**Hands-on #12: AWS Services Lambda & Glue & Athena : Serverless ETL Pipeline**

**SRIJA BANDA**

**801393582**

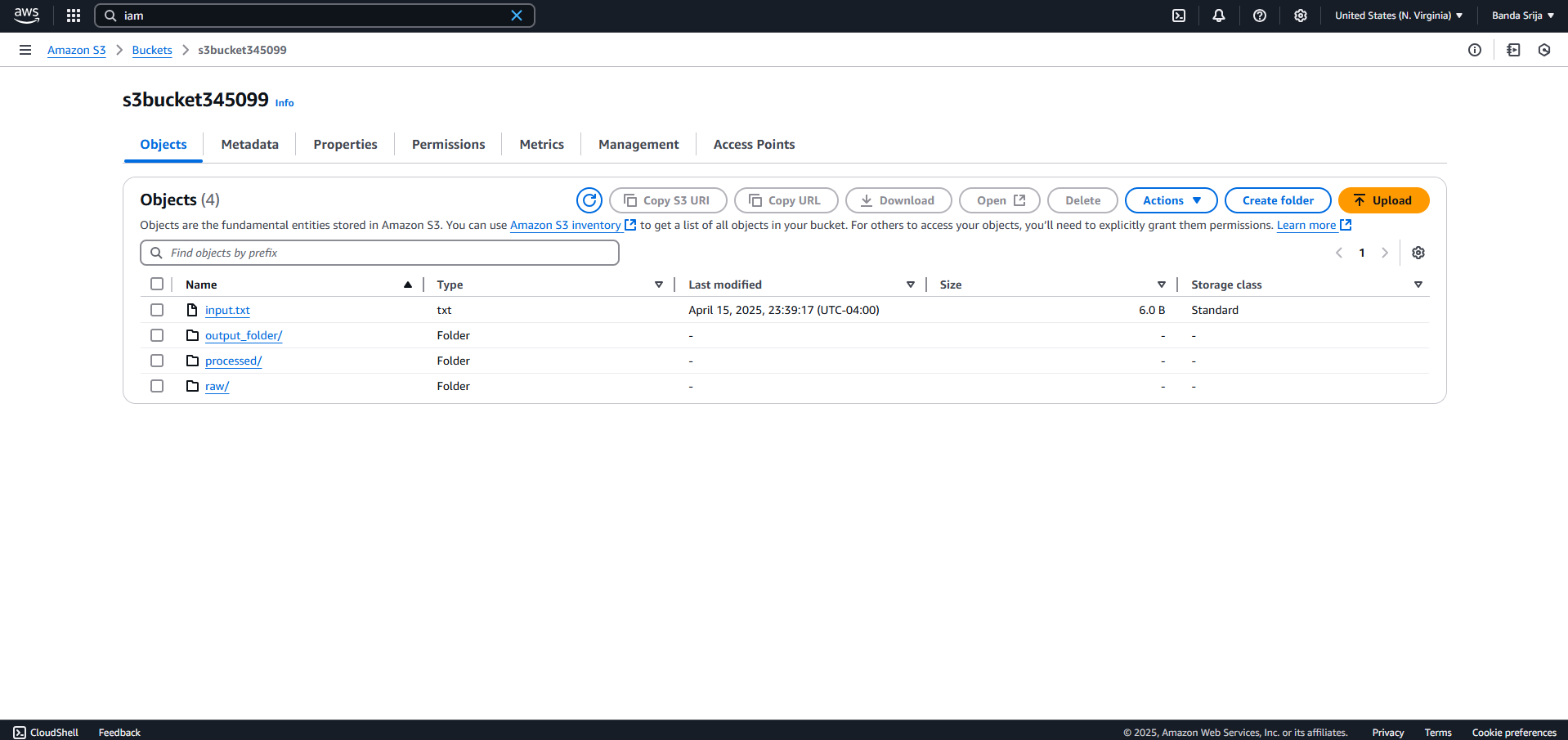
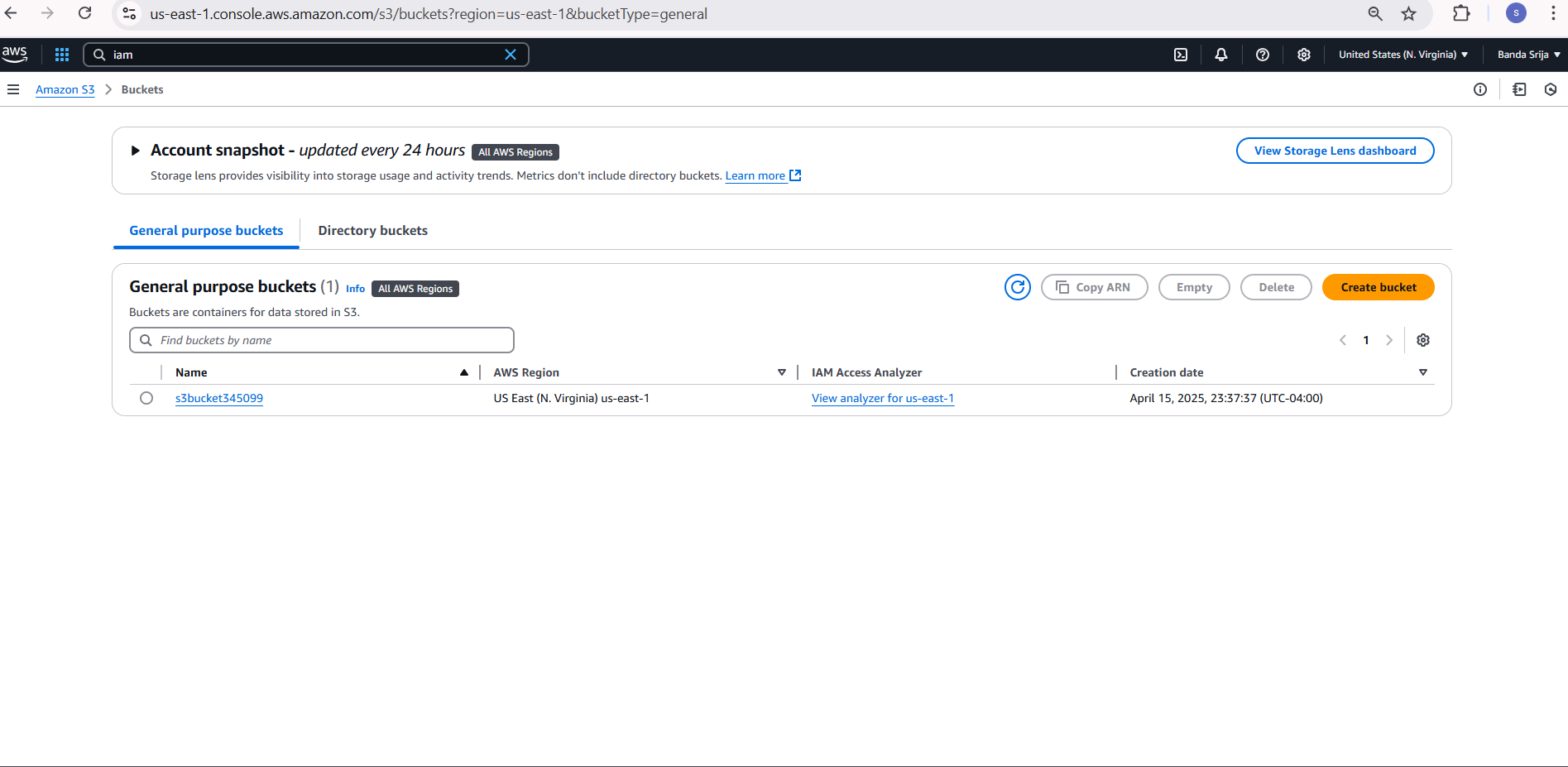
**GITHUB LINK-**

