Building a Real-Time Weather Data Pipeline for Weather Analytics
AWS Kinesis, Lambda, Redshift

Project overview:

This project addresses the growing need for real-time weather monitoring by providing data-driven insights into weather patterns and trends. The pipeline will support:

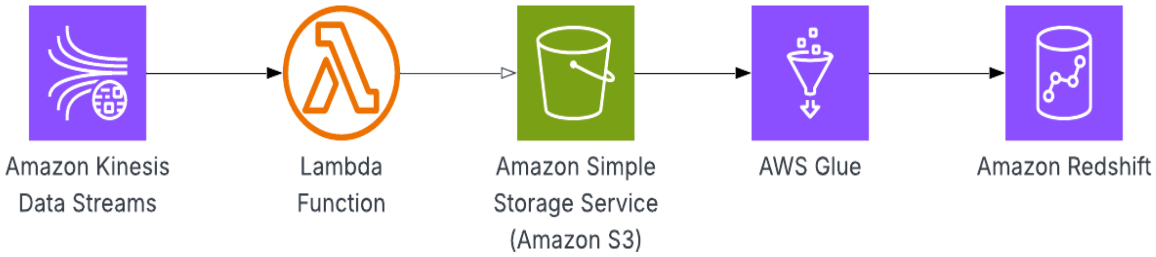
- Real-time monitoring – Immediate visibility into changing weather conditions
- Historical analysis – Trend identification for weather forecasting
- Operational efficiency – Automated data processing and storage
- Decision support – Visual dashboards for weather-dependent operations

The solution provides a scalable, serverless architecture for processing streaming weather data while maintaining cost-efficiency.

AWS Services Used:

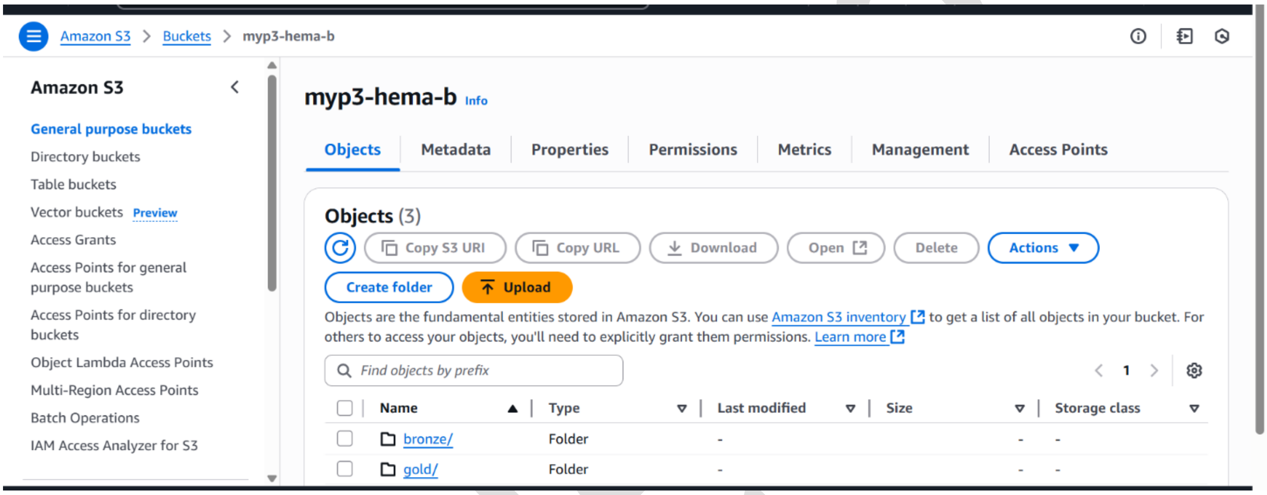
- **Amazon Kinesis Data Streams:** For ingesting weather data in real time.
- **AWS Lambda:** For processing incoming records and storing them in S3 (Bronze layer).
- **Amazon S3:** For storing raw (bronze) and processed (silver) data.
- **AWS Glue:** For running ETL jobs and creating tables in the Data Catalogue.
- **Amazon Redshift Serverless:** For querying structured weather data using SQL.
- **IAM:** For access control and permission management.
- **Cloud Shell:** To simulate data ingestion via Python script

Architecture Diagram


Amazon Kinesis Data Streams
Lambda Function
Amazon Simple Storage Service (Amazon S3)
AWS Glue
Amazon Redshift

Data Flow Description

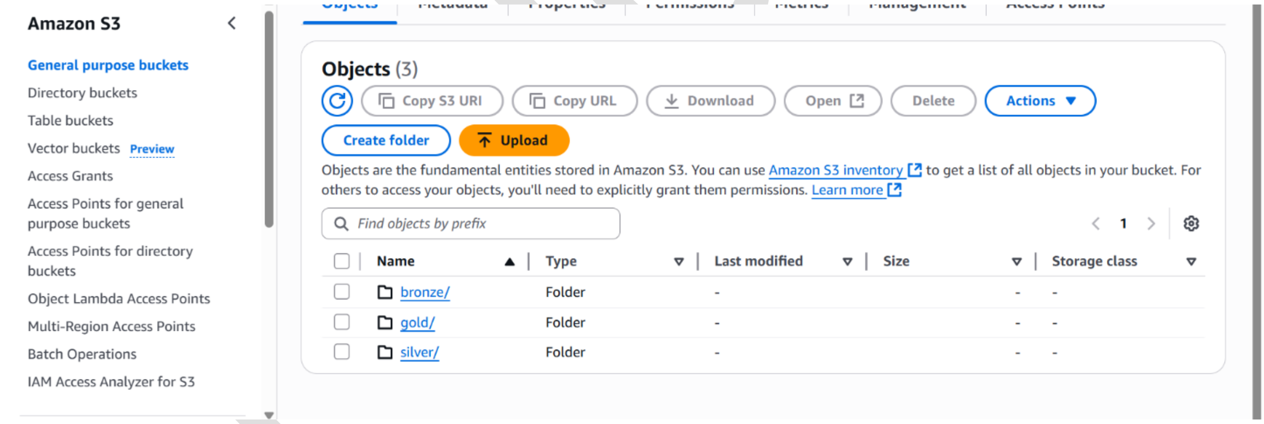
1. **Data Generation:**
   - A Python script runs on AWS Cloud Shell to simulate weather data and push it to the Kinesis Data Stream.
   - Creating bucket through terraform.
   - Name it as: “myp3-hema-b”, folders: “bronze”,” silver”,” gold” and weather_stream subfolders in it.



Objects (3)

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

□	Name	Type	Last modified	Size	Storage class
□	bronze/	Folder	-	-	-
□	gold/	Folder	-	-	-



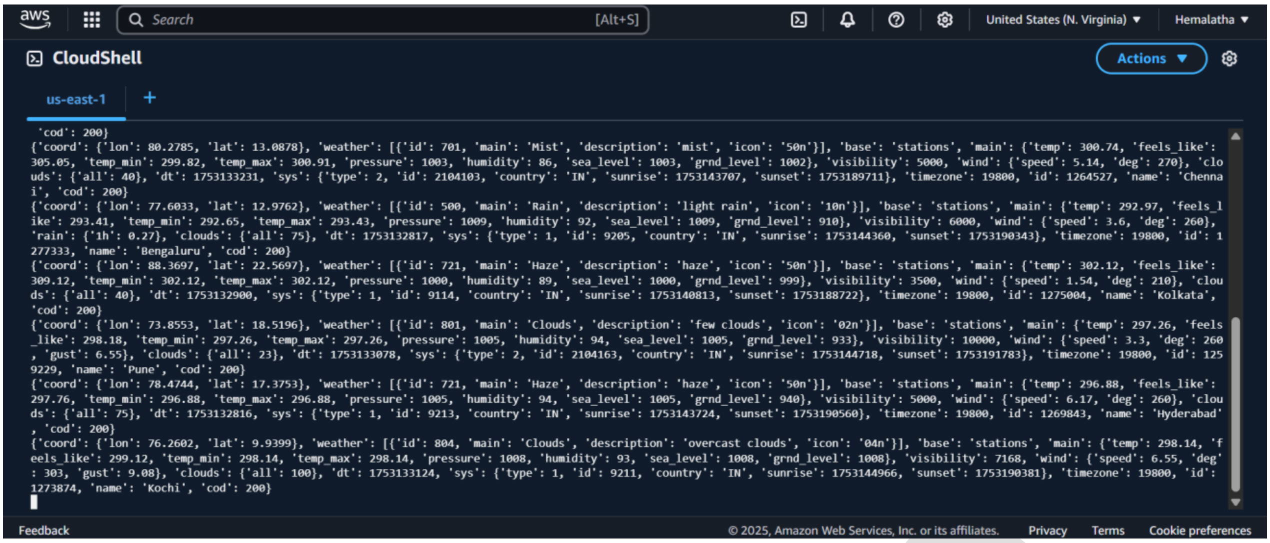
Objects (3)

