

# COVID-19 DATA ANALYSIS IN AWS BY USING SCALA

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

ScriptJob detailsRunsSchedules

ScriptInfo

```
32    val S3bucket_node1 = glueContext.getSourceWithFormat(formatOptions=JsonOptions("""{"quoteChar": "\"", "withHeader": true, "separator": ","}"""), connectionType
    ="s3", format="csv", options=JsonOptions("""{"paths": ["s3://training-labtarget1/TrainingLab/Monisha/INPUT/covid/INPUT/clean_covid_filtered.csv"],
    "recurse": true}"""), transformationContext="S3bucket_node1").getDynamicFrame()
33
34    // Script generated for node SQL
35    val SqlQuery0: String = """select total_cases,new_cases,total_deaths,(total_cases-total_deaths) as recovered from myDataSource
36    |""".stripMargin
37
38    val SQL_node1646623934398 = SparkSqlQuery.execute(glueContext = glueContext, sqlContext = sql, query = SqlQuery0, mapping = Map("myDataSource" ->
    S3bucket_node1)
39    )
40
41    // Script generated for node Amazon S3
42    val repartition=SQL_node1646623934398.repartition(1)
43    val AmazonS3_node1646624010085 = glueContext.getSinkWithFormat(connectionType="s3", options=JsonOptions("""{"path": "s3://training-labtarget1/TrainingLab
    /Monisha/OUTPUT/covid/scala_4/", "partitionKeys": []}"""), transformationContext="AmazonS3_node1646624010085", format="csv").writeDynamicFrame(repartition)
44
45    Job.commit()
46  }
```

ScalaLn 1, Col 1⊗ Errors: 0⚠ Warnings: 0⚙

Language: scala



mouni\_scala

Last Saved at 3/7/2022, 9:33:16 AM

Save

Delete

Run

Script

Job details

Runs

Schedules

Spark

▼

Glue version

[Info](#)

Glue 3.0 - Supports spark 3.1, Scala 2, Python 3

▼

Language

Scala

▼

Worker type

Set the type of predefined worker that is allowed when a job runs.

G.1X

▼

☐ Automatically scale the number of workers - *preview*

AWS Glue will optimize costs and resource usage by dynamically scaling the number of workers up and down throughout the job run.  
Requires Glue 3.0 or later.

Requested number of workers

The number of workers you want AWS Glue to allocate to this job.

10

☒ Generate job insights

## 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

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 386 ms

Bytes returned: 109 B

Raw

Formatted

< 1 >

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

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

mouni\_scala5

Last Saved at 3/7/2022, 9:48:34 AM

Save

Delete

Run

Script

Job details

Runs

Schedules

Script

Info

```
33
34 // Script generated for node SQL
35 val SqlQuery0: String = """select location
36 |from myDataSource
37 |where total_cases in (select max(total_cases)
38 |                        from myDataSource)
39 |""".stripMargin
40
41 val SQL_node1646626401342 = SparkSqlQuery.execute(glueContext = glueContext, sqlContext = sql, query = SqlQuery0, mapping = Map("myDataSource" ->
42   S3bucket_node1)
43 )
44
45 // Script generated for node Amazon S3
46 val repartition=SQL_node1646626401342.repartition(1)
47 val AmazonS3_node1646626517779 = glueContext.getSinkWithFormat(connectionType="s3", options=JsonOptions("""{"path": "s3://training-labtarget1/TrainingLab
48   /Monisha/OUTPUT/covid/scala5/", "partitionKeys": []}"""), transformationContext="AmazonS3_node1646626517779", format="csv").writeDynamicFrame(repartition)
49
50 Job.commit()
51 }
```

Scala

Ln 1, Col 1

Errors: 0

Warnings: 0

## 5.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 337 ms**



Bytes returned: 16 B

**Raw** | **Formatted**

< 1 >	
location	
Poland	

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




 mouni\_scala6 Last Saved at 3/7/2022, 10:01:45 AM Save Delete Run

 **Successfully created job**  
Successfully created job mouni\_scala6. To run the job choose the Run Job button. 

Script Job details Runs Schedules

**Script** [Info](#)

```
34 // Script generated for node SQL
35 val SqlQuery0: String = ""select sum(new_cases) from myDataSource where date between '2/17/2020' and '3/21/2020'
36 |
37 |"".stripMargin
38
39 val SQL_node1646627169403 = SparkSqlQuery.execute(glueContext = glueContext, sqlContext = sql, query = SqlQuery0, mapping = Map("myDataSource" ->
40   S3bucket_node1)
41 )
42 // Script generated for node Amazon S3
43 val repartition=SQL_node1646627169403.repartition(1)
44 val AmazonS3_node1646627254446 = glueContext.getSinkWithFormat(connectionType="s3", options=JsonOptions("""{"path": "s3://training-labtarget1/TrainingLab
45   /Monisha/OUTPUT/covid/scala_6/", "partitionKeys": []}"""), transformationContext="AmazonS3_node1646627254446", format="csv").writeDynamicFrame(repartition)
46
47 Job.commit()
```

