Using Counters and a Map-Only Job

**In this exercise you will create a Map-‐only MapReduce job.**

Your application will process a web server’s access log to count the number of times gifs, jpegs, and other resources have been retrieved. Your job will report three figures: number of gif requests, number of jpeg requests, and number of other requests.

**Hints**

1. You should use a Map-‐only MapReduce job, by setting the number of Reducers to 0 in the driver code.
2. For input data, use the Web access log file that you uploaded to the HDFS /~/weblog directory in the “Using HDFS” exercise.

|  |  |
| --- | --- |
| **Exercise Dir** | ~/workspace/counters |
| **Eclipse Proj** | counters |
| **Java Files** | ImageCounter.java (Driver)  ImageCounter.java (Mapper) |
| **JAR File** | counters.jar |
| **Test Data** | weblog  testlog |

NOTE: We suggest you test your code against the smaller version of the access log in the ~/testlog directory before you run your code against the full log in the ~/weblog directory.

1. Use a counter group such as ImageCounter, with names gif, jpeg and other.
2. In your driver code, retrieve the values of the counters after the job has completed and report them using System.out.println.
3. The output folder on HDFS will contain Mapper output files which are empty, because the Mappers did not write any data.

The results should look something like this:

JPG = 2629976

GIF = 73682

OTHER = 1774178

**END**