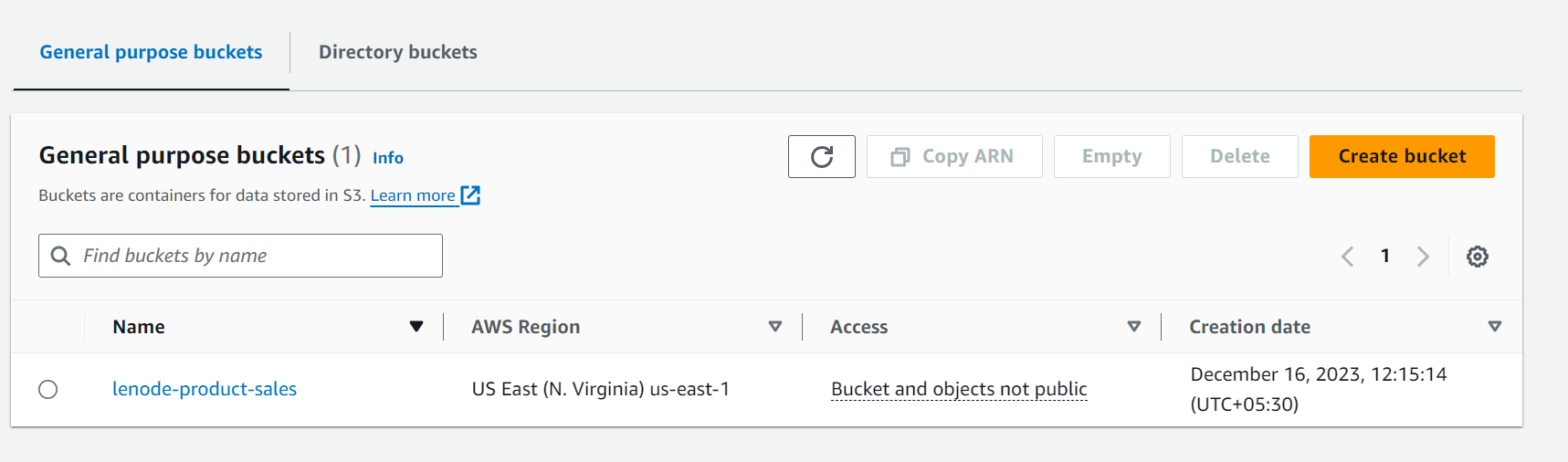
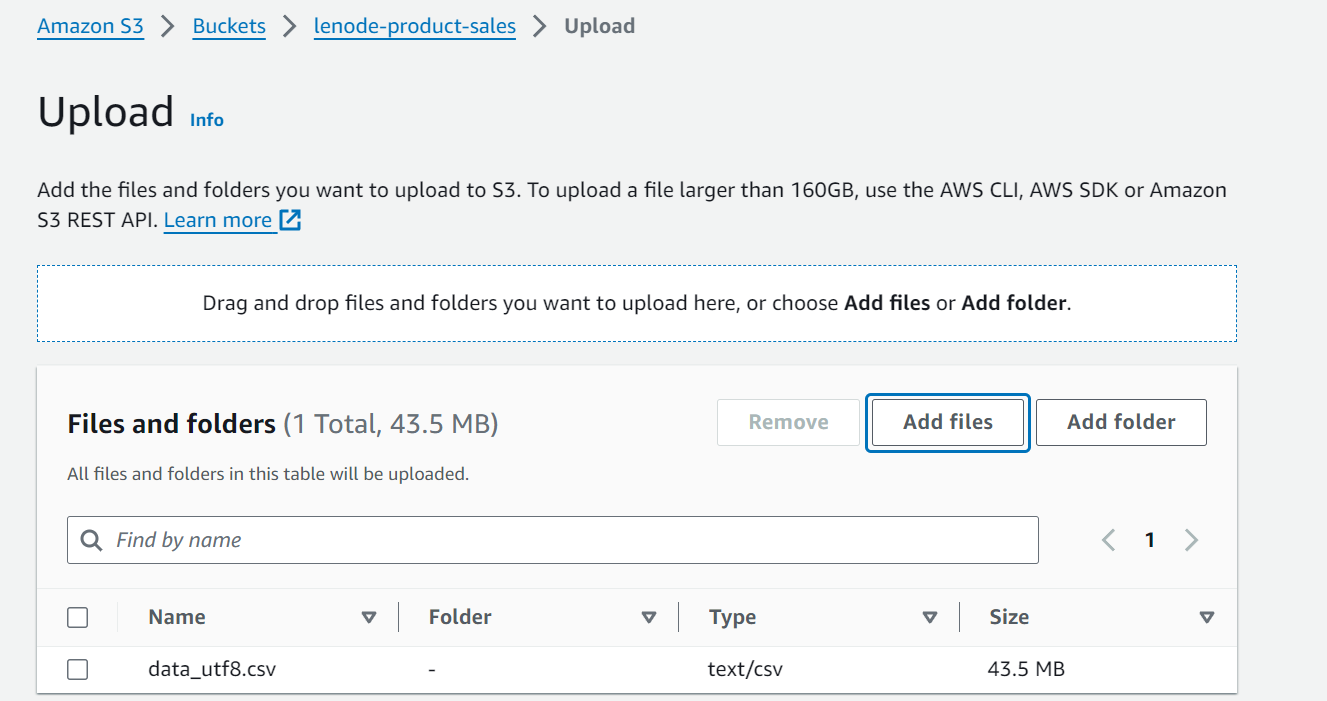
Lenodo Product Sales Analysis

# Step1: Create an S3 bucket with a unique name and upload the CSV file to the S3 bucket (ensure that the file is in UTF-8 format only)

* Create an S3 bucket.

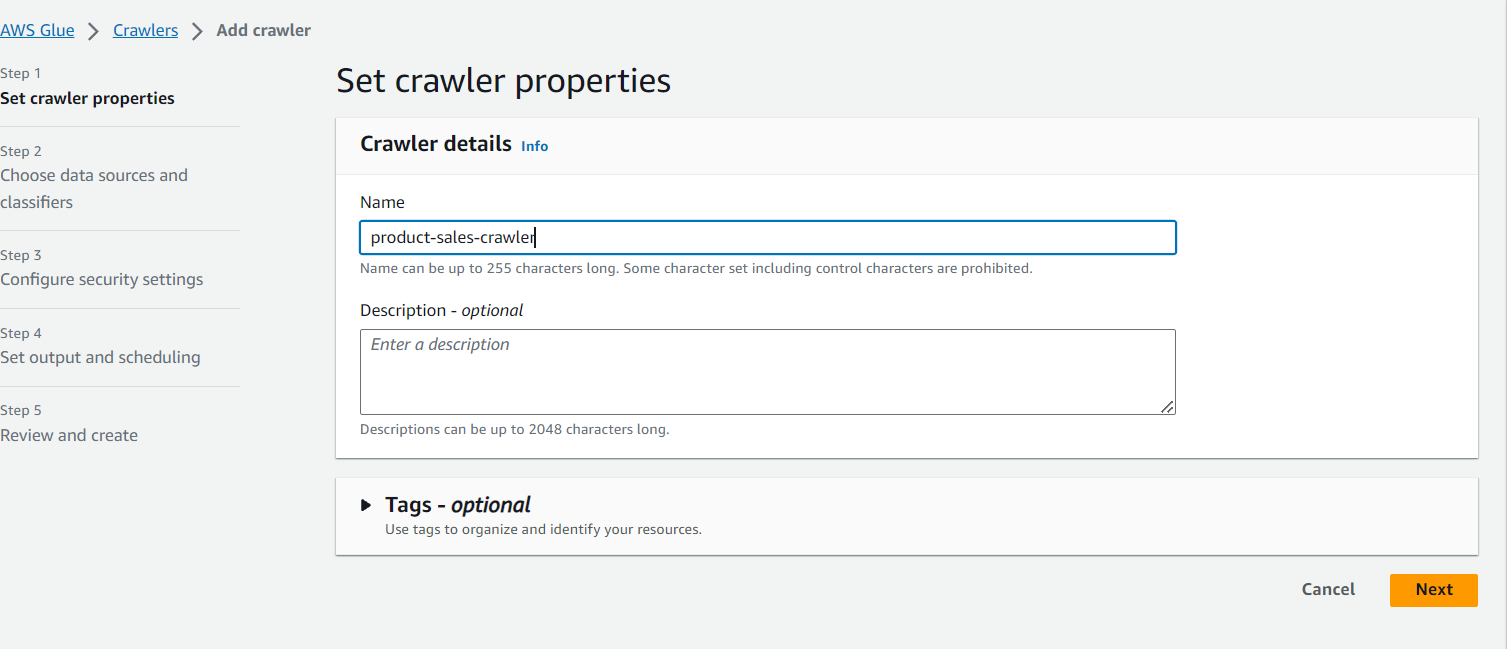


* Upload utf-8 data to created s3 bucket.

