# BigData Programming Assignment#1

Qing-Cheng Li
R01922024

## 1 Implementation Description

### 1.1 Sort

Use the number as key, NullWritable as the value of mapper, when reducer receive it, just output the key. Set IntWritable as the comparator so the result will output in numerical order, not in the alphabetical order.

### 1.2 PageRank

First, split the inpur string into *Node*, *PageRank value* and *output links*. For each output links, using the node output link points to as the key, $\frac{PageRank}{\#outputlink}$ as the value (I add a special character in the value for reducer to recognize it is a page rank contribute), and output the key-value pair $< Node, OutputLink >$. The reducer receive all the page rank and output links, after add all the page rank value, output the $< Node, PageRankValueAndOutputLinks >$ for next PageRank iteration.
For the $i^{th}$ iteration, the output path is *inputPath-iter-i*, and that is the input path of next iteration.

## 2 Problems Encountered

At the beginning, I had no machine that had installed the Hadoop to run MapReduce, so I take some time to read the documents of Hadoop, Java, Shell script and then install Hadoop in my personal computer. Finally, I can write, test those code on my desktop.