

# CS425: Computer Systems Architecture

## Homework Problem Set 1

Assignment Date: Wednesday 16/10/2024

**Due Date: Tuesday 29/10/2024 23:59:59**

**Instructions:** Solve all problems, create a .pdf file and send it via e-mail to HY425 course e-mail ([hy425@csd.uoc.gr](mailto:hy425@csd.uoc.gr)). Set the e-mail subject: HY425 - Homework 1

### Problem 1 (35 points)

Assume that you run four applications (X, Y, Z, W) on a 20-core processor system, given the application characteristics provided in the table below:

Application	X	Y	Z	W
% resources need	40	25	20	15
% parallelizable	50	80	60	90

- Calculate the speedup achieved when running application X on the entire system, compared to running it serially.
- Calculate the speedup achieved when running application W on the entire system, compared to running it serially.
- Application X needs 40% of the resources, so we assign it 40% of the cores. Find the overall speedup if X runs in parallel but remaining applications run serially.
- All applications are assigned the corresponding portion of the cores and all run in parallel. Find the overall speedup.

### Problem 2 (35 points)

The servers used by the major cloud providers provide adequate compute capacity to sustain the highest load demand. However, these servers operate at only 60% of their peak. The power consumption does not scale linearly with the load, so the consumption is approximately 90% of the maximum power when these servers operate at 60% of their peak. One solution would be to turn off the servers, but the problem would be long restart times in response to larger load demands. You propose a new system that does quick restart. The quick restart mechanism requires 20% of the maximum power while waiting in this state.

- Calculate the power savings in case of turning off 60% of the servers.
- Calculate the power savings when 60% of the servers are in the “quick restart” state.
- Calculate the power savings if the voltage is reduced by 20% and frequency is reduced by 40%.
- Calculate the power savings if 30% of the servers are in the “quick restart” state and 30% of them are completely off.

### Problem 3 (30 points)

Apart from scalability and throughput, availability is a key consideration when designing servers.

- Your institution has a processor chip which exhibits a Failure In Time of 10. What is the Mean Time To Failure for this system?
- If engineers need one day to fix and get the system running again, find the system’s availability.
- The institution wants to build a supercomputer with 2048 nodes using these chips, and assigned

to you the MTTF estimation of the following two scenarios:

- a.** To cut down costs, the institution will deploy unreliable computer units to set up the supercomputer, meaning that if one fails, they all fail.
- b.** The company will use reliable computer units to build the supercomputer. In this case, the system will experience catastrophic failure only if  $\frac{1}{3}$  of the units fail.

Note: Consider the worst case scenario of simultaneous failures.

Which scenario would you recommend if the extra upfront cost of the second scenario is 3x compared to the first one? Explain your calculations.