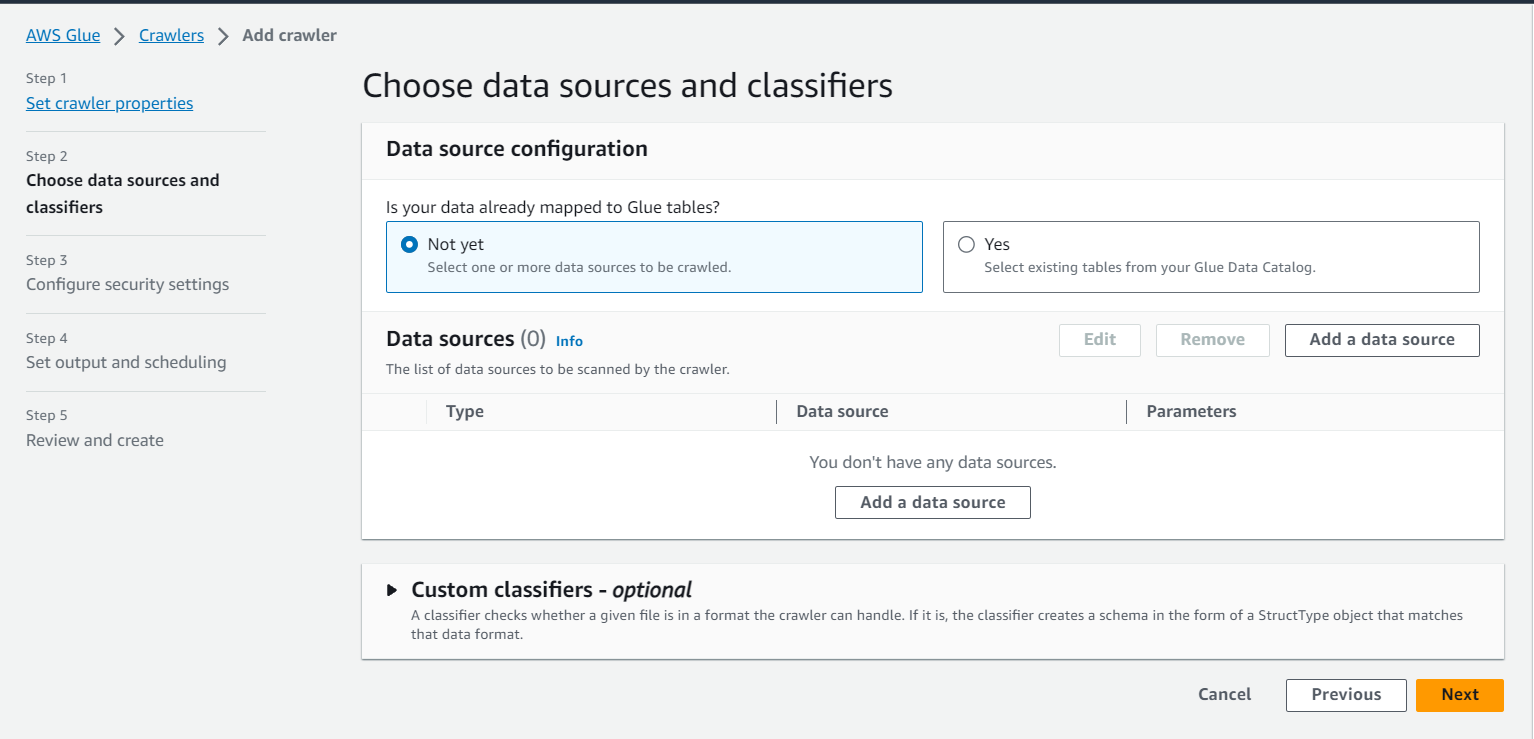


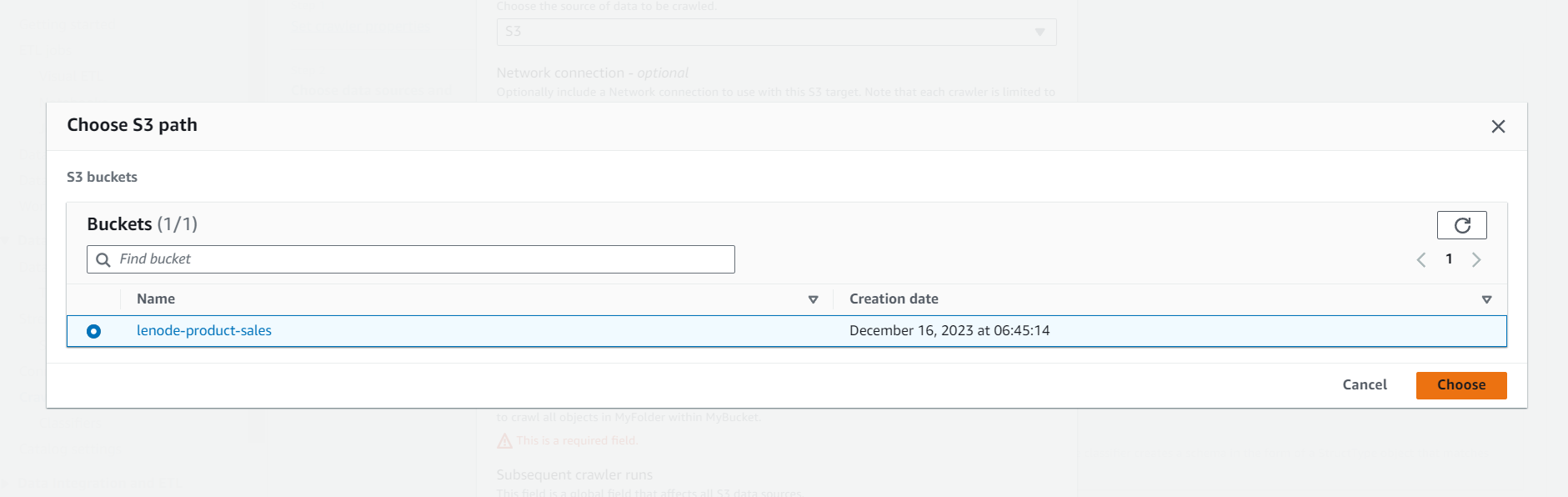
# Step 2: Create a crawler to crawl the CSV data and generate a metadata catalog.

* Provide name for crawler.

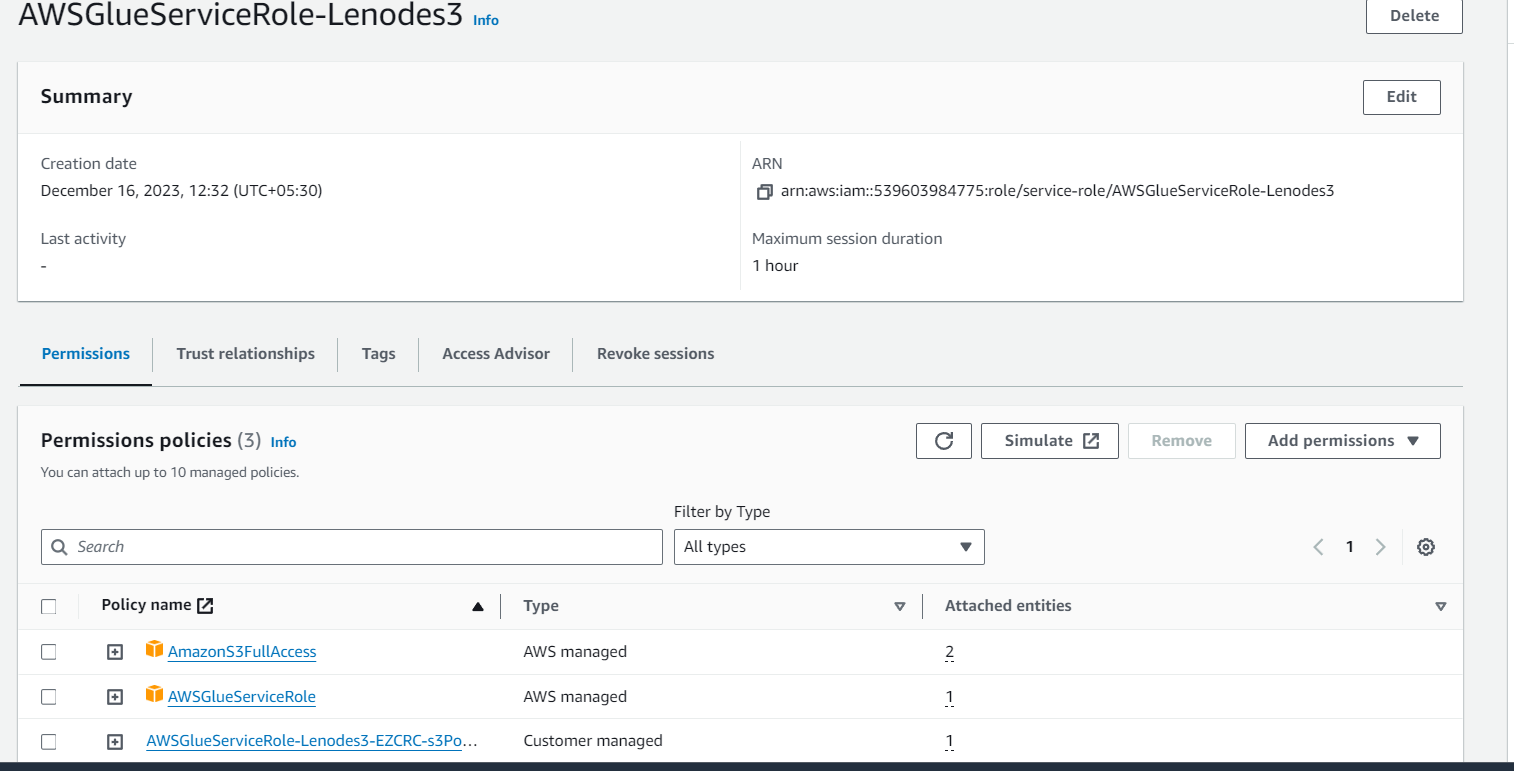


* 2.Adding datasource from s3 bucket with sales data.

