


HW4

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)

1. READING ASSIGNMENT

Read the following papers

- Big data as the new enabler in business and other intelligence

2. READING ASSIGNMENT

<https://learning.oreilly.com/library/view/hadoop-essentials/9781784396688/ch03.html>

<https://learning.oreilly.com/library/view/professional-hadoop/9781119267171/c03.xhtml#c3>

<https://ebookcentral.proquest.com/lib/northeastern-ebooks/reader.action?docID=1566514&ppg=173>

3. 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.

4. Redo Part3 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. You could measure the running time programmatically, or use your smartphone's timer.

5. Determine the average stock_price_adj_close value by the year.

Use a Writable object to pass count and local average in which a Reducer could be used as a Combiner.

6. Top 10 most visiting IP Addresses (HW3-Part3). This can be achieved by 2 jobs :

A map reduce job to find how many time each item is bought.

The second job, sort items based on number of times it was bought, and get top 10 items.

The second MapReduce job swaps key and value so that we get words sorted i