Attached Files:

- Big data as the new enabler in business and other intelligence.pdf Big data as the new enabler in business and other intelligence.pdf - Alternative Formats (2.787 MB)
- MapReduce is Good Enough.pdf MapReduce is Good Enough.pdf - Alternative Formats (1.049 MB)
- Big Data Processing with Hadoop-MapReduce in Cloud Systems.pdf Big Data Processing with Hadoop-MapReduce in Cloud Systems.pdf - Alternative Formats (556.006 KB)

1. READING ASSIGNMENT
Read the following papers
- Big data as the new enabler in business and other intelligence
- Big Data Processing with Hadoop-MapReduce in Cloud Systems.
- Mapreduce is good enough?

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. Redo Part2 of this assignment, but cram multiple values (max stock_volume, min stock_volume, max stock_price_adj_close) into a Text object with some delimiter. Use a Combiner. Compare the running time of Part 2 to Part 3.

4. Re do HW3-Part3, but use SecondarySorting to sort the values based on AccessDate in a Descending Order.

5. 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).

6. 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.

7. Redo Part 5 using Memory-Conscious Median and Standard Deviation implementation as explained in the Slides (MR Summarization Patterns Slides). Use a Combiner for optimization.