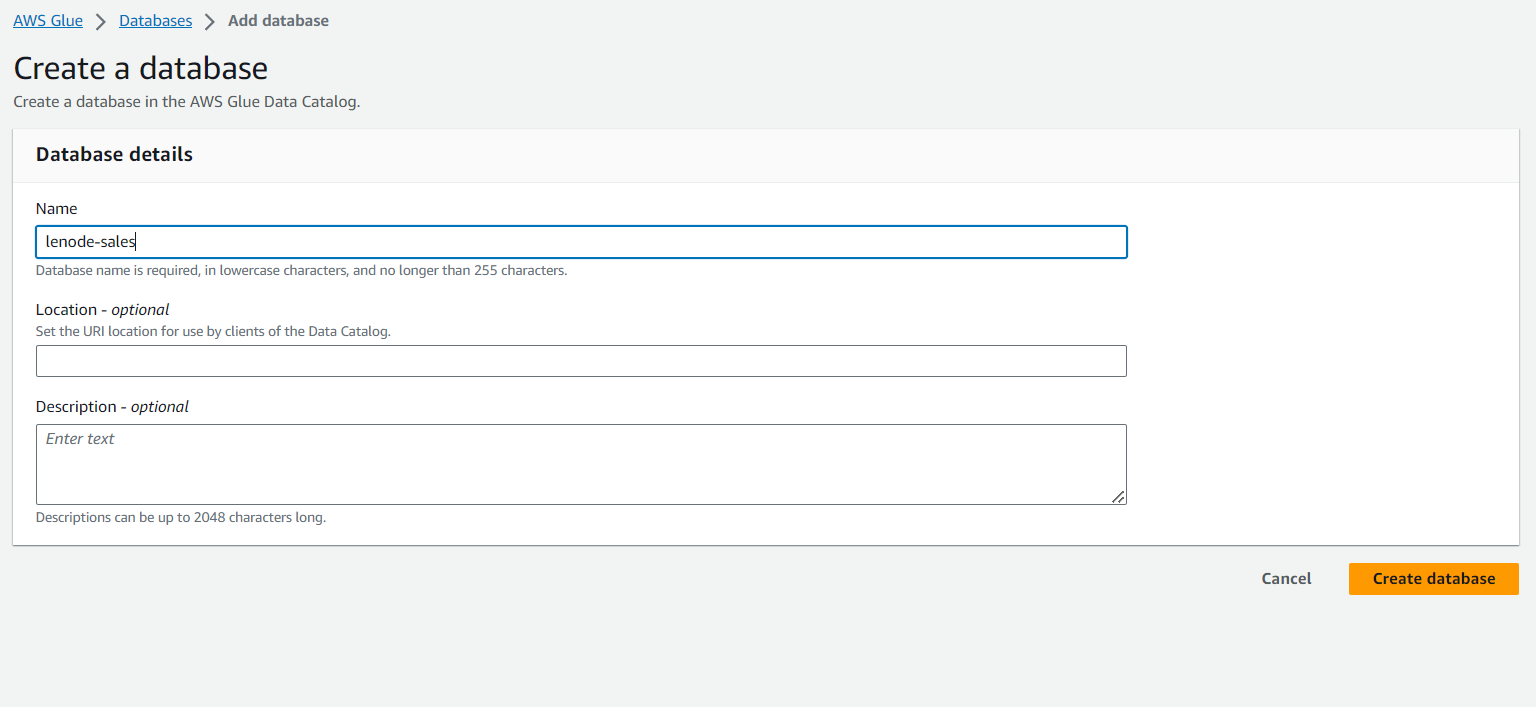


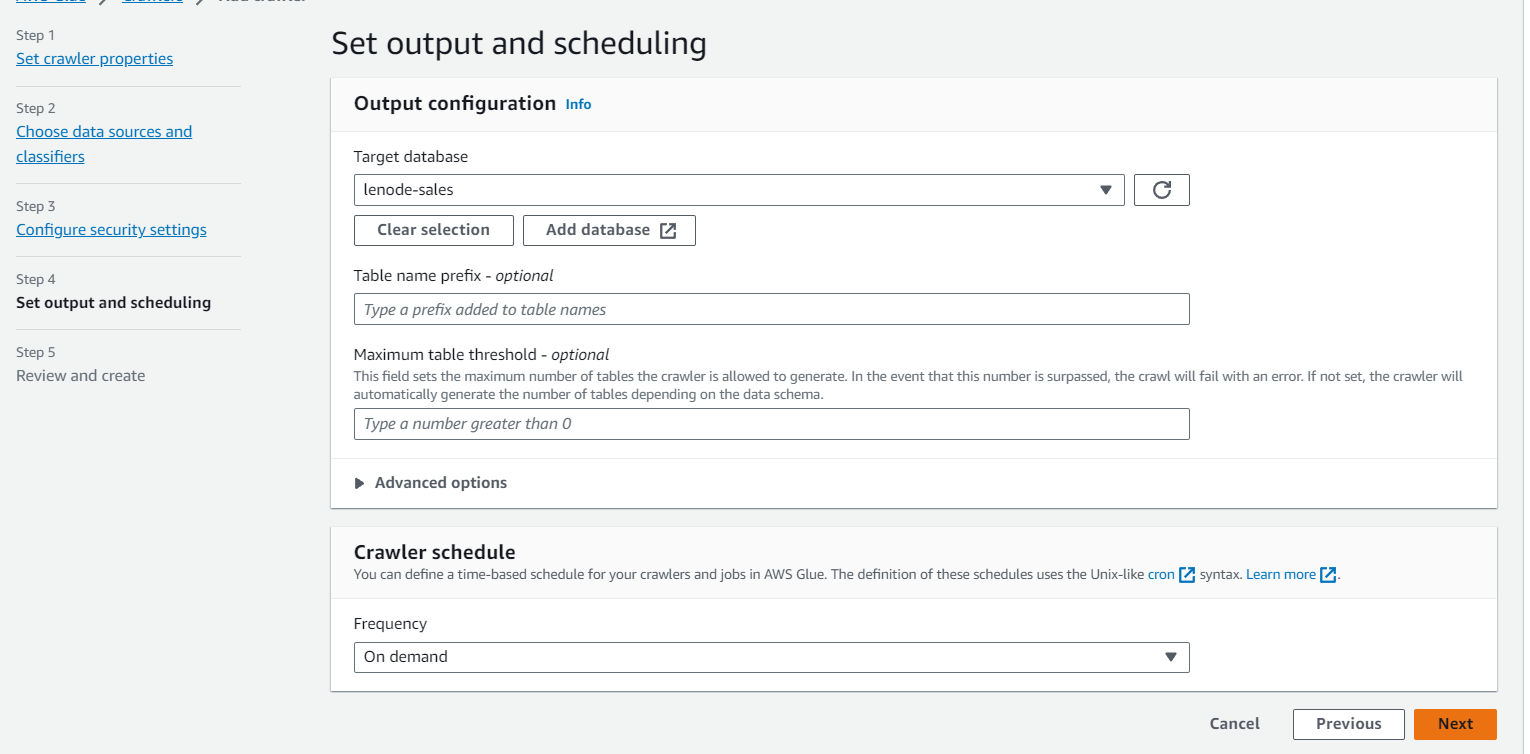


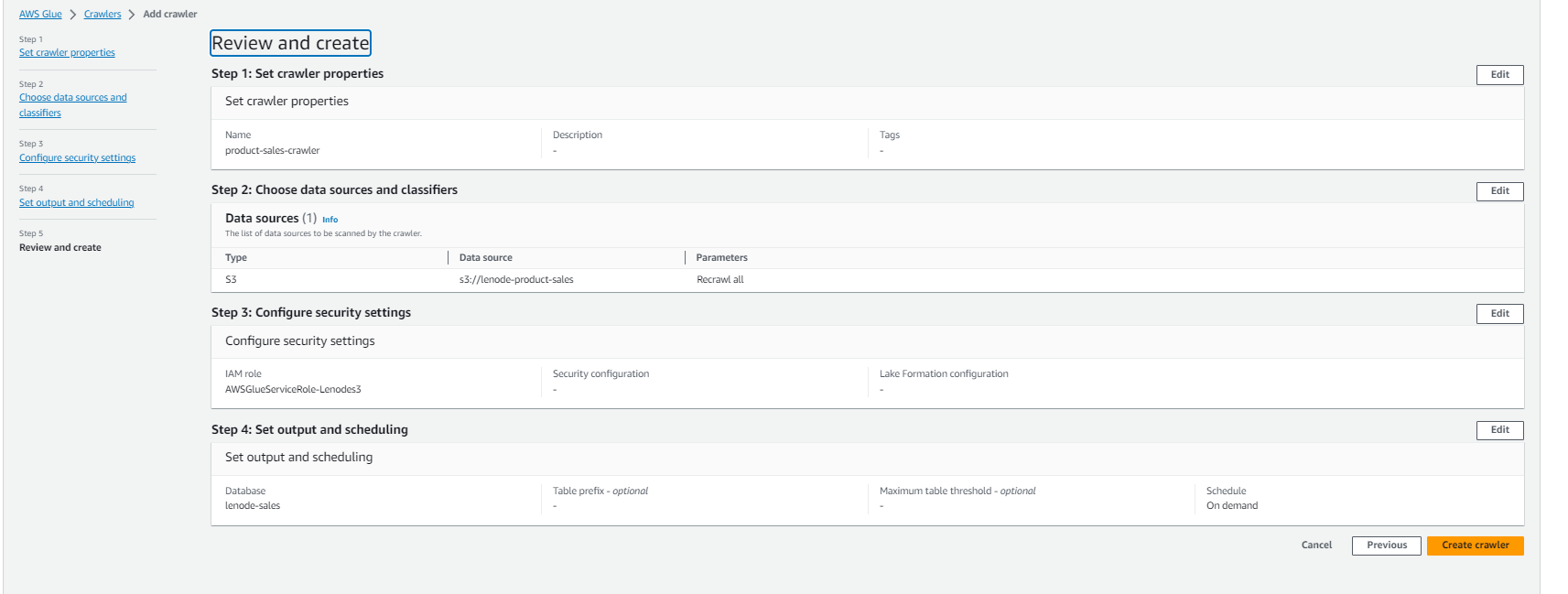
* Attaching role with required permissions



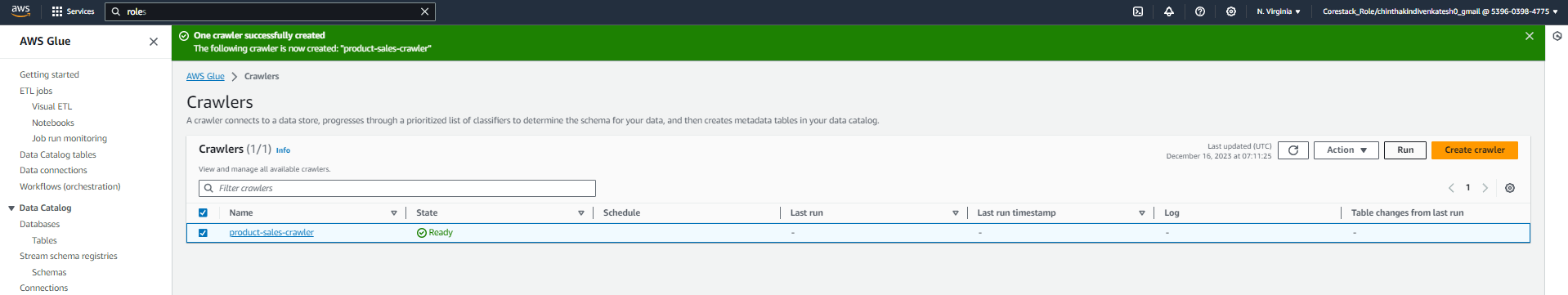
* creating a database to load the s3 data and to crawl

