

**Question 1: Parallel Hash Join.....[190 points]**

- (i) **[10 points] Parallel Join Algorithms:**  
Explain how to parallelize a nested-loop join.
- (ii) **[10 points] Parallel Join Algorithms:**  
Explain how to parallelize a hash join.
- (iii) **[10 points] Parallel Join Algorithms:**  
Distinguish between: hash join and sort-merge join. Which mechanism would you choose?
- (iv) **[10 points] Parallel Join Algorithms:**  
Distinguish between policy and mechanism.
- (v) **[10 points] Parallel Join Algorithms:**  
List the design goals for parallel join algorithms.
- (vi) **[10 points] Parallel Join Algorithms:**  
Define a TLB.
- (vii) **[10 points] Parallel Join Algorithms:**  
Distinguish between temporal and spatial locality.
- (viii) **[10 points] Parallel Join Algorithms:**  
List the phases of a hash join.
- (ix) **[10 points] Partition Phase:**  
Why is the partition phase optional?
- (x) **[10 points] Partition Phase:**  
What is the purpose of the partition phase?
- (xi) **[10 points] Partition Phase:**  
List two types of partitioning.
- (xii) **[10 points] Partition Phase:**  
List two types of non-blocking partitioning.
- (xiii) **[10 points] Partition Phase:**  
Why is it not necessary to synchronize in blocking partitioning?
- (xiv) **[10 points] Partition Phase:**  
Define these terms: (1) radix, and (2) prefix sum.
- (xv) **[10 points] Build Phase:**  
List the design decisions in a hash table.
- (xvi) **[10 points] Build Phase:**  
Distinguish between Robin Hood and Hopscotch hashing.
- (xvii) **[10 points] Probe Phase:**  
Explain the bloom filter optimization. Why is it known as sideways information passing?

(xviii) **[10 points] Evaluation:**

Explain why the radix partitioning hash join algorithm outperforms other algorithms on an uniform dataset.

(xix) **[10 points] Evaluation:**

Explain why the no partitioning hash join algorithm outperforms other algorithms on a skewed dataset.