[50 marks total]

Due Wednesday 21st Sept, midnight

Question 1 [15 marks]

(a) Which of the following types of system (i.e. Sequential, Concurrent, Parallel, and Distributed) would be most appropriate to implement a weather forecasting simulation? Why?

(b) Which of the following types of system (i.e. Sequential, Concurrent, Parallel, and Distributed) would be most appropriate to implement a web server that may receive hundreds of simultaneous requests for web pages? Why?

(c) Which of the following types of system (i.e. Sequential, Concurrent, Parallel, and Distributed) would be most appropriate to implement an airline booking system that allows users to check fares offered by many different travel agents? Why?

(d) Define the terms latency and bandwidth using the definitions from the book.

(e) A sequential program takes 1 day to finish execution; when parallelized and run using 10 nodes, it takes 2 day to finish execution. Calculate the speedup achieved, and explain your answer.

Question 2 [5 marks]

What is the key difference between a SMP (symmetric multiprocessor architecture) and a NUMA (non-uniform memory access architecture)?

Question 3 [5 marks]

Consider a NUMA architecture with three processing elements (PEs).

- PE1 is composed of R1, P1 and M1 (storing memory locations 0-199),
- PE2 is composed of R2, P2 and M2 (storing memory locations 200 299) and
- PE3 is composed of R3, P3 and M3 (storing memory locations 300-399).

Suppose that P1 needs to access location 100 whereas P3 needs to access location 210. Which memory access request is likely to take less time? Briefly explain why.

Question 4 [5 marks]

Consider a NUMA architecture and one processing element make a request related to a memory location managed by another processing element, do you think that a read request would take longer or the same time as a write request? Justify your answer.

Question 5 [5 marks]

Consider the pipelined matrix multiplication example from “Lecture 2 – Programmer’s World View” slides. At what point in the pipeline can computation begin. Explain your answer.

Question 6 [15 marks]

Consider the matrix multiplication example from “Lecture 2 – Programmer’s World View” slides (same as Question 5.) In that example, we use three workers to compute the product of a 3x3 matrix by another 3x3 matrix. Imagine that we still had three workers but now wanted to compute the product of a 6x3 matrix by a 3x3 matrix.

Assume the following configuration (the coordinator is playing the manager role):

Coordinator -> Worker 0 -> Worker 1 -> Worker 2 -> Coordinator

This is a closed pipeline so the Coordinators above are the same entity.

We would say that Worker 1 is downstream of Worker 0 and upstream of Worker 2.

Data is assumed to always to flow downwards and cannot take shortcuts i.e. data flowing from Worker 0 to the Coordinator must always pass through Worker 1 and Worker 2.

Describe how would this change how the pipeline is filled, the computations done by each worker and how the pipeline is drained in order to extract the result of the computation? You should identify the actions taken by the coordinator and each worker during each of these phases.