PROJECT 1 REPORT

1) A document with the following details

a) Transformation of data during the computations i.e. data type of key, value

The input to the map phase is raw data. LongWritable corresponds to Java Long, Text corresponds to Java String, IntWritable corresponds to Java Integer. The map() method is passed a key and a value. The text value containing the line of the input is converted to StringTokenizer. The map() method is also provided an instance of Context to write the output. It collects mapped keys and values. The reduce() method is passed a key and a value. The output types of the reduce function are Text and IntWritable.

The mapper reads one line at a time and splits the line into tokens separated by whitespaces. The output of the map function is a Key value pair.

The output of the first map:
< Monday, 1>
< Tuesday, 1>
< Wednesday, 1>
< Monday, 1>

The output of the second map:
< Monday, 1>
< Saturday, 1>
< Sunday, 1>
< Friday, 1>

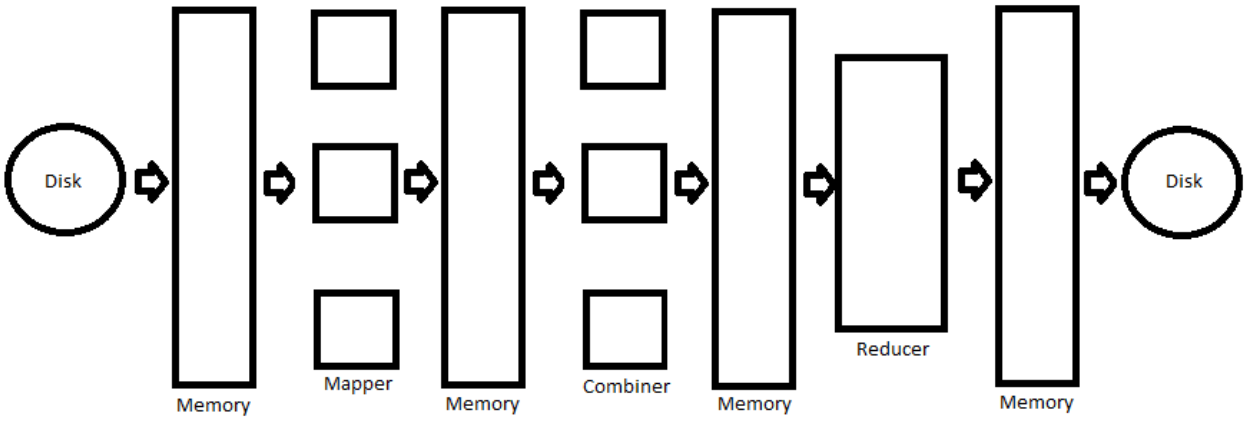
The output of each map is input to the combiner who performs sorting followed by aggregation on the keys.

The output of the first map:
< Monday, 2>
< Tuesday, 1>
< Wednesday, 1>

The output of the second map:
< Friday, 1>
< Monday, 2>
< Saturday, 1>
< Sunday, 1>

Reducer sums up the value count for each key.
The output of reducer:
< Friday, 1>
< Monday, 4>
< Saturday, 1>
< Sunday, 1>
< Tuesday, 1>
< Wednesday, 1>

b)



The input to the mapper is read from the file which is on the disk. The output of the map function is on the memory. The output of the mapper is input to the combiner. The output of the combiner is on the memory. The output of the combiner is input to the reducer. The output of the reducer is then written to the file which is on the disk.