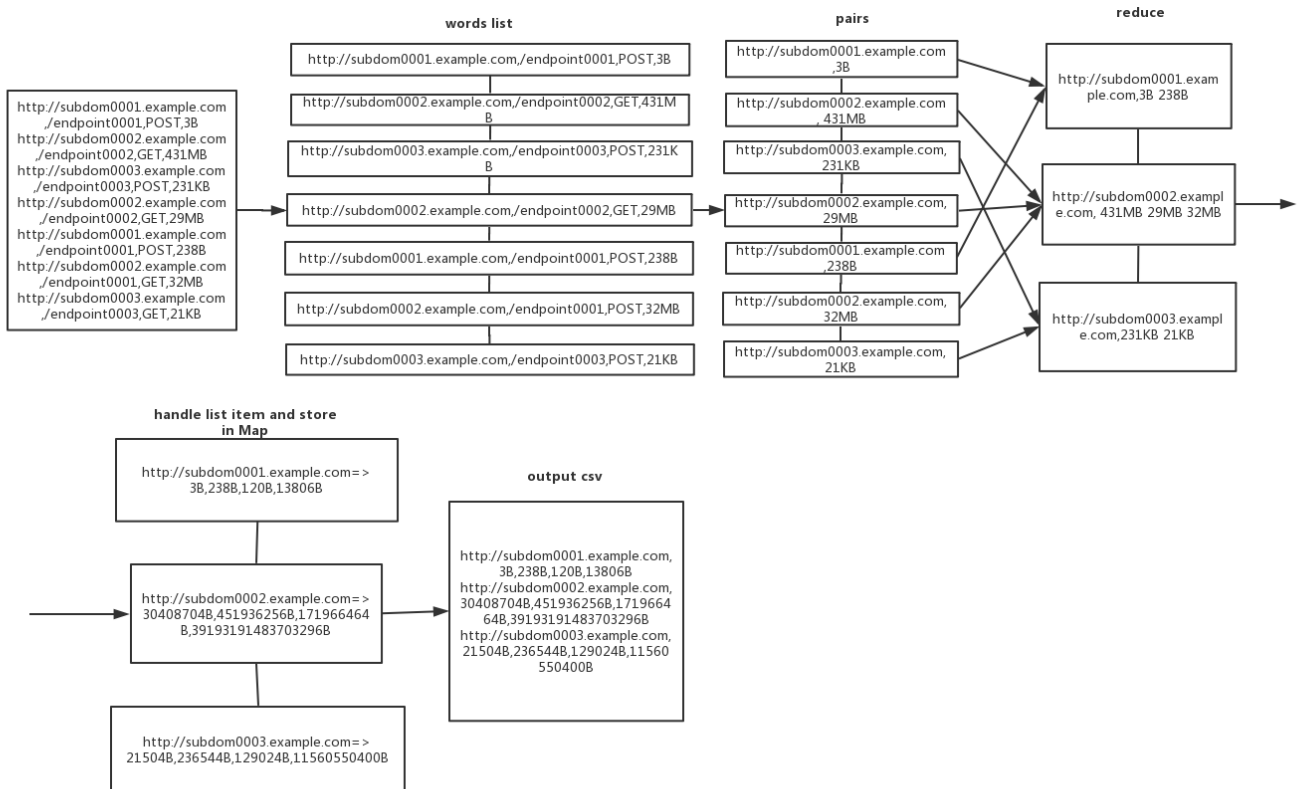


COMP9313 Assignment2 Solution

1. Processing graph



2. RDD operation

During this assignment, I use many RDD functions, such as map, flatmap and reduceByKey. First, load content from input csv file, and then pair web URL as key and payload as value. After that, combine pairs which have the same key. In this process, value is connected with a space. Next, calculate the average, maximum, minimum, and standard deviation of the value for each key value, and combine each of the result with the key into new pair. Finally, output all key-value pairs in given format to output csv file.