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

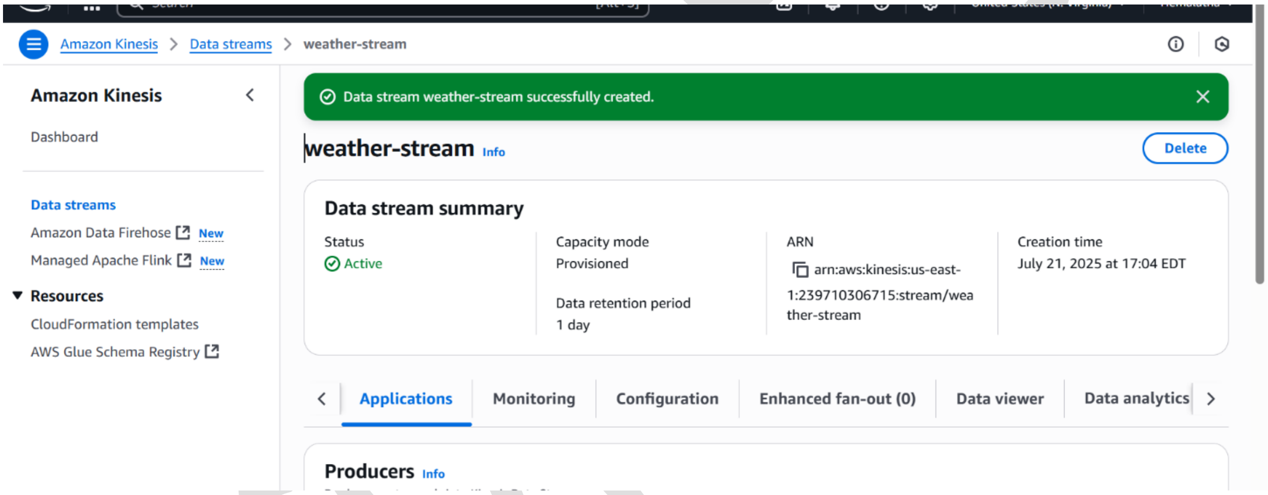
□	Name	Type	Last modified	Size	Storage class
□	bronze/	Folder	-	-	-
□	gold/	Folder	-	-	-
□	silver/	Folder	-	-	-

2. **Kinesis Stream:**

A stream named weather-stream receives real-time JSON records from Cloud Shell.



```
 'cod': 200}
{'coord': {'lon': 80.2785, 'lat': 13.0878}, 'weather': [{'id': 701, 'main': 'Mist', 'description': 'mist', 'icon': '50n'}], 'base': 'stations', 'main': {'temp': 300.74, 'feels_like':
305.05, 'temp_min': 299.82, 'temp_max': 300.91, 'pressure': 1003, 'humidity': 86, 'sea_level': 1003, 'grnd_level': 1002}, 'visibility': 5000, 'wind': {'speed': 5.14, 'deg': 270}, 'clo
uds': {'all': 40}, 'dt': 1753133231, 'sys': {'type': 2, 'id': 2104103, 'country': 'IN', 'sunrise': 1753143707, 'sunset': 1753189711}, 'timezone': 19800, 'id': 1264527, 'name': 'Chenna
i', 'cod': 200}
{'coord': {'lon': 77.6033, 'lat': 12.9762}, 'weather': [{'id': 500, 'main': 'Rain', 'description': 'light rain', 'icon': '10n'}], 'base': 'stations', 'main': {'temp': 292.97, 'feels_l
ike': 293.41, 'temp_min': 292.65, 'temp_max': 293.43, 'pressure': 1009, 'humidity': 92, 'sea_level': 1009, 'grnd_level': 910}, 'visibility': 6000, 'wind': {'speed': 3.6, 'deg': 260},
'rain': {'1h': 0.27}, 'clouds': {'all': 75}, 'dt': 1753132817, 'sys': {'type': 1, 'id': 9205, 'country': 'IN', 'sunrise': 1753144360, 'sunset': 1753190343}, 'timezone': 19800, 'id': 1
277333, 'name': 'Bengaluru', 'cod': 200}
{'coord': {'lon': 88.3697, 'lat': 22.5697}, 'weather': [{'id': 721, 'main': 'Haze', 'description': 'haze', 'icon': '50n'}], 'base': 'stations', 'main': {'temp': 302.12, 'feels_like':
309.12, 'temp_min': 302.12, 'temp_max': 302.12, 'pressure': 1000, 'humidity': 89, 'sea_level': 1000, 'grnd_level': 999}, 'visibility': 3500, 'wind': {'speed': 1.54, 'deg': 210}, 'clou
ds': {'all': 40}, 'dt': 1753132900, 'sys': {'type': 1, 'id': 9114, 'country': 'IN', 'sunrise': 1753140813, 'sunset': 1753188722}, 'timezone': 19800, 'id': 1275004, 'name': 'Kolkata',
'cod': 200}
{'coord': {'lon': 73.8553, 'lat': 18.5196}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02n'}], 'base': 'stations', 'main': {'temp': 297.26, 'feels
_like': 298.18, 'temp_min': 297.26, 'temp_max': 297.26, 'pressure': 1005, 'humidity': 94, 'sea_level': 1005, 'grnd_level': 933}, 'visibility': 10000, 'wind': {'speed': 3.3, 'deg': 260
, 'gust': 6.55}, 'clouds': {'all': 23}, 'dt': 1753133078, 'sys': {'type': 2, 'id': 2104163, 'country': 'IN', 'sunrise': 1753144718, 'sunset': 1753191783}, 'timezone': 19800, 'id': 125
9229, 'name': 'Pune', 'cod': 200}
{'coord': {'lon': 78.4744, 'lat': 17.3753}, 'weather': [{'id': 721, 'main': 'Haze', 'description': 'haze', 'icon': '50n'}], 'base': 'stations', 'main': {'temp': 296.88, 'feels_like':
297.76, 'temp_min': 296.88, 'temp_max': 296.88, 'pressure': 1005, 'humidity': 94, 'sea_level': 1005, 'grnd_level': 940}, 'visibility': 5000, 'wind': {'speed': 6.17, 'deg': 260}, 'clou
ds': {'all': 75}, 'dt': 1753132816, 'sys': {'type': 1, 'id': 9213, 'country': 'IN', 'sunrise': 1753143724, 'sunset': 1753190560}, 'timezone': 19800, 'id': 1269843, 'name': 'Hyderabad'
, 'cod': 200}
{'coord': {'lon': 76.2602, 'lat': 9.9399}, 'weather': [{'id': 804, 'main': 'Clouds', 'description': 'overcast clouds', 'icon': '04n'}], 'base': 'stations', 'main': {'temp': 298.14, 'f
eels_like': 299.12, 'temp_min': 298.14, 'temp_max': 298.14, 'pressure': 1008, 'humidity': 93, 'sea_level': 1008, 'grnd_level': 1008}, 'visibility': 7168, 'wind': {'speed': 6.55, 'deg'
: 303, 'gust': 9.08}, 'clouds': {'all': 100}, 'dt': 1753133124, 'sys': {'type': 1, 'id': 9211, 'country': 'IN', 'sunrise': 1753144966, 'sunset': 1753190381}, 'timezone': 19800, 'id':
1273874, 'name': 'Kochi', 'cod': 200}
█
```



Data stream weather-stream successfully created.

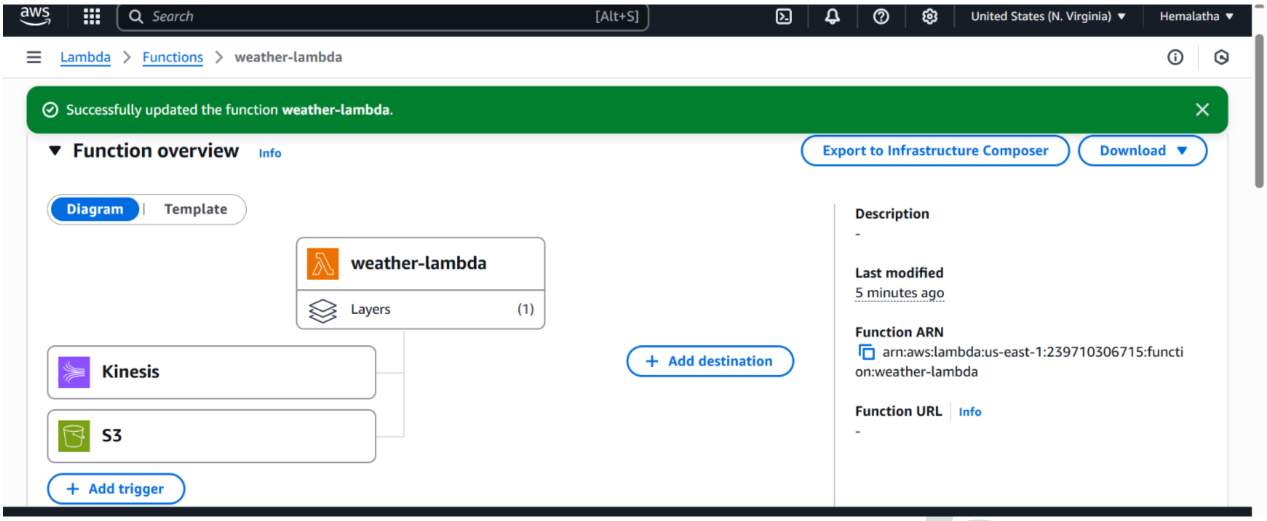
weather-stream Info

Data stream summary

Status	Capacity mode	ARN	Creation time
Active	Provisioned	arn:aws:kinesis:us-east-1:239710306715:stream/weather-stream	July 21, 2025 at 17:04 EDT
	Data retention period 1 day		

3. **AWS Lambda Trigger:**

A Lambda function processes each record and stores in the **bronze** S3 bucket and performs queries for data using lambda function, results in cleaned data stored in s3 **silver** bucket.



aws
Search [Alt+S]
United States (N. Virginia) ▾
Hemalatha ▾
Lambda > Functions > weather-lambda

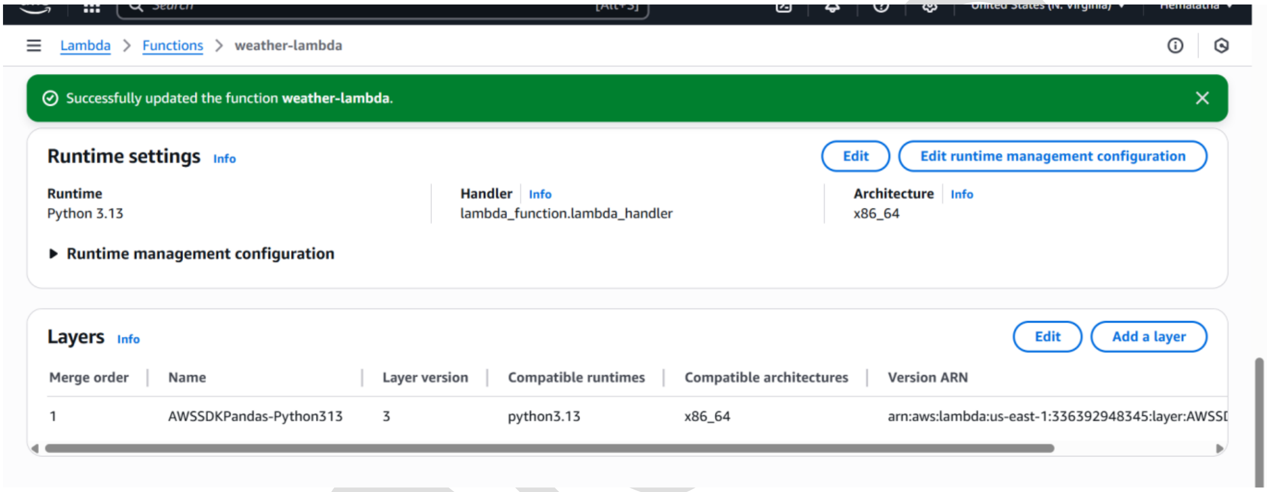
Successfully updated the function **weather-lambda**.

