Big Data Technologies

Polytech Nice Sophia – 2024/2025

Lab 3

1 - Perform the following tasks:

```
val textFile = sc.textFile("les-arbres.csv").take(1001) // array[String]

textFile.size

var collectTreeHeights = textFile.filter(line => !line.startsWith("IDBASE")).map(line =>
line.split(";")).map(fields => fields(13).toFloat).filter(height => height>0)

var countTrees = collectTreeHeights.map(element => 1).reduce((x, y) => x + y)

var totalHeight = collectTreeHeights.reduce((x, y) => x + y)

var maxHeight = collectTreeHeights.reduce((x, y) => Math.max(x,y))

System.out.println("Count: " + countTrees)

System.out.println("Total height: " + totalHeight)

System.out.println("Max height: " + maxHeight)

System.out.println("Average height " + totalHeight / countTrees)
```

2 - Study the following MapReduce program:

The proposed code computes an approximation of the number pi.

The program generates nb_samples couples (x,y) belonging to the square with vertices {(0,0), (1,0), (1,1), (0,1)}. Then, it computes an approximation of the probability of a couple (x,y) to belong to the unit circle. This probability is theoretically equal to pi/4. The approximate probability is finally multiplied by 4 to obtain an approximation of pi.