Project Report: AWS Serverless Data Pipeline – Online Orders Demo

By- Vivek Sai Chinna Burada

📝 Project Overview

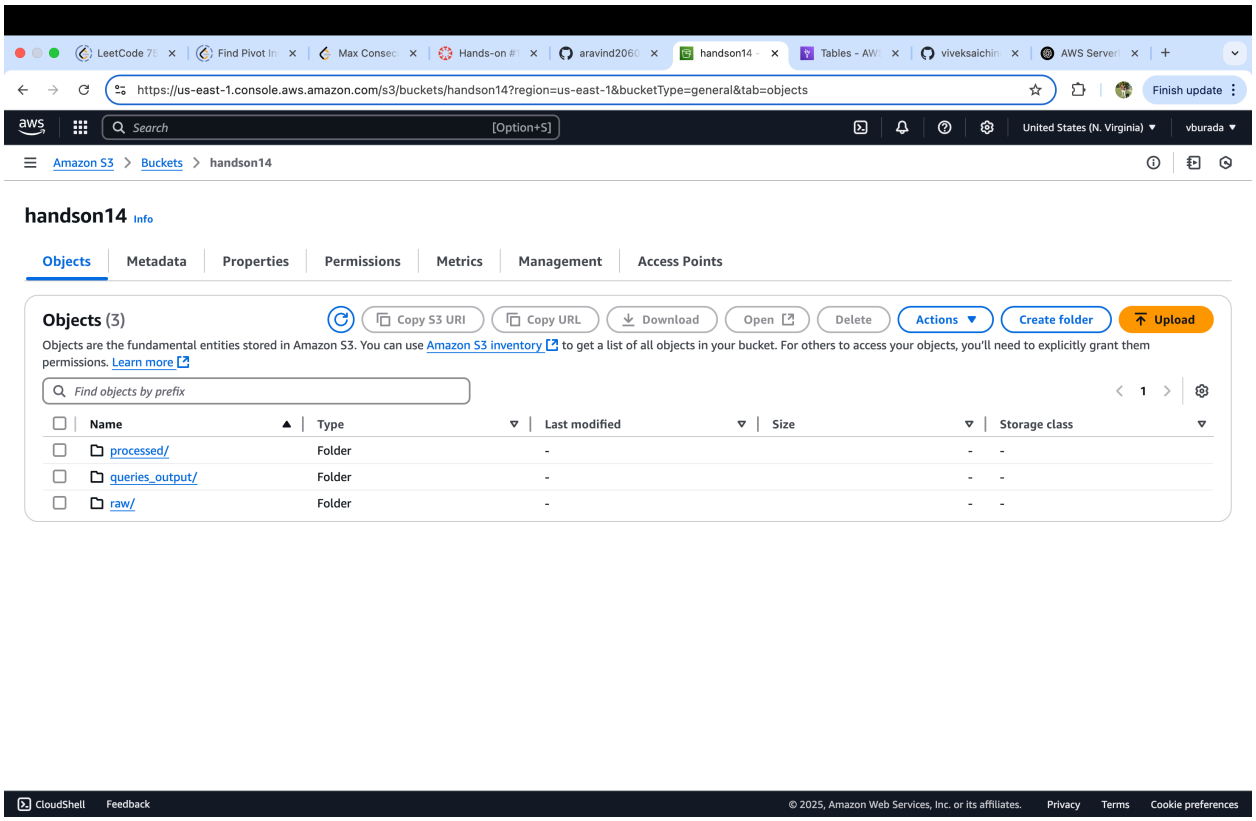
This project demonstrates a cost-effective serverless data pipeline using AWS services to process online order data. The pipeline ingests a CSV file of orders, filters outdated pending/cancelled entries, catalogs the cleaned data, and allows ad-hoc querying.

🛠️ AWS Services Used

- **Amazon S3** – Stores raw and processed order files.
- **AWS Lambda** – Processes data to filter unwanted records.
- **AWS Glue** – Crawls the processed data and creates a catalog.
- **Amazon Athena** – Executes SQL queries on cleaned data.
- **Amazon CloudWatch** – Logs Lambda execution details and metrics.

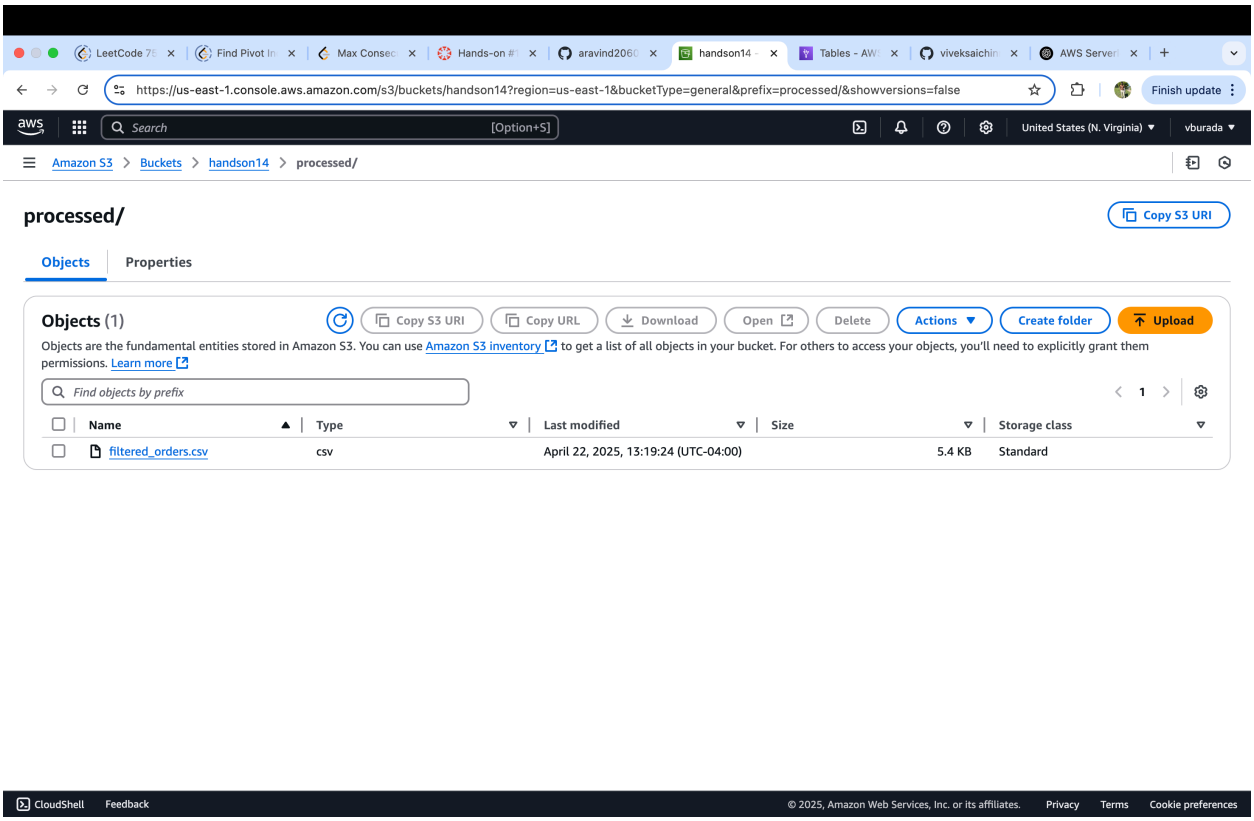
🎯 Goal & Data Flow

1. **Upload `orders.csv` to S3 → raw/**
2. **S3 trigger invokes Lambda**
3. **Lambda filters out pending/cancelled orders older than 30 days**
4. **Lambda saves filtered data to S3 → processed/**
5. **Glue crawler catalogs the cleaned data**
6. **Athena queries cleaned data**



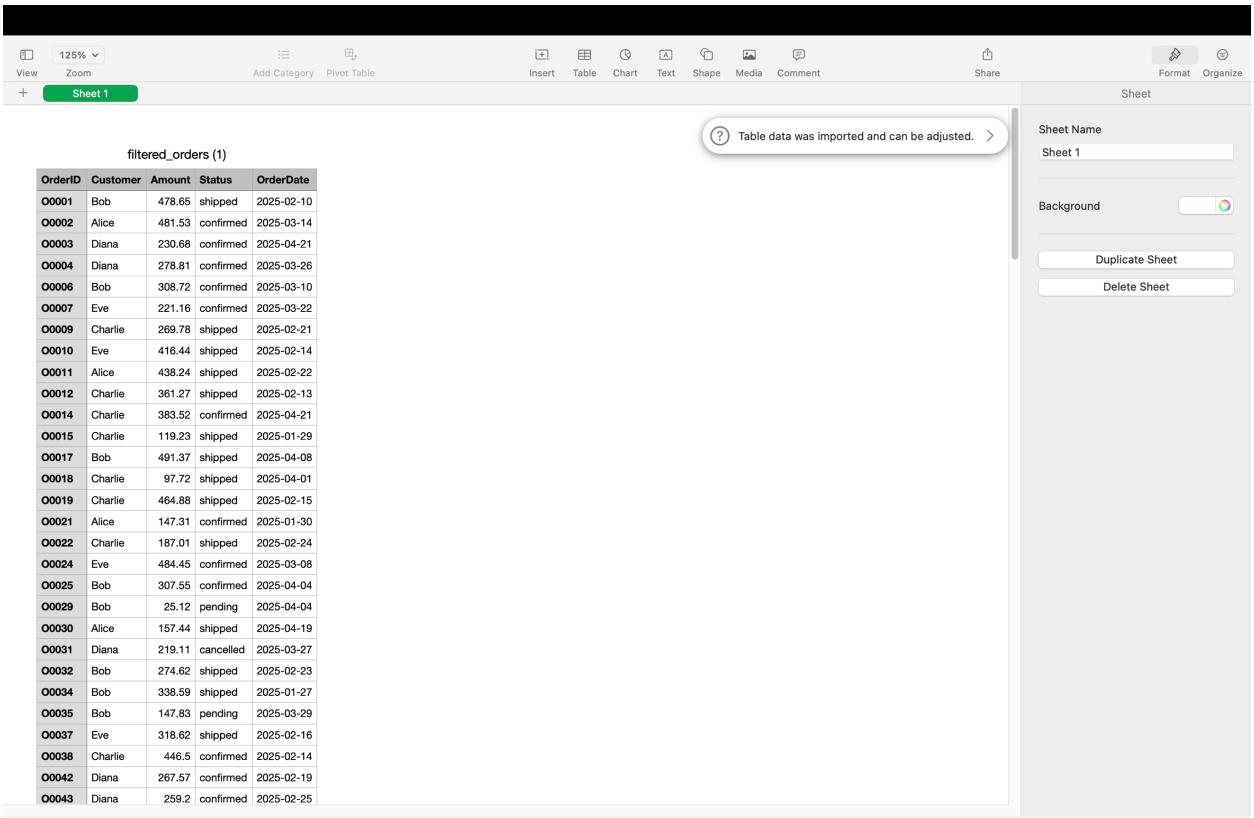
📂 S3 Bucket Setup

- Bucket name: `handson14`
- Folders:
  - `raw/`: contains uploaded `orders.csv`
  - `processed/`: receives Lambda output `filtered_orders.csv`



🧪 Data Generation (Local)

Python script generated `orders.csv` with 100 random orders and statuses (`confirmed`, `shipped`, `pending`, `cancelled`).

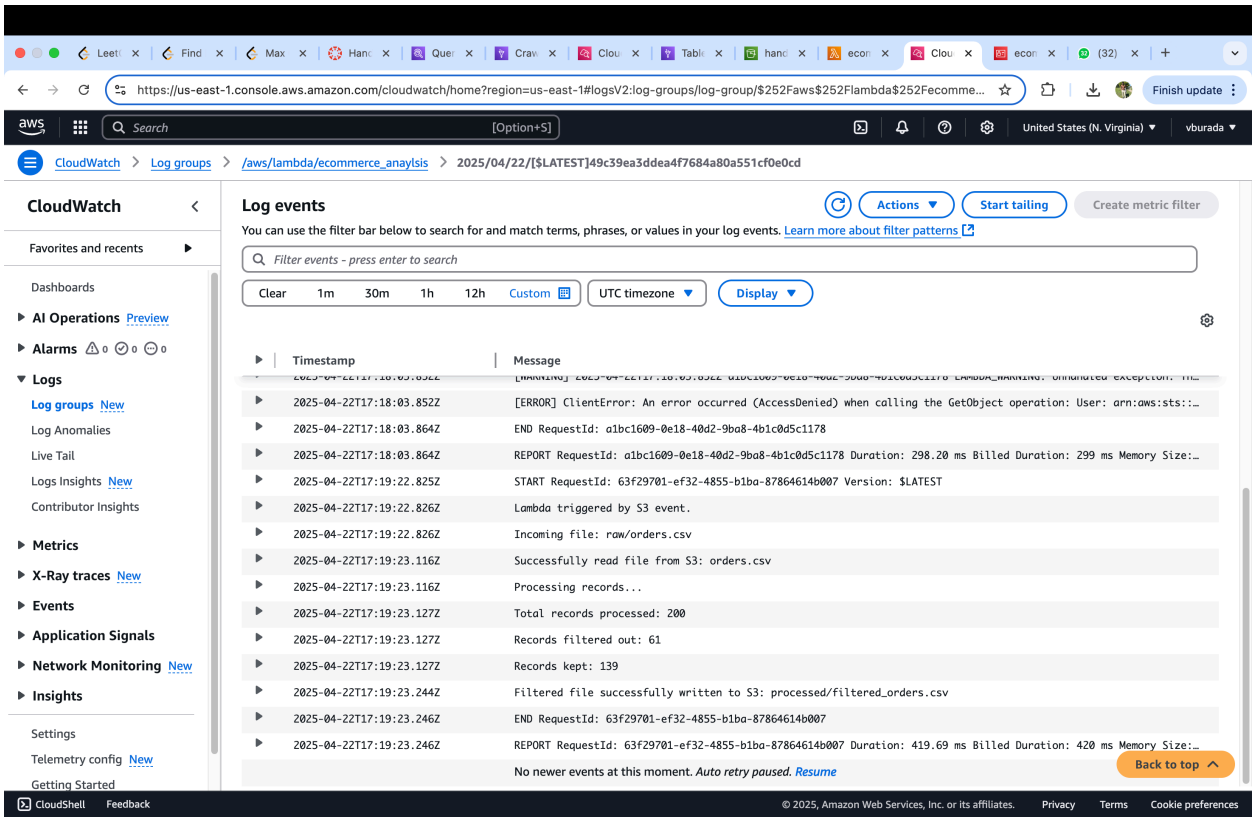


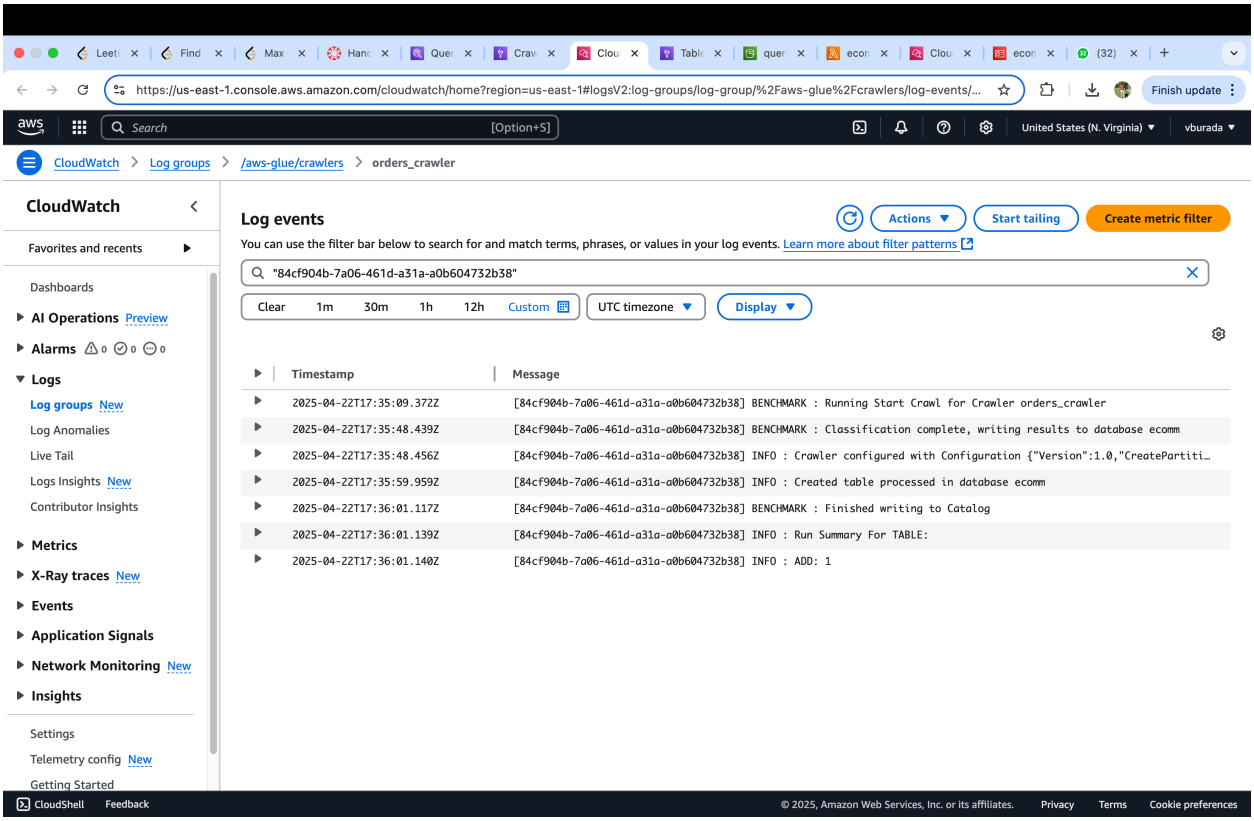
filtered_orders (1)

OrderID	Customer	Amount	Status	OrderDate
O0001	Bob	478.65	shipped	2025-02-10
O0002	Alice	481.53	confirmed	2025-03-14
O0003	Diana	230.68	confirmed	2025-04-21
O0004	Diana	278.81	confirmed	2025-03-26
O0006	Bob	308.72	confirmed	2025-03-10
O0007	Eve	221.16	confirmed	2025-03-22
O0009	Charlie	269.78	shipped	2025-02-21
O0010	Eve	416.44	shipped	2025-02-14
O0011	Alice	438.24	shipped	2025-02-22
O0012	Charlie	361.27	shipped	2025-02-13
O0014	Charlie	383.52	confirmed	2025-04-21
O0015	Charlie	119.23	shipped	2025-01-29
O0017	Bob	491.37	shipped	2025-04-08
O0018	Charlie	97.72	shipped	2025-04-01
O0019	Charlie	464.88	shipped	2025-02-15
O0021	Alice	147.31	confirmed	2025-01-30
O0022	Charlie	187.01	shipped	2025-02-24
O0024	Eve	484.45	confirmed	2025-03-08
O0025	Bob	307.55	confirmed	2025-04-04
O0029	Bob	25.12	pending	2025-04-04
O0030	Alice	157.44	shipped	2025-04-19
O0031	Diana	219.11	cancelled	2025-03-27
O0032	Bob	274.62	shipped	2025-02-23
O0034	Bob	338.59	shipped	2025-01-27
O0035	Bob	147.83	pending	2025-03-29
O0037	Eve	318.62	shipped	2025-02-16
O0038	Charlie	446.5	confirmed	2025-02-14
O0042	Diana	267.57	confirmed	2025-02-19
O0043	Diana	259.2	confirmed	2025-02-25

🧠 Lambda Function – Core Logic

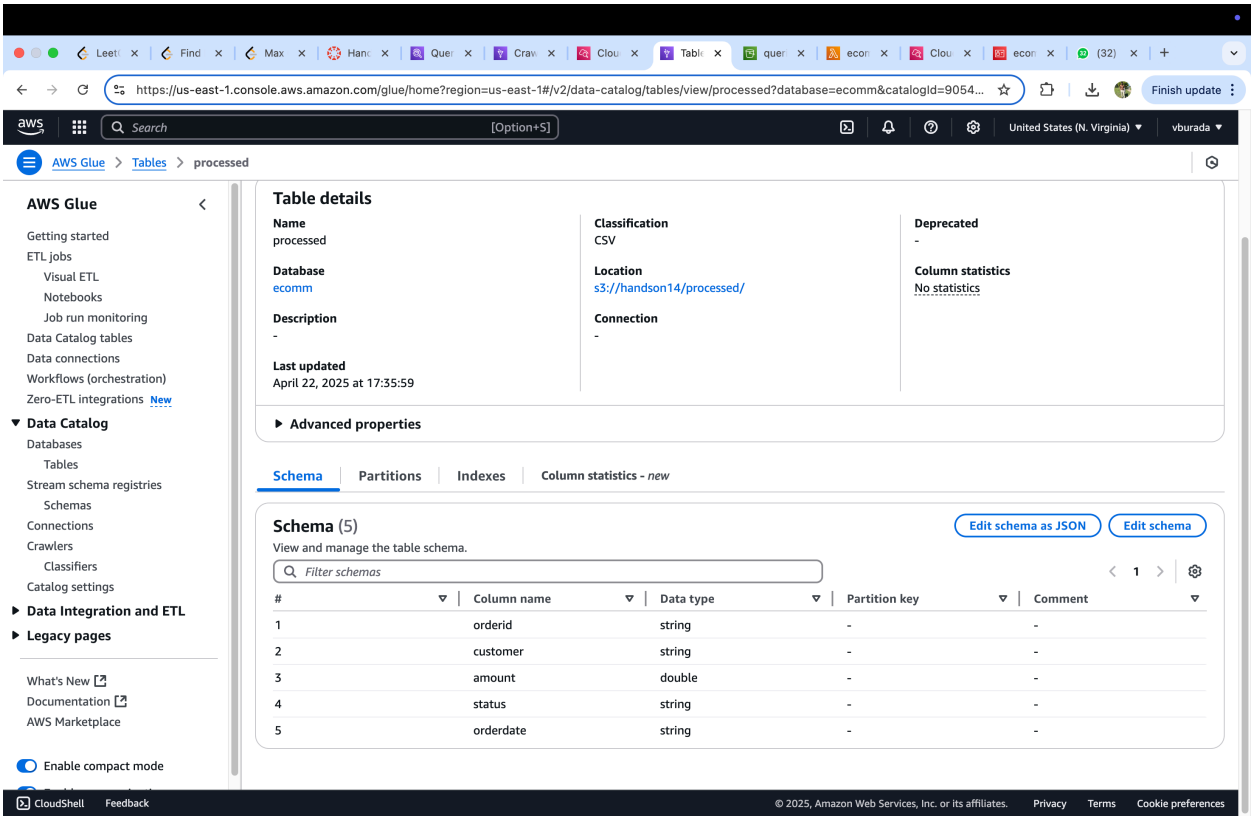
- Triggered by S3 event (`raw/` prefix)
- Filters out:
  - Orders with **status = pending/cancelled**
  - AND **older than 30 days**
- Writes filtered CSV to `processed/` folder


https://us-east-1.console.aws.amazon.com/cloudwatch/home?region=us-east-1#logsV2:log-groups/log-group/$252Faws$252Flambda$252Fecomme...
Finish update
Search
[Option+S]
United States (N. Virginia)
vburada
CloudWatch
Log groups
/aws/lambda/ecommerce_anaylsis
2025/04/22/[$LATEST]49c39ea3ddea4f7684a80a551cf0e0cd
CloudWatch
Favorites and recents
Dashboards
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
Network Monitoring New
Insights
Settings
Telemetry config New
Getting Started
Log events
Actions
Start tailing
Create metric filter
You can use the filter bar below to search for and match terms, phrases, or values in your log events. Learn more about filter patterns
Filter events - press enter to search
Clear
1m
30m
1h
12h
Custom
UTC timezone
Display
Timestamp
Message
2025-04-22T17:18:03.852Z [ERROR] ClientError: An error occurred (AccessDenied) when calling the GetObject operation: User: arn:aws:sts::…
2025-04-22T17:18:03.864Z END RequestId: a1bc1609-0e18-40d2-9ba8-4b1c0d5c1178
2025-04-22T17:18:03.864Z REPORT RequestId: a1bc1609-0e18-40d2-9ba8-4b1c0d5c1178 Duration: 298.20 ms Billed Duration: 299 ms Memory Size:…
2025-04-22T17:19:22.825Z START RequestId: 63f29701-ef32-4855-b1ba-87864614b007 Version: $LATEST
2025-04-22T17:19:22.826Z Lambda triggered by S3 event.
2025-04-22T17:19:22.826Z Incoming file: raw/orders.csv
2025-04-22T17:19:23.116Z Successfully read file from S3: orders.csv
2025-04-22T17:19:23.116Z Processing records...
2025-04-22T17:19:23.127Z Total records processed: 200
2025-04-22T17:19:23.127Z Records filtered out: 61
2025-04-22T17:19:23.127Z Records kept: 139
2025-04-22T17:19:23.244Z Filtered file successfully written to S3: processed/filtered_orders.csv
2025-04-22T17:19:23.246Z END RequestId: 63f29701-ef32-4855-b1ba-87864614b007
2025-04-22T17:19:23.246Z REPORT RequestId: 63f29701-ef32-4855-b1ba-87864614b007 Duration: 419.69 ms Billed Duration: 420 ms Memory Size:…
No newer events at this moment. Auto retry paused. Resume
Back to top
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


https://us-east-1.console.aws.amazon.com/cloudwatch/home?region=us-east-1#logsV2:log-groups/log-group/%2Faws-glue%2Fcrawlers/log-events/...
Finish update
Search
[Option+S]
United States (N. Virginia)
vburada
CloudWatch
Log groups
/aws-glue/crawlers
orders_crawler
CloudWatch
Favorites and recents
Dashboards
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
Network Monitoring New
Insights
Settings
Telemetry config New
Getting Started
Log events
Actions
Start tailing
Create metric filter
You can use the filter bar below to search for and match terms, phrases, or values in your log events. Learn more about filter patterns
"84cf904b-7a06-461d-a31a-a0b604732b38"
Clear
1m
30m
1h
12h
Custom
UTC timezone
Display
Timestamp
Message
2025-04-22T17:35:09.372Z [84cf904b-7a06-461d-a31a-a0b604732b38] BENCHMARK : Running Start Crawl for Crawler orders_crawler
2025-04-22T17:35:48.439Z [84cf904b-7a06-461d-a31a-a0b604732b38] BENCHMARK : Classification complete, writing results to database ecomm
2025-04-22T17:35:48.456Z [84cf904b-7a06-461d-a31a-a0b604732b38] INFO : Crawler configured with Configuration {"Version":1.0,"CreatePartiti…
2025-04-22T17:35:59.959Z [84cf904b-7a06-461d-a31a-a0b604732b38] INFO : Created table processed in database ecomm
2025-04-22T17:36:01.117Z [84cf904b-7a06-461d-a31a-a0b604732b38] BENCHMARK : Finished writing to Catalog
2025-04-22T17:36:01.139Z [84cf904b-7a06-461d-a31a-a0b604732b38] INFO : Run Summary For TABLE:
2025-04-22T17:36:01.140Z [84cf904b-7a06-461d-a31a-a0b604732b38] INFO : ADD: 1
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

AWS Glue – Data Catalog

- Glue crawler setup:
  - **Source**: `s3://handson14/processed/`
  - **Target DB**: `orders_crawler`
  - **Table name**: `processed`
- Run once after Lambda execution



Athena Queries & Output

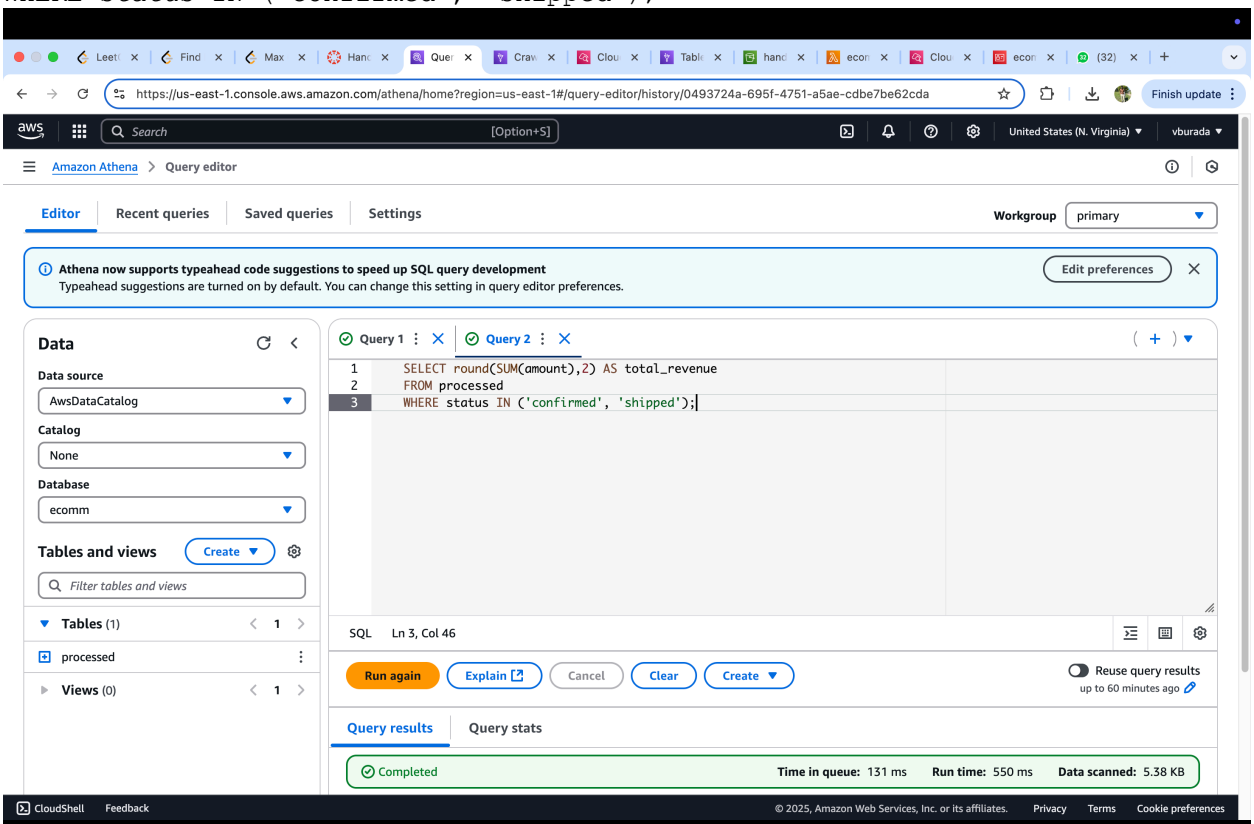
1. **Recent 10 Orders**

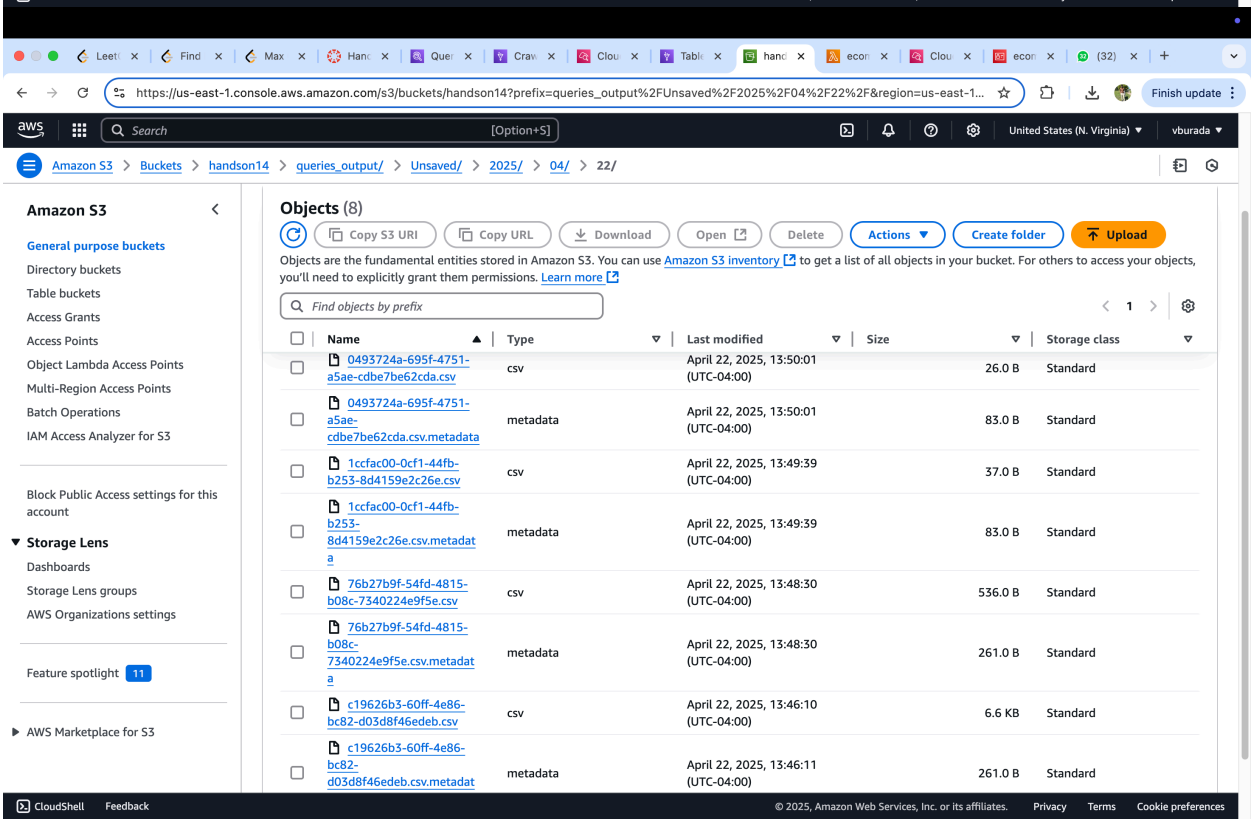
```
sql
CopyEdit
SELECT *
FROM orders_processed
ORDER BY orderdate DESC
LIMIT 10;
```

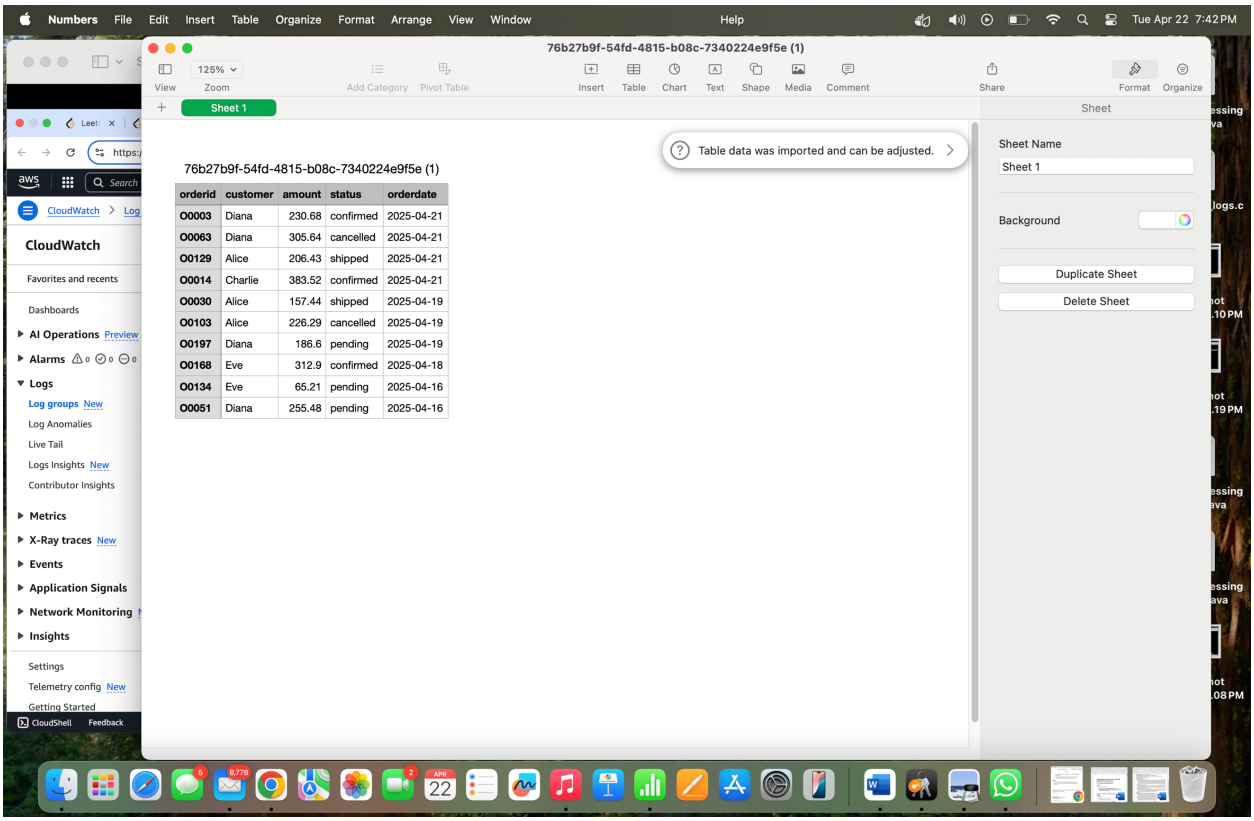
2. **Total Revenue of Fulfilled Orders**

```
sql
CopyEdit
```

```
SELECT SUM(amount) AS total_revenue
FROM orders_processed
WHERE status IN ('confirmed', 'shipped');
```

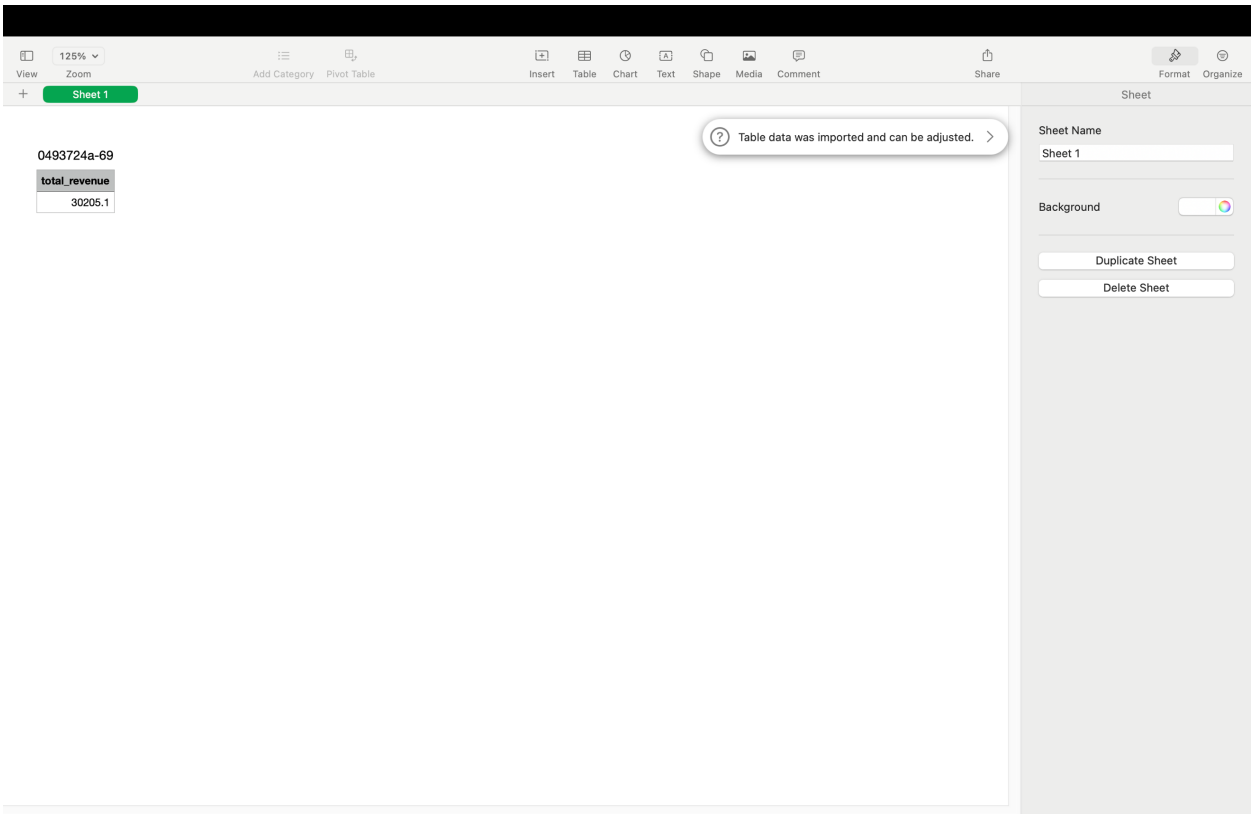






76b27b9f-54fd-4815-b08c-7340224e9f5e (1)

orderid	customer	amount	status	orderdate
O0003	Diana	230.68	confirmed	2025-04-21
O0063	Diana	305.64	cancelled	2025-04-21
O0129	Alice	206.43	shipped	2025-04-21
O0014	Charlie	383.52	confirmed	2025-04-21
O0030	Alice	157.44	shipped	2025-04-19
O0103	Alice	226.29	cancelled	2025-04-19
O0197	Diana	186.6	pending	2025-04-19
O0168	Eve	312.9	confirmed	2025-04-18
O0134	Eve	65.21	pending	2025-04-16
O0051	Diana	255.48	pending	2025-04-16



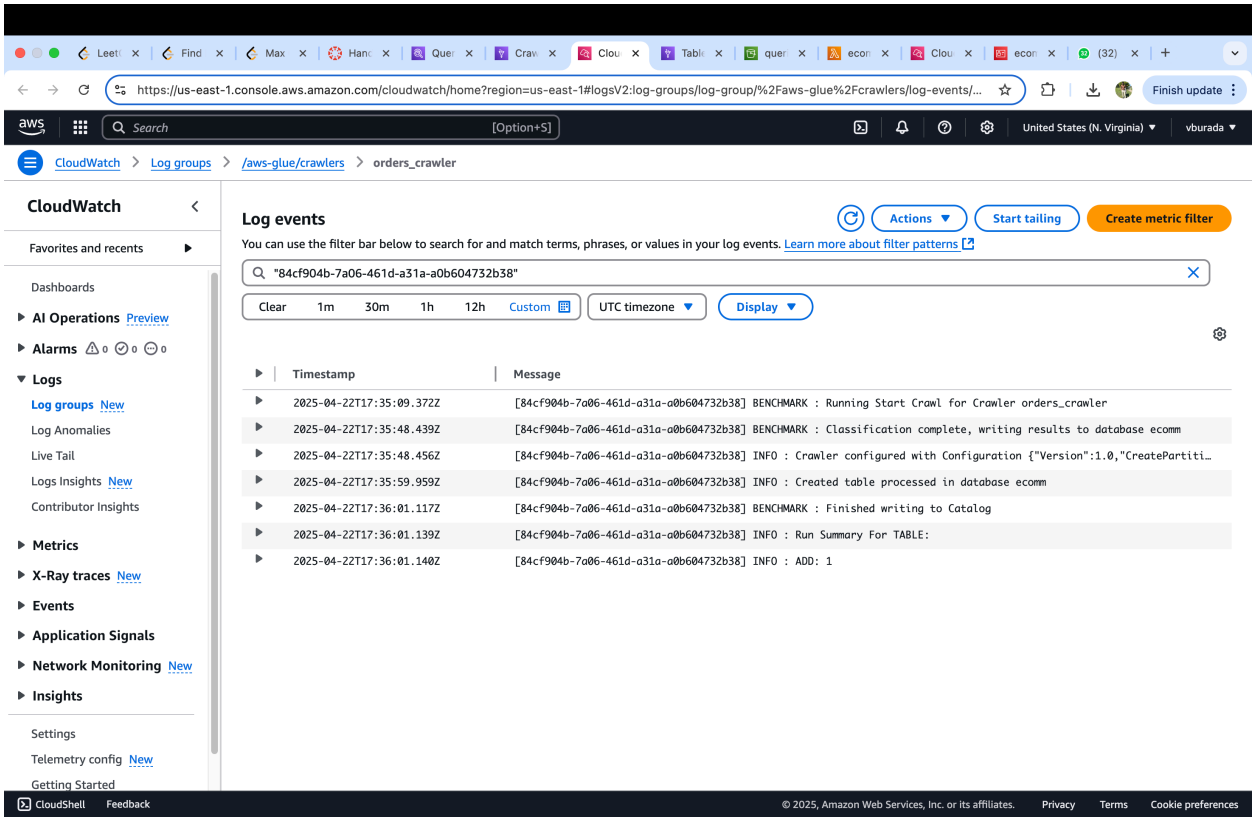
0493724a-69

total_revenue
30205.1

📈 CloudWatch Logs

Each Lambda run logs:

- File name and path
- Number of records kept vs. filtered
- Success or errors



✅ Summary

Stage	Outcome
Data Ingestion	orders.csv uploaded to S3
Data Processing	Lambda filters unwanted rows
Cataloging	Glue crawler creates Athena table
Querying	Athena SQL queries executed
Monitoring	CloudWatch logs Lambda execution
