

L3 Exercise 4 - Table Design - Solution

June 7, 2023

1 Exercise 4: Optimizing Redshift Table Design

```
In [1]: %load_ext sql
```

```
In [2]: from time import time
import configparser
import matplotlib.pyplot as plt
import pandas as pd
```

```
In [3]: config = configparser.ConfigParser()
config.read_file(open('dwh.cfg'))
KEY=config.get('AWS','key')
SECRET= config.get('AWS','secret')

DWH_DB= config.get("DWH","DWH_DB")
DWH_DB_USER= config.get("DWH","DWH_DB_USER")
DWH_DB_PASSWORD= config.get("DWH","DWH_DB_PASSWORD")
DWH_PORT = config.get("DWH","DWH_PORT")
```

2 STEP 1: Get the params of the created redshift cluster

- We need:
 - The redshift cluster endpoint
 - The IAM role ARN that give access to Redshift to read from S3

```
In [4]: # FILL IN THE REDSHIFT ENDPOINT HERE
# e.g. DWH_ENDPOINT="redshift-cluster-1.csmamz5zxmle.us-west-2.redshift.amazonaws.com"
DWH_ENDPOINT="dwhcluster.c4uipmqcjl1.us-west-2.redshift.amazonaws.com"

#FILL IN THE IAM ROLE ARN you got in step 2.2 of the previous exercise
#e.g DWH_ROLE_ARN="arn:aws:iam::988332130976:role/dwhRole"
DWH_ROLE_ARN="arn:aws:iam::918744264023:role/dwhRole"
```