Prof. Dr. Stefan Brass
Institut für Informatik
MLU Halle-Wittenberg

January 23, 2020

Databases IIB: DBMS-Implementation
— Exercise Sheet 11 —

Please read Part a) think about the answers, and mark questions which you want to discuss in class. You only have to submit Part c) to f). Please upload your solution into the StudIP file folder called "Hausaufgabe_11" in the StudIP entry of the lecture. The deadline is January 28 (the day before the next lecture).

Please also think about the implementation project. The implementation project will count minimally like three homeworks. Besides the program code, you will have to submit a short written description how your program works and how your data file is structured.

Repetition Questions

a) What would you answer to the following question in an oral exam?

- If you program your own DBMS, what would be the structure of ROWIDs, and what would be their size in bytes?
- How large would B$^+$-tree nodes be in your system?
- If you program a B$^+$-tree, how would the internal representation of the branch nodes look like? Please sketch the data structure that you use for a node. For simplicity, you can assume that the search keys are `int`-values (with a fixed length of 4 Bytes) and the index is a unique index (only one ROWID per search key value). Choose a reasonable number of bytes per ROWID (see the first question).
- Why is it good to link leaf nodes only in one direction, not as a doubly linked list? (Think about the splitting of nodes.)
- Please sketch the data structure for leaf nodes.
- The standard storage structure for a table is a heap file. Name at least one other data structure that can be used in Oracle to store table rows.
- What is an index-organized table? Name advantages and disadvantages compared with a heap file and a normal index. What is the problem if one wants to create additional indexes for an index-organized table? Explain a solution to this problem.
- What is a bitmap index? In which situations should one consider bitmap indexes?
- What is an advantage of a partitioned table?

In-Class Exercises

b) We will discuss the exam from 2015/16, which you find here:

[http://www.informatik.uni-halle.de/~brass/dbi17/exam15.pdf]

Homework Exercises

Consider a database with the following tables:

- `EMP(`<u>`EMPNO`</u>`, ENAME, SAL, DEPTNO→DEPT, MGR°→EMP)`
- `DEPT(`<u>`DEPTNO`</u>`, DNAME, LOC)`

This is a simplified version of the well-known example database from Oracle. Suppose that the employee table `EMP` contains 10000 rows, the department table `DEPT` contains 25 rows, and each department has approximately the same number of employees. A manager (`MGR`) supervises on average 10 subordinates. Furthermore assume that 10 rows are stored per block in both tables. The following query is given:

```
SELECT   EMPNO, SAL, DNAME
FROM     EMP E, DEPT D
WHERE    E.DEPTNO = D.DEPTNO
AND      E.MGR = 7839
ORDER BY SAL
```

Describe how the query can be evaluated if one has the following indexes:

c) None.

d) An index `I1` on `EMP(EMPNO)` and an index `I2` on `DEPT(DEPTNO)` (both are `UNIQUE`).

e) Like d), but in addition a third index `I3` on `EMP(MGR)`.

f) If you could select an index (including an index on a combination of attributes), which would you choose? The goal is that the above query is executed as fast as possible.

You do not have to draw query evaluation plans yet. Just describe in words how the query can be evaluated. Please submit your solution as either plain text or PDF (not in Microsoft Word format).