

**מבחן במסדי נתונים**

**מירב שקרון**

**7028010**

סמסטר קיץ מועד ב' כ חשון התשע"ח, 9.11.2017

**הנחיות כלליות:**

* משך הבחינה: 180 דקות.
* **יש לענות בגוף השאלון!** המחברת תשמש כטיוטא בלבד. על מענה במחברת יורדו נקודות!
* במבחן 9 שאלות, שימו לב שאתם עוברים על כולם
* אין להכניס שום חומר עזר.
* השימוש במחשבון **אסור**.
* בסיום הבחינה - נא למסור את השאלון ואת המחברת.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Total | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |  |
| 100 | 9 | 12 | 5 | 10 | 6 | 8 | 12 | 12 | 26 | Max points |
|  |  |  |  |  |  |  |  |  |  | Grade |

**ב ה צ ל ח ה!**

1. SQL + Relational algebra (26 נק')

נתון מבנה הנתונים להלן, של אולימפיאדת מקסיקו 2016:

Participant (ParticipantId, name, gender, yearOfBirth, country)

טבלת משתתפים (gender=1- נקבה; gender=0- זכר)

CompetitionTypes (competitionTypesId, name)  טבלת תחרויות

PrizesTypes (prizeTypeId, name, points)  טבלת סוגי פרסים וניקוד

CompetitionParticipant (competitionTypeId, ParticipantId, prizeTypeId , dayAsInt)

טבלת משתתפים בתחרות והפרס (הפרס יכול להיות null)

ניתן להניח ש- dayAsInt הוא מספר שלם המייצג את יום המשחק מאז שהחלה האולימפיאדה.

1. כתוב שאילתא שתחזיר את כל המדינות שבהן זכו נשים בשלוש סוגי פרסים שונים לפחות (8 נק')

Select p.country from Participant as p

join competitionParticipant as cp

on(p.participantId = cp.participantId)

where p.gender = 1

and count(distinct prizeTypeId)>2

group by p.country

1. כתוב שאילתא שתחזיר את סוגי התחרויות שבהם זכו במעל 100 נקודות בכל סוג (8 נק')

select competitionTypeId

from CompetitionParticipant

where CompetitionTypeId not in

(select cp.competitionTypeId

from PrizesTypes as pt join CompetitionParticipant as cp

on(pt.prizeTypeId = cp.prizeTypeId)

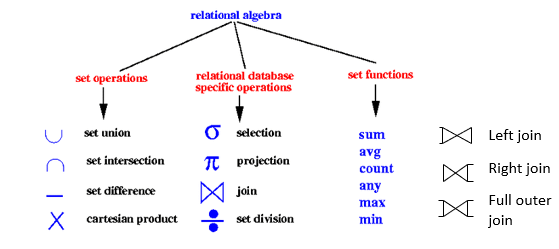
group by cp.competitionTypeId, pt.prizeTypeId

having sum(pt.points) < 101)

1. כתבו ביטוי רלציוני המחזיר את כל הנשים שזכו בפרס כלשהוא במהלך התחרות (10 נק')

πp .*name*  
 σ*NOT*(cp .*prizeid*=*NULL*)AND p .*gender*= 1  
  (ρp*participant* ⋈p .*id*=*cp*.*id*  
   ρ*cp**competitionparticipant*)

להזכירכם, אלה הסימונים שלמדנו בכיתה-



1. Normalization (12 נק')

נתונה הרלציה הבאה:

R (U, V, W, X, Y, Z)

נתונות התלויות שלה:

{Y, U}🡪{X, Z}

{W, X, U}🡪Y

{U, Y, Z}🡪V

1. מצא את כל ה-candidate keys האפשריים (5 נק')

(U,Y,W),(W,X,U)

1. מהי רמת הנירמול (NF) של הרלציה הנ"ל (1NF, 2NF, 3NF, BCNF, 4NF).