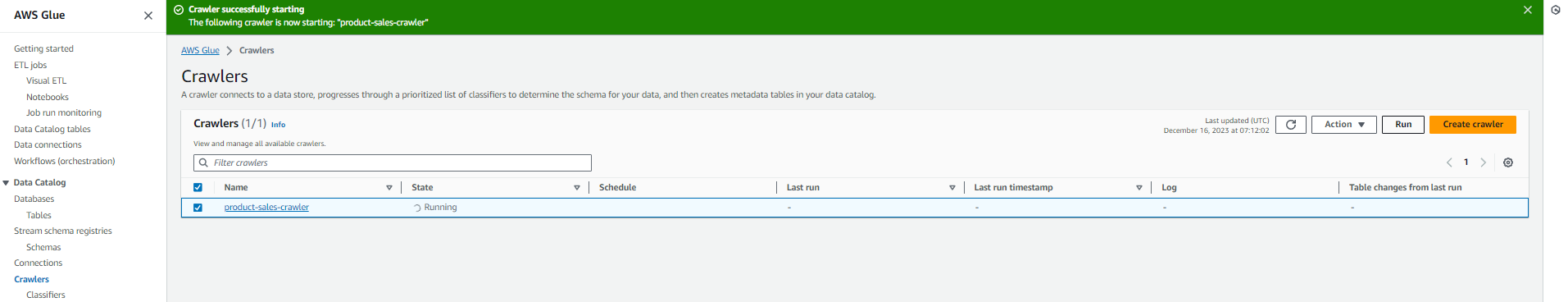




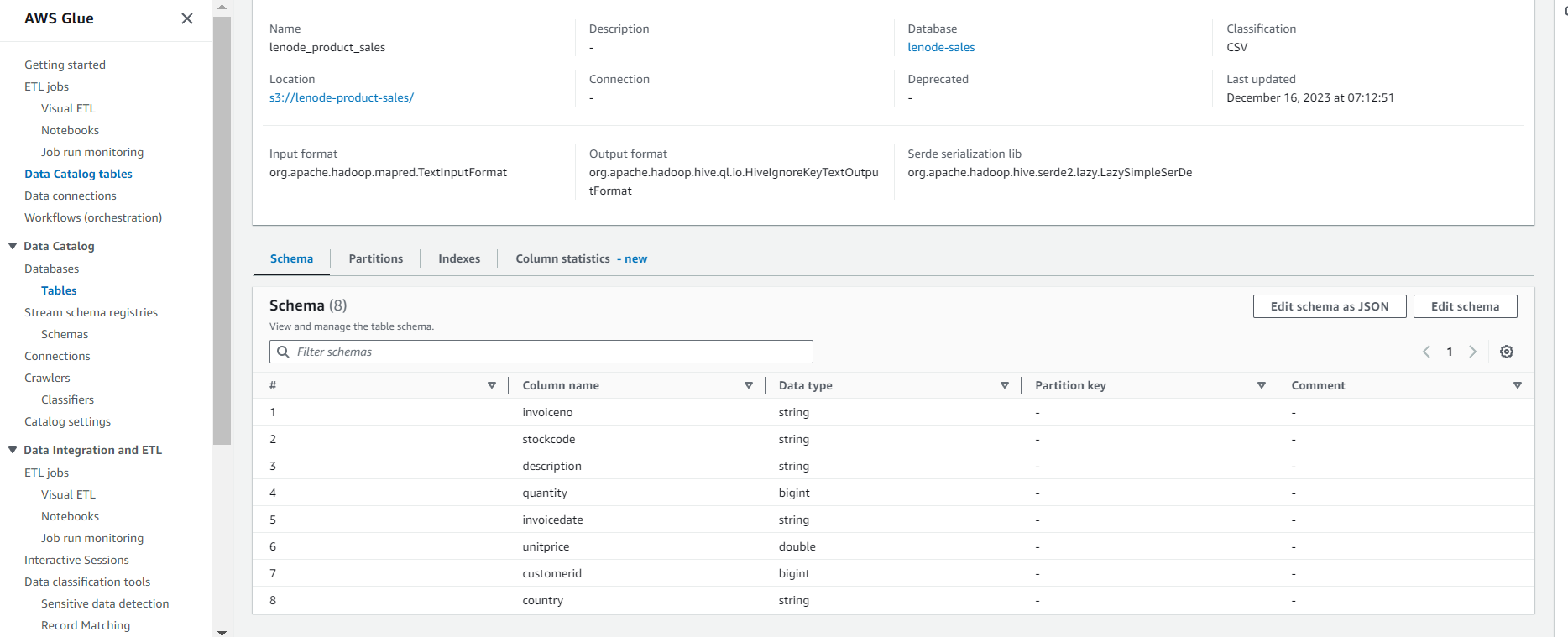


* Running the crawler for the sales data in s3



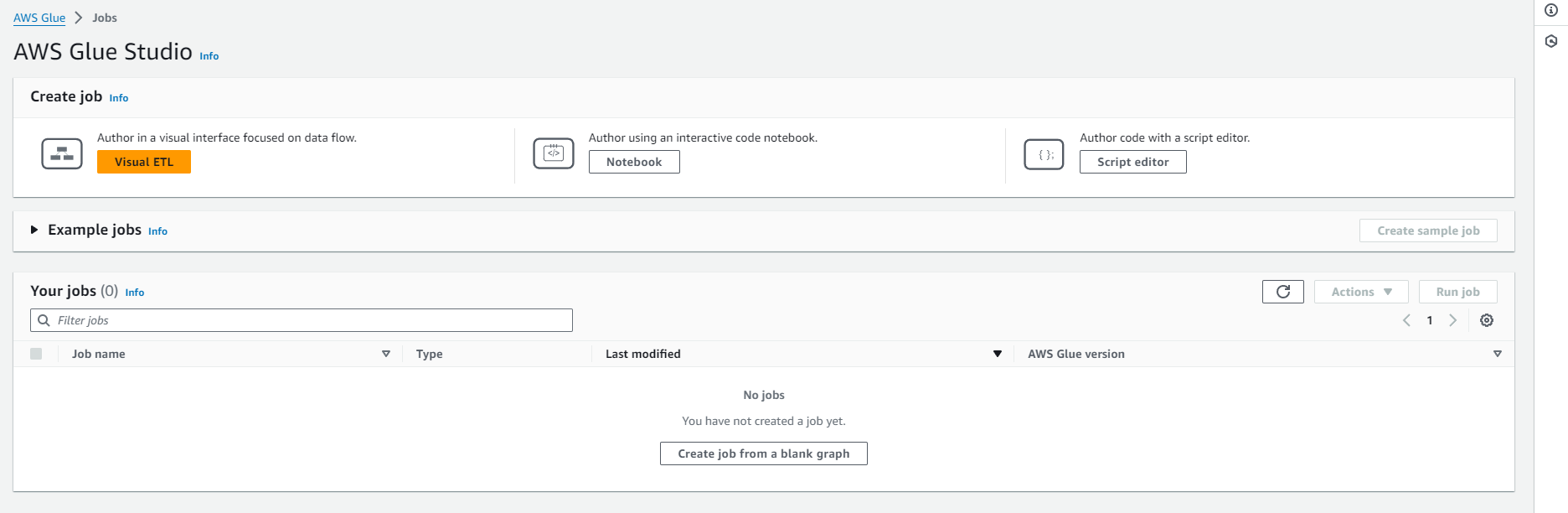


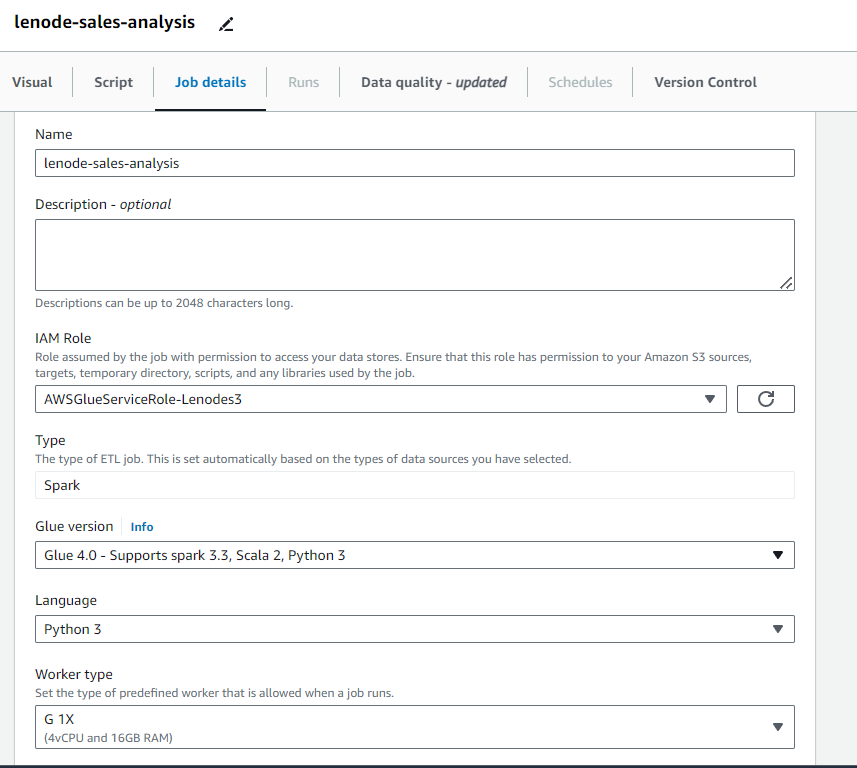
* This is the catalog generated after running crawler



