

מבחן במסדי נתונים, מועד א', תשע"ט

מס' הקורס 67506

תאריך הבחינה: ה' באדר א, תשע"ט (10.2.2019) מרצה: פרופ' שרה כהן

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 תשובות לא לעניין! ענו רק על מה שנשאלתם
- אפשר להשתמש במחשבוני פשוטים (ללא זיכרון לאגירת נוסחאות או מידע כלשהו)

בהצלחה!

שאלה 1 (6 נקודות)

בכל אחד מהסעיפים הבאים נתונה דיאגרמת ישויות קשרים. כמו כן, בכל הסעיפים ידוע שיש n ישויות בקבוצת הישויות S ו- m ישויות בקבוצת הישויות T . בכל סעיף, סמן את כל התשובות הנכונות.

א.



1. בהכרח $n \leq m$
2. בהכרח $m \leq n$
3. שתי האפשרויות הקודמות אינן נכונות

ב.



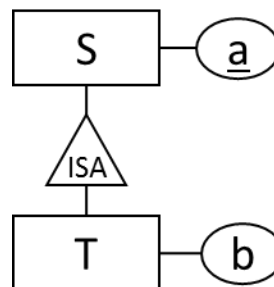
1. בהכרח $n \leq m$
2. בהכרח $m \leq n$
3. שתי האפשרויות הקודמות אינן נכונות

ג.



1. בהכרח $n \leq m$
2. בהכרח $m \leq n$
3. שתי האפשרויות הקודמות אינן נכונות

ד.



1. בהכרח $n \leq m$
2. בהכרח $m \leq n$
3. שתי האפשרויות הקודמות אינן נכונות

שאלה 2 (6 נקודות)

נתון היחס הבא, המציין שלשות של מספר סטודנט, מספר קורס וציון סופי של הסטודנט בקורס:

Study(sid, cid, grade)

ציון של פחות מ-60 הוא ציון נכשל בקורס. כתוב שאילתה באלגברה רלציונית אשר מחזירה את השורות (s,c,g) מתוך Study עבורם g ציון נכשל וגם הסטודנט s לא נכשל באף קורס אחר מלבד c.

שאלה 3 (6 נקודות)

מבצעים את סדרת הפעולות הבאות:

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

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

select R.a, R.b, R.c, count(*)
from R natural left outer join S
group by R.a, R.b, R.c;
```

אילו שורות יוחזרו על ידי השאילתה?

שאלה 4 (8 נקודות)

להלן הגדרת יחס בבסיס נתונים כלשהו :

`Studies(studentId, courseID)`

הטבלה Studies מכילה שורות שמציינות את הקורסים אשר סטודנטים לומדים. ניתן להניח שהטבלה אינה מכילה ערכי NULL.

כתוב את השאילתה הבאה בSQL :

מצא את כל הזוגות הסטודנטים אשר לומדים בדיוק את אותם קורסים.

שאלה 5 (6 נקודות)

סמנו את כל הטענות הנכונות:

- גודל החוצץ (buffer) לא משפיע על עלות החישוב של אלגוריתם Index-nested-loops join.
- בהנחה שקיים אינדקס על אטריבוט הצירוף, Index-nested-loops join בהכרח יהיה יותר יעיל מאלגוריתם Block-nested-loops-Join.
- גודל התוצאה של צירוף תלוי בדרך החישוב שנבחרה לחישוב הצירוף.
- בהנחה שהתנאים לאלגוריתם Hash-Join מתקיימים, הוא תמיד יעיל יותר מאלגוריתם Block-nested-loops-Join.
- אם גודל החוצץ (buffer) מספיק גדול כדי להכיל את שני הטבלאות R ו S ונשאר בלוק נוסף, אז הדרך היעילה ביותר לחשב את הצירוף של R ו S היא Block-nested-loops-Join.
- אף אחת מהטענות לא נכונה.

שאלה 6 (8 נקודות)

- נתון קובץ עם מיליון בלוקים ($N=1,000,000$). אנו רוצים למיין אותו בעזרת מיון חיצוני (External Sort). נתון שמספר הבלוקים בbuffer הוא B .
- מה הגודל המינימלי של B הנדרש כדי למיין את הקובץ הנ"ל בשני מעברים (passes) בלבד, כלומר מעבר אחד לחלוקה ומעבר אחד בלבד למיזוג?

- | | | | |
|----------|--------------|-------------------------------|----------|
| א. 32 | ב. 33 | ג. 34 | ד. 99 |
| ה. 101 | ו. 102 | ז. 999 | ח. 1,000 |
| ט. 1,001 | י. 1,000,000 | יא. אין מספיק נתונים כדי לחשב | |
| יב. אחר | | | |

- מה הגודל המינימלי של B הנדרש כדי למיין את הקובץ הנ"ל בשלושה מעברים (passes) בלבד, כלומר מעבר אחד לחלוקה ושני מעברים בלבד למיזוג?

- | | | | |
|----------|--------------|-------------------------------|----------|
| א. 32 | ב. 33 | ג. 34 | ד. 99 |
| ה. 101 | ו. 102 | ז. 999 | ח. 1,000 |
| ט. 1,001 | י. 1,000,000 | יא. אין מספיק נתונים כדי לחשב | |
| יב. אחר | | | |

- כמה איטרציות של מיזוג יידרשו בכדי למיין את הקובץ הנ"ל אם נתון ש $B=6$?



- מה עלות ה I/O הכוללת של מיון הקובץ אם נתון ש $B=6$?

- | | | |
|---------------|---------------|---------------|
| א. 8,000,000 | ב. 9,000,000 | ג. 16,000,000 |
| ד. 17,000,000 | ה. 18,000,000 | ו. 22,000,000 |

שאלה 7 (8 נקודות)

רוצים לחשב את הביטוי $\sigma_{A=11 \wedge C < 3} (R(A, B) \bowtie S(B, C))$.

גודלי היחסים הם $B(R)=5,000$, $B(S)=300$. בכל בלוק של R יש 10 רשומות של יחס R, ובכל בלוק של S יש 5 רשומות. ליחס R יש שני אינדקסים עם עלות גישה זניחה: אחד על אטריבוט A ואחד על אטריבוט B. כמו כן, ידוע ש B הוא מפתח ביחס R, וכן $V(S, B)=20$, $V(R, A)=100$ בחוצץ (buffer) יש 10 בלוקים.

1. מה יהיה גודל התוצאה בבלוקים של הביטוי $\sigma_{C < 3} S(B, C)$?

א. 50	ב. 100	ג. 200	ד. 300
ה. 500	ו. 1,500	ז. 5,000	ח. 50,000
ט. אין מספיק נתונים כדי לחשב	י. אחר		

