Database Systems (CPH) – Fall 2023
Lecturer: Yongluan Zhou
Exam Date: Jan 5, 2024

AALBORG UNIVERSITY
DENMARK

Preamble

Please read the following instructions carefully:

- This exam has eight questions and thirteen pages. Make sure your copy contains all exercises and all questions. The following table summarizes the questions in this exam and their weights.

Question	Points
1. ER Modeling	16
2. SQL	20
3. Relational Algebra	10
4. Design Theory	12
5. Physical Design	8
6. Query Processing and Optimization	12
7. Transactions and Concurrency Control	10
8. Crash Recovery	12

- 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 can consult aids such as books, notes, etc.
- As a reference, we recommend the course material we endorsed (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 intelligence, etc.) during the exam.
- The passing limit for the exam is **50** points.
- Good luck!

1. ER Modeling (16 points)

1.1 (2 points) Which data modeling process is the ER model used in? Choose one or more options:

(a) Requirement collection and analysis

(b) Conceptual design

(c) Data model mapping

(d) Physical design

(e) None of the above

1.2 (2 points) Which of the following statements about weak entity types and strong entity types are true? Choose one or more options:

(a) A weak entity has a primary key.

(b) Double boxes represent weak entity types.

(c) Double diamonds represent the identifying relationship for strong entity types.

(d) Weak entity type only has a partial key.

(e) None of the above

1.3 (6 points) Consider the relationship *supervise* where each student is supervised by one professor while each professor can supervise many students. Which of the following ER diagrams is correct? Choose one or more options:

(a)

(b)

(c)

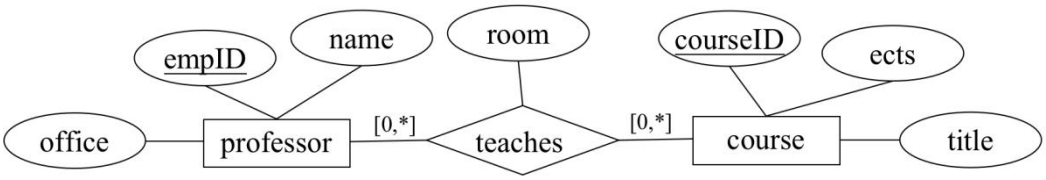
(d)

(e)

(f)

(g) None of the above

1.4 (6 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? Choose one or more options:

(a) office(professor)

(b) professor(empID, name, office)

(c) professor(empID, name, office, teaches)

(d) teaches(room)

(e) teaches(room, professor, course)

(f) teaches(empID → professor, courseID → course, room)

(g) teaches(empID → professor, courseID → course, room)

(h) course(courseID, ects, title)

(i) course(courseID, ects, title, teaches)

(j) title(course)

(k) None of the above

AALBORG UNIVERSITY
DENMARK

2. SQL (20 points)

Consider the following relational schema. Note that this does not describe the situation at Aalborg University.

We assume that tutors are responsible for one or multiple study groups, students individually (not per group) hand in solutions for exercise sheets and receive individual grades in terms of the number of points achieved per sheet. Some of the tutors are more experienced (senior) than others.

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]}
handsin: {[sid → student, eid → exercisesheet, achievedpoints: int]}
member: {[sid → student, gid → studygroup]}

2.1 (4 points) Which of the following queries can find the number of tutors whose first name is “Helle”? Choose one or more options:

(a) SELECT * FROM tutor WHERE tutor.firstname is “Helle”;

(b) SELECT COUNT(*) FROM tutor WHERE tutor.firstname = Helle;

(c) SELECT COUNT(tid) FROM tutor WHERE tutor.firstname = ‘Helle’;

(d) SELECT NUMBER(*) FROM tutor t WHERE t.firstname = ‘Helle’;

(e) SELECT NUMBER(tid) FROM tutor t WHERE t.firstname = Helle;

(f) SELECT COUNT(*) FROM tutor t WHERE t.firstname = ‘Helle’;