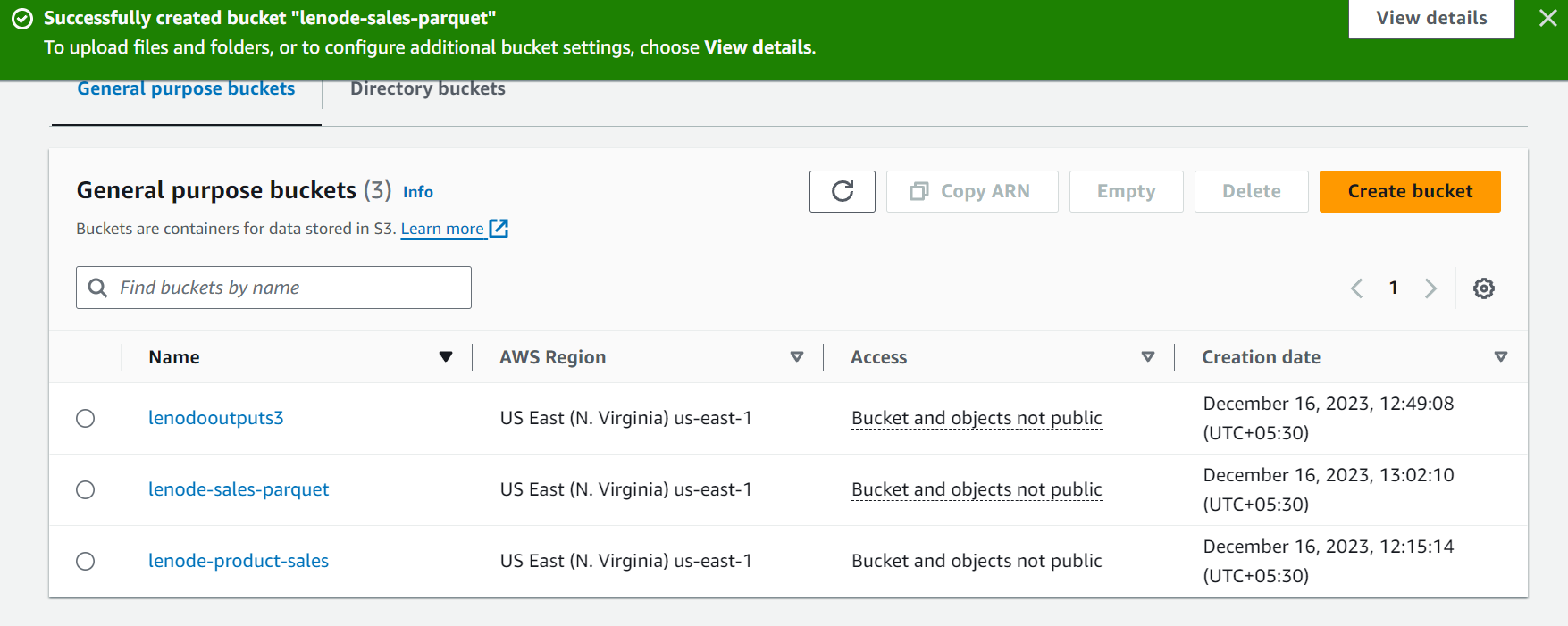
# Step 3: Create a Glue job to transform the data into the Parquet format as CSV is not optimal for data warehouse queries

* Setting up configuration to run Glue job.

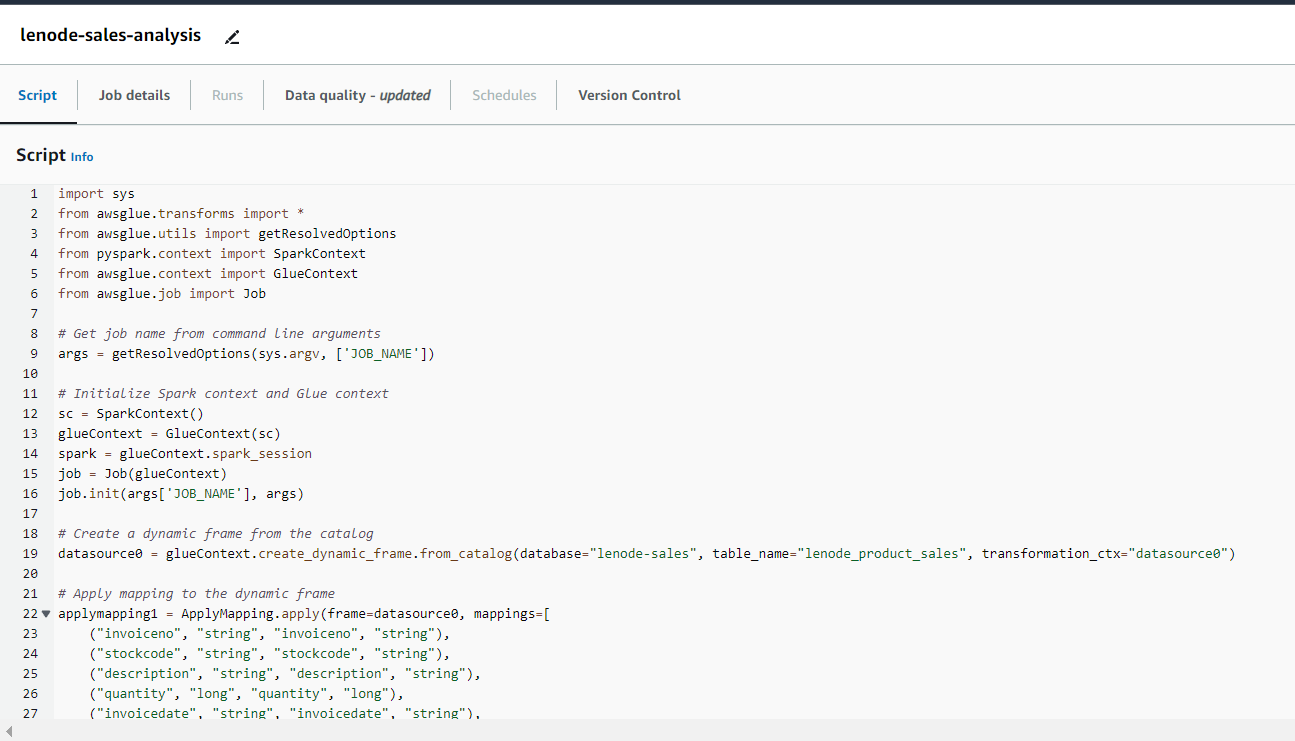


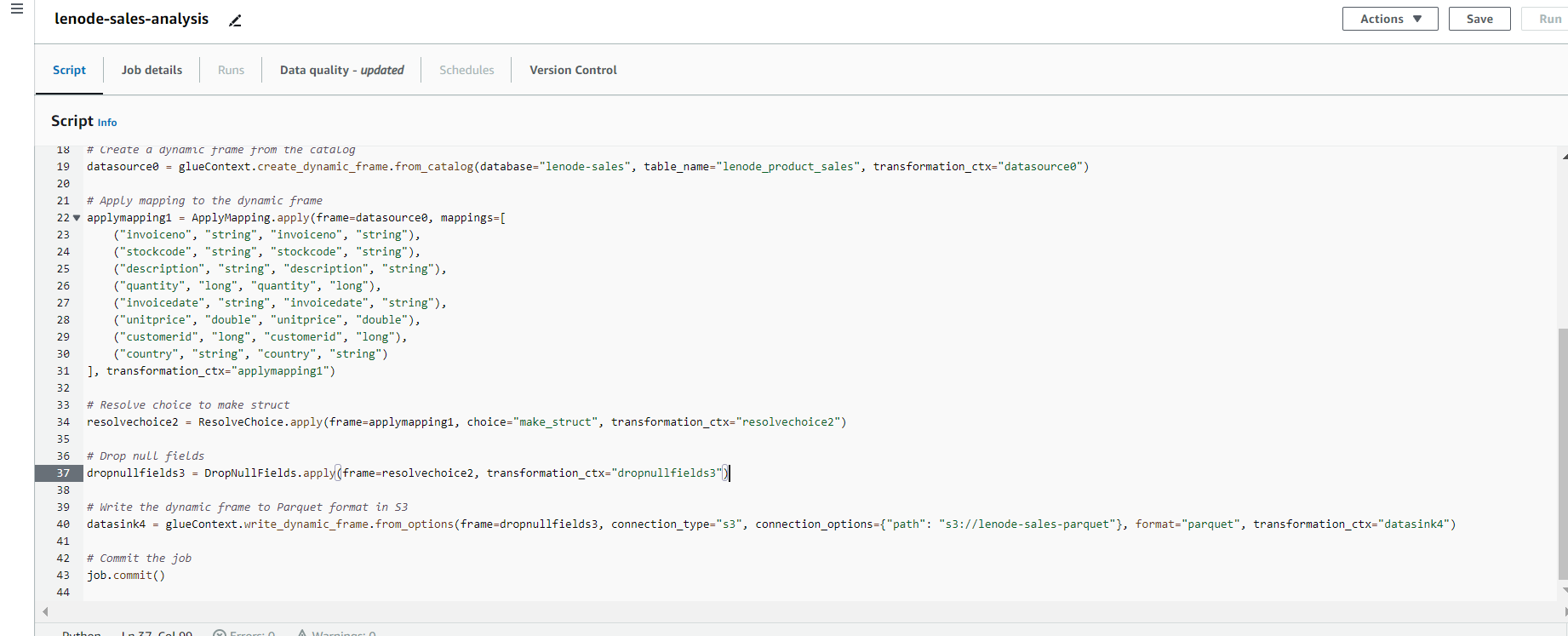


* creating s3 bucket to store output parquet files data

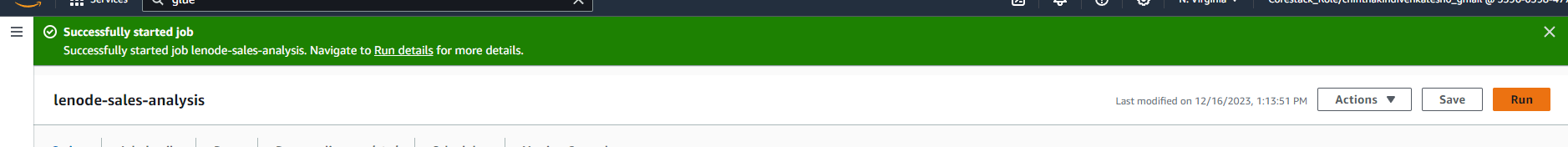


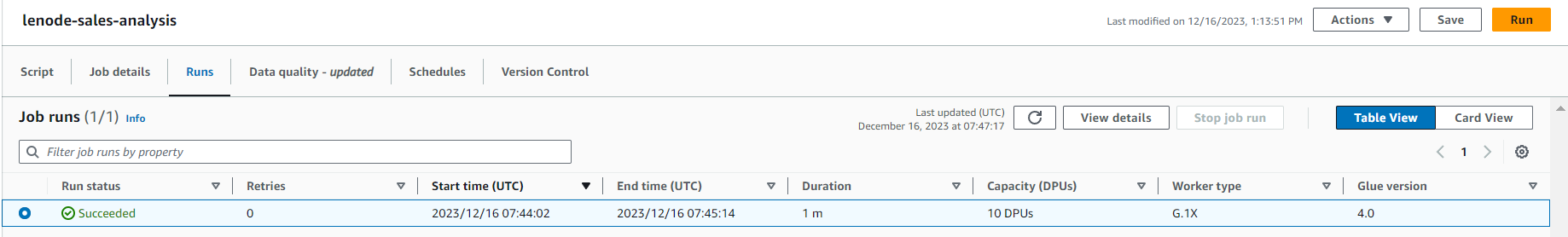
* Script used for transformation of data to parquet data from s3



