Questions from Chapter 2 of textbook: Data-Intensive Text Processing with MapReduce

Textbook can be downloaded from: https://lintool.github.io/MapReduceAlgorithms/

Read Chapter 2 and answer following questions:

2.1 Functional Programming Roots

a. What is a higher order function. How is it used in MapReduce? Give your own example.

2.2 Mappers and Reducers

Understand the purpose of mappers and reducers. (No need to write anything)

a. What are the restrictions on mappers and reducers. Can they have side effects? Explain

2.3 The Execution Framework

a. Write brief summary of the various responsibilities of the execution framework

2.4 Partitioners and Combiners

a. Write a brief summary of the roles of partitioners and combiners. Where are they executed and what are their input and output?

2.5 The Distributed File System

a. What are the responsibilities of the NameNode?

b. What assumptions about the operational environment of HDFS were made by the designers?

c. What are the advantages and disadvantages of the single-master design in HDFS?

2.6 Hadoop Cluster Architecture

a. Which of the following can a programmer control more precisely: the number of mappers (M) or the number of reducers (R)?

b. On what factors does the number of map tasks (M) depend?

c. How does the framework try to optimize the following:

- Number of input splits

- Location of map tasks

d. What guarantee does the framework make about the order of intermediate keys presented to the reduce method?