2 simple ways to write data to AWS RDS

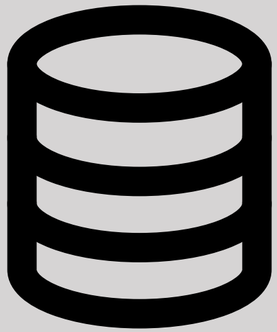
RDS Running Postgres DB

Made By

Kamalraj M M

What is a Database

LAYMAN DEFINITION



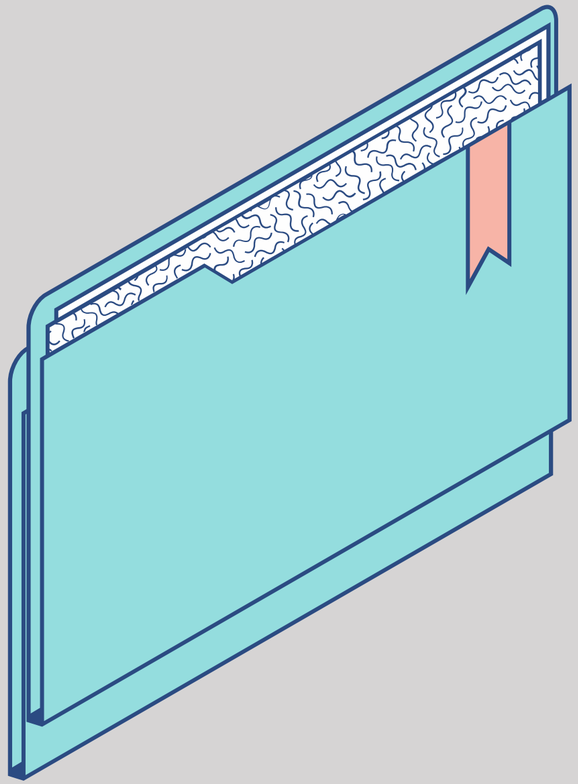
A General Purpose Database

Database is a server + application that stores data of any type and serves the stored data **when it is Queried**

Database Server has many database in it.

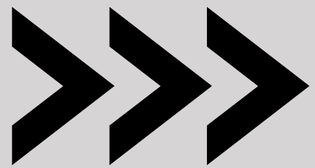
Each Database can have many tables inside them.

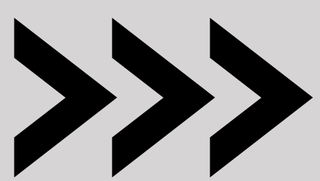
Each Tables stores data as Tuples

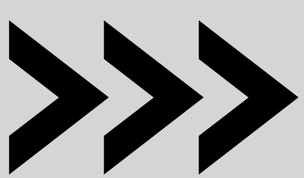


Pre-Requisites

- Comfortable with AWS Wrangler & Pandas
- Ready to use Psycopg2 and Pandas Library




RDS - Instance with Postgres

PSEUDOCODE

2 ways to write data to RDS

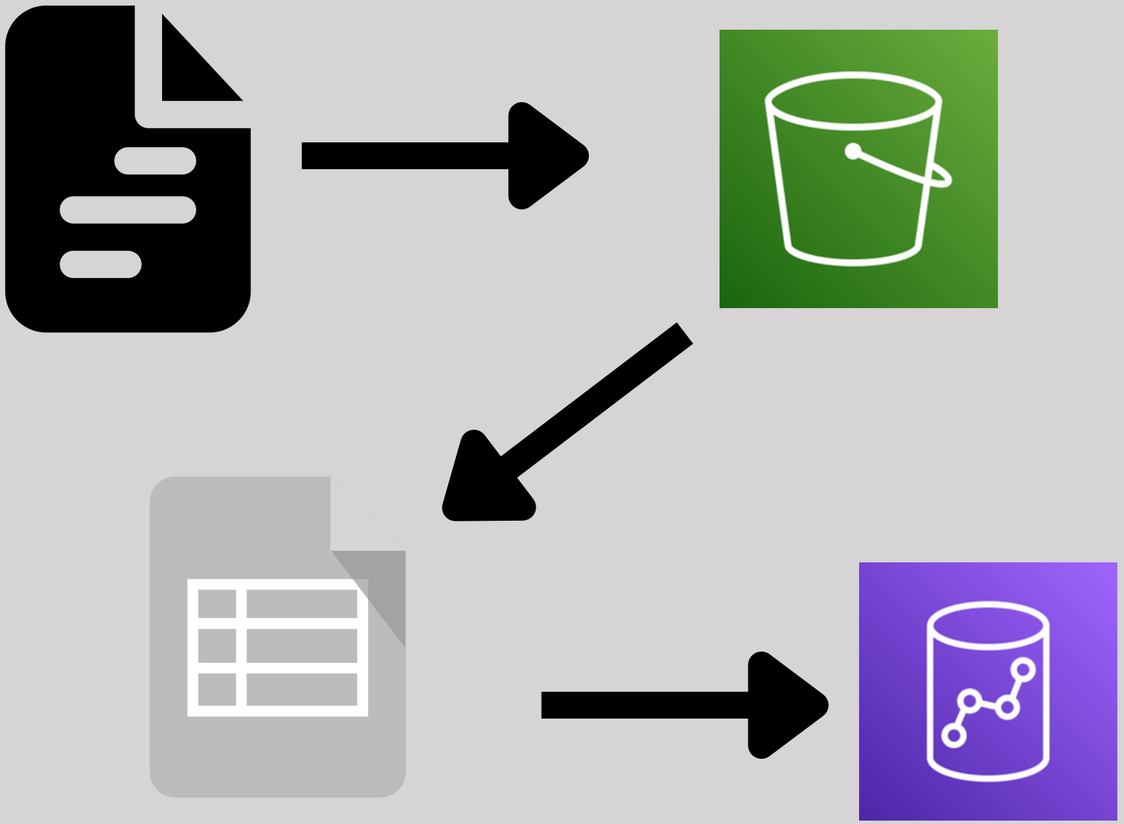
- Start the DB instance if not running
- Secure the DB Credentials in config file

There are **two efficient ways** to write data to DB. Both will require psycopg2 binary library for creating connection

1. Use pandas sql method to write data to RDS
2. Use psycopg2 to copy file from s3 to RDS

LIVE DEMO

SUPER CHARGING ATHENA



Terminal+ PSQL

1. Get the RDS Instance up and running.
2. Connect to Database over PSQL
3. Create table and check by querying

Method 1 : Pandas Route

File location : S3 / Local hard_drive

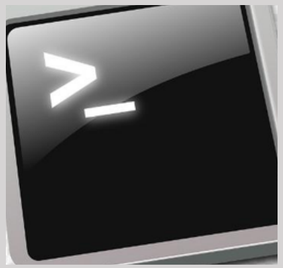
RDS Connection : Psycopg2 connection object + SQLAlchemy connection engine

Read local file into Pandas DataFrame

Ensure there is no Null or missing values in Dataframe

Write the dataframe using yourDataframe.to_sql(table_name, connection)

schema = pd.read_sql(f"""SELECT * FROM {Table_Name}""",con=credentials)



Method 2 : Copy Command Route

File location : Having it in local drive / server will be best

Connect to RDS DB using Psql and create Table

Use the \copy command to write the data to RDS database using the credentials

Use psql to query the RDS database instance

No involvement of S3 or Pandas. Everything happens over Terminal

THANK YOU