252-0834-00L *Information Systems for Engineers*

Final exam

ETH Zurich, Autumn Semester 2017
Friday 09.02.2018

First name: _____

Last name: _____

Legi number: _____

Signature: _____

You can fill out the above fields immediately, but **do not open the exam** before you are explicitly told so by the exam staff.

Please fill in your first and last name using only BLOCK CAPITAL LETTERS. It is your responsibility that your name and legi number are *clearly legible*.

- You have 90 minutes for the exam.
- There is no penalty for incorrect answers.
- Answers must be given in blue or black pen. Pencils, red pens, and green pens are not allowed.
- If you wish to change your answer, clearly strike out your previous solution, and write the new solution separately.

Problem	Points	Score
1	6	
2	3	
3	2	
4	2	
5	5	
6	4	
7	6	
8	24	
9	11	
10	7	
11	12	
12	9	
13	5	
14	2	
15	8	
Total:	106	

1. You are given the following relations:

```
Students(legi, first_name, last_name)
Results(legi, grade)
```

You can refer to them as S and R, respectively. For each of the SQL statements below, write down a corresponding relational algebra expression:

(a) (1)

```
1 SELECT legi
2 FROM Results
3 WHERE grade >= 5.00;
```

(b) (2)

```
1 SELECT grade, COUNT(*) AS grade_count
2 FROM Results
3 GROUP BY grade;
```

(c) (3)

```
1 SELECT first_name, last_name, grade
2 FROM ((SELECT * FROM Results
3         WHERE grade < 4.00)
4        INNER JOIN Students
5        ON Results.legi = Students.legi);
```

2. Mark all correct statements: (3)

All partial functions are also functions.	◯ True	◯ False
All functions are also partial functions.	◯ True	◯ False
All partial functions are also relations.	◯ True	◯ False
All relations are also partial functions.	◯ True	◯ False
All relations are also functions.	◯ True	◯ False
All functions are also relations.	◯ True	◯ False

3. If given a set of 4 partial functions with identical support that contains 2 attributes, in how many ways can we draw a corresponding table? (2)

4. For the following table of rivers, write down the corresponding set of partial functions, using the same syntax used in the lectures and exercise sessions. (2)

id	name	length
0	Tiber	406
1	Vistula	1047
2	Nile	6853

5. For each of the following descriptions of data fields, mark the most appropriate and efficient SQL data type for storing it that satisfies its precision requirements.

   (a) IATA airport codes (unique 3-character identifiers). (1)

   - ◯ `CHAR(3)`
   - ◯ `VARCHAR(3)`
   - ◯ `TINYTEXT`

   (b) Number of products sold of a certain type. (1)

   - ◯ `FLOAT`
   - ◯ `INT`
   - ◯ `NUMERIC(7,2)`

   (c) Your birthday. (1)

   - ◯ `TIME`
   - ◯ `DATE`
   - ◯ `DATETIME`

   (d) Prices of an online webshop. (1)

   - ◯ `FLOAT`
   - ◯ `INT`
   - ◯ `NUMERIC(7,2)`

   (e) When an order to an online shop was made. (1)

   - ◯ `TIME`
   - ◯ `DATE`
   - ◯ `DATETIME`

6. For the table of rivers presented in Question 4, write down a corresponding SQL `CREATE TABLE` statement using a fitting name and appropriate data types. "id" is the only key for the relation. No fields are allowed to be `NULL`. (4)

7. You are given the following mathematical definition of a relation. (6)

$$\begin{aligned}
&\{R \in \mathcal{P}(\mathbb{S} \nrightarrow \mathbb{V}) \mid \\
&(\forall t \in R,\ \text{support}(t) = \{\text{“country”}, \text{“product_line”}, \text{“product”}, \text{“margin”}\} \\
&\land t.\text{country} \in \mathbb{S} \land t.\text{product_line} \in \mathbb{S} \land t.\text{product} \in \mathbb{S} \land t.\text{margin} \in \mathbb{R} \cup \{\text{null}\}) \\
&\land (\forall t, s \in R,\ t.\text{country} = s.\text{country} \land t.\text{product_line} = s.\text{product_line} \\
&\land t.\text{product} = s.\text{product} \Rightarrow t = s)\}
\end{aligned}$$

Write down a corresponding SQL `CREATE TABLE` statement.

8. For this question we consider tables that satisfy the following schemas:

   `Customers(id:integer, name:string, age:integer)`
   `Products(id:integer, name:string, price:decimal)`
   `Orders(customer_id:integer, product_id:integer, time:datetime)`

   Primary keys are underlined. `customer_id` and `product_id` are foreign keys to the `Customers` and `Products` tables, respectively. For each of the following statements, write down (i.) a corresponding relational algebra expression and (ii.) a corresponding SQL query. Time constants should be on the form `09:12:34`, date constants should be on the form `2018-02-09`, and datetime constants should be on the form `2018-02-09T09:12:34`.

   (a) The name and price of all products cheaper than 20.00 CHF.

   i. Write down a relational algebra expression that returns the desired result. (2)

   ii. Write down an SQL query that returns the desired result. (2)

   (b) All orders placed before noon on New Year's day, 2017.

   i. Write down a relational algebra expression that returns the desired result. (2)

   ii. Write down an SQL query that returns the desired result. (2)

(c) For each product that has been ordered at least once, return its ID and how many times it has been ordered.

i. Write down a relational algebra expression that returns the desired result. (2)

ii. Write down an SQL query that returns the desired result. (2)

(d) All orders made by customers over the age of 60.

i. Write down a relational algebra expression that returns the desired result. (3)

ii. Write down an SQL query that returns the desired result. (3)

(e) For each customer who has made at least one order, return their name and the price of the most expensive product that they bought.

i. Write down a relational algebra expression that returns the desired result. (3)

ii. Write down an SQL query that returns the desired result. (3)

9. Consider a relation $R(A, B, C, D, E)$. Consider also its decomposition into the relations $R_1(A, B, C)$, $R_2(B, C, D)$ and $R_3(C, D, E)$. We are now investigating whether or not this decomposition is lossless, assuming various sets of functional dependencies.

   (a) Fill in the **starting** Chase algorithm tableau for the above relation and decomposition. This table is to be used as the starting point for each one of the parts (b) through (e). Assignments (b), (c), (d) and (e) are separate and independent from each other. (1)

A	**B**	**C**	**D**	**E**

   (b) Apply the Chase algorithm for the functional dependencies $B \to A$ and $A \to E$, **in this order**. Fill in the *updated* tableau for each step of the algorithm below each corresponding functional dependency. Mark whether the decomposition is lossless or lossy. (2)

$B \to A$

A	**B**	**C**	**D**	**E**

$A \to E$

A	**B**	**C**	**D**	**E**

   ◯ Lossless
   ◯ Lossy

   (c) Apply the Chase algorithm for the functional dependencies $D \to B$ and $C \to A$, **in this order**. Fill in the *updated* tableau for each step of the algorithm below each corresponding functional dependency. Mark whether the decomposition is lossless or lossy. (2)

$D \to B$

A	**B**	**C**	**D**	**E**

$C \to A$

A	**B**	**C**	**D**	**E**

   ◯ Lossless
   ◯ Lossy

(d) Apply the Chase algorithm for the functional dependencies $C \to E$, $D \to A$ and $B \to D$, **in this order**. Fill in the *updated* tableau for each step of the algorithm below each corresponding functional dependency. Mark whether the decomposition is lossless or lossy. (3)

$C \to E$

A	B	C	D	E

$D \to A$

A	B	C	D	E

$B \to D$

A	B	C	D	E

◯ Lossless
◯ Lossy

(e) Apply the Chase algorithm for the functional dependencies $CD \to A$, $D \to E$ and $A \to B$, **in this order**. Fill in the *updated* tableau for each step of the algorithm below each corresponding functional dependency. Mark whether the decomposition is lossless or lossy. (3)

$CD \to A$

A	B	C	D	E

$D \to E$

A	B	C	D	E

$A \to B$

A	B	C	D	E

◯ Lossless
◯ Lossy

10. Consider a relation $R(A, B, C, D, E)$. For each of the following sets of functional dependencies, compute the closure of the set of attributes on the left of each functional dependency, then mark whether the set of functional dependencies is in BCNF or not.

