assignment2 solution

This scala program is working for a count payload size of each base URL via RDDs in Spark framework.

Firstly, read the file in the function of SparkContext. The goal of this is converting text file to RDD object.

Secondly, before convert, each line of the file to key-value pair, use filter to remove the empty line from records. After filtering, use the map function to apply the customized function to each line for produce key-value pair (BaseURL, Payload size).

Thirdly, use 'groupByKey' to group the set of key-value pair by key for producing a sorted array containing payload size

Finally, calculate mean, variance, minimum and maximum via the two functions named 'mean' and 'variance'.