**Assignment 2**

**Stats Class:** a case class for calculating min, max, mean and variance

****

**Solution**

1. Determine conversion constant for unit conversion.



1. Create an RDD from input file and split each line by comma.



1. Transform to pair of base URL and size of payload RDD. Then, convert size of payload to byte unit.



1. Transform that RDD to a pair of base URL and stats object. Next, reduce key-value pair by key using aggregation function in stats class. Then, compute mean and variance by using *gen\_output* function in stats class and map a key,value pair to one string output. Next, return a new RDD that has exactly one partition.



1. Finally, write an output.