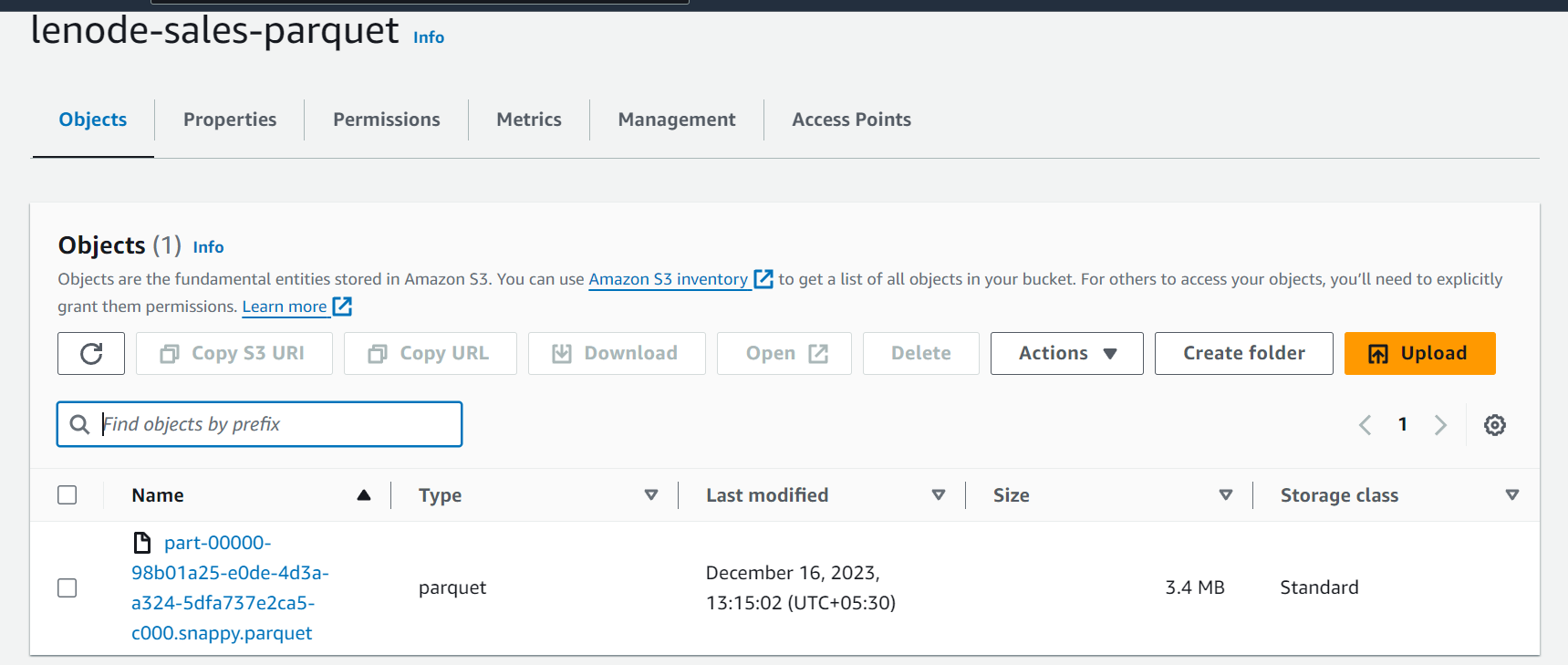


* Running the glue job



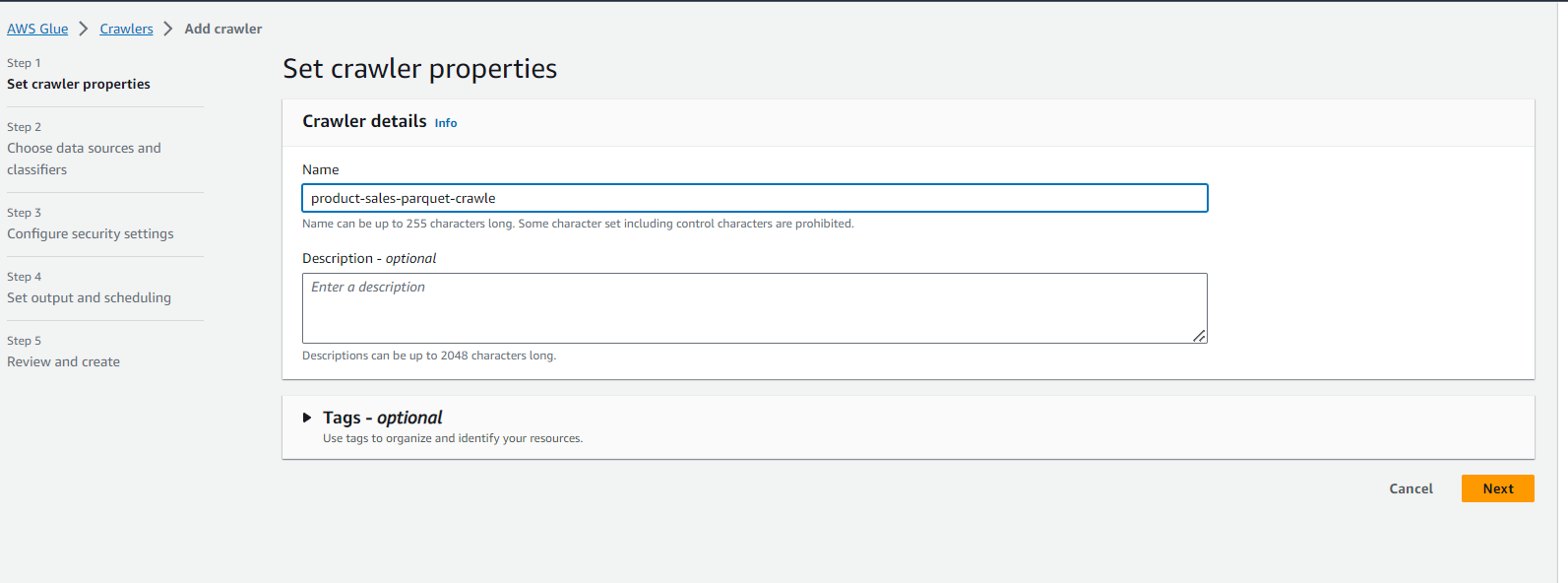


* Ouput parquet file in the target s3 bucket

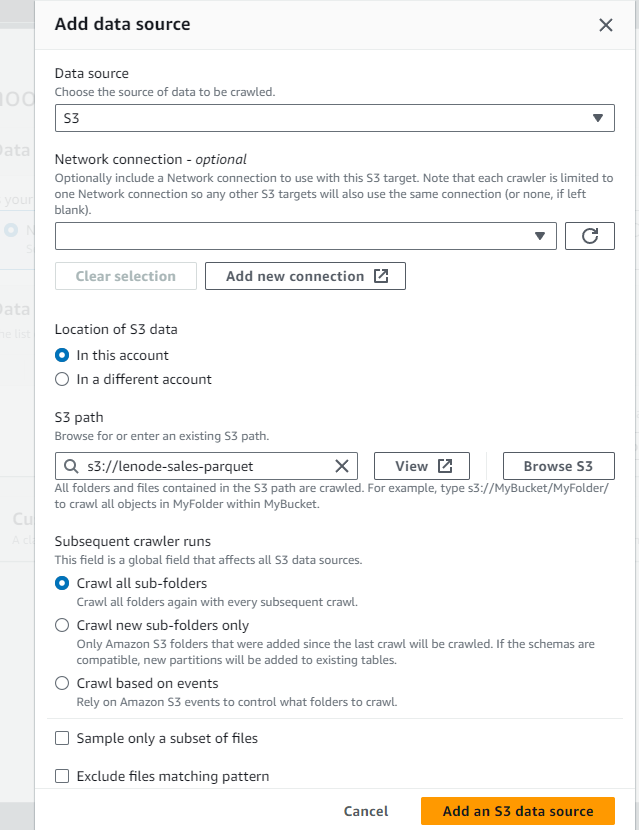


# Step 4: Add another crawler to crawl the Parquet data files to generate the metadata catalog of the Parquet file in order to query it with Athena

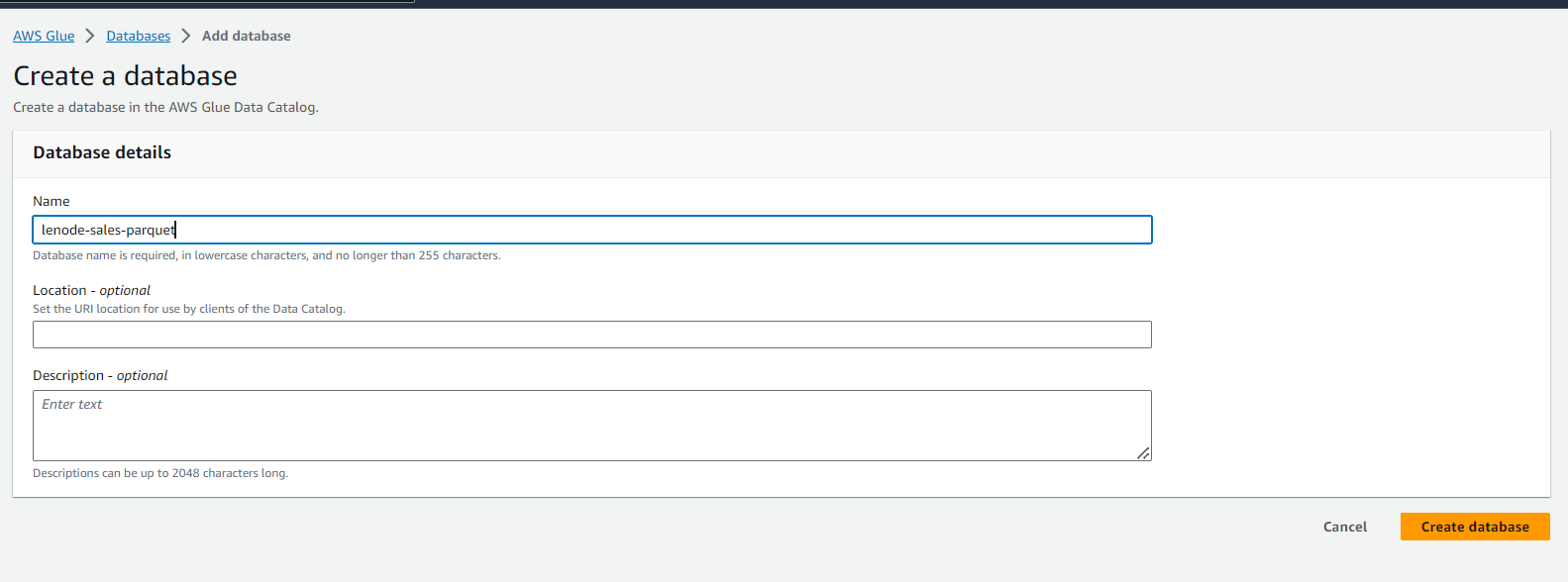
* Providing name for parquet data crawler.

