

מבחן במסדי נתונים, מועד א', תשע"ה
מס' הקורס 67506

מרצים: פרופ' שרה כהן
פרופ' יהושע שגיב

תאריך הבחינה: י"ח בשבט, תשע"ד (6.1.2015)

אסור לכתוב בשוליים!

הנחיות כלליות

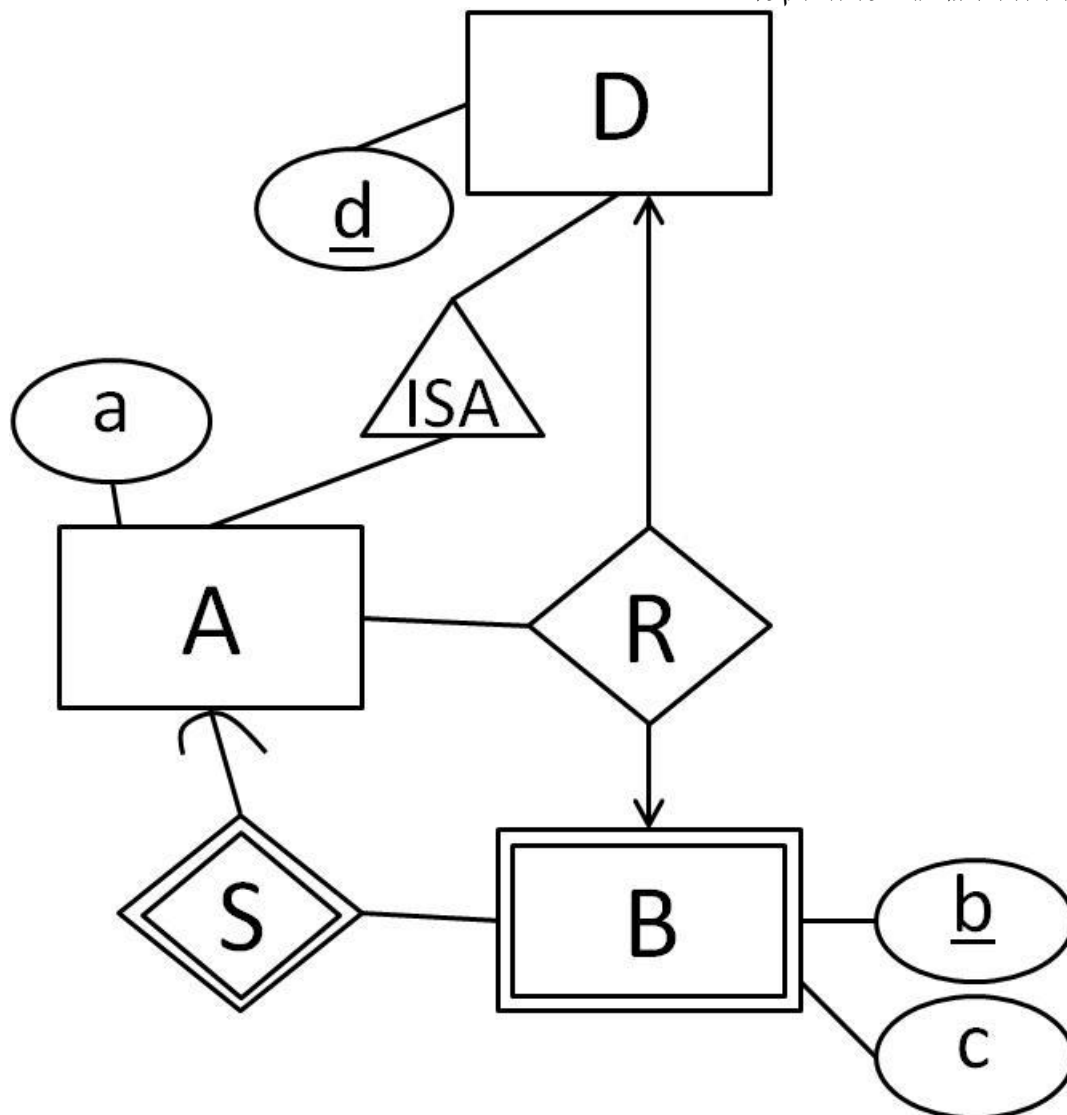
- משך הבחינה: 3 שעות. חומר סגור
- ענו על כל השאלות (בבחינה יש 14 שאלות)
- סמנו תשובה אחת בלבד כשצריך לבחור בין כמה אפשרויות, אלא אם נאמר אחרת
- רשמו מספר מזהה בראש כל דף
- אל תבזבזו זמן על מתן תשובות לא לעניין! ענו רק על מה שנשאלתם
- אפשר להשתמש במחשבוני פשוטים (ללא זיכרון לאגירת נוסחאות או מידע כלשהו)

שאלה	נקודות
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
סה"כ	

בהצלחה!

שאלה 1 (6 נקודות)

נתונה דיאגרמת ישויות וקשרים:



החץ לתוך A הוא מעוגל

תרגמו את הדיאגרמה ליחסים. בכל יחס יש לציין את שם היחס, שמות השדות וכן את התלויות הפונקציונליות. הלא טריויאליות שמתקיימות ביחס. (אין צורך לרשום תלויות הנובעות מתלויות אחרות שכבר רשמתם.)

D(d)
 A(a,d) $d \rightarrow a$
 B(b,c,d) $d, b \rightarrow c$
 R(dd, da, db, b) $dd, da \rightarrow db, b$ $db, b, da \rightarrow dd$

שאלה 2 (8 נקודות)

בכל אחד מהסעיפים הבאים, הוכיחו את הטענה הנתונה או תנו דוגמה נגדית פשוטה.

א. טענה: יהי R יחס ותהי $F = \{X \rightarrow A\}$ קבוצה המכילה תלות פונקציונלית בודדת. אזי, מתקיים ש- R ב-BCNF אם ורק אם R ב-3NF.

נכון. אם R ב-BCNF אז בוודאי (לפי ההגדרה) R ב-3NF. נניח ש- R ב-3NF ונראה ש- R ב-BCNF.

מקרה א': התלות $X \rightarrow A$ טריויאלית. אז בוודאי ש- R ב-BCNF לפי ההגדרה

מקרה ב': התלות $X \rightarrow A$ אינה טריויאלית. קל לראות שאם כך המפתח היחיד של R הוא $R - \{A\}$. מכיוון ש- R ב-3NF, התלות אינה טריויאלית ו- A אינו שדה במפתח, ניתן להסיק ש $X = R - \{A\}$, כלומר X מפתח על. לכן, לפי ההגדרה R ב-BCNF.

ב. טענה: יהי R יחס ותהי F קבוצה של k תלויות פונקציונליות. ליחס R יש לכל היותר $k+1$ מפתחות.

לא נכון. דוגמה נגדית: $R = (A, B, C, D, E, G)$
 F מכיל את התלויות:

- $A \rightarrow B$ •
- $B \rightarrow A$ •
- $C \rightarrow D$ •
- $D \rightarrow C$ •
- $E \rightarrow G$ •
- $G \rightarrow E$ •

אזי ל- R יש 8 מפתחות: $BDG, BDE, BCG, BCE, ADG, ADE, ACG, ACE$

שאלה 3 (8 נקודות)

נתונה סכמה $R=(A,B,C,D,E,G)$ ו- $F=\{C \rightarrow DE, D \rightarrow G, C \rightarrow G, E \rightarrow A, G \rightarrow C, CG \rightarrow A\}$.
כמו כן, נתון פירוק של R לתתי סכמות $R_1=(BCDE)$ $R_2=(ADG)$.

א. ציינו לכל אחד מהתלויות הבאות האם היא נשמרת על ידי הפירוק:

לא	כן	האם $C \rightarrow G$ נשמרת?
לא	כן	האם $E \rightarrow A$ נשמרת?
לא	כן	האם $G \rightarrow C$ נשמרת?
לא	כן	האם $CG \rightarrow A$ נשמרת?
לא	כן	ב. האם R_1 ב-3NF?
לא	כן	האם R_1 ב-BCNF?
לא	כן	ג. האם R_2 ב-3NF?
לא	כן	האם R_2 ב-BCNF?

שאלה 4 (6 נקודות)

מבצעים את סדרת הפעולות הבאה:

```
create table R(a integer, b integer);
create view V as select a+b as c from R where a < b+2;
insert into R values (1,1);
insert into R values (1,2);
insert into R values (1,3);
insert into R values (2,1);
insert into R values (3,1);

select a, count(*), sum(distinct c)
from R, V
where b < c and a != (select max(a) from R)
group by a;
```

אלו שורות יוחזרו על ידי השאילתה?

(2,4,9), (1,8,9)

שאלה 5 (8 נקודות)

נתונה סכמה $R=(A,B,C,D,E,G)$ וקבוצת תלויות פונקציונליות F :

$$BG \rightarrow CD, G \rightarrow A, CD \rightarrow GE, C \rightarrow AG, A \rightarrow D$$

א. מצאו פירוק של R ל-3NF שהוא ללא אובדן ומשמר תלויות.

BGC
GA
CE
AD

ב. האם כל תתי הסכמות של הפירוק שמצאת בסעיף ב הם ב-BCNF?

- אם כן, נמק מדוע זה כך.
- אם לא, מצא פירוק של R ל-BCNF שהוא ללא אובדן. האם הפירוק החדש משמר תלויות? נמק.

הפירוק אינו ב-BCNF כי BGC אינו ב-BCNF. התלות $C \rightarrow G$ סותרת את תנאי ה-BCNF.

פירוק ל-BCNF: BC, GA, CE, AD, CG

הפירוק אינו משמר את התלות $BG \rightarrow CD$

שאלה 6 (8 נקודות)

בבסיס נתונים קיימים היחסים הבאים:

Person(pId, age)

Friends(pId1,pId2)

ניתן להניח שהטבלאות אינן מכילות ערכי NULL, ושכל ערך ב-pId1 וב-pId2 מופיע גם בטבלת Person. כמו כן, ניתן להניח שיחס החברות הוא סימטרי, כלומר אם (p1,p2) שורה ביחס Friends אזי גם (p2, p1) שורה ביחס Friends.

בכל אחד משני הסעיפים, כיתבו את השאילתה ב-SQL (שימו לב שסעיף ב נמצא בעמוד הבא).

א. החזירו זוגות של אנשים שונים שיש להם את מספר החברים המשותפים הגדול ביותר. (שימו לב שאם יש יותר מזוג אחד עם מספר מקסימלי של חברים משותפים, אז יש להחזיר את כל הזוגות האלה).

```
Select F1.pid1, F2.pid1
From Friends F1, Friends F2
Where F1.pid1 <> F2.pid1 and F1.pid2 = F2.pid2
Group by F1.pid1, F2.pid1
Having Count(*) >= ALL(
    Select count(*)
    From Friends F3, Friends F4
    Where F3.pid1 <> F4.pid1 and F3.pid2 = F4.pid2
    Group by F3.pid1, F4.pid1)
```

ב. החזירו את האנשים שחברים עם כל האנשים שצעירים מהם.

```
Select pid
From Person P1
Where not exists (
    Select pid
    From Person P2
    Where P2.age < P1.age
Except
    Select pid2
    From Friends F
    Where F.pid1 = P1.pid)
```

שאלה 7 (6 נקודות)

מבוצע קטע הקוד הבא:

```
create table R(a integer, b integer);

create or replace function fun() returns trigger as $$
begin
    if (new.a < new.b) then
        insert into R values(new.a+4, new.b);
        return old;
    end if;
    if (new.a > new.b + 2) then
        insert into R values(new.a-2, new.b);
        return new;
    end if;
    new.a := new.a * 2;
    new.b := new.b * 2;
    return new;
end
$$ language plpgsql;

create trigger t1
before insert on R
for each row
execute procedure fun();

insert into R values(6,1);

insert into R values(1,2);
```

מהו תוכן הטבלה R לאחר ביצוע פעולת ההכנסה של (6,1)?

(6,1) (4,1) (4,2)

אלו שורות יתווספו ל-R לאחר ביצוע פעולת ההכנסה של (1,2)?

(5,2) (6,4)

שאלה 8 (8 נקודות)

בשאלה זו מניחים שכל הביטויים מוגדרים היטב.

- א. האם $(R \cap S) \bowtie (T \cap S)$ ו- $R \bowtie S \bowtie T$ הינם שקולים? כן לא
- ב. האם $\pi_A(\sigma_{B \neq 10}(R))$ ו- $\pi_A(R) - \pi_A(\sigma_{B=10}(R))$ הינם שקולים? כן לא

שאלה 9 (8 נקודות)

צריך לחשב את הביטוי $R(A,B) \bowtie S(B,C) \bowtie T(C,D)$. בבילוק יש 4,000 בתים וכל אטריבוט הוא שדה באורך של 20 בתים. $B(R)=B(S)=100$ ו- $B(T)=200$. כמו כן, $V(R,B)=100$, $V(S,B)=50$, $V(S,C)=1,000$ ו- $V(T,C)=10,000$.

א. כמה רשומות יש בתוצאה?

בבילוק יש 100 רשומות של שני אטריבוטים. מספר הרשומות בתוצאה הוא:

$$10,000 * 10,000 * 20,000 / (100 * 10,000) = 2,000,000$$

ב. כמה בלוקים יש בתוצאה?

בבילוק של התוצאה יש 50 רשומות, לכן גודל התוצאה 40,000 בלוקים

ג. תארו בקיצור איך מבוצע החישוב היעיל ביותר של הביטוי וציינו את מחירו, בהנחה שגודל החוצץ הוא 120 בלוקים.

קוראים את כל R לזיכרון ומחלקים את S ו- T לדליים לפי הערכים של C . בהמשך עושים צירוף

של כל זוג דליים תואמים של S ושל T ותוך כדי כך עושים גם את הצירוף עם R שכולו בזיכרון.

$$B(R) + 3B(S) + 3B(T) = 1,000$$

גודל החוצץ הנדרש הוא $B(R)$ ועוד השורש של $B(S)$ ועוד שניים, בסה"כ 112.

ד. מהו הגודל המינימאלי של החוצץ שנדרש כדי לבצע את החישוב שתיארתם בסעיף ג? (זכרו שגודל החוצץ הוא 120 בלוקים, אבל האם אפשר להגיע למחיר הטוב ביותר עם פחות מזה?)

שאלה 10 (6 נקודות)

סמנו את הסעיף הנכון.

- א. פעולת בחירה לפי תנאי $A=10$ תמיד עדיף לבצע על ידי שימוש באינדקס (אם יש כזה) ולא על ידי סריקת כל הקובץ ללא שימוש באינדקס.
- ב. בעץ $B+$ הגודל של צומת הוא בלוק אחד מכיוון שבתהליך חיפוש ממילא קוראים רק צומת אחד בלבד מכל רמה פנימית בעץ. הערה: כל רמה מלבד העלים היא פנימית.
- ג. כאשר עץ $B+$ הינו אינדקס על אטריבוט A , אז הרמה התחתונה שלו היא אינדקס דליל רק אם הקובץ ממוין על A .
- ד. אם יחס ממוין על אטריבוט A , אז אינדקס על A הוא מיותר, כי תמיד אפשר לעשות חיפוש בינארי על A .
- ה. אף אחד מהסעיפים הקודמים אינו נכון.

