

Databases for data analytics

מועד ב 2025

מבוא

משך הבחינה: 3 שעות

חומר עזר:

1. דף נוסחאות המגיע עם המבחן
2. מילון, לצרכי סטודנטים הרוצים להימנע משגיאות כתיב. שימו לב, זהו מבחן חשיבה. לא יורדו נקודות על שגיאות כתיב.

טופס זה משמש גם לכלל הסטודנטים וגם למילואימניקים.
אנא ציינו אופן בולט בתחילת המחברת את סטוטוס המילואים שלכם.
הזכאים לבחינה מותאמת מתבקשים צריכים לענות רק על שאלות ה SQL. תשובה לשאלות האחרות תהווה בונוס.

לכל שאלה יש בונוס של חמש נקודות. הבונוסים הם מבוססים על הנלמד בקורס אבל ברמה גבוהה מהרגיל.
מומלץ לנסות לענות על הבונוסים אולם לאחר שפתרתם את כל הבחינה.

בהצלחה!

סכמה

השאלות בבחינה תתייחסנה לסכמה של [Financial](#) המובאת בתרשים למטה.

1. ניתן להניח ששדות המכילים את המילה Id הם מפתחות בטבלה, או קשורים למפתח בטבלה אחרת.
2. פרשנות שמות השדה היא לרוב מילולי. אם אתם לא בטוחים בפרשנות, ציינו את הפרשנות שבחרתם מפורשות ונמקו אותה.

loan	
loan_id	int
account_id	int
date	date
amount	int
duration	int
payments	decimal
status	varchar

order	
order_id	int
account_id	int
bank_to	varchar
account_to	int
amount	decimal
k_symbol	varchar

trans	
trans_id	int
account_id	int
date	date
type	varchar
operation	varchar
amount	int
balance	int
k_symbol	varchar
bank	varchar
account	int

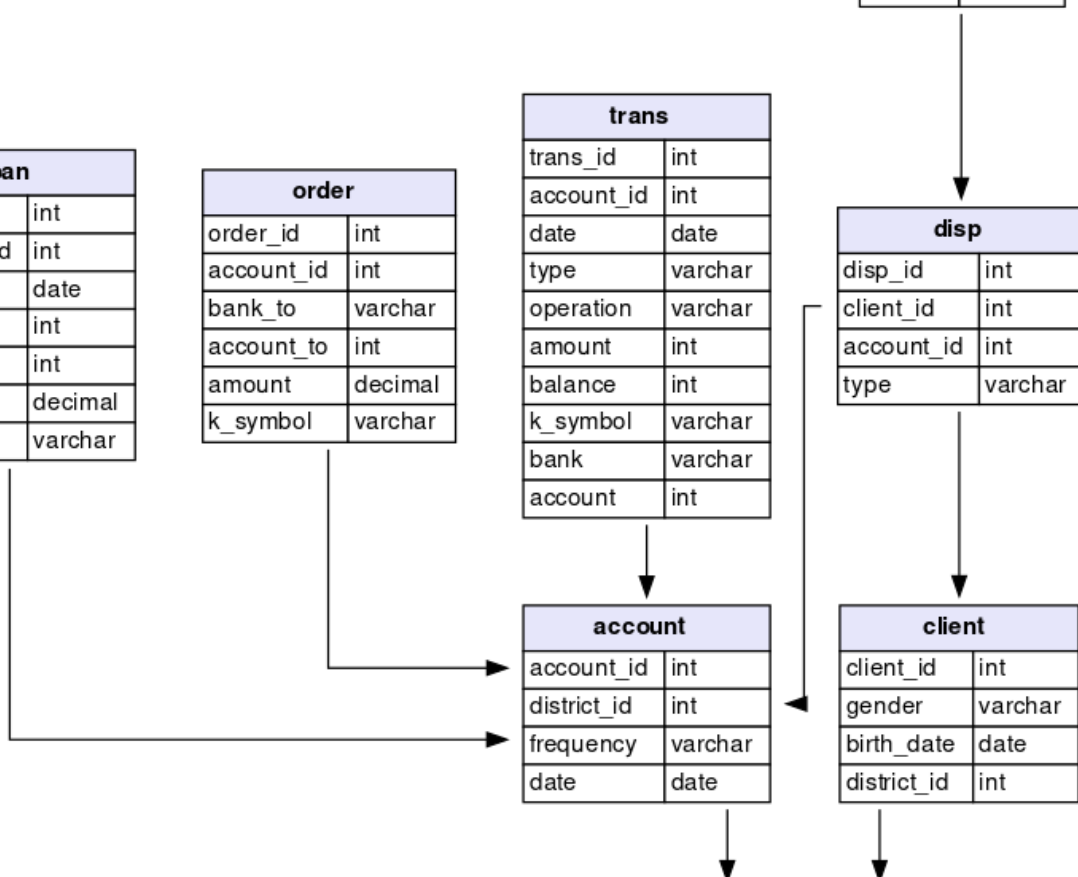
card	
card_id	int
disp_id	int
type	varchar
issued	date

