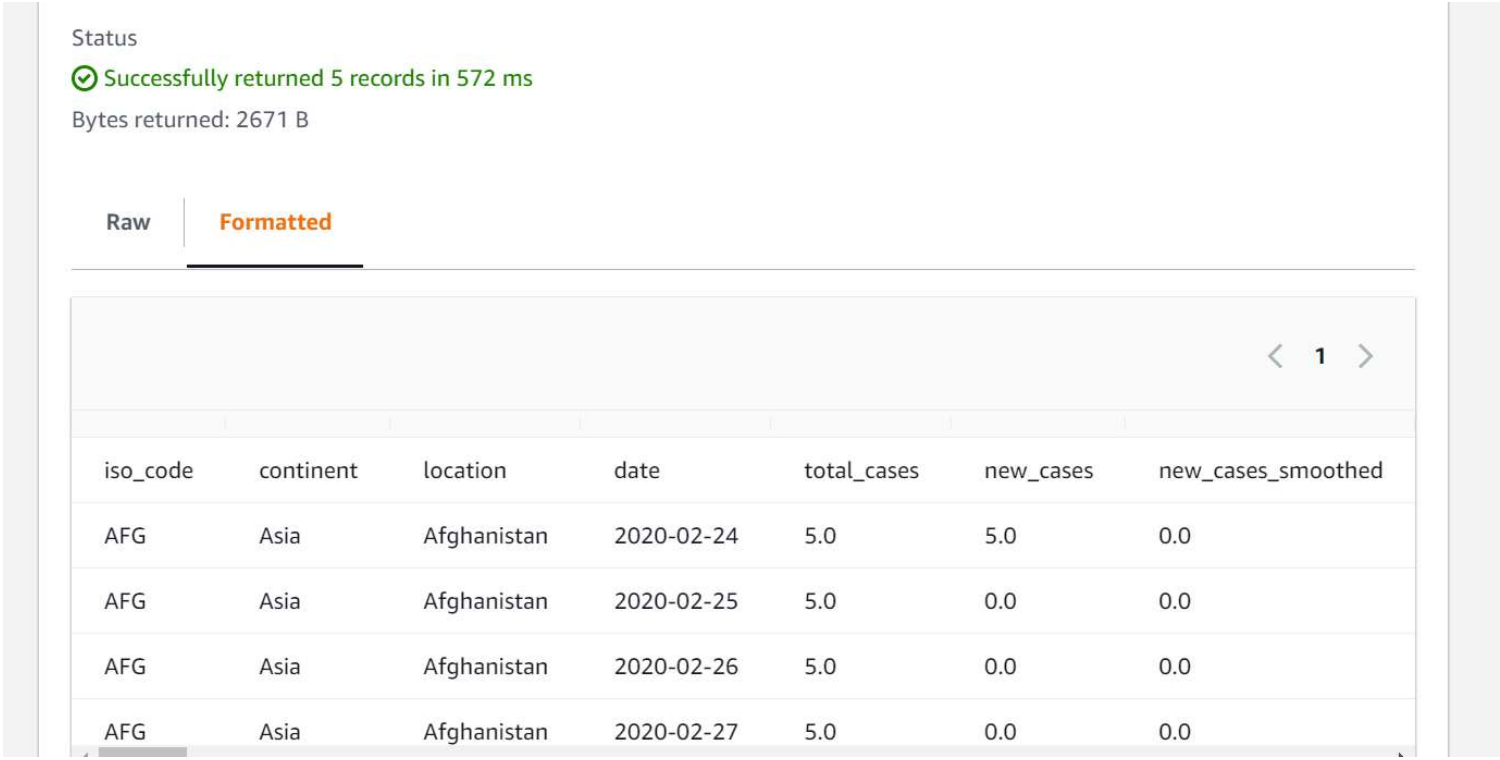
Problem Statements

1.Clean and transform data for processing

job name: Ammu_lambda1



1.Clean and transform data for processing Cont…



Status

Successfully returned 5 records in 572 ms

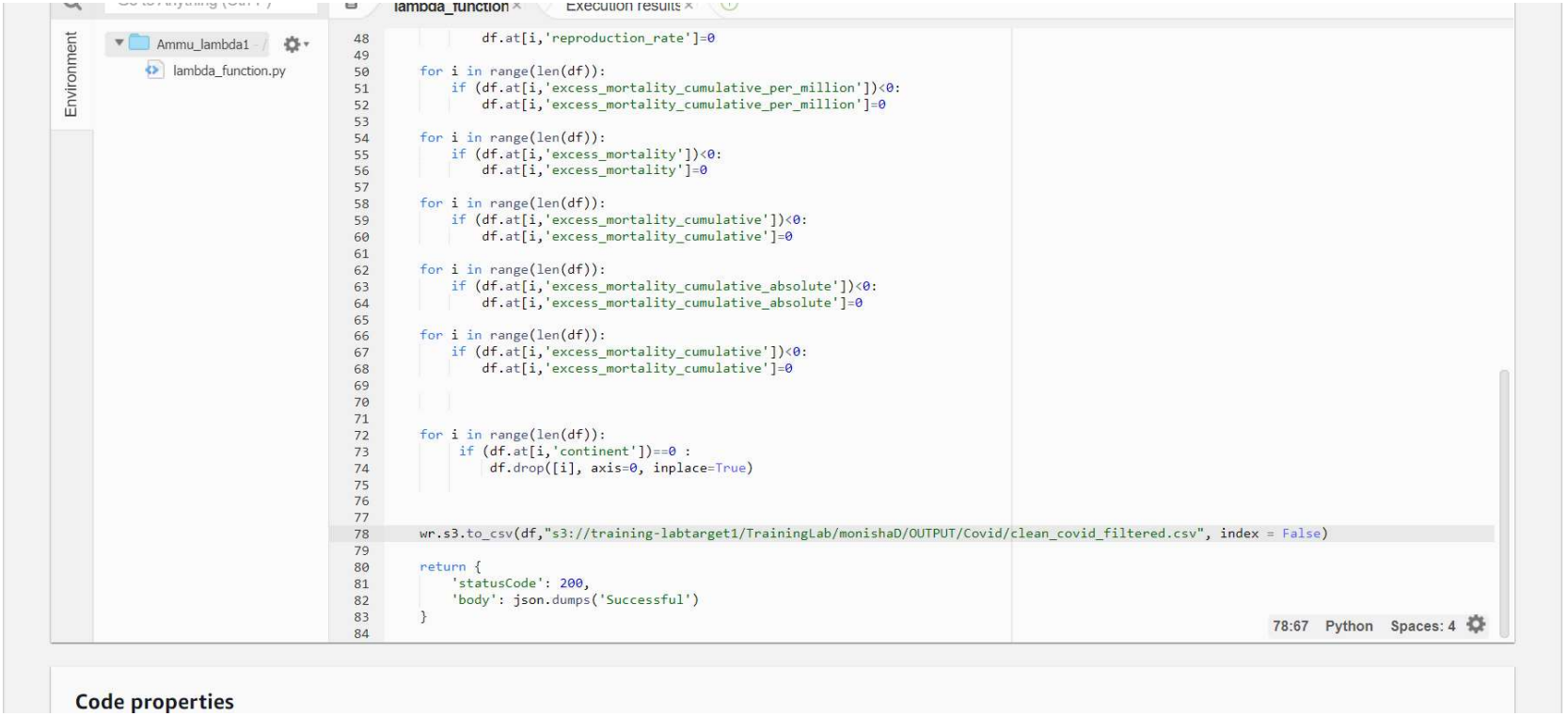
Bytes returned: 2671 B

Raw | Formatted

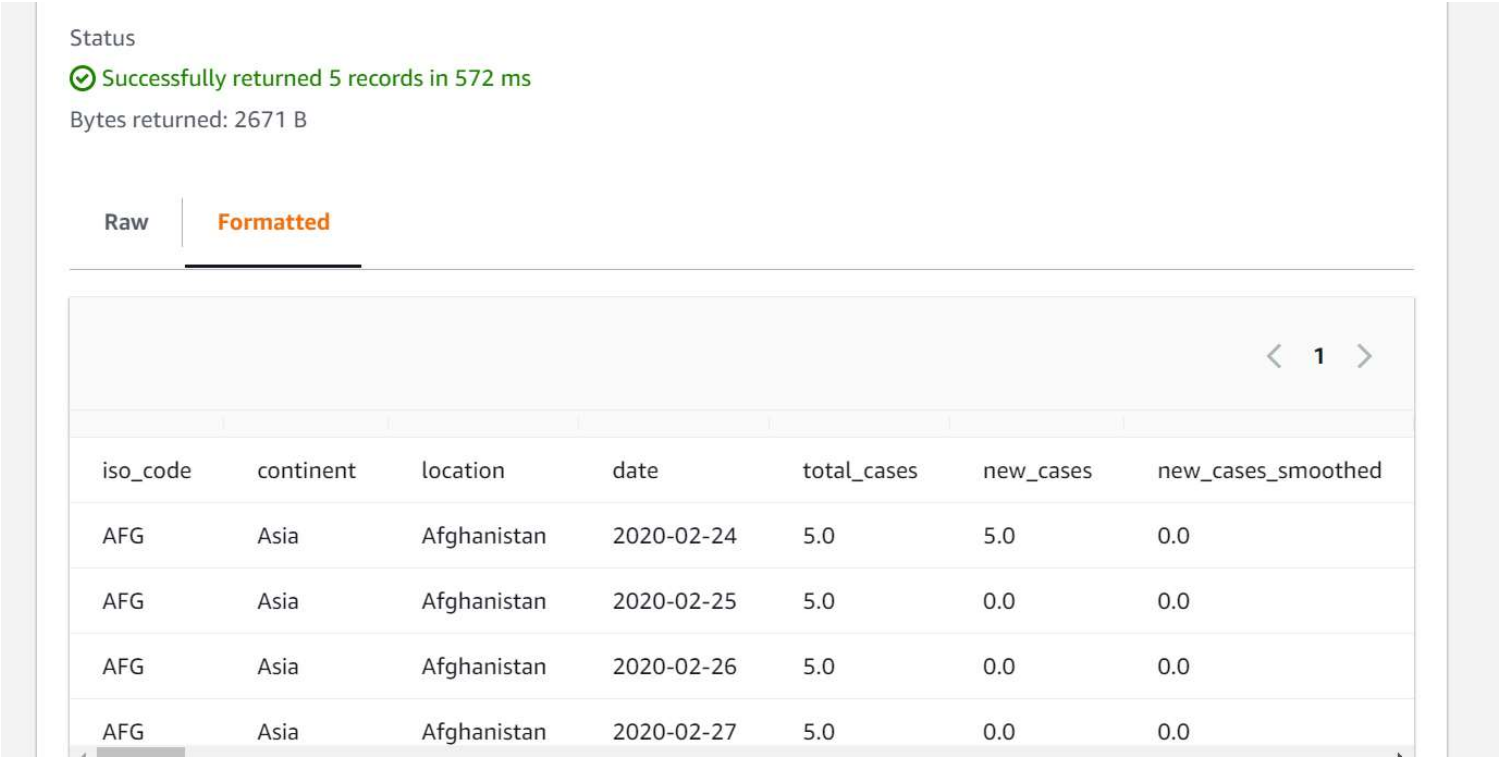
1

iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed
AFG	Asia	Afghanistan	2020-02-24	5.0	5.0	0.0
AFG	Asia	Afghanistan	2020-02-25	5.0	0.0	0.0
AFG	Asia	Afghanistan	2020-02-26	5.0	0.0	0.0
AFG	Asia	Afghanistan	2020-02-27	5.0	0.0	0.0

