

Data Intensive Computing Final Assignment

This program starts tokenizing the input file so that every line is mapped once. Each line gets split based on the commas, so that the call's source, destination, start time, and end time can be used for further calculation. Then the calls duration is calculated using the above information and the toMilliseconds method. Lastly the mapper maps the duration to the calls source. Next the reducer takes all of the saved durations for a single call source and adds them together. The reducer then outputs the source and its total duration to the output if that duration is greater than 60 minutes. The program runs the reducer for every source it was given from the mapper, and creates the final output.