Description
-

Last modified
5 minutes ago

Function ARN
arn:aws:lambda:us-east-1:239710306715:function:weather-lambda

Function URL Info
-



Lambda > Functions > weather-lambda

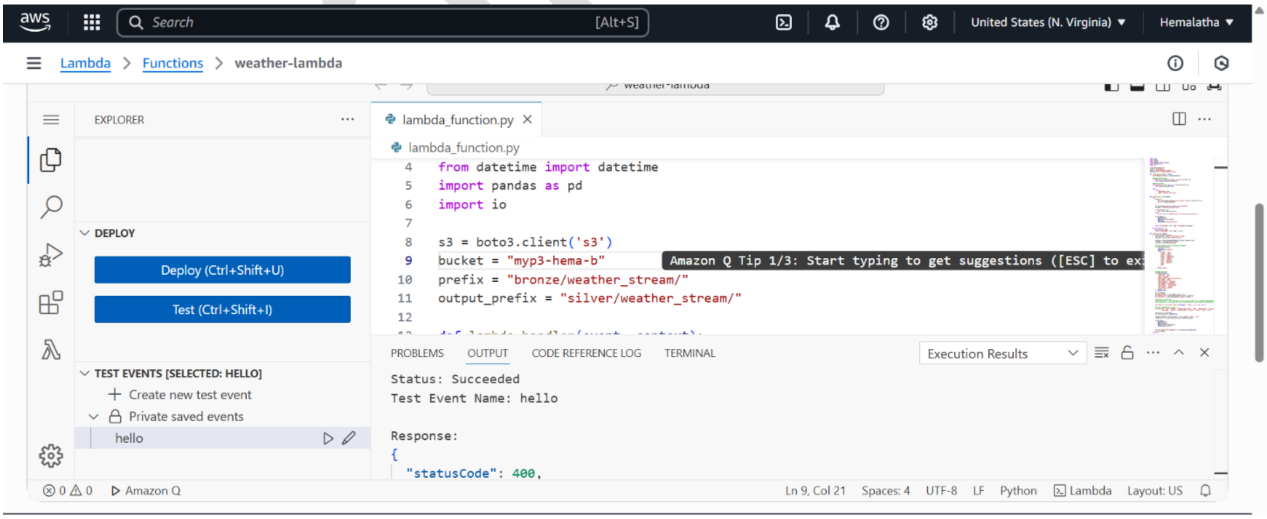
Successfully updated the function **weather-lambda**.

Runtime settings Info
Edit
Edit runtime management configuration

Runtime	Handler	Architecture
Python 3.13	lambda_function.lambda_handler	x86_64

▸ Runtime management configuration
Layers Info
Edit
Add a layer

Merge order	Name	Layer version	Compatible runtimes	Compatible architectures	Version ARN
1	AWSSDKPandas-Python313	3	python3.13	x86_64	arn:aws:lambda:us-east-1:336392948345:layer:AWSSD



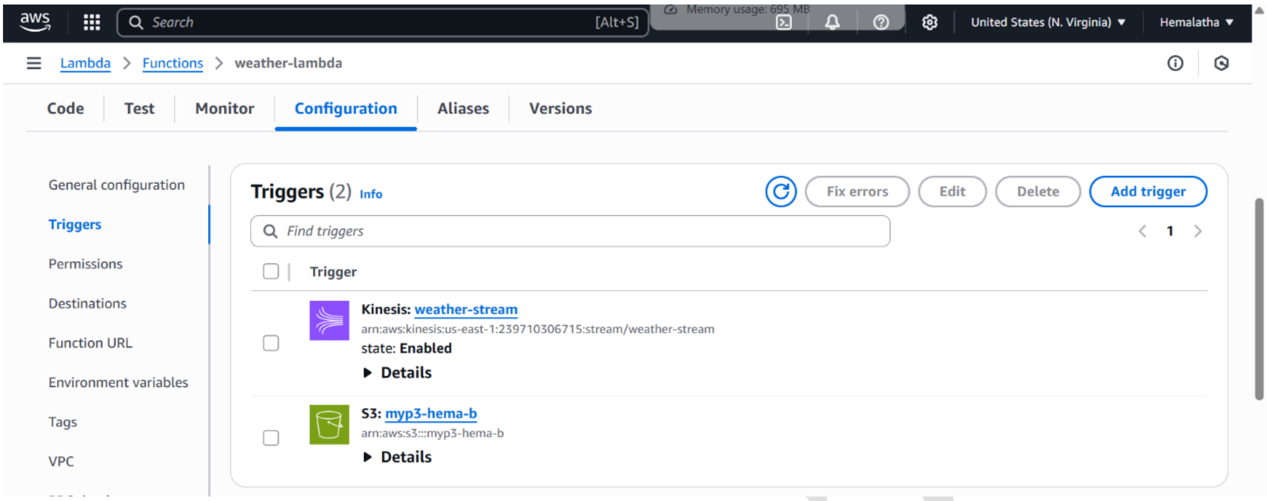
```python
from datetime import datetime
import pandas as pd
import io

s3 = boto3.client('s3')
bucket = "myp3-hema-b"
prefix = "bronze/weather_stream/"
output_prefix = "silver/weather_stream/"
```

Amazon Q Tip 1/3: Start typing to get suggestions ([ESC] to ex

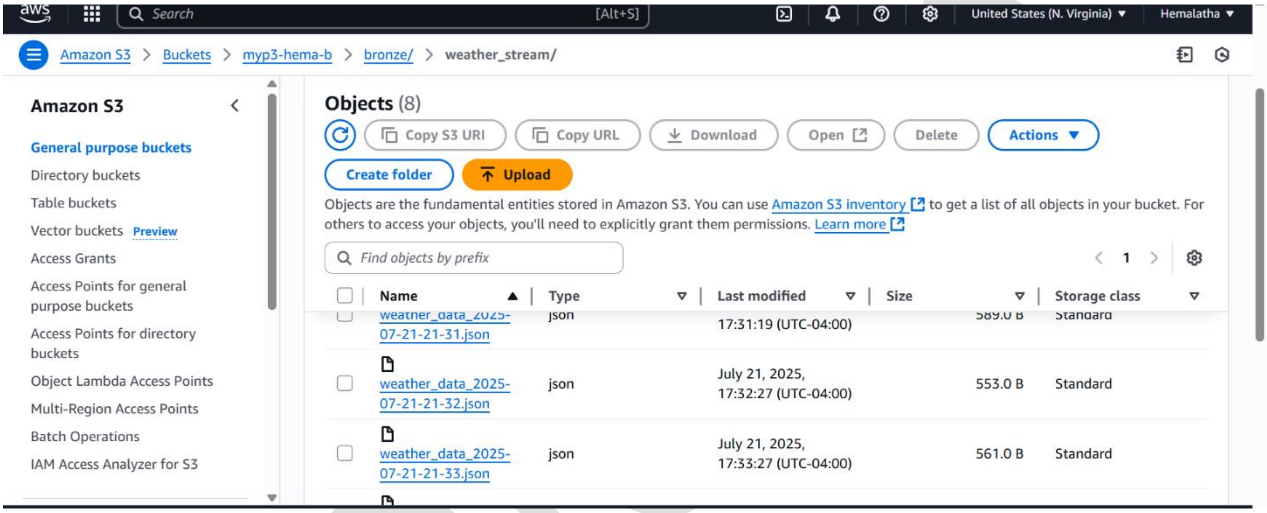
```
Status: Succeeded
Test Event Name: hello

Response:
{
  "statusCode": 400,
```



Triggers (2) Info

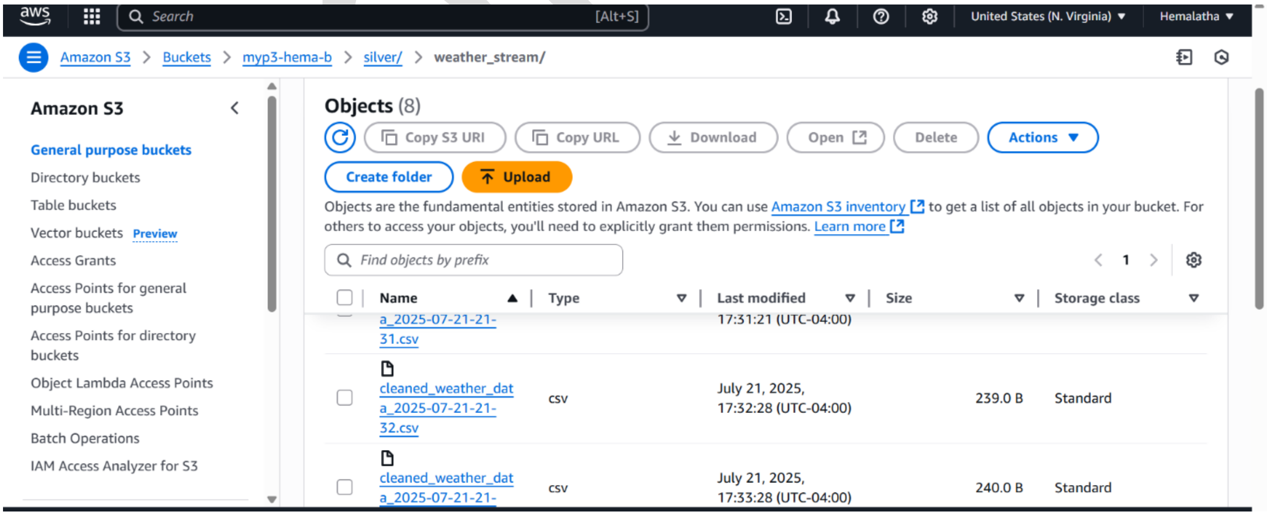
□	Trigger
□	**Kinesis:** weather-stream arn:aws:kinesis:us-east-1:239710306715:stream/weather-stream state: **Enabled** ▶ **Details**
□	**S3:** myp3-hema-b arn:aws:s3:::myp3-hema-b ▶ **Details**


aws
Search [Alt+S]
United States (N. Virginia) ▼
Hemalatha ▼
Amazon S3 > Buckets > myp3-hema-b > bronze/ > weather_stream/
Amazon S3
General purpose buckets
Directory buckets
Table buckets
Vector buckets Preview
Access Grants
Access Points for general purpose buckets
Access Points for directory buckets
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
IAM Access Analyzer for S3
Copy S3 URI
Copy URL
Download
Open
Delete
Actions ▼
Create folder
Upload
Find objects by prefix
1

