Multi-Threading Programming and Inter-Process Communication

M.Sc. course on "Technologies for Big Data Analysis" - Assignment 1

CHRISTOS BALAKTSIS (1234) and VASILEIOS PAPASTERGIOS (1234), Aristotle University, Greece

1 Introduction

The current document is a technical report for the first programming assignment in the M.Sc. course on *Technologies for Big Data Analysis*, offered by the *DWS M.Sc Program*[1] of the Aristotle University of Thessaloniki, Greece. The course is taught by Professor Apostolos Papadopoulos [2]. The authors attended the course during their first year of Ph.D. studies at the Institution.

The assignment contains 4 sub-problems and is part of a series, comprising 3 programming assignments on the following topics:

Assignment 1 Multi-threading Programming and Inter-Process Communication
Assignment 2 The Map-Reduce Programming Paradigm
Assignment 3 Big Data Analytics with Scala and Apache Spark

In this document we focus on Assignment 1 and its 4 sub-problems. We refer to them as *problems* in the rest of the document for simplicity. The source code of our solution has been made available at `https://github.com/Bilpapster/big-data-playground`.

Roadmap. The rest of our work is structured as follows. We devote one section for each one of the 4 problems. That means problems 1, 2, 3 and 4 are presented in sections 2, 3, 4 and 5 respectively. For each problem, we first provide the problem statement, as given by the assignment. Next, we thoroughly present the reasoning and/or methodology we have adopted to approach the problem and devise a solution. Wherever applicable, we also provide insights about the source code implementation we have developed. For problems 2 and 4, we complete the respective sections with a discussion about alternatives or improvements the solution could accept, in order to successfully support more complex requirements. Finally, we conclude our work in section 6.

2 Problem 1: Concurrent Array-Vector Multiplication

This is the section for the first problem.

2.1 Problem Statement

It is known that in Linear Algebra we can multiply a matrix with a vector from the right-hand side, provided that the number of columns in the matrix equals the number of rows in the vector. For instance, given a matrix $\mathbf{A}$ with dimensions $n \times m$ and a vector v with dimensions $m \times 1$, then the product $\mathbf{A} * \mathbf{v}$ is an $n \times 1$ vector, which results from the implementation of the well-known method of multiplying a matrix with a vector. An example is given following: **TODO**

Provided that we are capable of using k threads, where k is a power of 2 and the matrix has dimensions $n \times m$, where n is also a power of 2 and $n > k$, design a solution that computes the product $\mathbf{A} * \mathbf{v}$ using k threads with

[1] https://dws.csd.auth.gr/
[2] https://datalab-old.csd.auth.gr/~apostol/

Authors' Contact Information: Christos Balaktsis (1234), balaktsis@csd.auth.gr; Vasileios Papastergios (1234), papster@csd.auth.gr, Aristotle University, Thessaloniki, Greece.

the most efficient way. Your solution has to initialize both the matrix $\mathbf{A}$ and the vector $\mathbf{v}$ with random numbers in the range $[0, 10]$.

3 Problem 2: Simulating a pandemic

This is the section for the second problem.

3.1 Problem Statement

4 Problem 3: Key-value server store

This is the section for the third problem.

4.1 Problem Statement

5 Problem 4: Multi-server producer-consumer interaction

This is the section for the fourth problem.

5.1 Problem Statement

6 Conclusion

This is the section for the conclusion.