❋ HPC - Assignment - 06 ❋

Aim :- Use movies dataset. Write the map & reduce methods to determine the avg. ratings of movies. The i/p consists of a series of lines, each containing a movie number, user no., rating & a time stamp. The map should emit movie no. & list of rating & reduce should return for each movie no. a list of avg. rating.

Input :- In N-records of movies.

Output :- Top rated movie using map-reduce.

Theory :-

- The term "Top - N rated movies" means the first N movies sorted by the no. of ratings for each movie.
- If there is a tie in the last movie, all the movies will be counted except the last movie in the top list.
- If there is a tie in the last movie of the top list, all the movies with same no. of rating will show.

Example :-

Taking top, 5 rated movies for example -

I/p → A:5, B:4, C:4, D:3, E:3, F:3, G:2

O/p → A:5, B:4, C:4, D:3, E:3, F:3

o Explanation :-

- Movie A has the highest no. of rates :5, so now 4 to go.
- Movie B & C are ties, so now 2 to go because they are counted 2 movies.
- Movie D is included because its no. of rating is 3. Now 1 seat is left.
- Movie G is not included since there are actually 6 movies in top list now.

★ Conclusion :-

Found the top-N-rated movies using MapReduce.