

סמסטר חורף התש"פ

מרצה: פרופ' עודד שמואלי

מתרגלים: אסף ישורון

שובל לגזיאל

משה סבאג

מערכות מסד נתונים

236363

מועד א'

2 בפברואר 2020

פירוט החלקים והניקוד:

שאלה	נושא	ניקוד
1	ERD	10
2	RA	20
3	SQL	20
4	Design Theory	20
5	RDF	10
6	CC	20

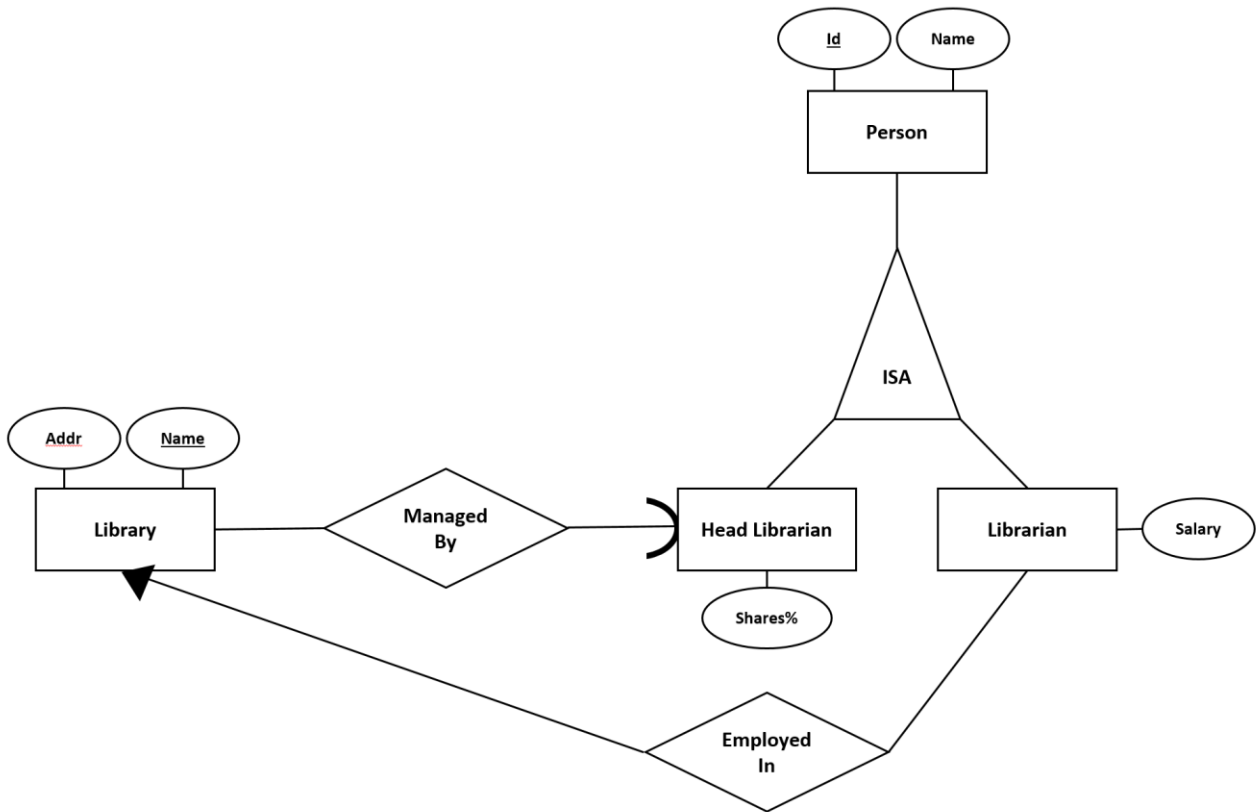
## הנחיות לנבחנים

1. כתבו את התשובות אך ורק בטופס הבחינה ובמקום המיועד להן, מחברת הטייטה לא תיבדק.
2. כל חומר עזר הכתוב על נייר בלבד מותר בשימוש.
3. אין לקבל או להעביר חומר כלשהו בזמן הבחינה.
4. יש להשתמש רק בסימנים או פונקציות שנלמדו בתרגול או בהרצאה בסמסטר זה והמופיעים בשקפים של הקורס. כל שימוש בסימון שאינו כזה מחייב הסבר מלא של משמעות הסימון.
5. משך הבחינה הינו שלוש שעות, תכננו את הזמן בהתאם.
6. אין לכתוב בעפרון.

