

מבחן במערכות בסיסי נתונים

פרופ' טובה מילוא

סמסטר א' מועד א'
11 בפברואר, 2013

הוראות (נא לקרוא!)

- משך הבחינה **שלוש שעות**, חלקו את זמנכם ביעילות.
- אסור השימוש בחומר עזר כלשהו.
- השימוש במחשבוניס מותר.
- יש לענות על כל השאלות **בגוף הבחינה** במקום המיועד לכך. יש לצרף את טופס המבחן למחברת הבחינה. מחברת ללא טופס עזר תיפסל. תשובות במחברת הבחינה לא ייבדקו. במידת הצורך ניתן לכתוב בגב טופס הבחינה.
- יש למלא מספר סידורי (מס' מחברת) ומספר ת.ז. על כל דף של טופס הבחינה. המבחן כולל 10 דפים כולל דף זה.

לשימוש הבודקים בלבד:

חלק	ציון
1	
2	
3	
סה"כ	

חלק 1 [40 נק']

בעולם הים מקפידים לשמור בצורה מסודרת מידע על החיות המופיעות לפני קהל. להלן חלק מן הסכמה בה הם משתמשים:

Animal (name, trainer, food, show_at)

Seal (name, training_hours, fish_won)

Dolphin (name, species, training_hours, fish_won)

Shark (name, does_bite)

הסכמה שומרת פרטים בסיסיים על כל חיה (שמה, המאמן שלה, מזון אהוב והשעה בה היא מופיעה). ניתן להניח כי לכל חיה בעולם הים יש שם ייחודי (ולכן זהו מפתח בטבלה). עבור חיות מסוגים מסוימים נשמרים פרטים נוספים: עבור כלבי-ים ודולפינים נשמרים מס' שעות אימון בשבוע ומס' הדגים הכולל בהם זכו; עבור כל דולפין נשמר גם זן הדולפין (species) אליו הוא משתייך; ועבור כרישים נשמר ביט המציין אם הם נושכים. הקישור בין הטבלאות מתבצע ע"י Foreign Key על name משלוש הטבלאות האחרונות ל-Animal.

א. [8 נק']

בעולם הים רוצים לתגמל מאמנים מצטיינים של כלבי ים. מאמן נחשב מצטיין אם ממוצע מס' הדגים בהם זכו כלבי-הים שלו גדול מממוצע מס' הדגים בקרב כלל אוכלוסיית כלבי-הים בעולם הים. כתבו שאילתה המחזירה עבור כל מאמן מצטיין את שמו ואת ממוצע מס' הדגים בהם זכו כלבי-הים שלו.

ב. [9 נק']

כתבו שאילתה המחזירה את שעות המופעים בהם לא מופיעה אף חיה "גדולה". החיות שנחשבות לגדולות בעולם הים הן כרישים, דולפינים מסוג "Common" ודולפינים מסוג "Bottlenose".

ג. [8 נק']

כתבו את אותה שאילתה באלגברה רלציונית.

ד. [15 נק']

התבוננו ב- view הנתון V1 ובשלוש השאילות Q1, Q2, Q3. לכל שאילתה סמנו מה מהבאים מתקיים:

(A) ניתן לשכתב את השאילתה תוך שימוש ב-V1 בלבד (אולי יותר מפעם אחת)

(B) ניתן לשכתב את השאילתה תוך שימוש ב-V2 בלבד (אולי יותר מפעם אחת)

(C) A ו-B נכונות

(D) אף תשובה מהנ"ל אינה נכונה

דוגמאות:

אם היינו מגדירים

CREATE VIEW V1 AS SELECT * FROM Animal WHERE trainer='Alice'

CREATE VIEW V2 AS SELECT * FROM Animal WHERE trainer='Bob'

אזי עבור השאילתה

SELECT * FROM Animal WHERE trainer='Alice' AND food LIKE '%fish%'

היה עליכם להחזיר A, ועבור השאילתה

SELECT A.name FROM Animal A, Dolphin D WHERE A.trainer='Alice' AND A.name = D.name

היה עליכם להחזיר D.

הגדרת V1:

CREATE VIEW V1 AS

SELECT A.show_at, A.trainer, D.training_hours, D.fish_won

FROM Animal A,

Dolphin D

WHERE A.name = D.name;

הגדרת V2:

CREATE VIEW V2 AS

SELECT A.name, A.trainer, D.training_hours, D.fish_won

FROM Animal A,

Dolphin D

WHERE A.name = D.name;

עבור כל שאילתה, סמנו את התשובה הנכונה.

Q1:

SELECT MAX(D.fish_won)

FROM Dolphin D WHERE D.training_hours > 4;

(A) (B) (C) (D)

Q2:

```
SELECT    A.trainer
FROM      Animal A
WHERE     A.name LIKE '%flip%'
GROUP BY  A.trainer
HAVING    COUNT(*) > 2;
```

(A) (B) (C) (D)

Q3:

```
SELECT DISTINCT A1.show_at
FROM Animal A1, Animal A2, Dolphin D1, Dolphin D2
WHERE A1.name = D1.name AND A2.name = D2.name AND A1.name <> A2.name AND
      A1.show_at=A2.show_at;
```

(A) (B) (C) (D)

חלק 2 [30 נק']

א. [18 נק']

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

$$A, B \rightarrow C$$

$$B, E \rightarrow C$$

$$C \rightarrow E$$

i. מצאו מפתח מינימאלי של הטבלה R .

ii. פרקו את R לצורת $3NF$. התשובה צריכה לכלול את הגזירה (עם צעדים והסבר לכל צעד), ואת הטבלאות שנוצרו עם השדות והמפתחות המינימאליים שלהן.

iii. עבור כל אחת מן הטבלאות בתוצאה הסופית של הפירוק, רשמו האם היא גם בצורת BCNF והסבירו בקצרה.

