

Report

1. Average

Map: First, I removed the first line by checking whether there is "userId", then separated the 4 columns and drew the second and the third. Put movieId into key and rating into value.

Reduce: Counted the number of value and calculated the sum of value. Finally put the movieId into key and the average rating into value.

2. ExtractUserMovies

Map: The same as above. Then filtered rating with an if.

Reduce: Put the values into only one String.

3. CountPairs

Map: Separated the key and value of the results of last question. The character between key and value is \t. Then put the movie with bigger id first to avoid duplicate.

Reduce: a simple count.

4. SelectPairs

Map: Separated the key and value of the results of last question. Then separated the two movies in key. Put the second movie id and the value into the new value.

Reduce: Traversed all those values to find the movieId with the biggest count number. When there are two movies have the same count number, we just kept the first and did nothing.

5. NameMovie

I wrote the two maps according to the subject. In order to distinguish the two values, I gave them a flag. For id, it's 0. For title, it's 1. Then we traversed the values. When we meet the value with the flag 0, we make it key. When we meet the value with flag 1, we make it value.

But my codes couldn't run.