CSE 891: Homework 4
Due date: Sunday, March 27, 2013 (before midnight)

1. In this homework exercise, you will write Pig Latin scripts for processing movie ratings data. You already downloaded the file earlier. It contains ~50.000 movies in a csv table listing (ID,name,rating,year, length).

. (a) Write a Pig Latin script that reads the movie data file (u.item) and finds the release date for the movie titled Star Trek.

. (b) Write a Pig Latin script that reads the movie data file and find what the lowest score is and how many movies have that lowest score.

. (c) Based on the result from the previous question, write a Pig Latin script that returns the title of those lowest rated movie.

. (d) Since we deal with big data, please write a Pig Latin Script to find out how many Star Trek the next generation episodes have "data" in the title.

2. The second part of this homework is all about Hive, so please redo all questions from above but instead of writing Pig Latin scripts, write Hive scripts to do so. You can write your own mapper/reducer special scripts, but I recommend not to do so, since that increases the difficulty of the task.

TIP: For some problems you need to search within strings, please check out the following link which explains regular expression usage in Pig:

http://pig.apache.org/docs/r0.8.1/piglatin_ref2.html#Expressions