Database Systems – Mini-project

sw613f14

Bruno Thalmann
Mikael E. Christensen
Mikkel S. Larsen
Stefan M. G. Micheelsen

Exam preparation guideline:
If you are well prepared for the exam, solving these exercises should take you not take you much longer than 1.5 hours. At the moment, however, you have not finished exam preparation, so it is expected that you are still slower.

1 Multiple Choice

Please mark all true statements!

1. Given the following relation: `student (SID, name)`
   as well as the following queries:

```
SELECT COUNT(*) AS cnt          SELECT COUNT(name) AS cnt
FROM student;                   FROM student;
```

   - ☐ The results of both queries are always the same, independent from the relations' extension.
   - ☒ The results of both queries are sometimes the same, sometimes not; depending on the relations' extension.
   - ☐ The results of both queries are never the same if the extension of the relations contains at least one tuple.
   - ☐ The results of both queries are never the same, independent from the relations' extension.

2. Given the following query:

```
SELECT 5-5=NULL AS useless;
```

   - ☐ The query result is `TRUE`.
   - ☐ The query result is `FALSE`.
   - ☐ The query result is `NULL`.
   - ☒ The query cannot be executed because it is invalid.

3. Given two relations with no common attributes. The natural join (`NATURAL JOIN`) between both relations...
   - ☒ is equivalent to the cross product.
   - ☐ always produces an empty result set.
   - ☐ never produces an empty result set.
   - ☐ cannot be executed.

4. Given relation $\mathcal{R}$ with two attributes a and b. Which of the following statements do not result in an erorr?

   - ☒ `SELECT a, b FROM R GROUP BY a, b;`
   - ☒ `SELECT a FROM R GROUP BY a, b;`
   - ☐ `SELECT b FROM R GROUP BY a;`
   - ☐ `SELECT a, b FROM R GROUP BY a;`

5. Given the following relations `R (a, x)` and `S (b, x)`, and the following query:

   ```
   (SELECT a FROM R) INTERSECT (SELECT b FROM S);
   ```

   Which of the following queries produces the same result?

   - ☒ `SELECT a FROM R JOIN S ON a=b;`
   - ☐ `(SELECT a FROM R) EXCEPT (SELECT b FROM S);`
   - ☐ `(SELECT a FROM R) UNION (SELECT b FROM S);`
   - ☐ `(SELECT a FROM R) UNION DISTINCT (SELECT b FROM S);`
   - ☐ `None of the above.`

6. By computing the attribute closure for a set of attributes α, we can determine if α

   - ☐ conains any prime attributes
   - ☐ contains any non-prime attributes
   - ☐ α is a candidate key
   - ☒ α is a super key

2 Integrity Constraints

Given the following database schema for a discussion forum:
`user (ID, name, age)`
`topic (ID, name, author → user)`
`post (ID, author → user, topic → topic, subject, content)`

Extend the following `CREATE TABLE` statements with the following integrity constraints:

- `ID` attributes uniquely identify rows in each of the relations.
- Users have a unique name.
- Users must be at least 16 years old and apart from that have a realistic age (e.g., < 150).
- If a user is deleted, then so are all his/her posts and all topics that he/she started.
- If a topic is deleted, then so are all its related posts.

```
CREATE TABLE user (
ID INTEGER PRIMARY KEY,
name VARCHAR(32) UNIQUE,
age INTEGER CHECK (age BETWEEN 16 AND 149)
);

CREATE TABLE topic (
ID INTEGER PRIMARY KEY,
name VARCHAR(32),
author INTEGER REFERENCES user(ID) ON DELETE CASCADE
);

CREATE TABLE post (
ID INTEGER PRIMARY KEY,
author INTEGER REFERENCES user(ID) ON DELETE CASCADE,
topic INTEGER REFERENCES topic(ID) ON DELETE CASCADE,
subject VARCHAR(32),
content VARCHAR(1024)
);
```

3 Recursion in SQL

Please write an SQL query that recursively computes 2^5. You may use the following template:

```
WITH RECURSIVE tempTable(number, counter) AS ( (...)
)
SELECT (...)
FROM tempTable;
```

```
WITH RECURSIVE tmpTable(number, counter) AS (
    VALUES(2,1)
  UNION
    SELECT number * 2, counter+1
    FROM tmpTable
    WHERE counter < 5
)
SELECT MAX(number)
FROM tmpTable
```

5 Normalization

Given the following list of relations with their functional dependencies, determine for each relation its candidate keys, then determine the highest fulfilled normal form (up to but including BCNF). You may assume that 1NF is always fulfilled. Explain your answer!

