Step 1) Scraped data from reddit based on movie subreddit and comments data related to movie date wise.

Step 2) Cleaned the data of empty lines, hyperlinks and timestamp.

Step 3) Used MapReduce to combine comments based on date to analyze the data later.

Step 4) Used MapReduce to Profile data. I have 2 columns where date is in range of a month before movie release and a month after it, while comment can be of any length.