

מבחן במסדי נתונים, מועד א', תש"פ

מס' הקורס 67506

מרצה: פרופ' שרה כהן

תאריך הבחינה: כג' בשבט, תש"פ (18.2.2020)

אסור לכתוב בשוליים!

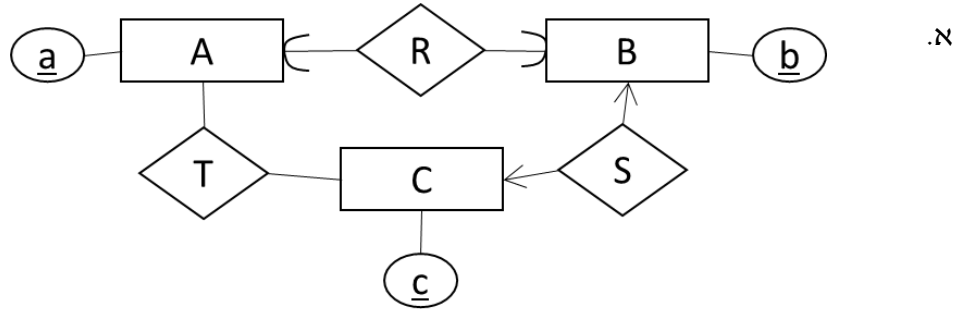
הנחיות כלליות

- משך הבחינה: 3 שעות. חומר סגור
- ענו על כל השאלות (בבחינה יש 14 שאלות רגילות ושאלת בונוס)
- סמנו תשובה אחת בלבד כשצריך לבחור בין כמה אפשרויות, אלא אם נאמר אחרת
- אל תבזבזו זמן על מתן הסברים אם לא התבקשתם! ענו רק על מה שנשאלתם
- אפשר להשתמש במחשבוני פשוטים (ללא זיכרון לאגירת נוסחאות)

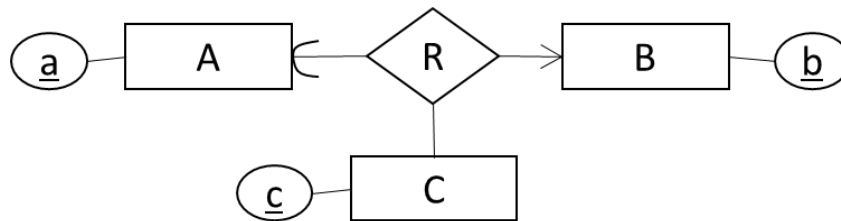
בהצלחה!

שאלה 1 (6 נקודות)

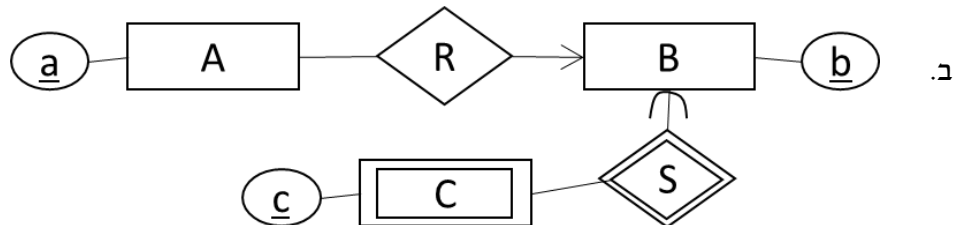
בכל אחד מהסעיפים הבאים נתונה דיאגרמת ישויות קשרים. כמו כן, בכל הסעיפים ידוע שיש x ישויות בקבוצת הישויות A , y ישויות בקבוצת הישויות B ו- z ישויות בקבוצת הישויות C . בכל סעיף, סמן את כל התשובות הנכונות. (למשל, אם בהכרח $x=z$ יש לסמן גם $z \leq x$ וגם $x \leq z$).



