1. [15 points] This question deals with System R query optimization.
   (a) [10 points] Why does the System R optimizer use the “only left deep trees” rule?
   (b) [5 points] What are two statistical assumptions that System R makes in doing selectivity estimation?
2. [15 points] In the DBMin algorithm, consider the case when query Q2 finds a page p in some other query Q1’s locality set. When Q2 accesses the page, DB Min does not update the statistics for this page in Q1’s locality set. Do you think this is a good idea? Justify your answer.
3. [15 points] Draw a picture of the operators and their connections in a parallel join of R and S on two processors. Assume that R and S are originally partitioned on attributes other than their join attributes.
4. [10 points] Give three options for distributed joins discussed in the R* optimizer validation paper.
5. [15 points] Suppose that instead of using the heuristic “accept the collection of bids that results in a point as far below the bid curve B(t) as possible”, we implemented a version of Mariposa with the heuristic “accept the collection of bids that results in a point as close to the bid curve B(t) as possible.” What effect if any would this have on the system?
6. [15 points] Consider the R-tree SplitNode() algorithm. What is the goal of this algorithm? (What is it trying to achieve with respect to the two nodes it produces?) What is the rationale for this goal?
7. [15 points] According to the Bucky Benchmark designers, what are the features that O/R DBMS add over RDBMS?