**IFT 512:** Advanced Big Data Analytics/AI

**Name:** Aishwarya Devi Akkim

**Assignment 2.1:** A2.1SparkIntro

**Instructor Name:** Robert Rucker

**Due Date:** 8/27/2023

Create cluster

A screenshot of a computer

Description automatically generated

// Read the file and set headers true

val flight2015datasummary = spark

.read

.option("inferSchema", "true")

.option("header", "true")

.csv("/FileStore/tables/2015\_summary-6.csv")

//Display the table

display(flight2015datasummary)

A screenshot of a computer

Description automatically generated

//Updated table column names

val updated\_names\_flight2015datasummary = flight2015datasummary

.withColumnRenamed("DEST\_COUNTRY\_NAME","D\_COUNTRY\_NAME")

.withColumnRenamed("ORIGIN\_COUNTRY\_NAME","O\_COUNTRY\_NAME")

// Display the updated table

display(updated\_names\_flight2015datasummary)

A screenshot of a computer

Description automatically generated

display(updated\_names\_flight2015datasummary.sort("COUNT").take(10))

A screenshot of a computer

Description automatically generated

//script to convert DF to SQL

case class flight2015datasummary(D\_COUNTRY\_NAME: String,O\_COUNTRY\_NAME: String,COUNT: Integer)

import spark.implicits.\_

val flight\_summary\_2015 = updated\_names\_flight2015datasummary.as[flight2015datasummary]

flight\_summary\_2015.createOrReplaceTempView("flight\_summary\_2015")

val df2sql = spark.sql(""" SELECT D\_COUNTRY\_NAME, O\_COUNTRY\_NAME, COUNT FROM flight\_summary\_2015 """)

df2sql.show(10)

A screenshot of a computer

Description automatically generated