Objects (8)

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

□	Name ▲	Type ▽	Last modified ▽	Size ▽	Storage class ▽
□	weather_data_2025-07-21-21-31.json	json	17:31:19 (UTC-04:00)	589.0 B	Standard
□	weather_data_2025-07-21-21-32.json	json	July 21, 2025, 17:32:27 (UTC-04:00)	553.0 B	Standard
□	weather_data_2025-07-21-21-33.json	json	July 21, 2025, 17:33:27 (UTC-04:00)	561.0 B	Standard

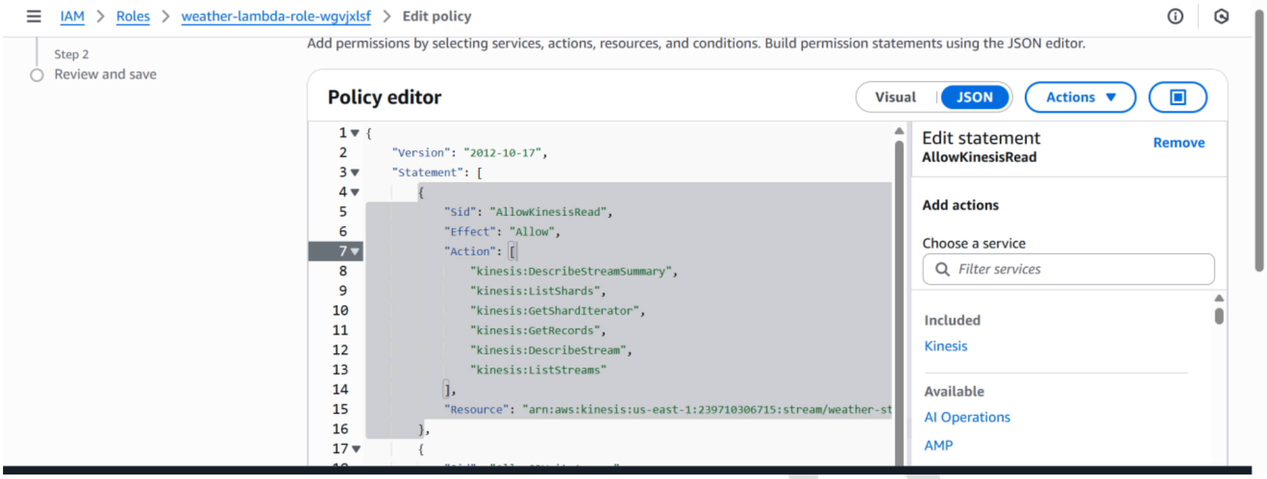

aws
Search [Alt+S]
United States (N. Virginia) ▼
Hemalatha ▼
Amazon S3 > Buckets > myp3-hema-b > silver/ > weather_stream/
Amazon S3
General purpose buckets
Directory buckets
Table buckets
Vector buckets Preview
Access Grants
Access Points for general purpose buckets
Access Points for directory buckets
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
IAM Access Analyzer for S3
Copy S3 URI
Copy URL
Download
Open
Delete
Actions ▼
Create folder
Upload
Find objects by prefix
1

Objects (8)

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

□	Name ▲	Type ▽	Last modified ▽	Size ▽	Storage class ▽
	a_2025-07-21-21-31.csv		17:31:21 (UTC-04:00)		
□	cleaned_weather_data_2025-07-21-21-32.csv	csv	July 21, 2025, 17:32:28 (UTC-04:00)	239.0 B	Standard
□	cleaned_weather_data_2025-07-21-21-	csv	July 21, 2025, 17:33:28 (UTC-04:00)	240.0 B	Standard

4. IAM Roles:


☰ IAM > Roles > weather-lambda-role-wgvjxlsf > Edit policy
Step 2
○ Review and save

Add permissions by selecting services, actions, resources, and conditions. Build permission statements using the JSON editor.

Policy editor

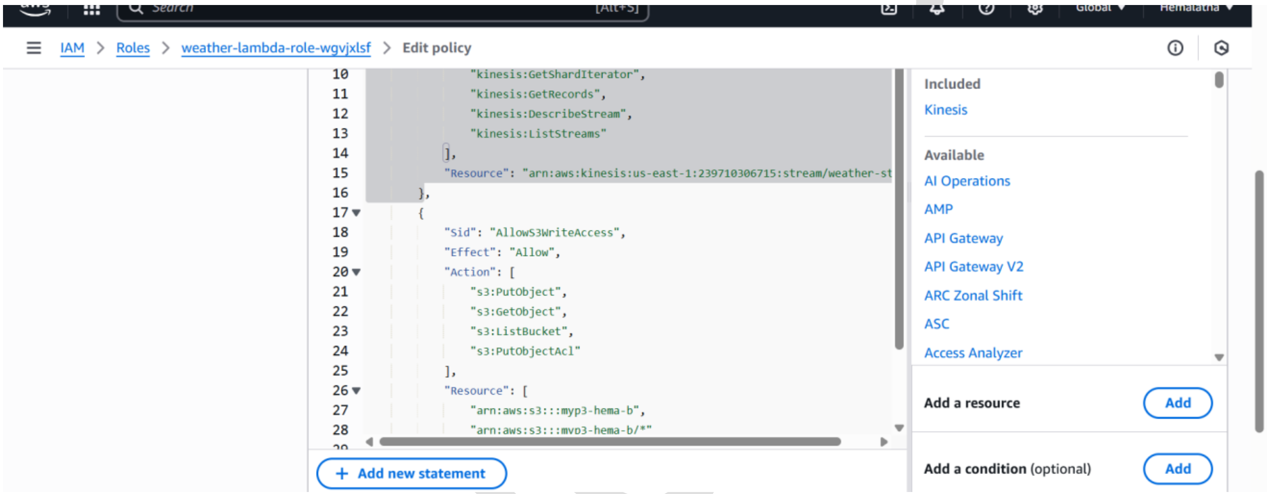
Visual | JSON | Actions ▼

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowKinesisRead",
            "Effect": "Allow",
            "Action": [
                "kinesis:DescribeStreamSummary",
                "kinesis:ListShards",
                "kinesis:GetShardIterator",
                "kinesis:GetRecords",
                "kinesis:DescribeStream",
                "kinesis:ListStreams"
            ],
            "Resource": "arn:aws:kinesis:us-east-1:239710306715:stream/weather-st
        },
        {
```

Edit statement
AllowKinesisRead

Add actions

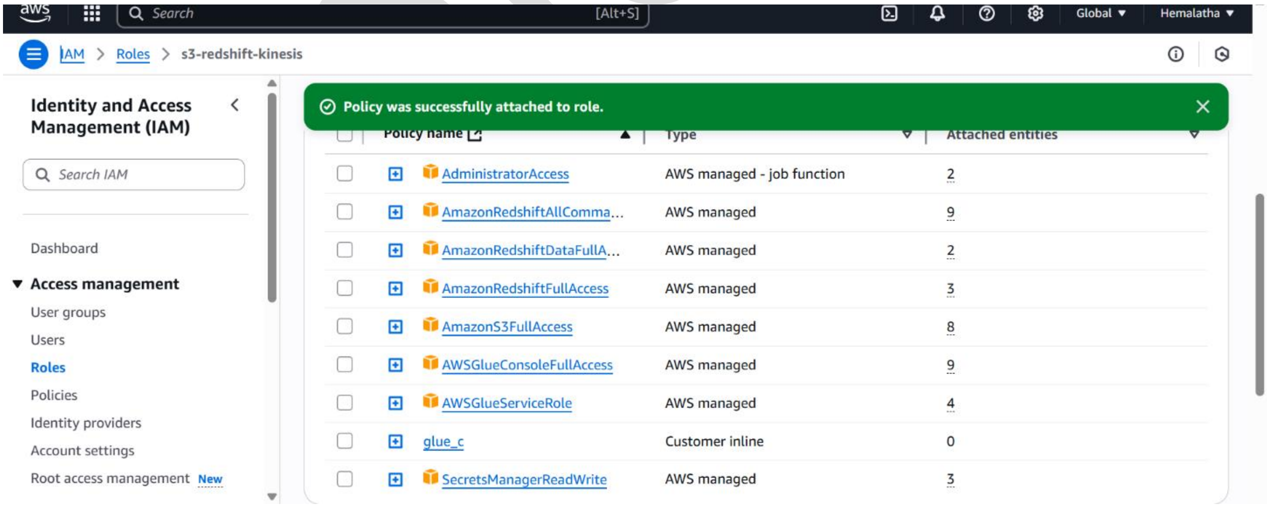
Choose a service


aws | Search [Alt+S] | Global ▼ | Hemalatha ▼
☰ IAM > Roles > weather-lambda-role-wgvjxlsf > Edit policy

