1. (15%) Consider the distinguishing characteristics between CPU-bound and I/O-bound processes when answering the following questions.
   (a) (5%) Define and clearly distinguish between CPU-bound processes and I/O-bound processes. Consider the following applications:
   - Training a Deep Neural Network with millions of parameters using gradient descent optimization.
   - Training a K-Nearest Neighbors (KNN) classifier on a large dataset stored externally.

   Explain clearly which scenario corresponds more closely to a CPU-bound process and which scenario corresponds to an I/O-bound process, and justify your reasoning.
   (b) (10%) If a system frequently trains I/O-bound Machine Learning algorithms (such as Deep Learning Recommendation Model on external datasets), suggest **two optimization techniques** or system improvements to enhance performance. Explain clearly why each proposed method would help alleviate the I/O bottleneck.

2. (10%) Parallelism and concurrency in modern computing.
   (a) Define clearly the terms **parallelism** and **concurrency**. Explain how concurrency can exist without parallelism, providing one illustrative example scenario.
   (b) Explain clearly the difference between shared memory programming and message passing programming models. Provide one illustrative example scenario clearly differentiating these two models.

3. (10%) Understand how memory management affects system performance.
   (a) A system currently has a page-fault service time of 8 ms, with a memory access time of 100 ns. If the effective access time must not exceed 200 ns, calculate clearly the maximum allowable page-fault rate.
   (b) Discuss how increasing the cache block size affects cache hit rate and memory performance. Explain clearly the trade-offs involved and suggest a scenario where increasing the block size would improve performance significantly, and one scenario where it would degrade performance significantly.

4. (15%) File systems and peripheral management.
   (a) (5%) Define the terms inode, file descriptor, and directory. Explain briefly how the Unix file system uses these to manage files.
   (b) (10%) Consider a system that frequently uses short-lived, high-speed I/O operations (e.g., high-frequency sensors or high-speed NICs). Discuss the performance trade-offs between using polling vs. interrupts in this context. Is there a threshold (e.g., event rate or latency requirement) where polling becomes more advantageous?

5. (20%) Suppose that the following processes arrive for execution at the times indicated. Each process will run the listed amount of time. You may make some reasonable assumptions and write them down explicitly, if they are necessary to answer the following questions.
   (a) Please draw Gantt charts that illustrate the execution of these processes using the following scheduling algorithms: FCFS, non-preemptive SJF, and preemptive SJF.
   (b) Which of the algorithms in (a) results in the minimum average turnarround time (over all processes)? Be sure to justify your answer.

| Process | Arrival Time | Burst Time |
|---|---|---|
| P1 | 0 | 10 |
| P2 | 5 | 2 |
| P3 | 2 | 5 |
| P4 | 4 | 4 |

6. (10%) Explain the key drawback of spinlocks with busy-waiting compared to binary semaphores with sleeping.

7. (20%) Are the following statements true or false? For each statement, you will get 4 points for correct answer, zero point for blank, or -2 point for incorrect answer.
   (a) A deadlock-free solution is guaranteed to be starvation-free.
   (b) A context switch always involves updating the page table, whether it is between processes or threads
   (c) FF:FF:FF:FF:FF:FF is a broadcast address in Ethernet networks.
   (d) Buffer-overflow attacks can occur within communication protocols.
   (e) Named pipes are bidirectional.