from pyspark.sql import HiveContext

from pyspark.sql.types import \*

from pyspark.sql import Row

# Carregando o arquivo pessoas.json para um dataframe

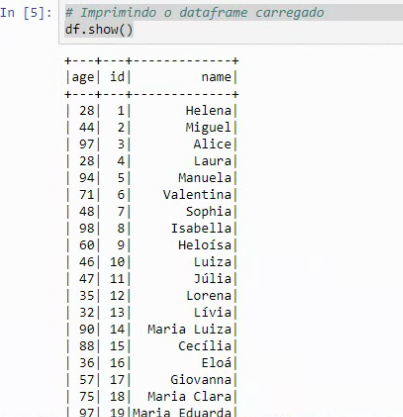
df = spark.read.json(path='/base/pessoas.json')

# Carregando o arquivo pessoas.json para um dataframe

df = spark.read.json(path='/base/pessoas.json')

# Imprimindo o dataframe carregado

df.show()

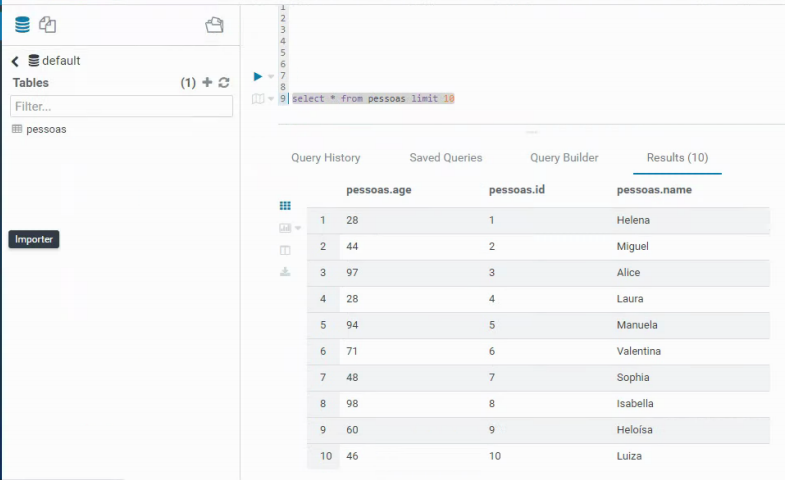


df.printSchema()

hc = HiveContext(sc)

hc.sql("show databases").show()

df.write.format("orc").mode("overwrite").saveAsTable("default.pessoas")



|  |
| --- |
|  |

sqoop import --table employees --connect jdbc:mysql://database/employees --username root --password secret --warehouse-dir /aula/hivetable/ --m 1 --hive-import --hive-table default.employees --hive-delims-replacement "SPECIAL"