```
                "kinesis:GetShardIterator",
                "kinesis:GetRecords",
                "kinesis:DescribeStream",
                "kinesis:ListStreams"
            ],
            "Resource": "arn:aws:kinesis:us-east-1:239710306715:stream/weather-st
        },
        {
            "Sid": "AllowS3WriteAccess",
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:ListBucket",
                "s3:PutObjectAcl"
            ],
            "Resource": [
                "arn:aws:s3:::myp3-hema-b",
                "arn:aws:s3:::mvp3-hema-b/*"
```

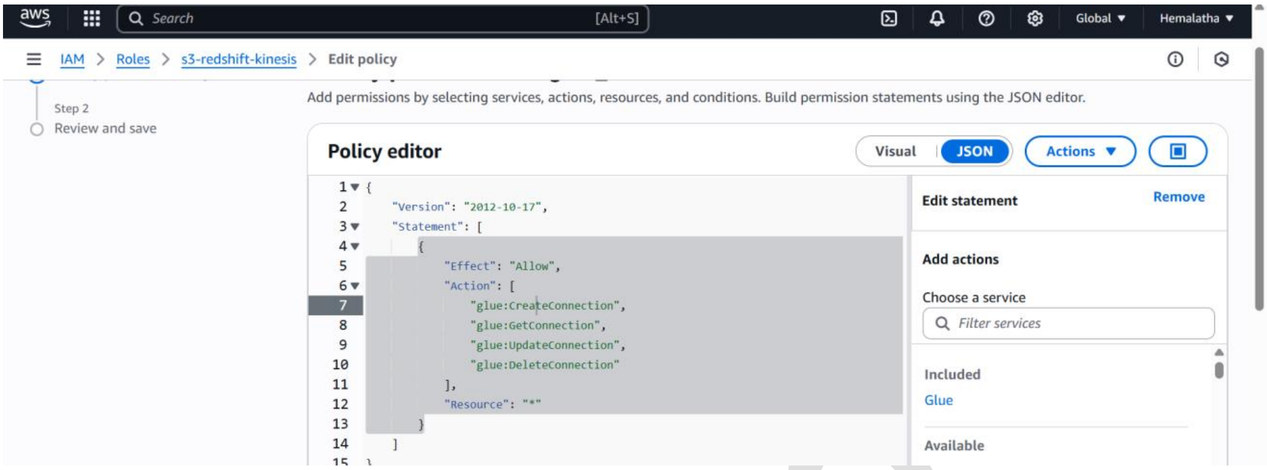
Add a resource

Add a condition (optional)


aws | Search [Alt+S] | Global ▼ | Hemalatha ▼
☰ IAM > Roles > s3-redshift-kinesis
Identity and Access Management (IAM)
🔍 Search IAM
Dashboard
▼ Access management
User groups
Users
Roles
Policies
Identity providers
Account settings
Root access management New

Policy was successfully attached to role.

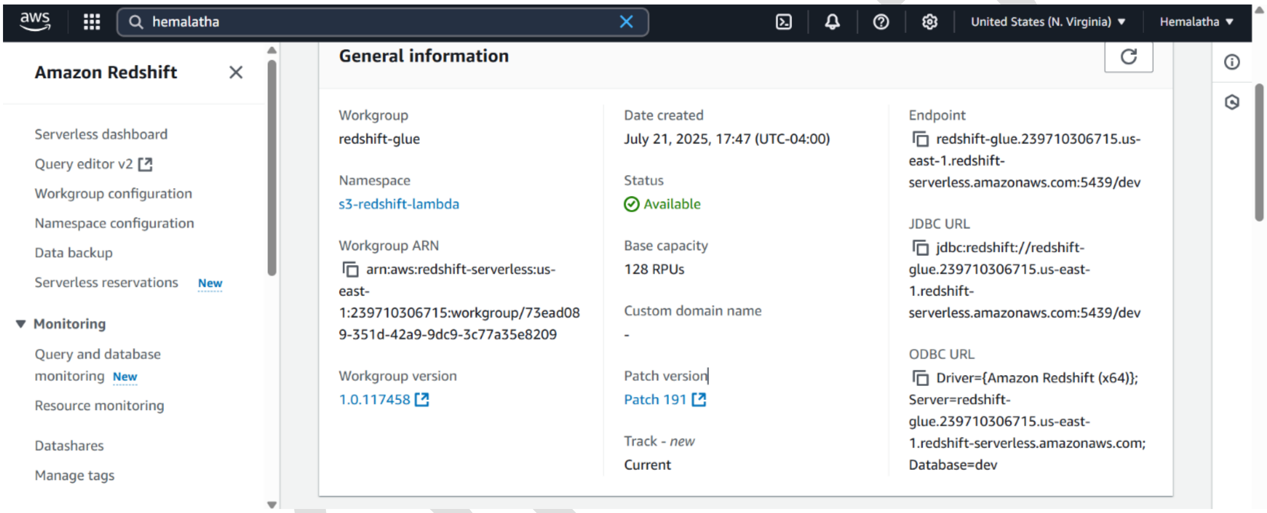
□	Policy name ▲	Type	Attached entities
□	AdministratorAccess	AWS managed - job function	2
□	AmazonRedshiftAllComma...	AWS managed	9
□	AmazonRedshiftDataFullA...	AWS managed	2
□	AmazonRedshiftFullAccess	AWS managed	3
□	AmazonS3FullAccess	AWS managed	8
□	AWSGlueConsoleFullAccess	AWS managed	9
□	AWSGlueServiceRole	AWS managed	4
□	glue_c	Customer inline	0
□	SecretsManagerReadWrite	AWS managed	3



Add permissions by selecting services, actions, resources, and conditions. Build permission statements using the JSON editor.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:CreateConnection",
                "glue:GetConnection",
                "glue:UpdateConnection",
                "glue:DeleteConnection"
            ],
            "Resource": "*"
        }
    ]
}
```

5. Redshift: Create a table and schema that connects with glue job script



General information

Workgroup
redshift-glue

Namespace
s3-redshift-lambda

Workgroup ARN
arn:aws:redshift-serverless:us-east-1:239710306715:workgroup/73ead089-351d-42a9-9dc9-3c77a35e8209

Workgroup version
1.0.117458

Date created
July 21, 2025, 17:47 (UTC-04:00)

Status
Available

Base capacity
128 RPUs

Custom domain name
-

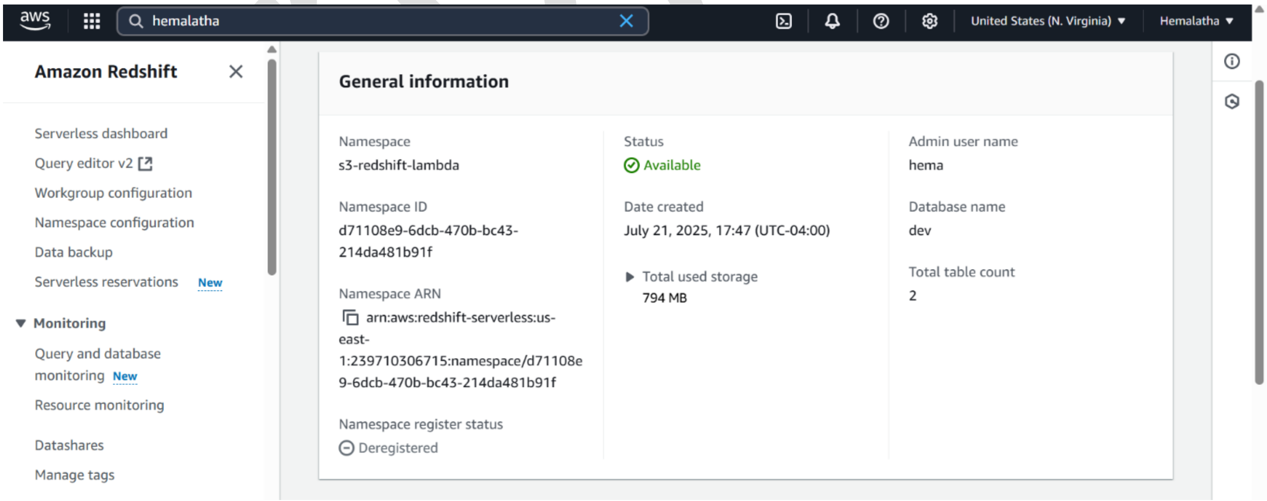
Patch version
Patch 191

Track - *new*
Current

Endpoint
redshift-glue.239710306715.us-east-1.redshift-serverless.amazonaws.com:5439/dev

JDBC URL
jdbc:redshift://redshift-glue.239710306715.us-east-1.redshift-serverless.amazonaws.com:5439/dev

ODBC URL
Driver={Amazon Redshift (x64)}; Server=redshift-glue.239710306715.us-east-1.redshift-serverless.amazonaws.com; Database=dev



General information

Namespace
s3-redshift-lambda

Namespace ID
d71108e9-6dcb-470b-bc43-214da481b91f

Namespace ARN
arn:aws:redshift-serverless:us-east-1:239710306715:namespace/d71108e9-6dcb-470b-bc43-214da481b91f

