HW5B- PIG – Due date is 14th Nov.

|  |  |
| --- | --- |
| Instructions | Points |
| * Download “[movies.txt.gz](https://snap.stanford.edu/data/movies.txt.gz)” From the link <https://snap.stanford.edu/data/web-Movies.html> and preprocess the data so that it can be loaded to PIG using the “LOAD” command. | 20%  joint |
| * Write a query that shows some analytics from the ‘review text’. | 15% (student 3) |
| * Write a query that shows some analytics from the ‘review summary’. | 10% (student3) |
| * Write an aggregate query that involves column ‘helpfulness’.   Use sort and limit also | 25% (Student 1) |
| * Write an aggregate query that involves column ‘score’. Show the result **excluding** score less than 2.0 | 25% (student 2) |