Zhihong Zhai CM2742

Question 1:

The basic idea of MapReduce is to divide a large task into several small independent tasks and run them in parrallel to complete the task faster. MapReduce has two phases: Map and Reduce. Take the MinTemperature as an example.

In Map phase, the original input is divided into several independent parts and is distributed to different nodes. Each node runs a Map method to generate ~~a~~ key-value pairs. For example, the mapper reads input one line a time, extracts the year and temperature. The year is key, temperature is value.

In Reduce phase, ~~the~~ a Reducer collects key-value pairs generated by mapper. All values that have the same key will be gathered together. Reducer can have algorithm to deal with the values to form the final output. For example, reduler combine the temperature value that has the same year number as a key. And find the minimum of temperature in that year and output the year as a key and the minimum temperature as value.

Question 2

Read: Clint requirs a file-read to the namenode. The namenode check the permission and return a block ID if the clint has the permission. The clint then start reading from that block ID. Clint perform a one-byte checksum every 512 bytes. If checksum mismatch, clint will inform namenode and throw checksum exception. The clint can read from another replica.

Write: Clint requirs a file-write to namenode ~~ID~~ Name Node checks the permission. Name Node returns block ID & locations. Clint breaks file in to packets. Clint send to ~~to~~ one target node. Then the node sends the ~~fast~~ packet to the next node. The final node will check the checksum. It will return ~~ackn~~ acknowlege if checksum passes.

Question 3.

1. hadoop fs -setrep [-R][w] <rep> <path/file>

This command sets the number of replica in the HDFS of a file. <rep> is the number and <path/file> is the file path.

hadoop fs -setrep 2 /tmp/text.txt, will set file /tmp/text.txt to have two replica in HDFS.

2. hadoop fs -chown [-R] [OWNER][:[Group]] PATH

It changes the owner and the group of a file.

hadoop fs -chown hadoop:hadoop text.txt. will change the owner of text.txt to hadoop and it belongs to hadoop group.

3. hadoop fs -touchz <File>

It creates a file that is 0 length. If the file exits and its length is not 0, throws error

hadoop fs -touchz tmp.txt. creates a blank Text file.

4. hadoop fs -mv <SRC> <DST>

Move <SRC> file to <DST> loction

hadoop fs -mv /tmp/text.txt /tmp2/ will move /tmp/text.txt to /tmp2/

5. hadoop fs -mkdir PATH

Makes a new directory <PATH>

hadoop fs -mkdir /tmp2/

will make a dirctory /tmp2/