2. ETL operations on dataset



2. ETL operations on dataset Cont...



Status

Successfully returned 5 records in 572 ms

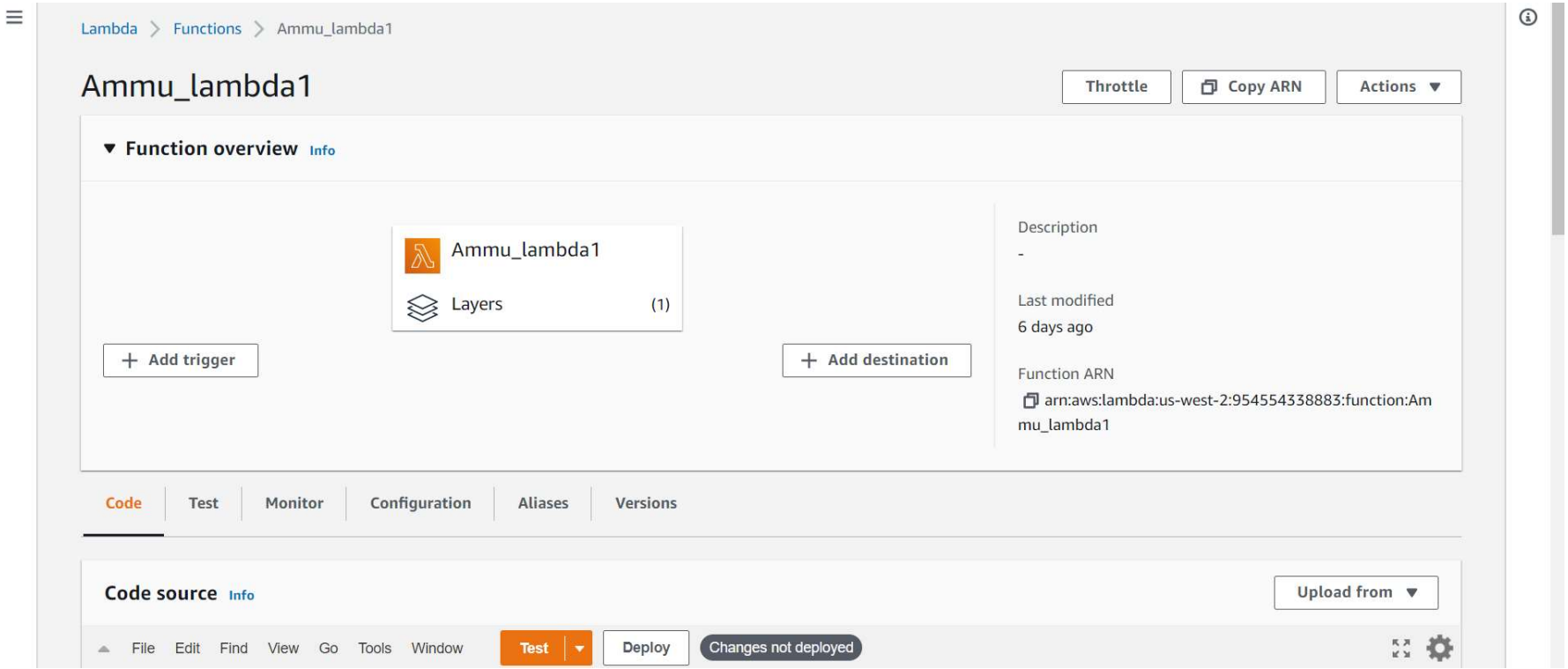
Bytes returned: 2671 B

Raw | Formatted

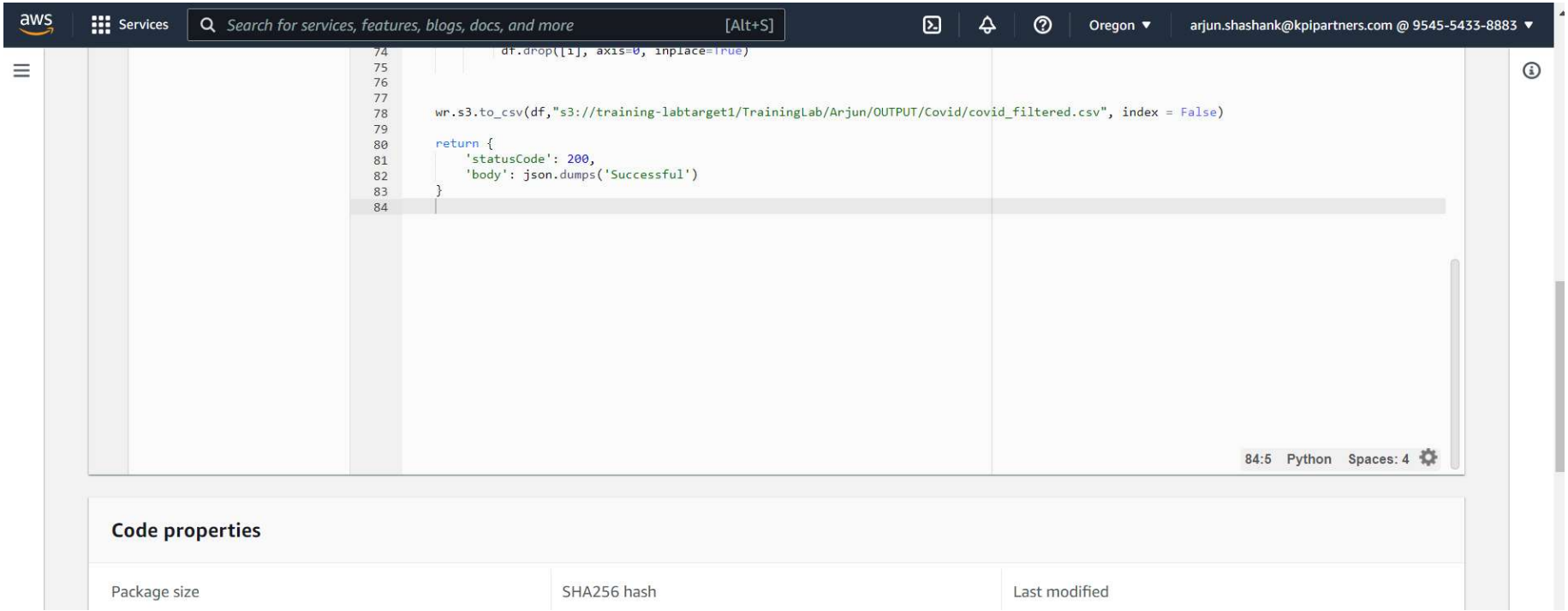
1

iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed
AFG	Asia	Afghanistan	2020-02-24	5.0	5.0	0.0
AFG	Asia	Afghanistan	2020-02-25	5.0	0.0	0.0
AFG	Asia	Afghanistan	2020-02-26	5.0	0.0	0.0
AFG	Asia	Afghanistan	2020-02-27	5.0	0.0	0.0