בהצלחה!

## ERD – 10 נקודות

התבוננו בתרשים ה-ERD הבא:



א. תרגמו את התרשים (על כל מרכיביו) לסכמות של יחסים. סמנו מפתחות בקו תחתון וציינו מפתחות זרים במקומות המתאימים (7 נקודות).

Person(Id, Name)

Librarian(Id, Salary), Id references Person.Id

Head Librarian(Id, Shares%), Id references Person.Id

Employed In(Id, Name), Id references Person.Id, Name references Library.Name

Library(Name, Addr, Id), Id references Head Librarian.Id

ב. הניחו כי התרשים תורגם לטבלאות אשר מהוות חלק ממסד נתונים. הקיפו בעיגול את כל המשפטים המתארים מצבי מסד נתונים אפשריים (3 נקודות):

- a. בטבלה Librarian קיימת בדיוק רשומה אחת, ובטבלה Head Librarian אין כלל רשומות.
- b. בטבלה Library קיימת בדיוק רשומה אחת, ובטבלה Person אין כלל רשומות.
- c. בטבלה Librarian קיימות בדיוק 2 רשומות, ובטבלה Library קיימת בדיוק רשומה אחת.
- d. בטבלה Librarian קיימות בדיוק 3 רשומות, בטבלה Head Librarian קיימות בדיוק 2 רשומות ובטבלה Person קיימות בדיוק 4 רשומות.

## RA, RC – 20 נקודות

נתונות סכמות  $R(A,B,C)$  ו-  $S(C,D,E)$ .

בנוסף נתון היחס  $R_{>}(a,b)$  שמתקיים עבור כל  $a > b$ . וכמו כן, היחס  $R_{\neq}(a,b)$  שמתקיים עבור כל  $a \neq b$ .

א. (4 נק') בהינתן שאילתת ה-SQL הבאה:

```
SELECT DISTINCT B, C, D
FROM R, S
WHERE R.C = S.C AND B > 2 AND E > 4
```

a. (2 נק') נסחו שאילתת RA שקולה:

$$\pi_{B,C,D}(\sigma_{B>2 \wedge E>4}(R \bowtie S))$$

b. (2 נק') נסחו שאילתת RC שקולה:

$$\{(b,c,d) \mid \exists a,e (R(a,b,c) \wedge S(c,d,e) \wedge R_{>}(b,2) \wedge R_{>}(e,4))\}$$

ב. (6 נק') בהינתן שאילתת ה-SQL הבאה:

```
(SELECT DISTINCT B, C
FROM R, S WHERE R.C = S.C AND B > 2 AND E > 4)
UNION
(SELECT DISTINCT B, C FROM R WHERE A=B)
```

a. (2 נק') נסחו שאילתת RA שקולה:

$$(\pi_{B,C} \sigma_{B>2 \wedge E>4}(R \bowtie S)) \cup (\pi_{B,C} \sigma_{A=B} R)$$

b. (4 נק') נסחו שאילתת RC שקולה:

$$\{(b,c) \mid \exists a,d,e (R(a,b,c) \wedge S(c,d,e) \wedge R_{>}(e,4) \wedge R_{>}(b,2) \vee R(b,b,c))\}$$

ג. (7 נק') נתון:

i. השאילתות:

```
SELECT COUNT(*) FROM R;
SELECT COUNT(*) FROM (SELECT DISTINCT * FROM R) AS RDistinct;
```

מחזירות שתייהן  $n$ .

ii. השאילתות:

```
SELECT COUNT(*) FROM S;
SELECT COUNT(*) FROM (SELECT DISTINCT * FROM S) AS SDistinct;
```

מחזירות שתייהן  $m$ .

iii. השאילתא:

$$\pi_C S$$

מחזירה  $k$  רשומות.

עבור כל אחת מהשאלות הבאות, ציינו אם הביטוי של השאלתא מוגדר היטב ב-RA, ואם כן כתבו מהו המספר המקסימלי של Nיות (tuples) שיכולות לחזור מהרצת השאלתא:

a.  $R \div S$  (1 נק')

מוגדר/ לא מוגדר (הקף),  
אם בחרתם מוגדר, ציינו את מספר ה-Nיות (tuples) המקסימלי:

b.  $R \div \pi_C S$  (3 נק')

מוגדר/ לא מוגדר (הקף),  
אם בחרתם מוגדר, ציינו את מספר ה-Nיות (tuples) המקסימלי:  $\begin{bmatrix} n \\ k \end{bmatrix}$

c.  $\sigma_{A>5} R \bowtie \pi_E S$  (3 נק')

מוגדר/ לא מוגדר (הקף),  
אם בחרתם מוגדר, ציינו את מספר ה-Nיות (tuples) המקסימלי:  $n$

ד. (3 נק') עבור שאלתת ה-RA הבאה:

$$\pi_B(\sigma_{A=D}(\pi_{A,B}R \times \pi_{D,E}S)) \setminus \pi_B(\sigma_{B=2}R)$$

בחרו מבין האפשרויות את שאלתת ה-RC השקולה לשאלתת הנתונה (בחרו אפשרות אחת בלבד):

a.  $\{b \mid \exists a, c, d, e \ R(a, b, c) \wedge S(c, d, e) \wedge R_{\neq}(b, 2)\}$

b.  $\{b \mid \exists a, c, e \ R(a, b, c) \wedge S(c, a, e) \wedge \neg R(a, 2, c)\}$

c.  $\{b \mid \exists a, c1, c2, e \ R(a, b, c1) \wedge S(c2, a, e) \wedge R_{\neq}(b, 2)\}$

d.  $\{b \mid \exists a, c1, c2, e \ R(a, b, c1) \wedge S(c2, a, e) \wedge \neg R(a, 2, c1)\}$

e. שאלתת ה-RA הנתונה לא ניתנת לביטוי על ידי שאלתת RC.

## SQL – 22 נקודות

לפניכם תיאור מסד נתונים ובו אתלטים וענפי ספורט. קראו אותו בעיון וענו על השאלות שאחריו.

Athletes Table – **Athlete**

<u>AthleteID</u> (INTEGER)	AthleteName (TEXT)	Country (TEXT)
----------------------------	--------------------	----------------

טבלת האתלטים, המכילה עבור כל אתלט את המזהה שלו, שמו ומדינתו.

Sports Table – **Sport**

<u>SportID</u> (INTEGER)	SportName (TEXT)
--------------------------	------------------

טבלת ענפי הספורט, המכילה עבור כל ענף ספורט את המזהה שלו ואת שמו.

Plays Table – **Plays**

<u>AthleteID</u> (INTEGER)	<u>SportID</u> (INTEGER)	Rank (INTEGER)
----------------------------	--------------------------	----------------

טבלת השתתפות של אתלטים בענפי ספורט, המקשרת בין אתלט לענף ספורט בו הוא משתתף. השדה Rank מציין את דירוגו העולמי של האתלט בענף הספורט.

**שימו לב, מפתחות הטבלאות מסומנים בקו תחתון.**

בעת יצירת מסד הנתונים הוגדרו ההגבלות הבאות:

i. `Plays.AthleteID`, `Plays.SportID` הינם מפתחות זרים ל-`Athlete.AthleteID`, `Sport.SportID`, בהתאמה.

ii. השדה "Rank" ב-`Play` מקיים את המגבלות "NOT NULL" ו-"CHECK (Rank > 0)".

**בנוסף, הניחו כי אף אחת מן הטבלאות אינה ריקה.**

א. (5 נקודות) כתבו שאילתת SQL המחזירה את כל שמות ענפי הספורט בהם משתתפים לכל היותר 10 אתלטים. על שמות הענפים להיות ממוינים על פי הדירוג הממוצע של האתלטים המשתתפים בהם,

ב

```
SELECT SportName
FROM Sport INNER JOIN Plays ON Sport.SportID = Plays.SportID
GROUP BY Sport.SportID
HAVING COUNT(*) <= 10
ORDER BY AVG(Rank) ASC;
```

ב. (5 נקודות) כתבו שאילתת SQL המחזירה את מספרם המזהה של האתלטים המשתתפים בהכי הרבה ענפי ספורט.

```
SELECT AthleteID
FROM Plays
GROUP BY AthleteID
HAVING COUNT(*) >= ALL
(SELECT COUNT(*) FROM Plays GROUP BY AthleteID);
```