(g) None of the above

2.2 (4 points) Which of the following queries can find IDs of studygroups supervised by a senior tutor. Choose one or more options:

(a)
```
SELECT gid
    FROM studygroup sg
            JOIN tutor t ON sg.tid = t.tid
    WHERE t.issenior is true;
```

(b)
```
SELECT gid
    FROM studygroup sg, tutor t
    WHERE sg.tid = t.tid AND t.issenior;
```

(c)
```
SELECT gid
    FROM tutor t
            JOIN studygroup sg ON t.tid = sg.tid
    WHERE t.issenior = true;
```

(d) SELECT gid
FROM studygroup sg JOIN tutor t
WHERE t.issenior;

(e) SELECT gid
FROM studygroup sg NATURAL JOIN tutor t
WHERE t.issenior;

(f) SELECT gid
FROM studygroup sg, tutor t
WHERE sg.tid = t.tid, t.issenior;

(g) None of the above

2.3 (4 points) Consider the following query:

SELECT * FROM handsin NATURAL JOIN exercisesheet;

What is the set of columns that compose the resulting table? Choose only one option:

(a) sid, eid, achievedpoints, int

(b) sid, eid, achievedpoints, eid, maxpoints

(c) sid, eid, achievedpoints, maxpoints

(d) eid, maxpoints, sid, eid, achievedpoints

(e) None of the above

2.3 (8 points) Identify the missing information that should go into the boxes of the following query: to find the IDs of all students who achieved more than 7 points and handed in at least 3 exercise sheets.

SELECT sid
FROM Box 1
WHERE Box 2
GROUP BY Box 3
HAVING Box 4;

AALBORG UNIVERSITY
DENMARK

3. Relational Algebra (10 points)

Consider the employee database.

Employee (empID, person_name, street, city)
Works (empID, comID, salary)
Company (comID, company_name, city)

Fill in the missing information that should go into the boxes to make the query compute the required information.

To enter the operation symbols and conditions into the answer sheet, please use the following replacements:

Operation symbol	What to write in the answer sheet
ρ	rho
γ	gamma
Π	pi
σ	sigma
⋈	join
×	x

3.1 (6 points) Find the ID and name of each employee who lives in the city where their company is located.

$\Pi_{empID,\ person\_name}$(Box 1 $_{Employee.city\ =\ Company.city}$ (Employee Box 2 Works Box 3 Company))

3.2 (4 points) Find the ID and name of each employee with the lowest salary in the database.

$\Pi_{empID,\ person\_name}$ (Employee) −
Π Box 4 (ρ_A (Employee ⋈ Works) Box 5 ρ_B (Employee ⋈ Works))

AALBORG UNIVERSITY
DENMARK

4. Design Theory (12 points)

Given a relational schema $R = (A, B, C, D, E)$ and the following set F of functional dependencies (FDs):

$$A \rightarrow C$$
$$BC \rightarrow D$$
$$C \rightarrow E$$
$$E \rightarrow B$$

4.1 (4 points) Which of the following attributes (sets) can be the **candidate key** of R? Choose one or multiple options:

(a) A
(b) B
(c) C
(d) BC
(e) D
(f) AD
(g) E
(h) AC
(i) CE
(j) ABCDE
(k) none of the above

4.2 (4 points) Which of the attribute set(s) in Question 4.1 can be the **superkey** of R? Choose one or more options from Question 4.1.

4.3 (4 points) Which of the following statements are correct? Choose one or more options:

(a) The relation R is in BCNF and 3NF.
(b) The relation R is in BCNF but not in 3NF.
(c) The relation R is in 3NF but not in BCNF.
(d) The relation R is not in BCNF or 3NF.
(e) none of the above

5. Physical Design (8 points)

Consider the following B+tree with degree $d = 4$.

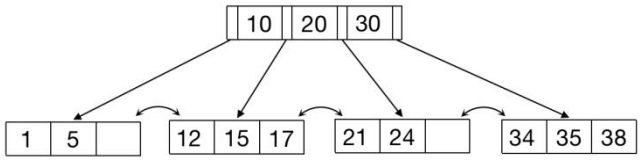


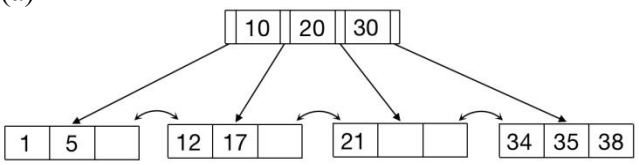
Figure 5: B+ Tree with $d = 4$

Please make the following assumptions:

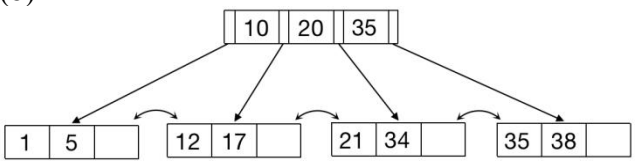
- The left pointer of a key k in an internal node leads towards keys less than k, while the right pointer leads towards keys greater than or equal to k.
- A leaf node underflows when the ***number of keys*** goes below $\lceil \frac{d-1}{2} \rceil$.
- An internal node underflows when the ***number of pointers*** goes below $\lceil \frac{d}{2} \rceil$.

5.1 (4 points) Delete 15^*, then delete 24^*. What is the resulting tree? Choose one or multiple options:

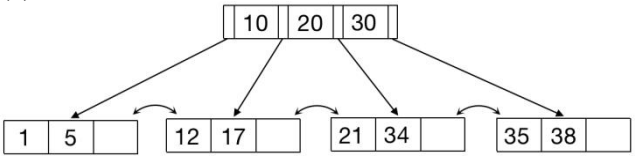
(a)



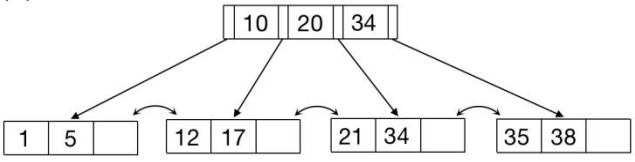
(b)



(c)



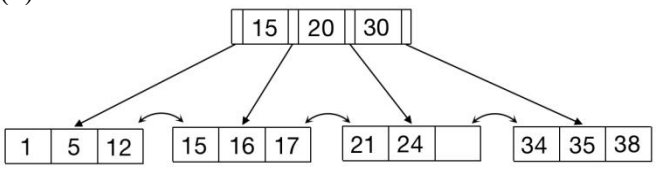
(d)



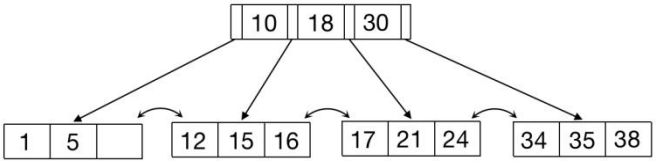
(e) None of the above

5.2 (4 points) Consider the original tree in Figure 5 (ignore the operation in Question 5.1). Insert 16^* into the tree. What is the resulting tree? Choose one or multiple options:

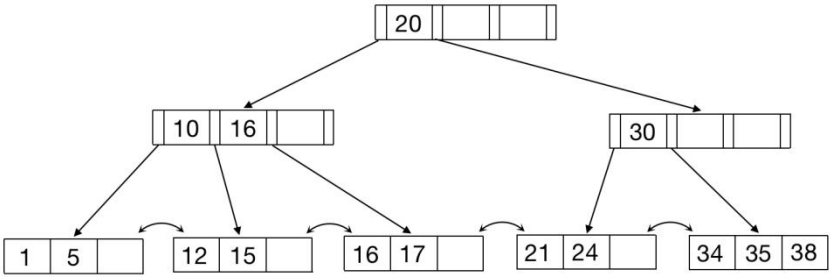
(a)