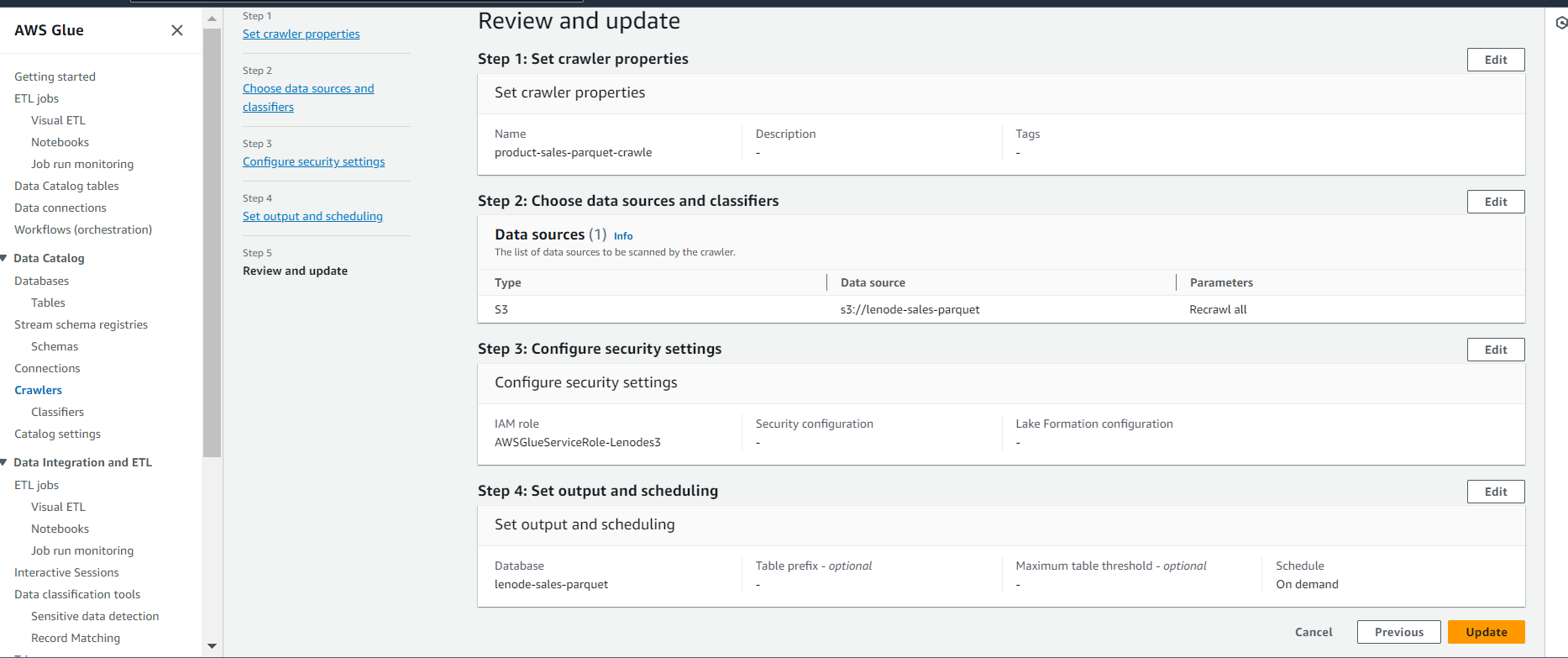


* Adding S3 bucket of parquet files as source for the crawler.

