Isabella Coraci   Darsh Joshi

CIT 322 Lab

This lab uses two Python scripts to illustrate the concept of distributed computing by utilizing different "workers." The work that is being divided in simple terms counts the number of times each word appears in a text. Below the two Python scripts are described more in depth.

First, in worker.py a TCP/IP socket is created and then a host and port are created so that they are ready for connections from clients. Namely, a serialized object is expected. The loop is ended if nothing is received. If something is received, pickle.loads(data) "unpickles" the data. If there is text still left, it is split into works using split() and the amount each word appears is recorded. "Done" stops it. Once this is completed, the dictionary of word counts is pickled and sent to the client.

In coordinator.py, the function send_chunk_and_get_counts(worker_host, worker_port, text_chunk) is what does the work of sending the pickled data to the workers and receiving the completed dictionary. The main function defines the host and port numbers. The text is divided into chunks which are sent to different workers. After this is done, the counter object aggregates the respective word counts. Once this is completed, a done signal is sent to each worker to turn off and the word counting work is completed using distributed computing.