1. Circle “true” or “false” for the following statements (10 points).

true     false  A DML provides a description of the metadata for a database.
true     false  A DDL provides for data retrieval and data update.
true     false  The three levels of a database architecture are indexed-sequential files, hashing, and B+-trees.
true     false  The technical concept “Implementation Independence” (also called “Data Independence”) has essentially the same meaning as “information hiding” for an ADT.
true     false  Relational algebra is a pair, (Set of Relations, Set of Relational Operators).
true     false  The set of relational-algebra operators \{∪, ρ, □\} is not closed.
true     false  For the integers, the set of arithmetic operators \{+ , −, ×, ÷\} is closed.
true     false  The integrity constraints of a database cause the database to be persistent.
true     false  If a company has separate views of a database for personnel and payroll, the database is not integrated.
true     false  Referential integrity is implemented using foreign keys.
(16 points) Produce appropriate relation schemes for this diagram. (“Appropriate” means to follow the procedure for generating schemes discussed in the book and in class.) Underline all candidate keys (be sure to make your underline continuous for composite candidate keys). Add a second underline for all primary keys, so that all primary keys are double underlined. Draw an arrow from each foreign key to the primary key it references.
3. Let \( r \) and \( s \) be relations as follows:

\[
\begin{array}{ccc}
\text{r} & \text{A} & \text{B} & \text{C} \\
1 & 2 & 3 \\
1 & 3 & 0 \\
4 & 7 & 0 \\
4 & 7 & 3 \\
\end{array}
\quad
\begin{array}{ccc}
\text{s} & \text{B} & \text{C} & \text{D} \\
2 & 3 & 4 \\
7 & 0 & 5 \\
7 & 3 & 6 \\
9 & 0 & 7 \\
0 & 5 & 8 \\
\end{array}
\]

Compute the relation for each of the following:

(a) (3 points) \( \pi_A r \)

(b) (3 points) \( \pi_B s - \pi_B r \)

(c) (3 points) \( \pi_{BC} r \cap \pi_{BC} s \)

(d) (3 points) \( \pi_{AD}(r \bowtie s) \)

(e) (3 points) \( (\pi_C s \bowtie (\pi_E \rho_{C \rightarrow E} r)) \)
4. Consider the following database scheme.

\[
yuppie(SSN, Name, City, State, Salary) \quad \text{key: } SSN \\
car(VIN, Make, Model, Year, Cost, Color) \quad \text{key: } VIN \\
owns(SSN, VIN) \quad \text{key: } \{SSN, VIN\}
\]

(a) (5 points) Write a relational-algebra expression to find the names of yuppies who own red cars and whose salary is less than the cost of the car.

(b) (5 points) Write an SQL query to find the VIN of cars that cost more than $50,000 but are not red.

(c) (5 points) Express the following SQL query in English.

\[
\text{select y1.SSN, y1.Name, y2.SSN, y2.Name} \\
\text{from yuppie y1, yuppie y2} \\
\text{where y1.City = y2.City and y1.SSN < y2.SSN}
\]
5. Consider relations \( r(ABC) \) and \( s(CDE) \) and suppose that \( r \) has 2,000 tuples stored 20 per block and that \( s \) has 6,000 tuples stored 30 per block. Assume that \( A \) is the primary key for \( r \) and that \( C \) is the primary key for \( s \). Assume that the average seek time is 8 milliseconds, the average rotational latency is 4 milliseconds, and the average block-transfer time is 68 microseconds.

(a) (4 points) Estimate the time required to execute \( \sigma_{C=1} r \) assuming that \( r \) is stored as a sequential file that is not contiguous and not sorted on \( C \).

(b) (4 points) Estimate the time required to execute \( \sigma_{A=1} r \) assuming that \( r \) has a B\(^+\)-tree index on the primary key. (Assume that the entire index is stored in memory.)

(c) (4 points) Estimate the time required to execute \( \sigma_{B=1} r \Join s \) assuming that both \( r \) and \( s \) are stored as sequential files that are contiguous but not sorted and that 5% of the tuples in \( r \) have a \( B \)-value of 1, a small enough number of \( r \) tuples that we have space in memory to hold them all.

(d) (4 points) Assume that \( s \) has a B\(^+\)-tree index on the primary key \( C \). If a node in the tree may have up to 20 pointers, give the maximum height of the tree. (For counting the height, a B\(^+\)-tree with only the root node is of height 1.)
6. Let \( r_1(ABC), r_2(CDE), \) and \( r_3(ACF) \), and assume that the domain for all attributes is the set of integers. Rewrite the following queries so that they are optimal in the sense that \( \pi \) and \( \sigma \) operations are done before binary operators when possible.

(a) (6 points) \( \sigma_{(A=4 \lor C=6) \land E>=3}(r_1 \bowtie r_2 \bowtie r_3) \)

(b) (6 points) \( \pi_A \sigma_{C>1}(\pi_{AC} r_1 \cap \pi_{AC} r_3) \)

(c) (6 points) \( \pi_{ABE}(r_1 \bowtie r_2) \)

7. (10 points) Consider the following transactions:

\[
\begin{array}{l}
T_0: & \text{Read}(A) \quad \text{Read}(A) \\
& A := A + 200 \quad A := A - 100 \\
& \text{Write}(A) \quad \text{Write}(A) \\
& \text{Read}(B) \quad \text{Read}(B) \\
& B := B + 100 \quad B := B + 100 \\
& \text{Write}(B) \quad \text{Write}(B)
\end{array}
\]

Let \( A \) initially be 200, and let \( B \) initially be 50. Assume that \( T_0 \) executes then \( T_1 \) executes. If the recovery scheme is using a log with deferred updates, show the state of the log if the system crashes just before “Write(B)”. Assume all possible log entries have been written to disk. Explain how the system recovers.