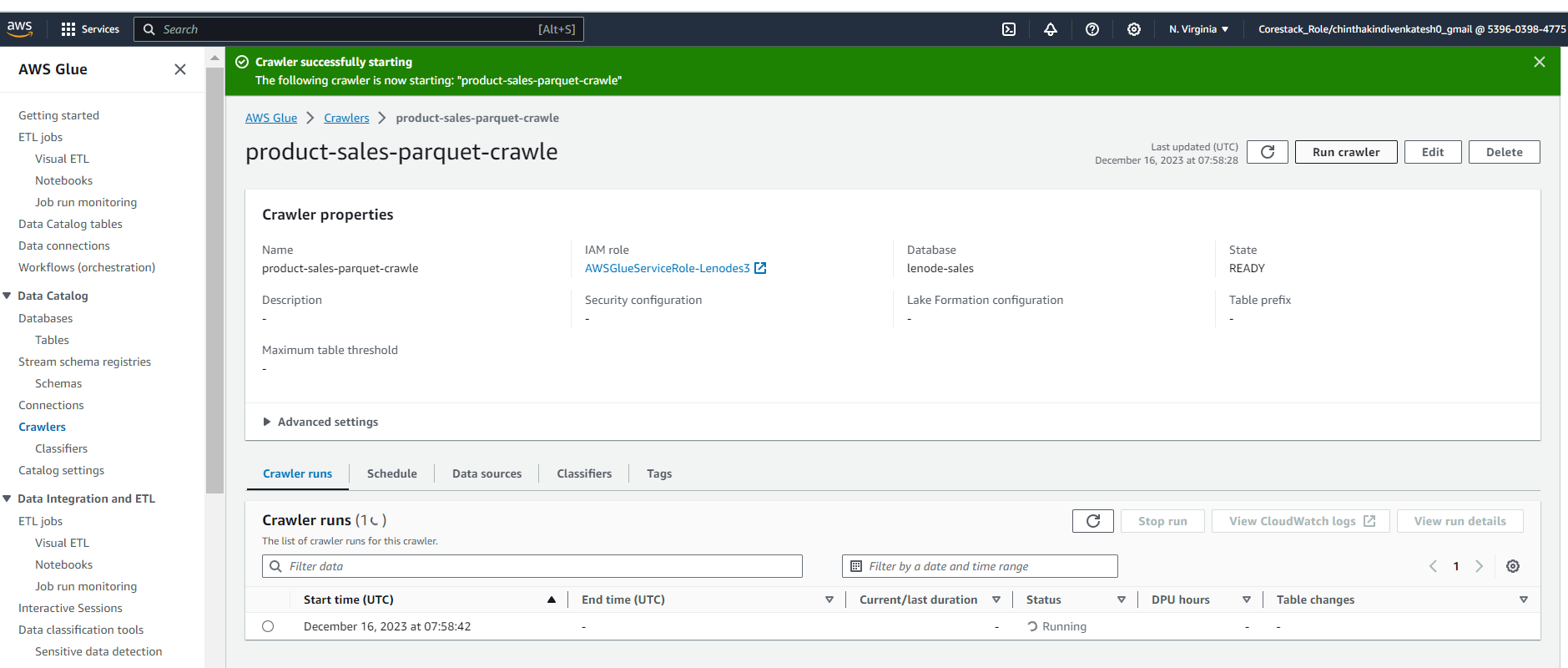


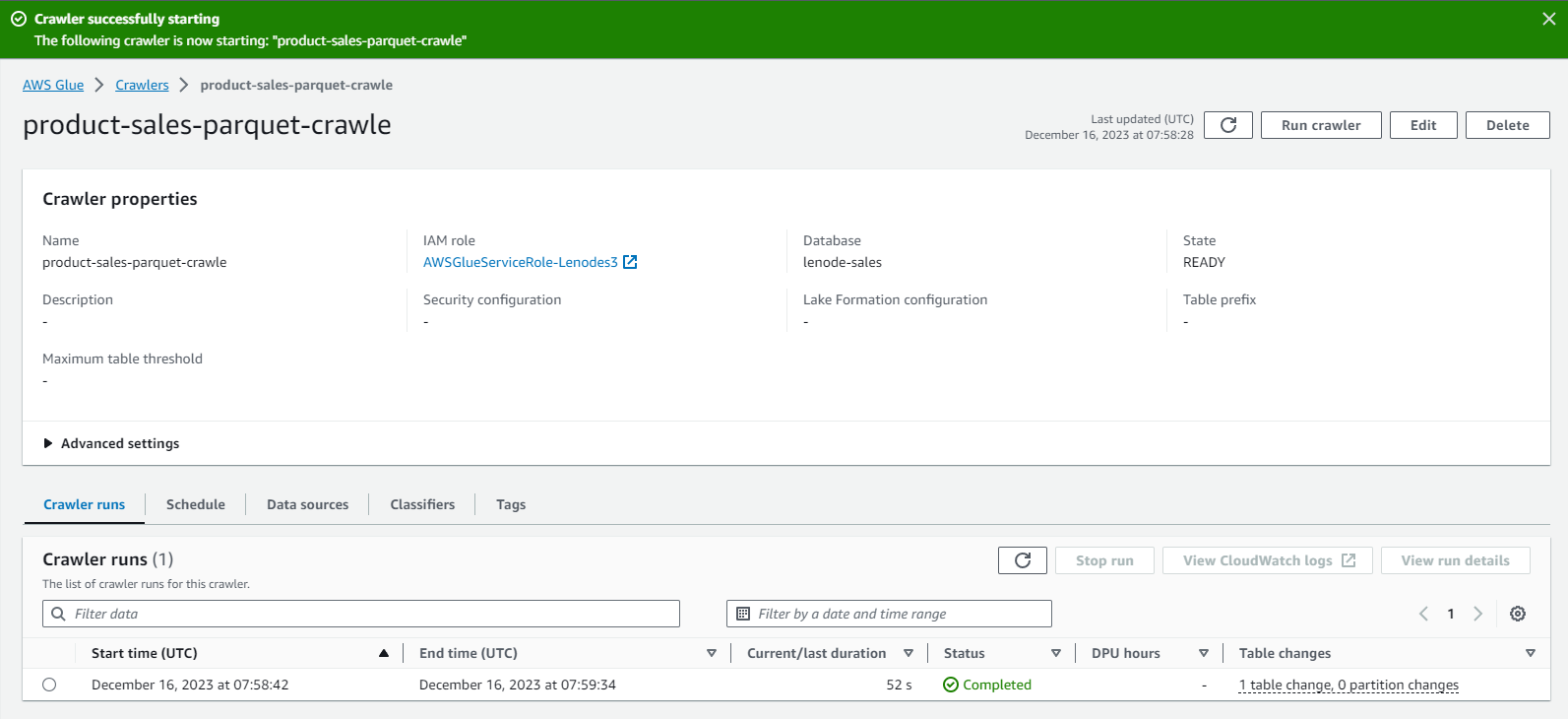
* Create database for this parquet data



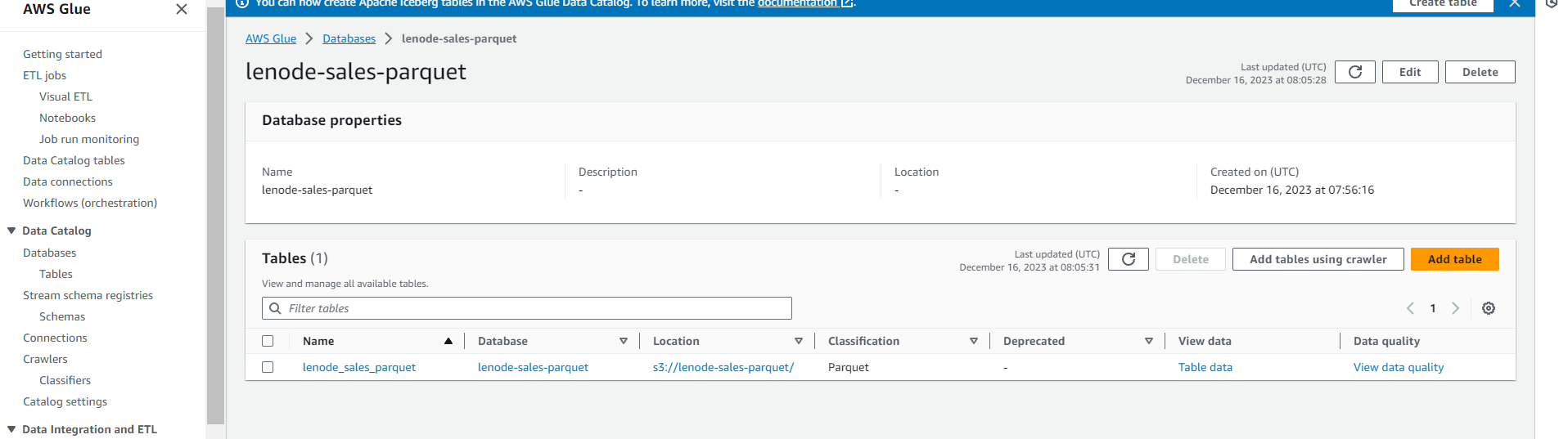


* Run the crawler

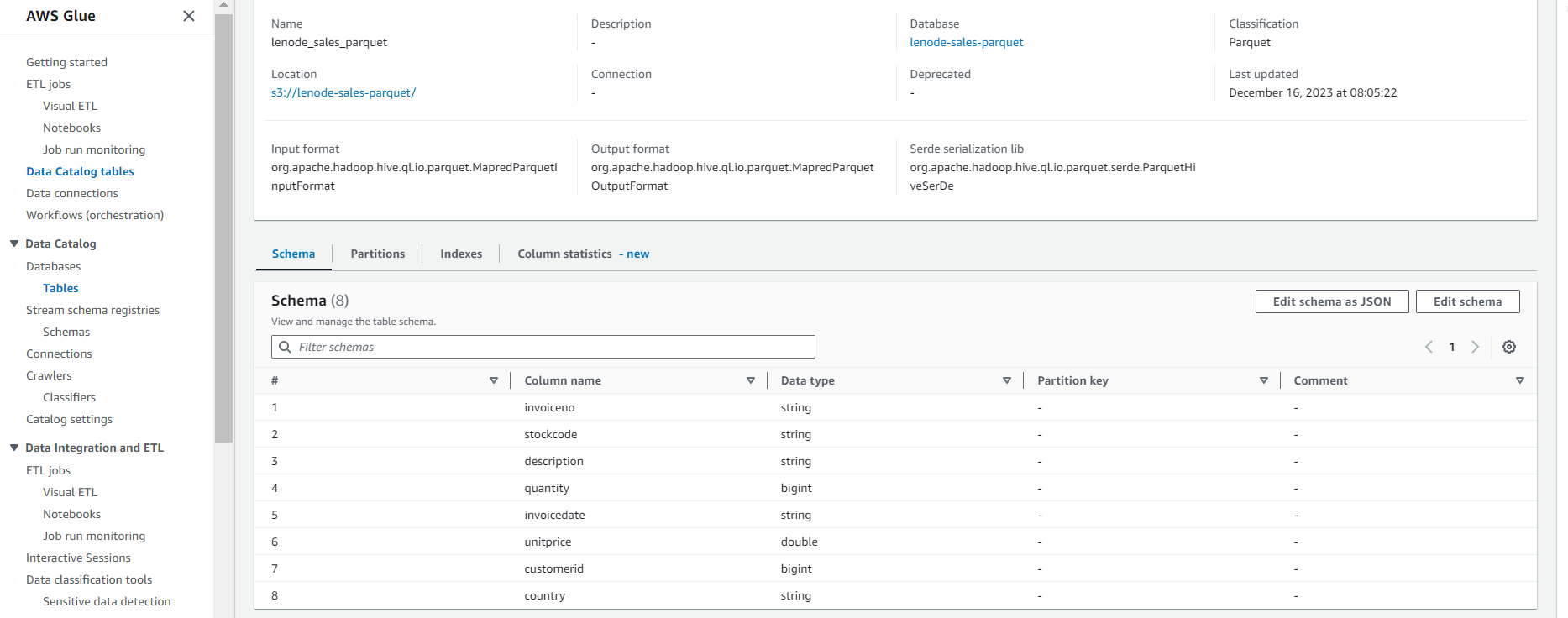




* Ouput data in Database

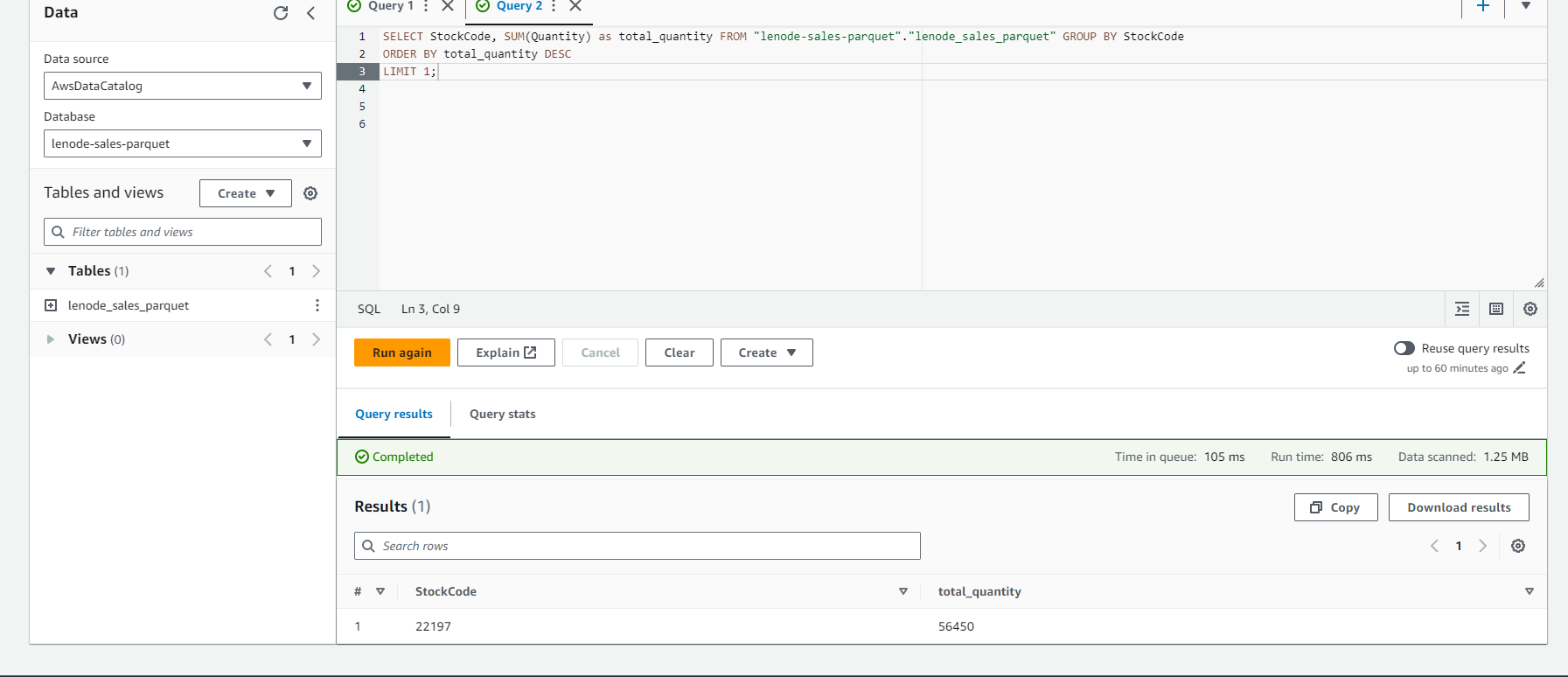


* This is the catalog generated after running crawler



# Step 5: Query the data to identify the best-selling item and countries where customers have bought the most-sold item using Athena

* To identify the best-selling item from sales:



* To find countries where customers bought the most-sold item (assuming 'StockCode' identifies the item):