ג. (8 נקודות) כתבו שאילתת SQL המחזירה את מספרם המזהה של האתלטים המשתתפים בדיוק באותם ענפי ספורט כמו אתלט בעל מספר מזהה 1 (הניחו כי קיים כזה). שימו לב: עליכם להחזיר, בין היתר, את 1 עצמו.

```
SELECT AthleteID FROM Plays
GROUP BY AthleteID
HAVING COUNT(*) = (SELECT COUNT(*) FROM Plays WHERE AthleteID = 1)
EXCEPT
SELECT AthleteID FROM Plays
WHERE SportID NOT IN (SELECT SportID FROM Plays WHERE AthleteID = 1);
```

הערה: השאילתה עובדת תחת הסטנדרטים PostgreSQL 9.3, PostgreSQL 9.6, אבל לא תחת הסטנדרט MySQL 5.6

ד. (2 נקודות) הסבירו בקצרה מה היו צריכים מתכנני מסד הנתונים לשנות בהגדרת הטבלאות על מנת שכל אתלט (Athlete) יוכל להשתתף (Play) בספורט (Sport) אחד לכל היותר? הציעו את השינוי הפשוט ביותר שידאג לכך.

לשנות בהגדרת הטבלה Play את המפתח להיות רק AthleteID.

## תלויות וצורות נורמליות (20 נק')

תהא  $U = \{A, B, C, D\}$ ,  $F = \{f1, f2, f3, f4\}$  סכמה, כאשר  $F$  מהווה כיסוי מינימלי של עצמה. תהא  $G = \{f2, f3, f4\}$  תת-קבוצה של  $F$  המתקבלת ע"י השמטת התלות  $f1$ .

א. ענו בקצרה על השאלות הבאות (3 נקודות):

a. האם ייתכן ש- $F$  מהווה כיסוי מינימלי של  $G$ ? כן לא

נימוק: כי אז התלות  $f1$  מיותרת ב- $F$ .

b. האם קיימת רלציה לא ריקה המספקת את  $F$  אך לא את  $G$ ? כן לא

נימוק: כל תלויות  $G$  נמצאות גם ב- $F$ .

c. האם קיימת רלציה לא ריקה המספקת את  $G$  אך לא את  $F$ ? כן לא

נימוק: נסמן  $f1 = X \rightarrow A$ . רלציה בעלת שתי שורות המסכימות על  $X^+$  להוציא את  $A$  תקיים את הנדרש.

ב. (3 נקודות) תהא  $\{U, F\}$  סכמה,  $Z \subseteq U$  קבוצת אטריביוטים, ונגדיר:

$G = \{Z \rightarrow A \mid A \in Z_F^+\}$ . באילו מקרים מובטח כי  $F^+ = G^+$ ? סמנו את כל האפשרויות הנכונות.

א.  $F = \emptyset$

ב.  $\{U, F\}$  בצורה הנורמלית 3NF.

ג.  $\{U, F\}$  בצורה הנורמלית BCNF.

ד. קיימת קבוצת תלויות  $H$  המקיימת  $F^+ = H^+$  ובה כל התלויות הן מהצורה  $Z \rightarrow Y$  (עבור  $Z$  הנתונה ו- $Y$  כלשהי).

ה.  $Z$  מהווה מפתח על (superkey) של  $\{U, F\}$ .

ו.  $Z$  מהווה מפתח קביל (key) של  $\{U, F\}$ .

ז. קיים כיסוי מינימלי של  $F$  בו כל התלויות הן מהצורה  $Z \rightarrow Y$  (עבור  $Z$  הנתונה ו- $Y$  כלשהי).

ג. (7 נקודות) תהא  $\{U, F\}$  סכמה. עבור קבוצת אטריביוטים  $X \subseteq U$  נגדיר:

$$X_F^- = \{Z \mid Z \subseteq U, F \models Z \rightarrow X\}$$

אינטואיטיבית -  $X_F^-$  הן כל קבוצות האטריביוטים כך שכל קבוצה מהן קובעת פונקציונלית את  $X$  בהינתן  $F$ .