(a) (1)

$A \to E$ _____
$BD \to C$ _____

◯ In BCNF
◯ Not in BCNF

(b) (1)

$A \to B$ _____
$B \to C$ _____
$C \to D$ _____
$D \to A$ _____

◯ In BCNF
◯ Not in BCNF

(c) (1)

$CD \to AB$ _____
$AE \to CB$ _____

◯ In BCNF
◯ Not in BCNF

(d)

$A \to E$ _____

$E \to B$ _____

$B \to CD$ _____

$CD \to A$ _____

◯ In BCNF

◯ Not in BCNF

(2)

(e)

$AB \to E$ _____

$DE \to C$ _____

$C \to A$ _____

$B \to A$ _____

◯ In BCNF

◯ Not in BCNF

(2)

11. Consider a relation $R(A, B, C, D, E)$ with functional dependencies $BC \to D$ and $E \to A$. Find a decomposition of this relation that is in Boyce-Codd Normal Form (BCNF) by solving all of the parts below, using the algorithm covered in the book, lectures and exercise sessions.

(a) Using the given functional dependencies, show that the relation R is not in BCNF. (1)

_____

_____

(b) Based on your result from part (a), decompose R to two relations R_1 and R_2. State their attributes. (2)

R_1(_____)

R_2(_____)

(c) Show that one of the new relations from part (b) is in BCNF, while the other is not. You do not need to execute the functional dependency projection algorithm exhaustively. Still, you should show how the relevant functional dependencies are projected, rendering the new relations to be in BCNF or not. (3)

_____

_____

_____

_____

_____

_____

_____

(d) Decompose the relation from part (c) that is not in BCNF to R_3 and R_4. State their attributes. (2)

R_3(_____)

R_4(_____)

(e) Show that both R_3 and R_4 from part (d) are in BCNF. You do not need to execute the functional dependency projection algorithm exhaustively. Still, you should show how the relevant functional dependencies are projected, rendering the new relations to be in BCNF. (3)

(f) Based on your answers above, write down the final solution. (1)

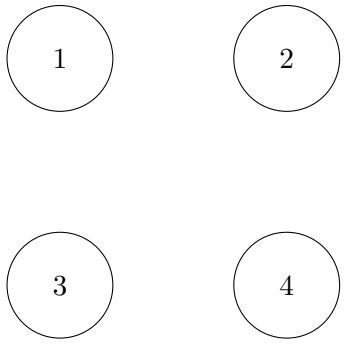
12. For this question we will determine if each of the schedules below are serializable. We use the following notation: $r_2(A)$ means that transaction 2 reads from location A; $w_1(B)$ means that transaction 1 writes to location B; etc.

(a) The table below has been populated with the following schedule: (1)

$$r_4(B), r_4(A), w_1(A), r_3(A), w_3(A), r_2(A), w_3(B), w_4(A)$$

Draw arrows between cells to mark each write-after-read, read-after-write and write-after-write dependency, then mark if the schedule is serializable or not. If you wish, you can use the diagram on the right to reach a conclusion.

1	**2**	**3**	**4**
			$r(B)$
			$r(A)$
$w(A)$			
		$r(A)$	
		$w(A)$	
	$r(A)$		
		$w(B)$	
			$w(A)$

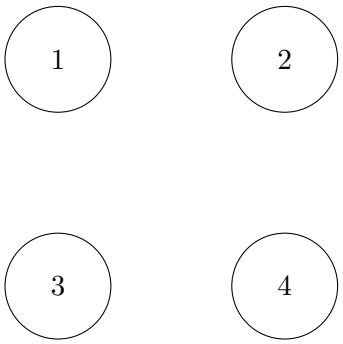


◯ Serializable. ◯ Not serializable.

(b) For the following schedule, fill in the table below, draw arrows to indicate sequential dependencies, then mark whether the schedule is serialiable or not: (2)

$$r_1(C), r_2(C), r_3(A), r_4(B), r_3(B), w_3(B), w_2(C), r_1(B), w_1(B), w_4(B)$$

1	2	3	4

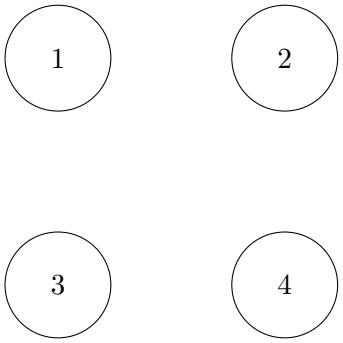


◯ Serializable. ◯ Not serializable.

