Assignment #8

1.
```
2015-04-18 00:29:53,825 [main] WARN  org.apache.hadoop.conf.Configuration -
2015-04-18 00:29:54,458 [main] WARN  org.apache.hadoop.conf.Configuration -
grunt> ls
hdfs://localhost:8020/user/ec2-user/mydat.txt<r 1>      54
grunt>
```

2. The phrase "USING PigStorage(',')" is necessary because it's telling pig which load function and delimiter to use when parsing the input file.

3. The "dump dat;" command will print out the contents of the dat alias.

4.
```
grunt> ls
hdfs://localhost:8020/user/ec2-user/data1<r 1>  36
hdfs://localhost:8020/user/ec2-user/data2<r 1>  28
hdfs://localhost:8020/user/ec2-user/mydat.txt<r 1>      54
grunt> cat data1
1,2,3
4,2,1
8,3,4
4,3,3
7,2,5
8,4,3
grunt> cat data2
2,4
8,9
1,3
2,7
2,9
4,6
4,9
grunt>
```

5. To verify that aliases are created OK, you can use the describe and dump commands that will print the schema and contents of the alias.

6. If the join was changed to left outer join, 8 records would be returned.

```
(1,2,3,1,3)
(4,3,3,4,9)
(4,3,3,4,6)
(4,2,1,4,9)
(4,2,1,4,6)
(7,2,5,,)
(8,4,3,8,9)
(8,3,4,8,9)
grunt>
```

7. We didn't use "USING PigStorage(',')" because we omitted the USING clause which causes pig to use the default parser which is in fact the PigStorage parser using the default delimiter of tabs.

8. The output of the describe command on the B alias is telling us that alias A will be grouped by the column containing the chararray type.

9.
```
(Car,3)
(River,2)
(Deer,2)
(Bear,2)
grunt>
```

10. Midterm question: What is the default parser when using the LOAD operator without the USING clause and it's default delimiter?