UNIVERSITÉ SIDI MOHAMED BEN ABDLLAH

FACULTÉ DES SCIENCES DHAR EL MAHRAZ DE FÈS

DÉPARTEMENT D'INFORMATIQUE

Département : Informatique

Master : Informatique Décisionnelle et Vision Intelligent

Titre :

TD / TP 2 With SPARK

Présenter par :
ABIBOU SOUKAYNA

Supervisé par :
Pr. Noura

Année universitaire : 2020/2021

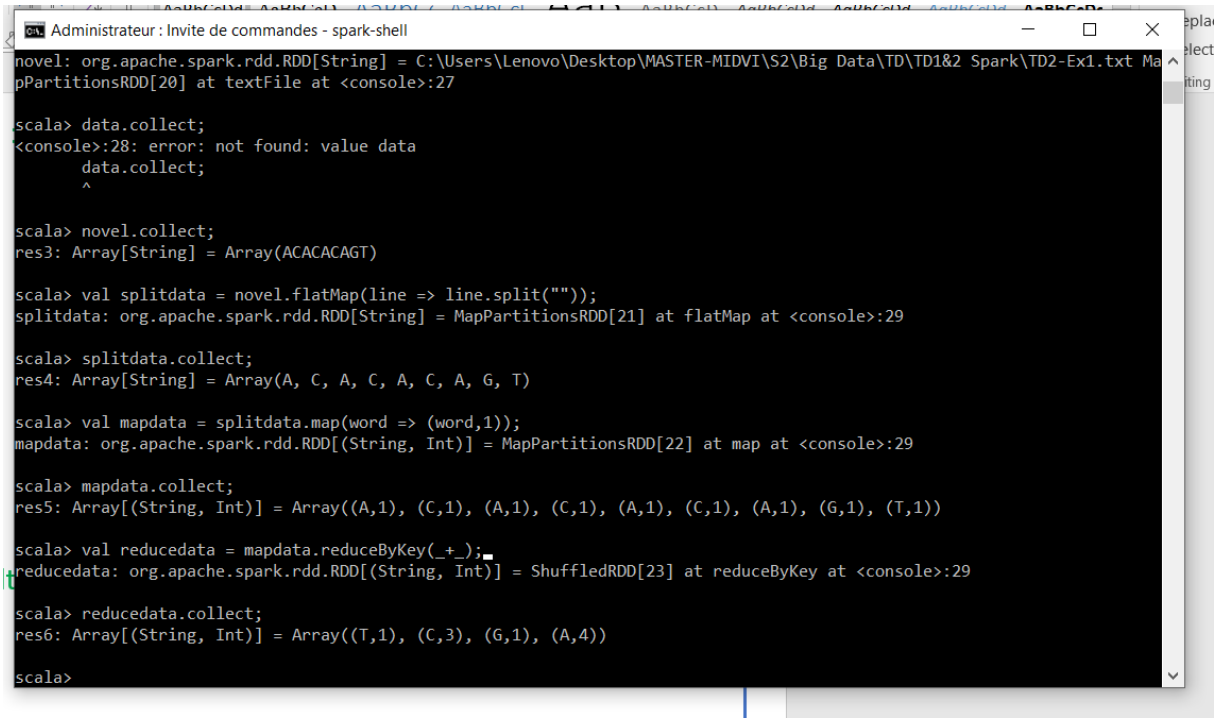
Filière MIDVI

Exercice 1 :

1- Code :

```
val data=sc.textFile("txtFile");
data.collect;
val splitdata = data.flatMap(line => line.split(""));
splitdata.collect;
val mapdata = splitdata.map(word => (word,1));
mapdata.collect;
val reducedata = mapdata.reduceByKey(_+_);
reducedata.collect;
```

2- Résultat d'exécution :



