

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

מרצים: פרופ' שרה כהן  
פרופ' יהושע שגיב

תאריך הבחינה: ז' שבט, תשע"ז (3.2.2017)

**אסור לכתוב בשוליים!**

**הנחיות כלליות**

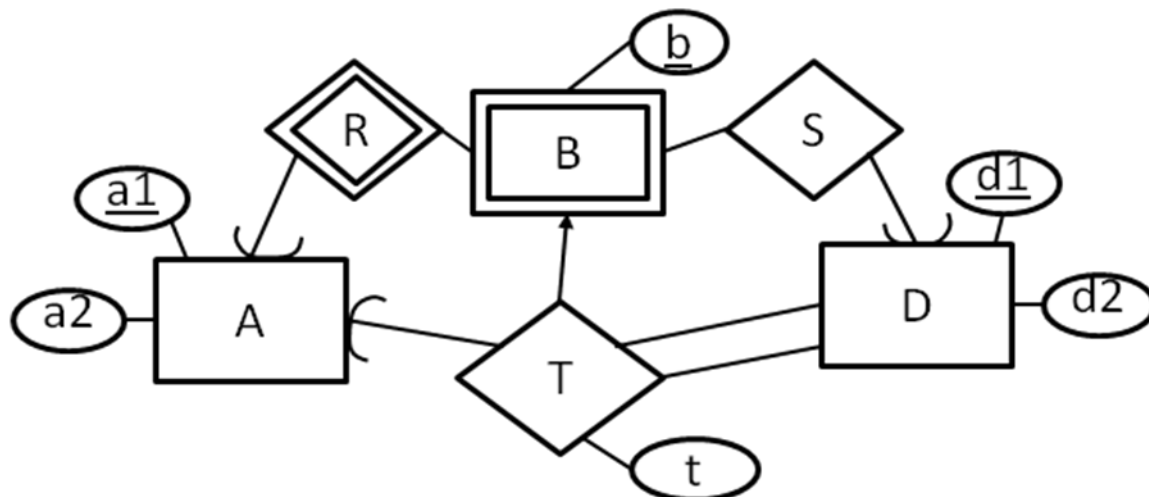
- משך הבחינה: 3 שעות. חומר סגור
- ענו על כל השאלות (בבחינה יש 14 שאלות)
- סמנו תשובה אחת בלבד כשצריך לבחור בין כמה אפשרויות, אלא אם נאמר אחרת
- רשמו מספר מזהה בראש כל דף
- אל תבזבזו זמן על מתן תשובות לא לעניין! ענו רק על מה שנשאלתם
- אפשר להשתמש במחשבוני פשוטים (ללא זיכרון לאגירת נוסחאות או מידע כלשהו)

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:



תרגם את הדיאגרמה ליחסים. בכל יחס יש לציין את שם היחס, שמות השדות וכן את המפתחות של היחס. אם יש כמה אפשרויות לבחור מפתח ליחס, יש לציין את כל האפשרויות.

$A(\underline{a1}, a2)$   
 $B(\underline{a1}, \underline{b}, d1)$  [ OR  $B(a1, \underline{b})$  and  $S(\underline{a1}, \underline{b}, d1)$  ]  
 $D(\underline{d1}, d2)$   
 $T(a1, a1B, b, d11, d12, t)$  with one of the following keys  
     1)  $a1B, b, d11, d12$   
     2)  $a1, d11, d12$

## שאלה 2 (8 נקודות)

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

יהי  $R=(A_1, \dots, A_n)$  יחס ו- $F$  קבוצה של תלויות פונקציונליות. נאמר שקבוצת אטריבוטים  $X$  היא סגורה אם  $X^+ = X$ .

א. ידוע שהתתי קבוצות היחידות הסגורות של  $A_1, \dots, A_n$  הם הקבוצה הריקה והקבוצה  $\{A_1, \dots, A_n\}$ . הוכח או תן דוגמה נגדית פשוטה:  $R$  בהכרח ב-BCNF.