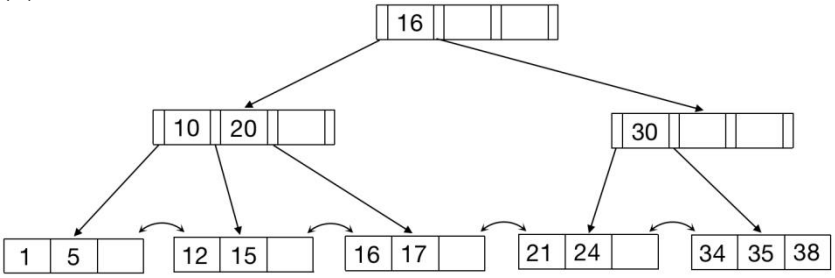
(b)



(c)



(d)



(e) None of the above

AALBORG UNIVERSITY
DENMARK

6. Query Processing and Optimization (12 points)

Consider a database containing a relation R with 1,100 pages, and a relation S with 10,000 pages, and a computer with 200 buffer pages.

Answer the following questions:

6.1 (2 points) Suppose you need to sort R using the multiway external merge sort algorithm. How many runs will be generated by Phase 1 of the algorithm? Choose only one option:
(a) 2
(b) 6
(c) 100
(d) 200
(e) 1,100
(f) None of the above

6.2 (3 points) Following Question 6.1, how many passes in total (including both Phase 1 and Phase 2) are needed to sort R? Choose only one option:
(a) 1
(b) 2
(c) 3
(d) 4
(e) None of the above

6.3 (3 points) Suppose you need to join R with S using the *most optimized* Sort-Merge Join algorithm. What is the number of I/O incurred by the join operation if we *do not* count the I/O of writing the final join results to the disk? Choose only one option:
(a) 11,100
(b) 33,300
(c) 55,500
(d) 77,700
(e) None of the above

6.4 (4 points) Continuing Question 6.3, suppose you can control the number of buffer pages allocated to the join algorithm. Which of the following number(s) of buffer pages would make Block Nested-Loop Join outperforming Sort-Merge Join regarding I/O cost? Choose one or more options:
(a) 250 pages
(b) 350 pages
(c) 450 pages
(d) 550 pages

7. Transactions and Concurrency Control (12 points)

7.1 (2 points) Which transaction properties are addressed by concurrency control protocols? Choose one or multiple options:
(a) Atomicity
(b) Consistency
(c) Isolation
(d) Durability
(e) None of the above

7.2 (2 points) Which transaction properties are addressed by crash recovery mechanisms? Choose one or multiple options:
(a) Atomicity
(b) Consistency
(c) Isolation
(d) Durability
(e) None of the above

7.3 We consider three transactions T1, T2, and T3 consisting of the following actions:
`T1: R(X) W(X) R(Y) W(Y) C` `T2: R(Y) W(Y) W(Z) C` `T3: R(X) R(Z) W(Z) C`
R(X) represents reading object X, W (X) represents writing object Y, and C represents Commit. Consider the following two schedules:

Schedule (1):

```
T1: R(X) W(X)                R(Y) W(Y) C
T2:                R(Y) W(Y)                          W(Z) C
T3:          R(X)                          R(Z) W(Z) C
```

Schedule (2):

```
T1:      R(X) W(X)              R(Y) W(Y) C
T2:                                          R(Y) W(Y) W(Z) C
T3: R(X)           R(Z) W(Z) C
```

Answer the following questions:

7.3.1 (3 points) Which type(s) of schedule does Schedule (1) belong to? Choose one or more options:
(a) Serial schedule
(b) Serializable schedule
(c) Conflict serializable schedule
(d) None of the above

7.3.2 (3 points) Which type(s) of schedule does Schedule (2) belong to? Choose one or more options:
(a) Serial schedule
(b) Serializable schedule
(c) Conflict serializable schedule
(d) None of the above

8. ~~Crash Recovery (12 points)~~