

# תרגיל 3 | DB (67329)

שם: רונאל חרדים, עומרי טויטו | ת"ז: 208917641, 208432361

## שאלה 1

השאלתה מצורפת:

בנוסף מצורף ניתוח השאלתה עם הפקודה *explain analyse*

```
explain analyse
select distinct year from authors
where institution = 'Hebrew University of Jerusalem'
```

```
Terminal
File Edit View Search Terminal Help

QUERY PLAN

-----
HashAggregate (cost=3904.46..3904.99 rows=53 width=4) (actual time=19.799..19.830 rows=43 loops=1)
  Group Key: year
    -> Seq Scan on authors (cost=0.00..3902.30 rows=863 width=4) (actual time=0.427..19.011 rows=929 loops=1)
        Filter: ((institution)::text = 'Hebrew University of Jerusalem'::text)
        Rows Removed by Filter: 163895
  Planning Time: 0.089 ms
  Execution Time: 19.902 ms
(7 rows)

(END)
```

(ב)

הסבר השאלתה מצורף בסעיף קודם.

זמן שלקח להריץ את השאלתה: 19.902 ms

אופן חישוב השאלתה: המערכת חישה את השאלתה בעזרת *hash* עם סריקה על הטבלה *authors* והעלאת השורות בהן מתקיים בתנאי לגבי המוסד.

(ג)

השאלתה עם אינדקס:

```

create index institution
on authors(institution );

explain analyse
select distinct year from authors
where institution = 'Hebrew University of Jerusalem';

```

```

----- QUERY PLAN -----
HashAggregate  (cost=1555.19..1555.72 rows=53 width=4) (actual time=1.656..1.688 rows=43 loops=1)
  Group Key: year
    -> Bitmap Heap Scan on authors  (cost=36.81..1553.03 rows=863 width=4) (actual time=0.139..0.902 rows=929 loops=1)
          Recheck Cond: ((institution)::text = 'Hebrew University of Jerusalem'::text)
          Heap Blocks: exact=53
          -> Bitmap Index Scan on institution  (cost=0.00..36.59 rows=863 width=0) (actual time=0.125..0.126 rows=929 loops=1)
Planning Time: 0.539 ms
Execution Time: 1.819 ms
(8 rows)

(END)

```

(ד)

הסבר השאילתה מצורף בסעיף קודם.

זמן שלקח להריץ את השאילתה:  $1.819\text{ ms}$

אופן חישוב השאילתה: המערכת סיננה את השורות עליהם התנאי מתקיים עם *index scan* על העמודה *institution*, אחכ חישבה את השאילתה בעזרת *hash* על השדה.

## שאלה 2

(א)

1. ראשית נחשב את מספר השורות בבלוק

$$\left\lfloor \frac{2000}{180} \right\rfloor = 11$$

כעת, נחשב את מספר הבלוקים בטבלה:

$$\left\lceil \frac{12000}{11} \right\rceil = 1091$$

לכן העלות היא 1091

2. נחשב את דרגת הפיצול של אינדקס על השדה  $count$ :

$$d = \left\lfloor \frac{2000 + 8}{8 + 8} \right\rfloor = \left\lfloor \frac{2008}{16} \right\rfloor = \lfloor 125.5 \rfloor = 125$$

3. ראשית נחשב את גובה העץ:

$$h = \left\lceil \log_{\frac{125}{2}}(12000) \right\rceil = 3$$

כעת נחשב את מספר העלים בהם נמצא המידע:

$$leavs = \left\lceil \frac{\frac{12000}{20}}{\left\lceil \frac{125}{2} \right\rceil - 1} \right\rceil = \left\lceil \frac{600}{62} \right\rceil = 10$$

עלות חישוב השאילתה עם דרגת הפיצול והאינדקס הנ"ל תהיה:

$$cost = 10 + 3 = 13$$

(ב)

1. ראשית נחשב את מספר השורות בבלוק

$$\left\lfloor \frac{2000}{180} \right\rfloor = 11$$

כעת, נחשב את מספר הבלוקים בטבלה:

$$\left\lceil \frac{12000}{11} \right\rceil = 1091$$

לכן העלות היא 1091

2. נחשב את דרגת הפיצול של אינדקס על השדה  $count$ :

$$d = \left\lfloor \frac{2000 + 8}{8 + 8} \right\rfloor = \left\lfloor \frac{2008}{16} \right\rfloor = \lfloor 125.5 \rfloor = 125$$

3. מספר השורות המתאימות הוא  $\frac{12000}{20} = 600$  אך אנו צריכים את השדה  $name$  לכן נצטרך לגשת ל 600 השורות המתאימות בנוסף לעלות החישוב שחישבנו בשאלה קודמת לכן

**סהכ:**  $600 + 13 = 613$

(ג)

1. עלות חישוב השאילתה: בדומה לסעיפים קודמים גודל הטבלה נכנס ב 1091 בלוקים ולכן זאת העלות.

2. דרגת הפיצול האופטימלית היא:

$$d = \left\lfloor \frac{2000 + 6 + 16 + 4}{6 + 16 + 4 + 8} \right\rfloor = \lfloor 59.6 \rfloor = 59$$

3. עלות חישוב השאילתה עם האינדקס:

נחשב את גובה העץ:

$$h = \left\lceil \log_{\frac{59}{2}} 12000 \right\rceil = 3$$

נחשב את עלות הטיול על העץ בטווח הערכים:

$$leavs = \left\lceil \frac{\frac{12000}{3 \cdot 3 \cdot 80}}{\left\lceil \frac{59}{2} \right\rceil - 1} \right\rceil = \left\lceil \frac{17}{29} \right\rceil = 1$$

סהכ:

$$3 + 1 = 4$$

(ד)

1. דרגת הפיצול האופטימית של האינדקס היא:

$$d = \left\lfloor \frac{2000 + 6 + 16}{6 + 16 + 8} \right\rfloor = \lfloor 67.4 \rfloor = 67$$

2. עלות החישוב עם האינדקס:

ראשית נחשב את גובה העץ:

$$h = \left\lceil \log_{\frac{67}{2}} 12000 \right\rceil = 3$$

נחשב את עלות הטיול בטווח:

$$leavs = \left\lceil \frac{\frac{12000}{80}}{\left\lceil \frac{67}{2} \right\rceil - 1} \right\rceil = \left\lceil \frac{150}{33} \right\rceil = 5$$

סה"כ: מכיוון שאנו מחפשים שני ערכים אזי אנו צריכים לעבור פעמיים על העץ בנוסף הערכים מתפלגים אחיד על 80 לכן

$$2 \cdot (3 + 5) = 16$$

(ה)

1. דרגת הפיצול האופטימית של האינדקס היא:

$$d = \left\lfloor \frac{2000 + 8}{8 + 8} \right\rfloor = \left\lfloor \frac{2008}{16} \right\rfloor = \lfloor 125.5 \rfloor = 125$$

2. עלות חישוב השאילתה בהינתן האינדקס היא:

ראשית נחשב את גובה העץ:

$$h = \left\lceil \log_{\frac{67}{2}} 12000 \right\rceil = 3$$

נחשב את עלות הטיול בטווח:

$$leavs = \left\lceil \frac{\frac{19 \cdot 12000}{20}}{\left\lceil \frac{125}{2} \right\rceil - 1} \right\rceil = \left\lceil \frac{11400}{62} \right\rceil = 184$$

**סהכ:** אנו צריכים להוציא את השדה *name* ולכן נצטרך לבצע גישות לזכרון כמספר השורות המתאימות = 1091 לכן סהכ

$$1091 + 3 + 184 = 1278$$