(c) For the following schedule, fill in the table below, draw arrows to indicate sequential dependencies, then mark whether the schedule is serialiable or not: (2)

$$r_4(A), w_4(A), r_1(C), w_1(C), r_2(B), w_2(B), r_2(A), w_3(A), w_3(B), w_3(C)$$

1	2	3	4

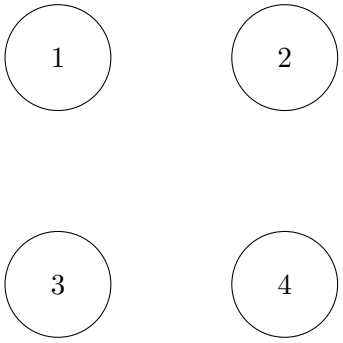


◯ Serializable. ◯ Not serializable.

(d) For the following schedule, fill in the table below, draw arrows to indicate sequential dependencies, then mark whether the schedule is serialiable or not: (2)

$$w_2(C), r_1(B), r_1(A), r_4(C), w_2(B), w_3(A), r_4(A), r_3(B)$$

1	**2**	**3**	**4**

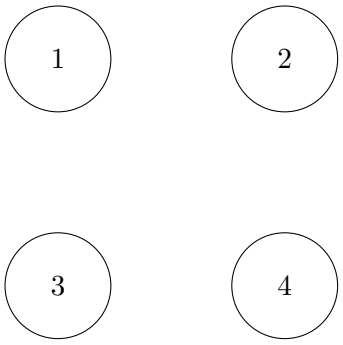


◯ Serializable. ◯ Not serializable.

(e) For the following schedule, fill in the table below, draw arrows to indicate sequential dependencies, then mark whether the schedule is serialiable or not: (2)

$$r_3(A), r_1(A), w_2(B), r_1(B), w_3(A), r_3(B), w_3(B), w_1(A)$$

1	**2**	**3**



◯ Serializable. ◯ Not serializable.

13. You are given the following fact table:

Country	Indicator type	Indicator name	Indicator value	Indicator year
Liechtenstein	Economic	GDP	4087	2005
Liechtenstein	Economic	GDP growth rate	4.8	2005
Liechtenstein	Economic	GDP per capita	117279.2	2005
Liechtenstein	Infrastructure	Mobile-cellular subscriptions	79.2	2005
Liechtenstein	Infrastructure	Individuals using the Internet	63.4	2005
Liechtenstein	Economic	GDP	5678	2010
Liechtenstein	Economic	GDP growth rate	7.4	2010
Liechtenstein	Economic	GDP per capita	156533.4	2010
Liechtenstein	Infrastructure	Mobile-cellular subscriptions	98.3	2010
Liechtenstein	Infrastructure	Individuals using the Internet	80.0	2010
Liechtenstein	Economic	GDP	6361	2017
Liechtenstein	Economic	GDP growth rate	1.4	2017
Liechtenstein	Economic	GDP per capita	169491.8	2017
Liechtenstein	Infrastructure	Mobile-cellular subscriptions	108.8	2017
Liechtenstein	Infrastructure	Individuals using the Internet	96.6	2017
Switzerland	Economic	GDP	407543	2005
Switzerland	Economic	GDP growth rate	3.0	2005
Switzerland	Economic	GDP per capita	55009.4	2005
Switzerland	Infrastructure	Mobile-cellular subscriptions	92.2	2005
Switzerland	Infrastructure	Individuals using the Internet	70.1	2005
Switzerland	Economic	GDP	581209	2010
Switzerland	Economic	GDP growth rate	3.0	2010
Switzerland	Economic	GDP per capita	74223.4	2010
Switzerland	Infrastructure	Mobile-cellular subscriptions	123.2	2010
Switzerland	Infrastructure	Individuals using the Internet	83.9	2010
Switzerland	Economic	GDP	670790	2017
Switzerland	Economic	GDP growth rate	0.8	2017
Switzerland	Economic	GDP per capita	80831.1	2017
Switzerland	Infrastructure	Mobile-cellular subscriptions	142.0	2017
Switzerland	Infrastructure	Individuals using the Internet	88.0	2017
USA	Economic	GDP	13093726	2005
USA	Economic	GDP growth rate	3.3	2005
USA	Economic	GDP per capita	44214.7	2005
USA	Infrastructure	Mobile-cellular subscriptions	68.3	2005
USA	Infrastructure	Individuals using the Internet	68.0	2005
USA	Economic	GDP	14964371	2010
USA	Economic	GDP growth rate	2.5	2010
USA	Economic	GDP per capita	48291.5	2010
USA	Infrastructure	Mobile-cellular subscriptions	91.3	2010
USA	Infrastructure	Individuals using the Internet	71.7	2010
USA	Economic	GDP	18036648	2017
USA	Economic	GDP growth rate	2.6	2017
USA	Economic	GDP per capita	56053.8	2017
USA	Infrastructure	Mobile-cellular subscriptions	117.6	2017
USA	Infrastructure	Individuals using the Internet	74.6	2017

