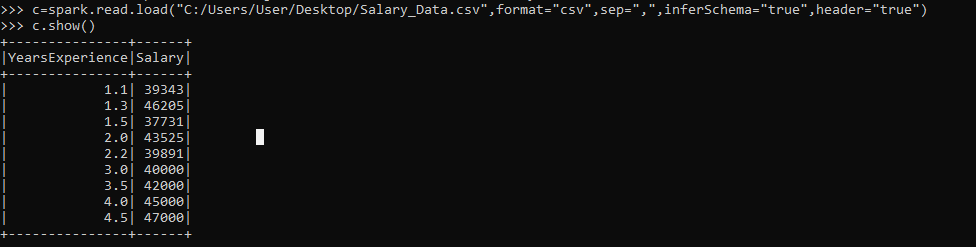
**Basic Functions of PySpark**

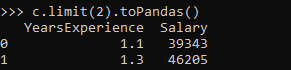
**Read:**

****

Spark.read.load() 🡺 used for load the file into spark shell

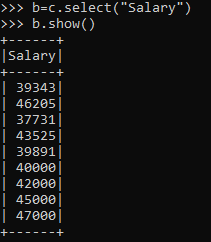
Show() 🡺 used for display the dataset.

**To Pandas:**

****

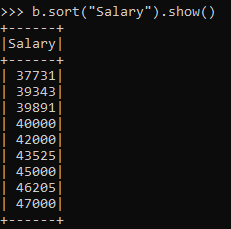
toPandas() 🡺 used for change the dataset into pandas data frame.

**Select By column:**

****

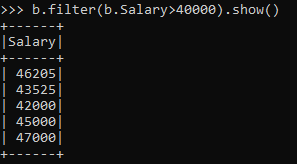
Select() 🡺 used for select specific column from dataset or dataframe.

**Sort:**

****

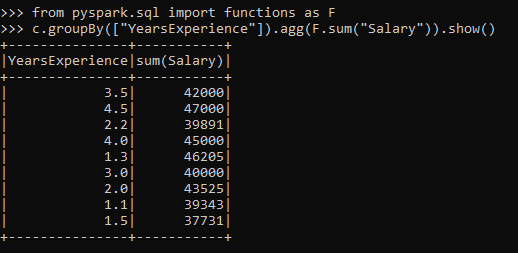
Sort() 🡺 used for sort the column.

**Filter:**

****

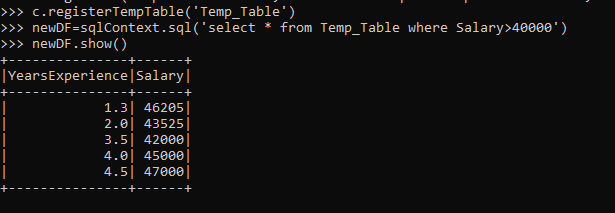
Filter() 🡺 used for filter dataset based on condition.

**Group By:**

****

groupBy([“column name”]) 🡺 used for group by based on columns.

**SQL commands on PySpark:**

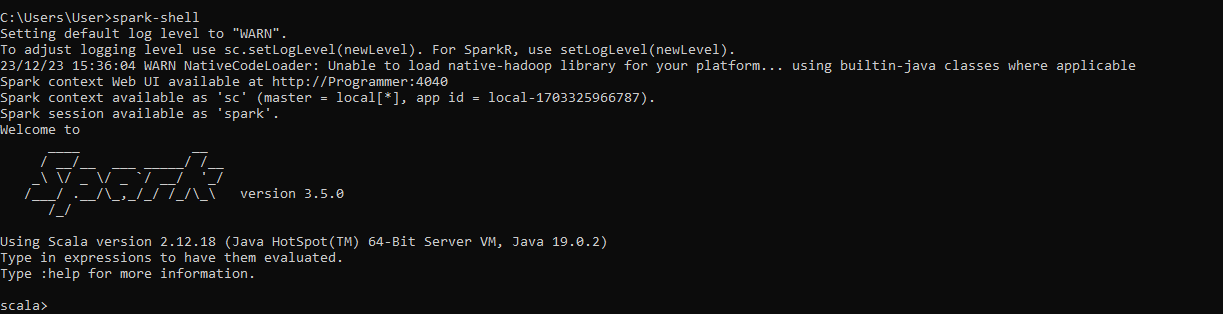
****

Var\_name.registerTempTable(‘table\_name’)🡺 change dataframe as tables for sql performance.

sqlContext.sql(‘query’)🡺 used to write sql query.

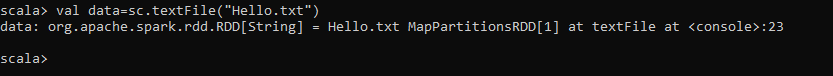
**Basic Spark Command**

**To start spark shell:**

****

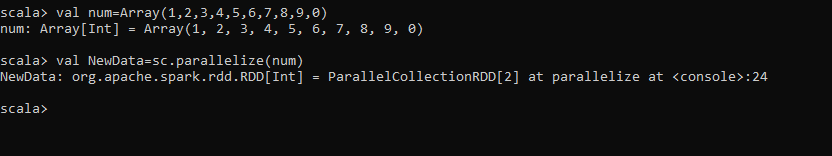
Spark-shell 🡺 used for start spark shell window.

**Read:**

****

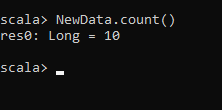
Sc.textFile()🡺 read the text file.

**Create RDD through parallelizing:**

****

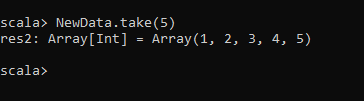
Sc.parallelize(var\_name)🡺 used for create RDD dataset.

**Count items in RDD:**

****

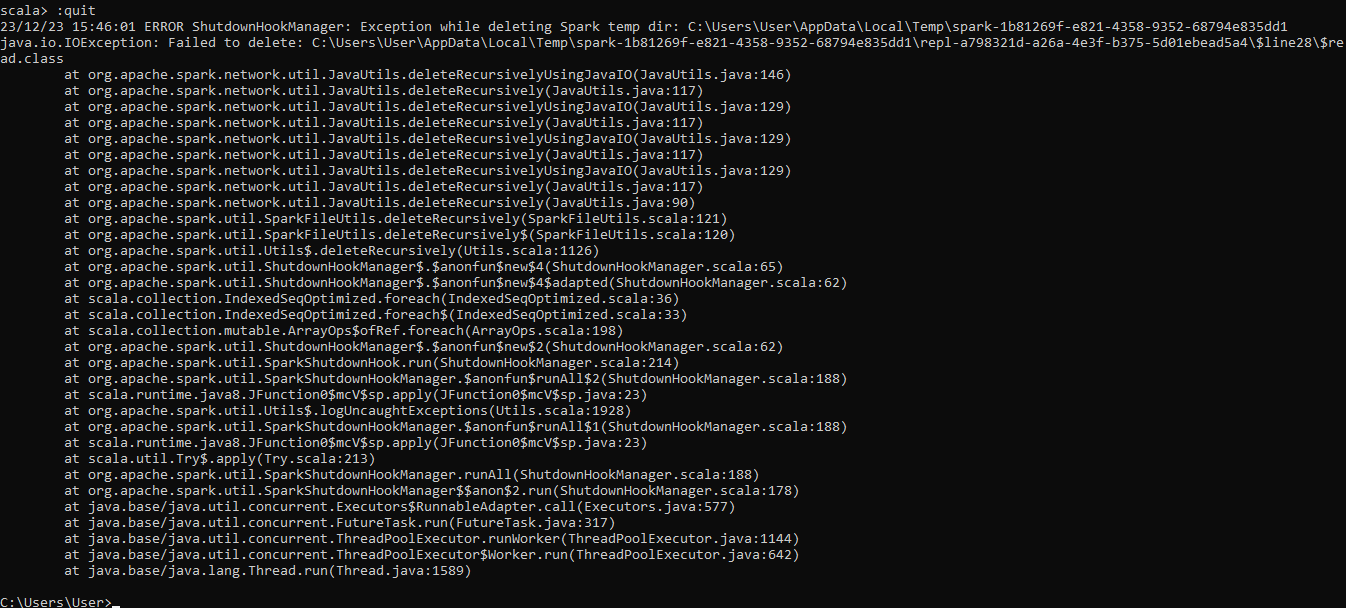
Count()🡺 gives count of records in RDD.

**Read data from RDD:**

****

Var\_name.take(range)🡺used for take data and display them.

**Quit:**

****

:q or :Quit 🡺 used for exit from spark shell.