3. Storing the modified data in AWS

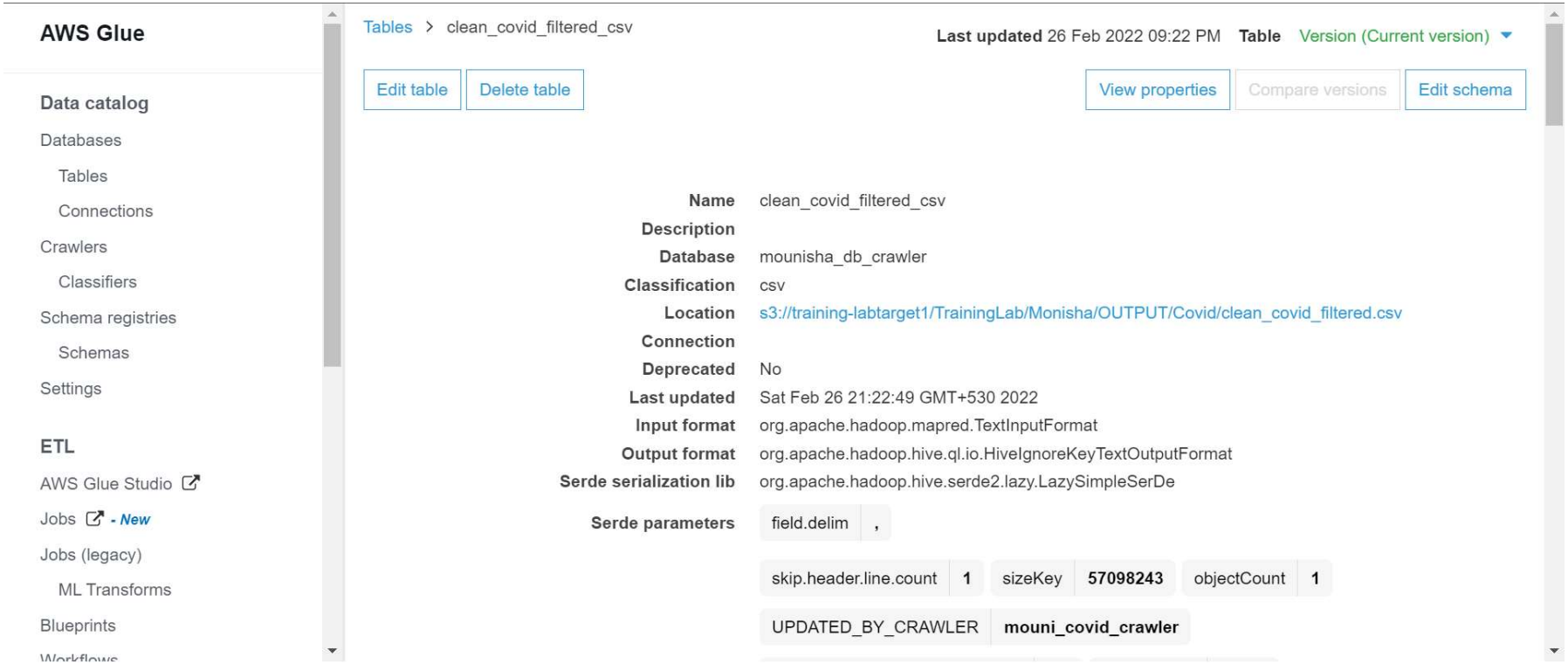


3. Storing the modified data in AWS Cont…



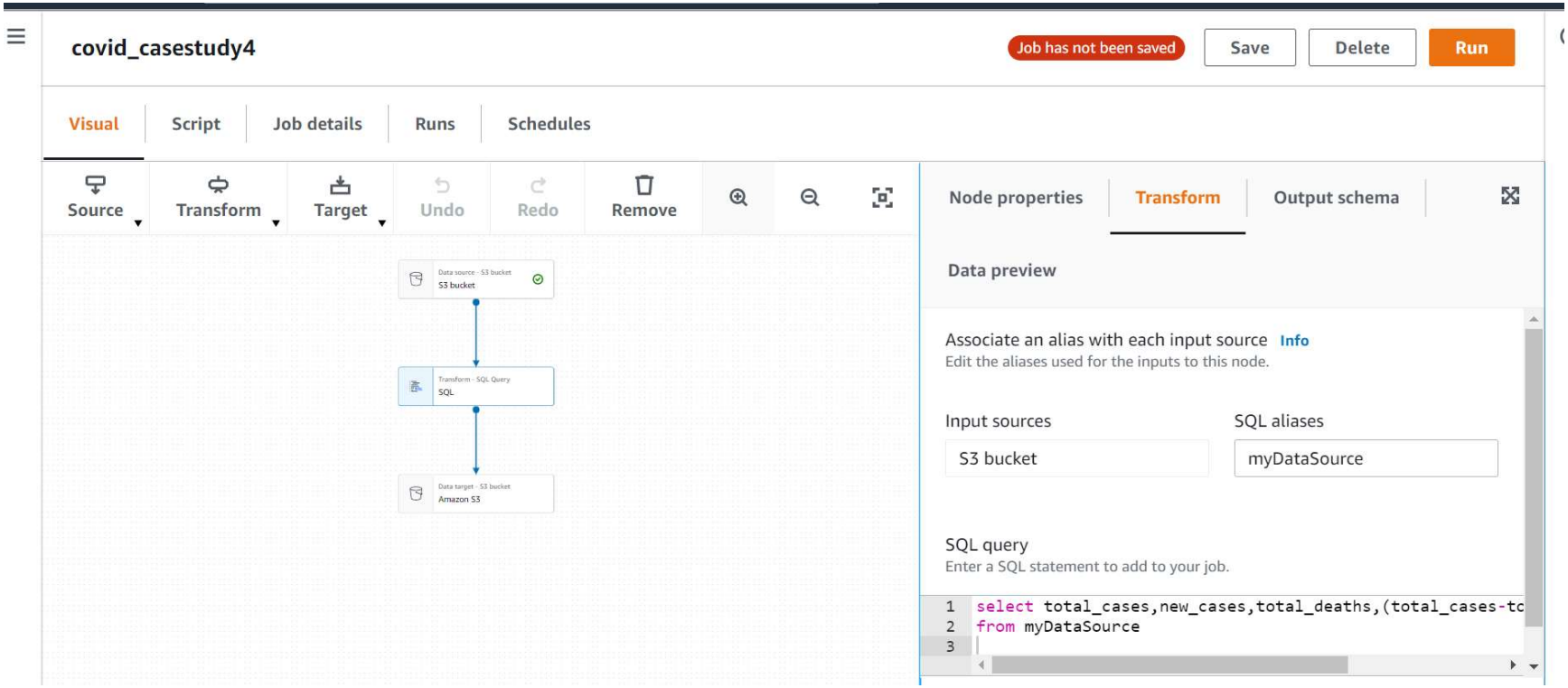
4.Display total cases ,new cases , recovered cases and deaths.

