Universität Konstanz

Introduction to Computing for the Social Sciences
Exercise Sheet for Session 12

Prof. Dr. David Garcia

Exercise 1: Parallelization

You are working on parallelizing a program in which 80% of the computation is parallelizable. Explain how you calculate your answers to the following questions:

a) What speedup factor will you get using 4 processors instead of 1?

b) How many processors do you need to get a speedup factor of 10?

Exercise 2: Conceptual questions

a) What is predicted by Moore's law? What is its limit and a workaround?

b) What is the difference between concurrency and parallelism? List two advantages and one disadvantage of concurrency in comparison to running tasks sequentially.

c) In your own words, explain speculative execution and one of its security risks.

d) You work for Melon Usk, who asks you to calculate the mean number of likes of 3 Billion X posts. Your servers only have memory for analyzing at most one Million posts. Describe in your own words how you would approach this problem. How would it differ if he asks you to calculate the standard deviation instead?

Exercise 3: JQ

We are going to learn how to use jq, a "secret" tool many data scientists use. jq is a lightweight and flexible command-line JSON processor that can save you lots of coding and runtime in your everyday data science.

We will go together through **Navendu Pottekkat's interactive tutorial:**
https://navendu.me/posts/jq-interactive-guide/

Strongly recommended: install jq locally for your Unix-like command line interface:
https://jqlang.org/