

# **CS 434: Parallel and Distributed Computing Lab 4**

Kweku Yamoah: 71712022

April 19, 2021

# Introduction

This report was generated as a requirement of Lab 4 for Parallel and Distributing computing. The report contains two sections; Part 1 and Part 2. Part 1 is the Lab where I experiment with the MaoReduce framework and my understanding in class. Part 2 is the project which focuses on implementing matrix multiplication using MapReduce algorithms.

## Part 1

### MapReduce Framework & Why

My selected MapReduce framework was MrJob. I chose MrJob because of the following reasons:

- it provides an easy route for writing python programs that run on Apache Hadoop.
- it provides an easy switch for input and output formats with a single line of code.
- All codes are written in a single python class which enforces the understanding of OPP concepts.
- it has an extensive documentation with examples as reference, which, makes learning of the framework easy.

Installing MrJob was easy because it require running a command with the python pip package. However, understanding the documentation and how to submit my first job took some time to understand.

### Word-count Algorithm

The word count algorithm inherited

## Part 2