

## תרגיל בית מס' 4

נושאי התרגיל: CC, Indexes, Neo4j

- מתרגלת אחראית: מעיין ארנברג.
- ההגשה **בזוגות ומוקלדת** בפורמט PDF בלבד, עד לתאריך 6/7/2023.
- שאלות על התרגיל יש לשאול בפורום הייעודי בפיאצה:  
<https://piazza.com/technion.ac.il/spring2023/236363>
- שאלות אדמיניסטרטיביות יש להפנות לאופיר פדר.
- הבהרות שנוספו לאחר פרסום התרגיל מסומנות **בצהוב**.

פרטי המגישים:

ת.ז.	
1.	
2.	

# שאלה 1 – Indexes (35 נק')

מסד הנתונים "Music Maestro" מחזיק מידע על אומנים ואלבומים.

המערכת מתחזקת את הטבלאות הבאות:

Artists(ID, Name, BirthDate, Nationality, Genre)	אומנים – מזהה, שם, תאריך לידה, מוצא, ז'אנר אופייני
Songs(ID, Title, ReleaseDate, ArtistID)	שירים – מזהה, שם, תאריך פרסום, ומזהה אומן

עבדו תחת ההנחות הבאות:

- כל בלוק מכיל 1024 בתים (bytes)
- כל רשומה בכל יחס תופסת 100 בתים
- רשומות אינן מפוצלות בין בלוקים
- במערכת רשומים 1000 אומנים ו-100,000 שירים
- אומן מוציא לכל היותר שיר אחד בחודש
- מסד הנתונים מעודכן לתאריך פרסום התרגיל (יוני 2023) ואינו מכיל שירים עם תאריך פרסום עתידי

1. התבוננו בשאילתה הבאה, למציאת השירים החדשים של כל אמני ה-Jazz:

```
SELECT A.ID, S.ID
FROM Artists A, Songs S
WHERE S.ArtistsID = A.ID AND
      A.Genre = 'Jazz' AND
      S.ReleaseDate >= TO_DATE('2023/01/01', 'YYYY/MM/DD');
```

הראו דוגמה לתכנון שאילתת nested loop join, ונתחו את ה-cost \O בהתאם. הניחו כי במסד הנתונים אין אינדקסים כלל.

כעת נוסיף למסד הנתונים את האינדקס הבא:

```
CREATE INDEX artist_date_idx ON Songs(ArtistID, ReleaseDate);
```

ידוע כי האינדקס ממומש כעץ B+ עם דרגה  $d=10$ .

2. לכל אחת מהשאלות הבאות, תארו במילים כיצד ניתן לבצע את השאילתה באופן יעיל תוך שימוש באינדקס, ונתחו את עלות ה- $I/O$ .

i. החזרת תאריך הפרסום של השיר החדש ביותר של אומן מסוים:

```
SELECT S.ReleaseDate
FROM Songs S
WHERE S.ArtistsID = '236363'
ORDER BY S.ReleaseDate Desc
LIMIT 1;
```

ii. החזרת כל השירים של אומן מסוים החל משנת 2023:

```
SELECT S.ID
FROM Songs S
WHERE S.ArtistsID = '236363'
AND S.ReleaseDate >= TO_DATE('2023/01/01', 'YYYY/MM/DD');
```

3. תארו ביצוע של השאילתה מסעיף 1, תוך שימוש באלגוריתם index nested loop join עם האינדקס הנתון:

```
SELECT A.ID, S.ID
FROM Artists A, Songs S
WHERE S.ArtistsID = A.ID
AND A.Genre = 'Jazz'
AND S.ReleaseDate >= TO_DATE('2023/01/01', 'YYYY/MM/DD');
```

## שאלה 2 – Currency Control (35 נק')

1. לכל אחד מהתזמונים הבאים, בדקו והסבירו:

- האם התזמון בר סידור מבט?
- האם התזמון בר סידור קונפליקט?

א.  $R_2(x)$   $W_1(x)$   $W_2(y)$   $R_3(y)$   $R_1(y)$   $W_2(x)$

ב.  $R_2(z)$   $W_3(z)$   $W_4(z)$   $R_2(y)$   $W_2(x)$   $R_1(z)$   $R_4(y)$

$$R_1(x) \ R_2(x) \ R_1(y) \ W_2(x) \ W_3(y) \ W_1(y) \ W_2(y) \quad .III$$

2. בכל סעיף, הביאו דוגמה לתזמון מתאים של שתי טרנזקציות אשר להן פעולות על משתנה משותף אחד לפחות.

I. תזמון אשר יכול להתקבל ע"י פרוטוקול Strict 2PL וגם ע"י Conservative 2PL, ואינו סדרתי.

II. תזמון בר סידור מבט, שאינו יכול להתקבל ע"י 2PL.

## שאלה 3 – Neo4j (30 נק')

מסד הנתונים PawsomeDB הוא מסד מסוג Neo4j, המחזיק מידע על גזעי חתולים ועל חתולי מחמד ובעליהם.

סוגי הצמתים:

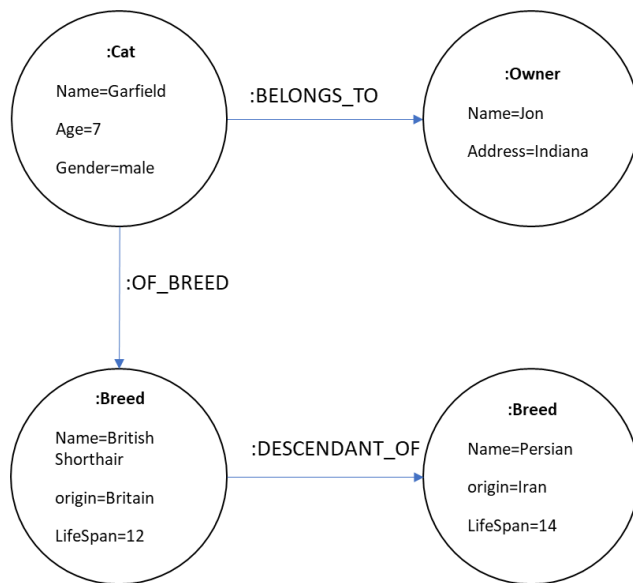
- **Label:** Breed **Properties:** name, origin, lifespan.
- **Label:** Cat **Properties:** name, age, gender.
- **Label:** Owner **Properties:** name, address.

סוגי היחסים:

- **Type:** BELONGS\_TO **Properties:** relationship\_length.
- **Type:** DESCENDANT\_OF.
- **Type:** OF\_BREED.

היחס BELONGS\_TO מתקיים תמיד בין Cat ל-Owner, היחס DESCENDANT\_OF בין Breed ל-Breed, והיחס OF\_BREED בין Cat ל-Breed.

להלן דוגמה לתוכן ה-DB:



ענו על כל אחד מהסעיפים הבאים באמצעות שאילתת cypher.

1. לכל גזע חתולים במסד הנתונים, החזירו את שם הגזע ואת מספר החתולים המשתייכים אליו באופן ישיר (OF\_BREED). אין צורך לכלול גזעים ללא חתולים.

מיינו את התוצאה בסדר יורד לפי מספר החתולים.

--

2. חזרו על סעיף 1, והפעם לכל גזע חתולים חשבו את מספר החתולים המשתייכים אליו באופן ישיר או משתייכים לגזע צאצא. למשל, בדוגמה, Garfield משתייך באופן ישיר לגזע British Shorthair ובאופן עקיף לגזע Persian.

--

3. החזירו את שמות האנשים שכל החתולים בבעלותם מאותו מין (gender). אין צורך להחזיר אנשים שאין בבעלותם חתולים.

--

4. החזירו את שמות האנשים שלהם חתול מכל גזע במסד הנתונים. בסעיף זה, כמו בסעיף 1, הכוונה להשתייכות ישירה לגזע (OF\_BREED).

--