CS5614 Final Exam

December 10, 1999

Please enter the following information:

- **Name:**

- **ID:**

GOOD LUCK and Happy Y2K!
Do not write below this line

Problem	Max Score	Score
1	30	
2	10	
3	20	
4	10	
5	20	
6	10	
7 (XC)	10	
Total	100	

1. (30 points) Short answer questions:

    (a) Identify the most common UNDO and REDO policies used by RDBMS vendors.

    (b) If R is declared to be a many-one relationship, under what conditions will R be symmetric? *Note:* A symmetric relationship is one which is its own inverse.

    (c) Inductive logic programming (ILP) is a viable data mining technique for obtaining what kinds of patterns?

    (d) How many `GROUP BY`s need to be `UNION`ed to realize a `CUBE` of n variables? Assume that each variable has a cardinality of 3.

    (e) Why can't a language like PROLOG (which supports knowledge based reasoning and AI techniques) be used for data mining?

    (f) Consider the relation `Car(maker,model,year,fuel)`. Write a query in Datalog to find those `Cars` that are not both made in 1999 and manufactured by *GM*.

2. • (5 points) Consider a relation $R(A_1, A_2, A_3, A_4, ..., A_n)$. What is the maximum number of keys R can have at the same time?

• (5 points) When we go from 3NF to BCNF, notice that we might be splitting a key across two decomposed relations. Why is this such a good idea? If we split the key attributes across two relations, isn't that a cause for worry? Why then, is BCNF such a good normal form?

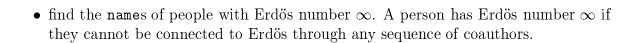3. (20 points) Consider the relational schema:

`Coauthor(name1,name2)`

which holds true if the person with `name1` has co-authored a paper with the person whose name is `name2` (you may assume that names are unique). A given paper's coauthors will only be listed once in the `Coauthor` relation (i.e., `X` and `Y` will either be recorded as `Coauthor(X,Y)` or `Coauthor(Y,X)` but not both).

The Erdös number project seeks to chronicle the professional network of the proligic Hungarian mathematician, Paul Erdös who has written hundreds of research papers. His Erdös number is 0. His immediate co-authors have Erdös number 1. People other than Erdös who have written a joint paper with someone with Erdös number 1 but not with Erdös, have Erdös number 2 and so on. Write queries in Datalog to:

- find the `name`s of people with Erdös number 2.

- find the `name`s of people with Erdös number > 2. You are welcome to reuse the results of the previous subpart, if necessary.

- find the `name`s of people with Erdös number ∞. A person has Erdös number ∞ if they cannot be connected to Erdös through any sequence of coauthors.

- find the `name`s of people with the highest Erdös number $< \infty$. This number is not known beforehand.

4. (10 points) Using the ideas learnt in class, explain how you would estimate the size of the relation $R(A_1, A_2, \cdots, A_n)$ after duplicates have been eliminated.

5. (20 points) Consider the following two transactions:

- **Transaction 1:**
  - read(A,x)
  - read(B,y)
  - if (x==0) then y=y+1
  - write(B,y)
- **Transaction 2:**
  - read(B,y)
  - read(A,x)
  - if (y==0) then x=x+1
  - write(A,x)

For the following questions, take into account all operations, even the `if`s. The consistency constraint is that `A==0 OR B==0`. The initial values of `A` and `B` are both zero. You may assume that the `x` and `y` in the two transactions are different. Calculate the number of :

- Interleavings of the above two transactions (all combinations)
- Serial Schedules
- Serializable Schedules
- Conflict-Serializable Schedules

6. • (5 points) Consider the compatibility matrix for S, X, IS, IX and SIX locks used for managing a hierarchy of database elements. Complete the following sentences in the most exhaustive way possible:

    – To obtain an IS lock on an element, a _____ lock must first be obtained on all its _____.

    – To obtain an S lock on an element, a _____ lock must first be obtained on all its _____.

    – To obtain a IX lock on an element, a _____ lock must first be obtained on all its _____.

    – To obtain an X lock on an element, a _____ lock must first be obtained on all its _____.

    – To obtain a SIX lock on an element, a _____ lock must first be obtained on all its _____.

    • (5 points) Using your answers to the above, explain how the standard two-phase locking mechanism can be extended to work with a hierarchy of database elements. (The traditional rule is that no new locks can be obtained after the first unlock; How can this be applied to a hierarchy of elements?)

7. (Extra Credit: 10 points) The `EXISTS R` operator is a boolean predicate that is true if and only if the relation `R` is not empty (`R` could be a subquery). It can be used in an SQL `SELECT` query in the `WHERE` clause. Some RDBMS vendors do not provide this facility, since it can be replaced by other operators. Explain how, *without* using the `COUNT` aggregation operator.