**ASSIGNMENT 19.2**

PROBLEM STATEMENT:

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**

Solutions:

val sportsDf=

sqlContext.read.option("header","true").format("/home/acadgild /Downloads/Sports\_Data.csv")

sportsDf.createOrReplaceTempView("sport")

def udfChangeName=org.apache.spark.sql.functions.udf((firstName: String, lastName: String, append: String) => {

s"${append}.${firstName.substring(0,2)} $lastName"

})

val f1 = sportsDf.withColumn("name", udfChangeName(sportsDf("firstname"),sportsDf("lastname"),org.apache.spark.sql.functions.lit("Mr")))

**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**

Solutions:

def udfRanking=org.apache.spark.sql.functions.udf((medal: String, age : Int) => {

medal match {

case "gold" => if(age >= 32) "pro" else "amateur"

case "silver" => if(age >= 32) "expert" else "rookie"

}

})

val f2 = sportsDf.withColumn("ranking", udfRanking(sportsDf("medal\_type"),sportsDf("age")))