**CSCE 4013/5013 Cloud Computing and Security**

**Quiz #12 (20 points) (Solution)**

**Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, ID:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Problem #1 (10 pts)**

1. Yes / No Do you participate in this quiz?

**Problem #2 (10 pts)**

**Given the Spark code below, specify the partitioners and the number of partitions associated with RDDs.**

import org.apache.spark.HashPartitioner

//assuming that there are thousands of pair elements in rdd1

val rdd1 = sc.parallelize(List(("a", 1), ("b", 1), ("a", 2), ...., ("ef", 2)), 100)

val rdd2 = rdd1.repartition(50)

val rdd3 = rdd2.partitionBy(new HashPartitioner(100))

val rdd4 = rdd3.map(x=>x)

val rdd5 = rdd4.reduceByKey((x,y)=>x+y)

|  |  |  |
| --- | --- | --- |
| **RDD** | **Partitioner** | **Number of Partitions** |
| rdd1 | None | 100 |
| rdd2 | None | 50 |
| rdd3 | HashPartitioner(100) | 100 |
| rdd4 | None | 100 |
| rdd5 | HashPartitioner(100) | 100 |