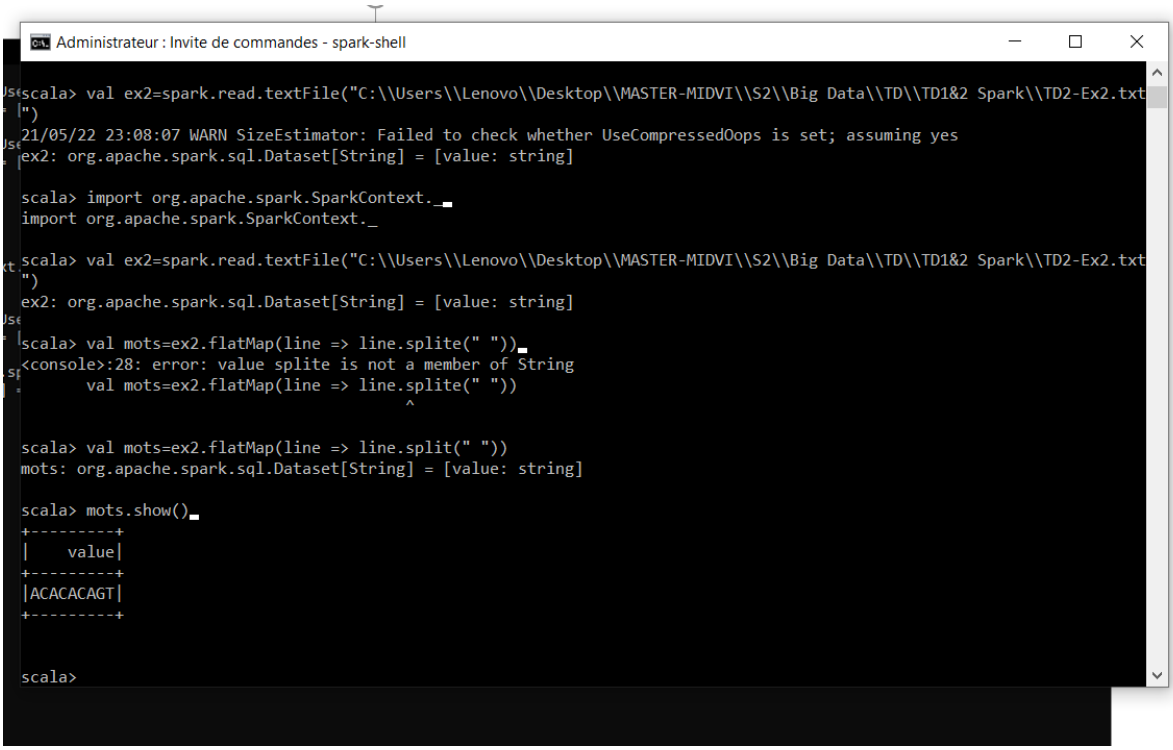
Exercice 2 :

1- Code :

```
import org.apache.spark.SparkContext._
var samplefile = sc.textFile(inputPath)
var ArrayFile = samplefile.collect()

var resultFile = ArrayBuffer[String]()
for ( i<- ArrayFile){
for (j <- 0 until i.length()-2){
if ( i.substring(j, j+3).equals("AAA") ||  i.substring(j, j+3).equals("TTT") )
resultFile += i.substring(j, j+3)
}
for(j <- 0 until i.length()-3){
if ( i.substring(j, j+4).equals("CCCC") ||  i.substring(j, j+4).equals("GGGG") )
resultFile += i.substring(j, j+4)
}
}
var rdd = sc.parallelize(resultFile)
.flatMap(line => line.split(" "))
.map(word => (word , 1))
.reduceByKey(_+_)
```

2- Résultat d’exécution :



Exercice 3 :

1- Code :

```
var samplefile = sc.textFile(inputPath)
var ArrayFile = samplefile.collect()
var resultFile = ArrayBuffer[String]()
for ( i<- ArrayFile){
for (j <- 0 until i.length()-2){
if ( i.substring(j, j+3).equals("ATG")){
for (k <- j+3 until i.length()-2){
if ( i.substring(k, k+3).equals("TAA") ||
i.substring(k, k+3).equals("TAG") || i.substring(k,
k+3).equals("TGA"))
resultFile += i.substring(j, k+3)  }   }  }
var rdd = sc.parallelize(resultFile)
.flatMap(line => line.split(" "))
.map(word => (word , 1))
.reduceByKey(_+_)
.sortByKey(true, 1)
```

2- Résultat d’exécution :