COMP9120 Relational Database Systems

Tutorial Week 11: Query Processing

Introduction

Suppose we have a schema Rel1(A, B, C) and Rel2(C, D). Each A field occupies 4 bytes, each B field occupies 12 bytes, each C field occupies 8 bytes, each D field occupies 8 bytes. Rel1 contains 100,000 records, and Rel2 contains 50,000 records. There are 100 different values for A represented in the database, 1000 different values for B, 50,000 different values for C, and 10,000 different values for D. Rel1 is stored with a primary (sparse, clustered) B+ tree index on the pair of attributes (A,B); assume this index has 2 levels, including the root page and excluding the leaf (record) pages. Rel2 is stored with a primary (sparse, clustered) B+ tree index on C; assume this index has 2 levels, including the root page and excluding the leaf (record) pages.

General features: assume that each page is 4K bytes, of which 250 bytes are taken for header and array of record pointers. Assume that no record is ever split across several pages. Assume that index entries in any index use the format of (search key, rowid), where rowid uses 4 bytes.

Exercise 1. Block-Nested Loops Join

Consider the following query:

```
SELECT Rel1.A, Rel1.B, Rel1.C
FROM Rel1, Rel2
WHERE Rel1.C = Rel2.C AND Rel2.D = 16;
```

and consider the query plan which calculates this as follows:

Form the equi-join of Rel1 and Rel2 by using a block-nested loops join with Rel1 as the outer relation, and then filter each tuple of the join to see if the value of D is 16; if so, output the values of A, B and C from that tuple of the join.

How many page I/Os are needed to compute this plan (assume that we have only the minimal space, say 2 pages worth, for buffering in memory)?

Exercise 2. Index-Nested Loops Join

Reconsider the query in the above Exercise. But, now consider a different query plan, where the join is processed as an index-nested loops join (using the primary index on Rel2.C), and then each tuple of the join is filtered to check the value of D. How does this affect the cost?