

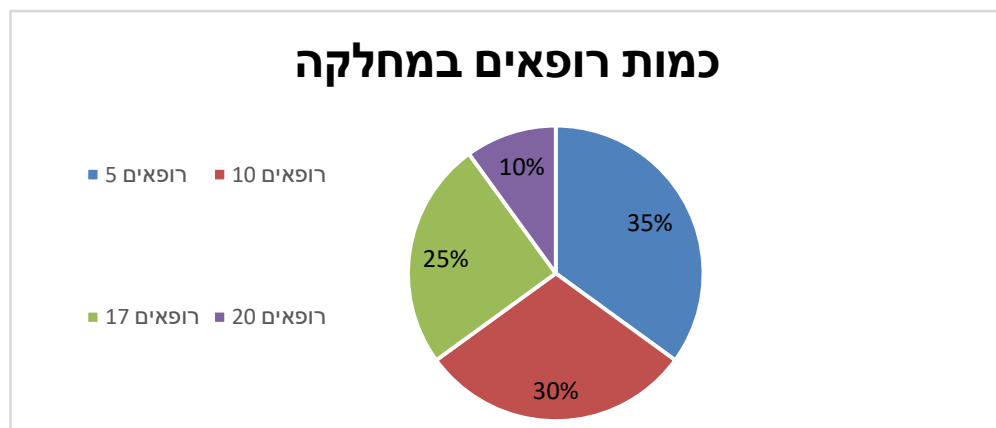
## תכן של מסדי נתונים תרגיל בית 3 - תשע"ח

תאריך הגשה: 20/05/2018

### נתונים לחישוב עלויות:

על פי נתוני משרד הבריאות ישנם 80 בתי חולים הרשומים במערכת הנתונים שירות רפואה לתושבי הארץ. ידוע שכ-30% מבתי החולים במערכת הם בתי חולים פרטיים, והשאר ציבוריים. בבית חולים ציבורי ישנן בממוצע כ-16 מחלקות שונות ובבית חולים פרטי-6 מחלקות שונות בממוצע.

התפלגות הרופאים למחלקות השונות נתונה בגרף הבא. (לדוגמא: ב-35% מהמחלקות בכלל הארץ ישנם 5 רופאים המאיישים את המחלקה).



ישנן 14 משמרות בשבוע אליהן הרופאים משובצים (יום א' בוקר, יום א' ערב, יום ב' בוקר, וכו'). רופאים מתמחים, המהווים 25% מכלל הרופאים, עובדים בממוצע 6 משמרות בשבוע, ורופאים ותיקים שסיימו את ההתמחות שלהם עובדים בממוצע 4 משמרות בשבוע. (רופאים מתמחים הינם רופאים אשר סיימו את לימודיהם ב-3 השנים האחרונות). בנוסף ידוע כי 40% מהרופאים הינם רופאים עמוסים העובדים במשמרת רצופה (לדוגמא: יום שלישי ערב ויום רביעי בוקר).

במידה ואין נתונים סטטיסטיים יש להניח התפלגות אחידה.

להלן נתונים על המחשב שעליו תותקן המערכת ונתונים על ה-RDBMS:

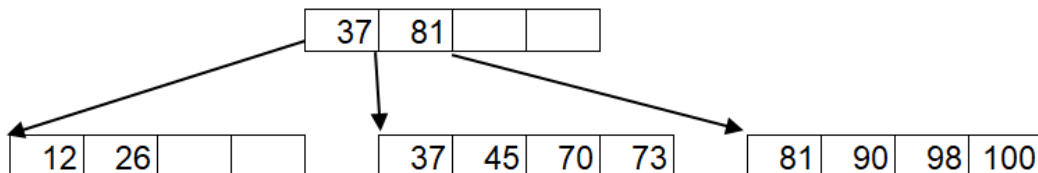
- ♣ גודל בלוק הוא 256 בתים (bytes).
- ♣ גודל הזיכרון הפנוי לשאילתה מסוימת הוא 100 בלוקים.
- ♣ עבור אינדקסים משניים ניתן להניח  $LB_i=6,000$ .
- ♣ כל טבלה ממוינת על פי המפתח הראשי שלה. במידה והמפתח הראשי מורכב ממס' שדות, אופן מיון הטבלה הינו עפ"י סדר הופעת העמודות בטבלה.

