



Amazon S3 console interface showing the 'Objects' tab for a bucket named 's3-cleaned-bucket-ishan35-dev'. The interface includes a sidebar with navigation options like Buckets, Access Points, and Storage Lens. The main content area displays a list of objects, including a file named '032460a2340a4205a86765243952d1e7.snappy.parquet' with a size of 6.1 KB and a storage class of Standard. The file was last modified on August 6, 2023, at 10:27:15 (UTC+05:30). The console also shows options for actions like Copy S3 URI, Copy URL, Download, Open, and Delete, as well as a search bar for finding objects by prefix.

Name	Type	Last modified	Size	Storage class
032460a2340a4205a86765243952d1e7.snappy.parquet	parquet	August 6, 2023, 10:27:15 (UTC+05:30)	6.1 KB	Standard

AWS Glue console interface showing the 'Databases' tab. The interface includes a sidebar with navigation options like Getting started, ETL jobs, and Data Catalog. The main content area displays a list of databases, including 'covid_x_ishan_data', 'de_youtube_cleaned', 'de_youtube_raw', 'default', and 'ishan-cloudtubex'. The 'default' database was last updated on August 8, 2023, at 10:13:23. The console also shows options for actions like Edit, Delete, and Add database.

Name	Description	Location URI	Created on (UTC)
covid_x_ishan_data	-	-	August 7, 2023 at 11:26:03
de_youtube_cleaned	-	-	August 5, 2023 at 21:11:18
de_youtube_raw	-	-	August 6, 2023 at 12:47:40
default	-	-	August 6, 2023 at 09:09:37
ishan-cloudtubex	-	-	August 5, 2023 at 18:36:52

Visual Editor

us-east-1.console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/bl...

Visual 1

Source Action Target Undo Redo Remove

Untitled job

Unsaved job found

We found an unsaved job, do you wish to restore it?

Restore

No node selected

Choose a node from the graph to view its configuration properties.

Data source - S3 bucket

Amazon S3

Transform - SQL Query

SQL Query

© 2023, Amazon Web Services India Private Limited or its affiliates. Privacy Terms Cookie preferences

Script Editor

us-east-1.console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/bl...

Script (Locked)

Generate classic script

Download script

Edit script

```
1 import sys
2 from aws glue.transforms import *
3 from aws glue.utils import getResolvedOptions
4 from pyspark.context import SparkContext
5 from aws glue.context import GlueContext
6 from aws glue.job import Job
7 from aws glue import DynamicFrame
8
9
10 def sparkSqlQuery(glueContext, query, mapping, transformation_ctx) -> DynamicFrame:
11     for alias, frame in mapping.items():
12         frame.toDF().createOrReplaceTempView(alias)
13         result = spark.sql(query)
14         return DynamicFrame.fromDF(result, glueContext, transformation_ctx)
15
16
17 args = getResolvedOptions(sys.argv, ["JOB_NAME"])
18 sc = SparkContext()
19 glueContext = GlueContext(sc)
20 spark = glueContext.spark_session
21 job = Job(glueContext)
22 job.init(args["JOB_NAME"], args)
23
24 # Script generated for node Amazon S3
25 AmazonS3_node1691489637560 = glueContext.create_dynamic_frame.from_options(
26     format_options={},
27     connection_type="s3",
28     format="parquet",
29     connection_options={"paths": [""], "recurse": True},
30     transformation_ctx=AmazonS3_node1691489637560,
31 )
```

https://us-east-1.console.aws.amazon.com/gluestudio/home?region=us-east-1#

© 2023, Amazon Web Services India Private Limited or its affiliates. Privacy Terms Cookie preferences