

Name

CWID

# Homework Assignment 5

May 5, 2016

## CS520 Results

---

*Please leave this empty!*

1.1

1.2

1.3

Sum

# Lab Part

- This part of the assignment helps you to practice the techniques we have introduced in class

## Part 1.1 Download & Install Perm (Provenance-aware Postgres) (Total: 0 Points)

- Perm is an extended Postgres server with extended SQL dialect that has new keywords for computing provenance
- Download Perm from sourceforge and install the server. See <http://www.cs.iit.edu/~dbgroup/research/perm.php> for instructions

## Part 1.2 Create example database (Total: 0 Points)

- Create university from the Silberschatz textbook (a script is shown at the end of this homework assignment).

## Part 1.3 Compute provenance (Total: 0 Points)

- Connect to the database using the psql commandline client or any other interface, e.g., JDBC
- Run some example SQL queries
- To compute the provenance of a query add the keyword **PROVENANCE** after the select clause

For example, get the names of instructors supervising students

```
SELECT i.name
FROM advisor a, student s, instructor i
WHERE a.i_id = i.id AND a.s_id = s.id;
```

Get the provenance:

```
SELECT PROVENANCE i.name
FROM advisor a, student s, instructor i
WHERE a.i_id = i.id AND a.s_id = s.id;
```

# Theory Part

- This part of the assignment helps you to practice the techniques we have introduced in class.

Consider the following database schema and instance:

## user

name	limitMB	
peter	50	$u_1$
hans	50	$u_2$
hilde	100	$u_3$
gertrud	1000	$u_4$
urs	1500	$u_5$

## file

id	name	sizeMB	format	
1	losung.pdf	1	PDF	$f_1$
2	urlaub.jpg	1	JPG	$f_2$
3	vorstellung.avi	347	Quicktime	$f_3$
4	stoiber.wmv	15	WindowsMediaVideo	$f_4$
5	faust.txt	2	Text	$f_5$
6	faust2.txt	1	Text	$f_6$
7	rambo3.avi	745	Quicktime	$f_7$

## downloads

userName	fileId	date	
peter	5	1.1.2007	$d_1$
peter	6	1.1.2007	$d_2$
hilde	7	25.7.2007	$d_3$
peter	2	1.1.2007	$d_4$
urs	1	4.9.2007	$d_5$

### Hints:

- Attribute with black background are the primary key of an relation
- The attribut *userName* of relation *downloads* is a foreign key to *name* in relation *user*.
- The attribute *fileId* of relation *downloads* is a foreign key to attribute *id* in relation *file*.

## **Part 1.4 Foundations of Provenance Models (Total: 0 Points)**

### **Question 1.4.1 (0 Points)**

Explain the **necessity** and **sufficiency** principles that are the foundation of most provenance models?

### **Question 1.4.2 (0 Points)**

List the provenance models you have seen so far.

### **Question 1.4.3 (0 Points)**

What are the advantages and disadvantages of the declarative and syntactic definitions of provenance?

### **Question 1.4.4 (0 Points)**

What is insensitivity to query rewrite and how does this property relate to query equivalence?

**Question 1.4.5 (0 Points)**

How is the concept of a witness used in Why-provenance defined?

**Question 1.4.6 (0 Points)**

How can insensitive Why-provenance be computed from Why-provenance?

**Question 1.4.7 (0 Points)**

The minimal why-provenance is insensitive to query rewrite. The proof is very straight-forward. Can you write it down?

**Question 1.4.8 (0 Points)**

What properties have to hold for  $(K, +, \times, 0, 1)$  to be a semiring?

**Question 1.4.9 (0 Points)**

Give definitions for the operators of the relational algebra over semiring-annotated relations.

**Question 1.4.10 (0 Points)**

Which are important semirings in the semiring annotated relational model and what is their correspondence in the standard relational model?

**Question 1.4.11 (0 Points)**

Define the semiring that corresponds to why-provenance and the semiring that corresponds to minimal why-provenance.

## Part 1.5 Provenance Computation (Total: 0 Points)

For the following the queries compute the provenance according to the following provenance models for their all result tuples.

- Why-Provenance
- Minimal Why-Provenance
- Provenance Polynomials

### Question 1.5.1 (0 Points)

$$\pi_{name}(user)$$

#### Solution

Result relation:

name	
peter	$t_1$
hans	$t_2$
hilde	$t_3$
gertrud	$t_4$
urs	$t_5$

Why provenance:

name	
peter	$\{\{u_1\}\}$
hans	$\{\{u_2\}\}$
hilde	$\{\{u_3\}\}$
gertrud	$\{\{u_4\}\}$
urs	$\{\{u_5\}\}$

Minimal Why provenance:

name	
peter	$\{\{u_1\}\}$
hans	$\{\{u_2\}\}$
hilde	$\{\{u_3\}\}$
gertrud	$\{\{u_4\}\}$
urs	$\{\{u_5\}\}$

Provenance Polynomials:

name	
peter	$u_1$
hans	$u_2$
hilde	$u_3$
gertrud	$u_4$
urs	$u_5$





### Question 1.5.2 (0 Points)

$$\pi_{name}(\sigma_{name=peter \vee name=gertrud}(user)) \bowtie \rho_{name \leftarrow userName}(downloads)$$

#### Solution

Result relation:

name
peter

 $t_1$ 

Why provenance:

name
peter

 $\{\{u_1, d_1\}, \{u_1, d_2\}, \{u_1, d_4\}\}$ 

Minimal Why provenance:

name
peter

 $\{\{u_1, d_1\}, \{u_1, d_2\}, \{u_1, d_4\}\}$ 

Provenance Polynomials:

name
peter

 $u_1 \times d_1 + u_1 \times d_2 + u_1 \times d_4$

**Question 1.5.3 (0 Points)**

$$\pi_{name}(\sigma_{format=Quicktime}(file)) \cup \pi_{name}(\sigma_{sizeMB>100}(file)) \bowtie \rho_{id \leftarrow fileId}(downloads)$$

**Solution**

**Result relation:**

name	
vorstellung.avi	$t_1$
rambo3.avi	$t_2$

**Why provenance:**

name	
vorstellung.avi	$\{\{f_3\}\}$
rambo3.avi	$\{\{f_7\}, \{f_7, d_3\}\}$

**Minimal Why provenance:**

name	
vorstellung.avi	$\{\{f_3\}\}$
rambo3.avi	$\{\{f_7\}\}$

**Provenance Polynomials:**

name	
vorstellung.avi	$f_3$
rambo3.avi	$f_3 + f_7 \times d_3$