**נמק** למה הרלציה לא יכולה להתאים ל-NF גבוה יותר (7 נק')

Non-prime: V,Z

2NF: doesn’t hold since Z depends on a subset of a super key (Y,U) is a subset of (U,Y,Z).

Then: R holds 1NF only

1. ERD (12 נק')

נתבקשת לתכנן מערכת הפצה ארגונית להפצת מכתבים ומסרים ללקוחות עבור חברת הביטוח "שוד ישיר"

* ישנם כ-50 סוגי מכתבים שונים .
* לכל סוג מכתב יש את ערוצי ההפצה המורשים עבורו .לדוגמא: מכתבים מסוימים (כמו הודעה על חוב) חייבים להישלח בדואר רשום בלבד, בעוד אחרים יכולים להישלח בדואר רגיל או בערוצי הפצה דיגיטליים.
* ערוצי ההפצה הקיימים הם: מסרון , דוא"ל , דואר רגיל ודואר רשום.

1. יש לשרטט ERD מנורמל לניהול יעיל של כללי ההפצה כפי שהוגדרו לעיל (6 נק')

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ירד מהחומר נכון לקיץ 2021....\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. הלקוח מעוניין לעקוב אחר ההפצות באמצעות דו"ח במבנה הבא:



עדכן את ה-ERD שכתבת על מנת לתמוך בדוח זה .

שימו לב: יש להוסיף ישויות ושדות ככל שנדרש ולהקפיד על נרמול (6 נק')

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ירד מהחומר נכון לקיץ 2021....\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. XML and XSD (8 נק')

כתבו את **כל** הסיבות מדוע ה- XML הבא, לא בתוקף (not valid) לפי הגדרת ה-XSD להלן:

Xml:

<CellPhone>

<company>Samsung</company>

<model>Galaxy S8</model>

<model>Galaxy S8 note </model>

<camera>

<type>primary</type>

<MP>12</MP>

</camera>

<camera>

<type>secondary</type>

<MP>8</MP>

</camera>

<memory>

<cardSlot>256 GB</cardSlot>

<internal>64 GB</internal>

</memory>

</CellPhone>

XSD:

<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified" xmlns:xs="http://www.w3.org/2001/XMLSchema">

<xs:element name="CellPhone">

<xs:complexType>

<xs:sequence>

<xs:element type="xs:string" name="company"/>

<xs:element type="xs:string" name="model"/>

<xs:element name="camera">

<xs:complexType>

<xs:sequence>

<xs:element type="xs:int" name="MP"/>

<xs:element type="xs:string" name="type"/>

</xs:sequence>

</xs:complexType>

</xs:element>

<xs:element name="memory">

<xs:complexType>

<xs:sequence>

<xs:element type="xs:string" name="cardSlot"/>

<xs:element type="xs:string" name="internal"/>

<xs:element type="xs:string" name="external" maxOccurs="3"/>

</xs:sequence>

</xs:complexType>

</xs:element>

</xs:sequence>

</xs:complexType>

</xs:element>

</xs:schema>

1. <model> occurs twice
2. <type> occurs before <MP>
3. <external> should occur at least once.
4. NoSQL (6 נק')

מה ה-DB מסוג NoSQL המתאים ביותר לדרישות הבאות:

(בחרו בסוג המתאים **ביותר** לפי הסוגים שנלמדו)

**נמקו**

1. אתם רוצים בסיס נתונים אשר ייתן לכם לבטא קשרים בין ישויות שונות בסביבה בה יש קשרים מרובים. מעניין אתכם מי מקושר למי גם אם הקשר עקיף .

Neo4J

1. אתם רוצים בסיס נתונים אשר יאחסן מאמרים, וייתן לכם אפשרות לקבל את המאמרים הרלוונטיים ביותר לפי מילות מפתח.

