

Chapter 5 – Advanced Pig Latin

Exercise

1. Find all movies that received a rating of 5. Use the files `u.data` and `u.item`. Only the movie name should be returned (and no duplicates). Store the results in HDFS.
2. Find the movies with an **average** score greater than 4.5. Hint: use `COGROUP`
3. Use `SPLIT` to read the `u.item` data and output two results. One should contain all the comedy movies (field `$10`) and the other should be documentaries (field `$12`).
4. Find the movies that were released in 1960. Hint: use `matches`