Name: 

SSN: 

Closed Book
Closed Notes
No Calculators
13 Questions

Instructor: David W. Embley
1. Let $q$ have scheme $ACD$, $r$ have scheme $AB$, $s$ have scheme $BCD$, and $t$ have scheme $BC$.

(a) (7 points) Use the rules of query optimization to rewrite the following query in an optimal form. (Push $\pi$ and $\sigma$ over $\bowtie$ and $\cup$ as far as possible.)

$$\pi_{AC} \sigma_{D=1}(\pi (r \bowtie s) \cup (q \bowtie t))$$

(b) Assume that on the average a seek (head move) takes 8 milliseconds, rotational latency (spin under) takes 4 milliseconds, a block transfer takes 68 microseconds. Assume that there are 64,000 records in $r$ and 128,000 records in $q$. A block has 512 bytes. All attributes are 4-byte integers. The key for $q$ is $A$, for $r$ is $AB$, for $s$ is $B$, and for $t$ is $BC$. (A formula with numbers plugged in is sufficient; you need not do the arithmetic.)

i. (3 points) Estimate the access time for $\sigma_{B=1} r$ if $r$ is stored contiguously as a sequential file. (A formula with numbers plugged in is sufficient; you need not do the arithmetic.)

ii. (3 points) Estimate the access time for $\sigma_{B=1} r$ if $r$ is a sequential file stored in random blocks scattered on the disk. (A formula with numbers plugged in is sufficient; you need not do the arithmetic.)

iii. (6 points) Estimate the access time for $\sigma_{A=1} q$ if $q$ has a B$^+$-tree index for attribute $A$ with only its root node in memory and all other nodes on disk. Assume that a B$^+$-tree pointer occupies 4 bytes and that the size of a B$^+$-tree node is the size of a block. Give the minimum and maximum estimate depending on the height of the B$^+$-tree. (A formula with numbers plugged in is sufficient; you need not do the arithmetic.)
2. (17 points) Create an OSM diagram for the following.

Students have an ID# and a single Name and a single Address; an ID# uniquely identifies a Student as does the Student’s Name and Address together. Undergraduate Students also have their Best ACT Score, if any. Graduate Students also have their Best GRE Score, if any. Undergraduate and Graduate Students are mutually exclusive and together constitute all the Students of interest. Faculty Members are identified by their SSN and have a single Name and a single Office#. There is only one Faculty Member in each office. Graduate Students have a single Faculty Advisor, and Faculty Advisors may advise many Graduate Students. Undergraduate Students do not have Faculty Advisors. Some of the Students are TA’s. Each TA has a Salary. The Salary of every Graduate Student exceeds the Salary of any Undergraduate Student.

In your diagram, use the singular of nouns or noun phrases that start with capital letters as object-set names. Be sure to name all relationship sets and give proper participation and co-occurrence constraints. Make appropriate use of generalization/specialization. You must also give a formal predicate-calculus statement with quantifiers to capture the salary constraint.
3. (13 points) Convert the following ORM diagram into an ORM hypergraph. Make your ORM hypergraph correspond to the results that would be obtained by the hypergraph generation algorithm (Algorithm 9.1).
4. (15 points) Reduce the following ORM hypergraph to a canonical hypergraph. Assume that semantic equivalence holds where you need it. You may do your reduction by clearly marking components to be deleted, by redirecting edges, by adding roles to make connections congruent, and by circling equivalence classes that should be lexicalized.
5. Assume that the following ORM hypergraph is canonical.

(a) (8 points) Give the set of schemes produced by the scheme synthesis algorithm (Algorithm 10.1). Also, add scheme names \( R_1, R_2, \ldots \) for the schemes.

(b) (4 points) Underline the keys in your schemes above.

