Tên: Huỳnh Lan Nhi

MSSV: 1050080028

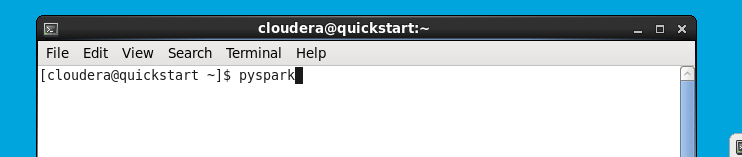
LAB trên lớp

**0.1 Raaghav Radhakrishnan - 246097**

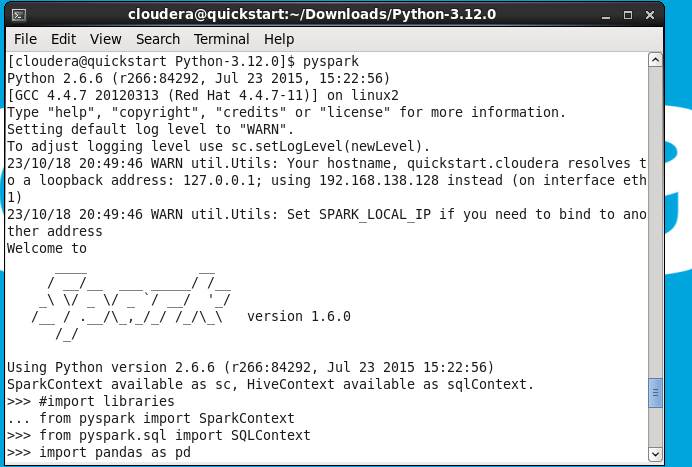
**0.1.1 Exercise 1 - Part (a)**

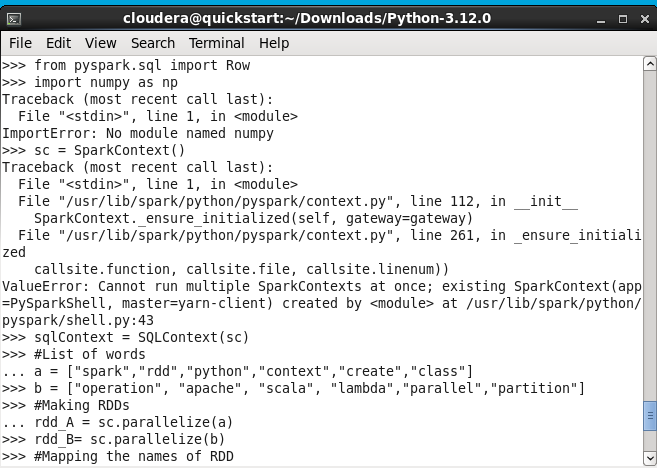
Để làm được bài tập trên cần mở pyspark

Để mở pyspark: Mở terminal nhập câu lệnh pyspark



Nhập các câu lệnh có sẵn trong bài lab





**0.1.2 Right outer join:**

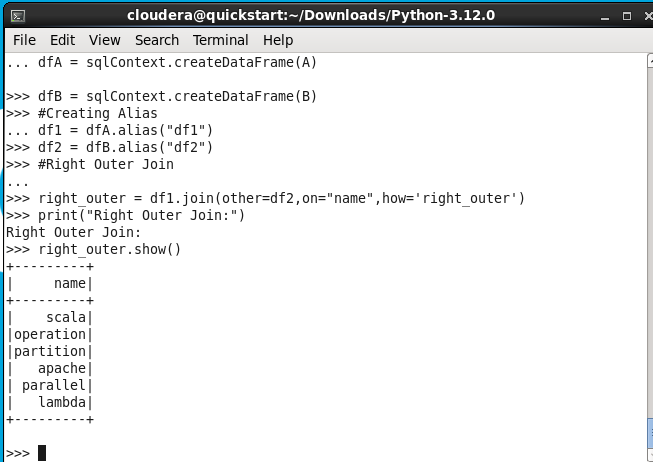
Nhập câu lệnh:

#Right Outer Join

right\_outer = df1.join(other=df2,on="name",how='right\_outer')

print("Right Outer Join:")

right\_outer.show()

****

**0.1.3 Full outer join:**

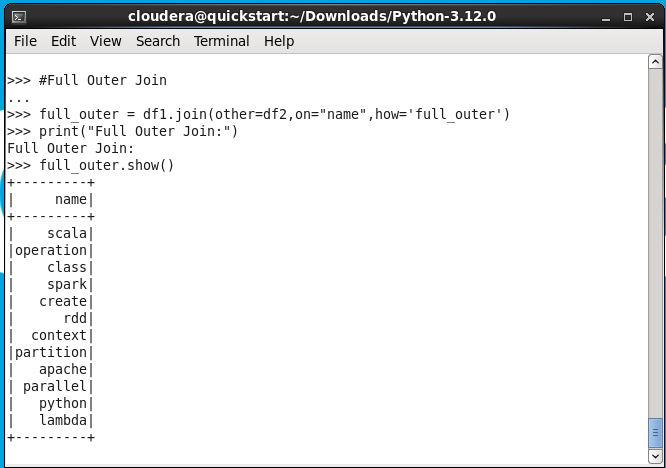
Nhập lệnh:

#Full Outer Join

full\_outer = df1.join(other=df2,on="name",how='full\_outer')

print("Full Outer Join:")

full\_outer.show()



Nhập lệnh:

map\_rdd = full\_outer.rdd.map(lambda x: sum([word.count('s') for word in x]))

map\_df = map\_rdd.map(lambda x: Row(name = x))

map\_df = sqlContext.createDataFrame(map\_df)

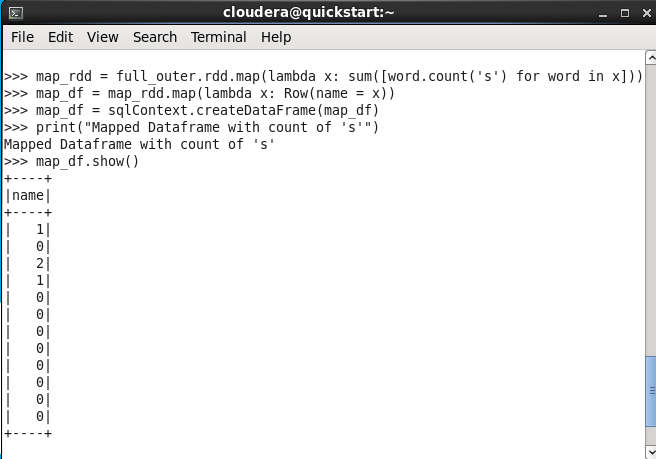
print("Mapped Dataframe with count of 's'")

map\_df.show()

#Reducing the RDD

reduce\_rdd=map\_rdd.reduce(lambda x,y: x+y)

print("Using Map-Reduce, the character \"s\" appears", reduce\_rdd, "times in all a and b.\n")



Nhập lệnh:

#Aggregate function

count = full\_outer.rdd.aggregate(0, lambda i, x: i + x[0].count('s'),

lambda i, j: i+j)

print("Using aggregate function, the character \"s\" appears",count,

"times in all a and b." )

