skill speed

for the serious learner

HOMEWORK ASSIGNMENT: MODULE 7

Complete the following tasks:

1. Construct RDD (x1) 15 values in a sequence using parallelize function. Using the transformation and Actions on RDD, return a collection whose values are 2 more than the thrice of the original values in x1 (3*x + 2)

2. Return the distinct elements of the below list by constructing the RDD on the elements List("aa","bb","ca","da","ef","dd","er","ff","gg","bb","ca","da","ef","qr","rr","ty")

3. You are given two lists -
   - List("hello","world","appearance","spark","scala","java","sparkSQL","eclipse","SparkCore")
   - List("spark","scala","actor","sparkSQL","SparkCore")

   Using RDDs and operations on it, generate the following output:
   Array(SparkCore, spark, sparkSQL, scala)

4. What are basics types of operations in Spark that are performed in RDD and list few of them under each category

5. Can we have 100 partition while creating a list of 100000 numbers? Can the length of the RDD can be checked?