Dev Ganesh

REPORT FOR HW2

Method 1: I wrote code in **R** to instantitate **22 SOCK** cores. Each of these cores put together frequency tables for its set of files. After this, the code put together all the frequency tables into one table.

I ran my code on Hilbert and it took the 22 processors **15.2 minutes** to process all 81 CSV files. Before this, I ran my code locally on my laptop with just 2 processors and it took ~ **4 hours** to process the files.

With my parallel code, I get the **mean** to be **6.98276,** the **standard deviation** to be **30.22053** and the **median** to be **0**.

Method 2: I wrote a mapper and reducer in Python. I ran the code locally to check if it Is working and it processed the CSV files fine. When I ran it on the Hadoop machine, I kept getting a map failure error. The cause for this requires further investigation.