CREATING CRAWLER

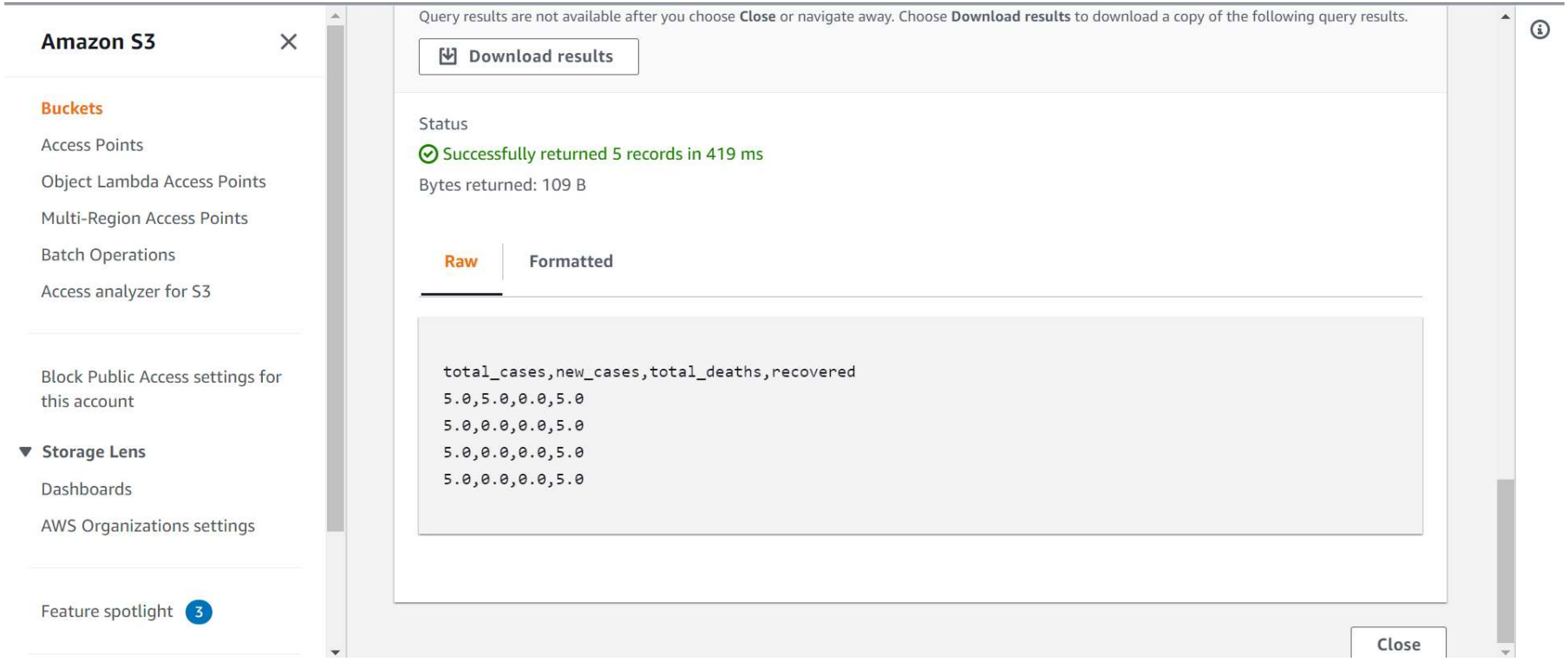


4. Display total cases ,new cases ,recovered cases and deaths.

Job name:covid_casestudy4

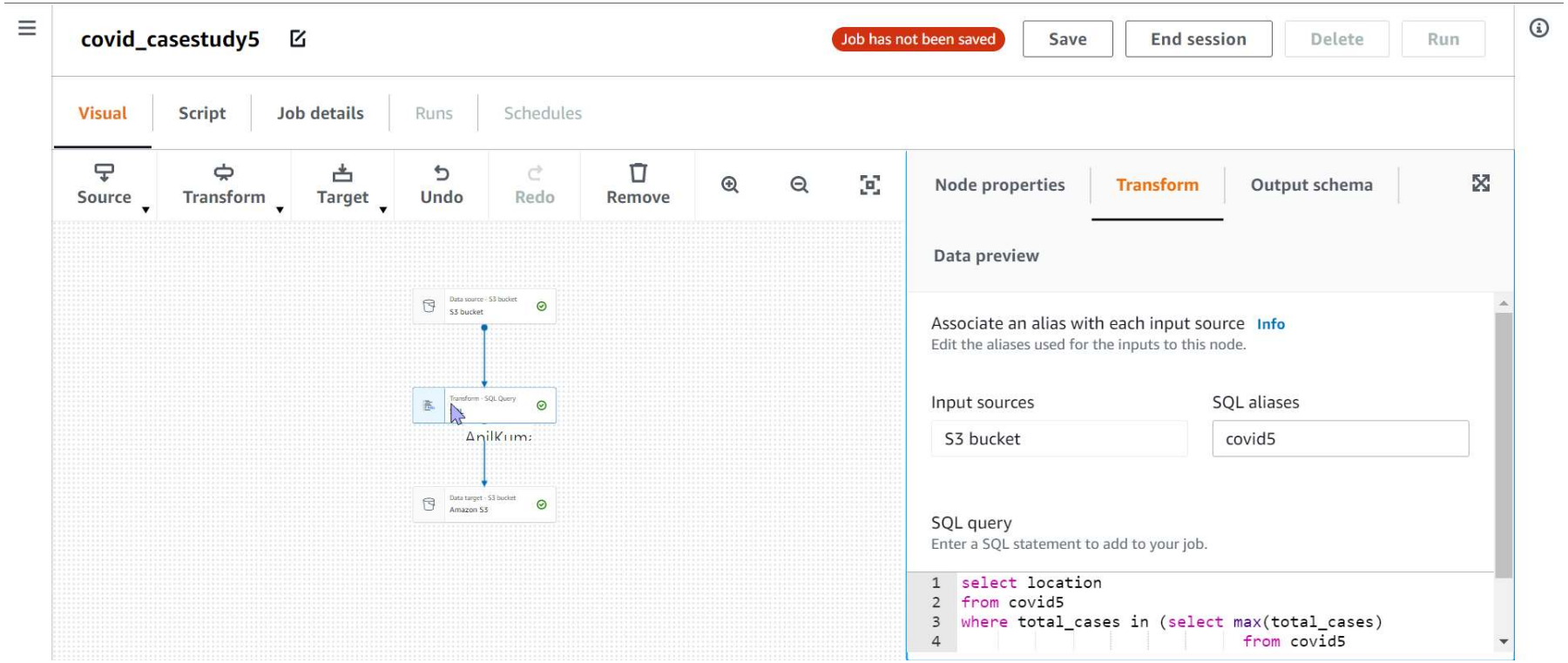


4.OUTPUT

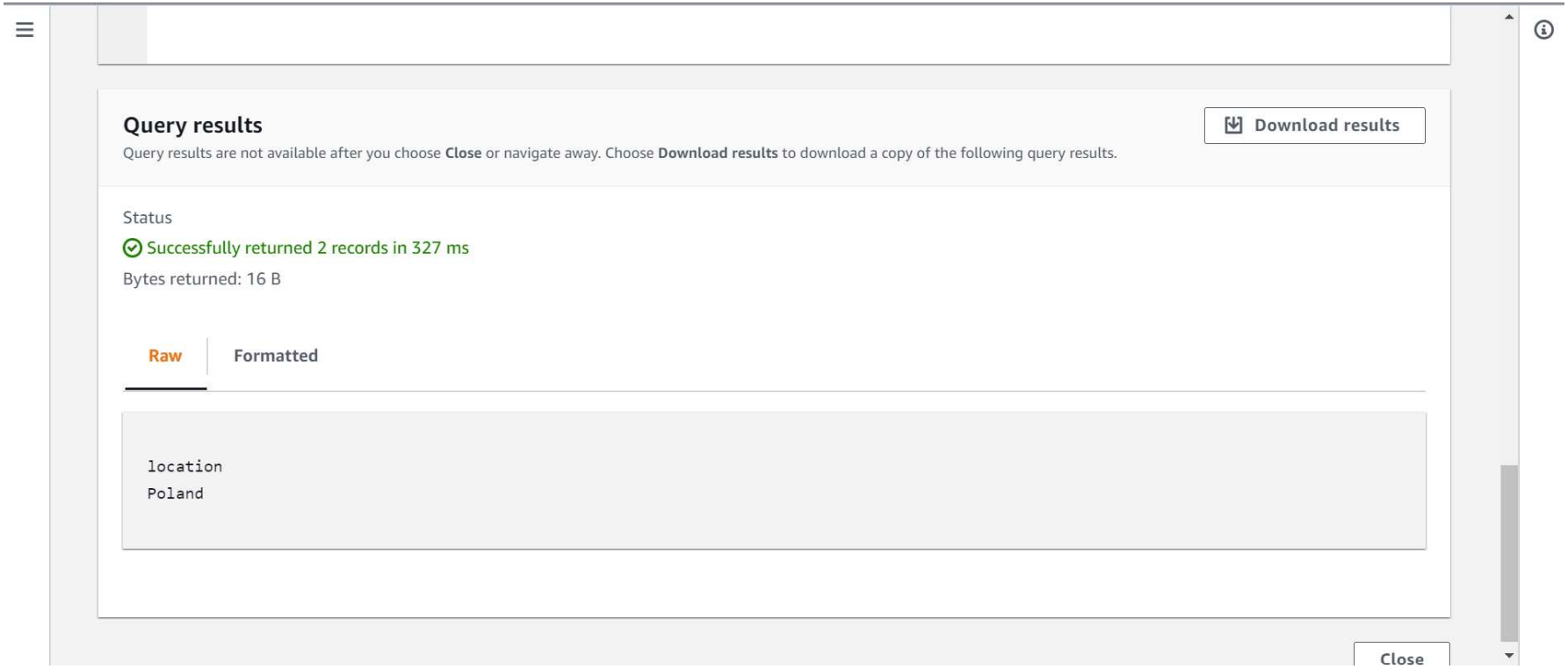

Amazon S3
Buckets
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
Access analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
AWS Organizations settings
Feature spotlight 3
Query results are not available after you choose Close or navigate away. Choose Download results to download a copy of the following query results.
Download results
Status
Successfully returned 5 records in 419 ms
Bytes returned: 109 B
Raw
Formatted
total_cases,new_cases,total_deaths,recovered
5.0,5.0,0.0,5.0
5.0,0.0,0.0,5.0
5.0,0.0,0.0,5.0
5.0,0.0,0.0,5.0
Close

5.Which country in Distinct WHO region highest cases in till date.

JOB NAME :covid_casestudy5

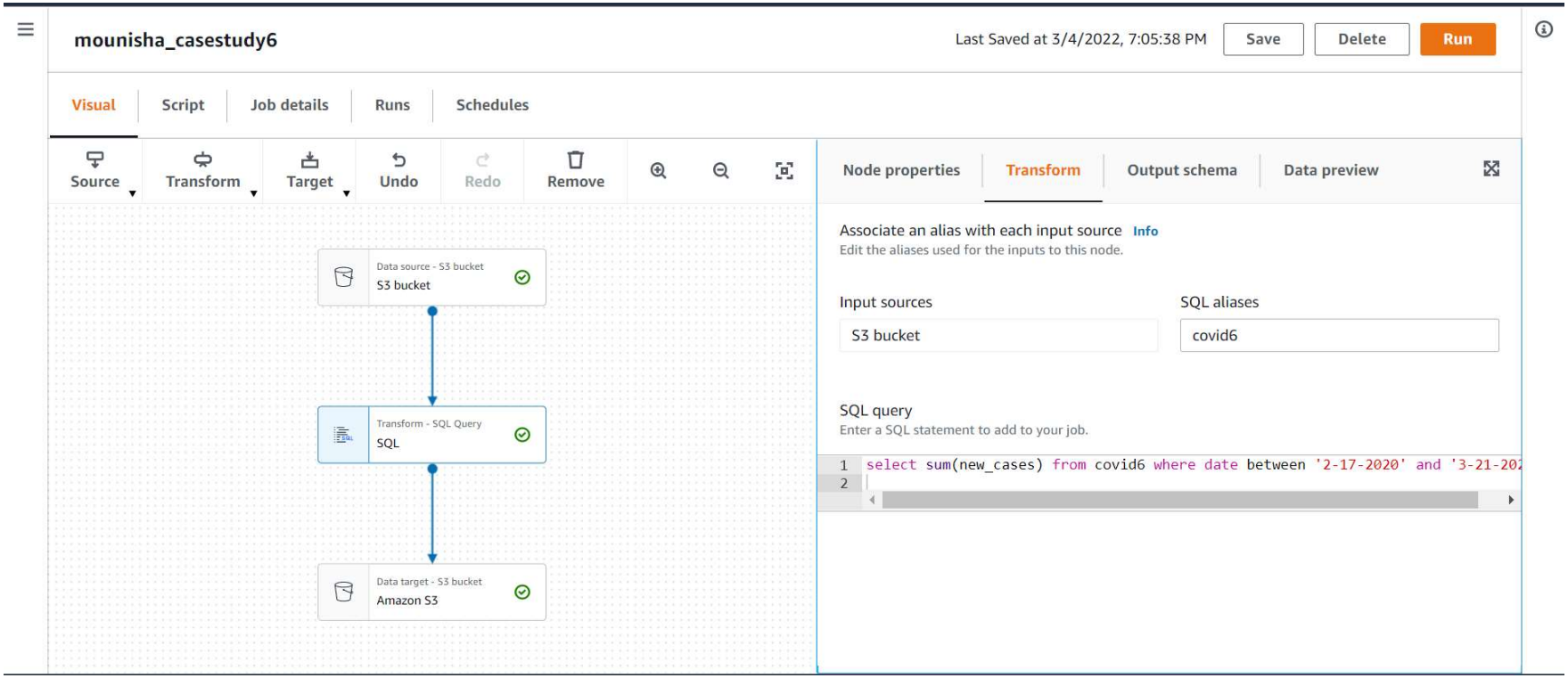


5.OUTPUT

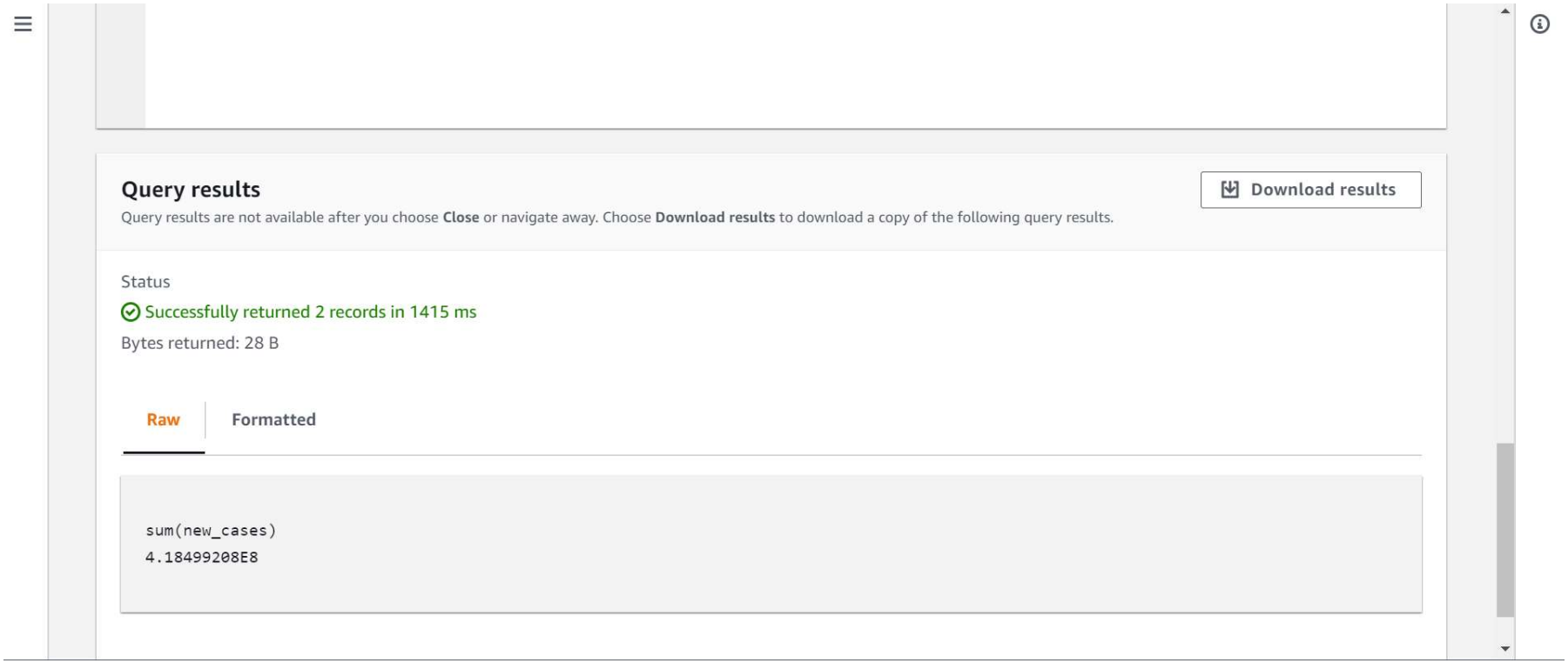

Query results
Query results are not available after you choose Close or navigate away. Choose Download results to download a copy of the following query results.
Download results
Status
Successfully returned 2 records in 327 ms
Bytes returned: 16 B
Raw
Formatted
location
Poland
Close

