MapReduce Exercises

Mining of Massive Datasets book (exercises of section 2.3)

Exercise 2.3.1

Design map-reduce algorithms to take a very large file of integers and produce as output:

(a) The largest integer.

Map function: It takes input and returns (key= 1, value= max(value)) for each entry.

Reduce function: Map Function send all inputs with key=1 thus Reduce Function need to only return Max value with key=1.

(b) The average of all the integers.

Map function: It takes input and return (key=1, value= (w,a)), where w is weight of chunk list and a is average.

Reduce function: Calculate weighted average on value list where the key is fixed:

$$\sum (w.a) / \sum (w)$$

(c) The same set of integers, but with each integer appearing only once.

Map function: It takes input and returns (key= input, value= input) for each entry.

Reduce function: Key list is output.

(d) The count of the number of distinct integers in the input.

Map function: It takes input and returns (key= input, value= input) for each entry.

Reduce function: Number of distinct integers:

$(\sum \text{(value)})$ / key