Scala Ln 44, Col 188  Errors: 0  Warnings: 0 

## 6.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 316 ms**


Bytes returned: 44 B

**Raw** | Formatted

```
sum(CAST(new_cases AS DOUBLE))
4.18499208E8
```



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

 **mouni\_scala8** Last Saved at 3/7/2022, 10:29:05 AM Save Delete Run 

Script Job details Runs Schedules

**Script** Info

```
32 val S3bucket_node1 = glueContext.getSourceWithFormat(formatOptions=JsonOptions("""{"quoteChar": "\"", "withHeader": true, "separator": ","}"""), connectionType
33   ="s3", format="csv", options=JsonOptions("""{"paths": ["s3://training-labtarget1/TrainingLab/Monisha/INPUT/covid/INPUT/clean_covid_filtered.csv"],
34   "recurse": true}"""), transformationContext="S3bucket_node1").getDynamicFrame()
35
36 // Script generated for node SQL
37 val SqlQuery0: String = """select first(date) from myDataSource where continent ='Asia'
38 | """.stripMargin
39
40 val SQL_node1646628900230 = SparkSqlQuery.execute(glueContext = glueContext, sqlContext = sql, query = SqlQuery0, mapping = Map("myDataSource" ->
41   S3bucket_node1)
42 )
43
44 // Script generated for node Amazon S3
45 val repartition=SQL_node1646628900230.repartition(1)
46 val AmazonS3_node1646628952501 = glueContext.getSinkWithFormat(connectionType="s3", options=JsonOptions("""{"path": "s3://training-labtarget1/TrainingLab
47   /Monisha/OUTPUT/covid/scala_8/", "partitionKeys": []}"""), transformationContext="AmazonS3_node1646628952501", format="csv").writeDynamicFrame(repartition)
48
49 Job.commit()
50 }
```

Scala Ln 1, Col 1  Errors: 0  Warnings: 0 

## 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 422 ms**

Bytes returned: 23 B

**Raw** | Formatted

first(date)
2020-02-24

Close



Python Ln 1, Col 1  Errors: 0  Warnings: 0

## 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 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 341 ms

Bytes returned: 88 B

**Raw** | Formatted

date,new_cases
2021-12-16,9999.0
2020-10-30,99970.0
2021-01-13,9997.0
2022-01-17,9996.0

Close

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



mouni\_py10

Last Saved at 3/7/2022, 11:25:06 AM

Save

Delete

Run

Script

Job details

Runs

Schedules

Script

Info

```
46     query=SqlQuery0,
47     mapping={"myDataSource": S3bucket_node1},
48     transformation_ctx="SQL_node1646631617149",
49 )
50
51 # Script generated for node Amazon S3
52 AmazonS3_node1646631753038 = glueContext.write_dynamic_frame.from_options(
53     frame=SQL_node1646631617149,
54     connection_type="s3",
55     format="csv",
56     connection_options={
57         "path": "s3://training-labtarget1/TrainingLab/Monisha/OUTPUT/covid/py_10/",
58         "partitionKeys": [],
59     },
60     transformation_ctx="AmazonS3_node1646631753038",
61 )
62
63 job.commit()
64
```

Python

Ln 55, Col 18

Errors: 0

Warnings: 0

## 10 . 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 5 records in 340 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

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

≡

mouni\_py13

Last Saved at 3/7/2022, 11:39:06 AM

Save

Delete

Run

ⓘ

Script

Job details

Runs

Schedules

Script

Info

```
46 query=SqlQuery0,
47 mapping={"myDataSource": S3bucket_node1},
48 transformation_ctx="SQL_node1646632987734",
49 )
50
51 # Script generated for node Amazon S3
52 AmazonS3_node1646633012576 = glueContext.write_dynamic_frame.from_options(
53     frame=SQL_node1646632987734,
54     connection_type="s3",
55     format="csv",
56     connection_options={
57         "path": "s3://training-labtarget1/TrainingLab/Monisha/OUTPUT/covid/py_13/",
58         "partitionKeys": [],
59     },
60     transformation_ctx="AmazonS3_node1646633012576",
61 )
62
63 job.commit()
64
```

Python

Ln 1, Col 1

⊗ Errors: 0

⚠ Warnings: 0

⚙

## 13.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 331 ms

Bytes returned: 121 B

Raw

Formatted

location,active\_cases,critical\_cases

Bhutan,944258.0,0.0

Cuba,2.00705355E8,0.0

Malta,1.5490864E7,4011.0

Tonga,1173.0,0.0

Close



## 13.output



aws

Services

Search for services, features, blogs, docs, and more [Alt+S]

Oregon

arjun.shashank@kpipartners.com @ 9545-5433-8883

employee100

employeeedetails

kavericovid

persons

sample

sample01

students

test

Query results

Table details

Query 2157972

Execution

Data

Visualize

Completed, started on March 04, 2022 at 15:32:55

ELAPSED TIME: 00 m 02 s

Rows returned (226)

Export

Search rows

< 1 2 3 4 5 6 7 ... 23 >

location	active_cases	critical_cases
Mongolia	247132732	0
Chad	4731688	0
Canada	1394964096	767420
Uruguay	268668010	0
Jersey	0	0
Jordan	681058224	0
"North Macedonia"	155562104	0



THANK YOU