

Mid-Quarter Examination

DUE in 60 minutes.

FOR COURSE:CS 5703

Quarter:Spring 1995

INSTRUCTIONS

1. It is an open book exam. You may use the textbook and your own notes. Do not use use any other book or material.
2. There are 3questions. Question 1 has 2 parts. Question 2 has 3 parts. Question 3 has 2 parts.
3. Please use the white space below the question to answer the question. If additional space is needed, use the back side of pages.
4. Budget proper time for each question. First answer the questions you know best.

| |
|-------------------------------------|
| NAME |
| UMN Identification Number |
| Account |
| Machine |
| Classification: (NTU UNITE Regular) |

| Question | Points |
|----------|--------|
| 1a. | |
| 1b. | |
| 2a. | |
| 2b. | |
| 2c. | |
| 3a. | |
| 3b. | |

QUESTION 1 Consider a database with the following relations: Emp(ssn, fname, minit, lname, bdate, address, sex, salary, superssn, dno), Project(pnumber, pname plocation, dnum) and Works_On(essn, pno, hours). The primary keys to these

relations are ssn, pnumber and <essn, pno> respectively. Consider the following SQL query:

```
SELECT  fname, lname, 1.1*salary
FROM    Emp, Works_On, Project
WHERE   ssn = essn AND pno = pnumber AND pname = 'productX'
```

Q 1a Draw the canonical syntax-tree for the query.

Q 1b Show how the query tree can be optimized via heuristic algorithms based on relational algebra rules. Show the intermediate trees. Identify the rules used for tree transformations. Briefly identify your assumptions.

QUESTION 2. A PARTS file has Part# as a key field. The records in PARTS file have the following values for keys: 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.

Q 2a. Draw a B+tree of order $p = 3$ containing the key values.

Q 2b. Draw a linear hashing file containing these key values. Assume bucket capacity = 2. Assume the family of hash functions is $h_0 = x \bmod 1$, $h_1 = x \bmod 2$, $h_2 = x \bmod 4$, ... Test your answer by applying the search procedure for several key-values.

Q 2c. How does a B-tree differ from a B+tree? Why is B+tree usually preferred as an access structure to a datafile?

QUESTION 3. Consider the cost-models for strategies for computing $\text{join}(R, S, R.A = S.B)$. Assume R has $b(R)$ block, $r(R)$ records and blocking factor of $\text{bfr}(R)$. S has $b(S)$ blocks, $r(S)$ records and blocking factor of $\text{bfr}(S)$. $R.A$ is unique. $S.B$ is unique. Assume minimal buffer space.

Q 3a. Assume R has a primary index on $R.A$ with depth $x(R.A)$ and S has a primary index on $S.B$ with depth $x(S.B)$. Derive an algebraic condition to characterize the dominance zone of nested-loop with indexed search over sort-merge. Identify your assumptions and justify your answer.

Q 3b. Assume R is hashed on $R.A$ and S is hashed on $S.B$. Derive an algebraic condition to characterize the dominance zone of nested-loop with indexed search over hybrid hash strategy. Identify your assumptions and justify your answer. QPSpace