# Databricks notebook source

from pyspark.sql.types import StructType, StringType, IntegerType

from pyspark.sql.functions import sum

# COMMAND ----------

schema = StructType().add("Id", StringType(), True)\

.add("Name", StringType(), True)\

.add("Age",IntegerType(),True)

# COMMAND ----------

df\_001=spark.read.option("header",True).option("mode","permisive").schema(schema).format("csv").load("/FileStore/gmde/db.csv")

# COMMAND ----------

df\_001.display()

# COMMAND ----------

dbutils.fs.mount(source = "wasbs://databricks@adlsgmde1710.blob.core.windows.net",

mount\_point ="/mnt/databricks",

extra\_configs = {"fs.azure.account.key.adlsgmde1710.blob.core.windows.net":dbutils.secrets.get(scope = "db\_scope", key = "adlssecret")})

# COMMAND ----------

dbutils.fs.ls("/mnt/databricks")

# COMMAND ----------

df\_002=spark.read.option("header",True).option("inferSchema",True).format("csv").load("/mnt/databricks/employee.csv")

# COMMAND ----------

df\_002.display()

# COMMAND ----------

grouped\_data = df\_002.groupBy("Department").agg(sum("Salary").alias("dep\_total"))

# COMMAND ----------

grouped\_data.display()

# COMMAND ----------

df\_002.write.parquet("/mnt/databricks/parquet/")

# COMMAND ----------

grouped\_data.write.format("delta").saveAsTable("store\_trans1")

# COMMAND ----------

df\_002.write.format("delta").saveAsTable("store\_trans")

# COMMAND ----------

df\_002.write.format("delta").save("/mnt/databricks/storage/")

# COMMAND ----------

# MAGIC %sql

# MAGIC CREATE TABLE ext\_store\_trans

# MAGIC USING Delta

# MAGIC OPTIONS (path "abfss://databricks@adlsgmde1710.dfs.core.windows.net/storage")

# COMMAND ----------

# COMMAND ----------

# MAGIC %sql

# MAGIC describe detail store\_trans

# COMMAND ----------

# MAGIC %sql

# MAGIC describe

# MAGIC store\_trans

# COMMAND ----------

# MAGIC %sql

# MAGIC describe extended store\_trans

# COMMAND ----------

# MAGIC %sql

# MAGIC describe history ext\_store\_trans

# COMMAND ----------

# MAGIC %sql

# MAGIC select \*from ext\_store\_trans

# COMMAND ----------

# MAGIC %sql

# MAGIC OPTIMIZE ext\_store\_trans

# MAGIC ZORDER BY (EmployeeID, Department);

# COMMAND ----------

# MAGIC %sql

# MAGIC OPTIMIZE ext\_store\_trans;

# COMMAND ----------

# MAGIC %sql

# MAGIC insert into ext\_store\_trans Values(83,"Frank Castle","fcaste","Design",76000,"2022-07-30");

# MAGIC

# COMMAND ----------

# MAGIC %sql

# MAGIC alter table ext\_store\_trans

# MAGIC cluster by(EmployeeID,Department);