Question B

1. For each Map function :

   Input : Part of log file

   Output : key-value pairs {“IP”, count (int) }, like (62.234.138.228, 5) (after combiner).

2. For each Reduce function :

   Input : key-values pairs produced by Map functions, like (62.234.138.228, 5).

   Output: key-values pairs that get the total number of each IP, like (62.234.138.228, 10).

3. The code has 3 files, master, mapper, reducer.

   Run the “master.py”, can get the result directly.

   Master divides the log file into 9 parts.

   9 Mappers runs at the same time. Each of them produces key-values pairs, like (62.234.138.228, 1). And then these Paris will be sorted and combined together, the final output key-values pairs will be (62.234.138.228, 5). And by partition, the pairs will be sent to the only one reducer.

   Reducer, get the output from mapper and add the counter together.

Taken together, it has 116 different IP, has 11087 connections.