Namespace register status
Deregistered

Status
Available

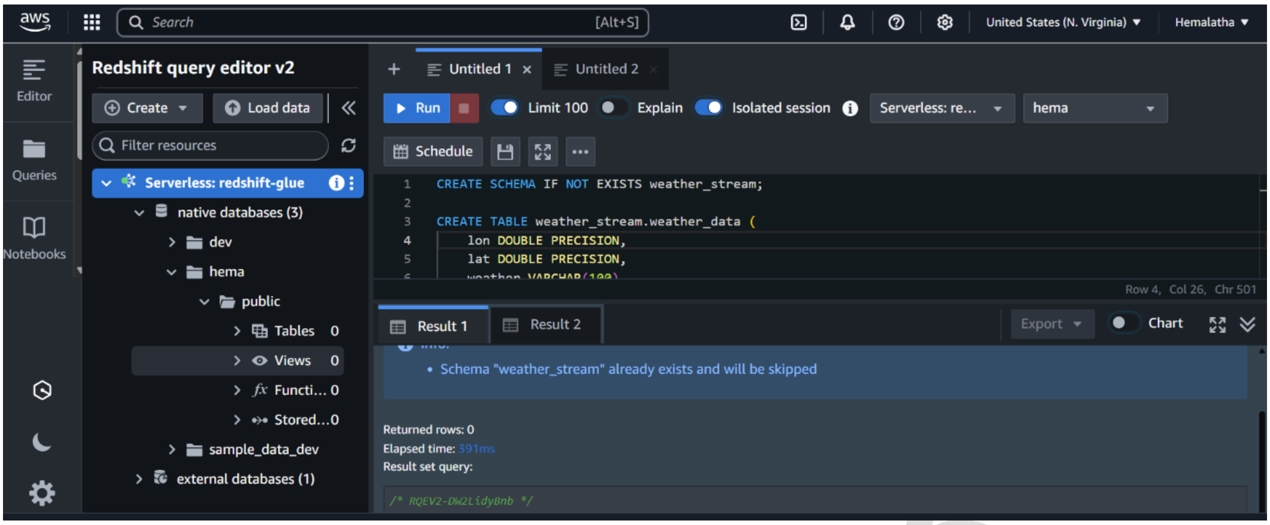
Date created
July 21, 2025, 17:47 (UTC-04:00)

▶ Total used storage
794 MB

Admin user name
hema

Database name
dev

Total table count
2



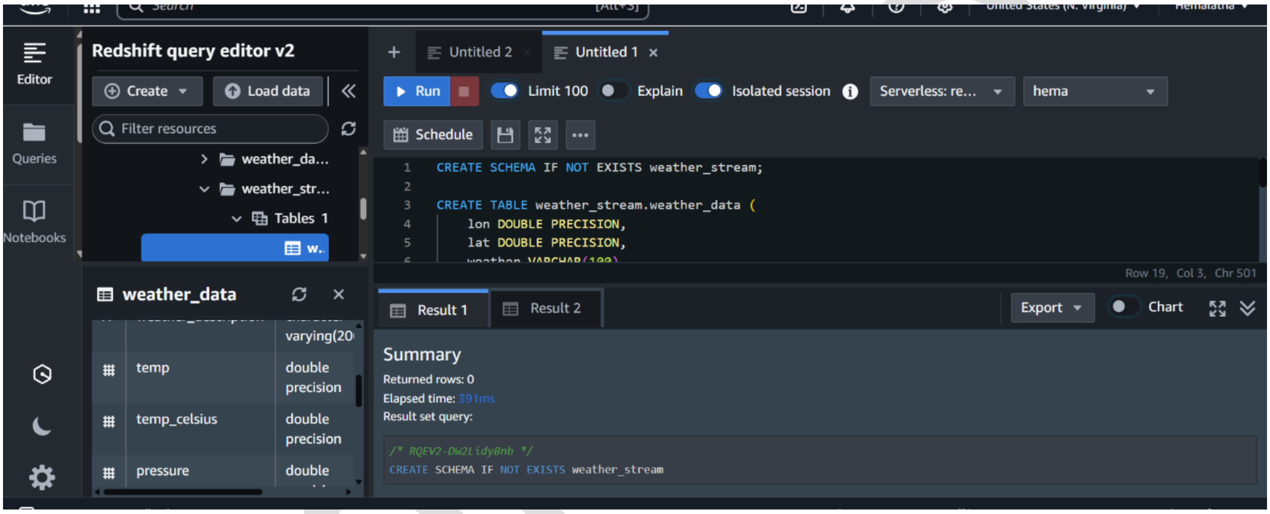
```
CREATE SCHEMA IF NOT EXISTS weather_stream;

CREATE TABLE weather_stream.weather_data (
    lon DOUBLE PRECISION,
    lat DOUBLE PRECISION,
    weather VARCHAR(100),
```

- Schema "weather_stream" already exists and will be skipped

Returned rows: 0
Elapsed time: 391ms
Result set query:

```
/* RQEV2-DW2LidyBnb */
```



```
CREATE SCHEMA IF NOT EXISTS weather_stream;

CREATE TABLE weather_stream.weather_data (
    lon DOUBLE PRECISION,
    lat DOUBLE PRECISION,
    weather VARCHAR(100),
```

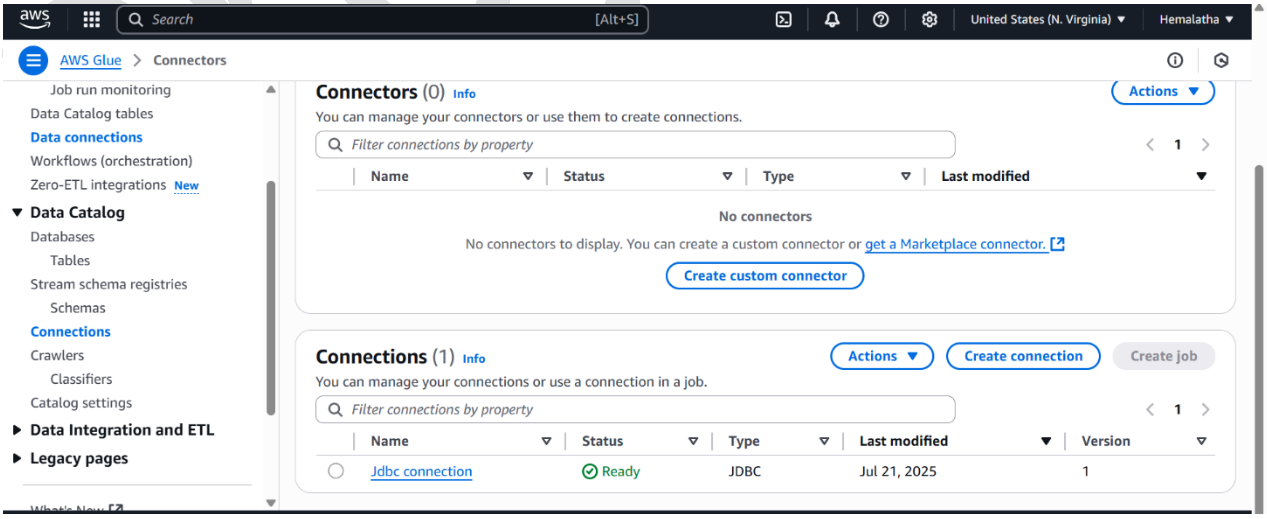
weather_data

		varying(20
#	temp	double precision
#	temp_celsius	double precision
#	pressure	double

Summary
Returned rows: 0
Elapsed time: 391ms
Result set query:

```
/* RQEV2-DW2LidyBnb */
CREATE SCHEMA IF NOT EXISTS weather_stream
```

6. JDBC connection:



Connectors (0) Info

You can manage your connectors or use them to create connections.

Filter connections by property

Name	Status	Type	Last modified

No connectors

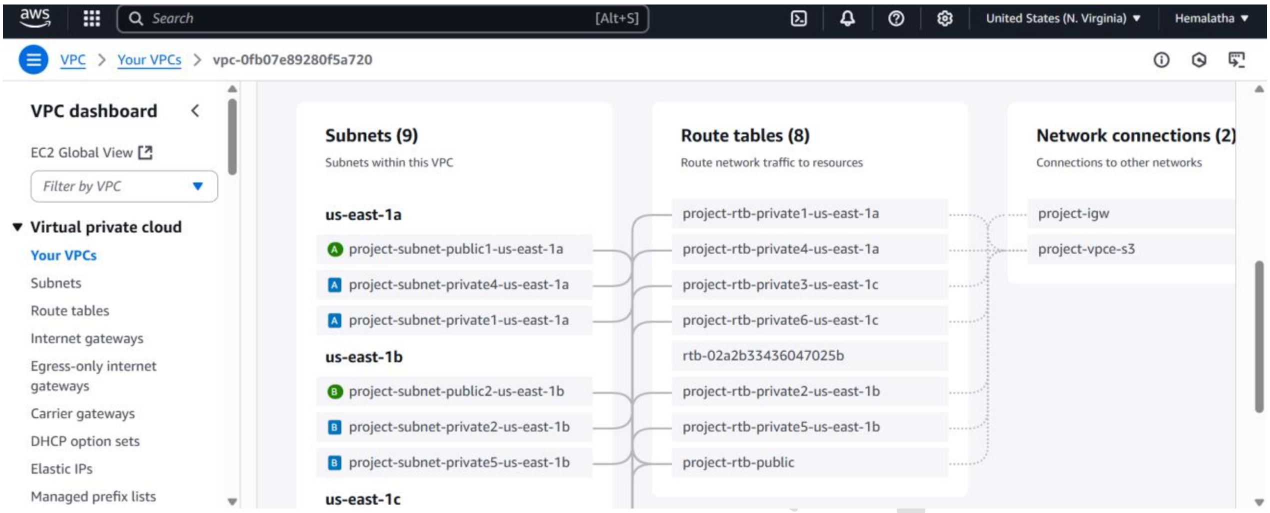
No connectors to display. You can create a custom connector or get a Marketplace connector.

Connections (1) Info

You can manage your connections or use a connection in a job.

