Cheat Sheet

df = sqlContext.read.text("/user/macksv17/mytest/orders\_text")

.json

.parquet

For csv

df = sqlContext.read \

.format("com.databricks.spark.csv") \

.option("header", "true") \

.option("mode", "DROPMALFORMED") \

.option("delimiter", "|") \

.load("/user/macksv17/practice4/question3/results\_output")

For ORC

DF.write.format("com.databricks.spark.csv").mode("Overwrite").option("delimiter", "|").save(

========== DataBricks ====================

com.databricks:spark-csv\_2.10:1.5.0 , com.databricks:spark-avro\_2.10:2.1.0  
import com.databricks.spark.avro.\_  
import com.databricks.spark.csv.\_

------------ Sql Context ------------------  
sqlContext.setConf(“spark.sql.shuffle.partitions”,“2”)  
sqlContext.setCont(“spark.sql.parquet.compression.codec”,“snappy”) =>{snappy,gzip,lzo}  
sqlContext.setCont(“spark.sql.avro.compression.codec”,“snappy”) =>{snappy,gzip,lzo}