

Issued on: March 30, 2014

Due by 23:59 PM, April 08, 2014

**Problem 1.** Attached PDF document `bigdatainfrastructure1-vm-....pdf` among other things on slide 15 and 16 describes a Hadoop Streaming job. Demonstrate that scripts `map.R` and `reduce.R` which could be found in the directory `airline\src\deptdelay_by_month\R\streaming` of the downloaded Git repository work.

Please note that your `hadoop-streaming.jar` file is most probably located at:

```
/usr/lib/hadoop-0.20-mapreduce/contrib/streaming/hadoop-streaming-2.0.0-mr1-cdh4.6.0.jar
```

**Problem 2.** Replace scripts `map.R` and `reduce.R` with scripts in any other scripting language of your choice. Demonstrate that they work.

Please describe every step of your work and present all intermediate and final results in a Word document. Please submit to the class drop box. For issues and comments please visit the class Discussion Board or send emails to [cscie63@fas.harvard.edu](mailto:cscie63@fas.harvard.edu)