COMP62421

Two hours

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

Querying Data on the Web

Date: Wednesday 27th January 2016

Time: 09:45 - 11:45

Please answer ALL Questions provided

This is a CLOSED book examination

The use of electronic calculators is permitted provided they are not programmable and do not store text

[PTO]

1. a) Briefly explain why, unlike general-purpose programming languages, database languages have three well-defined, separate sublanguages. (2 marks)

   b) State two of the three main subcomponents in a query processor. (1 mark)

   c) Assume the following database on beers that are served in bars where drinkers that frequent the bars are recorded as liking a beer or not:

```
CREATE TABLE Bar
   (name    VARCHAR(20) NOT NULL PRIMARY KEY,
    address VARCHAR(20));
CREATE TABLE Beer
   (name    VARCHAR(20) NOT NULL PRIMARY KEY,
    brewer  VARCHAR(20));
CREATE TABLE Drinker
   (name    VARCHAR(20) NOT NULL PRIMARY KEY,
    address VARCHAR(20));
CREATE TABLE Frequents
   (drinker VARCHAR(20) NOT NULL REFERENCES Drinker(name),
    bar     VARCHAR(20) NOT NULL REFERENCES Bar(name),
    times_a_week SMALLINT CHECK(times_a_week > 0),
    PRIMARY KEY(drinker, bar));
CREATE TABLE Likes
   (drinker VARCHAR(20) NOT NULL REFERENCES Drinker(name),
    beer    VARCHAR(20) NOT NULL REFERENCES Beer(name),
    PRIMARY KEY(drinker, beer));
CREATE TABLE Serves
   (bar     VARCHAR(20) NOT NULL REFERENCES Bar(name),
    beer    VARCHAR(20) NOT NULL REFERENCES Beer(name),
    price   DECIMAL(5,2) CHECK(price > 0),
    PRIMARY KEY(bar, beer));
```

   Now, consider the problem P stated as follows: *"Retrieve the name and address of drinkers who frequent a bar more than once a week as well as the name and address of such bars."*

   i) Write a tuple-relational calculus expression that computes the answer to P. (4 marks)

   ii) Write a relational algebraic expression that computes the answer to P. (3 marks)

2. a) Briefly describe the basic operator interaction protocols in query evaluation. (2 marks)

   b) Briefly describe two positive consequences for performance that the iterator model for physical operator design makes more probable. (2 marks)

   c) Given the database on beers, bars and drinkers in the previous question, write a SQL query that returns the name of each beer and its popularity (i.e., the count of how many drinkers like a given beer) from the most to the least popular. (3 marks)

   d) Assume that the answer to the previous item is stored as a column in an extended beer table, so that every beer tuple allows one to access the number of drinkers that like it. Further assume that the answer has revealed that one beer is overwhelmingly the most popular, i.e., all other beers are liked by significantly fewer drinkers. A colleague of yours suggested that range partitioning the `likes` table on the `popularity` column is a good idea. Would you recommend this? Explain why or why not. If you believe that relevant information is missing, then make explicit any assumptions that were required for you to answer. (3 marks)

3. a) Briefly describe two reasons, among those discussed in the course unit, why relational languages are not directly suitable for querying XML data. (2 marks)

b) Consider the following XQuery function:

```
declare function local:mystery
    ($y) as xs:integer+
    {for $x in $y
          return (if ($x mod 2 = 0)
                        then $x+1
                        else 1
                  )
    }
```

Recall that XQuery `mod` returns the remainder after division of one number by another function. Further recall that $f(x)$ is idempotent iff $f(f(x)) = f(x)$, for all x. State whether `local:mystery` is idempotent and argue (briefly and informally) that it is idempotent if you do believe it is, or else show with a counterexample that it is not idempotent. (3 marks)

c) Write out the expression that results from mapping the following XPath expression into XQuery Core:

```
$root/employees/emp/name[@title = "Dr"]
```

(5 marks)

4. a) Briefly explain why the triple store strategy for mapping RDF data into relational data creates obstacles for efficient evaluation. (2 marks)

   b) Use the property table strategy to map the following RDF triples into relational tables: (8 marks)

```
(b3, d-name,    billbarley )
(b3, d-address, fallowfield)
(b4, d-name,    helenhops)
(b4, d-address, chorlton )
(b5, be-name,   liberty)
(b5, be-brewer, marble )
(b6, be-name,   clementine)
(b6, be-brewer, prospect  )
(b4, likes, b5)
(b3, likes, b6)
(b3, likes, b3)
```

5. a) Briefly contrast ACID guarantees in SQL databases and BASE guarantees in NoSQL databases. (2 marks)

   b) Consider the following RA algebraic expression:

$$(R \bowtie S) \cup T$$

   where the schemas of the input relations are

$$R(a, b)$$
$$S(b, c)$$
$$T(a, b, c)$$

   For each non-leaf node in the logical operator tree, sketch the pseudocode of the mapper and reducer functions that would compute the correct value for the given algebraic expression. (8 marks)

END OF EXAMINATION