

1. READING ASSIGNMENT

Read the following papers

- Design patterns for efficient graph algorithms in MapReduce
- Big data as the new enabler in business and other intelligence

2. Create a Writable object that stores some fields from the the NYSE dataset to find

- the date of the max stock_volume
- the date of the min stock_volume
- the max stock_price_adj_close

This will be a custom writable class with the above fields.

Mapper will use this writable object as a value, and Reducer will use this writable object as a value.

3. Re do HW4-Part3, but use SecondarySorting to sort the values based on AccessDate in a Descending Order.

4. Determine the average stock_price_adj_close value by the year.

Choose an implementation in which a Reducer could be used as a Combiner. (discussed in the lecture, and available in the slides).

5. Using the Movielens dataset, determine the median and standard deviation of ratings per movie.

Iterate through the given set of values and add each value to an in-memory list. The iteration also calculates a running sum and count.

6. Redo Part 3 using Memory-Conscious Median and Standard Deviation implementation as explained in the Slides (MR Summarization Patterns Slides). Use a Combiner for optimization.

[Big data as the new enabler in business and other intelligence.pdf](#)

[Design patterns for efficient graph algorithms in MapReduce.pdf](#)