[**https://github.com/Srijabanda/ServerlessETLPipeline.git**](https://github.com/Srijabanda/ServerlessETLPipeline.git)

**Step‑by‑step setup**

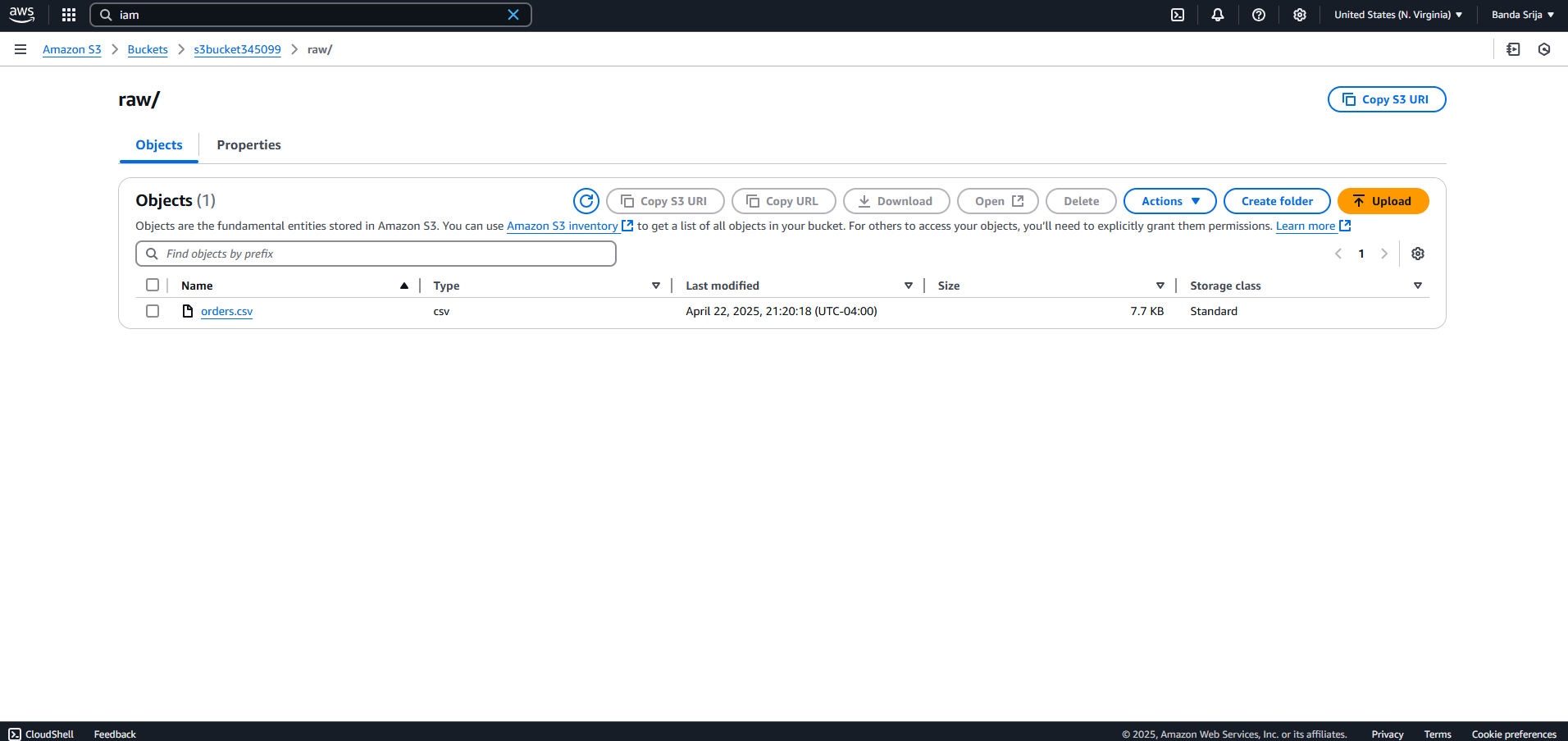
**1. Create the bucket and folders**

– Create an S3 bucket, for example online-orders-pipeline.  
– Inside it, add two empty folders named raw/ and processed/.



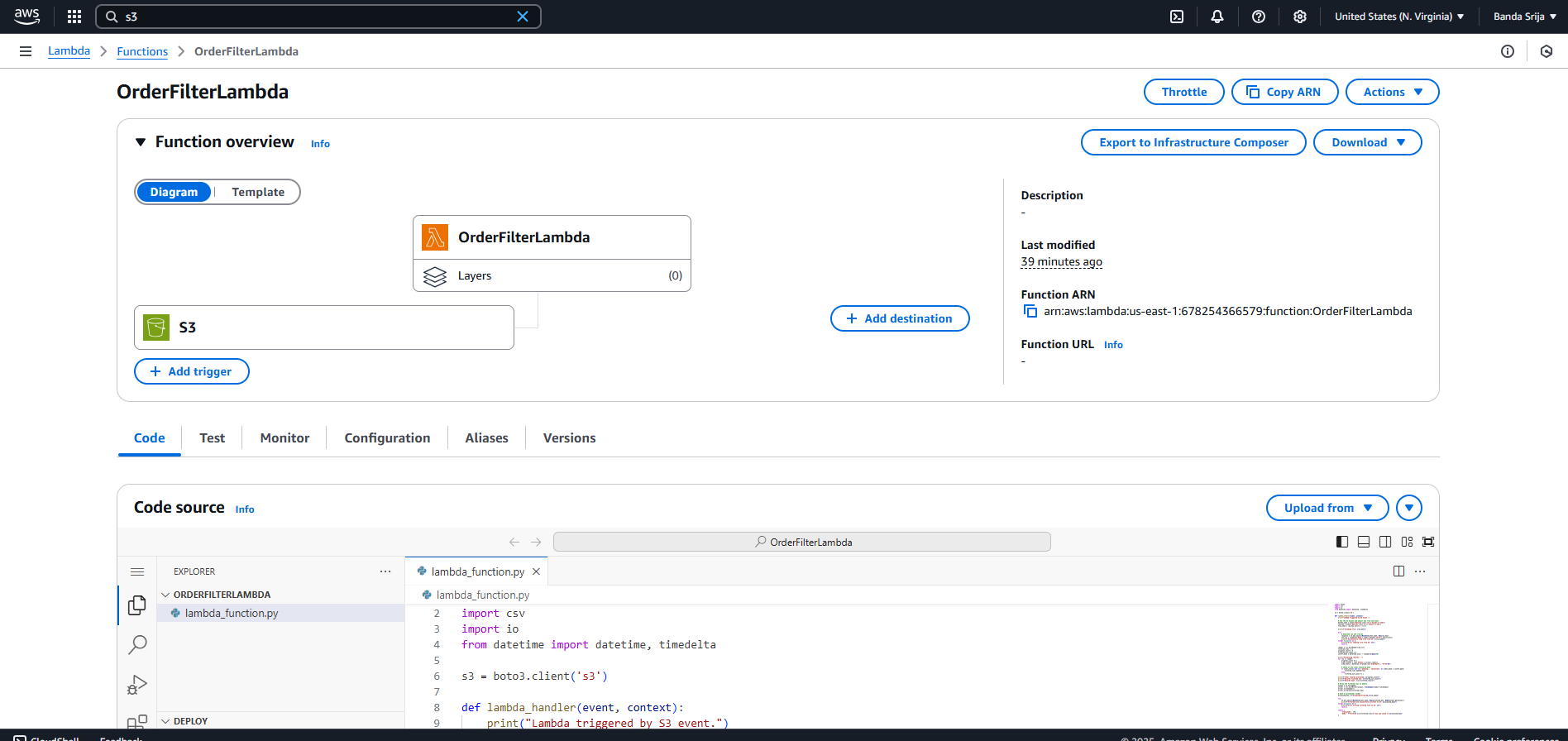
**2. Generate sample data locally**

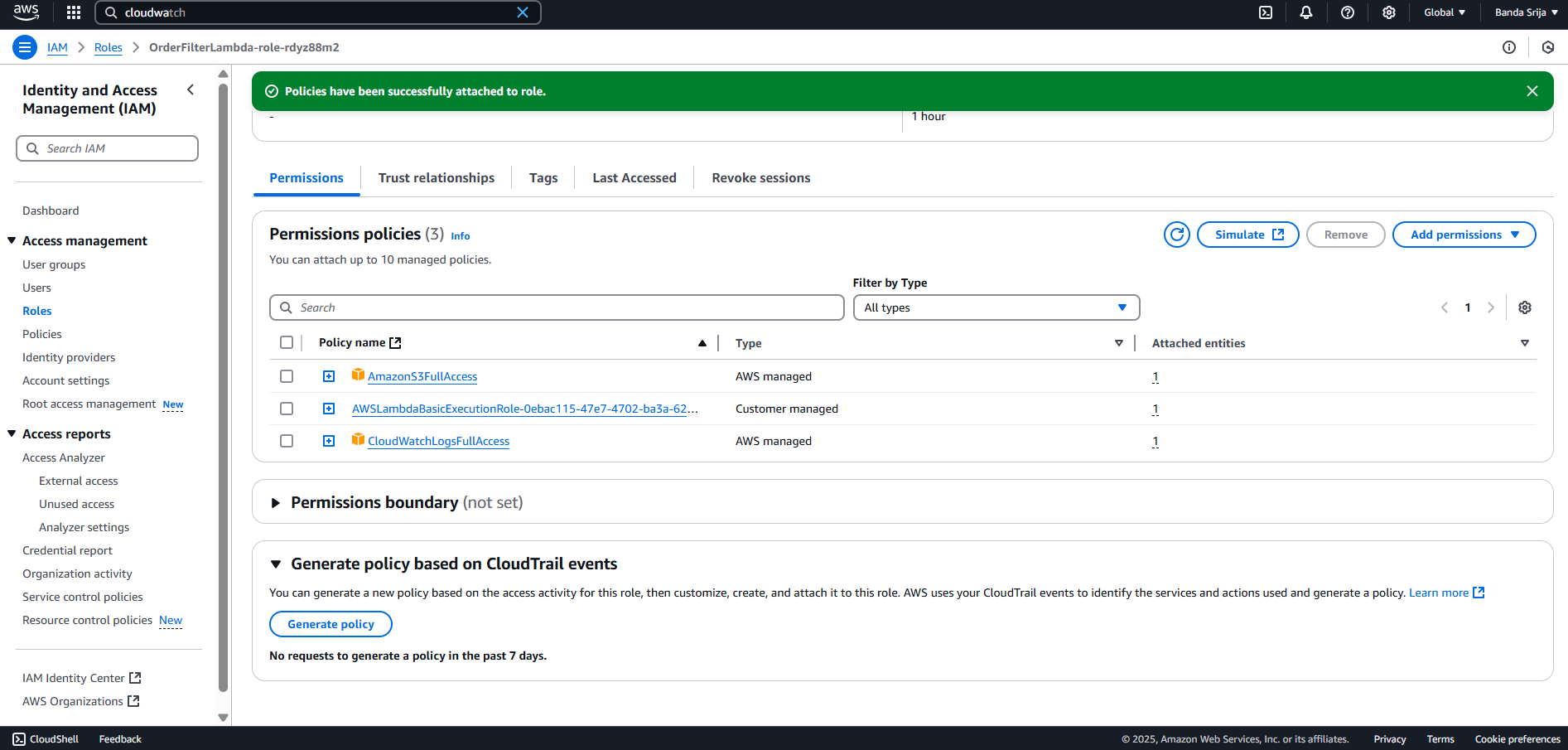
Run the following Python script; it creates orders.csv with 100 random rows.



