Assignment 2

1. Create an RDD from input, split each line by comma, get URL and payload columns

```scala
val lines = sc.textFile(inputFilePath, 1)
val columns = lines.map(line => (line.split(",")(0),line.split(",")(3)))
```

2. Convert payload with B, KB, MB unit to integer (convert method is self-defined)

```scala
val digits_convert = columns.mapValues(x => convert(x)).mapValues(x => List[Int](x))
```

3. ReduceByKey operation

```scala
val reducebykey = digits_convert.reduceByKey((a, b) => (a ::: b)).sortByKey()
```

4. Through list operations to get min, max, mean, variance (vvariance method is self-defined)

```scala
val pre_result = reducebykey.mapValues(x => List(x.min, x.max, x.sum/x.length, vvariance(x)))
```

5. Format the output to string with unit of "B", and save output to file

```scala
val tostring = pre_result.mapValues(_.map(_.toString + "B")).mapValues(_.mkString(","))
val final_result = tostring.map{case(k,v) => s"$k,$v"}
final_result.saveAsTextFile(outputDirPath)
```