Bc. Pavol Loffay (učo 374115)
Studium FI N-AP SSME [sem 2, roč 1]
jaro 2014 (jiné)

Odpovědník Home Assignment No. 2

Odpovědi k průchodu Út 22. 4. 2014 00:07.42, operaci Ne 4. 5. 2014 17:24.35, osobě P. Loffay, učo 374115

This assignment must be completed with meaningful answers when handing in.
Solutions with incomplete parts or filled with garbage will be automatically classified
as unsatisfactory without any later corrections possible! Work on the task individually.
You can work with the ROPOT repeatedly until the deadline, after which your answers
will be evaluated manually. There are some automatic value range test prepared just as
a pre-test for you -- if it fails, it may be a warning that you filled sth. in incorrectly.

• Klikněte: Ukaž Přehled nastavení parametrů odpovědníku.

Přehled nastavení parametrů odpovědníku

Kdy lze s odpovědníkem pracovat:

- od 21. 4. 2014 00:00 do 4. 5. 2014 23:59

Zobrazují se pouze správné odpovědi: ne
Test můžu skládat opakovaně: ano, pokaždé stejnou sadu otázek
Způsob bodování: neboduje se
Při vyplňování záleží na velikosti písmen: ne
Při vyplňování záleží na diakritice: ne
Při vyplňování nedovoluji zaměnit různé typy apostrofů a uvozovek: ne
Při vyplňování záleží na interpunkci: ne

Zeleně jsou vyznačeny správné odpovědi.

Query result size estimation

This task consists of three parts: 1) SQL query formulation, 2) query trasformation into relational algebra expression, and 3) final transformation to query evaluation plan. For each operation in the query evaluation plan (i.e. each partial query result) you will provide basic statistics (result set size, number of different attribute values, etc.).

In case you are not aware of relational algebra, please refer to a lecture of basic database course here.

Symbols used for relational algebra operations: *(for plain text typing)*

- `$[condition]` - selection - \sigma_{A=13}(relation)
- `#[attribute list]` - projection - \Pi_{A,B,D}(relation)
- `r * s` - Cartesian product - r \times s
- `r @ s` - natural join - \newcommand{\njoin}{\bowtie} r \njoin s
- `r ~@ s` - left outer join - \newcommand{\ljoin}{\sqsupset\hspace{-1pt}\bowtie} r \ljoin s
- `r @~ s` - right outer join - \newcommand{\rjoin}{\bowtie\hspace{-1pt}\sqsubset} r \rjoin s
- `r ~@~ s` - full outer join - \newcommand{\fjoin}{\sqsupset\!\bowtie\!\;\!\sqsubset} r \fjoin s

Query evaluation plan notation: *(for plain text typing)*

Relational algebra expression `$[A=B](r * s)` is transformed into the following query evaluation plan:

Task definition

Assume the following relations:

- student(studentId, name, surname, fieldOfStudy)
- enrollment(studentId, courseId, typeOfCompletion), where *studentId* is a foreign key into *student* and *courseId* is a foreign key into *course*
- course(courseId, name, numberOfCredits, recommendedCompletion)

The possible types of completion are credit ('z'), colloquium ('k'), and examination ('zk').

In the following assignments, work with a query that retrieves

- identification of students (studentId), who have enrolled for the course called *Matematika A* or *Matematika B* and have selected a type of completion equal to the recommended completion for these courses.

Remark: Prerequisites of these courses prevent students from registering for both the courses.

1. Formulate an SQL command for the query: (1 point) *(you should check your query on a data sample here)*

```
SELECT e.studentid
FROM enrollment AS e
LEFT JOIN course AS c ON e.courseid=c.courseid
WHERE (c.name='Matematika A' OR c.name='Matematika B')
AND e.typeofcompletion=c.recommendedcompletion
```

SELECT e.studentid FROM enrollment AS e LEFT JOIN course AS c ON e.courseid=c.courseid WHERE (c.name='Matematika A' OR c.name='Matematika B') AND e.typeofcompletion=c.recommendedcompletion

2. Transform your SQL command into a relational algebra expression: (0,5 point) *(see the rel. alg. syntax defined above; **you should the correctness of your expression on a data sample here**)*

```
#[studentid]($[(name='Matematika A' | name='Matematika B') & typeofcompletion=recommendedcompletion](enrollment @ course))
```

#[studentid]($[(name='Matematika A' | name='Matematika B') & typeofcompletion=recommendedcompletion](enrollment @ course))

