MapReduce System

By : Mahmoud Abdelhadi

In this system we have 5 main functions, as explained above:

* First thing I do is launching the mappers processes, and each map has its own socket, read and write Object, then preparing them to receive the data from the main server as explained below.

1- Splitting:

The splitting in my project done line by line, which means that the lines separated to the mappers equally, the lines will be sent to mappers processes using Sockets.

2- Mapping:

While the mappers receiving lines from the server, it starts mapping the received text, I used a TreeMap to save the result in the class in a KeyValue format as requested.

After done mapping, the mappers will send the data as a list of KeyValue items back to the access point, then we will start with step 3.

* And now it will launch the Reducers processes, and also every reducer has its own Socket, Object reader and writer, then it will be prepared to receive the data from the access point to reduce it as explained below

3- Shuffling: after the access point receive the data, it will spread it to the reducers process, also it will spread it equally so that each reducer gets as much data as the others

4- Reducing:

While receiving the data, the process will reduce the input using the code in run () function, after that it will save the data also as KeyValue format, then it will send the data back to the server.

5- Writing:

After receiving the data from the reducers, it will write it to the output file using IOWrite Object.

In this project, we have the following classes:

1- ArrayListSocket

2- CheckUtils

3- Commands

4-Demo (For Testing )

5- GUI

6- IO

7- IOReport

8- IOWrite

9- KeyVal

10- MapReduceSystem

11- ProtocolMsg

12- Reducer

13- ShuffleManager

14- Shuffler

15- SocketServer

16- SocketServerFactory

17- SocketServerMap

18- SocketServerReduce

19- SocketThread

20- Splitter

21- Writer

Thank you.

Mahmoud Abdelhadi