

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

מרצה: פרופ' שרה כהן

תאריך הבחינה: ג' באדר, תשע"ח (18.2.2018)

אסור לכתוב בשוליים!

הנחיות כלליות

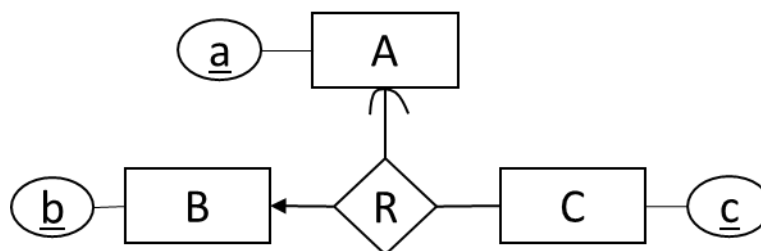
- משך הבחינה: 3 שעות. חומר סגור
- ענו על כל השאלות (בבחינה יש 14 שאלות)
- סמנו תשובה אחת בלבד כשצריך לבחור בין כמה אפשרויות, אלא אם נאמר אחרת
- רשמו מספר מזהה בראש כל דף
- אל תבזבזו זמן על מתן תשובות לא לעניין! ענו רק על מה שנשאלתם
- אפשר להשתמש במחשבוני פשוטים (ללא זיכרון לאגירת נוסחאות או מידע כלשהו)

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

א. לפי הדיאגרמה, האם יתכן ש $A=\{a_1, a_2\}$, $B=\{b_1, b_2\}$, $C=\{c_1\}$?

לא

כן

אם כן, תן דוגמה לתוכן אפשרי עבור R, ואם לא, נמק מדוע זה אינו אפשרי:

$$R = \{(a_1, b_1, c_1), (a_2, b_2, c_1)\}$$

1

אם ענית כן, כמה אפשרויות נוספות יש עבור התוכן של R?

ב. לפי הדיאגרמה, האם יתכן ש $A=\{a_1, a_2\}$, $B=\{b_1\}$, $C=\{c_1, c_2\}$?

לא

כן

אם כן, תן דוגמה לתוכן אפשרי עבור R, ואם לא, נמק מדוע זה אינו אפשרי:

$$R = \{(a_1, b_1, c_1), (a_2, b_1, c_2)\}$$

3

אם ענית כן, כמה אפשרויות נוספות יש עבור התוכן של R?

ג. לפי הדיאגרמה, האם יתכן ש $A=\{a\}$, $B=\{b_1, b_2\}$, $C=\{c_1, c_2\}$?

לא

כן

אם כן, תן דוגמה לתוכן אפשרי עבור R, ואם לא, נמק מדוע זה אינו אפשרי:

לא אפשרי. בגלל החץ לתוך A, נובע שהגודל של R הוא מכלפת הגדלים של B ו-C. כלומר, 4. בגלל החץ לתוך B הגודל של R הוא לכל היותר מכלפת הגדלים של A ו-C. כלומר 2.

אם ענית כן, כמה אפשרויות נוספות יש עבור התוכן של R?

שאלה 2 (8 נקודות)

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

בסעיפים הבאים, נתון $R = (A, B, C, D)$, וקבוצת תלויות פונקציונליות F כך שהקבוצות אטריבוטים הסגורות היחידות הם: $\emptyset, \{A, B\}, \{C\}, \{D\}, \{A, B, C, D\}$.

א. רשום תוכן אפשרי עבור F , אם ידוע ש R ב- $BCNF$. אם לא קיים F מתאים, הסבר מדוע F כזה אינו קיים.

לא יתכן ש R ב- $BCNF$. מעצם קיום הקבוצה $\{A, B\}$ נובע ש $A \rightarrow B$. אבל A אינו מפתח כי הקבוצה $\{A, B\}$ סגורה.

ב. רשום תוכן אפשרי עבור F , אם ידוע ש R ב- $3NF$ אך לא ב- $BCNF$. אם לא קיים F מתאים, הסבר מדוע F כזה אינו קיים.

$A \rightarrow B$
 $B \rightarrow A$
 $AC \rightarrow D$
 $AD \rightarrow C$
 $CD \rightarrow A$

שאלה 3 (8 נקודות)

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

$$F = \{A \rightarrow D, BD \rightarrow C, C \rightarrow E, E \rightarrow A, D \rightarrow C\}$$

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

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

$A \rightarrow C$
 $C \rightarrow A$

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 .

$D \rightarrow C$
 $C \rightarrow E$

$E \rightarrow D$

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);
create table S(b integer, c integer);
```

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

```
select R1.a, R2.a, R3.b,
       (select max(b) from S where R1.a = S.c or R1.b = S.b)
from R R1,
       (select a from R except select b from S) as R2,
       (select b from R except select c from S) as R3
where R1.a <> R2.a and R1.a + R2.a + R3.b < 10;
```

אילו שורות יוחזרו על ידי השאילתה?

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

שאלה 5 (8 נקודות)

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

$$A \rightarrow B, ABCD \rightarrow E, EG \rightarrow HI, ACDG \rightarrow EH$$

- א. האם R_1 ב-3NF? ☒ לא ☐ כן
- האם R_1 ב-BCNF? ☒ לא ☐ כן

ב. מצא כיסוי מינימאלי ל- F .

$A \rightarrow B$
 $ACD \rightarrow E$
 $EG \rightarrow H$
 $EG \rightarrow I$

ג. מצא פירוק של R ל-3NF שהוא ללא אובדן ומשמר תלויות.

(A,B) (A,C,D,E) (E,G,H) (E,G,I) (A,C,D,G)

ד. איך היתה משתנה התשובה לסעיף ג' אם $R=(A,B,C,D,E,G, H, I, J)$ (אבל F נשאר ללא שינוי).

נחליף את (A,C,D,G) עם (A,C,D,G,J)

שאלה 6 (8 נקודות)

להלן הגדרת יחסים בבסיס נתונים כלשהו:

Supplies(supplier, part)

Requires(job, part)

הטבלה Supplies מכילה שורות שמציינות את החלקים אשר ספק מספק. הטבלה Requires מכילה את החלקים שעבודה מסויימת צורך. ניתן להניח שהטבלאות אינן מכילות ערכי NULL.

כתוב את השאילתה הבאה בSQL:

מצא את כל הזוגות של ספק ועבודה כך שהספק מספק את כל החלקים הנדרשים עבור העבודה.

```
Select supplier, job
From Suppliers S, Requires R
Where (not exists (select part from Requires R1 where R1.job = R.job
                  Except
                  Select part from Suppliers S1 where S1.supplier = S.Supplier))
```

שאלה 7 (6 נקודות)

המטרה של פרוטוקול 2-phase commit הוא: (הקף את כל התשובות הנכונות)

- א. למנוע היווצרות של deadlocks.
- ב. לוודא שתוצאת החישוב של מספר טרנזקציות תהיה שקולה לתוצאה של ריצה סדרתית של הטרנזקציות.
- ג. ☒ לוודא אטומיות של טרנזקציות, כמו במסד רלציוני.
- ד. לוודא שכל המחשבים שמשתתפים בטרנזקציה יבצעו פעולת commit (אם כך הוחלט) בדיוק באותו רגע.
- ה. ☒ לוודא שאם מחשב אחד השותף לטרנזקציה מבצע abort, כך יעשו שאר המחשבים השותפים לטרנזקציה.
- ו. לאפשר למחשבים השייכים למערכת NoSQL להשיג את התכונה A ממשפט CAP.
- ז. ☒ לאפשר למחשבים השייכים למערכת NoSQL להשיג את התכונה C ממשפט CAP.

שאלה 8 (6 נקודות)

נתון היחס הבא, המציין שלשות של מספר מזהה, שנת לידה ומדינת מגורים.

Person(id, byear, country)

כתוב שאילתה באלגברה רלציונית אשר מחזירה לכל מדינה, את המזהה של האדם המבוגר ביותר במדינה. אם יש כמה אנשים מבוגרים ביותר באותו מדינה, יש להחזיר את האדם עם המספר מזהה הקטן ביותר מבין האנשים המבוגרים ביותר. שים לב: השאילתה צריכה להחזיר זוגות (id, country).

$$\pi_{id, country} \text{Person} - \pi_{id, country} (\sigma_{(byear > b \text{ or } (b = byear \text{ and } id > i))} (\text{Person} \bowtie \rho_{P(i, b, country)} \text{Person}))$$

שאלה 9 (8 נקודות)

רוצים לחשב את הביטוי $\pi_{A,C}(R(A,B) \bowtie S(B,C))$. ההטלה היא ללא מחיקת כפילויות. גודלי היחסים הם $B(R)=1,000$, $B(S)=10,000$. ליחס S יש אינדקס על האטריבוט B שזמן הגישה אליו זניח. גודל כל אחד מהאטריבוטים הוא 20 bytes וגודל בלוק הוא 4,000 bytes. הבלוקים הם מלאים. בחוצץ (buffer) יש 100 בלוקים. כמו כן, $V(R,A)=100$, $V(R,B)=20$. נחשב את התוצאה עם אלגוריתם *index-nested-loop-join*.

מה תהיה עלות חישוב התוצאה:

1. בהנחה שהאינדקס מקובץ ו $V(S,B)=1,000$.

א. 1,020	ב. 1,200	ג. 11,000	ד. 33,000
ה. 101,000	ו. 111,000	ז. 1,001,000	ח. 10,010,000
ט. אין מספיק נתונים כדי לחשב	י. אחר		
2. בהנחה שהאינדקס אינו מקובץ ו $V(S,B)=100,000$.

א. 1,020	ב. 1,200	ג. 11,000	ד. 33,000
ה. 101,000	ו. 111,000	ז. 1,001,000	ח. 10,010,000
ט. אין מספיק נתונים כדי לחשב	י. אחר		
3. בהנחה שאטריבוט B הוא מפתח ביחס S .

א. 1,020	ב. 1,200	ג. 11,000	ד. 33,000
ה. 101,000	ו. 111,000	ז. 1,001,000	ח. 10,010,000
ט. אין מספיק נתונים כדי לחשב	י. אחר		
4. בהנחה שאטריבוט B הוא מפתח ביחס S ובחוצץ (buffer) יש 1002 בלוקים?

א. 1,020	ב. 1,200	ג. 11,000	ד. 33,000
ה. 101,000	ו. 111,000	ז. 1,001,000	ח. 10,010,000
ט. אין מספיק נתונים כדי לחשב	י. אחר		

שאלה 10 (6 נקודות)

שורה ב-WAL היא בעלת התבנית הבאה: $\langle \text{row}, \text{tId}, \text{objId}, \text{old}, \text{new}, \text{prev} \rangle$ כאשר:

- row = מספר השורה בלוג.
- tId = מספר הטרינזקציה שבצעה שינוי.
- objId = מזהה האובייקט שהשתנה.
- old = ערך ישן של האובייקט.
- new = ערך חדש של האובייקט.
- prev = מספר שורה קודמת של הטרינזקציה בלוג.

לכל מדיניות של ניהול החוצץ (buffer), הקף את כל השדות בשורת ה-WAL שההכרחיות לשחזור נכון במקרה נפילה.

- | | | | | | | | |
|-----|-----|-------|-----|-----|------|---------------------|----|
| row | tId | objId | old | new | prev | : No-Steal No-Force | א. |
| row | tId | objId | old | new | prev | : No-Steal Force | ב. |
| row | tId | objId | old | new | prev | : Steal No-Force | ג. |
| row | tId | objId | old | new | prev | : Steal Force | ד. |

שאלה 11 (8 נקודות)

רוצים לחשב את הביטוי $\sigma_{A < 10 \wedge D < 5} \pi_{A,D} (R(A, B, C) \bowtie S(B, D))$. ההטלה היא ללא מחיקת

כפילויות. גודלי היחסים הם $B(S)=90$, $B(R)=1,000$. גודל כל אחד מהאטריבוטים הוא 10 bytes וגודל בלוק הוא 3,000 bytes. אין אינדקסים ואסור לבנות אותם. כמו כן, $V(S,B)=100$ וידוע ש B הוא מפתח ביחס R . בחוצץ (buffer) יש 22 בלוקים.

1. מה יהיה מספר הרשומות בתוצאה?

- א. 1,000 ב. 1,500 ג. 2,250 ד. 3,000
ה. 4,500 ו. 9,000 ז. 13,500 ח. 67,500
ט. אין מספיק נתונים כדי לחשב י. אחר _____

2. מה יהיה גודל התוצאה בבלוקים?

- א. 7 ב. 10 ג. 15 ד. 20
ה. 30 ו. 45 ז. 90 ח. 135
ט. אין מספיק נתונים כדי לחשב י. אחר _____

3. מהו האלגוריתם הכי יעיל לחישוב התוצאה? אם יש כמה תשובות נכונות יש להקיף את כולם.

- א. Block nested loop join
ב. Index nested loop join
ג. Sort merge join
ד. Hash join
ה. אחר _____

4. מה עלות החישוב היעיל ביותר?

- א. 1,090 ב. 1,596 ג. 1,818 ד. 3,270
ה. 4,590 ו. 5,090 ז. 5,500 ח. 7,450
ט. אין מספיק נתונים כדי לחשב י. אחר _____

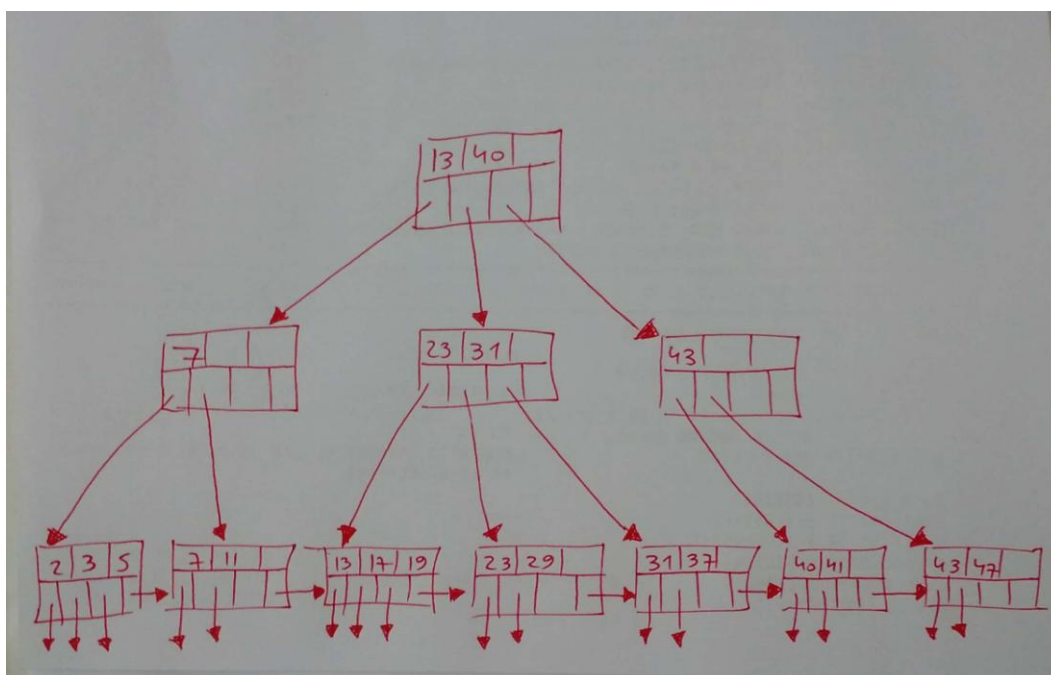
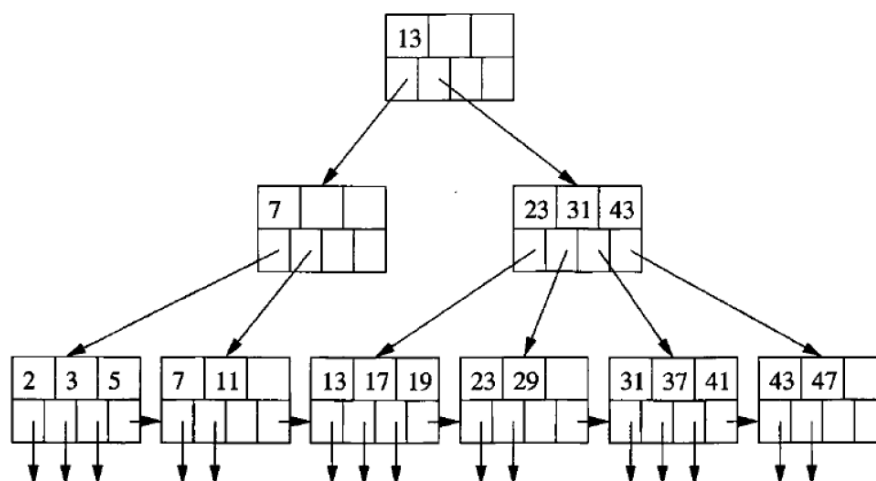
5. מה תהיה עלות החישוב היעיל ביותר אם $B(S)=60$.

- א. 1,090 ב. 1,596 ג. 1,818 ד. 3,270
ה. 4,590 ו. 5,090 ז. 5,500 ח. 7,450
ט. אין מספיק נתונים כדי לחשב י. 1060

שאלה 12 (6 נקודות)

1. תאר בקצרה מהו אינדקס דליל:
 אינדקס דליל הוא אינדקס מקובץ (כלומר הטבלה ממוינת לפי מפתח החיפוש של האינדקס) שבו
 עבור כל בלוק נתונים בטבלה יש כניסה/מצביע אחד מהאינדקס, ולא עבור כל שורה בטבלה בנפרד.
 כאשר מתבצע חיפוש של מפתח באינדקס, האינדקס מפנה אל הבלוק שבו המפתח אמור להופיע.

2. נתון ה B+ Tree הבא עליו מבצעים הכנסה של הערך 40.
 צייר את העץ המתקבל לאחר ההכנסה. (ניתן לבצע שינויים על העץ הנתון במקום לצייר מחדש)



שאלה 13 (8 נקודות)

א. תן דוגמה קצרה ביותר של תזמון שאינו בר-התאוששות (not recoverable) אבל כן ניתן להשגה על ידי 2PL. ציין גם את זמני הנעילה ושחרור של הפריטים. אם לא ניתן למצוא תזמון כזה, נמק בקצרה את תשובתך.

T1	T2
L(A)	
W(A)	
U(A)	
	L(A)
	R(A)
	U(A)
	Commit
Commit	

ב. תן דוגמה קצרה ביותר של תזמון שאינו בר-התאוששות (not recoverable) אבל כן ניתן להשגה על ידי Strict 2PL. ציין גם את זמני הנעילה ושחרור של הפריטים. אם לא ניתן למצוא תזמון כזה, נמק בקצרה את תשובתך.

אין תזמון כזה. מכיוון שטרנזקציה משחררת מנעולים רק כאשר היא מבצעת commit ב Strict 2PL כל טרנזקציה יכולה לתפוס מנעולים רק לאחר שטרנזקציה שהיא קראה את שינויה כבר בצעה commit. לכן בפרט, טרנזקציה תבצע commit רק לאחר שטרנזקציה שהיא קראה את שינויה כבר בצעה commit.

שאלה 14 (8 נקודות)

נתון התזמון הבא :

	T1	T2	T3	T4
1)	R(X)			
2)			R(X)	
3)		W(X)		
4)	W(X)			
5)			W(Y)	
6)		W(Y)		
7)				W(X)

א. האם התזמון בר סידור קונפליקטים (conflict serializable)? ☒ כן ☐ לא

ב. האם התזמון בר סידור מבטים (view serializable)? ☐ כן ☒ לא

ג. האם התזמון יכול להיווצר על ידי פרוטוקול חותמות הזמן כאשר

$$TS(T1) = 1, TS(T2) = 2, TS(T3) = 3, TS(T4) = 4$$

☒ כן ☐ לא

3

אם ענית לא, באיזה שורה הפרוטוקול ייכשל?

ד. האם התזמון יכול להיווצר על ידי פרוטוקול חותמות הזמן כאשר

$$TS(T1) = 2, TS(T2) = 3, TS(T3) = 1, TS(T4) = 4$$

☐ כן ☒ לא

אם ענית לא, באיזה שורה הפרוטוקול ייכשל?