6. Total no of confirmed cases over between a certain date.

JOB NAME : mounisha_casestudy6

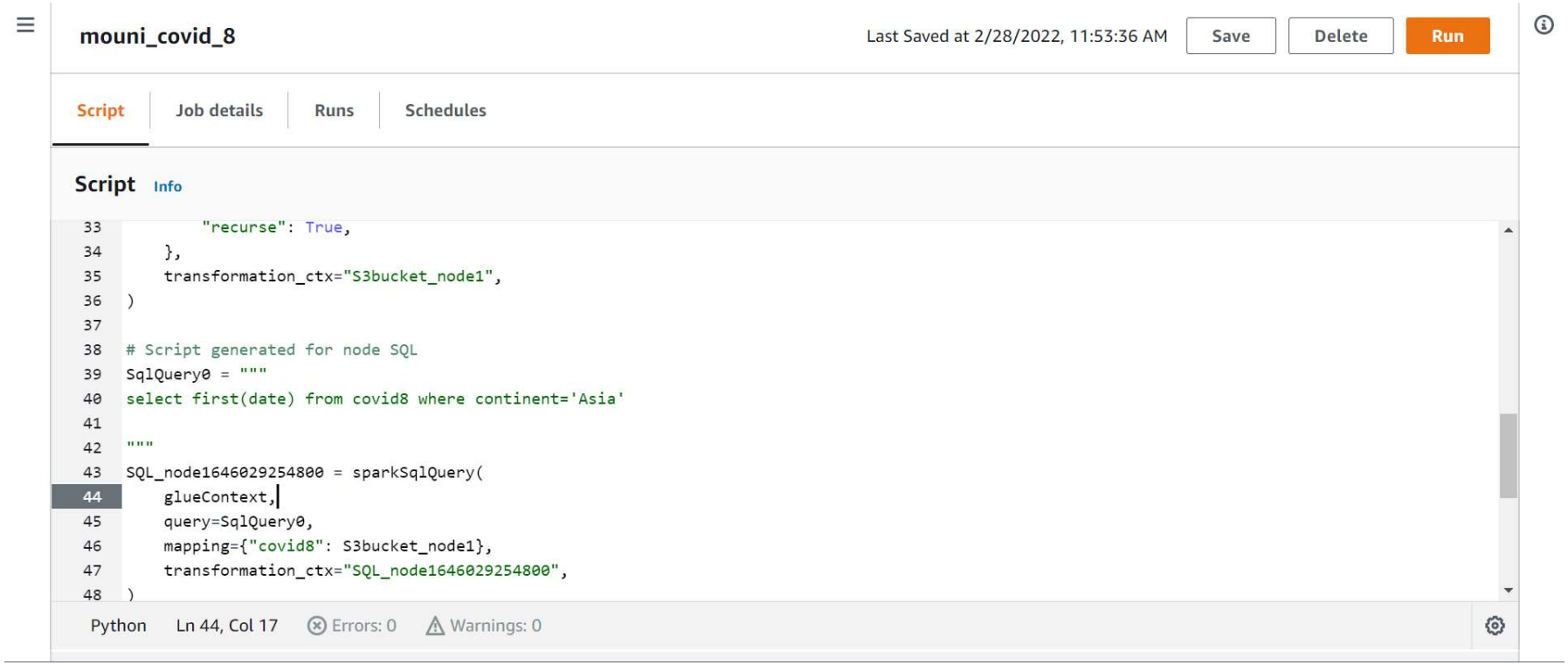


6.OUTPUT

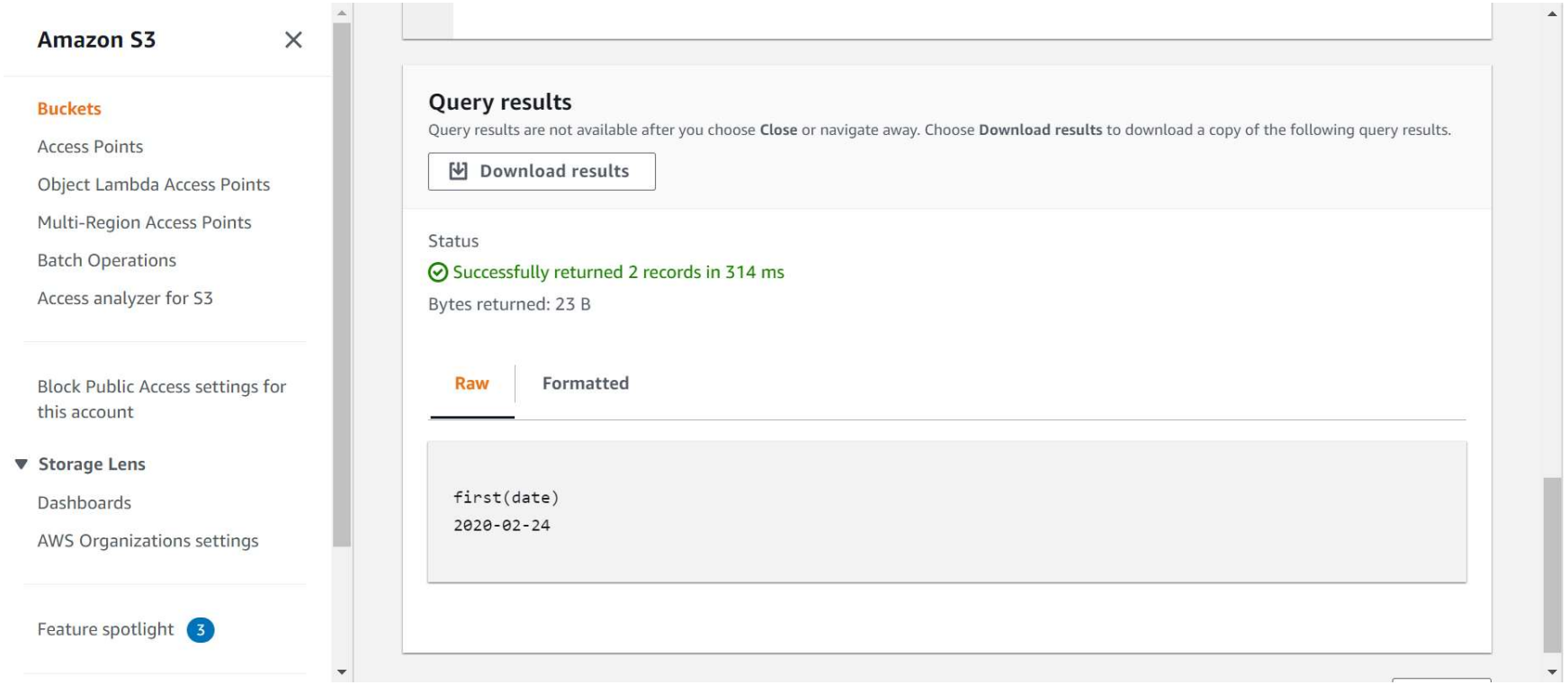

Query results
Download results
Query results are not available after you choose Close or navigate away. Choose Download results to download a copy of the following query results.
Status
Successfully returned 2 records in 1415 ms
Bytes returned: 28 B
Raw
Formatted
sum(new_cases)
4.18499208E8

8.Date of first confirmed case in a particular region.

