Cloud Computing
COMP30520/COMP41110

Practical 05: Map/Reduce Programming

Deliverable: ***ClassNo_Surname_FirstName_StudentNo_ Pracitcal5.zip***
(Ex: COMP41110_Smith_John_12345 _Practical.zip)

A. Description

Let's consider the following programming problem:

Log File Analysis

A common task in digital forensics is the analysis of log files. Each line (entry) of a web server's log file normally contains important information such as: IP address, data and time of request, request line, HTTP status, etc. Write a program that takes a log file as an input and then extracts the following information from the log files: the total number of connections to the server (i.e. total numbers of entries), the number of distinct IPs and the number of entries for each IP.

B. Question:

Suppose that we use Map/Reduce programming model to code the programming problems above, answer the following questions:

1. Define the input and the output of the Map function(s) for each programming problem. Justify your answer.
2. Define the input and the output of the Reduce function(s) for each programming problem. Justify your answer.
3. Implement your Map/Reduce functions in Java or Python.

Please notice you can have one or many Map /Reduce functions.

C. Submission:

Submission should take place via Moodle on or before the deadline. Submission should consist of one zip file which contains the answer of all questions (documents and source codes).