1. Find the average of numbers.

Initialize the base RDD with a list of 1,2,3,4,5,6,7,8,9,10

scala> val data = sc.parallelize(1 to 10)

data: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[11] at parallelize at <console>:27

scala> data.reduce(\_+\_)

res11: Int = 55

scala> data.reduce(\_+\_)/data.count()

res12: Long = 5

scala> data.reduce(\_+\_)/data.count().toDouble()

<console>:30: error: Double does not take parameters

data.reduce(\_+\_)/data.count().toDouble()

^

scala> data.reduce(\_+\_)/data.count().toDouble

res14: Double = 5.5

2. Find the average of numbers.

Read the numbers from a file, each number in separate line. 1,2,3,4,5,6,7,8,9,10

scala> val data = sc.textFile("file:///my-folder/src\_file/sess4\_assgn1\_src01.txt")

data: org.apache.spark.rdd.RDD[String] = file:///my-folder/src\_file/sess4\_assgn1\_src01.txt MapPartitionsRDD[13] at textFile at <console>:27

scala> data.map((\_.toInt)).reduce(\_+\_).toDouble/data.count()

res11: Double = 5.5