JOB NAME :mouni_covid_8

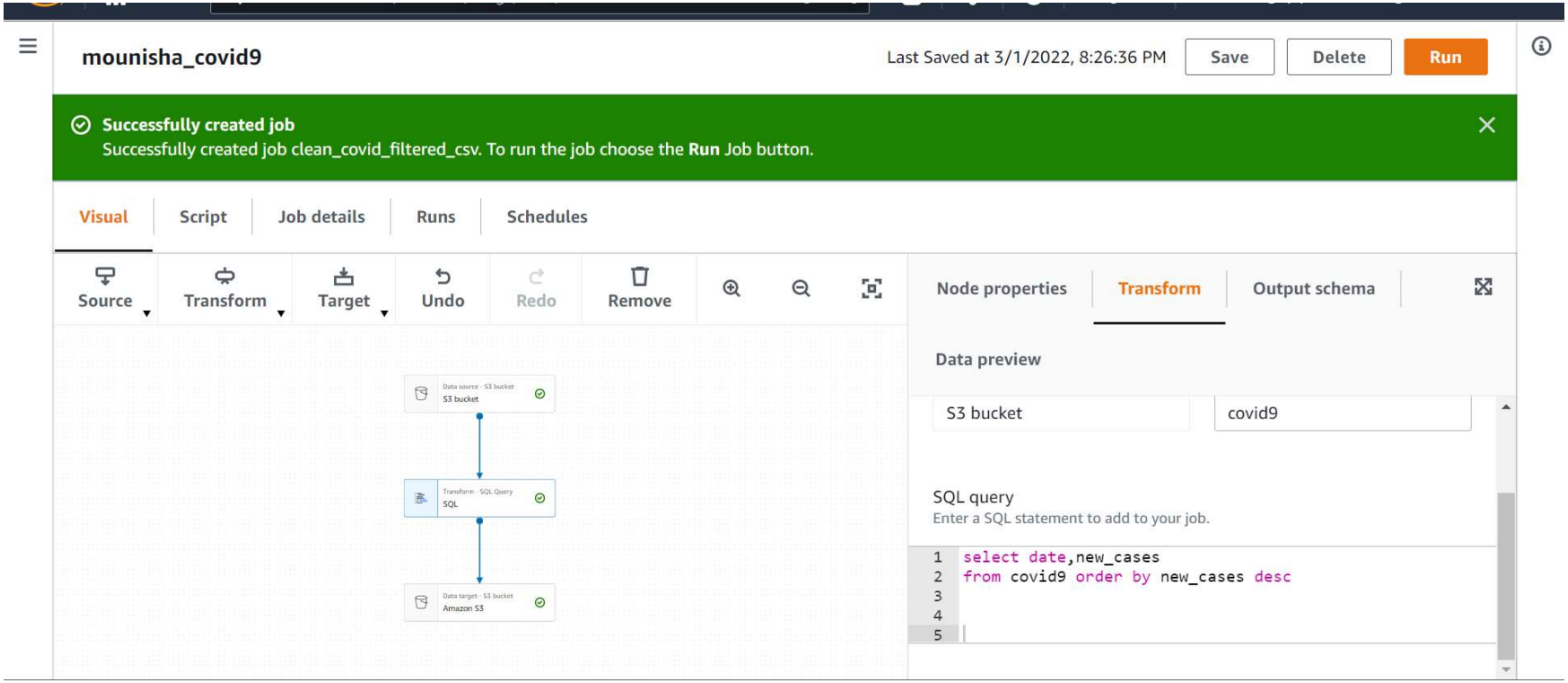


8.OUTPUT

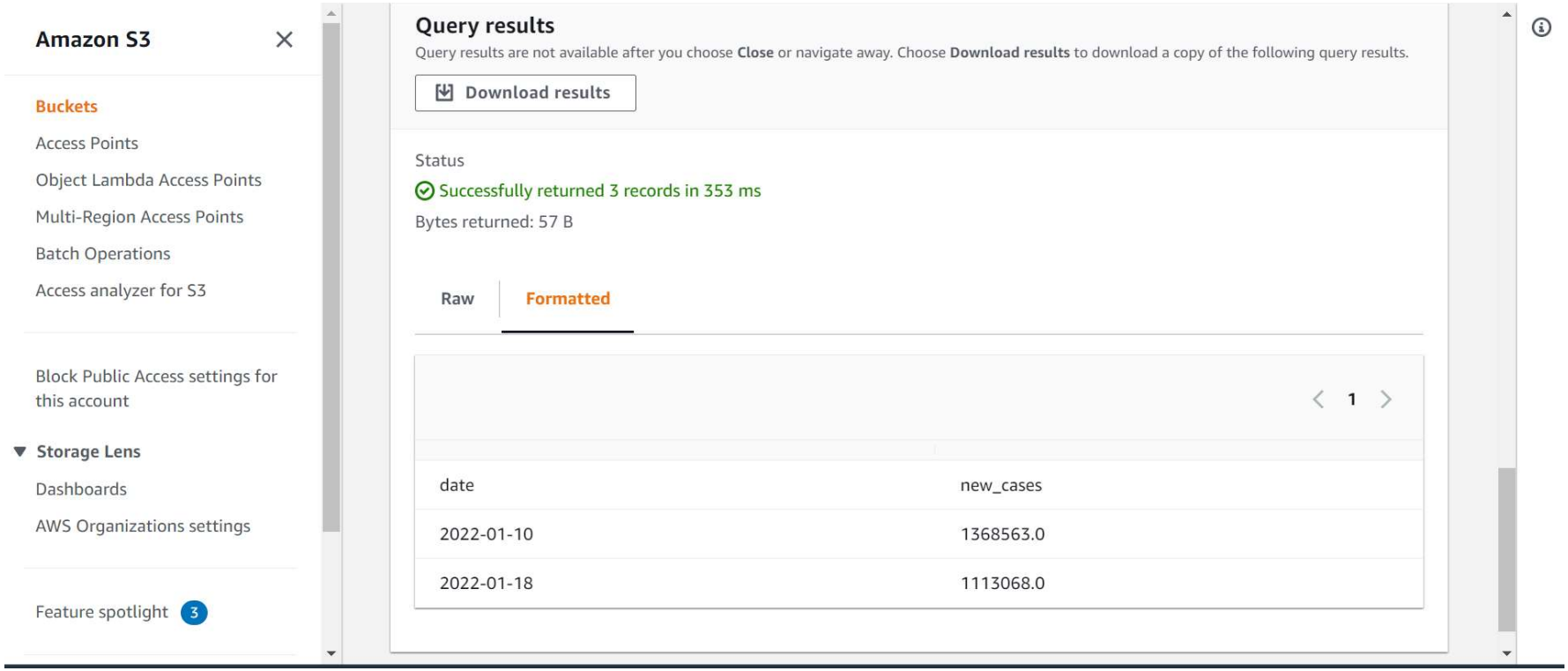

Amazon S3
Buckets
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
Access analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
AWS Organizations settings
Feature spotlight 3
Query results
Query results are not available after you choose Close or navigate away. Choose Download results to download a copy of the following query results.
Download results
Status
Successfully returned 2 records in 314 ms
Bytes returned: 23 B
Raw
Formatted
first(date)
2020-02-24

9.Date on which max no. cases were reported in a country.

JOB NAME : mounisha_covid9

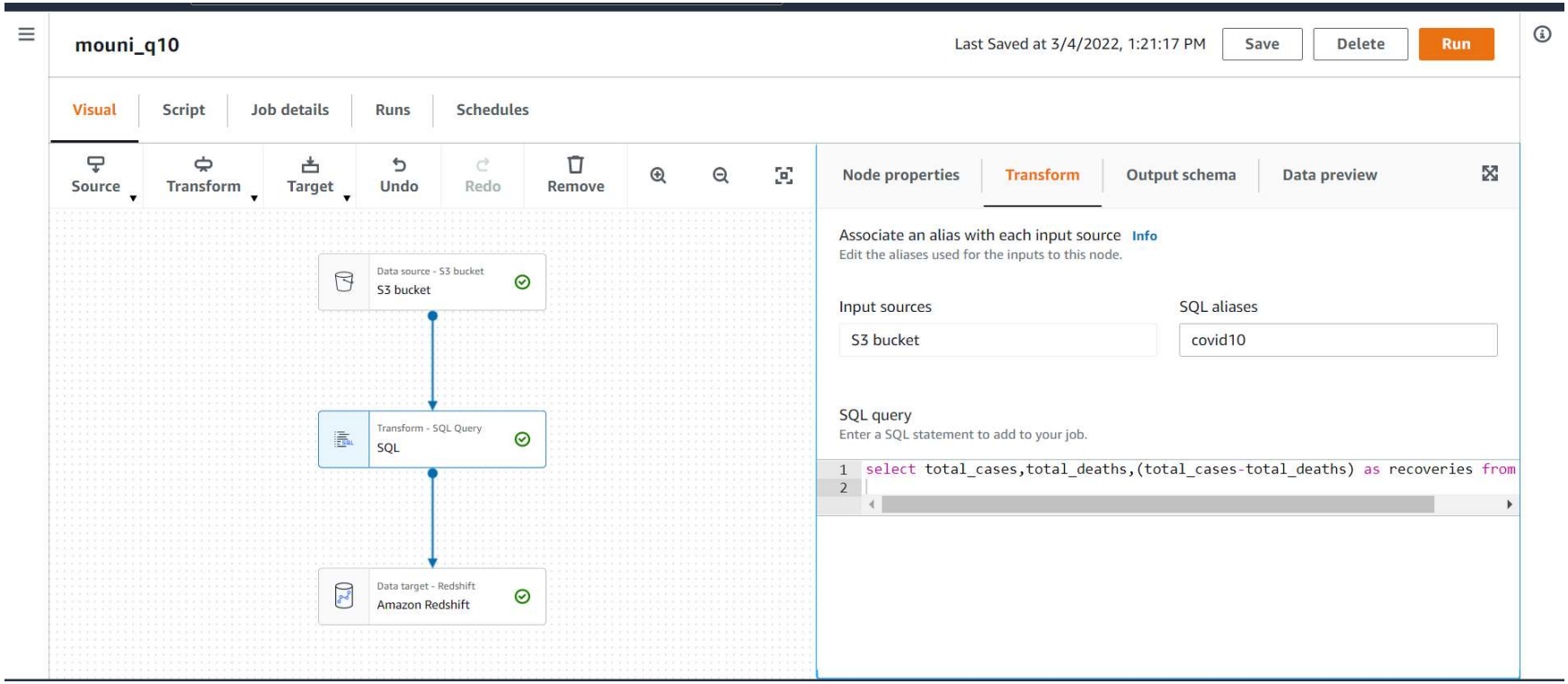


9.OUTPUT

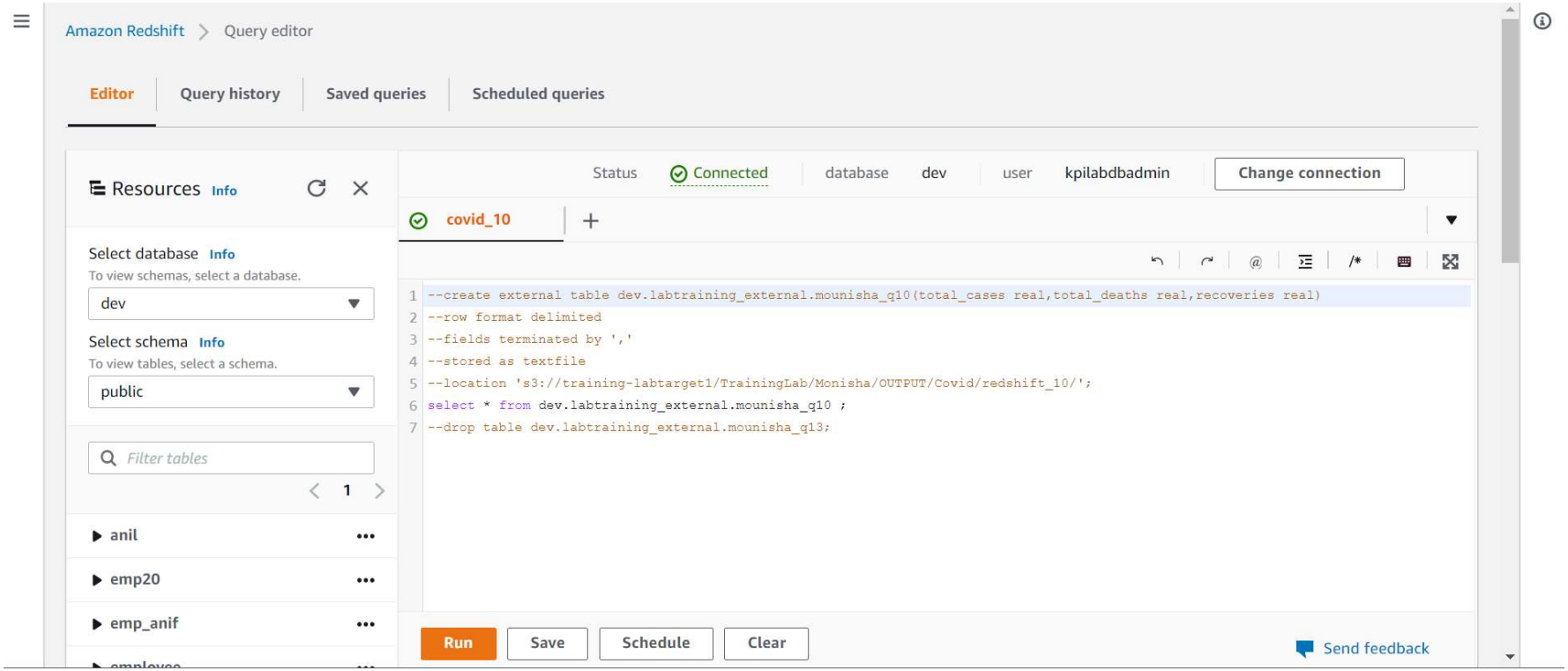

Amazon S3
Buckets
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
Access analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
AWS Organizations settings
Feature spotlight 3
Query results
Query results are not available after you choose Close or navigate away. Choose Download results to download a copy of the following query results.
Download results
Status
Successfully returned 3 records in 353 ms
Bytes returned: 57 B
Raw
Formatted
1
date
new_cases
2022-01-10
1368563.0
2022-01-18
1113068.0

10.Line chart showing total cases, deaths &recoveries of a particular country .

