Scott Berry

Running Hadoop Streaming Command (full logging was way too much too screenshot)

A picture containing text

Description automatically generated

With the output logging to the following file (Seems clear that cleaned data would yield more insightful results)

Text

Description automatically generated

The Hadoop Streaming command was enabled by creating a HDFS directory which I placed the IUB text file in. Along with providing the mapper and reducer python scripts as arguments, the output directory was created in the HDFS directory. With mapper, the IUB text was mapped into keys and values of individual words and counts of their occurrences. With reducer, the keys are aggregated and made unique with their corresponding count value. With this command and scripts, I was able to quickly create a count of every unique word in the IUB Wiki text file.