Assignment10: Ingest daily flight data in Redshift table
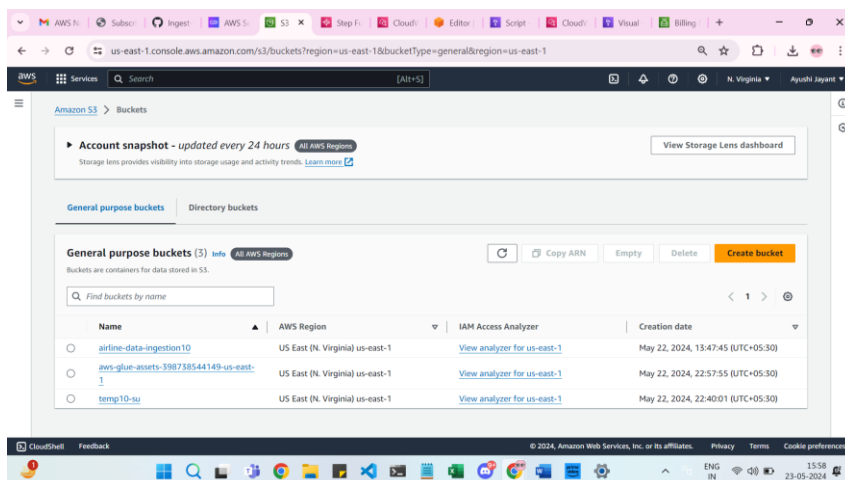
1)created s3 bucket with folders as airport_dim and flights_daily and uploaded the dataset into respective folders

2)Created redshift cluster along with proper Iam role and access key for redshift using secrets manager

3)We have given VPC connection for redshift and setup inbound and outbound rules along with endpoint as s3 for redshift port-5439

4)Setup all the relevant databases creation, crawlers creation(3 schemas)(2 redshift and 1 in s3 bucket) and created connections in glue for redshift and then prepared the visual etl

5)created state machine in step function to orchestrate pipeline on arrival of data in s3 bucket and set up sns on notification related to success or failure through mail

6)The data arrived in destination that is redshift editor where we can query the daily_flights_data that arrived after being processed successfully

Screenshot 1 — Redshift Cluster Details

us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1#/cluster-details?cluster=redshift-cluster-10

General information Info

Cluster identifier
redshift-cluster-10

Status
✓ Available

Node type
dc2.large

Endpoint
redshift-cluster-10.cjfqvqjuhxcw.us-east-1.redshift.amazonaws.com:5439/dev

Custom domain name
-

Date created
May 22, 2024, 15:29 (UTC+05:30)

Number of nodes
2

JDBC URL
jdbc:redshift://redshift-cluster-10.cjfqvqjuhxcw.us-east-1.redshift.amazonaws.com:5439/dev

Cluster namespace ARN
arn:aws:redshift:us-east-1:398738544149:namespace:256f5451-9358-4a85-af60-e9723e54d9b8

Storage used
0.04% (0.13 of 320 GB used)

Patch version
Patch 180

ODBC URL
Driver={Amazon Redshift (x64)};
Server=redshift-cluster-10.cjfqvqjuhxcw.us-east-1.redshift.amazonaws.com;
Database=dev

Multi-AZ
No

Cluster configuration
Production

Cluster performance | Query monitoring | Databases | Datashares | Zero-ETL integrations | Resource Policy | Schedules | Maintenance | Properties

▶ **Recommendations** (0)
To improve performance and decrease operating costs, the Amazon Redshift Advisor provides recommendations.

Screenshot 2 — AWS Secrets Manager

us-east-1.console.aws.amazon.com/secretsmanager/secret?name=Used-for-Redshift&region=us-east-1

AWS Secrets Manager > Secrets > Used-for-Redshift

Used-for-Redshift

Secret details

Encryption key
aws/secretsmanager

Secret description
-

Secret name
Used-for-Redshift

Secret ARN
arn:aws:secretsmanager:us-east-1:398738544149:secret:Used-for-Redshift-gubjPq

Overview | Rotation | Versions | Replication | Tags

Secret value Info
Retrieve and view the secret value.

Retrieve secret value

Screenshot 3 — IAM Roles

us-east-1.console.aws.amazon.com/iam/home?region=us-east-1#/roles

Identity and Access Management (IAM)

Dashboard

▼ Access management
- User groups
- Users
- Roles
- Policies
- Identity providers
- Account settings

▼ Access reports
- Access Analyzer
  - External access
  - Unused access

IAM > Roles

Roles (7) Info
An IAM role is an identity you can create that has specific permissions with credentials that are valid for short durations. Roles can be assumed by entities that you trust.

Delete | Create role

Role name	Trusted entities	Last activity
AmazonRedshift-CommandsAccessRole-20240522T152317	AWS Service: redshift, and 2 more.	24 hours ago
Assignment10	AWS Service: glue	1 hour ago
AWSServiceRoleForDevOpsGuru	AWS Service: devops-guru (Service-L	1 hour ago
AWSServiceRoleForRedshift	AWS Service: redshift (Service-Linked	24 minutes ago
AWSServiceRoleForSupport	AWS Service: support (Service-Linked	9 days ago
AWSServiceRoleForTrustedAdvisor	AWS Service: trustedadvisor (Service	-
StepFunctions-MyStateMachine-1761xjqcn-role-o8dot79w1	AWS Service: states	1 hour ago

Roles Anywhere Info

https://us-east-1.console.aws.amazon.com/iam/home?region=us-east-1#/ro...

Screenshot 1 — VPC Details

vpc-05b4468cda5e28d5f

VPC > Your VPCs > vpc-05b4468cda5e28d5f

Details Info

VPC ID	State	DNS hostnames	DNS resolution
vpc-05b4468cda5e28d5f	Available	Enabled	Enabled

Tenancy	DHCP option set	Main route table	Main network ACL
Default	dopt-0aa9ba26a28094844	rtb-0dfeb8f32d597ae50	acl-0521fc7f630a90580

Default VPC	IPv4 CIDR	IPv6 pool	IPv6 CIDR (Network border group)
Yes	172.31.0.0/16	–	–

Network Address Usage metrics	Route 53 Resolver DNS Firewall rule groups	Owner ID	
Disabled	–	398738544149	

Resource map | CIDRs | Flow logs | Tags | Integrations

Resource map Info

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
- Endpoints
- Endpoint services
- NAT gateways
- Peering connections

Security

Screenshot 2 — Security Group

Details

Security group name	Security group ID	Description	VPC ID
default	sg-09200ca87f6faf6c8	default VPC security group	vpc-05b4468cda5e28d5f

Owner	Inbound rules count	Outbound rules count	
398738544149	2 Permission entries	1 Permission entry	

Inbound rules | Outbound rules | Tags

Inbound rules (2)

	Name	Security group rule...	IP version	Type	Protocol
	–	sgr-00eab7d1e3702cce7	IPv4	Redshift	TCP
	–	sgr-0e570b8971cf86c64	–	All traffic	All

Screenshot 3 — Endpoints

Endpoints (1/2) Info

Actions | Create endpoint

	Name	VPC endpoint ID	VPC ID	Service name
	–	vpce-0f0a42c4e7e3cf459	vpc-05b4468cda5e28d5f	com.amazonaws.vpce.us-east-1.vpce-svc-0271423...
✓	assignmnet10	vpce-0be69b9518f4c00a3	vpc-05b4468cda5e28d5f	com.amazonaws.us-east-1.s3

Details

Endpoint ID	Status	Creation time	Endpoint type
vpce-0be69b9518f4c00a3	Available	Wednesday, May 22, 2024 at 15:47:51 GMT+5:30	Gateway

VPC ID	Status message	Service name	Private DNS names enabled
vpc-05b4468cda5e28d5f	–	com.amazonaws.us-east-1.s3	No

Screenshot 4 — AWS Glue Crawlers

AWS Glue > Crawlers

Crawlers

A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog.

Crawlers (3) Info

Action | Run | Create crawler

View and manage all available crawlers.

	Name	State	Schedule	Last run	Last run tim...	Log	Table cha...
	airline_dimcraw...	Ready		Succeeded	May 23, 2024 at...	View log	1 created
	flights-data-cra...	Ready		Succeeded	May 23, 2024 at...	View log	1 created
	flights-datas-cr...	Ready		Succeeded	May 23, 2024 at...	View log	1 updated

AWS Glue
- Getting started
- ETL jobs
  - Visual ETL
  - Notebooks
  - Job run monitoring
- Data Catalog tables
- Data connections
- Workflows (orchestration)
- **Data Catalog**
  - Databases
  - Tables
  - Stream schema registries
  - Schemas
  - Connections
  - Crawlers
  - Classifiers
  - Catalog settings
- Data Integration and ETL
- Legacy pages

Execution started successfully

Choose a step to view its details.

Start
StartCrawler
GetCrawler
Is_running?
Wait   Glue StartJobRun
Choice
success_notification   failed_notification

In progress   Failed   Caught error   Cancelled   Succe

Event view   State view

© 2024, Amazon Web Services, Inc. or its affiliates.   Privacy   Terms   Cookie preferences

---

Gmail

Compose

Inbox   3,719
Starred
Snoozed
Sent
Drafts
More

Labels

AWS Notification Message   Inbox

AWS Notifications   {"Error":"States.TaskFailed","Cause":"{\"AllocatedCapacity\":2,\"Attempt\":0,\"CompletedOn\":1716...1:59 PM (2 hours ago)

AWS Notifications   {"Error":"States.TaskFailed","Cause":"{\"AllocatedCapacity\":2,\"Attempt\":0,\"CompletedOn\":17164... 2:11 PM (1 hour ago)

AWS Notifications <no-reply@sns.amazonaws.com>   3:51 PM (15 minutes ago)
to me

{"AllocatedCapacity":2,"Attempt":0,"CompletedOn":1716459669534,"ExecutionClass":"STANDARD","ExecutionTime":127," GlueVersion":"4.0","Id":"jr_5c5ac5af057583bbb1883d0fc378f528d078fb5705c5cb2a2fd1c39a24e60dc","JobName":"flight_data_ing estion10","JobRunState":"SUCCEEDED","LastModifiedOn":1716459669534,"LogGroupName":"/aws-glue/jobs","MaxCapacity":2.0," NumberOfWorkers":2,"PredecessorRuns":[],"StartedOn":1716459521276,"Timeout":2880,"WorkerType":"G.1X"}

© 2024, Amazon Web Services, Inc. or its affiliates.

Redshift query editor v2

Editor

Create   Load data

Queries

Filter resources

redshift-cluster-10
awsdatacatalog
dev
airlines
Tables   2
airports_dim
daily_flights_f...
Views   0
Functions   0
Stored procedures   0
public
sample_data_dev

Notebooks

Charts

History

Scheduled queries

Untitled 1

Run   Limit 100   Explain   Isolated session   redshift-clust...   dev

Schedule

```
1    select * from airlines.daily_flights_fact limit 10;
```

Row 1, Col 51, Chr 51

Result 1 (10)   Export   Chart

	carrier	dep_airport	arr_airport	dep_city	arr_city
	DL	Albuquerque International...	Hartsfield-Jackson Atlanta...	Albuquerque	Atlanta
	DL	Dallas/Fort Worth Internat...	Hartsfield-Jackson Atlanta...	Dallas/Fort Worth	Atlanta
	DL	Dallas/Fort Worth Internat...	Hartsfield-Jackson Atlanta...	Dallas/Fort Worth	Atlanta
	AA	Dallas/Fort Worth Internat...	Hartsfield-Jackson Atlanta...	Dallas/Fort Worth	Atlanta
	AA	Dallas/Fort Worth Internat...	Hartsfield-Jackson Atlanta...	Dallas/Fort Worth	Atlanta
	AA	Dallas/Fort Worth Internat...	Hartsfield-Jackson Atlanta...	Dallas/Fort Worth	Atlanta
	AA	Dallas/Fort Worth Internat...	Hartsfield-Jackson Atlanta...	Dallas/Fort Worth	Atlanta
	AA	Dallas/Fort Worth Internat...	Hartsfield-Jackson Atlanta...	Dallas/Fort Worth	Atlanta
	DL	Dallas/Fort Worth Internat...	Hartsfield-Jackson Atlanta...	Dallas/Fort Worth	Atlanta
	DL	Dallas/Fort Worth Internat...	Hartsfield-Jackson Atlanta...	Dallas/Fort Worth	Atlanta

Query ID 220034   Elapsed time: 80 ms   Total rows: 10

CloudShell   Feedback   © 2024, Amazon Web Services, Inc. or its affiliates.   Privacy   Terms   Cookie preferences