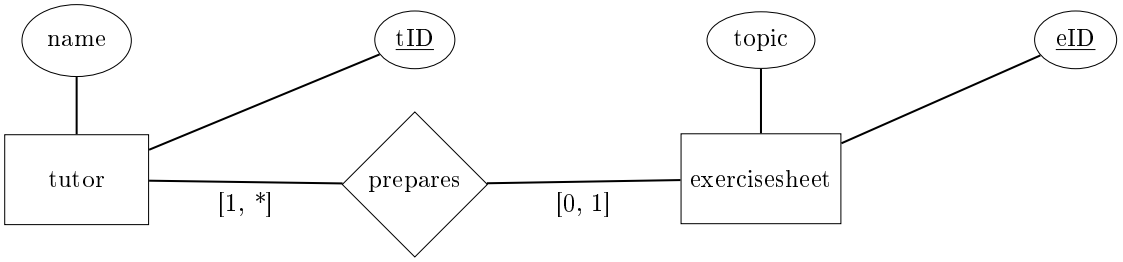
Database Systems – Autumn 2021

Please read the following instructions carefully:

- This exam has 7 exercises and 11 pages. Make sure your copy contains all exercises and all questions.
- Write your answers directly in the provided answer document (best directly on your computer, alternatively by printing and scanning) and upload your answer to Digital Exam. We recommend that you upload your answer as a spreadsheet file (e.g., .xlsx or .ods), but you can upload a PDF file if it is easier for you.
- Read the text of each exercise carefully before solving it.
- If you get stuck on an exercise, proceed with another one, and come back later.
- If your solution is based on an assumption that you have made, please add a comment for clarification.
- Aids are allowed during the exam, i.e., you are allowed to consult aids such as books, notes, etc.
- As a reference we recommend the course material endorsed by us (book, slides, exercises, etc.) as it complies with the notation we introduced in the course, which is the same one used in the exam.
- It is NOT allowed to seek help from or consult others (human beings, artificial intelligences, etc.) during the exam.
- The time limit is 75 minutes.
- The passing limit for the exam is 50 points.
- Good luck!

1 ER Modeling (15 points)

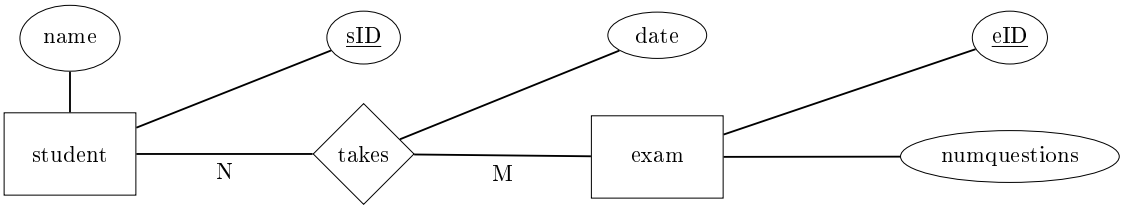
1.1. **(7.5 points)** Consider the following ER diagram:



Which of the following statements are true about the relationship between exercise sheets and tutors? *You can select one or multiple statements.*

(a) Each exercise sheet is prepared by at least one tutor
(b) Each tutor prepares at least one exercise sheet
(c) A tutor can prepare more than one exercise sheet
(d) An exercise sheet can be prepared by more than one tutor

1.2. **(7.5 points)** Consider the following ER diagram:



Which of the following relations are included in the result of mapping this ER diagram to a relational database schema? *You can select one or multiple relations.*

(a) `student(sID, name)`
(b) `takes(date)`
(c) `takes(sID → student, date)`
(d) `takes(eID → exam, date)`
(e) `takes(sID → student, eID → exam, date)`
(f) `takes(sID → student.sID, eID → exam.eID, name → student.name, date, numquestions → exam.numquestions)`
(g) `exam(eID, numquestions)`
(h) `exam(eID, numquestions, date)`

2 SQL (25 points)

Consider the following relations capturing information about students, study groups, tutors, and exercises:

student: {[sid: int, firstname: string, lastname: string, semester: int, birthdate: date]}

tutor: {[tid: int, firstname: string, lastname: string, issenior: boolean]}

studygroup: {[gid: int, tid → tutor, weekday: string, room: string, starttime: time]}

exercisesheet: {[eid: int, maxpoints: int, duedate: date]}

handsin: {[sid → student, eid → exercisesheet, achievedpoints: real]}

member: {[sid → student, gid → studygroup]}

where the primary keys are underlined and the foreign keys denoted by arrows.

We assume that tutors are responsible for one or multiple study groups, students individually hand in solutions for exercise sheets and receive individual grades in terms of the number of achieved points per sheet. Some of the tutors are more experienced (senior) than others.

