Problem02

September 26, 2019

```
[5]:  ################################################################################
      #
      # Ulysses Carlos 09/25/2019
      # Problem 02
      #
      # Using the Users.xml file, provide the count for all users which joined
      # (CreationDate) in 2017. (30 points). Output this to the screen.
      ################################################################################


      sc.stop()
      from pyspark import SparkConf, SparkContext
      conf = SparkConf().setMaster("local").setAppName("My App")
      sc = SparkContext(conf = conf)

      lines = sc.textFile("Users.xml")

      #filter_01 = lines.filter(lambda line : "=\"2017-" in line)
      filter_01 = lines.filter(lambda line : "CreationDate=\"2017" in line)
      filter_01.collect()

      print ("Total size of Users.xml : " + str(lines.count()) + "\n")
      print ("The total amount of users created in 2017 is " + str(filter_01.count()))
```

```
Total size of Users.xml : 66954

The total amount of users created in 2017 is 14239
```

[ ]: