CS585

Database Systems

Summer 2006

Final Exam

Name: _____

Student ID: _____

	Maximum	Received
Problem 1	10	
Problem 2	15	
Problem 3	15	
Problem 4	15	
Problem 5	15	
Problem 6	15	
Problem 7	15	

1) 10 pts

True/False questions

1. _____ Elements are **defined** within an XML schema.

2. _____ Attribute groups are **declared** within an XML schema.

3. _____ Types are **defined** within an XML schema.

4. _____ XQuery has some programming language features but SQL has not.

5. _____ SQL and XQuery are used to query data and considered functional languages.

6. _____ SQL and XQuery have Update capability.

7. _____ A distributed database requires distributed processing.

8. _____ Distributed database systems communicate with one another through a communication network.

9. _____ Thru query optimization, distributed databases always find the most efficient method of processing for a given query.

10. _____ Transaction data such as sales orders or inventory information is always aggregated before storing into OLAP cubes.

2) 20 pts
Short answer questions. Please answer briefly and to the point. 3-5 sentences will be enough to answer for each question.

a. Describe how and under what circumstances a distributed spatial database could have better performance than a central spatial database when computing the following query:
SELECT city.name FROM city WHERE city.center INSIDE California.region

b. Give two reasons why you would replicate or not replicate the data at local nodes in a distributed database.

Reasons to replicate
1)

2)

Reasons not to replicate
1)

2)

c. What is the difference between data mining and decision support using OLAP?

d. Can we express “if conditions” in XML language? Justify your answer.

3) 15 pts

Write a DTD that describes the following XML data:

```
<bib>
  <vendor id="id0_1">
    <name>Amazon</name>
    <email>webmaster@amazon.com</email>
    <phone>1-800-555-9999</phone>
    <book>
        <title>Unix Network Programming</title>
        <publisher>Addison Wesley</publisher>
        <year>1995</year>
        <author>
            <firstname>Richard</firstname>
            <lastname>Stevens</lastname>
        </author>
        <price>38.68</price>
    </book>
    <book>
        <title>An Introduction to O-O Design</title>
        <publisher>Addison Wesley</publisher>
        <year>1996</year>
        <author>
            <firstname>Jo</firstname>
            <lastname>Levin</lastname>
        </author>
        <author>
            <firstname>Harold</firstname>
            <lastname>Perry</lastname>
        </author>
        <price>11.55</price>
    </book>
 </vendor>
</bib>
```

Additional Space.

4) 15 pts
Consider the following XML document type definition (DTD) for a product catalog:

```
<!DOCTYPE CATALOG [
<!ELEMENT CATALOG (TOOL | TOY)+>
<!ELEMENT TOOL (NAME,SPECIFICATIONS+,OPTIONS?)>
<!ELEMENT NAME (#PCDATA)>
<!ELEMENT SPECIFICATIONS (#PCDATA)>
<!ELEMENT OPTIONS (#PCDATA)>
<!ELEMENT TOY (NAME,PRICE?)>
<!ELEMENT PRICE (#PCDATA)>]>
```

We designed a relational database schema that captures the same information as the Catalog DTD as below:

```
Tools(toolid: integer, name: string, options: string)
Specifications(toolid: integer, specno: integer, spec: string)
Toys(toyid: integer, name: string, price: string)
```

a. Given your relational database schema, translate the following SQL query into an equivalent
XML query:

```
SELECT S.spec
FROM Tools T, Specifications S
WHERE T.toolid = S.toolid AND S.specno = 1 AND T.name = “Hammer”
```

b. Given your relational database schema, translate the following SQL query into an equivalent
XML query:

```
SELECT name FROM Tools
UNION
SELECT name FROM Toys
```

5) 15 pts

Consider the following XML Schema . Is the following allowed according to the XML schema specification rules? If not, list all errors and explain what rules are violated.

```
<xsd:complexType name="ssnType" minOccurs="1">
    <xsd:restriction base="xsd:string">
        <xsd:pattern value="\d{3}-\d{2}-\d{4}"/>
        <xsd:pattern value="\d{9}"/>
    </xsd:restriction>
</xsd:complexType>


<xsd:element name="Company">
    <xsd:complexType>
     <xsd:sequence>
      <xsd:element name="Employee"  type="xsd:string"
                   maxOccurs="unbounded">
       <xsd:complexType>
          <xsd:sequence>
             <xsd:element name="SSN" type="ssnType"/>
             <xsd:element name="name" type="xsd:string" />
            <xsd:element name="BirthDate" type="xsd:gDate" use="optional"/>
             <xsd:element name="address" minOccurs="0">
                <xsd:simpleType>
                 <xsd:attribute name="city" type="xsd:string" maxOccurs="1”/>
                 <xsd:attribute name="state" type="xsd:string" maxOccurs="1"/>
                 <xsd:attribute name="zip" type="xsd:integer" maxOccurs="1"/>
                </xsd:simpleType>
              </xsd:element>
          </xsd:sequence>
         </xsd:complexType>
      </xsd:element>
     </xsd:sequence>
  </xsd:complexType>
</xsd:element>
```

Additional space

6) 15 pts
Given the distributed database design below:

Actors (Name: String,
Gender: String,
Agent: String)

Actors relation split across 4 nodes with
- Node 1 keeps male actors with Name < “M”
- Node 2 keeps males actors with Name >= “M”
- Node 3 keeps female actors with Name < “M”
- Node 4 keeps females actors with Name >= “M”

AppearsIn(Name: String,
Title: String)

AppearsIn relation split across 2 nodes
- Node 5 keeps tuples with Name < “M”
- Node 6 keeps tuples with Name >= “M”

Movies (Title: String,
Producer: String,
Year: Integer,
Cost: Integer,
Location: String,
Director: String)

Movies relation split across 2 ndoes
- Node 7 keeps movies with Year > 1980
- Node 8 keeps movies with Year <= 1980

And the query

```
SELECT    Actors.Agent
FROM      Actors, AppearsIn, Movies
WHERE     Actors.Name=AppearsIn.Name AND
          Movies.Tile=AppearsIn.Title AND
          Movies.Cost > 30,000,000 AND
          Movies.Location=”LA”
```

And the following assumptions:
- AppearsIn relation is of size 3000
- Actors relation is of size 1000
- Movies relation is of size 200
- Costs are between $100,000 and $200,000,000
- 10% of movies are made in LA

Specify an efficient order of operations for the above query. You must also show all transfer of data from site to site. State any additional assumptions you need to make.

7) 15 pts

Given the relation

Movies (Title: String,
Producer: String,
Year: Integer,
Cost: Integer,
Director: String)

And the queries

q1: SELECT Title FROM Movies
q2: SELECT Producer,Year FROM Movies
q3: SELECT Title, Cost FROM Movies
q4: SELECT Director FROM Movies

With the following access frequency from sites 1 to 3

	S1	S2	S3
q1	20	0	5
q2	0	10	0
q3	5	5	20
q4	0	10	5

a- Create the attribute affinity matrix

b- Determine an optimal vertical fragmentation of this relation into 3 fragments.

Additional Space

Additional Space