שאלה 11 (8 נקודות)

צריך לחשב הביטוי $\sigma_{A=7 \wedge B=3}(R(A,B) \bowtie S(B,C))$. נתון שהערך של אטריבוט הוא באורך 20 בתים ובבלוק יש 4,000 בתים. כמו כן, $B(R)=B(S)=100$, $V(R,A)=20$, $V(R,B)=100$ ו- $V(S,B)=10$. ליחס R יש אינדקס לא מקבץ (non-clustering) על A וליחס S יש אינדקס מקבץ (clustering) על B . בחוצץ יש 7 בלוקים.

א. תארו את הדרך היעילה ביותר לחישוב הביטוי וציינו את מחירה (זכרו: בחוצץ 7 בלוקים)

קוראים את R באופן סדרתי ומבצעים את שתי הבחירות. התוצאה תופסת בלוק אחד בלבד
 בזכרון (וקריאת R מצריכה בלוק נוסף בחוצץ). בהמשך קוראים, דרך האינדקס של S , את
 הרשומות שמקיימות $B=3$ ומבצעים את הצירוף. קריאת S מצריכה בלוק אחד בחוצץ ועוד בלוק
 לתוצאה הסופית (וזאת בנוסף לבלוק עבור תוצאת הביניים של הפעלת שתי הבחירות על R),
 בסה"כ 3 בלוקים. המחיר הוא $111=100+11$.

ב. האם, כדי להשיג את המחיר הטוב ביותר אפשר להסתפק בפחות מ- 7 בלוקים בחוצץ?
 אם כן כמה?

שאלה 12 (8 נקודות)

נתון התזמון הבא של שלוש עסקאות:

$R_1(A), W_1(A), R_2(A), W_2(A), R_3(B), W_3(B), R_1(B), W_1(B)$

- | | | | |
|----|--|-------------------------------------|-------------------------------------|
| א. | האם התזמון בר סדרתיות מבטית? | <input checked="" type="radio"/> כן | <input type="radio"/> לא |
| ב. | האם התזמון יכול להיווצר ע"י 2PL? | <input type="radio"/> כן | <input checked="" type="radio"/> לא |
| ג. | האם התזמון יכול להיווצר ע"י 2PL מחמיר? | <input type="radio"/> כן | <input checked="" type="radio"/> לא |
| ד. | האם יש אפשרות שהתזמון נוצר ע"י פרוטוקול חותמות הזמן תוך הפעלה (לפחות פעם אחת) של הכלל של תומס? | <input type="radio"/> כן | <input checked="" type="radio"/> לא |

שאלה 13 (8 נקודות)

סמנו **כן** אם הטענה נכונה ; אחרת סמנו **לא**.

- | | | | |
|----|---|-------------------------------------|-------------------------------------|
| א. | פרוטוקול האימות יכול ליצור תזמון שבו כל העסקאות מתחייבות, כך שתזמון זה אינו יכול להיווצר ע"י 2PL. | <input checked="" type="radio"/> כן | <input type="radio"/> לא |
| ב. | פרוטוקול 2PL מבטיח שאין קריאות מלוכלכות. | <input type="radio"/> כן | <input checked="" type="radio"/> לא |
| ג. | קיפאון (deadlock) יכול להיווצר ב- 2PL אבל לא ב- 2PL מחמיר. | <input type="radio"/> כן | <input checked="" type="radio"/> לא |
| ד. | ב- 2PL יכול להיווצר קיפאון אפילו אם כל הכתיבות הן עיוורות. | <input checked="" type="radio"/> כן | <input type="radio"/> לא |

שאלה 14 (6 נקודות)

סמנו את הסעיף שאינו נכון.

- | | |
|----|---|
| א. | שקילות קונפליקטים גוררת שקילות מבטים. |
| ב. | למרות שהלוג דורש כתיבה כפולה לדיסק (פעם ללוג ופעם למסד), הוא מייעל את המערכת. |
| ג. | עסקה מתחייבת כאשר רשומת התחייבות נרשמת עבורה בלוג על הדיסק. |
| ד. | בלוג רושמים מתי מתבצעת קריאה מלוכלכת. |