Prof. Li-Yan Yuan

CMPUT 391: Database Management Systems

Assignment 1

18:00, Feb. 10, 2014, at the 391 Drop Box

1. Present a real-life example (Not using ABCD, etc) to show differences between BCNF and 4NF. [5]

2. Consider a database consisting of the following tables with obvious meanings:

```
employee( employee_name, street, city )
works( employee_name, company_name, salary )
company( company_name, city )
```

   Write triggers, according to Oracle style, to enforce the constraint that no employee is allowed to work for a company NOT located in the city where the employee lives. One needs to (1) give a list of events that must be monitored for imposing the constraint, and (2) present one trigger for any event in (1). [20]

3. Consider a relation schema $R = ABCDE$, functional dependencies

$$A \rightarrow C$$
$$B \rightarrow C$$
$$C \rightarrow D$$
$$DE \rightarrow A$$

   and a decomposition $D = \{AC, AD, BE, AB\}$ of R.

   Prove or disprove that D is a join lossless decomposition of R with respect the given set of FDs. A proof can be done by finding an appropriate decomposition tree, and a counter-example is good enough for a disproof. [10]

4. Consider $R = ABCDEGHK$ and the following set F of functional dependencies:

$$E \rightarrow DK$$
$$C \rightarrow K$$
$$EK \rightarrow BC$$
$$HK \rightarrow A$$

   (a) Find a join loss-less, dependency preserving and 3NF decomposition of R.

   (b) Indicate whether your database schema is in BCNF with respect to F. Explain. [20]

5. Consider the following two algorithms for computing $R \bowtie S$:

   (a) index,

   (b) merge-sort join.

   Assume that (1) both tables are stored in B-tree, (2) the main memory can accommodate 502 pages, (3) R is stored in 20,000 pages with 20 tuples per page, and S is stored in 50,000 pages with 5 tuples per page, (4) there exists a non-clustered index on the join attribute of both R and S; and (5) each index search takes **1.6** disk accesses on average. Note that since the index on the join attribute of each table is non-clustered, neither R nor S is sorted on the join attribute.

   For each algorithm, estimate the number of page accesses, excluding outputting the joined result table, with an optimized configuration. By an optimized configuration we mean the arrangement of all the relevant factors such that the estimated number of disk accesses is less than that of other arrangements using the same algorithm. Note that at most 8-ways merge-sort is allowed. (That is, it is not allowed to merge more than 8 sub-sorted files in one merge operation.) [20]