Filter connections by property

	Name	Status	Type	Last modified	Version
○	Jdbc connection	✓ Ready	JDBC	Jul 21, 2025	1



VPC > Your VPCs > vpc-0fb07e89280f5a720

VPC dashboard

EC2 Global View

Filter by VPC

Virtual private cloud
- Your VPCs
- Subnets
- Route tables
- Internet gateways
- Egress-only internet gateways
- Carrier gateways
- DHCP option sets
- Elastic IPs
- Managed prefix lists

Subnets (9)
Subnets within this VPC

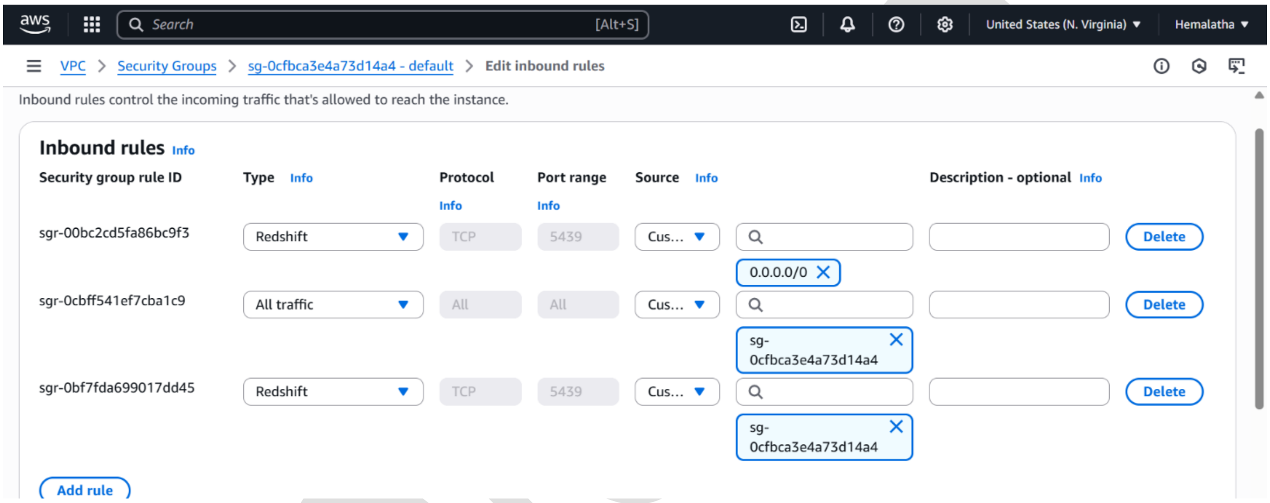
us-east-1a
- project-subnet-public1-us-east-1a
- project-subnet-private4-us-east-1a
- project-subnet-private1-us-east-1a

us-east-1b
- project-subnet-public2-us-east-1b
- project-subnet-private2-us-east-1b
- project-subnet-private5-us-east-1b

us-east-1c

Route tables (8)
Route network traffic to resources
- project-rtb-private1-us-east-1a
- project-rtb-private4-us-east-1a
- project-rtb-private3-us-east-1c
- project-rtb-private6-us-east-1c
- rtb-02a2b33436047025b
- project-rtb-private2-us-east-1b
- project-rtb-private5-us-east-1b
- project-rtb-public

Network connections (2)
Connections to other networks
- project-igw
- project-vpce-s3



VPC > Security Groups > sg-0cfbca3e4a73d14a4 - default > Edit inbound rules

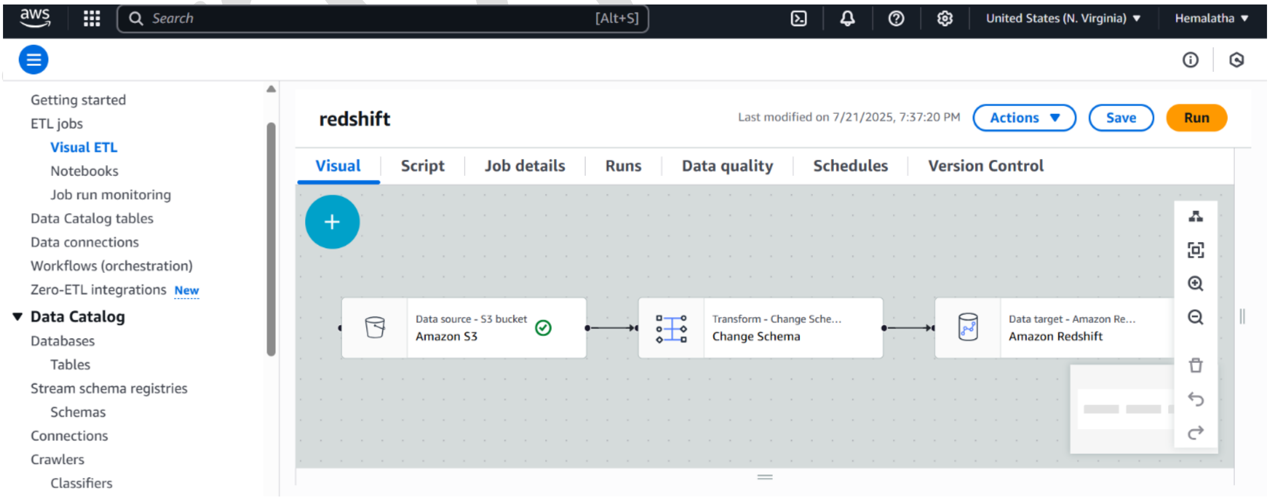
Inbound rules control the incoming traffic that's allowed to reach the instance.

Inbound rules Info

Security group rule ID	Type	Protocol	Port range	Source		Description - optional
sgr-00bc2cd5fa86bc9f3	Redshift	TCP	5439	Cus...	0.0.0.0/0	
sgr-0cbff541ef7cba1c9	All traffic	All	All	Cus...	sg-0cfbca3e4a73d14a4	
sgr-0bf7fda699017dd45	Redshift	TCP	5439	Cus...	sg-0cfbca3e4a73d14a4	

Add rule

7. Glue job: connect s3 -> change schema -> redshift



- Getting started
- ETL jobs
  - Visual ETL
  - Notebooks
  - Job run monitoring
- Data Catalog tables
- Data connections
- Workflows (orchestration)
- Zero-ETL integrations New
- Data Catalog
  - Databases
    - Tables
  - Stream schema registries
    - Schemas
  - Connections
  - Crawlers
    - Classifiers

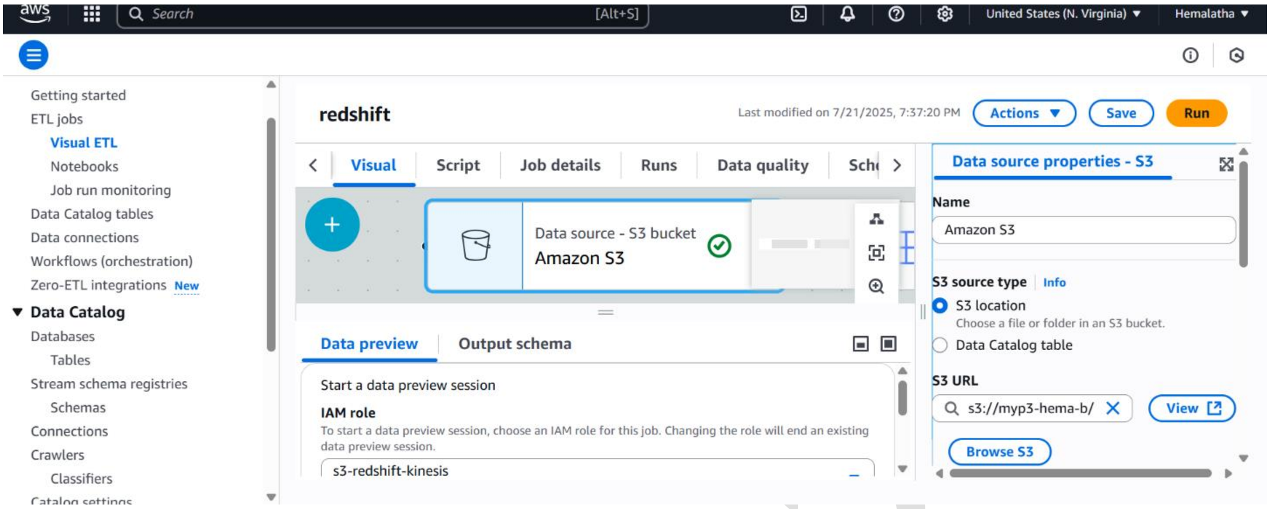
redshift

Last modified on 7/21/2025, 7:37:20 PM

Actions | Save | Run

Visual | Script | Job details | Runs | Data quality | Schedules | Version Control

Data source - S3 bucket: Amazon S3 → Transform - Change Sche...: Change Schema → Data target - Amazon Re...: Amazon Redshift



redshift

Last modified on 7/21/2025, 7:37:20 PM

Start a data preview session

IAM role

To start a data preview session, choose an IAM role for this job. Changing the role will end an existing data preview session.

s3-redshift-kinesis

Data source properties - S3

Name

Amazon S3

S3 source type Info

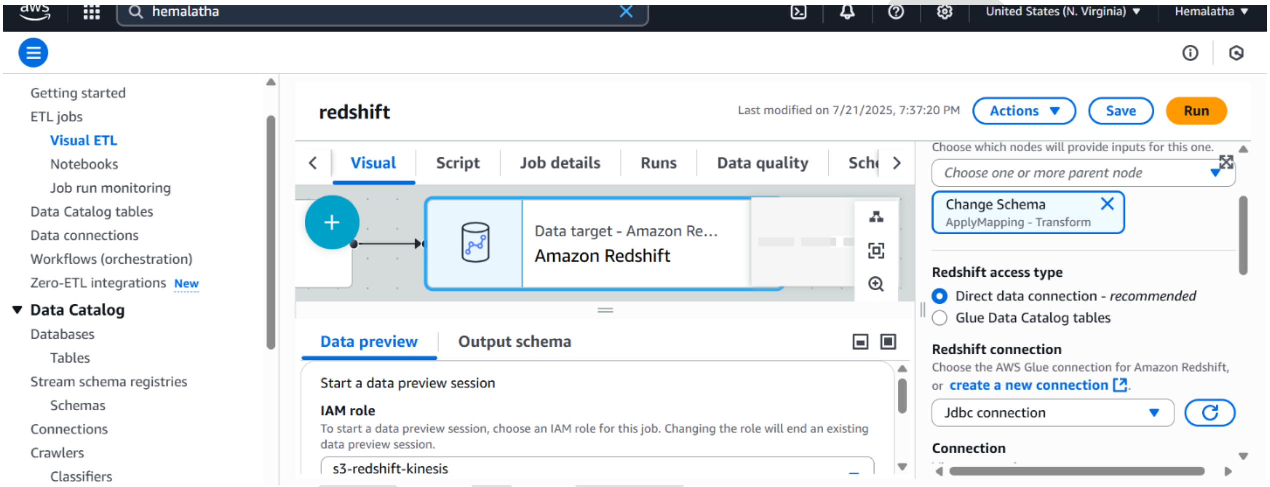
● S3 location

Choose a file or folder in an S3 bucket.