**3. Create the Lambda function**

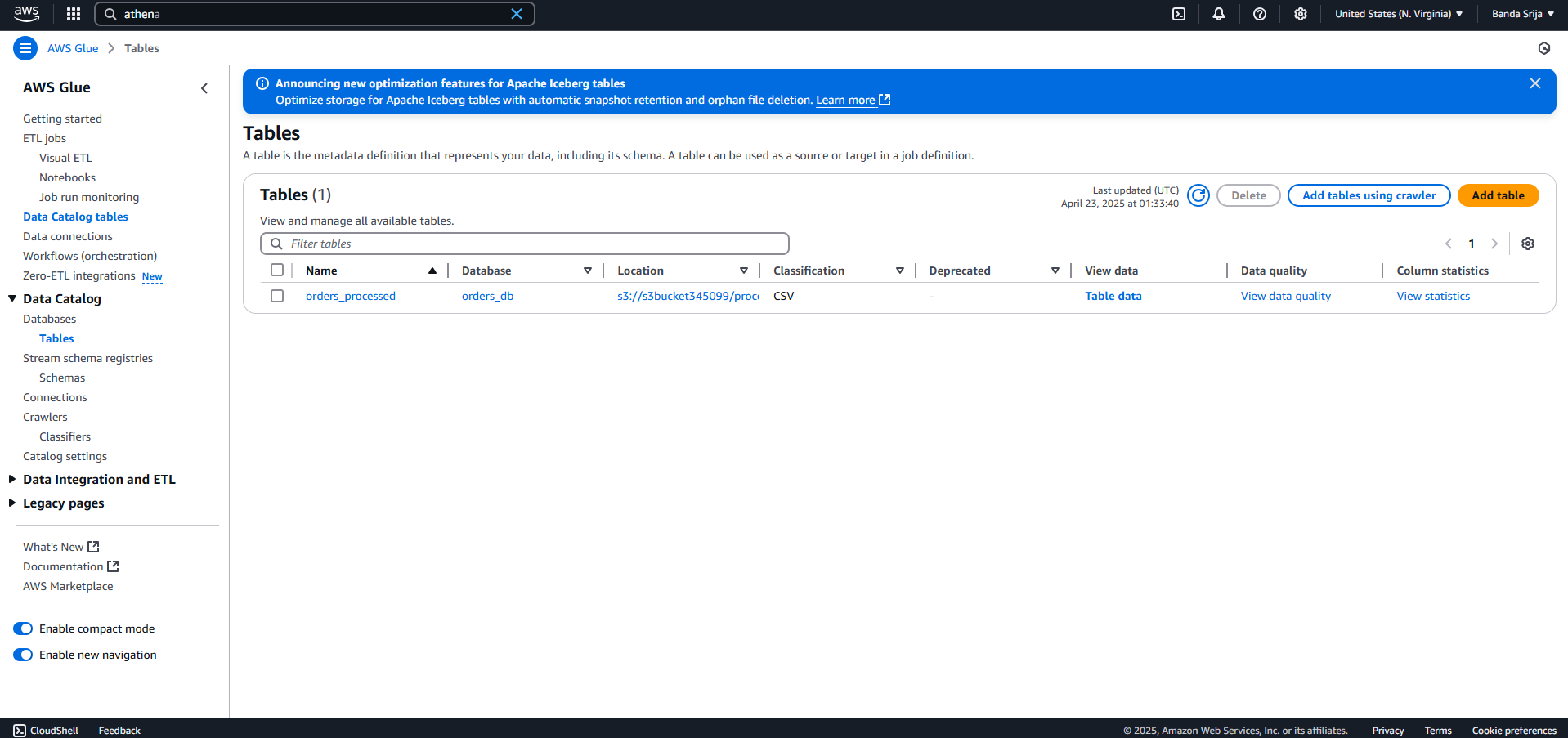
Handler (Python 3.9 or newer)