(c) (6 points) Give the inclusion dependencies as produced by Algorithm 10.2. Use \( r_1, r_2, \ldots \) to designate the relations associated respectively with \( R_1, R_2, \ldots \).
6. Consider the following generic relational database scheme.

\[
\begin{aligned}
\text{Person}(\text{Name, Address, DLNumber, PhoneNr}) \\
\text{Primary Key: } \text{DLNumber} \\
\text{Key: Name, Address} \\
\text{Vehicle}(\text{License Plate Nr, Price, Color}) \\
\text{Primary Key: } \text{License Plate Nr} \\
\text{Owns}(\text{DLNumber, License Plate Nr}) \\
\text{Primary Key: } \text{DLNumber, License Plate Nr} \\
\text{Owns}[\text{DLNumber}] \subseteq \text{Person}[\text{DLNumber}] \\
\text{Owns}[\text{License Plate Nr}] \subseteq \text{Vehicle}[\text{License Plate Nr}] \\
\end{aligned}
\]

(a) (7 points) Give the SQL create statement for the Owns relation scheme. Declare appropriate types for each attribute.

(b) (6 points) Write an SQL query for this scheme to list the \text{Name} and \text{Address} of car owners who own red vehicles costing more than $10,000.
7. For each pair of sets of FDs $F$ and $G$ below determine whether $F \equiv G$. If so, show how to derive each FD of $G$ not already in $F$ from $F$ and each FD of $F$ not already in $G$ from $G$. (You can use the closure of a set of attributes for this if you wish.) If $F$ is not equivalent to $G$, give all FDs of $G$ that cannot be derived from $F$ (if any) and all FDs of $F$ that cannot be derived from $G$ (if any).

(a) (6 points) $F = \{A \rightarrow B, CD \rightarrow A, C \rightarrow D\}$ and $G = \{C \rightarrow B, A \rightarrow B, C \rightarrow AD\}$

(b) (6 points) $F = \{A \rightarrow B, AB \rightarrow C, B \rightarrow CD, A \rightarrow C\}$ and $G = \{A \rightarrow BC, B \rightarrow D, C \rightarrow A\}$
8. Let $F = \{ A \rightarrow B, B \rightarrow AC, AC \rightarrow E, BD \rightarrow C, E \rightarrow I, G \rightarrow EI, GE \rightarrow H \}$.

(a) (3 points) Produce the ORM hypergraph corresponding to $F$. Each FD should correspond to one edge.

(b) (10 points) Give the ORM hypergraph after making head and tail reductions.
9. Let $U = ABCDE$ and let $F = \{AB \rightarrow C, C \rightarrow E, E \rightarrow BD, B \rightarrow D\}$.

(a) (5 points) List all candidate keys for $ABCDE$.

(b) (10 points) Decompose $ABCDE$ into BCNF. (Hint: consider using a hypergraph.)

(c) (2 points) Is it possible to have a database scheme for $U$ that is both in BCNF and dependency preserving?
10. Let $R = ABCDE$ and let $F = \{ABC \rightarrow D, D \rightarrow AB, A \rightarrow E\}$ be a set of FDs over $R$.

(a) (2 points) Give the hypergraph for $F$. Each FD should correspond to one edge.

(b) (3 points) Identify all properly embedded FDs.

(c) (7 points) Reduce the hypergraph so that it is canonical. Be sure that your reduction preserves information and constraints. (Assume semantic equivalence.)

(d) (10 points) Use scheme synthesis (Algorithm 10.1) to produce schemes. Underline the keys in each scheme.

(e) (3 points) Which schemes must the implementation join together to check the FD $ABC \rightarrow D$?
11. Consider the following canonical hypergraph.

(a) (2 points) Which object set is included in the most functional closures (of other object sets)?

(b) (7 points) Beginning with $D$, apply the nested scheme generation algorithm (Algorithm 10.3) to the hypergraph. For any succeeding tree, select the root to be the (or any one of the) not-already-marked object sets included in the most functional closures.

(c) (7 points) Beginning with $C$, apply the nested scheme generation algorithm (Algorithm 10.3 in the text) to the hypergraph. For any succeeding tree, select the root to be the (or any one of the) not-already-marked object sets included in the most functional closures.
12. (10 points) Based on the Data Extraction Presentation, do the following: (1) In a single sentence, state the problem to be solved. (2) In at most three sentences, describe the proposed solution.

13. (12 points) Based on the Family History Presentations, do the following: (1) In a single sentence, name one of the database problems that needs a resolution to improve technology for supporting family history. (2) In a single sentence, state what benefit would be gained by solving the problem you named. (3) In at most three sentences, propose a specific solution to the problem you named. (4) In a single sentence, say what would need to be done to convince product developers that your proposed solution is viable and should be adopted into a suite of products to support family-history technology.