

COMP9120 Relational Database Systems

Semester 2, 2016

Tutorial Week 10: Storage and Indexing

This is mainly a paper-based exercise, similar to what you might encounter in an exam. You are welcome to work in groups, but make sure you understand what your team is doing.

Exercise 1. Calculating space and time

Suppose we have a table Rel1(A, B, C). Each A field occupies 4 bytes, each B field occupies 12 bytes, each C field occupies 8 bytes. Rel1 contains 100,000 records. There are 100 different values for A represented in the database, 1000 different values for B, and 50,000 different values for C. Rel1 is stored with a primary, sparse, clustered index on the composite key consisting of the pair of columns (A,B); assume this index has 2 levels. Assume that in this database, each block is 4K bytes, of which 250 bytes are taken for header information. Record locations within the index use a 4-byte *rowid*. Assume that reading a disk block into memory takes 150 msec, and that the time needed for any query can be approximated by the time spent doing disk I/O.

1a) Calculate the space needed for the data of Rel1.

Hint: How much space is used by a single row? Multiply by the number of rows in the table to get a quick estimate of space. [A more accurate answer can be done using the fact that rows are not split across blocks, and each block has a header; so we could work out how many rows per block, then how many blocks for the relation, and then convert this back to space.]

1b) Calculate the time taken to perform a table scan through the relation Rel1.

Hint: How many blocks are needed to make up the space occupied by the relation? How many blocks will be read from disk during a scan?

1c) Calculate the time taken, using the primary index, to execute the following query

```
SELECT C
FROM Rel1
WHERE A = 'AQG' and (B between 'WPQ' and 'XYZ');
```

1d) Suppose that we often need to process a query

```
SELECT A, B
FROM Rel1
WHERE (B between 'WPQ' and 'XYZ') and C = 'UBMJ';
```

Why can't the primary index be useful in processing this? What index should be created to speed this up? How long will the query take using the extra index, assuming that the extra index has 2 levels?

Exercise 2. Index creation with Oracle

Find which indices already exist for the tables you defined in early tutorials and assignments. Identify an extra index, which would be useful for one of the queries you wrote (supposing that the database became much bigger, with many more rows in each table!). Create this index.

Note: To check which indexes are present in your schema (plus some details), you can use the following SQL commands in Oracle:

```
SELECT * FROM USER_INDEXES;  
SELECT * FROM USER_IND_COLUMNS;
```

Exercise 3. Assignment 3

Your team should have started looking into assignment 3 now. Use any remaining time in the tutorial to discuss plans for completing the assignment. You should also let your tutor know ASAP if you wish to change groups.