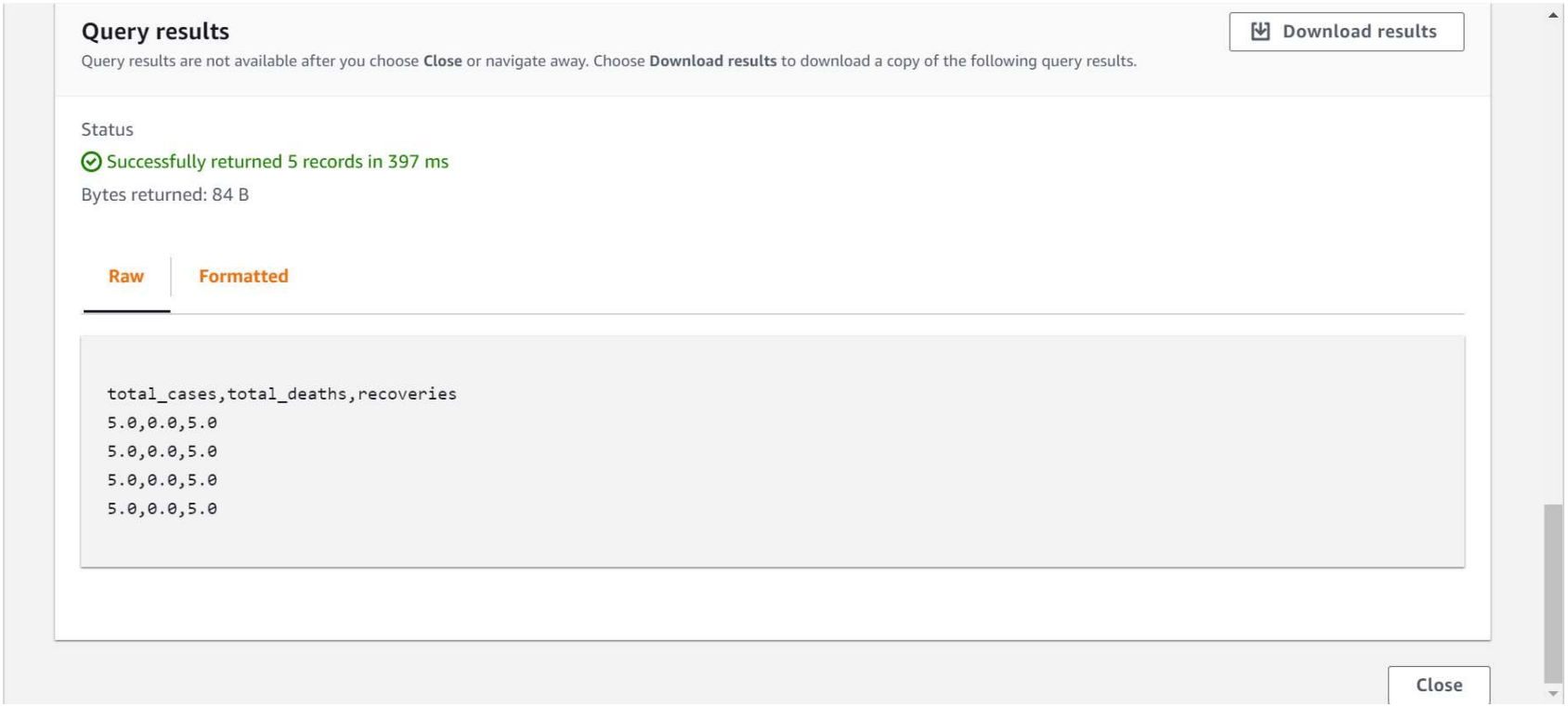
JOB NAME: mouni_q10



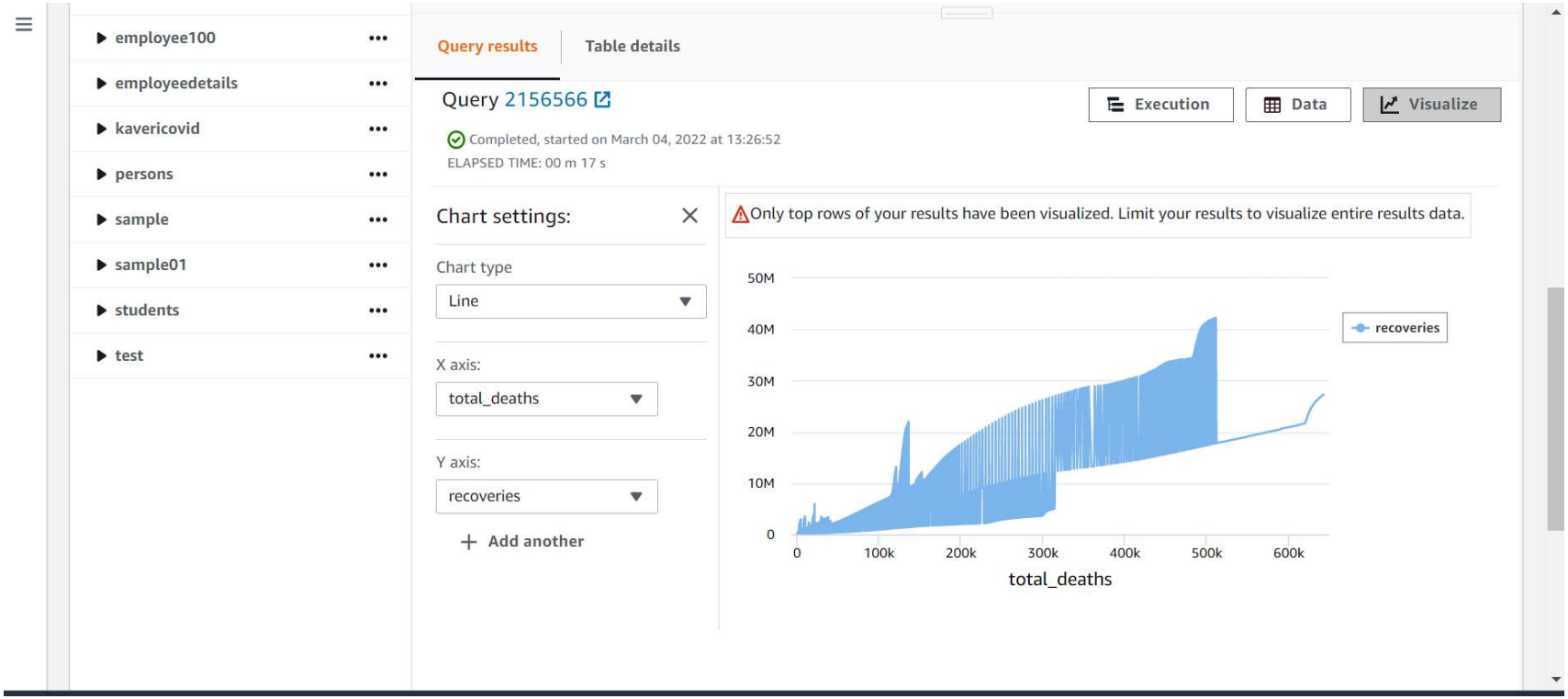
10 . Creating external table



10. OUTPUT

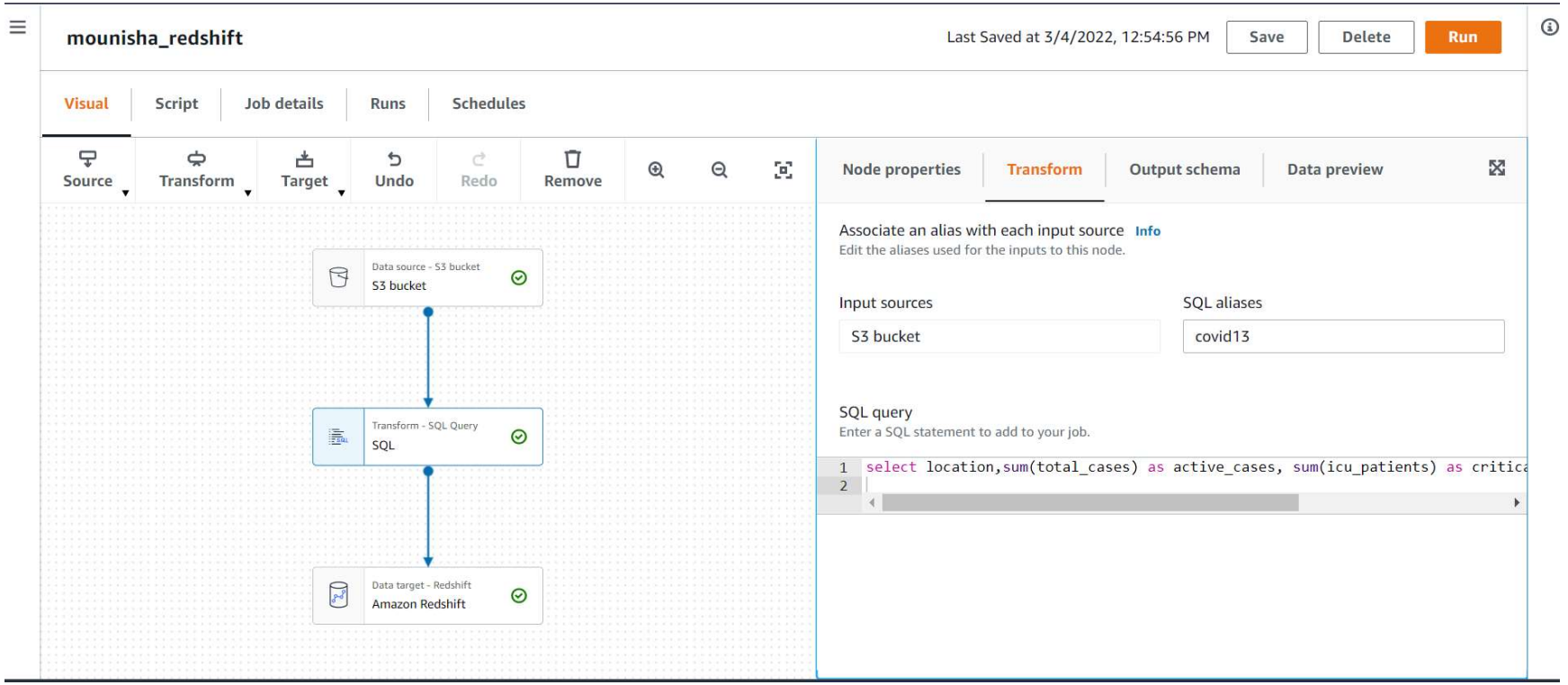

Query results
Download results
Query results are not available after you choose Close or navigate away. Choose Download results to download a copy of the following query results.
Status
Successfully returned 5 records in 397 ms
Bytes returned: 84 B
Raw
Formatted
total_cases,total_deaths,recoveries
5.0,0.0,5.0
5.0,0.0,5.0
5.0,0.0,5.0
5.0,0.0,5.0
Close