disp	
disp_id	int
client_id	int
account_id	int
type	varchar

account	
account_id	int
district_id	int
frequency	varchar
date	date

client	
client_id	int
gender	varchar
birth_date	date
district_id	int

district	
district_id	int
A2	varchar
A3	varchar
A4	int
A5	int
A6	int
A7	int
A8	int
A9	int
A10	decimal
A11	int
A12	decimal
A13	decimal
A14	int
A15	int
A16	int



1. שליפות (35 נקודות)

במהלך הבחינה מותר, ואפילו רצוי, להשתמש בשאלות שכתבתם קודם.
כדאי לעשות זאת אנא השתמשו ב views וציינו בשאלתה המשתמשת את המקום בו הגדרתם אותה.

במידה ויש סעיף שהתקשתם בפיתורו, אתם יכולים לציין שם של view שפותר אותו, לכתוב מה השדות בו, ולהשתמש בו בסעיפים אחרים.

- a. שילפו את ה district_id של המחוזות עבורם A2 הוא Center
- b. שילפו את ה amount המקסימלי של loan
- c. שילפו את כל ה cards של clients שמשייכים למחוז עבורו $A4 = 5$
- d. שילפו את ה account שלא לקחו loan ב amount גדול ממליון.
- e. שילפו את מספר ה transactions לכל account, כולל אלו שאין להם transactions
- f. עבור accounts שיש להם transactions, שילפו את ה transactions per account distribution. כלומר, מספר הפעמים ש accounts ביצע טרנזקציה 1, 2, 3 וכדומה ממוינים בסדר עולה.
- g. הציגו את זוגות ה clients שנולדו באותו היום. שילפו באופן סימטרי, כלומר (a,b) וגם (b,a). האם כדאי לשלוח זוגות רפלקסיבים?
- h. **בנוס:**
 - i. שילפו כמחרוזת את סוגי הכרטיסים של לקוח בעזרת group_concat (חתימת כרטיסים).
 - ii. שילפו את מספר הלקוחות לחתימת כרטיסים.
 - iii. שילפו את ה distribution של הסעיף הקודם.
 - iv. **רמז:** אבסטרקציה היא חברה שיכולה לעזור לכם. השתמשו ב views כדי להוריד את המורכבות.
 - v. **המלצה:** זו שאלה קשה. גשו אליה רק אחרי שעניתם ובדקתם היטב את המבחן.

2. תקינות נתונים (30 נקודות)

- a. שדה to_bank בטבלת order הוא מסוג varchar. האם כדאי להחליפו בשדה המפוענח כ foreign key מטבלה חדשה של בנקים? שימו לב כי גם בטבלת transactions יש שדה bank. ציינו יתרונות וחסרונות ליצוג הנוכחי וליצוג בטבלת בנקים. אנא התייחסו ל-
 - i. חשיבות השדה
 - ii. סט הערכים אפשרי
 - iii. משמעות הטעות ואופן הטיפול בה
 - iv. אפשרויות אחרות למנוע שגיאות
- b.
 - 1. איזה סוג יחס יש בין client ו loan (אחד לאחד, אחד לרבים, רבים לרבים)?
 - 2. שילפו את סכום ההלוואות ל client. שימו לב לכלל הטבלאות שאתם צריכים להשתמש בהן.

