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

Tarun Kumar Kanakala

801384590

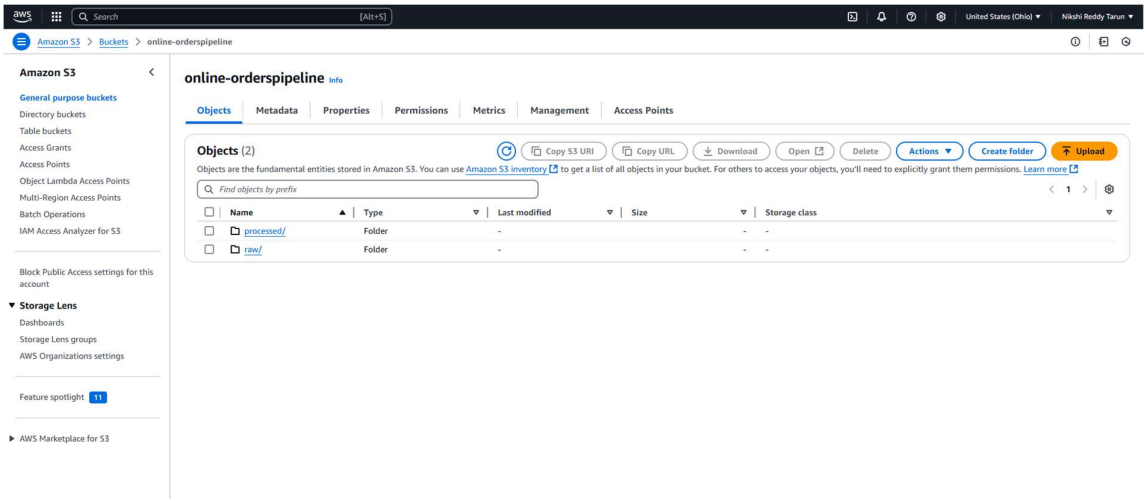
GITHUB LINK-

https://github.com/K-Tarunkumar/ServerlessETLPipeline.git

Step-by-step setup

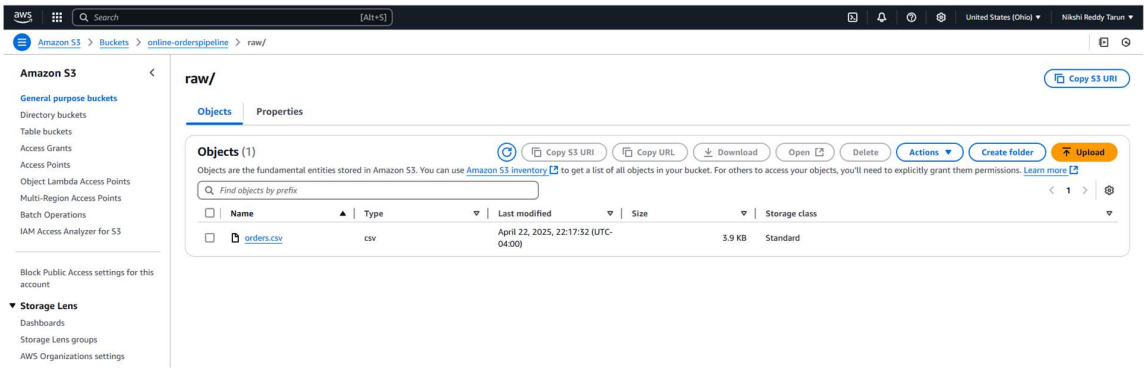
1. Create the bucket and folders

