# CPSC 313: Distributed and Cloud Computing Spring 2022

### Lab 1: Count words in files

#### Summary

You will create a single python code file named "count\_words.py" that takes a list of files as input, and count all the words in the list of files, printing out the list of words with their counts.

```
You should have the following constants & functions: STOP_WORDS =
```

```
['a', 'an', 'and', 'are', 'as', 'be', 'by', 'for', 'if', 'in', 'is', 'it', 'of', 'or', 'py', 'rst', 'that', 'the', 'to', 'with',]
```

def count\_words(file\_name): # take a file and output the word list with the counts for that file as a tuple (or list). This may look like: [['foo', 22], ['bar', 13], ...]

def main(): # main function: get the files from input (or \*args). Iterate through the files counting words in each file, then combine results from each

```
if __name__ == '__main__':
    main()
```

#### Requirements

- You should have a test file that tests both your count\_words function and your main loop.
- You must log relevant information from your working code into a log file: "count\_words.log"
- You must compute the total time it takes to count all the words in all the files, and output that clearly to the log file.
- You must follow the style guide (document is in Teams) with the following exceptions:
  - o naming: Single letter variable or function names are NOT allowed.
    - All names must be meaningful and descriptive. Instead of "for x in list:" you may use "for loop\_control in list" or something more meaningful.
  - You may use tabs instead of spaces, and your indentation is encouraged to be tab or 4 spaces instead of 2 spaces.

## Submission

- Your python file (count\_words.py)
  Your test file (test\_CW.py)
  A log file that shows the output of your test run and any log statements from the code