- | | | |
|---------------------|---------------------|---------------------|
| 1. בהכרח $x \leq y$ | 2. בהכרח $y \leq x$ | 3. בהכרח $z \leq x$ |
| 4. בהכרח $x \leq z$ | 5. בהכרח $y \leq z$ | 6. בהכרח $z \leq y$ |
7. אף אחת מהאפשרויות הקודמות אינן נכונות



- | | | |
|---------------------|---------------------|---------------------|
| 1. בהכרח $x \leq y$ | 2. בהכרח $y \leq x$ | 3. בהכרח $z \leq x$ |
| 4. בהכרח $x \leq z$ | 5. בהכרח $y \leq z$ | 6. בהכרח $z \leq y$ |
7. אף אחת מהאפשרויות הקודמות אינן נכונות



- | | | |
|---------------------|---------------------|---------------------|
| 1. בהכרח $x \leq y$ | 2. בהכרח $y \leq x$ | 3. בהכרח $z \leq x$ |
| 4. בהכרח $x \leq z$ | 5. בהכרח $y \leq z$ | 6. בהכרח $z \leq y$ |
7. אף אחת מהאפשרויות הקודמות אינן נכונות

שאלה 2 (6 נקודות)

נתון היחס הבא, המציין שלשות של שם של בן אדם, שם של תחביב, וכמות השעות השבועיות שהאדם משקיע בתחביב:

Hobby(pname, hname, time)

כתוב שאילתה באלגברה רלציונית אשר מחזירה את שמות האנשים העוסקים בכל התחביבים של "Donald Trump" ובכל תחביב כזה, משקיעים לפחות אותו כמות שעות כמו "Donald Trump".

שאלה 3 (8 נקודות)

נתייחס שוב לטבלה מהשאלה הקודמת :

Hobby(pname, hname, time)

ניתן להניח שהטבלה אינה ריקה ואין בה ערכי NULL.

נאמר שתחביב הוא **נפוץ** אם לפחות חצי מהאנשים בטבלה עוסקים בתחביב. כתבו שאילתה ב- SQL אשר מחזירה את שמות האנשים אשר עוסקים רק בתחביבים נפוצים.

שאלה 4 (6 נקודות)
מבצעים את סדרת הפעולות התקינות הבאה:

```
CREATE table R(a integer, v integer);
INSERT into R VALUES (2, 20);
INSERT into R VALUES (2, 30);
INSERT into R VALUES (4, 40);
INSERT into R VALUES (4);

SELECT a, c, count(*)
FROM   (SELECT R1.a as a, count(distinct R1.v) as c
        FROM R R1, R R2
        WHERE R1.a >= R2.a
        GROUP BY R1.a, R2.a
        HAVING count(*) > 1) T
GROUP BY a, c;
```

אילו שורות יוחזרו על ידי השאילתה?

שאלה 5 (6 נקודות)

נתון יחס $R(A, B, C)$, ורוצים לבנות אינדקס שייעל את זמן הריצה של שאילתות על הטבלה. ניתן לבנות כל אינדקס עם כל מספר של אטריבוטים.

1. סמנו את כל האינדקסים שעשויים לייעל את חישוב השאילתה $(R)_{\pi_A \sigma_{C=c1}}$?

- | | | | | |
|------------|---------------|---------------|------------------|------------------|
| א. $R(A)$ | ב. $R(A, B)$ | ג. $R(A, C)$ | ד. $R(A, B, C)$ | ה. $R(A, C, B)$ |
| ו. $R(B)$ | ז. $R(B, A)$ | ח. $R(B, C)$ | ט. $R(B, A, C)$ | י. $R(B, C, A)$ |
| יא. $R(C)$ | יב. $R(C, A)$ | יג. $R(C, B)$ | יד. $R(C, A, B)$ | טו. $R(C, B, A)$ |

2. סמנו את האינדקס שעשוי להיות הטוב ביותר לייעל את חישוב השאילתה $(R)_{\pi_A \sigma_{C=c1}}$. אם יש כמה אינדקסים שיהיו, לפי דעתך, הטובים ביותר במידה שווה, סמנו את כולם.

- | | | | | |
|------------|---------------|---------------|------------------|------------------|
| א. $R(A)$ | ב. $R(A, B)$ | ג. $R(A, C)$ | ד. $R(A, B, C)$ | ה. $R(A, C, B)$ |
| ו. $R(B)$ | ז. $R(B, A)$ | ח. $R(B, C)$ | ט. $R(B, A, C)$ | י. $R(B, C, A)$ |
| יא. $R(C)$ | יב. $R(C, A)$ | יג. $R(C, B)$ | יד. $R(C, A, B)$ | טו. $R(C, B, A)$ |

3. סמנו את האינדקס שעשוי להיות הטוב ביותר לייעל את חישוב השאילתה $(R)_{\pi_A \sigma_{B < 10 \wedge C = c1}}$. אם יש כמה אינדקסים שיהיו, לפי דעתך, הטובים ביותר במידה שווה, סמנו את כולם.

- | | | | | |
|------------|---------------|---------------|------------------|------------------|
| א. $R(A)$ | ב. $R(A, B)$ | ג. $R(A, C)$ | ד. $R(A, B, C)$ | ה. $R(A, C, B)$ |
| ו. $R(B)$ | ז. $R(B, A)$ | ח. $R(B, C)$ | ט. $R(B, A, C)$ | י. $R(B, C, A)$ |
| יא. $R(C)$ | יב. $R(C, A)$ | יג. $R(C, B)$ | יד. $R(C, A, B)$ | טו. $R(C, B, A)$ |

שאלה 6 (8 נקודות)

נרצה לחשב את הביטוי $\sigma_{A < 10} R(A, C) \bowtie S(C, D)$.

נתון $B(R)=1,000$, $B(S)=200$. ידוע שכל בלוק של S מכיל 100 שורות, וכל בלוק של R מכיל 20 שורות. כמו כן, נתון שמספר הבלוקים בחוצץ (buffer) הוא 33.

1. מה תהיה עלות החישוב היעיל ביותר של הביטוי הנ"ל בעזרת אלגוריתם Block-Nested-Loops-Join?

א. 2,662	ב. 2,268	ג. 3,000	ד. 3,200	ה. 3,204
ו. 3,332	ז. 3,336	ח. 3,600	ט. 6,000	י. 6,200
יא. 7,200	יב. לא ניתן להשתמש באלגוריתם זה	יג. אחר _____		

2. מה תהיה עלות החישוב היעיל ביותר של הביטוי הנ"ל בעזרת אלגוריתם Hash-Join?

א. 2,662	ב. 2,268	ג. 3,000	ד. 3,200	ה. 3,204
ו. 3,332	ז. 3,336	ח. 3,600	ט. 6,000	י. 6,200
יא. 7,200	יב. לא ניתן להשתמש באלגוריתם זה	יג. אחר _____		

3. מה תהיה עלות החישוב היעיל ביותר של הביטוי הנ"ל בעזרת אלגוריתם Sort-Merge-Join?

א. 2,662	ב. 2,268	ג. 3,000	ד. 3,200	ה. 3,204
ו. 3,332	ז. 3,336	ח. 3,600	ט. 6,000	י. 6,200
יא. 7,200	יב. לא ניתן להשתמש באלגוריתם זה	יג. אחר _____		

שאלה 7 (8 נקודות)

רוצים לחשב את הביטוי $(R(A, C) \bowtie S(C, D, E)) \pi_{A,D} \sigma_{A=8}$. ההטלה היא ללא מחיקת כפילויות. גודלי

היחסים הם $B(S)=100$, $B(R)=3,000$. גודל כל אחד מהאטריבוטים הוא 10 bytes וגודל בלוק הוא 1,000 bytes. אין אינדקסים ואסור לבנות אותם. כמו כן, $V(S,C)=10$ $V(R,A)=5$ $V(S,D)=100$ $V(R,C)=200$. בחוצץ (buffer) יש 68 בלוקים.

1. מה יהיה מספר הרשומות בתוצאה?

- א. 3,300 ב. 3,400 ג. 30,000 ד. 150,000
ה. 495,000 ו. 510,000 ז. 2,475,000 ח. 99,000,000

ט. אין מספיק נתונים כדי לחשב י. אחר _____

2. מה יהיה גודל התוצאה בבלוקים?

- א. 100 ב. 3,000 ג. 9,900 ד. 10,200
ה. 14,559 ו. 15,000 ז. 19,800 ח. 49,500

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. 3,100 ב. 4,000 ג. 4,432 ד. 4,500
ה. 5,764 ו. 6,100 ז. 9,300 ח. 15,500

ט. אין מספיק נתונים כדי לחשב י. אחר _____

שאלה 8 (8 נקודות)

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

$BCG \rightarrow D, A \rightarrow D, AD \rightarrow E, DE \rightarrow B, B \rightarrow GA, G \rightarrow A$

א. כתבו את כל המפתחות של R .

ב. האם R ב-3NF? כן לא

האם R ב-BCNF? כן לא

ג. מצאו כיסוי מינימאלי של F :

שאלה 9 (8 נקודות)

נתונה אותה סכמה $R=(A,B,C,D,E,G)$ ואותם תלויות פונקציונליות מהשאלה הקודמת :

$$BCG \rightarrow D, A \rightarrow D, AD \rightarrow E, DE \rightarrow B, B \rightarrow GA, G \rightarrow A$$

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

א. האם R_1 ב-3NF?	כן	לא
האם R_1 ב-BCNF?	כן	לא
ב. האם R_2 ב-3NF?	כן	לא
האם R_2 ב-BCNF?	כן	לא

ג. לכל אחד מהתלויות, הקף "כן" אם התלות נשמרת על ידי הפירוק או "לא" אם התלות אינה נשמרת על ידי הפירוק :

1. $BCG \rightarrow D$	כן	לא
2. $A \rightarrow D$	כן	לא
3. $AD \rightarrow E$	כן	לא
4. $DE \rightarrow B$	כן	לא
5. $B \rightarrow GA$	כן	לא
6. $G \rightarrow A$	כן	לא

שאלה 10 (8 נקודות)

א. נתון יחס $R=(A, B, C, D, E)$ וקבוצה F של תלויות פונקציונליים F . ידוע שצד ימין של כל אחד מהתלויות ב- F הוא בדיוק האטריבט A . הוכח או תן דוגמה נגדית: R ב- $BCNF$ אם ורק אם R ב- $3NF$.

ב. נתון יחס $R=(A, B, C, D, E)$ וקבוצה F של תלויות פונקציונליים F . ידוע שצד שמאל של כל אחד מהתלויות ב- F הוא בדיוק האטריבט A . הוכח או תן דוגמה נגדית: R ב- $BCNF$ אם ורק אם R ב- $3NF$.

שאלה 11 (8 נקודות)

נתון התזמון הבא :

	T1	T2	T3
1)	BEGIN		
2)		BEGIN	
3)			BEGIN
4)	W(A=1)		
5)		R(A)	
6)			W(A=3)
7)		W(A=A+5)	
8)			Commit
9)		Commit	
10)	Commit		

בתזמון ציינו במפורש את הערכים הנכתבים בפעולות הכתיבה. נניח שהערך המקורי של A לפני תחילת התזמון הוא 0.

א. נניח שמסד הנתונים מריץ את פרוטוקול חותמות זמן הפשוט ללא כלל הכתיבה של Thomas.

1. האם כל הטריזקציות יסיימו בהצלחה? אם לא, אלו טריזקציות יפיל (יעשה abort) המסד? הקף את כל התשובות הנכונות.

יסיים בהצלחה T1 יופל T2 יופל T3 יופל

2. מה יהיה הערך הסופי של A לאחר סיום התזמון? _____

ב. נניח שמסד הנתונים מריץ את פרוטוקול חותמות זמן הפשוט עם כלל הכתיבה של Thomas.

1. האם כל הטריזקציות יסיימו בהצלחה? אם לא, אלו טריזקציות יפיל (יעשה abort) המסד? הקף את כל התשובות הנכונות.

יסיים בהצלחה T1 יופל T2 יופל T3 יופל

2. מה יהיה הערך הסופי של A לאחר סיום התזמון? _____

ג. נניח שמסד הנתונים מריץ את הפרוטוקול האופטימי. נניח גם ששלבי ה- validation וה- apply נעשים ביחד ברגע ה commit.

1. האם כל הטריזקציות יסיימו בהצלחה? אם לא, אלו טריזקציות יפיל (יעשה abort) המסד? הקף את כל התשובות הנכונות.

יסיים בהצלחה T1 יופל T2 יופל T3 יופל

2. מה יהיה הערך הסופי של A לאחר סיום התזמון? _____

שאלה 12 (8 נקודות)

נתון תזמון עם פעולה אחת חסרה :

$$W_1(A), R_2(B), ?, W_1(C), R_2(A)$$

א. ידוע שהתזמון אינו בר סידור קונפליקטים. רשום את **כל** האפשרויות עבור הפעולה החסרה. לכל אפשרות יש לציין את הפעולה (קריאה או כתיבה), את האובייקט שעליו מבוצע הפעולה (A, B, C) או אובייקט חדש (D) ואת הטרגיקציה שמבצעת את הפעולה (T1, T2 או טרגיקציה חדשה T3). אם אין אפשרות שמקיימת את ההנחה על התזמון, יש לרשום "אין אפשרות".

ב. ידוע שהתזמון יכול להתקבל על ידי פרוטוקול 2PL אבל לא יכול להתקבל על ידי פרוטוקול 2PL המחמיר (Strict 2PL). רשום את **כל** האפשרויות עבור הפעולה החסרה. לכל אפשרות יש לציין את הפעולה (קריאה או כתיבה), את האובייקט שעליו מבוצע הפעולה (A, B, C או אובייקט חדש D) ואת הטרגיקציה שמבצעת את הפעולה (T1, T2 או טרגיקציה חדשה T3). אם אין אפשרות שמקיימת את ההנחה על התזמון, יש לרשום "אין אפשרות".

שאלה 13 (6 נקודות)

מסד נתונים משתמש בשיטת ההתאוששות של ARIES. יהיה T קבוצת הטרנזקציות בטבלת הטרנזקציות הפעילות (ATT) לפני נפילת המערכת ויהיה T' קבוצת הטרנזקציות בטבלת הטרנזקציות הפעילות בסוף שלב האנליזה בתהליך ההתאוששות. באופן דומה, יהיה P קבוצת הדפים המלוכלכים בטבלת הדפים המלוכלכים (DPT) לפני נפילת המערכת ויהיה P' קבוצת הדפים המלוכלכים בטבלת הדפים המלוכלכים בסוף שלב האנליזה בתהליך ההתאוששות.

לכל אחד מהסעיפים הבאים ציין אם הטענה נכונה.

א. בהכרח $T \subseteq T'$	נכון	לא נכון
ב. בהכרח $T' \subseteq T$	נכון	לא נכון
ג. בהכרח $P \subseteq P'$	נכון	לא נכון
ד. בהכרח $P' \subseteq P$	נכון	לא נכון

שאלה 14 (6 נקודות)

נתון מעגל chord בגודל 64, כלומר עם מזההים מ-0 עד 63. נניח שהמחשבים 5, 10, 15, 18, 23, 26, 30, 40 קיימים ברגע נתון במעגל.

רשום את תוכן טבלת האצבע (finger table) עבור מחשב (=קדקד) 23.



שאלת בונוס (3 נקודות)

הסבר אחד מבין המושגים הבאים : sparse index ,clustered index

המושג: _____

ההסבר:

