• This is a 120 minute, closed-book exam. You are allowed to use handwritten notes on both sides of a US letter size sheet of paper.

• This exam contains 8 single-sided sheets of paper (including this one)

• Write all answers on these pages, preferably on the white space in the problem statement. Continue on the back page if running out of space but clearly number your answers if doing so.

• Make sure you attack every problem; partial credit will be awarded for incomplete or partially correct results.

| 1 | 25 |
| 2 | 25 |
| 3 | 15 |
| 4 | 10 |
| 5 | 10 |
| 6 | 15 |
| Total | 100 |
1. Relational Algebra [25]

Consider the following relational schema:

\[
\begin{align*}
\text{employee} & (\text{empno}, \text{name}, \text{office}, \text{age}) \\
\text{books} & (\text{isbn}, \text{title}, \text{authors}, \text{publisher}) \\
\text{loan} & (\text{empno}, \text{isbn}, \text{date})
\end{align*}
\]

[5] Identify the keys (by underlying the attributes) for all three relations above.

Write the following queries in relational algebra:


b. [5] Find the names of employees who have borrowed all books published by McGraw-Hill.

c. [5] Find the names of employees who have borrowed more than five different books published by McGraw-Hill.

d. [5] For each publisher, find the names of employees who have borrowed more than five books of that publisher.
2. SQL [25]

Consider the employee database with the schema:

- Employee: \( \text{employee(} \text{employee\_name, street, city)} \)
- Works: \( \text{works(} \text{employee\_name, company\_name, salary)} \)
- Company: \( \text{company(} \text{company\_name, city)} \)
- Manages: \( \text{manages(} \text{employee\_name, manager\_name)} \)

Give an expression in SQL for each of the following queries:

a. [5] Find the names, street address, and cities of residence of all employees who work for First Bank Corporation and earn more than $10000.

b. [5] Find all employees in the database who earn more than each employee of Small Bank Corporation (do not use aggregates).

c. [5] Assume that the companies may be located in several cities. Find all companies located in every city in which Small Bank Corporation is located.
d. [5] Find the company that has the most employees.

e. [5] Find those companies whose employees earn a higher salary, on average, than the average salary at *First Bank Corporation*. 
3. XML [15]

Consider the following XML document:

```xml
<? xml version='1.0' ... >
<BOOKLIST>
  <BOOK GENRE='Science' FORMAT='Hardcover'>
    <AUTHOR>
      <FIRSTNAME>Richard</FIRSTNAME>
      <LASTNAME>Feynman</LASTNAME>
    </AUTHOR>
    <TITLE>The Character of Physical Law</TITLE>
    <PUBLISHED>1980</PUBLISHED>
  </BOOK>
  <BOOK GENRE='Fiction'>
    <AUTHOR>
      <FIRSTNAME>R.K.</FIRSTNAME>
      <LASTNAME>Narayan</LASTNAME>
    </AUTHOR>
    <TITLE>Waiting for the Mahatma</TITLE>
    <PUBLISHED>1981</PUBLISHED>
  </BOOK>
  <BOOK GENRE='Fiction'>
    <AUTHOR>
      <FIRSTNAME>R.K.</FIRSTNAME>
      <LASTNAME>Narayan</LASTNAME>
    </AUTHOR>
    <TITLE>The English Teacher</TITLE>
    <PUBLISHED>1980</PUBLISHED>
  </BOOK>
</BOOKLIST>
```

Answer the following questions:

a. [5] Give a DTD of this document so that the above document is correct but the DTD is as constrained as possible.
b. [5] Give a XPath expression that selects all authors from this document (no need to eliminate duplicates).

c. [5] What is the result of the execution of the following XQuery query?

```
FOR $p$ IN //BOOK
LET $q := $p/AUTHOR
LET $r := $p/PUBLISHED
LET $t := $p/@GENRE
RETURN
<PUBLICATION>
  <AUTHORNAME> {$q/FIRSNNAME} {$q/LASTNAME} </AUTHORNAME>
  {$r}
  <GENRE>{$t}</GENRE>
</PUBLICATION>
```

4. Transactions [10]

Consider the following two transactions:

T1: READ(A, t); t:=t+2; WRITE(A, t); READ(B, t); t:=t*3; WRITE(B, t);
T2: READ(B, s); s:=s*2; WRITE(B, s); READ(A, s); s:=s+3; WRITE(A, s);

where READ(A, t) means read element A into local variable t and WRITE(A, t) is similarly defined.

With respect to modification of the database elements A, B, answer the following questions:

a. [5] Pick values for t, s and an ordering of the execution of the operations in the two transactions to show that the result could be different than a serial execution of the transactions (remember that there are two possible serial executions, T1, T2 and T2, T1).
b. [5] Suppose the transactions are guarded by \textit{begin transaction} and \textit{commit} (i.e. transaction isolation is enforced by the database system). Pick two of the four SQL isolation levels discussed in class and explain differences in the execution of the two transactions under the two isolation levels.

5. Data-mining [10]

a. [5] Explain what is the difference between \textsc{ROLAP} and \textsc{MOLAP}.

b. [5] Briefly describe the \textit{frequent itemset problem} and how the \textit{A-priori} algorithm finds such frequent itemsets.


a. [3] Describe the difference in meaning between the terms \textit{relation} and \textit{relation schema}
b. [3] Why are functional dependencies important? Why should they be preserved?

c. [3] Why database systems have their own security and access control mechanisms (as opposed to using the operating systems mechanisms).

d. [3] Explain the differences, from a database perspective, between Object-Relational and Object Oriented.

e. [3] Why is concurrency control important?