10. Line chart

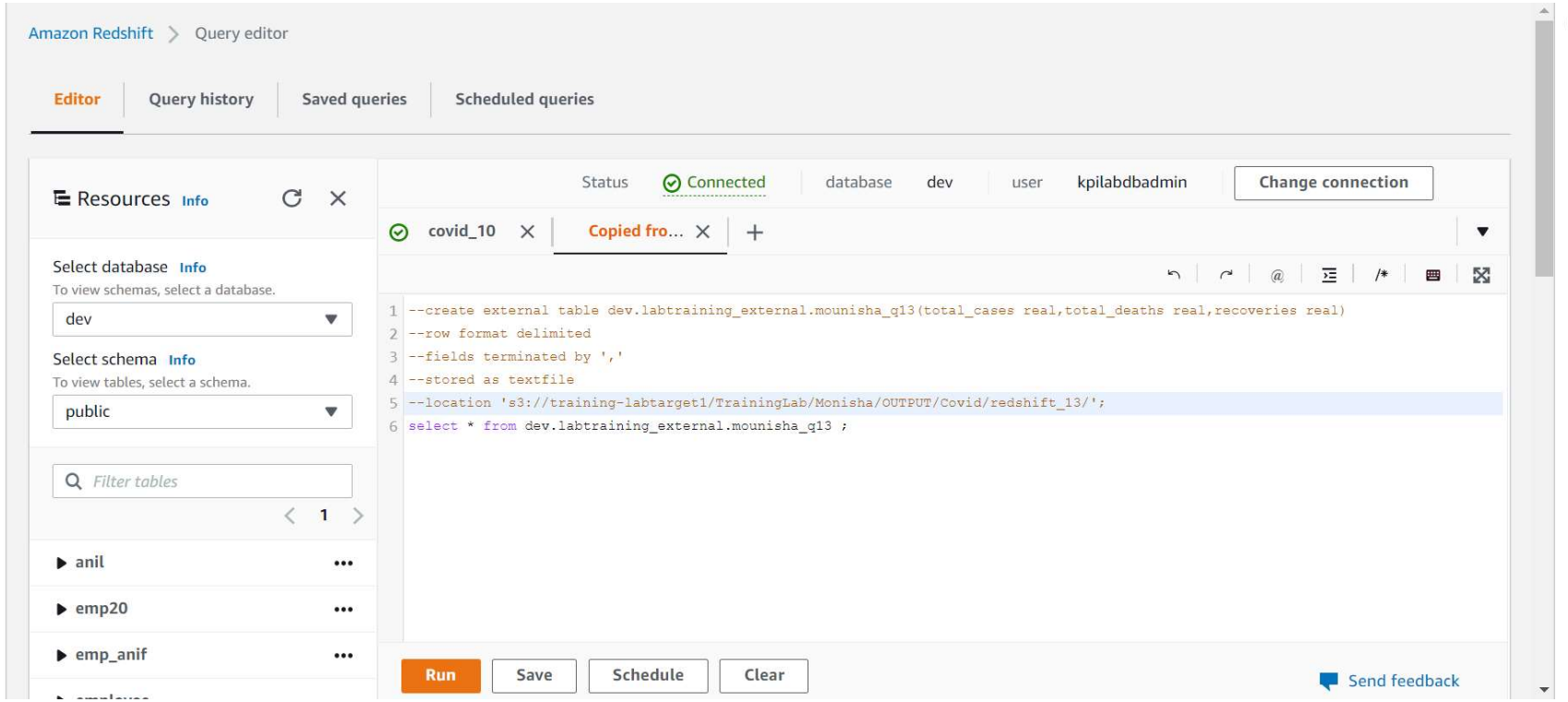


13.No. of active cases vs critical cases of a country.

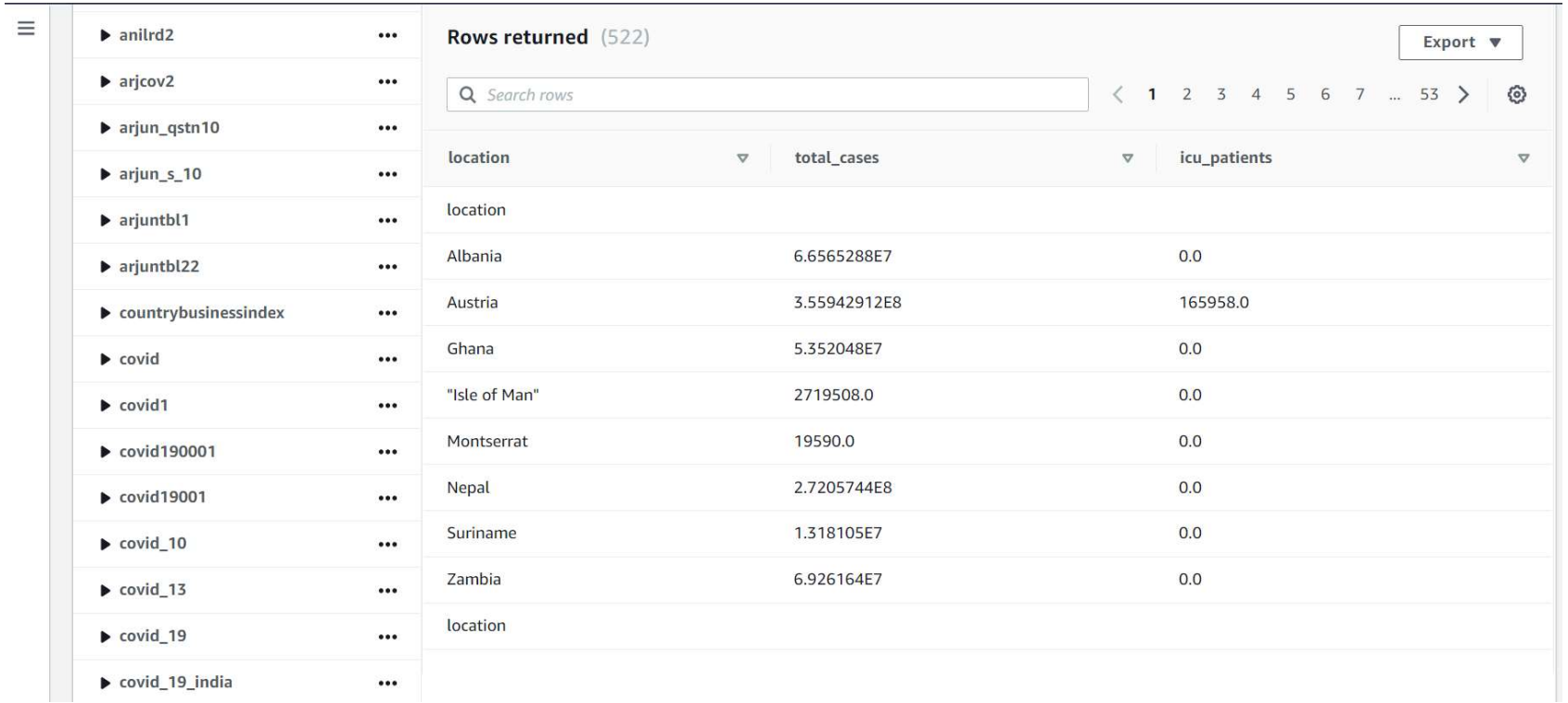
JOB NAME: mounisha_redshift



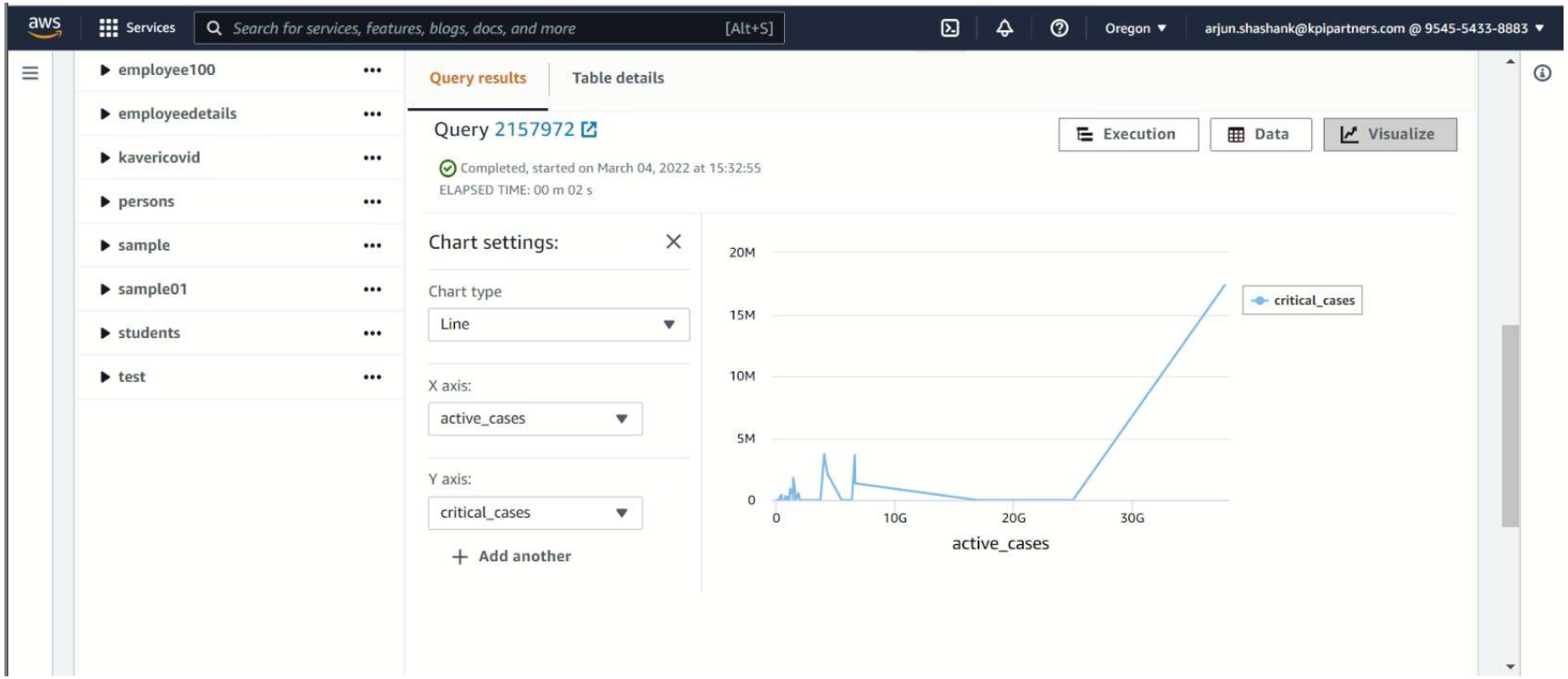
13. Creating external table



13. OUTPUT


anilrd2
arjcov2
arjun_qstn10
arjun_s_10
arjuntbl1
arjuntbl22
countrybusinessindex
covid
covid1
covid190001
covid19001
covid_10
covid_13
covid_19
covid_19_india
Rows returned (522)
Export
Search rows
1 2 3 4 5 6 7 ... 53
location
total_cases
icu_patients
location
Albania 6.6565288E7 0.0
Austria 3.55942912E8 165958.0
Ghana 5.352048E7 0.0
"Isle of Man" 2719508.0 0.0
Montserrat 19590.0 0.0
Nepal 2.7205744E8 0.0
Suriname 1.318105E7 0.0
Zambia 6.926164E7 0.0
location

13. active cases vs critical cases



THANK YOU