--

ב. [12 נק']

השאלות הבאות עוסקות במפתחות מינימאליים. X, Y ו- Z הן קב' זרות בזוגות של שדות, כלומר, החיתוכים ביניהן ריקים. עבור כל אחת מהשאלות הבאות הקיפו בעיגול את התשובה הנכונה -- נכון \ לא נכון. אין צורך בהסבר.

i. אם X הוא מפתח מינימאלי של X, U, Y, U, Z , אז X הוא גם מפתח מינימאלי של X, U, Y . נכון \ לא נכון

ii. אם X הוא מפתח מינימאלי של X, U, Y , אז X הוא גם מפתח מינימאלי של X, U, Y, U, Z . נכון \ לא נכון

iii. אם X, U, Y הוא מפתח מינימאלי של X, U, Y , אז X הוא מפתח מינימאלי של X . נכון \ לא נכון

iv. אם X, U, Y הוא מפתח מינימאלי של X, U, Y, U, Z , אז X הוא מפתח מינימאלי של X, U, Z . נכון \ לא נכון

חלק 3 [30 נק']**א. [5 נק']**

ברצוננו לבצע את ה-selection $\sigma_{a=3}(R)$ כאשר לא נתון מידע סטטיסטי על התפלגויות הערכים ב-R, וידוע שקיים אינדקס על השדה a. האם ניתן לשנות את האינדקס מ-clustered ל"רגיל" מבלי לפגוע בביצועי המערכת? אם כן, הסבירו באילו תנאים, אם לא הסבירו מדוע. ענו בקצרה (שלוש שורות לכל היותר)

ב. [5 נק']

עבור כל שתי טבלאות R ו-S שגודלן בדפים הוא B(R) ו-B(S) בהתאמה, מהי סיבוכיות הרצת אלגוריתם Block Nested Loop Join, בהנחה ש (א) הלולאה החיצונית רצה על R; (ב) הזיכרון יכול להכיל M דפים בסה"כ; ו- (ג) אנו מודדים קריאות IO בלבד, כולל קריאת הקלט מהדיסק אך ללא כתיבת הפלט לדיסק. ענו בקצרה (3 שורות לכל היותר)

ג. [20 נק']

נתונות שלוש טבלאות $R1(A, B, C)$ $R2(D, E)$ $R3(F, G)$ והשאלתה הבאה:

$$P = ((\sigma_{C=0}(R1)) \bowtie_{B=E} (R2 \bowtie_{E=G} R3))$$

הערות:

- שימו לב ל**סדר הפעולות** המוכתב ע"י הסוגריים, אשר יהיה קבוע בכל תוכניות הביצוע להלן.
- האינדקסים יושבים בזיכרון בשלמותם, כך שגישה לאינדקס לוקחת 0 פעולות IO.
- השאלתה מחושבת ב-pipeline, כלומר תוצאת אופרטור אחד ניתנת ישירות לאופרטור הבא (בלי להיכתב לדיסק בדרך).
- ניתן להניח שאפשר למיין את כל הטבלאות, כולל תוצאות ביניים, **במעבר אחד** (phase), כאשר תוצאת המיין נכתבת לדיסק ולא מועברת ישירות לאופרטור הבא.
- בכל מקום בו משתמשים ב-hash join, ניתן להניח שכמות הזיכרון וגדלי ה-buckets מאפשרים את הרצת האלגוריתם כפי שנלמד בכיתה (ללא חלוקה חוזרת ל-buckets וכדומה).
- אין לנו מידע על התפלגות ערכי השדות ואין אינדקסים מלבד אלו המוגדרים בכל סעיף בנפרד.

נתון המידע הסטטיסטי הבא:

	מס' ערכים בטבלה בעמודה נתונה	מספר דפים בטבלה	מס' רשומות בטבלה
R1	$V(R1, A) = 30,000$; $V(R1, B) = 25$; $V(R1, C) = 5$	$B(R1) = 100$	$T(R1) = 30,000$
R2	$V(R2, D) = 20,000$; $V(R2, E) = 10$	$B(R2) = 500$	$T(R2) = 20,000$
R3	$V(R3, F) = 24,000$; $V(R3, G) = 15$	$B(R3) = 300$	$T(R3) = 24,000$

בכל אחד משלושת הסעיפים הבאים, חשבו את מחיר תוכנית הביצוע במונחים של פעולות IO, לא כולל כתיבת התוצאה הסופית לדיסק. בטאו כל אחת מהתשובות כפונקציה של הסטטיסטיקות הנתונות. הקפידו על סדר פעולות חשבון וסוגריים בביטוי שתכתבו כדי להבטיח את קריאותו ונכונותו.
לדוגמא: $COST(P) = B(R1)B(R2)/V(R1, A)$ (לא התשובה האמיתית!)
 הסבירו כל תשובה בקצרה (3 שורות מקסימום).

i. יש אינדקס sorted clustered על C, ושתי פעולות ה-join ממומשות כ-sort-merge join.

ii. יש אינדקס sorted clustered על E ו-G, ושתי פעולות ה-join ממומשות כ- sort-merge join.

iii. יש אינדקס sorted clustered על C, ושתי פעולות ה-join ממומשות כ- hash join.

iv. ציינו מהי התוכנית (סעיפים i-iii) הזולה ביותר לחישוב השאילתה, ואת מחירה (מבוטא כמספר).

בהצלחה!