More Practice With MapReduce Java Programs

|  |  |
| --- | --- |
| **Exercise Dir** | ~/workspace/log\_file\_analysis |
| **Eclipse Proj** | log\_file\_analysis |
| **Java Files** | SumReducer.java (Reducer)  LogFileMapper.java (Mapper)  ProcessLogs.java (Driver) |
| **JAR File** | processlog.jar |
| **Test Data** | weblog testlog |

**In this exercise, you will analyze a log file from a web server to count the number of hits made from each unique IP address.**

Your task is to count the number of hits made from each IP address in the sample (anonymized) web server log file that you uploaded to the ~/weblog directory in HDFS when you completed the “Using HDFS” exercise.

In the log\_file\_analysis directory, you will find fixme for the Mapper and Driver.

1. Using the fixme files in the log\_file\_analysis project directory, write Mapper and Driver code to count the number of hits made from each IP address in the access log file. Your final result should be a file in HDFS containing each IP address, and the count of log hits from that address.

NOTE: The Reducer for this exercise performs the exact same function as the one in the WordCount program you ran earlier. You can reuse that code or you can write your own if you prefer.

1. Build your application jar file following the steps in the previous exercise.
2. Test your code using the sample log data in the ~/weblog directory.

NOTE: You may wish to test your code against the smaller version of the access log you created in a prior exercise (located in the ~/testlog HDFS directory) before you run your code against the full log which can be quite time consuming.

**END**