**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**

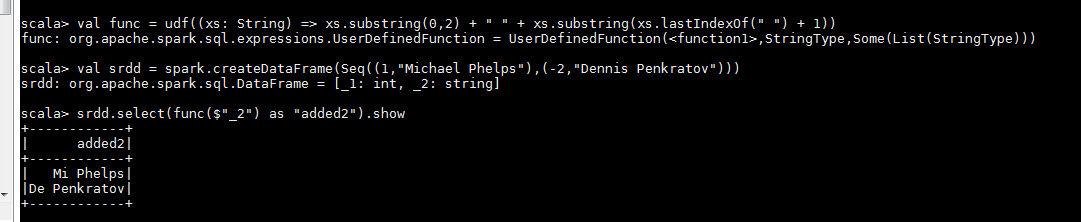
import org.apache.spark.sql.functions.udf

import spark.implicits.\_

val func = udf((xs: String) => xs.substring(0,2) + " " + xs.substring(xs.lastIndexOf(" ") + 1))

val srdd = spark.createDataFrame(Seq((1,"Michael Phelps"),(2,"Dennis Penkratov")))

srdd.select(func($"\_2") as "Name").show



**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**

import sqlContext.implicits.\_

import org.apache.spark.sql.functions.\_

val srdd = sc.parallelize(Seq((32,"Michael Phelps", "GoldMedallist"),(33,"Dennis Penkratov","GoldMedallist"))).toDF("age","name","medaltype")

val newDf = srdd.withColumn("ranking", when($"age" >= 32 && $"medaltype".equals("GoldMedallist"), "Pro").otherwise("Amateur"))

newDf.show()