○ Data Catalog table

S3 URL

s3://myp3-hema-b/



redshift

Last modified on 7/21/2025, 7:37:20 PM

Start a data preview session

IAM role

To start a data preview session, choose an IAM role for this job. Changing the role will end an existing data preview session.

s3-redshift-kinesis

Choose which nodes will provide inputs for this one.

Choose one or more parent node

Change Schema
ApplyMapping - Transform

Redshift access type

● Direct data connection - *recommended*

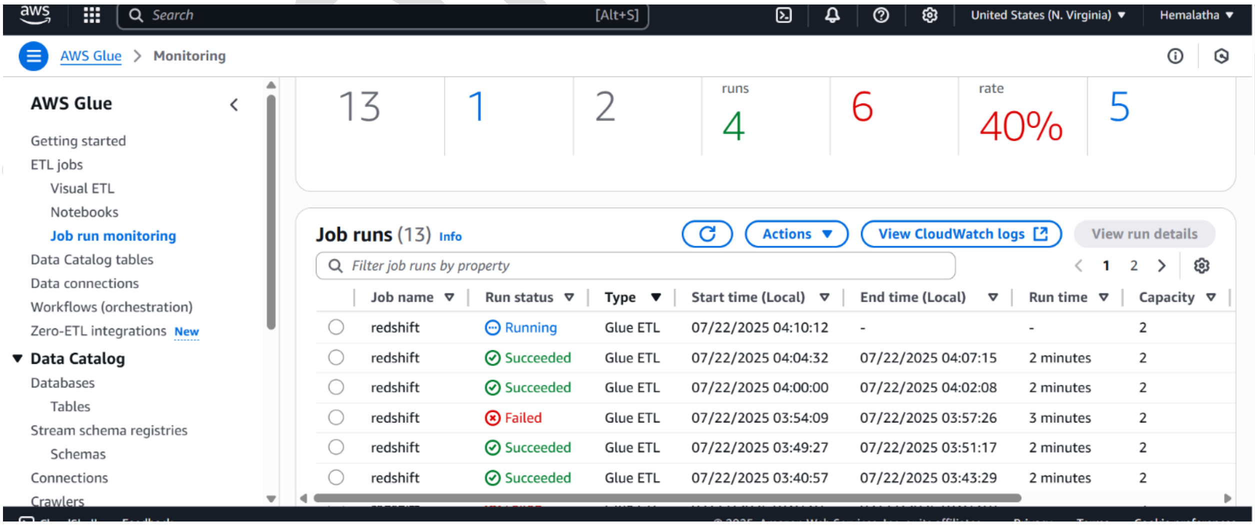
○ Glue Data Catalog tables

Redshift connection

Choose the AWS Glue connection for Amazon Redshift, or create a new connection.

Jdbc connection

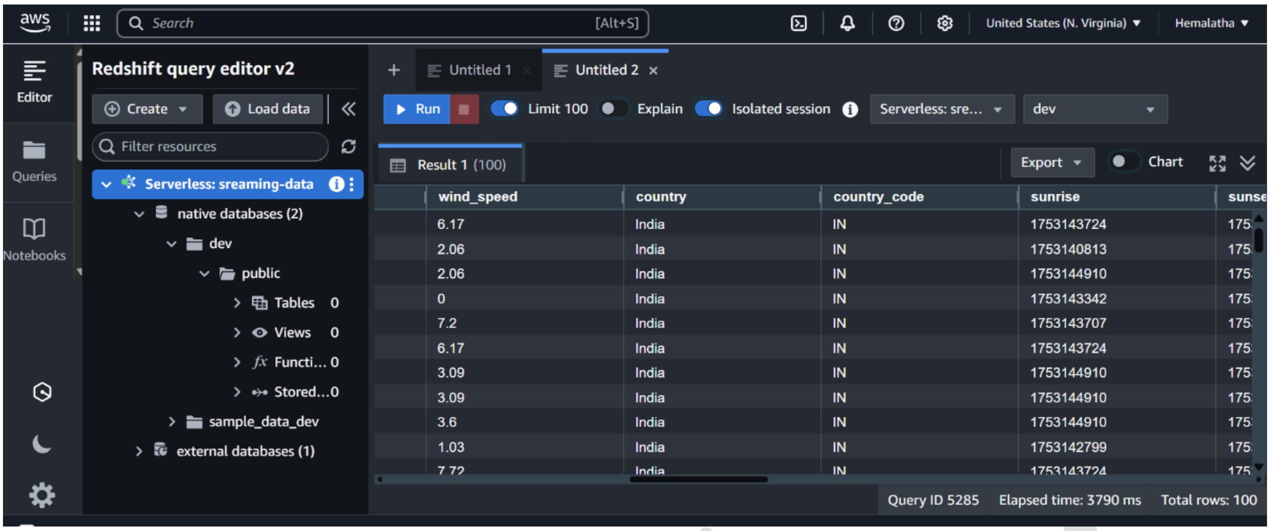
Connection



Job runs (13) Info

	Job name	Run status	Type	Start time (Local)	End time (Local)	Run time	Capacity
○	redshift	Running	Glue ETL	07/22/2025 04:10:12	-	-	2
○	redshift	✓ Succeeded	Glue ETL	07/22/2025 04:04:32	07/22/2025 04:07:15	2 minutes	2
○	redshift	✓ Succeeded	Glue ETL	07/22/2025 04:00:00	07/22/2025 04:02:08	2 minutes	2
○	redshift	☒ Failed	Glue ETL	07/22/2025 03:54:09	07/22/2025 03:57:26	3 minutes	2
○	redshift	✓ Succeeded	Glue ETL	07/22/2025 03:49:27	07/22/2025 03:51:17	2 minutes	2
○	redshift	✓ Succeeded	Glue ETL	07/22/2025 03:40:57	07/22/2025 03:43:29	2 minutes	2

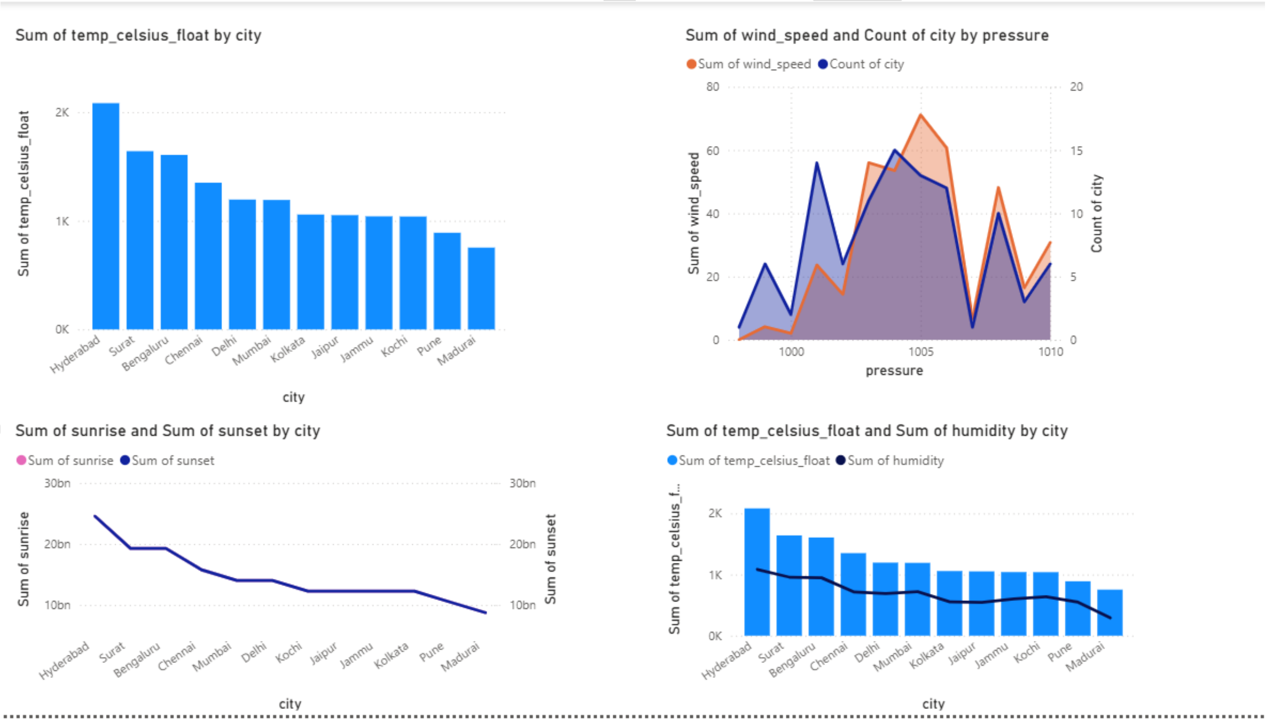
8. Output:



wind_speed	country	country_code	sunrise	sunse
6.17	India	IN	1753143724	175
2.06	India	IN	1753140813	175
2.06	India	IN	1753144910	175
0	India	IN	1753143342	175
7.2	India	IN	1753143707	175
6.17	India	IN	1753143724	175
3.09	India	IN	1753144910	175
3.09	India	IN	1753144910	175
3.6	India	IN	1753144910	175
1.03	India	IN	1753142799	175
7.72	India	IN	1753143724	175

Query ID 5285 Elapsed time: 3790 ms Total rows: 100

9. Visualisation:



Sum of temp_celsius_float by city

Sum of wind_speed and Count of city by pressure

Sum of sunrise and Sum of sunset by city

Sum of temp_celsius_float and Sum of humidity by city