Log processing is one of most common tasks in analytics. The issue is that the log files are huge and the records in the log file do not always comply with a fix schema. Download the file WindowsUpdate.log from BlackBoard and move it to ~/hive directory in your VM. You will use this file as your data set.

Study the file and make sure you understand what the first five fields in this file mean.

What we need to do is to find useful information in this log file. Specifically, we are after the following:

1. How many total events does this log file contain?
2. What are the event types and how many of each type does the file contain?
3. How many events have happened per day?
4. How many downloads were made in total?
5. How many downloads per day were made?
6. The date and time information about each download on 2014-03-11.
7. Which day had the highest number of downloads.

**Setup:** Use the WindowsUpdate.log as your data set.

**To Do:**  Use Hive and write Hive QL statements that produce the answers to each of the above questions.

**Deliver:** Email your Hive scripts – print and turn in the scripts and their answers, including the table(s) creation statements and all quires.

**Note:** Please run Hive is silent mode to avoid printing the MapReduce output on your handout.