3. **Prepare an evaluation plan** for the relational algebra expression formulated in the previous subtask. (1 point)
Draw the plan as a tree structure, as suggested below. Assign a unique identifier to each operation in the plan (this will be needed in the following subtask) like in this exa

```
Y:   $[A=B]
     |
     |
X:   *
    / \
   /   \
  r     s
```

```
:#[studentid]($[(name='Matematika A' | name='Matematika B') & typeofcompletion=recommendedcompletion](enrollment @ course))

Z         #[studentid]
          |
          |
Y         $[(name='Matematika A' | name='Matematika B') & typeofcompletion= recommendedcompletion]
          |
          |
X   ------@-------
    |             |
    |             |
enrollment    course

Tento vyraz sa da upravit na:
#[studentid]($[typeofcompletion=recommendedcompletion](($[name='Matematika A'](#[courseid,name,recommendedcompletion](course)) + $[na
[courseid,name,recommendedcompletion](course))) @ enrollment))

- 1. sa spravi projekciu nad course(atributy courseid,name,recommendedcompletion) a nasledne selekciu 'Matematika A' to iste ale s s
- 2. spravi sa zjednotenie tychto dvoch dotazov
- 3. spravi sa natural join s vysledkom a tabulkou enrollment
- 4. vyberem vysledky kde typeofcompetion=recommendedcompletion
- 4. projekcia studentid



F                                   #[studentid]
                                    |
                                    |
E                                   $[typeofcompetion=recommendedcompletion]
                                    |
                                    |
D                                   @ enrollment
                                    |
                                    |
C             ----------------------  +   -------------------------
              |                                                    |
              |                                                    |
B             $[name='Matematika A']                               $[name='Matematika B']
              |                                                    |
              |                                                    |
A             #[courseid,name,recommendedcompletion]               #[courseid,name,recommendedcompletion]
              |                                                    |
              |                                                    |
            course                                               course
```

:#[studentid]($[(name='Matematika A' | name='Matematika B') & typeofcompletion=recommendedcompletion](enrollment @ course)) Z #[studentid] | | Y $[(name='Matem typeofcompletion= recommendedcompletion] | | X ------@------- | | | | enrollment course Tento vyraz sa da upravit na: #[studentid]($[typeofcompletion=recommendedcomp [courseid,name,recommendedcompletion](course)) + $[name='Matematika B'](#[courseid,name,recommendedcompletion](course))) @ enrollment)) - 1. sa spravi projekci courseid,name,recommendedcompletion) a nasledne selekciu 'Matematika A' to iste ale s selekciou 'Matematika 2' - 2. spravi sa zjednotenie tychto dvoch dotazov - 3. spra enrollment - 4. vyberem vysledky kde typeofcompetion=recommendedcompletion - 4. projekcia studentid F #[studentid] | | E $[typeofcompetion=recommendedcompletio ---------- | | | | B $[name='Matematika A'] $[name='Matematika B'] | | | | A #[courseid,name,recommendedcompletion] #[courseid,name,recommendedcompletion

4. **Statistics about the relations** (fill in the missing values): (0.5 point)

T(student)	V(student,studentId)	V(student,name)	V(student,surname)	V(student,fieldOfStudy)
6100	6100 ✔ (6100)	3000	5900	10
T(enrollment)	V(enrollment,studentId)	V(enrollment,courseId)	V(enrollment,typeOfCompletion)	
27000	5000	90	3 ✔ (3)	
T(course)	V(course,courseId)	V(course,name)	V(course,numberOfCredits)	V(course,recommendedCompletion)

100	100 ✔ (100)	100	8	3 ✔ (3)

In addition, we know that **10% students do not have the mathematics courses mandatory**, so they can select other type of completion than the recommended one. **Half of these students do so.**

Hint: What is the primary key of the relation? | Hint: How many possible types of completion are there? | Hint: What is the primary key of the relation? | Hint: How many possible types of completion are there?

5. **Estimate the number of records in the query result.** (2 points)
   Compute the necessary statistics for each operation in the query evaluation plan (i.e. each unique identifier), starting from leaves and moving up.
   (describe your approach including all the formulas you use)

```
A: T(A) =  100...................100
B: T(B) =  100/100 = 1...........1
C: T(C) =  2
D: natural join sa urobi na atribute courseid -> T(enrollment,courseid) = 90
 T(priebezny_vysledok) = 2 s toho plynie 2 < 90

T(W) = T(R1) * T(R2)/V(R2,spojovaci_atribut)
T(D) = 2 * 27000/90 = 600
T(E) = 600/3 = 200   (T(relaceX)/V(relaceX,atribut))
T(F) = 200
```

A: T(A) = 100...................100 B: T(B) = 100/100 = 1...........1 C: T(C) = 2 D: natural join sa urobi na atribute courseid -> T(enrollment,courseid) = 90 T(priebezny_vysledok) = 2 s toho plynie 2 < 90 T(W) = T(R1) * T(R2)/V(R2,spojovaci_atribut) T(D) = 2 * 27000/90 = 600 T(E) = 600/3 = 200 (T(relaceX)/V(relaceX,atribut)) T(F) = 200

- Zpět na výběr operace

Bez uložení: