

Qualifying Exam: Database

Closed books, 60 min., 2012.1

1. Explain the following: (30%)

- 1) foreign key
- 2) data independence
- 3) IDREFS in XML
- 4) What are the four major concepts of the dataflow model?
- 5) Why is the normalization necessary (What are the problems of the unnormalized table)?
- 6) completeness of the D(data)-schema with respect to the F(functional)-schema

2. Construct a B+ -tree for the following set of key values:

(2, 3, 5, 7, 11, 17, 19, 23, 29, 31)

Assume that the tree is initially empty and values are added in the ascending order. Construct the B+ -tree for the case that the number of pointers that will fit in one node is four. Show the tree after each insertion. (20 %)

3. Assume two base relations $r(A,B)$ and $s(A,C)$ with $|r| = 15,000,000$ and $|s| = 800,000$. The block (page) size is 2,000 bytes, and the tuple size is 400 bytes for both relations. The values of the integer attribute A are uniformly distributed in relation r and $|r.A| = 500,000$, where $r.A$ is the set of distinct values of A in r. There is a B+ - tree index on attribute A in relation r, where each node contains up to 200 index entries.

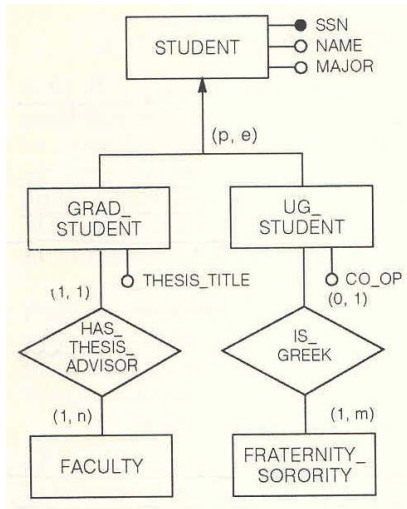
1) For $x \in r.A$, determine the number of blocks to read for $\sigma_{A=x}(r)$ if the index is used.

- a) the index is clustered. (10%)
- b) the index is unclustered. (8%)

2) Assume that A is the key of s and a foreign key of r, estimate the number of tuples for $r \bowtie s$ and explain why. (8%)

4. Consider the following two schemas. Remove the generalization hierarchy. (24%)

1) Using the superset entity (p: partial, e: exclusive)



2) Using the subset entities (t: total)

