CPSC 479 Project 2: Introduction to HPC - Data Science

Prof. Doina Bein, CSU Fullerton

dbein@fullerton.edu

Introduction

In this project you will design and implement an algorithm that solves some data science problem. The topics one can consider are:

Some data clustering techniques
Supervised learning and nearest neighbor classification
Core-sets on large datasets; parallel algorithms seeking core-sets
Finding dense subgraphs or large graphs

The program needs to use either MPI, openMP, or CUDA programming. No sequential programming is considered as an acceptable solution for this project.

What to do

1. Describe the problem you are solving and write clear pseudocode for the algorithm, include what parameters are needed to execute the program and how to execute the program. Include all in a PDF file.

2. Implement your algorithm using MPI/openMP/CUDA primitives in the language of choice

3. Compile and execute the program.

4. Submit the source code

5. Do an oral presentation of about 10 minutes on your project in week 15 (Monday or Wednesday) or week 16 after the final exam. The oral presentation will count for 20 points towards the final exam score. If you decide not to do the oral presentation, that you will lose the 20 points from the final exam score.