For each of the following statements, **identify the missing information** that should go into the boxes to let the query retrieve the requested information.

2.1. **(10 points)** Find the ids, first names, and last names of all the students that handed in all the exercise sheets.

```
SELECT s.sid, s.firstname, s.lastname
FROM student s JOIN [ Box 1 ] ON s.sid = [ Box 2 ]
GROUP BY [ Box 3 ]
HAVING [ Box 4 ] = (SELECT COUNT(*) FROM exercisesheet);
```

Box 1:

Box 2:

Box 3:

Box 4:

2.2. **(15 points)** For each exercise sheet handed in by at least one student, compute the passing rate, i.e., the percentage of handins that achieved at least 50% of the total points

```
WITH numpassing AS (
  SELECT e.eid, [ Box 1 ] AS numpass
  FROM exercisesheet e LEFT OUTER JOIN [ Box 2 ]
    ON e.eid = [ Box 3 ] AND [ Box 4 ] >= e.maxpoints * 0.5
  GROUP BY e.eid
)
SELECT e.eid, ((p.numpass * 100)/ COUNT(*)) AS passingRate
FROM numpassing p NATURAL JOIN exercisesheet e NATURAL JOIN [ Box 5 ]
GROUP BY [ Box 6 ]
```

Box 1:

Box 2:

Box 3:

Box 4:

Box 5:

Box 6:

As a reminder, in this question you should consider the following use case about students, study groups, tutors, and exercises:

student: {[sid: int, firstname: string, lastname: string, semester: int, birthdate: date]}

tutor: {[tid: int, firstname: string, lastname: string, issenior: boolean]}

studygroup: {[gid: int, tid → tutor, weekday: string, room: string, starttime: time]}

exercisesheet: {[eid: int, maxpoints: int, duedate: date]}

handsin: {[sid → student, eid → exercisesheet, achievedpoints: real]}

member: {[sid → student, gid → studygroup]}

3 Relational Algebra (10 points)

Consider the relational schema from Exercise 2 again, which is:

student: {[sid: int, firstname: string, lastname: string, semester: int, birthdate: date]}
tutor: {[tid: int, firstname: string, lastname: string, issenior: boolean]}
studygroup: {[gid: int, tid → tutor, weekday: string, room: string, starttime: time]}
exercisesheet: {[eid: int, maxpoints: int, duedate: date]}
handsin: {[sid → student, eid → exercisesheet, achievedpoints: real]}
member: {[sid → student, gid → studygroup]}

where the primary keys are underlined and the foreign keys denoted by arrows.

We assume that tutors are responsible for one or multiple study groups, students individually hand in solutions for exercise sheets and receive individual grades in terms of the number of achieved points per sheet. Some of the tutors are more experienced (senior) than others.

Fill in the missing information that should go into the boxes to make the query compute the requested information.
Hint: If the correct solution requires a box to remain empty, please write "EMPTY".
Hint given during the exam: *You can assume that projections can contain aggregate functions.*

3.1. **(10 points)** In **how many exercise sheets** did the student with **sid 3** achieve **at least half of the maximum number of points of that exercise sheet**?

$\pi_{\boxed{\text{Box 1}}}(\sigma_{\boxed{\text{Box 2}}}(student \bowtie_{\boxed{\text{Box 3}}} \boxed{\text{Box 4}} \bowtie_{\boxed{\text{Box 5}}} \boxed{\text{Box 6}}))$

Box 1:

Box 2:

Box 3:

Box 4:

Box 5:

Box 6:

4 Design Theory (20 points)

Functional Dependencies (7 points)

4.1. Given the following table

mid	title	year	country	genre	actor	director
8371	A Night to Remember	1942	USA	Comedy	Sidney Toler	Randall Wallace
8371	A Night to Remember	1942	USA	Mystery	Sidney Toler	Randall Wallace
8371	A Night to Remember	1942	USA	Romance	Sidney Toler	Randall Wallace
16676	Dark Alibi	1946	USA	Crime	Sidney Toler	Phil Karlson
16676	Dark Alibi	1946	USA	Drama	Sidney Toler	Phil Karlson
16676	Dark Alibi	1946	USA	Mystery	Sidney Toler	Phil Karlson
16676	Dark Alibi	1946	USA	Thriller	Sidney Toler	Phil Karlson
15696	Coney Island	1943	USA	Comedy	Yakima Canutt	Tracy Casper Lang
15696	Coney Island	1943	USA	Musical	Yakima Canutt	Tracy Casper Lang

Identify which of the following **functional dependencies** are satisfied by this table. *You can select one or multiple functional dependencies.*

(a) year, actor $\rightarrow$ title

(b) title $\rightarrow$ year, actor

(c) actor $\rightarrow$ title

(d) title $\rightarrow$ actor

Keys (13 points)

For the next two questions, consider the following relational schema

$$\mathcal{R} = \{[A, B, C, D, E]\}$$

and the following functional dependencies

$\{A\} \rightarrow \{B, D\}$
$\{B\} \rightarrow \{C, E\}$
$\{C\} \rightarrow \{B\}$
$\{E\} \rightarrow \{D\}$

4.2. **(6.5 points)** Which of the following are **super keys**? *You can select one or multiple set(s) of attributes.*

(a) { A }
(b) { B }
(c) { C }
(d) { E }
(e) { A, B }
(f) { C, E }
(g) { A, C }
(h) { B, E }
(i) { A, E }
(j) { B, C }

4.3. **(6.5 points)** Which of the following are **candidate keys**? *You can select one or multiple set(s) of attributes.*

(a) { A }
(b) { B }
(c) { C }
(d) { E }
(e) { A, B }
(f) { C, E }
(g) { A, C }
(h) { B, E }
(i) { A, E }
(j) { B, C }

5 Physical Design (5 points)

(5 points) A B+-tree with fanout=10 and three levels (root level, one inner level, leaf level) has at most ? pointers to file records. *Select the option that should go in ? to correctly complete this statement*

(a) 99

(b) 100

(c) 900

(d) 990

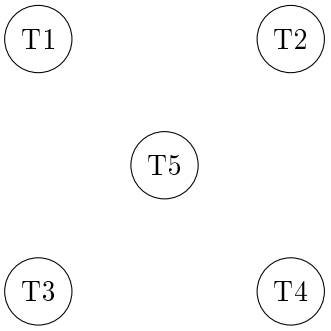
(e) 1000

6 Transactions (15 points)

6.1. **(10 points)** Consider five transactions, T1...T5, that are concurrently running on a database, and nine data objects, a ... i. The following table shows the data objects that are currently locked by the transactions as well as which releases of the locks the transactions are waiting for.

transaction	holds lock on	waits for release of lock on
T1	a	b
T2	b, i	e, c
T3	c, g, h	a
T4	d, f	g, c
T5	e	i

Please indicate which of the following edges are part of the **wait-for graph** (you can select one or multiple edge(s))



(a) (T1, T2)
(b) (T1, T3)
(c) (T1, T4)
(d) (T1, T5)
(e) (T2, T1)
(f) (T2, T3)
(g) (T2, T4)
(h) (T2, T5)
(i) (T3, T1)
(j) (T3, T2)
(k) (T3, T4)
(l) (T3, T5)
(m) (T4, T1)
(n) (T4, T2)
(o) (T4, T3)
(p) (T4, T5)
(q) (T5, T1)
(r) (T5, T2)
(s) (T5, T3)
(t) (T5, T4)

6.2. **(5 points)** Is there a deadlock? If your answer is Yes, list all the transactions involved in the deadlock(s)

7 Query Optimization (10 points)

7.1. **(2.5 points)** Consider the following SQL statement:

```
1  SELECT m.title
2  FROM movie m
3    JOIN involved i ON i.movieid = m.id
4    JOIN person p ON p.id = i.personid
5    JOIN ratings r ON m.id = r.movieid
6  WHERE p.name = 'John Travolta' AND i.role = 'actor'
7  GROUP BY m.id, m.title
8  ORDER BY AVG(r.rating) DESC;
```

Which clause is evaluated **second**?

(a) The SELECT clause (line 1)
(b) The FROM clause (lines 2-5)
(c) The WHERE clause (line 6)
(d) The GROUP BY clause (line 7)
(e) The ORDER BY clause (line 8)

7.2. **(2.5 points)** Consider the same SQL statement given in Question 7.1.:

```
1  SELECT m.title
2  FROM movie m
3    JOIN involved i ON i.movieid = m.id
4    JOIN person p ON p.id = i.personid
5    JOIN ratings r ON m.id = r.movieid
6  WHERE p.name = 'John Travolta' AND i.role = 'actor'
7  GROUP BY m.id, m.title
8  ORDER BY AVG(r.rating) DESC;
```

Which clause is evaluated **last**?

(a) The SELECT clause (line 1)
(b) The FROM clause (lines 2-5)
(c) The WHERE clause (line 6)
(d) The GROUP BY clause (line 7)
(e) The ORDER BY clause (line 8)

7.3. **(5 points)** What are the main heuristics that logical query optimization follows? *You can select one or multiple statements.*

(a) Execute joins with input from selections after other joins

(b) Execute joins with input from selections before other joins

(c) Introduce joins by combining selections and cross products

(d) Introduce selections and cross products to replace joins

(e) Push down joins

(f) Push down selections