

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

מרצים: פרופ' שרה כהן  
פרופ' יהושע שגיב

תאריך הבחינה: כ"ו בשבט, תשע"ו (5.2.2016)

**אסור לכתוב בשוליים!**

**הנחיות כלליות**

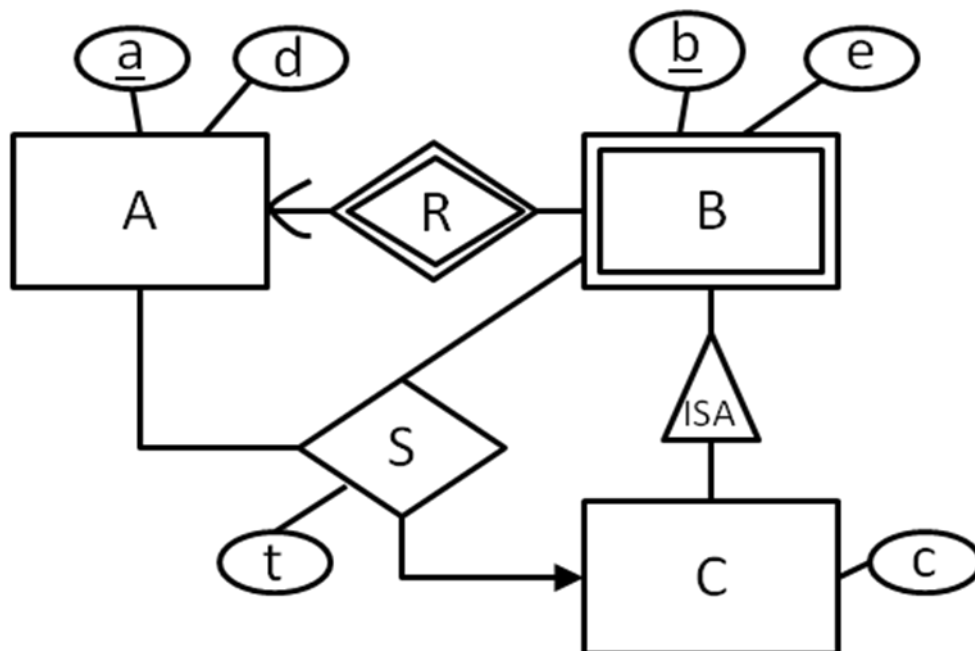
- משך הבחינה: 3 שעות. חומר סגור
- ענו על כל השאלות (בבחינה יש 14 שאלות)
- סמנו תשובה אחת בלבד כשצריך לבחור בין כמה אפשרויות, אלא אם נאמר אחרת
- רשמו מספר מזהה בראש כל דף
- אל תבזבזו זמן על מתן תשובות לא לעניין! ענו רק על מה שנשאלתם
- אפשר להשתמש במחשבוני פשוטים (ללא זיכרון לאגירת נוסחאות או מידע כלשהו)

| 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**

תרגם את הדיאגרמה ליחסים. בכל יחס יש לציין את שם היחס, שמות השדות וכן את המפתחות של היחס.

A(a, d)  
 B(a,b,e)  
 C(a,b,c) or C(a,b, c, e)  
 S(aa,ab,bb,ac,bc,t)  
 where aa is the attribute a from A,  
 ab, bb are the attributes a and b from B  
 ac, bc are the attributes a and b from C

**שאלה 2 (8 נקודות)**

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

א. יהי  $R$  יחס ו- $F$  קבוצה של תלויות פונקציונליות. ידוע שיש ל- $R$  בדיוק מפתח אחד. הוכח או תן דוגמה נגדית פשוטה:  $R$  בהכרח ב-BCNF.

$F = \{A \rightarrow B\}$ ,  $R = (A, B, C)$

The only key is AC and the relation is not in BCNF

ב. יהי  $R$  יחס ו- $F$  קבוצה של תלויות פונקציונליות. ידוע שיש ל- $R$  בדיוק מפתח אחד. הוכח או תן דוגמה נגדית פשוטה:  $R$  ב-BCNF אם ורק אם  $R$  ב-3NF.

לפי ההגדרה, אם  $R$  ב-BCNF אזי  $R$  ב-3NF. נשאר להראות את הכיוון השני. נניח על דרך השלילה ש- $R$  ב-3NF ולא ב-BCNF. יהי  $X \rightarrow A$  תלות ב- $F$ . יש 3 אפשרויות:

- 1) התלות טריויאלית ולא מהווה סתירה לתנאי ה-BCNF
- 2)  $X$  מפתח על והתלות לא מהווה סתירה לתנאי ה-BCNF
- 3)  $A$  שדה במפתח. (ואפשרויות 1 ו-2 לא מתקיימים). יהי  $K$  המפתח היחדי של היחס. נסתכל על הקבוצה  $K - \{A\} \cup X = K'$ . קל לראות ש- $K'$  הוא מפתח על כי בסגור שלו יש את כל  $K$ . הקבוצה  $K'$  לא מכיל את  $K$  כי  $A$  נמצא ב- $K$  ולא ב- $K'$ . לכן  $K'$  מכיל מפתח שונה מ- $K$  ולכן ליחס יש יותר ממפתח אחד בסתירה לנתון.

