

1. AWS lambda psycpg2 package:

https://github.com/qiaoqiaohejianjian/ETL--Project/tree/master/lambda_redshift

2. AWS redshift and lambda configuration:

S3: after generating processed data, to trigger lambda function "load2redshift" to load data to Redshift, we set Events in my bucket's Properties:

Name	Events	Filter	Type
DEproject_raw_data	All object create events	rawData/, .csv	Lambda

DEproject_processed_data

Name ⓘ

DEproject_processed_data

Events ⓘ

☒ PUT ☐ POST ☐ COPY ☐ Multipart upload completed ☐ All object create events ☐ Object in RRS lost ☐ Permanently deleted ☐ Delete marker created ☐ All object delete events ☐ Restore initiated ☐ Restore completed ☐ Replication time missed threshold ☐ Replication time completed after threshold ☐ Replication time not tracked ☐ Replication failed

Prefix ⓘ

processed/

Suffix ⓘ

.csv

Send to ⓘ

Lambda Function

Lambda

load2redshift

Lambda: a) wrap up code and package, and then upload in Function code -> Actions -> Upload a zip file

b) Set Environment Variables for your redshift database information, such as: database name, host(end point:), password etc.

- (New console) In the navigation pane, choose **CLUSTERS**. Then choose the cluster name from the list to open its details. On the **Properties** tab, in the **Database configurations** section, record the **Database name** and **Port**. View the **Connection details** section and record the **Endpoint** which is in the following form:

endpoint:port/databasename

This is the endpoint,
Leave out
:/port/database name

c) set up lambda function basic info

d) choose same VPC as your Redshift cluster, and same subnet

e) In Permissions: attach s3 Read-only and AWS Lambda VPC Access Execution Role policies

Redshift: a) create your cluster and store your database login information: user and password in a safe place

In your cluster -> properties, you can find more detail information

b) cluster permissions: attach a s3 read-only policy to the cluster

c) for the Network and security:

edit Security Group and Network ACLs, make sure related IP address can access you cluster with your database information.

d) create an endpoint for S3 service.

3. Connect your Redshift cluster:

<https://towardsdatascience.com/redshift-from-the-command-line-5d6b3233f649>

<https://docs.aws.amazon.com/redshift/latest/mgmt/connecting-from-psql.html>

<https://forums.aws.amazon.com/thread.jspa?threadID=310081>

Cluster permissions	
Your cluster needs permissions to access other AWS services on your behalf. For the req "redshift.amazonaws.com". You can associate up to 10 IAM roles with this cluster. Learn	
Attached IAM roles	Status
redshiftRole 🔗 arn:aws:iam::318140223133:role/redshiftRole	in-sync