TOP GEAR ASSIGNMENTS FOR APACHE SPARK ASSIGNMENTS – L1

Assignments – 1 (Creation of RDD and operations on RDDs)
Note : Identify your spark context and answer the following. Wherever applicable create a new RDD.

Consider the following array.
{10,21,90,34,40,98,21,44,59,21,90,34,29,19, 21,90,34,29,49,78 } . Create a new RDD for each of the following assignments.

1. For the above array,

a) Create a RDD for the above array.

```
scala> val inputRdd = sc.parallelize(Array(10,21,90,34,40,98,21,44,59,21,90,34,29,19,
21,90,34,29,49,78))
```

```
scala> val inputRdd = sc.parallelize(Array(10,21,90,34,40,98,21,44,59,21,90,34,29,19,
21,90,34,29,49,78))
inputRdd: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[0] at parallelize at <
console>:24
```

b) Display the array

```
scala> inputRdd.collect.foreach(println)
```

or

```
scala> inputRdd.take(50)
```

```
scala> inputRdd.take(50)
res1: Array[Int] = Array(10, 21, 90, 34, 40, 98, 21, 44, 59, 21, 90, 34, 29, 19, 21, 9
0, 34, 29, 49, 78)
```

c) Display the first element of the array

```
scala> inputRdd.take(1)
```

2. Consider the above array

a) Display the sorted output (ascending and descending) through an RDD.
b) Display the distinct elements of the array using an RDD
c) Display distinct elements without using a new RDD.

3. Consider the above array

a) Display maximum and minimum of given array using RDD.
b) Display top 5 list elements using RDD
c) Combine above array with a new array { 30,35,45,60,75,85} and display output.
d) Provide the sum of the array elements using reduce with distinct values.
e) Provide the sum of the array elements using reduce.