Map Reduce Algorithm

MR01 MovieTitlesGenresByID (Mapper Only)

Map out: List<K2,V2> _> list<movield, title \t genres>

movies.csv
<movield,title,genres>

MR02 MovieRatingsByID

Map out: List<K2,V2> _> list<movield,rating>

Red out: List<K3,V3> _> list<movield, sumRatings|countRatings>

MR09 UserRatingFrequency

Red out: <K3,V3> _> <userId, ratingsCount>

Map out: List(K2, V2) _> list(ratingCounts, 1)
Red out: <K3,V3> _> <ratingCount, frequency>

MR10 SortUserRatingFrequency (IntComparator)

Map out: List(K2, V2) _: list(ratingCount, frequency)
Red(1) out: <K3,V3> _> <ratingCount, frequency>

MR03 JoinMovieTitlesGenresRatings (TextPair, FirstComparator)

Reduce side join

Map1 out: List<K2,V2> _> list<(movield, 0), title \t genres> Map2 out: List<K2,V2> _> list<(movield, 1), sumRatings|countRatings>

Red out: List<K3,V3> _> list<movield, title \t genres \t sumRatings | countRatings>

MR04 CalculateMovieTitleRatings

Map out: List<K2,V2>_> list<title, sumRatings | countRatings>

Red out: List<K3,V3> _> list<title, avgRating>

MR05 SortMovieTitlesByRating (Partitioner, DoubleComparator)

Map out: List<K2,V2> _> list<avgRating, title>
Red(6) out: List<K3,V3> _> list<avgRating, title>

MR06 CalculateMovieGenreRatings

Map out: List<K2,V2> _> list<genre, sumRatings|countRatings>

Red(1) out: List<K3,V3> _> list<genre, avgRating>

MR07 SortMovieGenresByRating (Partitioner, DoubleComparator)

Map out: List<K2,V2>_> list<avgRating, genre> Red(1) out: List<K3,V3>_> list<avgRating, genre>

MR11 MovieGenresTitlesRatings (TextPair)

Map out: List<K2,V2>_> list<(genre,title), sumRating | countRatings> Red out: List<K3,V3>_> list<(genre, title), avgRating>

MR12 MovieRatingsTitlesByGenre (Multiple Outputs)

Each genre in its own directory

Map out: List<K2,V2> _> < list<(genre, title \t avgRating>> Red out: List<K3,V3> > < NullWritable, genre \t avgRating \t title>

MR13 SortMoviesByRatingWithinGenre (DoubleComparator)

Run for each genre directory (20x)

Map out: List<K2,V2> _> list<avgRating, title>
Red(1) out: List<K3,V3> _> list<avgRating, title>

Reducer specification legend

No Reducer

Reducers in Driver

Reducers in Shellscript