



Homework 2

Due by Feb 3 Tue 24:00

Subject: Combiner and In-Mapper Combining (Local Aggregation)

In this assignment you will develop better MapReduce programs using “*local aggregation*”. Please read section 3.1 of the textbook (Lin&Dyer book).

In the previous assignment you developed a **WordLengthHistogram** program that generated a histogram of word lengths (in 4 categories) on a large text corpus. Your program is probably a basic MapReduce program that can be improved using (1) combiner and (2) in-mapper combining.

Make these two changes on the program, (1) add a combiner (WordLengthHistogramWithCombiner) first and then (2) add an in-mapper combining (WordLengthHistogramWithInMapperCombining).

Test all 3 programs on your test dataset and report the time improvement in a report.

Submit your homework via google drive. Use the following header in your programs.

BiL401/501 Distributed Data Processing and Analysis (Spring 2015)

Homework

Date: ...

Name: **Firstname Lastname**