

Notes on *Query Evaluation Techniques for Large Databases*

COS 480/580: Database Management Systems

Sudarshan S. Chawathe
University of Maine

Fall 2005

The following review questions are based on Goetz Graefe's paper titled *Query Evaluation Techniques for Large Databases* (*ACM Computing Surveys*, 25(2):73–170, June 1993).

1. What is the *Halloween problem*? Illustrate it by using an example that is *not* discussed in the paper.
2. Outline the principal steps leading from a text query to an executable query plan. For each step, clearly indicate the input and output, and briefly describe its action.
3. Describe the main ideas of an iterator-based query-engine architecture.
4. Describe two alternatives to an iterator-based architecture.
5. What is the multiplicity of the mapping between logical and physical operators? Justify your answer using examples.
6. What is the difference between a left-deep plan and a bushy plan?
7. What is the difference between a left-deep plan and a right-deep plan (other than graphical representation)?
8. Describe two methods for creating level-0 runs. Comment on their performance characteristics. Justify your comments using examples.
9. Comment on the validity of the following statement: A hash join is preferable to a sort-merge join because the sorting-based must write the entire input to run files.
10. Provide an algorithm for generating optimized merge trees in the spirit of Figure 6 (page 88) and its accompanying discussion.
11. Provide a detailed example that illustrates bucket tuning and dynamic destaging during hashing. Your example must clearly indicate details such as hash buckets, memory buffers, and disk-resident data.
12. Describe hybrid hashing using a detailed illustrative example.
13. Compare hybrid hashing with other forms of dynamic hashing, such as extendible hashing and linear hashing. Highlight the most important similarities and differences.
14. What is a good block-size for an ext2-style filesystem built on a 300 GB disk that will be used mainly for storing typical mp3 files? Justify your answer.
15. Explain the following adjectives as they apply to indexes: clustering, nonclustering, dense, and sparse.
16. Provide a concrete example illustrating how a standard operating-system policy for buffer management, such as LRU replacement, may be particularly ill-suited for database workloads.
17. Provide pseudocode for a nested-loop aggregation operator (with grouping).
18. Write a standard SQL92 (SQL2) query that is equivalent to the query described in the first paragraph of Section 4.5 (page 103).
19. Given a table $R(a_1, a_2, \dots, a_k)$, how many non-equivalent queries will the following query template generate, where L may be replaced by any expression that results in a valid query?
`select sum(a_1) from R group by L ;`

20. Provide pseudocode for a pointer-based join based on merging.
21. Highlight the key similarities and differences between pointer-based hybrid-hash joins and value-based hybrid-hash joins.
22. How does the difference in access-times for L2 caches and main memory affect main-memory management in physical operators, such as a hybrid-hash join?