3. כיצד היה ניתן ליצג את ישירות הלוואות של לקוח?
4. מימוש סכום ההלוואות ללקוח היה עצוב. אנא ציינו יתרון של שימוש ביצוג הנוכחי על היצוג שהצעתם.
- c. בטבלה account יש שדה בשם frequency מסוג varchar.
- מה לדעתכם משמעות השדה? כל תשובה קבילה כל עוד יש נימוק טוב.
 - מה לדעתכם הערכים בשדה?
 - כיצד הייתם מייצגים את השדה כדי לחשב ממוצעי תדירויות?
 - כיצד הייתם מייצגים את השדה כדי לאפשר תדירות שעתית, יומית ושבועית.
 - כיצד הייתם מייצגים את השדה כדי לאפשר את שני הסעיפים הקודמים יחד.
- d.
- בנוסף:** בבניית ה GS הזנתם זוגות של סרטים וציון. על פי רוב ההמלצות סימטריות, כלומר אם AB זוכה לציון גבוה, כך גם BA. האם צריך להזין את שתי האפשרויות? ציינו יתרונות וחסרונות ונמקו המלצתכם.
 - ישנם סוגות בהן ההמלצה אינה סימטרית. לדוגמה, הצד הלא מומלץ הוא ההמשך המיותר בסדרה, היוצר שטרם הבשיל, או שלהי הקריירה המוקדשים לרווח. האם צריך להזין את שתי האפשרויות? ציינו יתרונות וחסרונות ונמקו המלצתכם.
 - המליצו על דרך לאתר את זוגות שוברי סימטריה שבהן ההמלצה שלילית וממשו אותה.
 - מה החשיבות של זוגות אלו?

3. עיצוב (35 נקודות)

עצבו בסיס נתונים עבור מערכת שעות בית ספרית. המערכת צריכה לתמוך בשימושים הבאים.

- הצגת מערכת שעות (schedule) כיתתית. המערכת תציג לכל כיתה (class) (קבוצת תלמידים, כמו י"ב 6): חדר (classroom), מקצוע (course), מורה, ויום ושעה.
- הצגת כל המורים בשכבה (grade).
- תמיכה במורים המלמדים כמה מקצועות שונים.
- הצגת המורה של כל כיתה בכל מקצוע (ניתן להניח שמורה יחיד מלמד כיתה מסוימת מקצוע מסויים אולם אז יש לאכוף זאת).

- כיתבו קוד ליצירת הטבלאות הנדרשות. הקפידו על נרמול.
 - פרטו נמקו את כל ההחלטות העיצוביות שקיבלתם:
- ציינו מה כל הישויות שאתם מייצגים, מה התרחיש בשבילו הם נדרשים, ומה היצוג שנבחר. אם יש ישויות רלוונטיות שאתם לא מייצגים ציינו והסבירו למה.
 - הסבירו את השדות שבחרתם, התרחיש לו הם נדרשים, והטיפוס שבחרתם להם.
 - נמקו את המפתחות שבחרתם, האם הם טבעיים (קיימים בעולם האמיתי) או פנימיים (לדוגמה auto increment).
 - ציינו מה היחסים בין הישויות (1:1, n:1, או n:n) בעולם האמיתי. אם בחרתם ליצג יחס אחר, ציינו ונמקו. מספיק לציין יחסים רק בין ישויות שכנות, שיש בניהן קשר ישיר.
 - פרטו מה ה constraints (לדוגמה: unique, FK, not null, check) שבחרתם ומה ההגנה שהם מספקים. אם יש נקודות שלא מוגנות על ידי בסיס הנתונים ציינו אותן ונמקו איך להגן.

3. נמקו את האופן בו אתם מגנים על שעות וימי הלימוד.
4. נמקו את היצוג של היחס (מורה, כיתה, מקצוע). פרטו מה היתרונות והחסרונות של בחירתכם.
5. מה הייתם צריכים לשנות במערכת כדי לתמוך במקרים מתחלף מורה שמלמד כיתה מסויימת?
6. **בנוס:** ממשו את השאילתה "הצגת שעות החלון בבית הספר". שעת חלון היא שעה בה הכיתה לא לומדת כאשר יש שיעור לפניה ואחריה באותו היום.