Based on the fact table on the previous page, the following pivot table was created:

Country	**Indicator name**		
	GDP	**GDP growth rate**	**GDP per capita**
Liechtenstein	5678	7.4	156533.4
Switzerland	581209	3.0	74223.4
USA	14964371	2.5	48291.5
Total	15551258	12.9	140594.1

How was the above pivot table display built from the fact table? For each column in the fact table, state whether, in order to produce the above display, the column acts as: a slicer dimension; as a dicer dimension with an aggregation; as a dicer dimension without any aggregation; or whether it contains the fact values (measure). If it is a slicer dimension, you also need to specify which dimension value the data cube is sliced on.

(a) Country: (1)
◯ Measure ◯ Dicer without aggregation ◯ Dicer with aggregation ◯ Slicer on: _____

(b) Indicator type: (1)
◯ Measure ◯ Dicer without aggregation ◯ Dicer with aggregation ◯ Slicer on: _____

(c) Indicator name: (1)
◯ Measure ◯ Dicer without aggregation ◯ Dicer with aggregation ◯ Slicer on: _____

(d) Indicator value: (1)
◯ Measure ◯ Dicer without aggregation ◯ Dicer with aggregation ◯ Slicer on: _____

(e) Indicator year: (1)
◯ Measure ◯ Dicer without aggregation ◯ Dicer with aggregation ◯ Slicer on: _____

14. Assuming that you wrote an SQL query to produce the above pivot table, which of the following keywords could you have used for this purpose? Your query should produce **exactly** the values shown on the table. Check all that apply. (2)

☐ `ROLLUP`

☐ `CUBE`

☐ `GROUPING SETS`

15. Based on the fact table (we no longer consider the pivot table generated in Question 13), the following SQL query was written:

```
SELECT indicator_year, country, AVG(indicator_value)
FROM fact_table
WHERE indicator_name = "Individuals using the Internet" AND indicator_year > 2005
GROUP BY ROLLUP(indicator_year, country)
ORDER BY AVG(indicator_value) DESC
```

Answer the following questions:

(a) How many tuples would be returned if you executed the above query? (2)

(b) How many of the tuples returned by the query will include `None`/`NULL` values? (2)

(c) Which tuple would be returned first if you executed the above query? (Select one answer) (1)

- ◯ (2017, Liechtenstein, 96.60)
- ◯ (2010, Switzerland, 83.90)
- ◯ (None, Switzerland, 98.70)
- ◯ (None, None, 82.47)

(d) Which tuple would be returned last if you executed the above query? (Select one answer) (1)

- ◯ (2010, Liechtenstein, 80.00)
- ◯ (None, USA, 71.50)
- ◯ (2010, None, 78.53)
- ◯ (2010, USA, 71.70)

(e) Assuming that the keyword `ROLLUP` is substituted for the keyword `CUBE`, how many of the tuples returned by the updated query will include `None`/`NULL` values? (2)
