### Université Constantine2 Avril 2024

### Faculté NTIC, dépt. IFA M1 -SDIA

### Interrogation 01

**Nom:**

**Prénom:**

**Groupe:**

### (TDSAI)

### Exercice 1

Utilisez les RDDs pour écrire un programme PySpark permettant de calculer la somme des nombre entiers positifs et le produit des nombres entiers négatifs.

from pyspark.sql import SparkSession

sc = SparkSession.builder.master("local[\*]").getOrCreate()

x=spark.sparkContext.parallelize([2,5,8,-9,10,13,-15,16,-21])

def test(x):

  if (x>0):

    return ("positif",x)

  else:

    return ("negatif",x)

y=x.map(lambda x:test(x))

rdd1 = y.filter(lambda x : x[0]=='positif')

pos=rdd1.reduceByKey(lambda x,y: x+y)

print(pos.collect())

rdd2 = y.filter(lambda x : x[0]=='negatif')

neg=rdd2.reduceByKey(lambda x,y: x\*y)

print(neg.collect())

### NB. Vous pouvez écrire un autre code sans utiliser la fonction test(). Le plus important c’est le bon choix des opérations appliquées sur les RDDs.

### Questions : cochez la ou (les) bonne(s) réponse(s)

1. Dans Spark, l’opération distinct() est une :

a) transformation b) action

c) les deux en même temps d) aucune réponse

1. En Spark ‘createDataFrame()’ permet de créer un DataFrame à partir de:

a) SGBD relationnelle b) RDDs c) SGBD NoSQL d) fichiers .json

e) fichiers xml g) DataFrame Pandas h) aucune réponse

1. les APIs de Spark sont :

a) Python b) Java c) Python uniquement d) toutes les réponses

e) Scala f) R g) Julia h) C++

**Bonne chance**