

Write the name of the course, the date of exam, your name), your date of birth and your signature in each separate answer paper.

Give the answers for tasks 1, 2, and 3 on separate papers.

1. Consider a disk unit that has 20 surfaces and 1000 cylinders. The average capacity of a track is 50 blocks, each 4KB. The average seek time is 4 ms. The disk makes 12000 rotations in a minute, i.e. one rotation takes 5 ms. Table T has 200 000 rows. It is implemented as a hash file without a bucket index. The key of the table is used as the hash key. The average length of a record is 400 bytes. Buckets correspond to blocks and the size of the original hash area is 40000 buckets. The hash function distributes the keys quite evenly. 10000 pages have been allocated for overflows but only 200 are in use.
 - a) What are the factors of the access time when a record is searched by the key? Give your estimate of the access time.
 - b) How should the file be accessed, if some other search criterion than the key is used. What is the total access time in this case? (6p)

2. Consider the tables R(A, B,...) and S(X, ..., Z)à R). (Z is a foreign key referring to R). A row of table R is connected to about 20 rows of S, in average. The size of table R is about 500 rows (the file has 200 data pages of size 4KB). Table S has 10000 rows (the file has 5000 data pages). The average length of values in column A is 30 bytes. Page size is 4KB and the length of address is 10 bytes. Both tables are implemented as heaps. 150 pages of buffer memory may be used in processing queries. Consider the query
select * from R,S where R.A=S.Z and S.Z=x;
 - a) Estimate the size of its result
 - b) How would you implement the query, if no indexes were available (describe the principle)? How many disk accesses are needed?
 - c) How do the indexes based on the keys and the foreign keys of these tables affect the query processing? What is the most effective plan and how many disk accesses are needed in that case? (10p)

3.
 - a) What is a checkpoint?
 - b) Describe the contents of a database log.
 - c) What happens to the log in commit (9p)

-
4. **Home task.** *You should answer this task only if you want to substitute your credits obtained by doing exercises.* Send the answer by e-mail to Harri.Laine@cs.helsinki.fi by **Friday 23rd of May at 18.00**. The answer may be embedded in the body of the mail. It may be sent as an attachment, too. You may also send only a link to the answer. The following file formats are accepted: plain ASCII text, MSWord's doc or rtf format, Postscript, PDF or HTML.

Find out and explain how database management systems store big (clob, blob) and variable length (varchar) column values. Include in your report at least three well known database management systems. The expected length of the report is about 3 pages. (5p)