1. $\mathcal{R} \rightarrow \{A, B, C, D\}$
   $A \rightarrow B$
   $B \rightarrow A$
   $C \rightarrow D$
   $D \rightarrow C$

1) candidate keys: AC, AD, BC, BD

3NF. right side of all FDs are prime

2. $\mathcal{R} \rightarrow \{A, B, C, D, E, F\}$
   $AB \rightarrow CDF$
   $CF \rightarrow AB$
   $E \rightarrow F$

2) candidate keys: ABE, CE

1NF, because the two non-prime attributes D, F are not fully functionn dependent on each candidate key

3. $\mathcal{R} \rightarrow \{A, B, C, D, E\}$
   $A \rightarrow EB$
   $C \rightarrow AD$
   $B \rightarrow C$
   $D \rightarrow B$

3) candidate keys: A, C, B, D

BCNF, since all FD left-sides are super keys

4 SQL Queries

Given the following relational schema for an internet discussion forum:

`user (UID, name, registrationDate)`
`category (CID, name)`
`topic (TID, CID → category, title)`
`post (BID, TID → topic, text, creator → user, creationDate)`

Formulate equivalent SQL statements on the above schema for the following queries.

1. Find the UIDs and the number of posts for the 10 users with the most posts, sorted in descending order by the number of posts.
2. Find the title(s) of the topic(s) with the most posts. Attention: there might be only one or multiple such topics.
3. Find the titles of the 10 topics with the latest posts.
4. Find the UIDs of all users who contributed more than 30 posts.

1)
```
SELECT creator UID, COUNT(*) posts
FROM post
GROUP BY UID
ORDER BY posts DESC
FETCH FIRST 10 ROWS ONLY
```

2)
```
SELECT title
FROM topic
WHERE TID IN (
    SELECT TID, COUNT(*) AS c
    FROM post
    WHERE c = (
        SELECT MAX(c) FROM(
            SELECT count(*) AS c
            FROM post
            GROUP BY TID
            )
        )
    GROUP BY TID
    )
```

3)
```
SELECT DISTINCT topic
FROM topic, post
WHERE topic.TID = post.TID
ORDER BY creationDate DESC
FETCH FIRST 10 ROWS ONLY
```

4)
```
SELECT creator AS UID
FROM post
GROUP BY UID
HAVING COUNT(*) >= 30
```

6 Functional Dependencies

Given the following extension of $\mathcal{R} = \{A, B, C, D, E\}$.

A	B	C	D	E
a1	b1	c1	d1	e1
a1	b2	c1	d2	e1
a2	b2	c2	d2	e1
a3	b2	c2	d3	e2
a4	b2	c3	d3	e2
a4	b3	c4	d3	e3
a4	b3	c5	d3	e4

1. Please indicate whether the following dependencies hold for $\mathcal{R}$ (you may omit an explanation of your answer).
   - $AB \rightarrow C$ ÷
   - $CDE \rightarrow B$ ✓
   - $CE \rightarrow A$ ✓
   - $BC \rightarrow E$ ÷
   - $BCDE \rightarrow A$ ✓
   - $D \rightarrow AE$ ÷
2. Please indicate which of the following attribute sets are candidate keys and which are super keys (you may omit an explanation of your answer).
   - BCDE super
   - ABE candidate
   - DE —
   - CD candidate

Instructions to prepare the report

We will try to have a double-blinded review process, i.e., you do not know the group whose report you are reviewing and you do not know which group is reviewing your own report. Hence, state your group name as well as the members of your group on the first page only. Do not repeat your group name and the names of group members on other pages. What will be handed over to the peer group is then everything except the first page.

Course goals covered by this self study

- Make use of SQL to create, modify, and query relational databases
- Create and evaluate a database schema that adheres to normal forms (logical design)

Instructions for peer review

After the deadline to hand in your own reports (12.05.2014), you will receive a report handed in from another group. Your task is to "grade" the handed in solutions and decide whether they are correct and how many points should be awarded.

To give you some guidance, we will provide you with a set of solutions for the exercises. Your task is then to decide how correct the solutions of your peer group are. As there is often more than one correct solution, you first need to decide on the correctness and then identify how much weight you want to give to coverage and mistakes in the solutions (regarding both the semantics of the solution and the notation).

Hence, for each exercise and handed in solution, you should give a short statement on how correct (in the range of 0% to 100%) the solution is. Furthermore, you should explain how and why you gave less then 100% – be explicit, e.g., we gave 80% because..., we reduced by 5% because.... In addition, please also give a final grade to the report that you were given – use the 7-trins-skala and add a short explanation.

As a consequence of the double-blinded review process, for the feedback please again prepare a first page listing group name and members and do not repeat this information on other pages.

Selfstudy: 06.05.2014
The report must be handed in via Moodle no later than
12.05.2014, 23:55 CET
Peer reviews must be handed in via Moodle no later than
16.05.2014, 23:55 CET