Institute of Informatics
System and Network Engineering Lab

Stephen Swatman - s.n.swatman@uva.nl
Ana-Lucia Varbanescu - a.l.varbanescu@uva.nl
May 9th, 2022

Performance Engineering

Assignment 3: Data-centric/Statistical modeling

Assignment 3.1: Statistical modeling: Matrix Multiplication (2p)

Select a high-performance parallel version of your matrix multiplication (either for a multi-core CPU or for the GPU), and build a statistical model for this application. Empirically validate your model and discuss its accuracy. Compare the accuracy to that of the previous models.

Assignment 3.2: Parallel SpMV (2p)

You are provided a sequential version of a Sparse Matrix - Vector (SpMV) multiplication (Canvas, DAS5) using a few sparse storage formats (COO, CSR, CSC) - note that other such formats do exist (e.g., ELLPACK, HYB). Parallelize this kernel using two such formats (given or new) *either* for a multi-core CPU or a GPU. Benchmark the two paralell versions (i.e., two formats = two versions) using the provided matrices (a3_matrices.txt). Which one of the two data formats is better and why?

Note Through the data formats, we **actually use** the fact that the matrices are sparse, as a data-driven optimization.

Assignment 3.3: Analytical modeling: SpMV (2p)

Build an analytical model for one of the two SpMV parallel kernels. Calibrate the model as much as possible (e.g., using microbenchmarking and/or reusing results from assignment 2).

Assignment 3.4: Statistical modeling: SpMV (3p)

Build a statistical model for the same parallel SpMV kernel.

Assignment 3.5: Validation and accuracy checking: SpMV (2p)

Validate your models empirically. What is their accuracy? Compare the two modeling techniques for SpMV and discuss advantages and disadvantages.

Submission due date: May 29, 2022