

מבחן במערכות בסיסי נתונים
פרופ' טובה מילוא

סמסטר א', מועד א'
18 בפברואר, 2007
חומר סגור, שלוש שעות
סה"כ 100 נקודות

נא לכתוב את כל התשובות על הדפים
(המבחן כולל 11 דפים, כולל דף זה)

מס מחברת: _____ מס סטודנט: _____

חלק	ציון
1	
2	
3	
4	
סה"כ	

1. [30 נקודות]

מחלקת כוח אדם שומרת מידע על העובדים בטבלה הבאה
Empolyee(eid, name, dept, manager, gender)
השדה **manager** הוא מפתח זר (**foreign key**) לשדה **eid** בטבלה
Empolyee, ויכול להיות null.

חבר-לעבודה (co-worker) של עובד הוא עובד אחר בעל אותו מנהל.
כתבו שאילתת SQL שמחשבת לכל עובד במחלקת המכירות (dept="sales")
את המספר הכולל של חבריו לעבודה (בכל המחלקות). השאילתא צריכה
להחזיר את התשובה בסדר יורד של מזהי העובדים (eid).

כתבו שאילתת SQL שמחזירה את כל המנהלים שמנהלים רק עובדים
במחלקות שונות משל עצמם. (לדוגמא, אם משה הוא במחלקת מכירות ומנהל
שלושה עובדים, שניים ממחלקת שיווק ואחד ממחלקת פיתוח, עליכם להחזיר
את משה, אבל אם הוא גם מנהל עובד רביעי במחלקת מכירות אז אינכם
צריכים להחזיר אותו).

2/11

בסעיפים הבאים נתונות קבוצות של שאילתות Q1, Q2, Q3. לכל קבוצה,
ציינו אילו שאילתות הן שקולות ואיזה לא. מבנה תשובתכם צריך להיות,

לדוגמא, $Q1=Q2=Q3$, $Q1=Q2 \neq Q3$, $Q1 \neq Q2 \neq Q3 \neq Q1$, או כל צרוף
אחר במבנה דומה.

א.

Q1: **select** dept
 from Employee
 where gender = 'F'
 group by dept
 having count(*) < 5

Q2: **select distinct** x.dept
 from Employee x
 where (select count(*)
 from Employee y
 where x.dept = y.dept **AND**
 y.gender='F') < 5

Q3: **select distinct** x.dept
 from Employee x
 where x.gender = 'F' **AND**
 (select count(*)
 from Employee y
 where x.dept = y.dept **AND**
 y.gender='F') < 5

תשובה:

3/11

ב.

Q1: **select distinct** x.dept
from Employee x, Employee y, Employee z
where x.manager = y.eid **AND**
z.manager = y.eid **AND**
x.gender = 'F' **AND**
y.gender = 'F' **AND**
z.dept = 'Sales'

Q2: **select distinct** x.dept
from Employee x, Employee y, Employee z
where x.manager = y.eid **AND**
z.manager = y.eid **AND**
z.gender = 'F' **AND**
y.gender = 'F' **AND**
z.dept = 'Sales'

Q3: **select distinct** x.dept
from Employee x, Employee y, Employee z, Employee u
where x.manager = y.eid **AND**
z.manager = y.eid **AND**
u.gender = z.gender **AND**
x.gender = 'F' **AND**
y.gender = 'F' **AND**
z.dept = 'Sales' **AND**
u.dept = 'Sales'

תשובה:

Supervise(eid, name, num)

השדה **num** מציין את נספר העובדים המנוהלים ע"י העובד שמספרו **eid** ויכול להיות 0.

כתבו שאילתת SQL שממלאת את הטבלה **Supervise** על ידי שימוש בנתונים שב- **Empolyee**. (התשובה צריכה להיות פקודה (או פקודות) insert של SQL)

2.1 נתונה הטבלה $R(A, B, C, D, E)$ עם התלויות הפונקציונאליות הבאות:

$AB \rightarrow C$

$BE \rightarrow A$

$D \rightarrow E$

פרקו את הטבלה לטבלאות ב-BCNF. התשובה צריכה להכיל

(א) רשימת טבלאות והשדות בכל טבלה

(ב) המפתחות בכל אחת מהטבלאות

6/11

2.2 נתונה קבוצת תלויות פונקציונאליות FD.

נסמן ב X^+ את הסגור של קבוצת השדות X .
לכל אחת מהטענות הבאות ציינו האם היא נכונה או לא. אין צורך לנמק.

א. לכל הקבוצות Y, X : $(X \cup Y)^+ = X^+ \cup Y^+$

ב. לכל הקבוצות Y, X : אם $X \subseteq Y$ אז $X^+ \subseteq Y^+$

ב. לכל הקבוצות Y, X : אם $X^+ = X$ ו $Y^+ = Y$, אז $(X \cap Y)^+ = X \cap Y$

3.2 נניח כעת שכל התלויות הפונקציונאליות ב FD הם מהצורה $A \rightarrow B$

כש A ו B הם שמות של שדות. (כלומר גם הצד השמאלי וגם הצד הימני בתלויות מכיל שדה אחד בלבד). איזה מהתשובות הנ"ל (אם בכלל) היו משתנות ?

נתון מסמך XML המתאר את המחלקות בפקולטות השונות באוניברסיטה.
 לכל פקולטה יש שם, דיקן, וקבוצת מחלקות. לכל מחלקה יש שם, יושב-ראש,
 ורשימת סטודנטים, כמו בדוגמא הבאה:

```
<university>
  <faculty>
    <name> Exact Sciences </name>
    <dean> Wolfson </name>
    <department>
      <name> Computer Science </name>
      <chair> Fiat </chair>
      <student> Moshe Cohen </student>
      <student> Sara Levi</student>
      . . . . .
    </department>
    <department>
      <name> Mathematics </name>
      . . . . .
    </department>
    . . . . .
  </faculty>
  <faculty> (another faculty) . . . . .</faculty>
  . . . . .
</university>
```

3.1 כתוב ביטוי XPath שמחזיר את כל הסטודנטים במחלקה למדעי המחשב.

2.3 כתוב ביטוי ב XQuery שמחשב, עבור כל סטודנט, בכמה מחלקות הוא/היא מופיעים.

נתונות שתי טבלאות $R(\underline{A}, B, C)$ ו- $S(\underline{D}, E, F)$ ושאלת ה-SQL הבאה:

```
Select R.A, sum(S.F)
From R, S
Where R.B = '1234' and R.C= S.D and S.E='5678'
Group by R.A
```

יש לנו שני אינדקסים clustered על R.A ו-S.D, ושני אינדקסים unclustered על R.B ו-S.E. נניח ששתי הטבלאות גדולות ולא ניתן להחזיק אותן בזיכרון. נתונות שלוש תוכניות הביצוע הלוגיות הבאות עבור השאילתא. (האופרטור gamma כאן מבצע את ה-grouping לפי השדה A ואת סיכום ערכי השדה F לכל קבוצה)

```
P1= gammaA,sum(F) (selectB=1234 (R)) JoinC=D (selectE=5678(S))
P2= gammaA,sum(F) (selectB=1234 (R JoinC=D (selectE=5678(S))))
P3= gammaA,sum(F) (selectE=5678 (selectB=1234 (R JoinC=D S)))
```

ענו על השאלות הבאות (נמקו בקצרה - שתי שורות לכל היותר!)

4.1 באילו מתוכניות הביצוע הנ"ל ניתן להיעזר באינדקסים הקיימים לחישוב ה-join (index-join)?

4.2 באילו מתוכניות הביצוע הנ"ל ניתן להשתמש ב-sorted-merge join בלי שנצטרך למיין קודם חלק מהקלט ל-merge? (להזכירכם, לאלגוריתם שני שלבים, הראשון ממיין את רלציות הקלט והשני מבצע את ה-join. באיזה מהתוכניות אין צורך בשלב הראשון עבור אחת או יותר מהרלציות?)

10/11

4.3 אם נחליט לממש את אופרטור ה-join ע"י hash join, איזה מתוכניות הביצוע הנ"ל תהייה תמיד יותר טובה מהאחרות?

4.4 נתבונן בתוכנית הביצוע הלוגית P1 ובשתי וריאציות של ביצועים אפשריים שלה: P1' שבה ה-join מבוצע ע"י nested-loop join, ו-P1'' שבה ה-join מבוצע ע"י אלגוריתם hash-join (נניח במקרה זה שאם הקלט ל-join קטן, ה hash-join יכול להתבצע בשלמותו בזיכרון ללא צורך בכתיבת הקבוצות לדיסק). נניח שכל שאר האופרטורים מלבד ה-join ממומשים ב P1' ו P1'' באותו אופן ומחירם זהה.

באיזה מהמקרים הבאים ניתן לדעת בוודאות ש- P1' יותר יעילה מ P1'' בהתבסס אך ורק על הנתונים בשאלה וללא תלות בנתונים סטטיסטיים אחרים על הנתונים? סמנו את התשובה הנכונה בכל סעיף.

(א) אם כל הרשומות ב-R מקיימות את התנאי $R.B = '1234'$ ניתן לדעת
לא ניתן לדעת

(ב) אם יש רק רשומה אחת ב-R שמקיימת את התנאי $R.B = '1234'$ ניתן לדעת
לא ניתן לדעת

(ג) אם כל הרשומות ב-S מקיימות את התנאי $R.E = '5678'$ ניתן לדעת
לא ניתן לדעת

(ד) אם יש רק רשומה אחת ב-S שמקיימת את התנאי $R.E = '5678'$ ניתן לדעת
לא ניתן לדעת