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



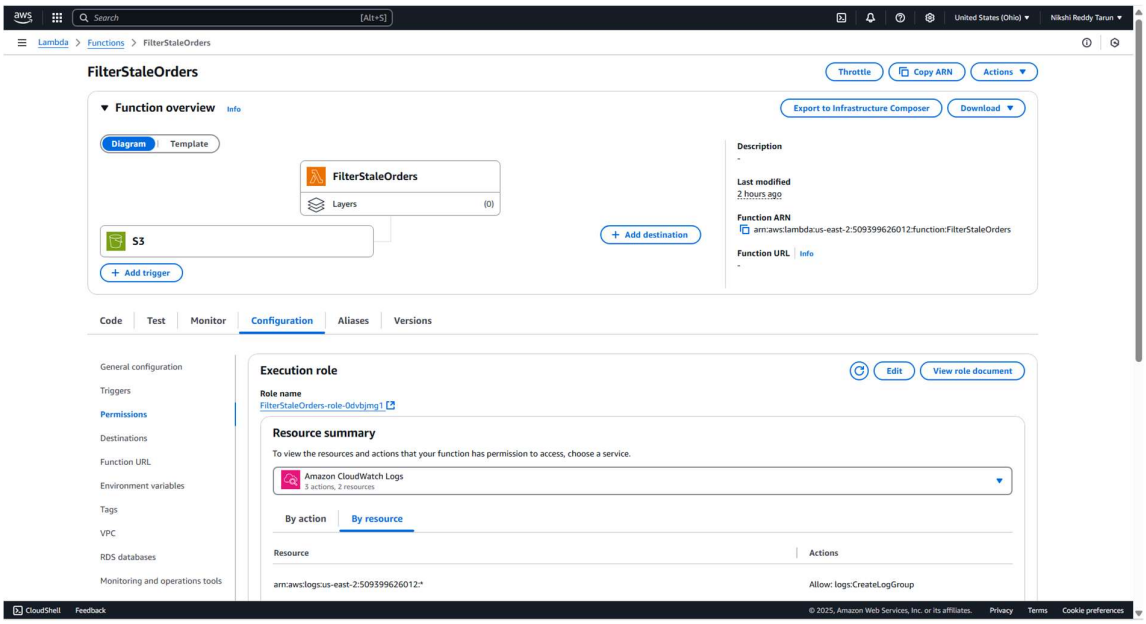
2. Generate sample data locally

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



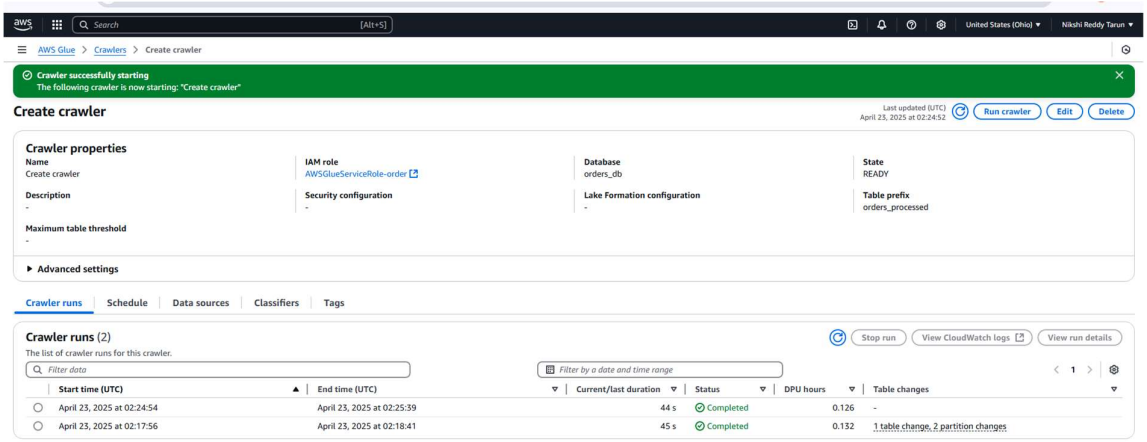
3. Create the Lambda function

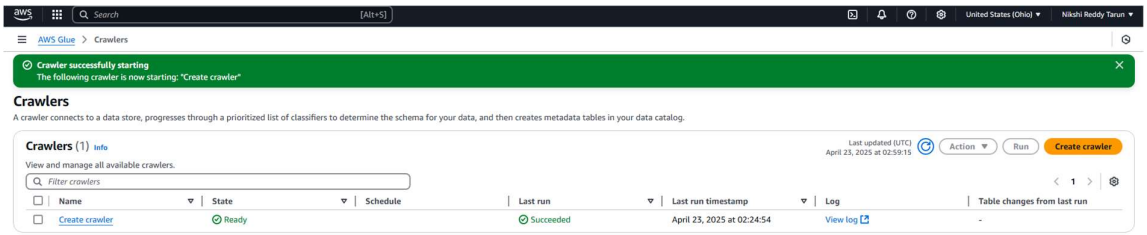
Handler (Python 3.9 or newer)



4. Set up a Glue crawler

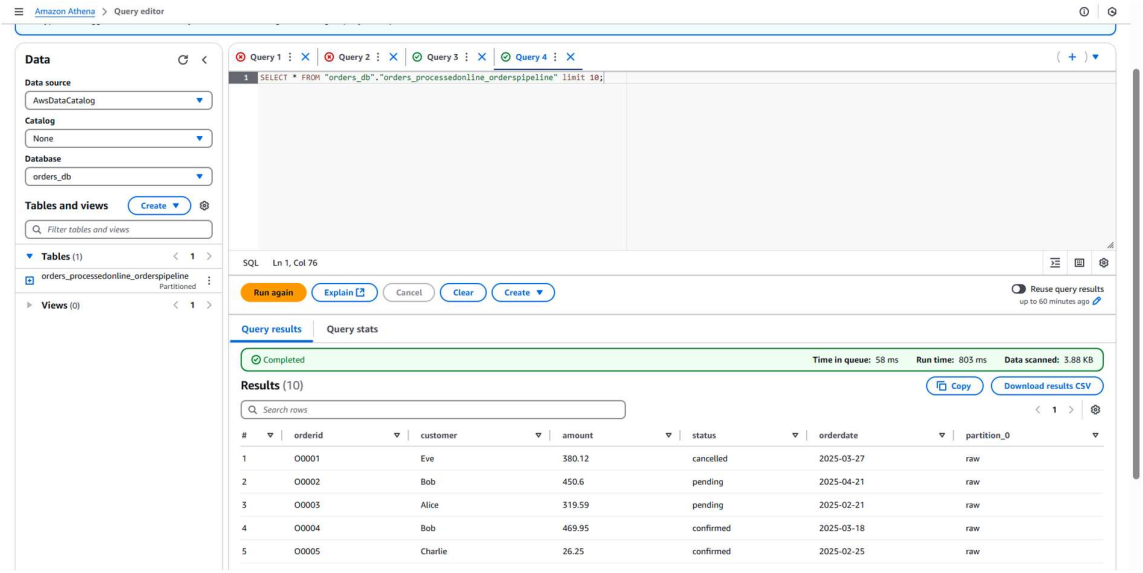
1. AWS Glue ➜ Crawlers ➜ Create.
2. Source: s3://<bucket>/processed/
3. Target database: create (or reuse) one, e.g. orders_db.
4. Table name: orders_processed.
5. Run the crawler after Lambda finishes.

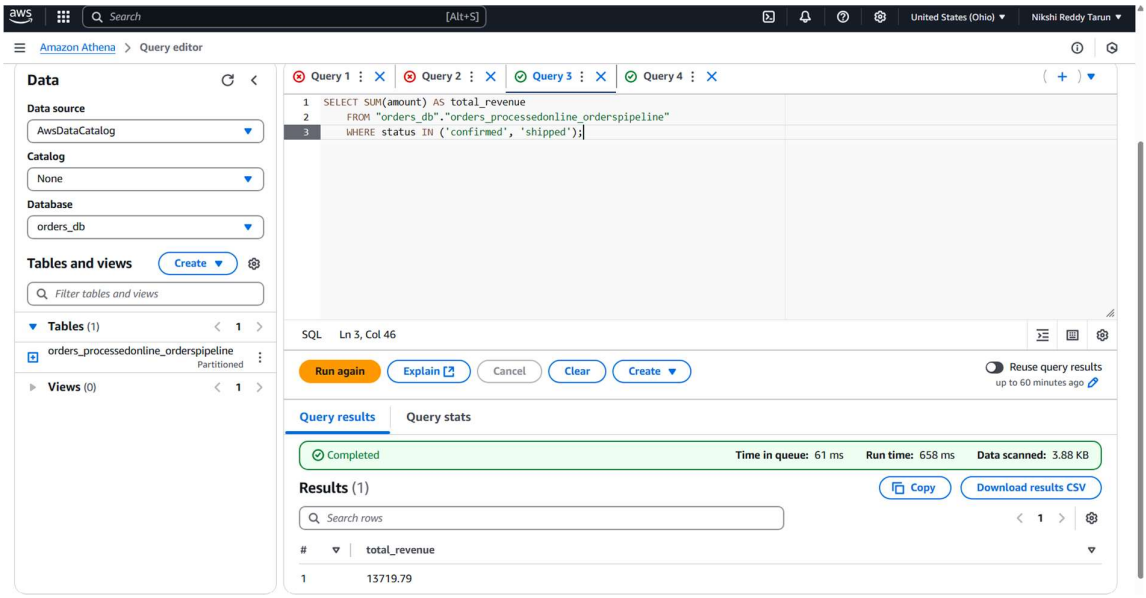




5. Query the data with Athena

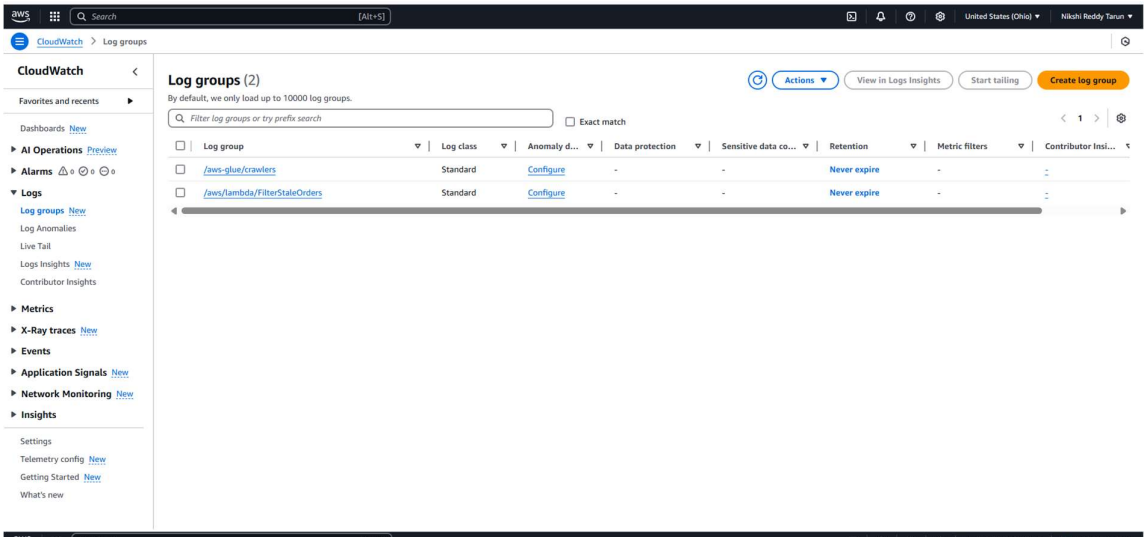
Make sure Athena's query-result location is set, then run:

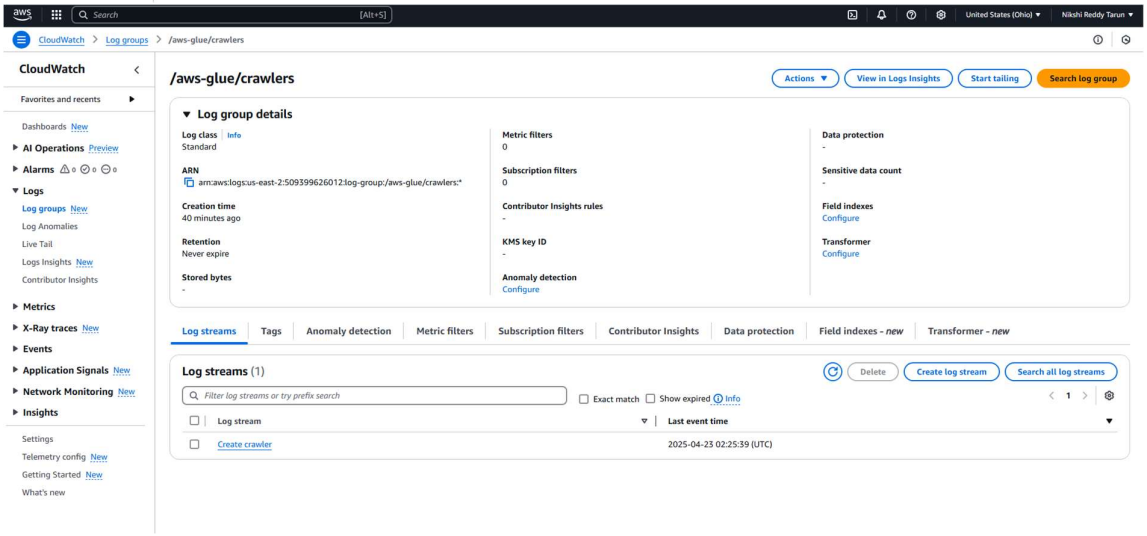




6. CloudWatch logs

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


Search
[Alt+S]
United States (Ohio)
Nikshi Reddy Tarun
CloudWatch > Log groups
CloudWatch
Favorites and recents
Dashboards New
AI Operations Preview
Alarms
Logs
Log groups New
Log Anomalies
Live Tail
Logs Insights New
Contributor Insights
Metrics
X-Ray traces New
Events
Application Signals New
Network Monitoring New
Insights
Settings
Telemetry config New
Getting Started New
What's new
Log groups (2)
By default, we only load up to 10000 log groups.
Actions
View in Logs Insights
Start tailing
Create log group
Filter log groups or try prefix search
Exact match
Log group
Log class
Anomaly d...
Data protection
Sensitive data co...
Retention
Metric filters
Contributor Insi...
/aws-glue/crawlers
Standard
Configure
Never expire
/aws/lambda/FilterStaleOrders
Standard
Configure
Never expire


Search
[Alt+S]
United States (Ohio)
Nikshi Reddy Tarun
CloudWatch > Log groups > /aws-glue/crawlers
CloudWatch
Favorites and recents
Dashboards New
AI Operations Preview
Alarms
Logs
Log groups New
Log Anomalies
Live Tail
Logs Insights New
Contributor Insights
Metrics
X-Ray traces New
Events
Application Signals New
Network Monitoring New
Insights
Settings
Telemetry config New
Getting Started New
What's new
/aws-glue/crawlers
Actions
View in Logs Insights
Start tailing
Search log group
Log group details
Log class Info
Standard
ARN
arn:aws:logs:us-east-2:509399626012:log-group:/aws-glue/crawlers:*
Creation time
40 minutes ago
Retention
Never expire
Stored bytes
-
Metric filters
0
Subscription filters
0
Contributor Insights rules
-
KMS key ID
-
Anomaly detection
Configure
Data protection
-
Sensitive data count
-
Field indexes
Configure
Transformer
Configure
Log streams
Tags
Anomaly detection
Metric filters
Subscription filters
Contributor Insights
Data protection
Field indexes - new
Transformer - new
Log streams (1)
Delete
Create log stream
Search all log streams
Filter log streams or try prefix search
Exact match
Show expired Info
Log stream
Last event time
Create crawler
2025-04-23 02:25:39 (UTC)