גדלי סוגי המשתנים נתונים בטבלה הבאה ( שימו לב כי התכונות וסוגיהם מתבססים על פתרון תרגיל בית 1 המוצע):

Data type	Description	Size
integer	Integer (whole number) data from $-2^{31}$ (-2,147,483,648) through $2^{31}-1$ (2,147,483,647)	4 bytes
smallint	Integer data from -32,768 to 32,767	2 bytes
tinyInt	Integer data from -32,768 to 32,767	1 byte
bigInt	Integer (whole number) data from $-2^{63}$ (-9,223,372,036,854,775,808) through $2^{63}-1$ (9,223,372,036,854,775,807)	8 bytes
bit	Integer data with a value of either 1 or 0	1 byte (actually 1 bit)
float	Floating point number data from $-1.79E+308$ through $1.79E+308$	8 bytes
real	Floating precision number data from $-3.40E+38$ through $3.40E+38$	4 bytes
date	YYYY is four digits from 0001 to 9999 that represent a year. MM is two digits from 01 to 12 that represent a month in the specified year. DD is two digits from 01 to 31, depending on the month, that represent a day of the specified month.	3 bytes
time	hh is two digits, ranging from 0 to 23, that represent the hour. mm is two digits, ranging from 0 to 59, that represent the minute. ss is two digits, ranging from 0 to 59, that represent the second. n* is zero to seven digits, ranging from 0 to 9999999, that represent the fractional seconds.	5 bytes, <u>unless</u> something else specified
varchar(n)	Variable-length, non-Unicode string data. n defines the string length and can be a value from 1 through 8,000. <b>Max</b> indicates that the maximum storage size is $2^{31}-1$ bytes (2 GB)	The storage size is the actual length of the data entered + 2 bytes.
Char(n)	Fixed-length Unicode character data of n characters entered characters. N must be a value from 1 through 4,000	n bytes
text	Variable-length data with a maximum length of $2^{31}-1$ characters	256 bytes
Varbinary(n)	Variable-length binary data with a maximum length of 8000 bytes. Default length = 1. Storage size varies.	It is the length of the value in bytes
Pointer	An address (long Integer)	10 bytes

### שאלה 1:

נתון עץ B+ עם  $B=5$  מהמתאר את האינדקס על שדה HospitalID בטבלת Hospital (שימו לב שאלה זאת כללית ואיננה קשורה לנתונים שתוארו עד כה):



### הנחות:

- המצביע הימני הינו עבור ערכים גדולים או שווים לערך בצומת ה"אב" והמצביע השמאלי עבור ערכים הקטנים מהערך בצומת ה"אב".
- בעת פיצול של צמתים,  $\lceil \frac{n}{2} \rceil$  ערכים יועברו לצומת השמאלי.
- בעת מחיקת ערך, במידה ונוצר מצב בו צומת מכיל כמות קטנה מהמותר לפי חוקי העץ, יש לנסות 'להלוות' מהאח הימני בלבד. עבור הצומת הימני ביותר יש 'להלוות' מהאח השמאלי.

### דרישות:

- הסבר כיצד נראה הקשר בין העץ לטבלה בדיסק.
- הראה את העץ לאחר כל ביצוע פעולה לפי סדר הפעולות הבא:
  - הכנסת הערך 94
  - הכנסת ערך 72
  - הכנסת ערך 95
  - הכנסת ערך 96
  - מחיקת הערך 81
  - מחיקת הערך 72
  - מחיקת הערך 100

### שאלה 2:

חשב את גדלי כל אחת מהטבלאות המשתתפות בשאלות 3 ו-4. עבור כל טבלה יש להראות את כמות הרשומות ( $n_r$ ), גודל כל רשומה (record size), כמות הרשומות שנכנסות בבולוק יחיד  $f_r$  וכמות הבולוקים שהיא תופסת  $b_r$ . במקרה בו כמות הרשומות בטבלה לא נתון בצורה מפורשת בשאלה, יש להסביר כיצד הערכתם את כמות הרשומות בה.

### שאלה 3:

נתונה השאילתה הבאה:

```
Select dp.hospitalID ,dp.departmentName,d.doctorID
from Department dp join Doctor d on dp.hospitalID=d.hospitalID and
dp.departmentID=d.departmentID
where dp.hospitalID=5
```

- א. חשב את עלות השאילתה ללא שימוש באינדקסים. יש להראות את כלל האפשרויות לחישוב העלות ולציין איזו תוכנית ביצוע עדיפה.
- ב. חשב את עלות השאילתה תוך שימוש באינדקס מסוג עץ B+ בעל רמת פיצול של 67% על השדה hospitalID בטבלת Department. האם השימוש באינדקס מועיל?
- ג. חשב את עלות השאילתה תוך שימוש באינדקס מסוג עץ B+ בעל רמת פיצול של 67% על השדה departmentID בטבלת Department. האם השימוש באינדקס מועיל?
- ד. חשב את עלות השאילתה תוך שימוש באינדקס משולב (על שני שדות) מסוג עץ B+ בעל רמת פיצול של 67%, על השדות hospitalID ו-departmentID (לפי hospitalID ולאחר מכן departmentID), בטבלת doctor.
- ה. נניח שהתנאי ב- WHERE שונה באופן הבא. איך תשתנה עלות השאילתה מסעיף א (ללא אינדקסים)?

```
Where dp.hospitalID=5 OR dp.hospitalID=6
```

### שאלה 4:

נתונה השאילתה הבאה:

```
Select D.doctorID, D.hospitalID
From worksInShift W join Doctor D on W.doctorID=D.doctorID
Where not exists(select w1.doctorID
                  From worksInShift w1 join worksInShift w2 on
                  w1.doctorID=w2.doctorID
                  where w1.doctorID=w.doctorID and
                  w1.shiftNumber+1=w2.shiftNumber)
group by w.doctorID
having COUNT(w.shiftNumber)= 4
order by D.hospitalID
```

- א. הסבר בלשון ברורה (משפט או שניים) את משמעות השאילתה.
- ב. עליך לחשב את עלות השאילתה ללא שימוש באינדקסים. יש להעריך את כלל האפשרויות לביצוע השאילתה ולציין איזו שיטה עדיפה (זולה ביותר). לצורך זה:
- חשב תחילה את עלות השאילתה הפנימית.
  - חשב את עלות השאילתה המלאה.
- ג. חשב כיצד ישפיע על עלות השאילתה שימוש באינדקס מסוג עץ B+ בעל רמת פיצול של 67% על השדה doctorID בטבלת WorksInShift.
- ד. עבור כל טבלה בשאילתה (WIS ו-Doctor), ציין איזה אינדקסים ניתן להגדיר לצורך שיפור עלות השאילתה הנוכחית, ועל איזו פעולה ישפיעו. אין צורך בחישוב העלות באמצעותם. עבור כל אינדקס, יש לפרט את המאפיינים שלו (ראשי/משני, קביל/לא קביל).

### תוצרי הגשה:

יש להעלות את פורמט ההגשה למערכת הגשת התרגילים במודל (moodle) ולהגיש עותק

קשיח לתא הקורס ליד המזכירות.

**תאריך הגשה (במודל + תא): יום רביעי ה-20/05/2018 עד השעה 12:00 בבוקר.**

**הגשה באיחור תתקבל עד יום חמישי ה-27/05/2018 בשעה 12:00 בבוקר.**

### הנחיות כלליות:

1. יש להגיש את התרגיל בזוגות. לא תאושרנה שלשות!
2. את תרגילי הבית יש להגיש עם תעודות זהות בלבד!
3. כאמור, את כל התרגילים יש להגיש הן באמצעות מערכת הגשת התרגילים של הקורס והן כהדפסה לתא הקורס. על כל התוצרים להגיע ליעדם (במערכת ולתא) עד המועד שנקבע. מאחריותכם לוודא שהגשתם את כל התוצרים כנדרש ושהם הגיעו ליעדם. לא תתאפשר בדיקה מחודשת של העבודה עקב טעויות בעת ההגשה!
4. במהלך תקופת הסמסטר יתכן ויפורסמו הנחיות, הערות והכוונות נוספות באתר הקורס- מאחריותכם לעקוב אחר הודעות אלו.
5. ניתן לפנות למרצה ולמתרגלות בכל שאלה בשעות הקבלה או באמצעות הדואר האלקטרוני.

## בהצלחה!