

CS W186 - Spring 2024

Exam Prep Section 5

Sorting/Hashing

1 Sorting and Hashing

Suppose the size of a page is 4KB, and the size of the memory buffer is 1 MB (1024 KB).

1. We have a relation of size 800 KB. How many page IOs are required to sort this relation?
2. We have a relation of size 5000 KB. How many page IOs are required to sort this relation?
3. What is the size of the largest relation that would need two passes to sort?
4. What is the size of the largest relation we can possibly hash in two passes (i.e. with just one partitioning phase)?
5. Now suppose we were executing a GROUP BY on age instead. Would sorting or hashing be better here, and why?

2 External Sorting

Assume our buffer pool has 8 frames. In this question, we'll externally sort a 500 page file.

1. How many passes will it take to sort this file?
2. Given the number of passes you calculated in 1.1, how many I/Os are necessary to externally sort the file?
3. What is the minimum number of additional frames needed to reduce the number of passes found in 1.1 by 1?
4. What is the minimum number of additional frames needed to sort the file in one pass?

3 Hashing

1. Suppose the size of each page is 4KB, and the size of our memory buffer is 64KB. What would be the I/O cost of hashing a file of 128 pages, assuming that the first hash function creates 2 partitions of 32 pages, and all other partitions are uniformly partitioned?