

מדריך בסיסי לחוקי הנרמול

מחבר: איציק סיון

1 מבוא

כל בסיס נתונים טבלאי בנוי ממספר רב של טבלאות כאשר בכל טבלה מספר העמודות קבוע ולכל טבלה יש מפתח חד ערכי.

חוקי הנרמול, המוגדרים בתיאורית מסדי הנתונים הרלציוניים (הנקראים גם בסיסי- נתונים טבלאיים או יחסיים - Relational database theory), מייצגים עקרונות המשמשים לעיצוב טבלאות. קיימים 6 חוקי נרמול, אבל מעשית רק 3 החוקים הראשונים מעשיים עבור מנתחי המערכות.

העקרונות הנובעים מחוקי הנרמול מהראשון ועד השלישי, מוצגים כאן בצורה שאינה דורשת הבנה של התיאוריה הרלציונית. יתרה מכך, עקרונות אלה הם רבי ערך, אף למי שאינו משתמש במסד נתונים רלציוני.

חוקי הנרמול נועדו למנוע פערים בעדכון הנתונים וחוסר עקביות בהם. על פי חוקי הנרמול כל נתון שאינו מפתח, חייב להופיע רק פעם אחת במסד הנתונים ולכן אין חשש של חוסר עקיבות שיכול להיווצר כאשר נתון מסויים מופיע במקומות שונים ומעודכן על ידי יישומים שונים.

חוקי הנרמול הוגדרו ע"י Edgar Codd בשנת 1971.

2 חוק הנרמול הראשון (1NF-First Normal Form)

חוק הנרמול הראשון, דן ב"צורה" של הטבלה.

חוק הנרמול הראשון

- לכל שורה יש מפתח חד ערכי
- לכל משבצת בטבלה יש רק ערך אחד בלבד

לדוגמא: בטבלת הלקוחות המצורפת יש 4 עמודות

מספר לקוח	שם משפחה	שם פרטי	טלפון
1023	פוטיין	ולדימיר	0523-666678
1029	אובמה	חוסיין	0528-112234
2528	נתניהו	בנימין	0544-909090

ע"פ החוק הראשון, פיסית לא ייתכן המצב הבא:

מספר לקוח	שם משפחה	שם פרטי	טלפון
1023	פוטיין	ולדימיר	0523-666678
1029	אובמה	חוסיין	0528-112234
2528	נתניהו	בנימין	0544-909090 0528-333666 0528-101101

פתרון אפשרי א:

מזהה לקוח	שם משפחה	שם פרטי	טלפון 1	טלפון 2	טלפון 3
1023	פוטין	ולדימיר	0523-666678		
1029	אובמה	חוסיין	0528-112234		
2528	נתניהו	בנימין	0544-909090	0528-333666	0528-101101

טכנית הפתרון נכון, אבל לאורך זמן הפתרון בעיתי. מה יקרה אם בעוד חודש יהיו לפוטין 4 טלפונים? ולאחר מכן לאובמה 5 טלפונים. אנחנו עלולים למצוא את עצמנו מעדכנים את מבנה מסד הנתונים בתדירות גבוהה! וכמו כן אם נאפשר לדוגמא לכל לקוח 5 טלפונים הרי חלק ניכר ממספרי הטלפון בכל מסד הנתונים ישאר ריק!

פתרון מומלץ: פתרון באמצעות שתי טבלאות:

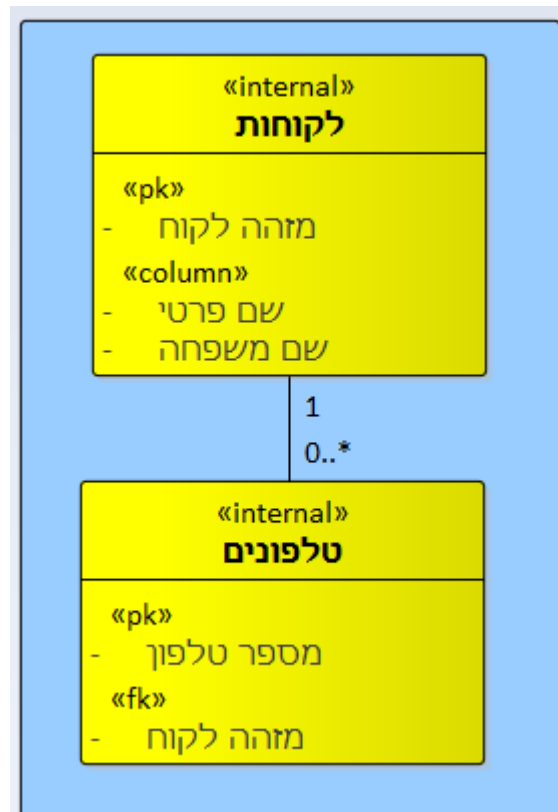
טבלת לקוחות

מזהה לקוח	שם משפחה	שם פרטי
1023	פוטין	ולדימיר
1029	אובמה	חוסיין
2528	נתניהו	בנימין

מספר טלפון	מספר לקוח
0523-666678	1023
0528-101101	2528
0528-112234	1029
0528-333666	2528
0544-909090	2528

טבלת טלפונים של לקוחות

המודל של הטבלאות ב UML נראה כך:



3 חוקי הנרמול השני והשלישי

חוקי הנרמול השני והשלישי דנים בקשר שבין שדות רגילים, לבין שדות שמהווים את מפתח הטבלה (Primary Key-PK).

שדה נחשב לשדה רגיל רק אם הוא אינו שדה המפתח או חלק ממנו וכן אם הוא אינו שדה שהיה עשוי היה להיבחר כמפתח (מפתח מועמד-Candidate Key)

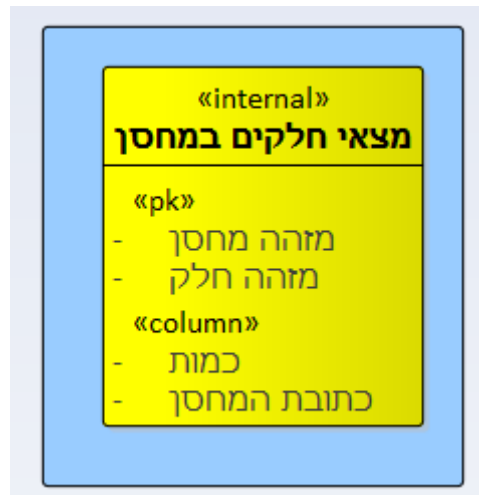
יש לזכור שמפתח של טבלה יכול להיות מורכב משדה אחד או מצירוף של מספר שדות

3.1 חוק הנרמול השני (2NF)

חוק הנרמול השני:

- הטבלה צריכה לעמוד בחוק הנרמול הראשון
- שדה רגיל צריך שיאפיין (יוסיף עובדה – fact) את כל שדות המפתח

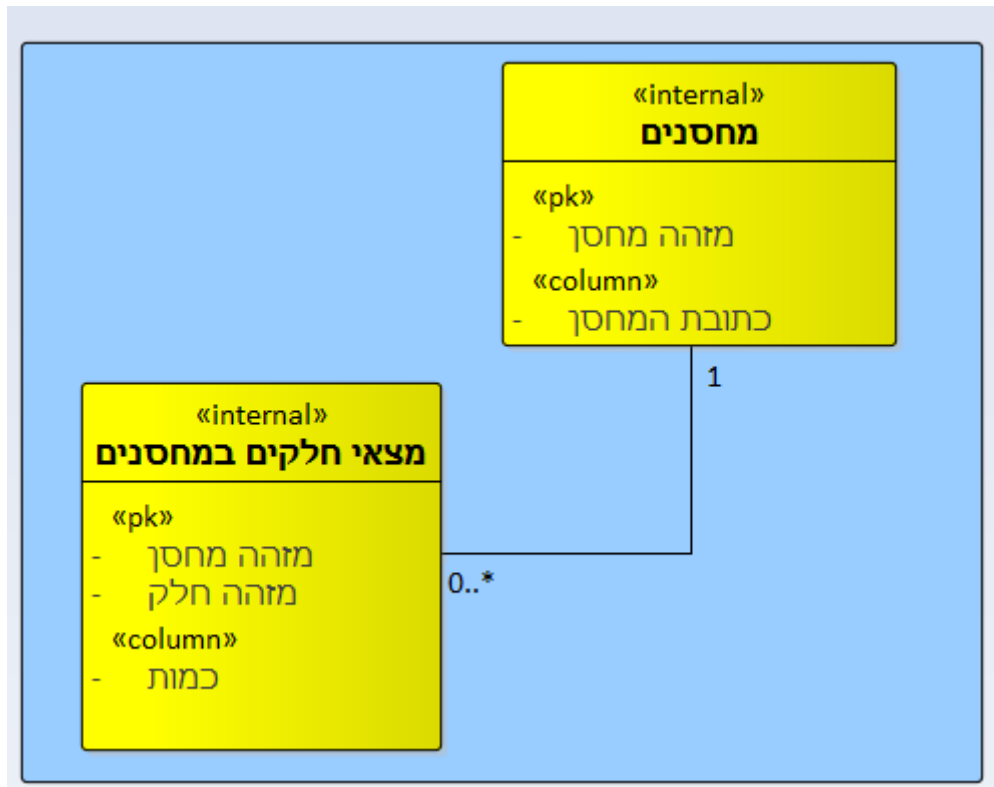
חוק הנרמול השני מופר כאשר שדה רגיל מאפיין רק חלק מהמפתח של הטבלה. הדבר רלוונטי כאשר המפתח של הטבלה מורכב ממספר שדות. בטבלת המצאי הבאה:



המפתח כאן מורכב מהשדות **חלק ו-מחסן**, אבל **כתובת-המחסן** הוא מאפיין של **המחסן** בלבד. הבעיות העיקריות בעיצוב זה הן:

- כתובת המחסן חוזרת בכל רשומה של חלק במחסן.
- אם כתובת המחסן משתנה, יש לעדכן כל אחת מהרשומות של החלקים המאוחסנים במחסן זה.
- בגלל הכפילות, יכולה להיווצר חוסר עקביות בנתון זה, כאשר ברשומות שונות יהיו כתובות שונות של אותו מחסן.
- אם בנקודת זמן מסוימת לא יהיו חלקים במחסן זה, יתכן שלא יהיה שדה בו ניתן לשמור את כתובת המחסן.

כדי לעמוד בחוק הנרמול השני, יש לפרק את הטבלה הנ"ל, ולהחליפה בשתי טבלאות:



תהליך זה, שבו משנים את עיצוב הנתונים ע"י החלפת טבלאות לא-מנורמלות בטבלאות מנורמלות, נקרא נרמול (normalization).

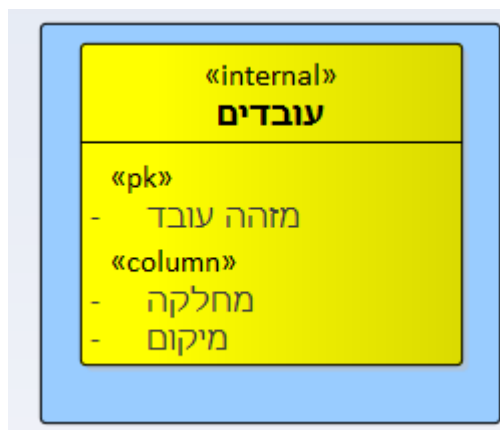
חוק הנרמול השלישי (3NF)

חוק הנרמול השלישי:

- א. הטבלה צריכה לעמוד בחוק הנרמול השני
- ב. שדה רגיל לא יכול להיות נגזר/מחושב משדות רגילים אחרים בטבלה הנוכחית או בטבלה אחרת כלשהיא

חוק הנרמול השלישי מופר כאשר שדה רגיל הוא מאפיין של שדה רגיל אחר

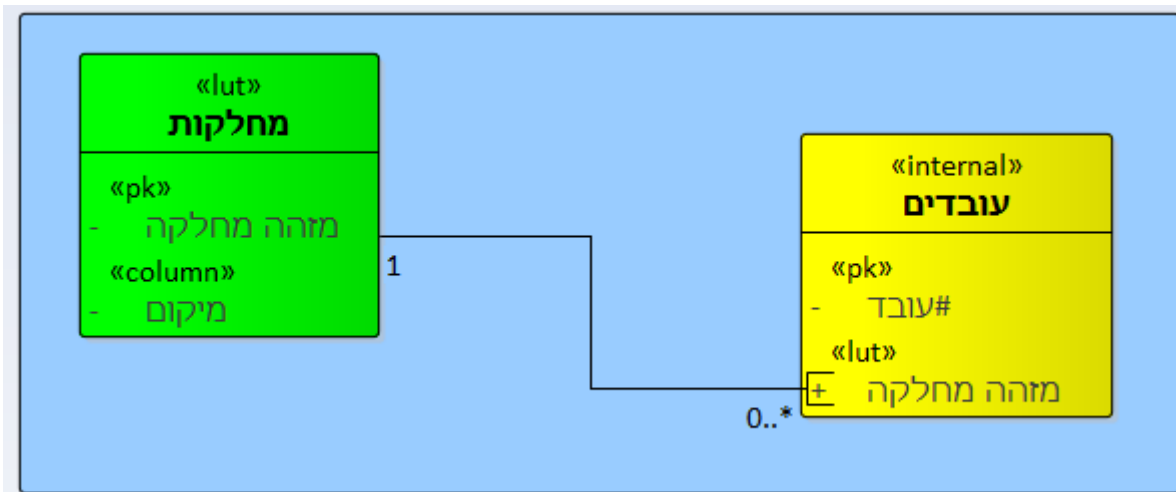
דוגמא 1



השדה **עובד** הוא שדה מפתח. אם כל מחלקה ממוקמת במקום אחר, אז השדה **מיקום** הוא מאפיין של השדה **מחלקה** – בנוסף על היותו מאפיין של **העובד**. הבעיות בעיצוב זה הן כמו אלה הנוצרות מהפרת חוק הנרמול השני:

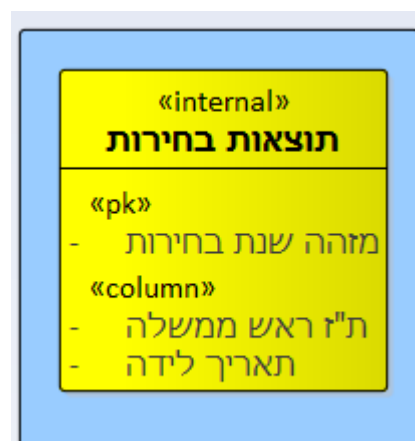
- מיקום המחלקה חוזר בכל רשומה של עובד המשויך למחלקה.
- אם מיקום המחלקה משתנה, כל רשומה כזאת צריכה להשתנות בהתאם.
- בגלל הכפילות, הנתון עלול להפוך לבלתי עקבי (inconsistent), כאשר רשומות שונות יציגו מיקומים שונים של אותה מחלקה.
- כאשר למחלקה אין עובדים, יתכן שלא יהיה שדה שישמור את מיקום המחלקה.

כדי לעמוד בחוק הנרמול השלישי, יש לפרק את הטבלה הנ"ל לשתי טבלאות:



לסיכום, טבלה היא מנורמלת על פי החוק השני והשלישי, אם כל שדה רגיל בה, תלוי בכל שדות המפתח ולא בשום שדה אחר

דוגמא 2:



הטבלה אינה עומדת בחוק השלישי כיון שתאריך הלידה תלוי אך ורק בת"ז של ראש הממשלה הפתרון הנכון:

