

Step 1 – Buckets and Folders:

hjuneau-online-orders-pipeline [Info](#)

Objects | Metadata | Properties | Permissions | Metrics | Management | Access Points

Objects (3) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#) [Actions](#) [Create folder](#) [Upload](#)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	processed/	Folder	-	-	-
<input type="checkbox"/>	raw/	Folder	-	-	-
<input type="checkbox"/>	Unsaved/	Folder	-	-	-

Step 2 – Create Lambda Function:

Diagram | Template

hjuneau-online-orders-lambda

Layers (0)

S3

[+ Add destination](#)

[+ Add trigger](#)

Description

Last modified: 27 minutes ago

Function ARN: [arn:aws:lambda:us-east-2:357275763906:function:hjuneau-online-orders-lambda](#)

Function URL: [Info](#)

Code | Test | Monitor | Configuration | Aliases | Versions

Code source [Info](#) [Upload from](#)

EXPLORER

- HJUNEAU-ONLINE-ORDERS-LAMBDA
 - lambda_function.py

DEPLOY

Deploy (Ctrl+Shift+U)

Test (Ctrl+Shift+I)

```
8 def lambda_handler(event, context):
34     for row in reader:
35         if order_status not in ['pending', 'cancelled'] or order_date > cutoff_date:
41             filtered_rows.append(row)
42         else:
43             filtered_out_count += 1
44
45     print(f"Total records processed: {original_count}")
46     print(f"Records filtered out: {filtered_out_count}")
47     print(f"Records kept: {len(filtered_rows)}")
48
49     # Write the filtered rows to memory
50     output = io.StringIO()
51     writer = csv.DictWriter(output, fieldnames=reader.fieldnames)
52     writer.writeheader()
53     writer.writerows(filtered_rows)
54
55     # Save to processed/ folder
56     processed_key = f"processed/filtered_{file_name}"
57
58     try:
```

hjuneau-online-orders-lambda-role-e84nw3k9 [Info](#)

[Delete](#)

Summary

Creation date
April 22, 2025, 18:20 (UTC-04:00)

Last activity
[15 minutes ago](#)

ARN
[arn:aws:iam::357275763906:role/service-role/hjuneau-online-orders-lambda-role-e84nw3k9](#)

Maximum session duration
1 hour

[Edit](#)

Permissions | Trust relationships | Tags | Last Accessed | Revoke sessions

Permissions policies (2) [Info](#)

[Simulate](#)[Remove](#)[Add permissions](#)

You can attach up to 10 managed policies.

Search	Filter by Type	All types	
<input type="checkbox"/>	Policy name	Type	Attached entities
<input type="checkbox"/>	AmazonS3FullAccess	AWS managed	3
<input type="checkbox"/>	AWSLambdaBasicExecutionRole-402b6bca-c287-4eb7-8057-92...	Customer managed	1

Step 3 – Set up Crawler

Review and update

Step 1: Set crawler properties

[Edit](#)

Set crawler properties

Name	Description	Tags
hjuneau-online-orders-crawler	-	-

Step 2: Choose data sources and classifiers

[Edit](#)

Data sources (1) [info](#)

The list of data sources to be scanned by the crawler.

Type	Data source	Parameters
S3	s3://hjuneau-online-orders-pipeline/processed/	Recrawl all

Step 3: Configure security settings

[Edit](#)

Configure security settings

IAM role	Security configuration	Lake Formation configuration
AWSGlueServiceRole-hjuneau-glue-service-role	-	-

Step 4: Set output and scheduling

[Edit](#)

Set output and scheduling

Database	Table prefix - optional	Maximum table threshold - optional	Schedule
orders_processed	-	-	On demand

[Cancel](#)[Previous](#)[Update](#)

Step 4 – Query Data:

Query 1

```
1 SELECT SUM(amount) AS total_revenue
2 FROM processed
3 WHERE status IN ('confirmed', 'shipped');
```

SQL Ln 3, Col 42

Run again

Explain

Cancel

Clear

Create

Reuse query results

up to 60 minutes ago

Query results

Query stats

Completed

Time in queue: 60 ms

Run time: 720 ms

Data scanned: 5.11 KB

Results (1)

Copy

Download results CSV

Search rows

#

total_revenue

1

24280.62

Query outputs are included in the [Github Repo](#)