Elastic search

1. הנתונים שלכם בנויים ממסמכים בהם יש פרטים מספריים רבים. אתם מעוניינים במסד נתונים שיאפשר לכם שאילתות עם חישובים מורכבים על פרטים מתוך המסמכים.

mongodb

1. Neo4j (10 נק')

נתון DB שיש בו סטודנטים (Student) וקורסים (Courses) .

סטודנטים יכולים להרשם (Register) לקורס.

חלק מהסטודנטים עוברים את הקורס (Passed) וחלק לא עוברים (Failed)

1. כתבו שאילתא ב- Cypherשמחזירה את כל הקורסים שלהם לפחות שני סטודנטים שנרשמו אליו.(5 נק')

Match (s:Student)-[:Register]-(c:Course) with c, count(distinct s) as regs where regs >1 return c

1. כתבו שאילתא ב- Cypherשמחזירה את הקורסים שבהם כל מי שנרשם עבר. (5 נק')

Match (s:Student)-[:Register]->(c:Course), (m)-[:Passed]->(c) where s=m return c

1. JAVA Streams (5 נק')

מה יהיה הפלט של תוכנית ה-Java stream הבאה:

List<String> myList11 = Arrays.asList("Sunday", "horse", "Tuesday", "honor", "goodbye", "hello", "hi");  
myList11.stream()  
 .map(String::toUpperCase)  
 .filter(s -> s.startsWith("H"))  
 .sorted()  
 .reduce((a,b) -> a + " \*\*\* " + b)  
 .ifPresent(s -> System.out.println(s));

HELLO \*\*\* HI \*\*\* HONOR \*\*\* HORSE

1. Linear regression (12 נק')

נזכיר שעבור linear regression הגדרנו את פונקציית ה-loss הבאה:

1. https://lh6.googleusercontent.com/3XpcoJ4CjAchf3zlaoBtbvtlgsTaT7XRBnua7zwOoQnEJlPjhdKJxuDh7kGJ8BHP0y9nuD_nRYtasoBXciRjU4c_PvwtADrhhDbnh2kJcwvYrmF9c2giMMan6_-FBZauV6pgHM5E

נזכיר שקו הניבוי ב- linear regression הוא: h(x) = xw + b

בהינתן הנתונים הבאים:

X = [3, 4, 0, -3]

Y = [2, 1, 1, -2]

1. מה הטעות כאשר w=0,b=0? (4 נק')
2. מצא ערכים עבור w ו-b שמביאים טעות קטנה יותר (לאו דווקא מינימלית) (4 נק')

1. שרטט את קו הניבוי (4 נק')

נקודת החיתוך עם ציר הy היא 0 (b=0) ובx=3 קל לראות ש אז יש לנו 2 נקודות:

(0,0), (3,2) ואפשר לצייר את הישר:

1. Spark (9 נק')

מה הפלט שיתקבל עבור קטע הקוד הבא:

lines = sc.textFile("test.txt")

lines.map(lambda s: (len(s), 1))

.reduceByKey(lambda a, b: a + b)

.sortByKey(false)

.collect()

אם התוכן המסמך test.txt הוא:

hello world

who are you !!!

hello hello

**נמק את תשובתך!**

[‘hello world’,’who are you !!!’,’hello hello’] -> [(11,1),(15,1),(11,1)] ->[(11,2),(15,1)]

* [(15,1),(11,2)] -> {15:1, 11:2}

ספירת שורות לפי אורכן. יש 2 שורות עם 11 אותיות ושורה אחת עם 15 אותיות.

הערה: למעשה הקוד יחזיר syntax error בגלל השימוש במילה false ולא False

בנוסף, נראה מתחכם ולא הוגן וגם לא נכון מבחינת מוסכמת קוד לכתוב sortByKey(false)

במקום sortByKey(ascending=False). האם המרצה בא לבדוק אם הסטודנט שינן את סדר האיברים שהפונקציה מקבלת? האם פחד שהקוד יהיה קריא וברור?