# Create AWS account:

## Create an Amazon AWS account [here](https://signin.aws.amazon.com/signin?redirect_uri=https%3A%2F%2Fportal.aws.amazon.com%2Fbilling%2Fsignup%2Fresume&client_id=signup&code_challenge_method=SHA-256&code_challenge=lXAcI95OkeMZCqYjgKvL3Jx5HpD_lyLTHTERjIauJD8)

# Create Amazon Redshift cluster & configure the cluster

## Sign into AWS management console and search for ‘AWS Redshift’ in search bar and click ‘Amazon Redshift’

## Create the Amazon Redshift cluster with the specified cluster configuration by specifying Cluster identifier, node type (dc2.large), and number of nodes (nodes: 2)

Graphical user interface, text, application, email

Description automatically generated

# Secure the cluster

## Under Cluster permissions, in Manage IAM roles, choose “Create IAM role” and specify “No additional Amazon S3 bucket” to allow the created IAM role

## This gives “CommandAccessRole” for the any resource that accesses the cluster

## Click “Create Cluster”

Graphical user interface, text, application, email

Description automatically generated

# Load sample dataset from S3 to cluster:

## Choose any sample dataset

## 

## Create an S3 bucket in AWS and upload the housing data

Graphical user interface, text, application, email

Description automatically generated

## Create IAM role for cluster access by choosing IAM from AWS console and create a role and select the below configurations to enable the cluster to have full access to S3 bucket, finally click on create IAM role

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

## Associate IAM role to the cluster by navigating to Redshift cluster and changing cluster permissions to associate the recently created IAM role

Graphical user interface, text, application, email

Description automatically generated

# Set up SQL Workbench/J & connecting to Redshift cluster

## Install SQL Workbench/ J following this [link](https://www.sql-workbench.eu/getting-started.html)

## In the connection profile of SQL Workbench/J, select the below configurations for driver, URL (from cluster details), username and password set for the cluster, select auto commit

Graphical user interface, text, application, email

Description automatically generated

## Click test button to ensure a successful connection to the Redshift cluster

Graphical user interface, text, application, email

Description automatically generated

# Create table schema and move data to the cluster

## Create table by creating relevant schema and copy the data from S3 to the cluster through relevant SQL queries

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

# Analyse the data:

## Find total bedrooms and total rooms in the dataset

Graphical user interface, text, application, email

Description automatically generated

## Find average age of people considering the ocean proximity

Graphical user interface, text, application, email

Description automatically generated

# Reference links:

1. AWS Getting started manual: <https://docs.aws.amazon.com/redshift/latest/gsg/redshift-gsg.pdf#getting-started>
2. Analyse query: <https://docs.aws.amazon.com/redshift/latest/gsg/rs-gsg-sample-data-load-query.html>