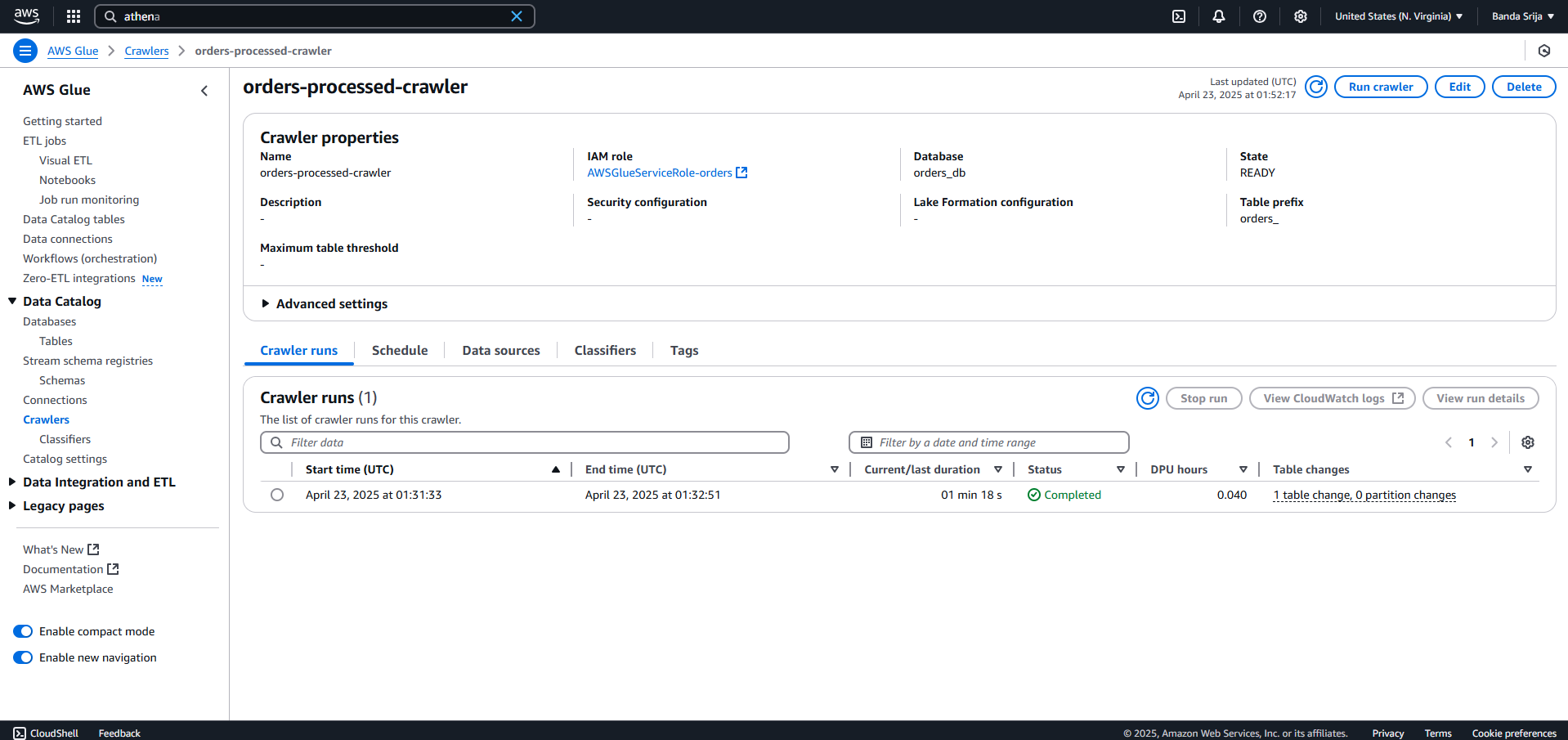
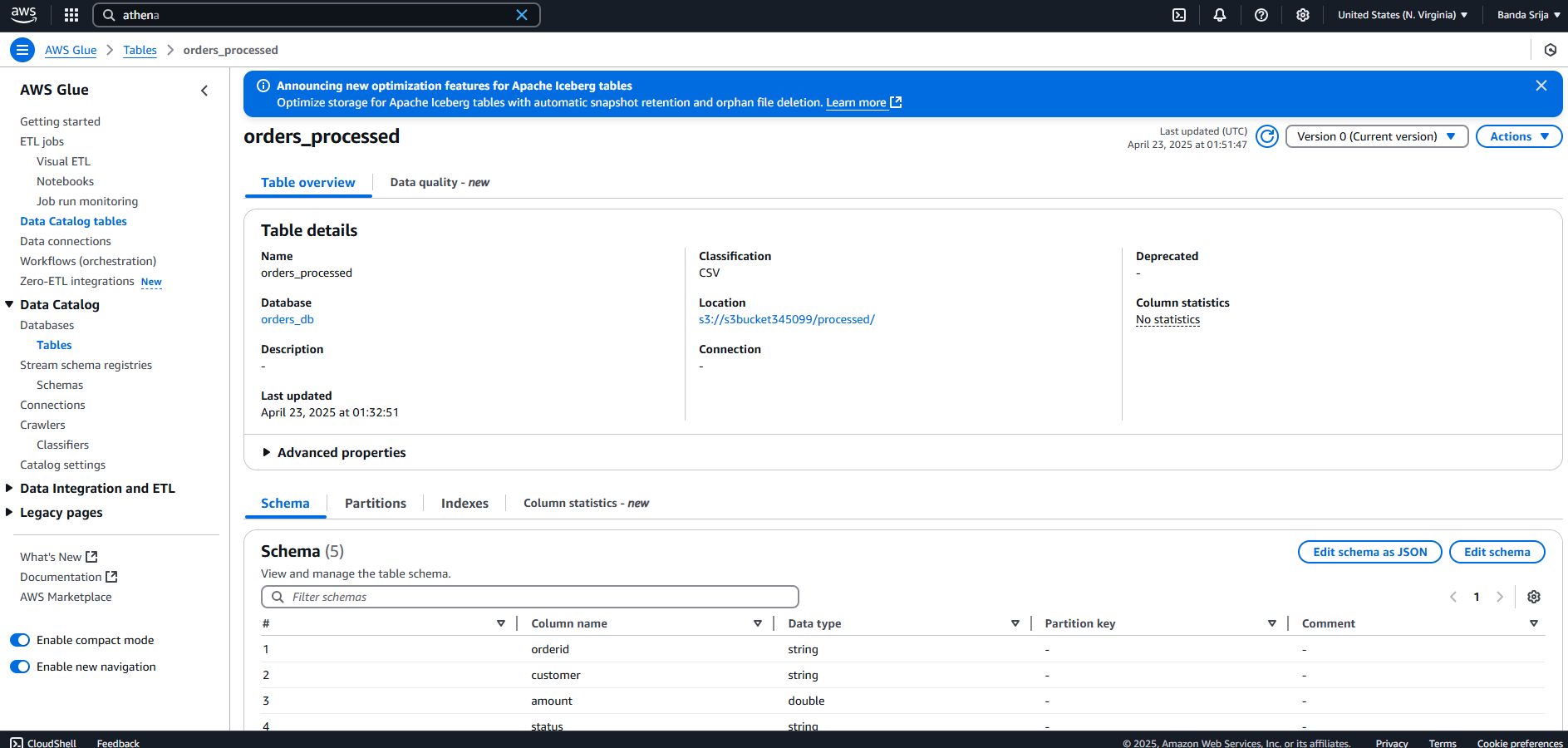