2. מה יהיה גודל התוצאה בבלוקים של הביטוי $\sigma_{A=11} R(A, B)$?

א. 50	ב. 100	ג. 200	ד. 300
ה. 500	ו. 1,500	ז. 5,000	ח. 50,000
ט. אין מספיק נתונים כדי לחשב	י. אחר		

3. מה יהיה מספר השורות בתוצאה של הביטוי כולו $\sigma_{A=11 \wedge C < 3} (R(A, B) \bowtie S(B, C))$?

א. 5	ב. 15	ג. 30	ד. 50
ה. 100	ו. 500	ז. 1,500	ח. 12,500
ט. אין מספיק נתונים כדי לחשב	י. אחר		

4. 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

5. מה עלות החישוב היעיל ביותר?

א. 400	ב. 800	ג. 1,800	ד. 2,600
ה. 6,800	ו. 7,100	ז. 42,500	ח. 65,300
ט. אין מספיק נתונים כדי לחשב	י. אחר		

שאלה 8 (8 נקודות)

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

$$F = \{CD \rightarrow AB, C \rightarrow D, D \rightarrow EH, AE \rightarrow C, A \rightarrow C, B \rightarrow D, G \rightarrow A, H \rightarrow E\}$$

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

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

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 .

האם R_2 ב-3NF? כן לא

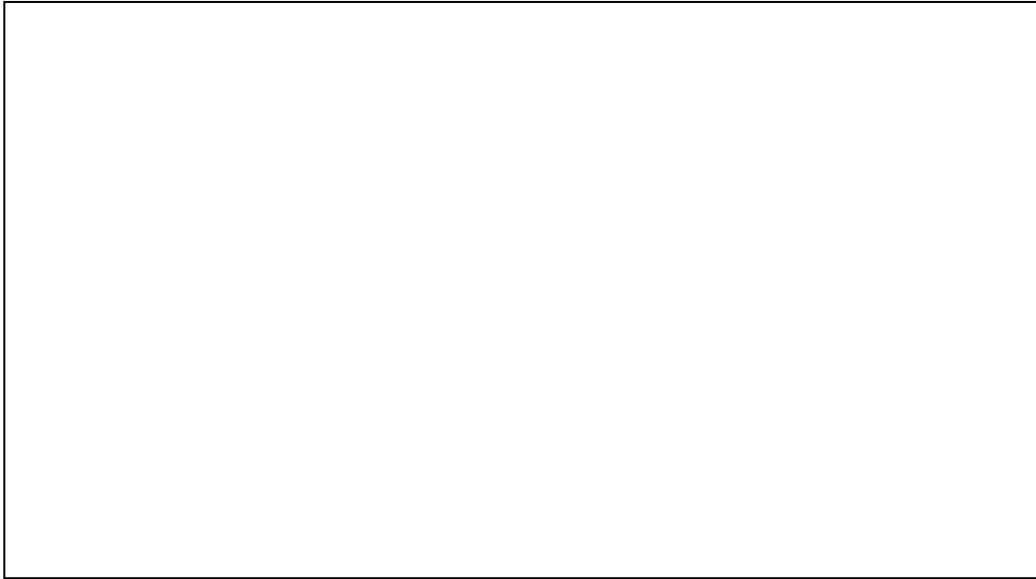
האם R_2 ב-BCNF? כן לא

שאלה 9 (8 נקודות)

א. נתון $R = (A, B, C)$ ו- F קבוצה של תלויות פונקציונליות שמכילה

$$A \rightarrow B, B \rightarrow C$$

ואולי תלויות נוספות. האם יתכן ש- R ב-3NF אך לא ב-BCNF? אם כן, תן דוגמה לתוכן של F שמקיים את הדרישה. אחרת, הוכח שזה לא ייתכן.



א. נתון $R = (A, B, C, D)$ ו- F קבוצה של תלויות פונקציונליות שמכילה

$$A \rightarrow B, B \rightarrow C$$

ואולי תלויות נוספות. האם יתכן ש- R ב-3NF אך לא ב-BCNF? אם כן, תן דוגמה לתוכן של F שמקיים את הדרישה. אחרת, הוכח שזה לא ייתכן. (שים לב להבדל ב- R לעומת הסעיף הקודם).



שאלה 10 (8 נקודות)

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

$$B \rightarrow DC, GA \rightarrow BE, C \rightarrow DGA, EB \rightarrow A, DC \rightarrow E, E \rightarrow G$$

א. כתבו את כל המפתחות של R .

ב. האם R ב-3NF? כן לא

האם R ב-BCNF? כן לא

ג. מצאו פירוק של R ל-BCNF שהוא ללא אובדן בעזרת האלגוריתם שנלמד בכיתה. יש לציין גם את הפירוק הסופי וגם את התלויות בהם השתמשתם על מנת לייצור את הפירוק.

שאלה 11 (8 נקודות)

נתון התזמון הבא :

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

בתזמון ציינו במפורש את הערכים הנכתבים בפעולות הכתיבה.

א. האם התזמון יכול להיווצר על ידי פרוטוקול 2PL? **כן** **לא**

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

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

כן **לא**

אם ענית לא, באיזה שורה הפרוטוקול ייכשל?

אם ענית כן, מה יהיו הערכים הסופיים של X ושל Y?

$$Y = \boxed{} \quad X = \boxed{}$$

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

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

כן **לא**

אם ענית לא, באיזה שורה הפרוטוקול ייכשל?

אם ענית כן, מה יהיו הערכים הסופיים של X ושל Y?

$$Y = \boxed{} \quad X = \boxed{}$$

שאלה 12 (8 נקודות)

א. תן דוגמה קצרה של תזמון ש

- מכיל לפחות קריאה מלוכלכת אחת (dirty read),
- מכיל לפחות קריאה אחת שלא ניתנת לשחזור (unrepeatable read),
- שבו כל פעולות הקריאה והכתיבה הם על אובייקט בודד A.

והתזמון הוא בר סידור קונפליקטים (conflict serializable). אם לא ניתן למצוא תזמון כזה, נמק בקצרה אך בדיוקנות את תשובתך.

ב. תן דוגמה קצרה של תזמון ש

- מכיל לפחות קריאה מלוכלכת אחת (dirty read),
- מכיל לפחות קריאה אחת שלא ניתנת לשחזור (unrepeatable read),
- שבו כל פעולות הקריאה והכתיבה הם על אובייקט בודד A.

והתזמון אינו בר סידור קונפליקטים (conflict serializable). אם לא ניתן למצוא תזמון כזה, נמק בקצרה אך בדיוקנות את תשובתך.

שאלה 13 (6 נקודות)

לכל אחד מהסעיפים הבאים ציין אם הטענה נכונה.

א. במדינות של NO-STEAL + FORCE, המסד לעולם לא יצטרך לבצע מחדש (redo) שינויים של טרנזקציה שביצעה commit בזמן התאוששות מנפילה.

נכון	לא נכון
------	---------

ב. במדינות של NO-STEAL + FORCE, המסד יצטרך לבטל (undo) שינויים של טרנזקציה שביצעה abort בזמן התאוששות מנפילה.

נכון	לא נכון
------	---------

ג. במדינות של NO-STEAL, מסד הנתונים יצטרך לשמור את כל הטבלה בזיכרון המרכזי אם טרנזקציה מעדכנת את כל השורות של הטבלה.

נכון	לא נכון
------	---------

ד. בשימוש ב-WAL, אם לא נאפשר למסד לכתוב דפים מלוכלכים לדיסק לפני שטרנזקציה ביצעה commit, אין צורך לשמור את הערכים המקוריים של אובייקטים בלוג עבור update.

נכון	לא נכון
------	---------

ה. בשיטת ההתאוששות ARIES, רושמים את הלוג לדיסק מיד לאחר כתיבת כל שורה בלוג.

נכון	לא נכון
------	---------

ו. רב מסדי הנתונים משתמשים ב-STEAL+NO-FORCE מכיוון ששיטה זו מאפשרת התאוששות מהירה ביותר.

נכון	לא נכון
------	---------

שאלה 14 (6 נקודות)

נתון שהמסד נתונים משתמש בארכיטקטורת Master-Slave על מנת לשמור העתקים של הנתונים, ובפרוטוקול 2-phase commit על מנת לשמור על עקביות ההעתקים. יש שלושה מחשבים שנסמן ב-A, B, C כאשר

- כל פריטי המידע נמצאים בשלושת המחשבים
- A הוא מחשב ה-Master

תוכנית מנסה לשנות את הערך של אובייקט X שנמצא במסד נתונים מ-1 ל-2. מחשב A מבקש ממחשבים B ו-C לשנות את הערך של X בהתאם בעזרת פרוטוקול 2-phase commit.

א. נניח ש:

- B קבל את הבקשה מ-A לעשות את השינוי
- B שלח הסכמה ל-A
- B עדיין לא קיבל הודעת commit או abort מ-A, אבל B מתקשר עם C ומגלה ש-C קיבל הודעת commit.

מה הפעולה הנכונה עבור B כעת? אם יש כמה אפשרויות שנכונות לפי הפרוטוקול, סמן את כולם:

1. B ישנה את הערך של X ל-2, כלומר יבצע commit.
2. B ו-C יחליטו ביחד ששניהם ישאירו את X בערך 1, כלומר יבצעו abort.
3. B יחכה להודעה מ-A.

ב. נניח ש:

- B קבל את הבקשה מ-A לעשות את השינוי
- B שלח הסכמה ל-A
- B עדיין לא קיבל הודעת commit או abort מ-A, אבל B מתקשר עם C ומגלה ש-C שלח הסכמה.

1. B ישנה את הערך של X ל-2, כלומר יבצע commit.
2. B ו-C יחליטו ביחד ששניהם ישאירו את X בערך 1, כלומר יבצעו abort.
3. B יחכה להודעה מ-A.

ג. נניח ש:

- B קבל את הבקשה מ-A לעשות את השינוי
- B שלח הסכמה ל-A
- B עדיין לא קיבל הודעת commit או abort מ-A, אבל B מתקשר עם C ומגלה ש-C עדיין לא שלח הודעת הסכמה או אי-הסכמה.

1. B ישנה את הערך של X ל-2, כלומר יבצע commit.
2. B ו-C יחליטו ביחד ששניהם ישאירו את X בערך 1, כלומר יבצעו abort.
3. B יחכה להודעה מ-A.

שאלת בונוס (3 נקודות)

מהו column-oriented database? תן דוגמה לסוג של שאילתה עבורו column-oriented database יהיה מהיר יותר מאשר במסד נתונים רלציוני רגיל.