**שאלה 3** (8 נקודות)

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

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

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

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

|                                                                                   |
|-----------------------------------------------------------------------------------|
| $C \rightarrow A$<br>$C \rightarrow B$<br>$D \rightarrow A$<br>$BD \rightarrow C$ |
|-----------------------------------------------------------------------------------|

האם  $R_1$  ב-3NF? ☒ כן ☐ לא

האם  $R_1$  ב-BCNF? ☐ כן ☒ לא

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

|                                         |
|-----------------------------------------|
| $E \rightarrow G$<br>$DG \rightarrow E$ |
|-----------------------------------------|

האם  $R_2$  ב-3NF? ☒ כן ☐ לא

האם  $R_2$  ב-BCNF? ☐ כן ☒ לא

**שאלה 4** (6 נקודות)

מבצעים את סדרת הפעולות הבאות:

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

```
insert into R values(0, 1);
insert into R values(1, 1);
insert into S values(0, 2);
insert into S values(1, 3);
insert into S values(1, 4);
```

```
with recursive T(a, b) as
((select * from R)
 union all
 (select a, b + 1 from T where b <
 (select max(b) from S where a = T.a)))
select * from T;
```

אילו שורות יוחזרו על ידי השאילתה?

מספר מזהה: \_\_\_\_\_ עמוד 5 מתוך 12

(0,1), (1,1), (0,2), (1,2), (1,3), (1,4)

**שאלה 5** (8 נקודות)

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

$B \rightarrow E, AB \rightarrow G, AC \rightarrow BD, AE \rightarrow CD, DE \rightarrow C$

א. כתבו את כל המפתחות של  $R$ .

AB, AC, AE

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

$R$  ב-3NF אך לא ב-BCNF. התלות  $B \rightarrow E$  סותר את תנאי ה-BCNF כי  $B$  לא מפתח על.

בכל התלויות או שצד ימין שדה במפתח או שצד שמאל מפתח על ולכן  $R$  ב-3NF.

ג. מצאו פירוק של  $R$  ל-BCNF שהוא ללא אובדן.

ABDG  
BE  
DEC

Or

ABDG  
DBC  
BE

**שאלה 6** (8 נקודות)

להלן הגדרת יחס בבסיס נתונים כלשהו :

**Weather(city, date, temp)**

ניתן להניח שהטבלה אינה מכילה ערכי NULL.

כתוב את השאילתות הבאות בSQL : (שימו לב שסעיף ב נמצא בעמוד הבא)

א. החזר את כל הערים  $c$  בהם בכל תאריך בו נרשם מזג אויר בירושלים, נרשם גם מזג האויר בעיר  $c$  ובירושלים היה קר יותר מאשר ב- $c$ . אם העיר ירושלים לא מופיעה בטבלה Weather, כל הערים שמופיעות ב-Weather מקיימות את התנאי באופן ריק וצריכות להיות מוחזרות.

```
select city
from Weather W1
where not exists (select date from Weather W2 where city = 'Jerusalem'
                  minus select date from Weather W3 where W3.city = W1.city)
and not exists (select * from Weather W4, Weather W5
                  where W4.date = W5.date and W4.city = 'Jerusalem' and
                  W5.city = W1.city and W4.temp > W5.temp)
```

ב. החזר את התאריך הקר ביותר בהיסטוריה, כלומר התאריך בו הטמפרטורה הממוצעת היא הנמוכה ביותר מבין התאריכים המופיעים בטבלה. במילים אחרות, לכל תאריך יש לחשב את הממוצע של כל הטמפרטורות המופיעות בטבלה עבור תאריך זה. יש להחזיר את כל התאריכים שבהם הממוצע הוא הנמוך ביותר.

```
select date
from Weather
group by date
having avg(temp) <= ALL (select avg(temp) from Weather group by date)
```



שאלה 7 (6 נקודות)

מבוצע קטע הקוד הבא:

```
create table R(a integer);

create or replace function yyy() returns trigger as$$
declare
    c integer;
begin
    select count(*) into c from R where R.a = new.a;
    if (c>2) then
        return null;
    end if;
    if (new.a < 0) then
        insert into R values(new.a + 1);
        return old;
    end if;
    if (new.a > 0) then
        insert into R values(new.a - 2);
        return new;
    end if;
    return new;
end
$$language plpgsql;

create trigger zzz
before insert on R for each row
execute procedure yyy();

insert into R values(-2);
insert into R values(2);
insert into R values(3);
```

מה הוא תוכן הטבלה R לאחר ביצוע פעולת ההכנסה של (-2)?

(0)

---

אלו שורות מתווספות לטבלה לאחר ביצוע פעולת ההכנסה של (2)?

(2) (0)

---

אלו שורות מתווספות לטבלה לאחר ביצוע פעולת ההכנסה של (3)?

(3) (1) (0)

---

**שאלה 8** (6 נקודות)

נתונים היחסים  $R(A,B,C)$  ו-  $S(B,C,D)$ . בכל אחד משני הסעיפים, סמנו 1 אם שני הביטויים שקולים, סמנו 2 אם יש הכלה בדיוק באחד משני הכיוונים וסמנו 3 אם אין הכלה באף כיוון.

- א.  $\pi_{ACD}(R \bowtie S)$  ו-  $(\pi_{CD}S) \bowtie (\pi_{AC}R)$       1      2      3
- ב.  $R - R$  ו-  $(R \cup R) - (R \bowtie R)$       1      2      3

**שאלה 9** (8 נקודות)

צריך לחשב את הצירוף  $R(A,B,C) \bowtie S(B,C,D)$ . היחס  $R$  ממוין על אטריבוט  $C$  בלבד, בעוד שהיחס  $S$  כלל אינו ממוין.  $V(S,B)=1,000$ ,  $V(R,C)=1,000$ ,  $V(R,B)=100$  ו-  $V(S,C)=100$ . נתון שהתפלגות הערכים היא אחידה. כל אחד משני היחסים תופס 10,000 בלוקים. כל ערך של אטריבוט  $A$  או  $D$  הוא באורך של 30 בתים, בעוד שכל ערך של אטריבוט  $B$  או  $C$  הוא באורך של 10 בתים. בבלוק יש מקום ל- 4,000 בתים. בחוצץ (buffer) יש 1,000 בלוקים. אין אינדקסים ואסור לבנות אותם. החליטו מהו החישוב היעיל ביותר של הביטוי ובהתאם לכך ענו על הסעיפים הבאים.

- א. מה מספר הרשומות ומה מספר הבלוקים בתוצאה?  
640,000 רשומות ו- 12,800 בלוקים

---

---

---

---

- ב. תארו את השיטה היעילה ביותר שלפיה יחושב הצירוף וכיתבו מהו המחיר?

מבצעים רק מיון של  $S$  על  $C$  ואחרי זה מבצעים מיזוג, כאשר  $R$  הוא רשימה אחת בלבד. בשלב המיזוג גם מוודאים שיש שוויון על  $B$ . לפיכך, קוראים את  $R$  פעם אחת בלבד ואת  $S$  שלוש פעמים. המחיר הכולל הוא 40,000

---

---

---

---

---

- ג. בהנחה שמבצעים את החישוב כפי שתיארתם בסעיף הקודם, מהו הגודל המינימאלי של החוצץ שמאפשר להגיע למחיר שכתבתם בתשובה (כזכור התפלגות הערכים היא אחידה)?  
גודל החוצץ הנדרש הוא שורש הגודל של  $S$  ועוד 2 ועוד  $P$ . במקרה הזה  $P=10$ . לכן גודל החוצץ הנדרש הוא 112

---

---

---

## שאלה 10 (6 נקודות)

סמנו את הסעיף שהטענה הכתובה בו נכונה.

- א. השיטה של חוצץ כפול (double buffering) נועדה כדי לקרוא מהדיסק קובץ דרך אינדקס, כאשר בלוק אחד בחוצץ הוא עבור האינדקס ובלוק שני נועד לקובץ עצמו.
- ב. בלוקים של overflow באינדקס דליל (sparse index) פוגעים ביעילות, כי הם מגדילים את המרחק מהשורש לחלק מהעלים.
- ג. כאשר קוראים שני בלוקים מדיסק, תמיד יש השהייה (latency) בין קריאת הראשון לשני, בגלל הסיבוב של הדיסק.
- ד. ב-B+tree נהוג לפצל צמתים, כדי להשיג יעילות גדולה יותר ע"י שימוש בחוצץ כפול (double buffering).
- ה. אף אחת מהתשובות הקודמות אינה נכונה.

## שאלה 11 (8 נקודות)

צריך לחשב את הביטוי  $\sigma_{A=3} \pi_{AC}(R(A,B) \bowtie S(B,C))$ .  $B(R)=B(S)=1,000,000$ . כמו כן,  $V(S,B)=10,000$ ,  $V(R,A)=100$  ואטריבוט B הוא מפתח ביחס R. ערך של אטריבוט הוא תמיד באורך 20 בתים ובלוק יש מקום ל-4,000 בתים. אין אינדקסים ואסור לבנות אותם. בחוצץ (buffer) יש 1,100 בלוקים. עליכם להחליט מהו החישוב היעיל ביותר של הביטוי ובהתאם לכך לענות על הסעיפים הבאים.

- א. מה מספר הרשומות ומה מספר הבלוקים בתוצאה?  
בתוצאה יש מיליון רשומות ו-10,000 בלוקים

- ב. איזה פעולה אלגברית תבוצע ראשונה בחישוב היעיל ביותר של הביטוי?  
בחירה

- ג. תארו מהי הדרך היעילה ביותר לחישוב הביטוי וכיתבו מהו המחיר?

משתמשים ב-hash join תוך ביצוע הבחירה מיד כשקוראים את R בפעם הראשונה. לכן, החלוקה לדליים היא של יחס בגודל 10,000 בלוקים. לפיכך, פעם אחת קוראים את R המקורי ועוד פעמיים קוראים רק את החלק של R לאחר הבחירה. את היחס S קוראים בשלמותו שלוש פעמים. לכן, המחיר הכולל הוא 4,020,000

- ד. בהנחה שמבצעים את החישוב כפי שתואר בסעיף הקודם, מהו הגודל המינימאלי של החוצץ שנדרש כדי להגיע למחיר שכתבתם בתשובה?

גודל החוצץ הנדרש הוא שורש הגודל של R אחרי הבחירה ועוד 2, כלומר 102

## שאלה 12 (8 נקודות)

נתון התזמון הבא :

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

בכל אחד מהסעיפים, סמנו את התשובה הנכונה.

- א. האם התזמון בר סדרתיות קונפליקטית? ☒ כן ☐ לא
- ב. כמה מבין התזמונים הסדרתיים של שלושת העסקאות הינם שקולי מבטים לתזמון הנתון? התשובה חייבת להיות רק מספר, שאותו יש לרשום מעל הקו: \_\_\_\_\_<sup>1</sup>
- ג. האם התזמון הנתון יכול להיווצר ע"י פרוטוקול 2PL? ☒ כן ☐ לא
- ד. נניח שחותמות הזמן ההתחלתיות של כל הפריטים הן אפס. האם אפשר לתת חותמות זמן לעסקאות, כך שפרוטוקול חותמות הזמן יבצע את התזמון בהצלחה ובנוסף גם יפעיל את הכלל של תומס לגבי פעולת כתיבה אחת לפחות? ☒ כן ☐ לא

## שאלה 13 (8 נקודות)

בכל אחד מהסעיפים, סמנו את התשובה הנכונה.

- א. אם תזמון נוצר ע"י פרוטוקול חותמות הזמן ואין בו כתיבות עיוורות, אז הוא בר סדרתיות קונפליקטית. ☒ כן ☐ לא
- ב. בפרוטוקול האופטימי, אם עסקה מתבטלת אז חייבים להשאיר לה את חותמת הזמן המקורית כאשר היא מתחילה מחדש. ☒ כן ☐ לא
- ג. אם תזמון הוא בר סדרתיות קונפליקטית, אז אין בו כתיבות חסרות השפעה. ☒ כן ☐ לא
- ד. בגלל כתיבות עיוורות, יתכן שעסקה קוראת עבור פריט A ערכים שונים בשני תזמונים שקולי מבטים. ☒ כן ☐ לא

## שאלה 14 (6 נקודות)

סמנו את הסעיף שהטענה הכתובה בו נכונה.

- א. ביטול עסקה בודדת מתבצע ע"י מעבר על רשומות הלוג שלה לפי הסדר, מהראשונה תחילה ועד לאחרונה, ועבור כל רשומת עדכון מחזירים לפריט במסד את ערכו לפני השינוי. ☐ כן ☒ לא
- ב. השימוש בלוג פוגע בביצועים של המערכת (כלומר, עושה אותה לפחות יעילה), כי כל שינוי צריך לכתוב פעמיים לדיסק. ☐ כן ☒ לא
- ג. לא מרשים קריאות מלוכלכות, כדי למנוע מצב שבו עסקה צריכה לשחרר מנעולים לפני שהיא מתחייבת. ☐ כן ☒ לא
- ד. בפרוטוקול האימות אין צורך לעשות שום שינויים במסד הנתונים כאשר מבטלים עסקה בודדת (בהנחה שהמערכת ממשיכה לפעול בצורה תקינה). ☒ כן ☐ לא
- ה. אף אחת מהתשובות הקודמות אינה נכונה. ☐ כן ☒ לא