**4. Set up a Glue crawler**

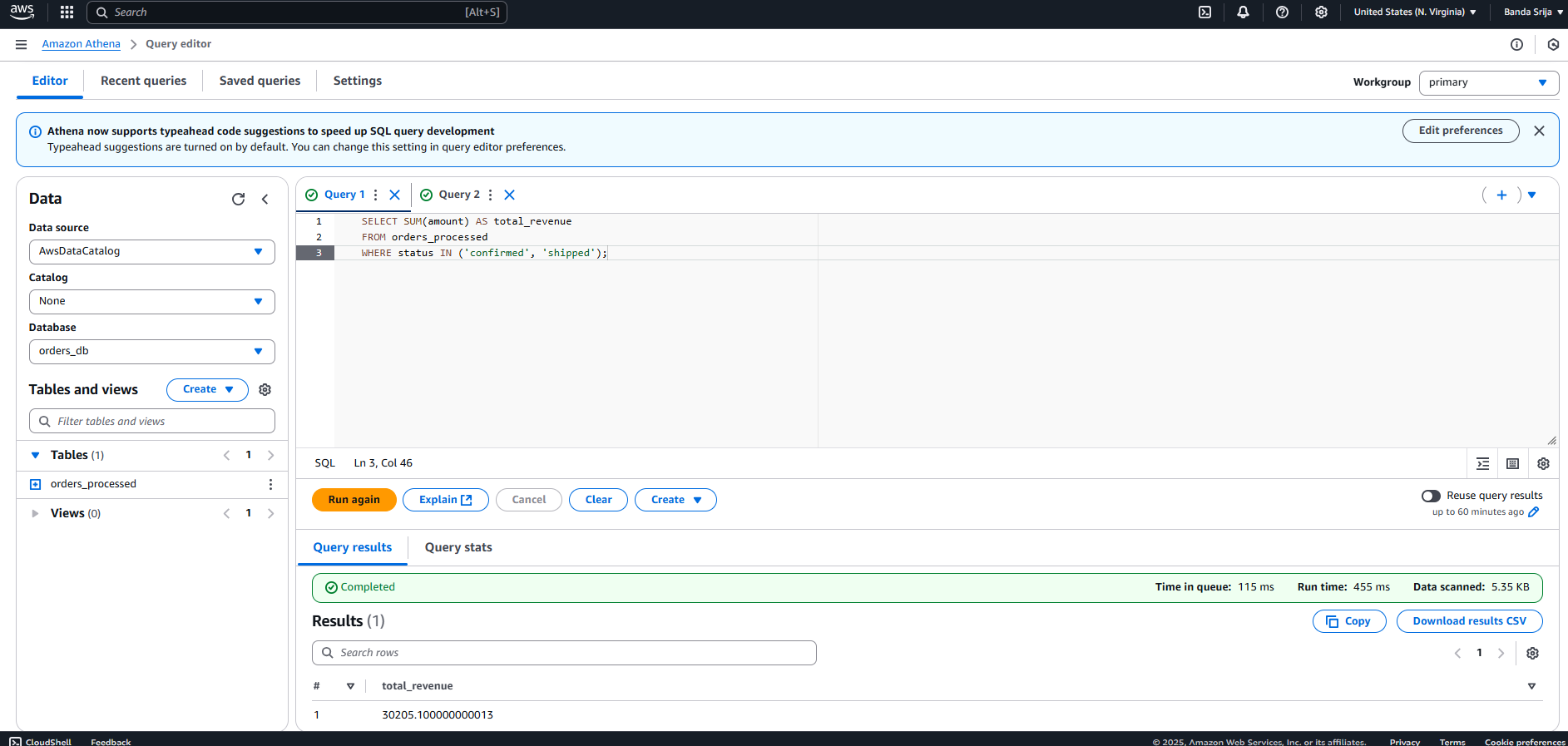
1. AWS Glue ➜ Crawlers ➜ Create.
2. Source: s3://<bucket>/processed/
3. Target database: create (or reuse) one, e.g. orders\_db.
4. Table name: orders\_processed.
5. Run the crawler after Lambda finishes.