לדוגמא, עבור הסכמה  $\{U = \{A, B, C\}, F = \{A \rightarrow BC, C \rightarrow B\}\}$  מתקיים:

$$BC_F^- = \{A, AB, AC, ABC, BC, C\}$$

הוכח/הפוך את הטענות הבאות:

a. עבור  $Y \in Z_F^-$ ,  $Z \subseteq U$ , מתקיים בהכרח  $Y_F^+ = Z$ .

לא נכון. בדוגמא למעלה נוכל לקחת  $Z = BC, Y = ABC$  והטענה לא מתקיימת.



b.  $U_F^-$  היא בהכרח קבוצת כל מפתחות העל (superkeys) של  $\{U, F\}$ .  
 נכון.  $U_F^-$  מכילה את כל קבוצות האטריביוטים שהסגור שלהן הוא  $U$ .

c. עבור  $Z \subseteq U$  מתקיים בהכרח  $Z_F^- = (Z_F^+)_F^-$ .  
 נכון. כל קבוצות האטריביוטים שהסגור שלהן מכיל את  $Z$ , הן קבוצות אטריביוטים שהסגור שלהם מכיל גם את  $Z_F^+$ .

ד. (7 נקודות) נתונה הסמכה:  
 $\{U = ABCDEHG, F = \{A \rightarrow G, H \rightarrow DE, G \rightarrow DB, D \rightarrow C, E \rightarrow D\}\}$

**מצאו** פירוק  $BCNF$  המשמר מידע עבור הסמכה. הוכיחו כי כל סכמה בפירוק אכן ב- $BCNF$ .

$R1(A,G), R2(D,C), R3(D,G,B), R4(H,E), R5(H,A).$

$abcdegh \rightarrow abcdg, ahe$

$ahe \rightarrow ae, ah$

$agdbc \rightarrow ga, gbcd$

$gbcd \rightarrow gdb, dc$

## RDF – 10 נקודות

א. בהינתן מסד RDF כללי, כתבו שאילתת SPARQL שמחזירה את כל השלשות במסד (1 נקודות):

```
SELECT ?subject ?predicate ?object {  
  ?subject ?predicate ?object  
}
```

התבוננו בגרף ה-RDF הבא:

PREFIX pre:< http://www.ontotext.com/explicit>

pre: user1	pre: friendOf	pre: user2
pre: user2	pre: friendOf	pre: user1
pre: user1	pre: posted	pre: post1
pre: user1	pre: posted	pre: post2
pre: user2	pre: posted	pre: post3
pre: user1	pre: likes	pre: post1
pre: user2	pre: likes	pre: post1
pre: user2	pre: likes	pre: post2
pre: user2	pre: likes	pre: post3
pre: user3	pre: likes	pre: post3

ב. רשמו במקומות המתאימים את תוצאת הפעלת השאילתות הבאות על הגרף הנתון:

a. (3 נקודות)

```
select ?u ?u3 where {  
  ?u pre:likes ?p .  
  ?u2 pre:posted ?p .  
  OPTIONAL {  
    ?u pre:friendOf ?u3  
  }  
}
```

<u>u</u>	<u>u3</u>
user1	user2
user2	user1
user2	user1
user2	user1
user3	

b. (3 נקודות)

```
select ?u ?p where {
  ?u pre:likes ?p
  MINUS {
    ?u pre:posted ?p1
  }
}
```

<u>u</u>	<u>p</u>
user3	post3

ג. נסחו שאילתת SPARQL שמניבה את התוצאה הבאה (3 נקודות):

<u>u</u>	<u>f</u>	<u>l</u>
user1	user2	user1
user2	user1	user2
user3		

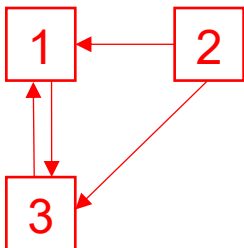
תשובה:

```
PREFIX pre:<http://www.ontotext.com/explicit/>
select distinct ?u ?f ?l where {
  ?u pre:likes ?p.
  OPTIONAL {
    ?u pre:friendOf ?f.
    ?f pre:friendOf ?l
  }
}
```

**בקרת מקביליות – 20 נקודות**

א. בהינתן  $s = W_3(z)R_1(y)R_2(y)W_1(y)W_3(y)W_1(z)$  :

a. צייר את ה-*precedence graph* של  $s$  (1 נקודות).