יהי  $X \rightarrow Y$  תלות לא טריואלית ב- $F$ . נניח ל דרך השלילה ש- $X$  אינו מפתח על. נסתכל על הקבוצה  $X^+$ . קבוצה זו סגורה, על פי ההגדרה. היא לא ריקה כי  $X$  אינו ריק. היא אינה שווה ל- $\{A_1, \dots, A_n\}$  כי  $X$  אינו מפתח. אז, מצאנו קבוצה סגורה נוספת, בסתירה לנתון.

ב. ידוע שכל התתי קבוצות של  $A_1, \dots, A_n$  הם סגורות. הוכח או תן דוגמה נגדית פשוטה:  $R$  בהכרח ב-BCNF.

יהי  $X \rightarrow Y$  תלות ב- $F$ . נראה שהתלות בהכרח טריואלית ולכן היחס ב-BCNF. נניח על דרך השלילה שהתלות אינה טריואלית. נסתכל על הקבוצה  $X^+$ . מכיוון שהתלות אינה טריואלית, בהכרח  $X^+$  שונה מ- $X$ . לכן,  $X$  אינה סגורה, בסתירה לנתון.

### שאלה 3 (8 נקודות)

נתונה סכמה  $R=(A,B,C,D,E,G)$  ותלויות פונקציונליות

$$F = \{ADB \rightarrow E, E \rightarrow GC, CB \rightarrow D, A \rightarrow C, G \rightarrow AB, D \rightarrow G\}$$

כמו כן, ניתן פירוק של  $R$  לתתי סכמות  $R_1 = (A,B,C,E)$  ו-  $R_2 = (C,D,G)$ .

א. רשום כיסוי מינימאלי של ההטלה של  $F$  על  $R_1$ .

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

לא

**כן**

האם  $R_1$  ב-3NF?

**לא**

כן

האם  $R_1$  ב-BCNF?

ב. רשום כיסוי מינימאלי של ההטלה של  $F$  על  $R_2$ .

$D \rightarrow C, D \rightarrow G, G \rightarrow D$

לא

**כן**

האם  $R_2$  ב-3NF?

לא

**כן**

האם  $R_2$  ב-BCNF?

### שאלה 4 (6 נקודות)

מבצעים את סדרת הפעולות הבאות:

```
create table R(a integer, b integer);
insert into R values(1,1);
insert into R values(1,2);
insert into R values(2,1);
insert into R values(2,2);
insert into R values(2,3);
insert into R values(3,1);

select a from R R1
where not exists(select b from R R2
                except
                select b from R R3 where R3.a = R1.a)

union all

select a from R R4
where not exists (select * from R R5
                where R5.a not in (select R6.a from R R6
                                   where R6.b=R4.b) );
```

אילו שורות יוחזרו על ידי השאילתה?

**2, 2, 2, 1, 2, 3**

**שאלה 5** (8 נקודות)

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

$$A \rightarrow B, BC \rightarrow G, GE \rightarrow CD, B \rightarrow E, D \rightarrow A$$

א. כתבו את כל המפתחות של  $R$ .

GA, GB, GD, GE, CD, AC, BC

ב. האם  $R$  ב-3NF? נמק בקצרה את תשובתך.

כן. כל האטריבוטים מופיעים במפתח ובפרט צד שמאל של כל תלות הוא שדה במפתח.

ג. מצאו פירוק של  $R$  ל-BCNF שהוא ללא אובדן, בעזרת האלגוריתם שנלמד בכיתה. ציינו את התלויות הפונקציונליות שלפיהן התבצע הפירוק.

AB, BE, DA, CDG

$A \rightarrow B, B \rightarrow E, D \rightarrow A$

**שאלה 6** (8 נקודות)

להלן הגדרת יחסים בבסיס נתונים כלשהו:

**People(pId, name)**

**SickDays(pId, month, year, numdays)**

הטבלה People מכילה מידע על אנשים – מספר מזהה, ושם. הטבלה SickDays מכילה שורות שמציינות את מספר ימי המחלה של אנשים לפי חודש ושנה. ניתן להניח שהטבלה אינה מכילה ערכי NULL.

כתוב את השאילתה הבאה בSQL:

החזר את שם האיש (או אנשים, במקרה של שוויון) שהיו חולים בשנת 2016 הכי הרבה ימים, בתנאי שסכום ימי המחלה שלהם, בכל השנים ביחד, הוא לכל היותר 100. במילים אחרות, שם של איש  $p$  צריך להיות בתוצאה אם  $p$  היה חולה לכל היותר 100 ימים בסה"כ וגם  $p$  היה חולה הכי הרבה ימים בשנת 2016 (כלומר, הכי הרבה ימים בהשוואה לכל אחד אחר, כולל אלה שהיו חולים יותר מ-100 ימים בסה"כ).

```
select name  
from People  
where pid in (  
    select pid  
    from SickDays  
    group by pid  
    having sum(numdays)<=100  
intersect  
    select pid  
    from SickDays  
    where year = 2016  
    group by pid  
    having sum(numdays)>=all (select sum(numdays)  
                                from SickDays  
                                where year = 2016  
                                group by pid)  
)
```

**שאלה 7 (6 נקודות)**

מבוצע קטע הקוד הבא:

```
create table R(a integer);

create function rrr() returns trigger as$$
declare
    c integer;
begin
    select count(*) into c from R;
    if (c < new.a) then
        update R set a = c where a = new.a;
        insert into R values(new.a);
        return new;
    end if;
    if (c > new.a) then
        delete from R where a >= new.a;
        insert into R values(new.a);
    end if;
    return new;
end
$$language plpgsql;

create trigger RRR
after insert on R
for each row
execute procedure rrr();

insert into R values(5);
```

מה הוא תוכן הטבלה R לאחר ביצוע כל הקוד הרשום למעלה?

**השורות 1, 2, 3, 4, 5**

---

מה יקרה אם כעת נבצע את הפעולה הנוספת הבאה?

```
insert into R values(3);
```

**כעת יהיו בטבלה השורות 1, 2, 3**

---

מה יקרה אם כעת נבצע את הפעולה הנוספת הבאה?

```
insert into R values(-1);
```

**הטריגר יכנס ללולאה אינסופית**

---

### שאלה 8 (6 נקודות)

בשני הסעיפים הבאים,  $X$  היא קבוצה של אטריבוטים ומניחים שכל הביטויים מוגדרים היטב. בכל אחד משני הסעיפים, **סמנו 1** אם שני הביטויים שקולים, **סמנו 2** אם יש הכלה בדיוק באחד משני הכיוונים ו**סמנו 3** אם אין הכלה באף כיוון.

- א.  $\pi_X(R \bowtie S) \bowtie T$  ו-  $R \bowtie \pi_X(S \bowtie T)$       1      2      3
- ב.  $\pi_X(R \cup S) - T$  ו-  $\pi_X(R - T) \cup \pi_X(S - T)$       1      2      3

### שאלה 9 (8 נקודות)

צריך לחשב את הביטוי  $\pi_{ACD}(R(A,B) \bowtie S(B,C) \bowtie T(C,D))$ . אטריבוט  $B$  הוא מפתח ביחס  $S$  (כלומר, מפתח בהקשר של תלויות פונקציונליות). כמו כן,  $V(S,C)=50$ ,  $V(R,B)=100$  ו-  $V(T,C)=20$ . גודלו של כל אחד משלושת היחסים הוא 100 בלוקים. ערך של אטריבוט הוא תמיד באורך 20 בתים ובבלוק יש מקום ל- 4,000 בתים. בחוצץ (buffer) יש 103 בלוקים.

- סמנו את הצירוף שמספר הרשומות בתוצאה שלו קטן ממש מזה שבכל צירוף אחר של שניים מבין היחסים  
 א.  $R \bowtie S$       ב.  $S \bowtie T$       ג.  $R \bowtie T$       ד. אין צירוף כזה
- סמנו את הצירוף שאם עושים אותו ראשון, אז מיד לאחריו (ולפני שעושים את הצירוף השני) אפשר לבצע הטלה שמוחקת חלק מהעמודות  
 א.  $R \bowtie S$       ב.  $S \bowtie T$       ג.  $R \bowtie T$       ד. אין צירוף כזה
- אם מחשבים את הביטוי בדרך שמייצרת תוצאת ביניים, מה הגודל בבלוקים של תוצאת הביניים הקטנה ביותר שבכלל אפשר לייצר?  
 א. 50      ב. 100      ג. 150      ד. 20,000  
 ה. אם כל התשובות הקודמות אינן נכונות, כתבו כאן את הנכונה \_\_\_\_\_
- בהנחה שאת שתי פעולות הצירוף עושים בשיטה של block nested loop join, מה המחיר הטוב ביותר שאפשר להשיג בדרך זו (בהתאם לכל הנתונים בתחילת השאלה)?  
 א. 300      ב. 400      ג. 450      ד. 500      ה. 600  
 ו. 650      ז. 700      ח. 750      ט. 900      י. 1200  
 יא. אם כל התשובות הקודמות אינן נכונות, כתבו כאן את הנכונה \_\_\_\_\_
- בהנחה שאת שתי פעולות הצירוף עושים בשיטה של hash join, מה המחיר הטוב ביותר שאפשר להשיג בדרך זו (בהתאם לכל הנתונים בתחילת השאלה)?  
 א. 600      ב. 700      ג. 900      ד. 1,000      ה. 1,100  
 ו. 1,200      ז. 1,300      ח. 1,400      ט. 1,500      י. 1,600  
 יא. אם כל התשובות הקודמות אינן נכונות, כתבו כאן את הנכונה \_\_\_\_\_



**שאלה 10** (6 נקודות)

סמנו את הסעיף שהטענה הכתובה בו נכונה.

- א. העדיפות הכי נמוכה לשמירת בלוקים בחוצץ (buffer) ניתנת לצמתים של עצי  $B+$ , כי רוב הזמן קוראים בלוקים של היחסים ולא של עצים אלה.
- ב. הבעיה עם בלוקים של overflow באינדקס דליל (sparse index) היא שבשום מקרה אי אפשר למלא אותם עד הסוף, כי הקובץ שעליו מצביע האינדקס חייב להיות ממוין.
- ג. כשיש אינדקס דליל עם חזרות, אז כשמחפשים את כל הרשומות עם ערך נתון, ממשיכים לבלוק הבא של הקובץ (שעליו מצביע האינדקס) אם ורק אם לרשומה האחרונה בבלוק יש את הערך שמחפשים.
- ד. נתון אינדקס אחד על אטריבוט  $A$  ואינדקס שני על  $B$ . רוצים למצוא את כל הרשומות של היחס  $R(A,B)$  שמקיימות  $A=10 \wedge B=25$ . אפשר לעשות זאת ע"י כך שקודם כל מבצעים חיתוך של קבוצות מצביעים (pointers) אם ורק אם לפחות אחד משני האינדקסים הוא דליל.
- ה. כל התשובות הקודמות אינן נכונות.

**שאלה 11** (8 נקודות)

צריך לחשב את הביטוי  $(R(A,B) \bowtie S(B,C))$ .  $\pi_{AC} \sigma_{B=3}$ .  $B(R)=400$  ו-  $B(S)=10,000$ . כמו כן,  $V(R,B)=4$  ו-  $V(S,B)=100$ . היחס  $S$  ממוין על  $B$ . ערך של אטריבוט הוא תמיד באורך 20 בתים ובבלוק יש מקום ל- 4,000 בתים. בחוצץ (buffer) יש 70 בלוקים.

1. מה מספר הרשומות בתוצאה, בהנחה שלא מבטלים כפילויות אחרי ההטלה?  
 א. 25,000,000    ב. 50,000,000    ג. 100,000,000    ד. 400,000,000  
 ה. כל התשובות הקודמות אינן נכונות
2. איזה פעולה תבוצע ראשונה בחישוב היעיל ביותר של כל הביטוי?  
 א. צירוף    ב. הטלה    ג. בחירה
3. בהנחה שמחשבים את הצירוף בשיטה של block nested loop join, מה המחיר הטוב ביותר שאפשר להשיג בדרך זו (בהתאם לכל הנתונים בתחילת השאלה)?  
 א. 10,400    ב. 10,600    ג. 10,700    ד. 20,400    ה. 20,600  
 ו. 20,700    ז. 30,400    ח. 40,400    ט. 60,400    י. 67,200  
 יא. אם כל התשובות הקודמות אינן נכונות, כתבו כאן את הנכונה \_\_\_\_\_
4. האם אפשר להשתמש בשיטה של sort-merge join במקרה הזה?  
 א. כן    ב. לא    ג. אי אפשר לדעת לפי הנתונים בשאלה
5. האם אפשר להשתמש בשיטה של hash join במקרה הזה?  
 א. כן    ב. לא    ג. אי אפשר לדעת לפי הנתונים בשאלה

## שאלה 12 (8 נקודות)

נתון התזמון הבא:

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

בכל אחד מהסעיפים, סמנו את התשובה הנכונה.

- האם התזמון בר סדרתיות קונפליקטית? ☒ כן ☐ לא
- האם התזמון יכול להיווצר ע"י פרוטוקול 2PL? ☐ כן ☒ לא
- האם אפשר למצוא חותמות זמן התחלתיות של העסקאות ושל כל הפריטים, כך שהתזמון יתקבל על ידי פרוטוקול חותמות הזמן? ☐ כן ☒ לא
- האם אפשר למצוא ערכים עבור חותמות הזמן הנדרשות, כך שהתזמון יתקבל על ידי פרוטוקול האימות? ☐ כן ☒ לא

## שאלה 13 (8 נקודות)

בכל אחד מהסעיפים, סמנו את התשובה הנכונה.

- אם תזמון הוא ללא כתיבות עיוורות ויש בו גם כתיבות חסרות השפעה, אז הוא לא יכול להיות בר-סדרתיות מבטית. ☐ כן ☒ לא
- הפרוטוקול האופטימי אינו מאפשר קריאות מלוכלכות, אבל זה אפשרי בפרוטוקול האימות. ☐ כן ☒ לא
- בשיטה wait/die למניעת מצבי קיפאון (deadlock), כשמנעול משתחרר אז חייבים לתת אותו לצעירה ביותר מבין כל העסקאות שמחכות לו, כדי למנוע היווצרות מצב של הרעבה (starvation). ☐ כן ☒ לא
- לבעיה האם שני תזמונים הינם שקולי מבטים יש אלגוריתם פולינומיאלי. ☐ כן ☒ לא

## שאלה 14 (6 נקודות)

סמנו את הסעיף שהטענה הכתובה בו נכונה.

- כשמבטלים עיסקה בודדת צריך להחזיר לערכם המקורי רק את הפריטים ששונו (ע"י העיסקה) בדיסק של המסד. מה שנכתב רק בזכרון הפנימי (אבל לא על הדיסק) הוא חסר השפעה. ☐ א. ☐ ב. ☐ ג. ☒ ד. ☐ ה.
- במערכות שמשממשות בלוג אין צורך לבצע גיבוי של הנתונים. ☐ א. ☐ ב. ☐ ג. ☐ ד. ☐ ה.
- אם תזמון הוא ללא קריאות מלוכלכות אז הוא בהכרח בר-סדרתיות מבטית. ☐ א. ☐ ב. ☐ ג. ☐ ד. ☐ ה.
- בגלל השימוש בלוג אפשר לכתוב לדיסק של המסד בלוקים שיש בהם ערכים שנכתבו על ידי עסקאות שטרם התחייבו. ☐ א. ☐ ב. ☐ ג. ☒ ד. ☐ ה.
- כל התשובות הקודמות אינן נכונות. ☐ א. ☐ ב. ☐ ג. ☐ ד. ☐ ה.