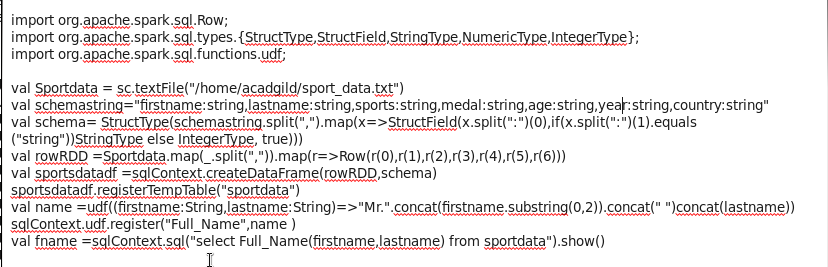
Using udfs on dataframe

1. Change firstname, lastname columns into

Mr.first\_two\_letters\_of\_firstname<space>lastname

for example - michael, phelps becomes Mr.mi phelps



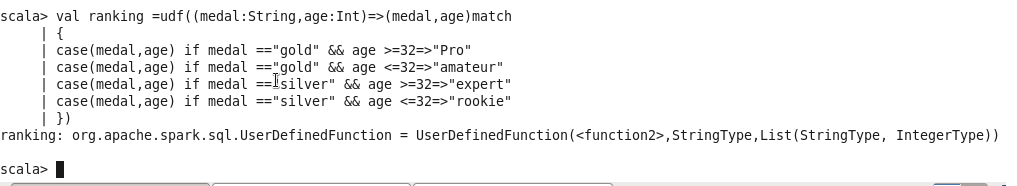
2. Add a new column called ranking using udfs on dataframe, where :

gold medalist, with age >= 32 are ranked as pro

gold medalists, with age <= 31 are ranked amateur

silver medalist, with age >= 32 are ranked as expert

silver medalists, with age <= 31 are ranked rookie



sqlContext.udf.register(“ranks”,ranking)

val Ranking RDD=sportdataDF.withColumn(“ranks”,Ranking(sportdataDF.col(“medal”),sportdataDF.col(“age”)))