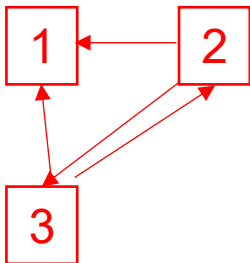


b. השלם את הטבלה הבאה (2 נקודות):

הסבר במשפט קצר	מתקיימת ב- (Y or N) ?s	התכונה
כי הגרף מעגלי	N	Conflict Serializable
בשל תנועות 1 ו-3 והפריטים $y$ ו- $z$ .	N	View Serializable

ב. בהינתן  $s' = W_3(z)R_2(z)W_2(y)R_3(y)R_1(y)W_1(y)R_1(z)W_1(z)$  :

a. (2 נקודות) צייר את ה-*precedence graph* של  $s'$



b. השלם את הטבלה הבאה (4 נקודות):

הסבר במשפט קצר	מתקיימת ב- (Y or N) ?s'	התכונה
כי הגרף מעגלי.	N	Conflict Serializable
בשל תנועות 2 ו-3 והפריטים y ו-z.	N	View Serializable

ג. (5 נקודות) האם ייתכן ש-s'' (המוצג למטה) נוצר ע"י תזמון באמצעות אלגוריתם 2-Phase Locking (פעולות הקשורות לנעילה אינן מוצגות). שימו לב, תנועה יכולה לבקש מנעול בכל נקודת זמן טרם הפעולה אותה המנעול מאפשר. נמקו את תשובתכם.

$$s'' = R_1(u)W_2(x)R_3(x)W_3(v)R_4(y)W_4(u)R_1(v)$$

אכן.
$LR_1(u) R_1(u) LW_2(x) W_2(x) UW_2(x) LR_3(x) R_3(x) LW_3(v) W_3(v) UW_3(v) UR_3(x) LR_4(y) R_4(y)$ $LR_1(v) UR_1(u) LW_4(u) W_4(u) UW_4(u) UR_4(y) R_1(v) UR_1(v)$

ד. התבונן בתהליך הביצוע החלקי הבא המתוזמן ע"י אלגוריתם 2-Phase Locking (הנח שמספר התנועה הוא תאריך התחלתה, קטן יותר אומר שהתאריך מוקדם יותר):

$$s = RL_1(x)R_1(x)WL_1(y)W_1(y)WL_2(z)W_2(z)RL_3(x)R_3(x) (*)$$

בנקודה (\*), תנועות 1, 2 ו-3 עדיין פעילות ותנועה מספר 2 מתעתדת לכתוב ל-x ולכן מבקשת  $WL_2(x)$  מהמתזמן.

(4 נקודות) מה יקרה תחת Wound-wait? בחר בתשובה אחת בלבד.

- תנועה מספר 2 תחכה ותנועה מספר 3 תופסק (restarted).
- תנועה מספר 1 תופסק (restarted).
- תנועה מספר 2 תחכה ותנועה מספר 1 תופסק (restarted).
- תנועה מספר 2 תקבל מידית את המנעול על x ותמשיך בפעילותה.
- התשובות א-ה אינן נכונות. התשובה הנכונה היא:

ה. (2 נקודות) הקיפו את הטענות הנכונות. עבור תהליכי ביצוע (schedule)  $s$  ו- $t$  כלשהם:

- a. אם  $s$  שקול מבט (view equivalent) ל- $t$  אז  $t$  בהכרח שקול מבט ל- $s$ .
- b. אם  $s$  שקול קונפליקט (conflict equivalent) ל- $t$ , ו- $t$  הוא תהליך ביצוע סדרתי, אז  $t$  בהכרח שקול מבט ל- $s$ .
- c. אם ה-precedence graph של  $s$  מעגלי אז  $s$  בהכרח אינו בר-סידור מבט (not view serializable).
- d. אם  $s$  ו- $t$  אינם שקולי מבט אז  $s$  ו- $t$  בהכרח אינם שקולי קונפליקט.

**מקום לתשובות נוספות:**

שאלה:	סעיף:

שאלה:	סעיף:

**מקום לתשובות נוספות:**

שאלה:

סעיף:

שאלה:

סעיף:

שאלה:

סעיף: