Tracy Michaels

Tmichaels1 002430918

Big Data Programming Assignment 4

1. - The program takes a file with a list ‘webpages’ and their out going links to other ‘webpages’

* The list is printed

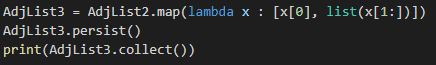




* The initial text file is split into a list of integers using a lambda function



* Using another lambda function to separate the pages from their outlinks into tuples which is them printed to the console





* Then the initial page rank values are calculated and stored as values in a list which is then printed to the console





* The main loop then starts of the pagerank algorithm
* A new list is created which contains the pagerank values added to the tuples of the pages and outlinks which is then printed to the console (screen shots are from iteration 1 out of 30 total)





* Each outlink is then mapped via a lambda function to its corresponding page and these are stored as tuples in a new list called contributions and printed to the console



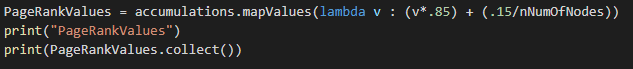


* The values of same keys are summed via another lambda function in stored as a list of tuples called accumlations which is then printed to the console



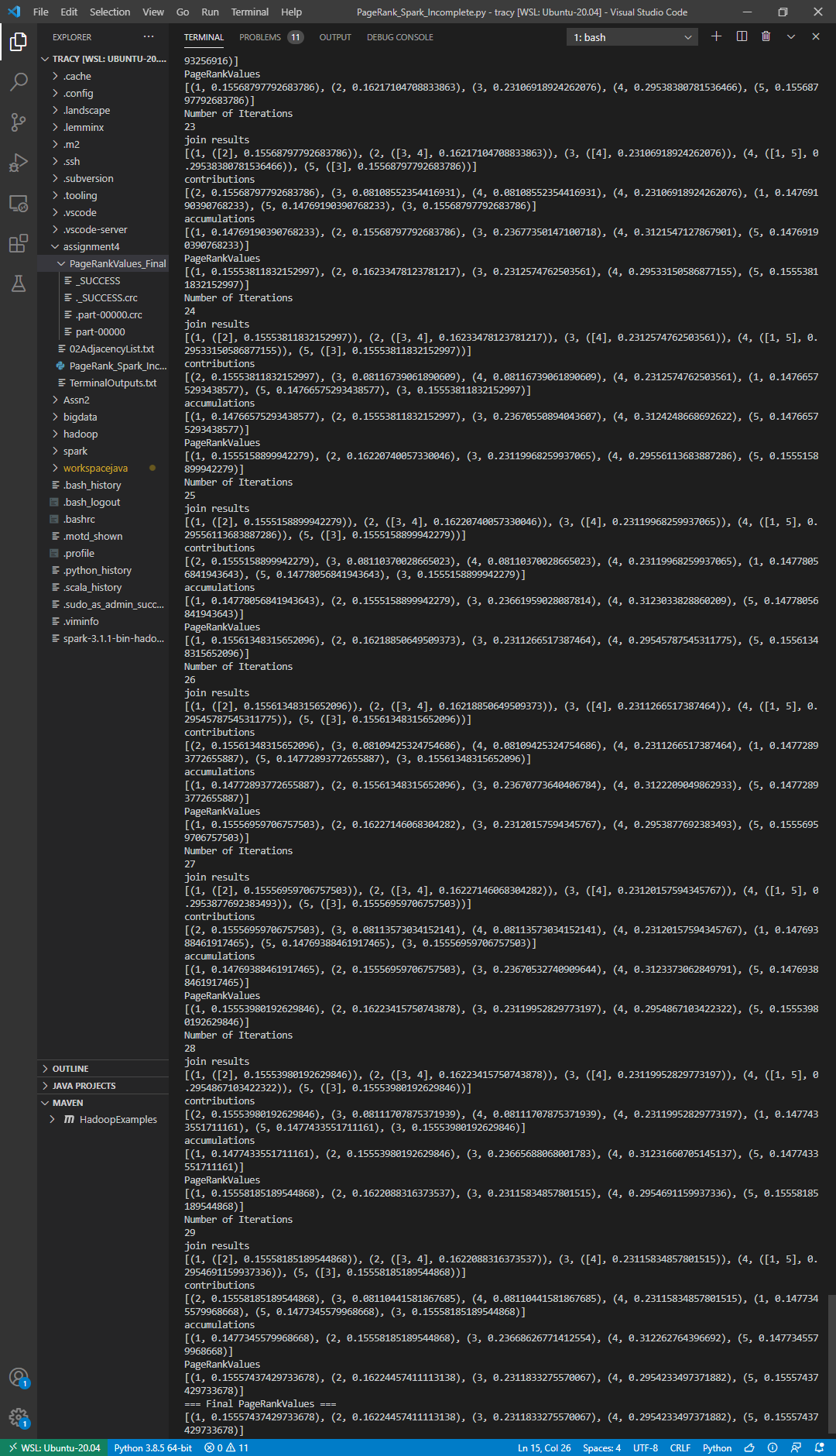
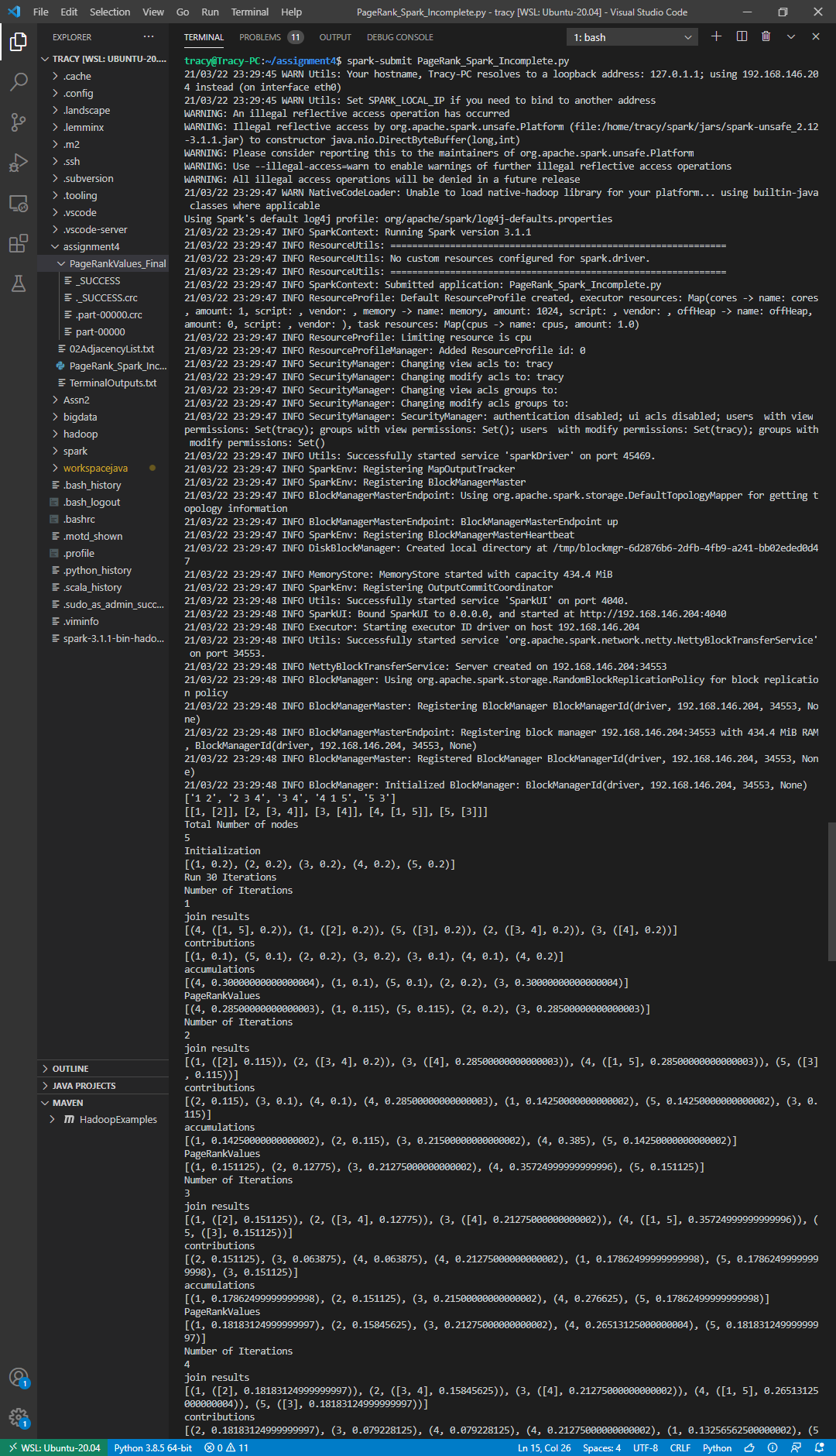


* The page rank values are then recalculated and stored back in the pagerankvalues list and printed to the console





* This process is then repeated 29 more times with the new page rank values

1. I skipped a few iterations of the output to save space in this document (full output provided in .txt file with submission
2. 

2.2) each iteration gave intermediate values for the page ranks of each page which converged over time to their final results