Mid-Term Exam

⊕ **We will not grade the back of the sheets.**

⊕ **Avoid asking questions** - If something is unclear, document your assumptions and move on.

⊕ Do not forget to write your name on the first page. Initial each subsequent page.

⊕ **Be neat and precise. We will not grade answers we cannot read.**

⊕ If you have written something incorrect along with the correct answer, you should not expect to get all the points. I will grade based upon what you wrote, not what you meant.

⊕ You should draw simple figures if you think it will make your answers clearer.

⊕ **Good luck and remember, brevity is the soul of wit.**

<table>
<thead>
<tr>
<th>Num</th>
<th>Poss</th>
<th>Pts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>
1. (25 points) Suppose you are given the following requirements for the National Basketball Association (NBA):

- the NBA has many teams
- each team has a name, city, coach, its captain, and players
- each player is on one team
- the captain is a player
- each player has name, position, and salary
- a game is played between “home” and “away” teams, and has a date and score

(a) (15 pts) Construct a concise E-R diagram for this database.

answer:

```
[Diagram]
```

```
```
(b) (10 points) Turn the E-R diagram into a relational schema. Combine and eliminate schemas where possible.

answer:

\text{team(name, city, coach, captain)}

\text{player(name, t\textunderscore name, position, salary)}

\text{game(home, date, away, score) though could use away and date for key}
2. (25 points) Assume the following movie table, with all rating values unique.

```
create table movies (name varchar(10), rating int);
```

and an example table is as follows:

<table>
<thead>
<tr>
<th>Title</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avatar</td>
<td>75</td>
</tr>
<tr>
<td>Blade Runner</td>
<td>80</td>
</tr>
<tr>
<td>Existence</td>
<td>42</td>
</tr>
<tr>
<td>Groundhog Day</td>
<td>100</td>
</tr>
<tr>
<td>Guardians of the Galaxy</td>
<td>73</td>
</tr>
<tr>
<td>Guardians of the Galaxy II</td>
<td>25</td>
</tr>
<tr>
<td>Her</td>
<td>55</td>
</tr>
<tr>
<td>March of the Penguins</td>
<td>60</td>
</tr>
<tr>
<td>Star Trek</td>
<td>72</td>
</tr>
<tr>
<td>Titanic</td>
<td>44</td>
</tr>
<tr>
<td>Twentieth Century Women</td>
<td>83</td>
</tr>
</tbody>
</table>

(a) **(10 pts)** Write an SQL query to return the fifth highest-rated movie.

**answer:**

```
(SELECT name FROM movies m1 WHERE 4=(SELECT COUNT(name) FROM movies m2 WHERE m2.rating > m1.rating);
```

(b) **(10 pts)** Write an SQL query to return the two movies whose ratings are closest. There will always be a unique answer.

**answer:**

```
WITH t(an,bn,diff) AS (SELECT a.name,b.name,a.rating - b.rating
FROM movies a, movies b
WHERE a.rating > b.rating)
SELECT an,bn,diff FROM t tout
WHERE NOT EXISTS
(SELECT an FROM t ti WHERE ti.diff < tout.diff);
```
(c) **(5 pts)** Re-write the previous query in relational algebra.

**answer:**

$$ t \leftarrow \rho_{t_1(a,b,d)}(π_{a.name,b.name,a.rating-b.rating}(σ_{a.rating>b.rating}(ρ_a(M) × ρ_b(M)))) $$

$$ π_{a,b}(t) - π_{t.a,t.b}(σ_{t.d>t_2.d}(t × ρ_{t_2}(t))) $$
3. (25 points) Given the relation schema: \( R(A, B, C, D, E, F) \), and FDs on it:
\[
\begin{align*}
A &\rightarrow CF \\
BD &\rightarrow AE \\
F &\rightarrow B.
\end{align*}
\]

(a) (5 points) List all candidate keys.
answer: D must be there, so DB, DA, DF

(b) (5 points) Is the relation in BCNF? List all FDs that violate it if not.
answer: \( BD \rightarrow AE \) does not violate, as BD is a superkey. The other two violate BCNF.

(c) (10 points) Decompose the relation into BCNF if it is not already in it. For each sub-relation, list all attributes, carried dependencies, candidate keys, and whether in BCNF.
answer:

- **\( A \rightarrow CF \)**
  - attrs = ACF
  - cand = A
  - \{A->CF\}
  - yes

- **\( BD \rightarrow AE \)**
  - attrs = ABDE
  - cand = BD
  - \{BD->AE\}
  - yes
(d) (5 points) Is your decomposition into BCNF dependency-preserving? List one FD that is not carried if it is not.

**answer:**
No. We lose $F \rightarrow B$.
If choose $F \rightarrow B$, lose $BD \rightarrow AE$
4. (25 points) Relational Algebra

\begin{tabular}{|c|c|c|c|}
\hline
A & B & C & D \\
\hline
1 & a & 1 & x \\
2 & b & 2 & y \\
3 & c & 3 & x \\
1 & b & 4 & y \\
2 & c & & \\
2 & a & & \\
\hline
\end{tabular}

(a) $R$

(b) $S$

(c) $T$

Give the results of the following queries:

(a) $\pi_{A,D}(R \times S)$

answer: $(1,x),(1,y),(2,x),(2,y),(3,x),(3,y)$

(b) $\pi_{A,D}(R \bowtie S)$

answer: $(1,x),(1,y),(2,x),(2,y),(3,y)$

(c) $\pi_E(T) - \pi_E(S \bowtie T)$

answer: $(3)$

(d) $\sigma_{E>1}(S \bowtie T)$

answer: $(b,2,y,2),(c,4,y,2),(null,null,z,3)$

(e) $\pi_C(S) - \pi_{S,C}(\sigma_{S,C<s_{2,C}}(S \bowtie \rho_{S_2}(S)))$
answer: (4) More generally, this relational algebra expression finds the largest value of C in S.