

Databases II

Exercise sheet 2 – WiSe 2019/20

4. Create and draw a 2-level secondary index on the Dept. attribute for the following table. This means that the index has 2 index levels, where the first level should be dense and the second level sparse. Per index block 10 entries can be stored. (Since it is a secondary index, an index entry points to a tuple).

Name	Dept	CourseNo
Donetta	CS	457
Annabelle	Arch.	45
Roosevelt	CS	27
Allyson	Socio.	470
Debra	Psych.	457
Bobbie	Psych.	27
Bradly	CS	11
Marcell	CS	470
Amanda	Psych.	470
Danelle	Socio.	470
Michael	Arch.	125
Ann	Path.	350
Tom	Gen.	291
Camilla	Pol.	11
Abdul	CS	27
Carrie	CS	27
Ewa	Psych.	125
Conrad	CS	350
Lucille	Socio.	350
Roberto	Arch.	125

5. A block can store 500 index entries or 80 records of relation R . R contains 10.000.000 records.
- How many blocks are needed for a (flat) *dense* index on R ?
 - How many blocks are required for a (flat) *sparse* index on R which contains one entry per block of data file?
6. Given a relation $R[A, B, \dots]$ with the following properties:
- $|R| = 1.000.000$ tuples,
 - the values of attribute A are equally distributed in interval $[1, 100.000.000]$,
 - the values of attribute B are equally distributed in interval $[1, 1.000]$,
 - there is a sparse primary index on A (i.e., each index entry points to a data block of R),
 - there is a dense secondary index on B (i.e., each index entry points to a tuple of R),
 - one block stores 500 index entries or 50 records.

The following queries are asked on R :

$$\text{Q1: } \sigma_{A > 60.000.000}(R), \quad \text{Q2: } \sigma_{B > 600}(R)$$

How many blocks must be read if the indices are not used to answer Q1 or Q2.

7. Assuming the same specification as in exercise 6, how many blocks must be read if the indices are used to answer Q1 or Q2?
8. Given a relation $R[A, B, C, E]$ with the following properties:
- $|R| = 10.000.000$ tuples,
 - the values of attribute A are equally distributed in interval $[1, 100.000.000]$ and unique,
 - the values of attribute B are equally distributed in interval $[1, 5.000]$,
 - the values of attribute C are equally distributed in interval $[1, 25.000]$,
 - the values of attribute E are equally distributed in interval $[1, 500.000]$,
 - there is a flat sparse index on attribute A ,
 - there is a flat dense index on each of attributes B, C and E ,
 - a data block can store 25 entries,
 - an index block can store 100 entries.

The following queries are asked:

$$\text{Q1: } \sigma_{C=20.000 \wedge A < 50.000.001}(R) \quad \text{Q2: } \sigma_{C=20.000 \vee A < 50.000.001}(R)$$

Describe the steps required to answer Q1 or Q2 as efficiently as possible and specify the required number of block accesses for the described steps.