Scott Berry

Running MapReduce(way too much logging to include all):

Text

Description automatically generated

Text

Description automatically generated

With the following output:

Text

Description automatically generatedText

Description automatically generated

This Hadoop job is tasked with taking in StackOverflow posts and their metadata and compiling the counts of every tag’s appearance over each month. This job successfully created an output of 8100 unique month:tag combinations and recorded the number of appearances of each tag in each month. The mapper.py logic takes in the csv of StackOverflow posts and tracks every post’s creation date and tags to be sent to the reducer. The reducer then reads in every line and gets the month from the creation date using decision tree logic from the scraped month int as well as a list of tags from the tag field using text cleaning logic. Iterating over each tag, the tag count is added or incremented to a dictionary tracking all tag data for the corresponding month. The output is then iterated over for each month with every tag and count printed.