**Assignment 19.2 Spark**

1. Change firstname, lastname columns into

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

for example - michael, phelps becomes Mr.mi phelps

<https://drive.google.com/drive/folders/0B_P3pWagdIrrTEVTVE1VWXI0YVk>

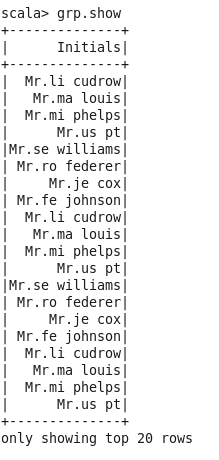
case class SPORTS (firstname: String , lastname: String, sports: String, medal: String , age: String, year: String, country: String)

val spt = sc.textFile("sport\_data.txt").map(\_.split(",")).map(p => SPORTS (p(0), p(1), p(2), p(3), p(4),p(5),p(6))).toDF()

spt.registerTempTable("sports");

val sport = sqlContext.sql("SELECT \* FROM sports")

scala> val grp = sqlContext.sql("SELECT CONCAT( 'Mr' , '.' ,substr(firstname,1,2) ,' ' , lastname ) Initials from sports ")



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

val grp = sqlContext.sql("SELECT firstname, lastname, medal, country , age, case when medal == 'gold' and age >= '32' then 'pro' when medal == 'gold' and age <= '31' then 'amateur' when medal == 'silver' and age >= '32' then 'expert' when medal == 'silver' and age <= '31' then 'rookie' end ranked from sports ")