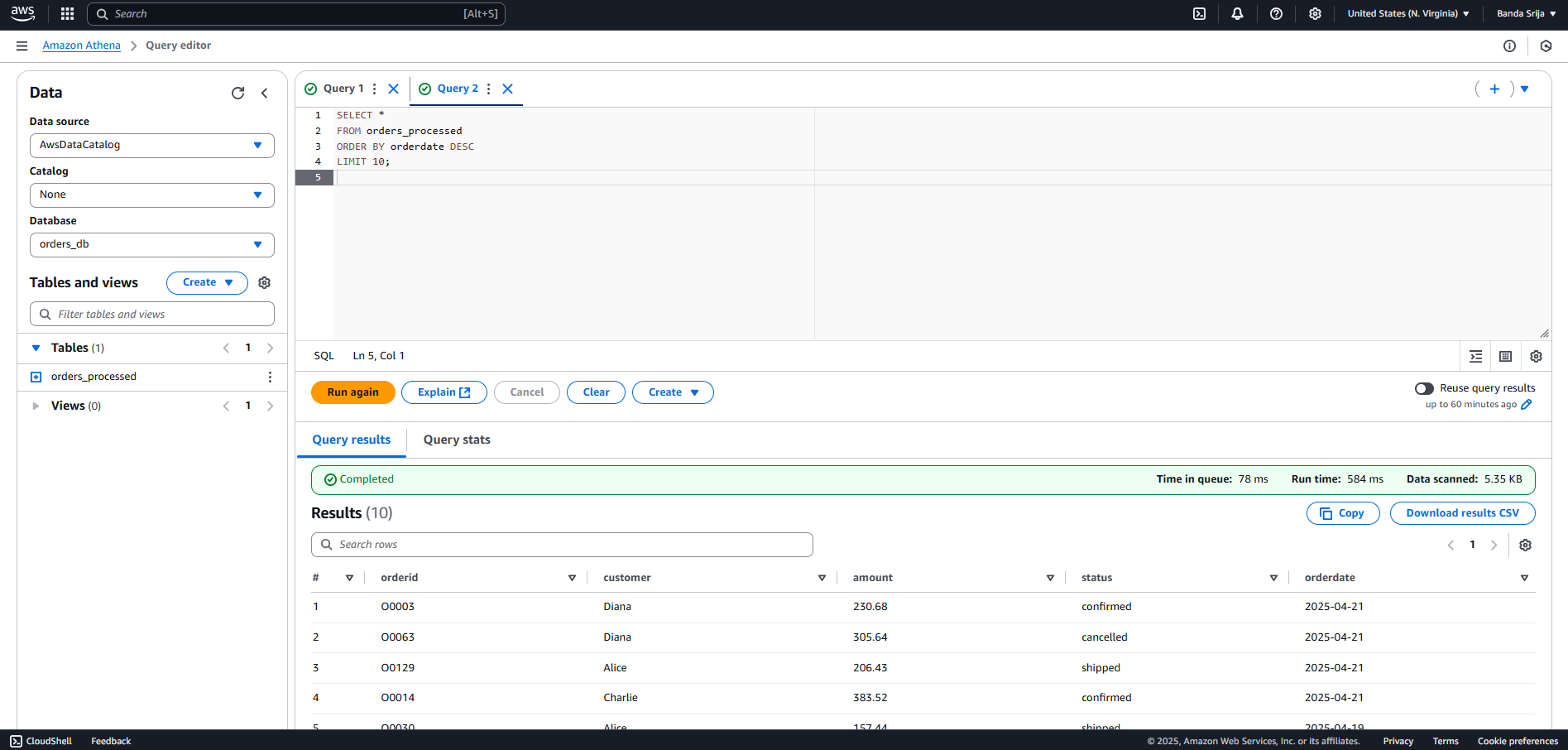




**5. Query the data with Athena**

Make sure Athena’s query‑result location is set, then run:





6. **CloudWatch logs**

CloudWatch Logs ➜ Log groups ➜ /aws/lambda/<function‑name>