# Relational Database Design

לאחר שעיצבנו מסד נתונים, איך יודעים אם בנינו אותו נכון?
אם עיצוב הוא לא טוב, אז הוא יכול לגרום להמון בעיות(אנומליות). לכן אנו רוצים לבנות עיצוב טוב, עיצוב ללא בעיות.
בשביל זה צריך להבין את הסיבות לבעיות, ובאמצעות הידע הזה ניתן לשפר את העיצוב.
צריך קודם כל לבדוק את ה**הנחות** על הנתונים, ולפי זה לבנות תלויות פונקציונאליות.
בדוגמה שלנו:

| Flight | Class | Price | Airpl | Num |
|---|---|---|---|---|
| | | | | |

יש תלויות:

$$\text{Flight} + \text{Class} \to \text{Price}$$

$$\text{Flight} + \text{Price} \to \text{Class}$$

$$\text{Flight} \to \text{Airpl}$$

$$\text{Airpl} \to \text{Num}$$

נמצא מפתח ראשי:

$$K = \text{Flight,Class}$$

ואז

$$\begin{array}{lcl} \text{Flight}, \text{Class} \to \text{Price} & - & \text{FFD - Full FD} \\ \left.\begin{array}{l}\text{Flight,Class} \to \text{Airpl} \\ \text{Flight}, \text{Class} \to \text{Num}\end{array}\right\} & - & \text{PFD - Partial FD} \end{array}$$

יש לנו דברים שתלויים רק בחלק מהמפתח. נרצה לפרק את המפתח:

$$K - \text{key}$$

$$K' \subset K$$

$$K' \to X \qquad K \to X \text{ - PFD}$$

כלומר מפרקים למפתח, וכל מה שתלוי במפתח בצורה מלאה:

$$\begin{array}{lll} 1) & R1 & (K, S - K - X) \\ 2) & R2 & (K', X) \end{array}$$

$X$ זה מה שתלוי בחלק של המפתח(במקרה שלנו - רק בFlight.
עכשיו נחלק לשני טבלאות:

| R1 Flight | Class | Price |
|---|---|---|
| | | |

| R2 Flight | Airpl | Num |
|---|---|---|
| | | |

## עוד דוגמה

| a | b | c | d |
|---|---|---|---|
| | | | |

$a \to b$
$c \to d$

הפעם צריך לפרק לג יחסים(המפתח בכל יחס מסומן בקו תחתון):

$R1\,(\underline{a}, \underline{c})$ $\quad R2\,(\underline{a}, b)$ $\quad R3\,(\underline{c}, d)$

## עוד דוגמה

| a | b | c | d | e |
|---|---|---|---|---|
| | | | | |

$a, b \to c$
$a \to d$
$b \to e$

$R1\,(\underline{a}, \underline{b}, c)$ $\quad R2\,(\underline{a}, d)$ $\quad R3\,(\underline{b}, e)$

## נחזור לדוגמה המקורית

נשים לב שביחס $R2$ יש תלות טרנזיטיבית - Flight $\to$ Airpl $\to$ Num. לכן ניתן לפרק אותו:

| Flight | Airpl | Num |
|---|---|---|
| | | |

$\Rightarrow$

| Airpl | Num |
|---|---|
| | |

,

| Flight | Airpl |
|---|---|
| | |

# BCNF - Boyce-Codd Normal Form

## הגדרה

יחס $R$ נמצא בBCNF אם ורק אם לכל תלות פונקציונלית לא טריוויאלית $X \to Y \in R$, $X$ הוא superkey.

מתברר שתנאי כזה מבטיח שלא יהיו בעיות כמו כפילויות או אנומליות ביחס $R$.

## בדוגמה שלנו

התלויות הן

| | | |
|---|---|---|
| Flight,Class | $\to$ | Price |
| Flight,Price | $\to$ | Class |
| Flight | $\to$ | Airpl |
| Airpl | $\to$ | Num |

האם היחס הראשון בBCNF?

| Flight | Class | Price | Airpl | Num |
|---|---|---|---|---|

נבדוק את התלויות:

| | | | |
|---|---|---|---|
| $\surd$ | Flight,Class | $\to$ | Price |
| $\surd$ | Flight,Price | $\to$ | Class |
| $\chi$ | Flight | $\to$ | Airpl |
| $\chi$ | Airpl | $\to$ | Num |

לכן היחס אינו בBCNF.

נבדוק את הפירוק השני

| Flight | Class | Price |
|---|---|---|

| Flight | Airpl | Num |
|---|---|---|

את שני התלויות הראשונות לא צריך לבדוק, כי הם היו בסדר גם לפני הפירוק.

| | | | |
|---|---|---|---|
| $\surd$ | Flight | $\to$ | Airpl |
| $\chi$ | Airpl | $\to$ | Num |

גם לא בBCNF. נבדוק את השלישי:

| Flight | Class | Price |
|---|---|---|

| Flight | Airpl |
|---|---|

| Flight | Num |
|---|---|

| | | | |
|---|---|---|---|
| $\surd$ | Airpl | $\to$ | Num |

לכן היחס השלישי בBCNF!

נשים לב שכדי להגיע לBCNF, היינו צריכים לפרק את היחס שוב ושוב.

## משפט

אם יחס $R$ הוא בינארי, אזי $R \in \text{BCNF}$

### הוכחה

נבדוק את כל המקרים עבור היחס הבינארי $R(A,B)$

1. אין תלויות לא טריוויאליות - $F = \phi$ מכיוון שאין תלויות לא טריוויאליות, אין מה שיפר את התנאי של BCNF, ולכן $R \in \text{BCNF}$.
במקרה הזה $K = A, B$, אבל זה לא נחוץ להוכחה.
2. $A \to B, B \nrightarrow A$. אזי $K = A$, ולכן זה ממלא את התנאי של BCNF.
3. $B \to A, B \nrightarrow A$. אזי $K = B$, ולכן זה ממלא את התנאי של BCNF.
4. $A \to B, B \to A$. במקרה הזה יש שני מפתחות אפשריים:

$$K_1 = A, K_2 = B$$

התלות הראשונה מתאימה למפתח $K_1$. התלות השנייה מתאימה למפתח $K_2$. לכן שתיהן עונות על התנאי של BCNF.

### הערה

תלות לא צריכה להכיל את כל המפתחות האפשריים - מספיק שהיא מכילה אחד מהם כדי למלא את התנאי.

## אלגוריתם Decomposition

נתון יחס $R0\,(S0)$ וקבוצת תלויות פונקציונליות $F0$. קוראים לפונציה

```
dec_BCNF(S0,F0);
```

המוגדרת ע"י

```
dec_BCNF(S,F)
    if(S∈BCNF) return S;
    else
    if(X → Y ∈ F violates BCNF)
    {
        compute X+;
        S1=X+;S2=(X, S − X+);
        compute F1; compute F2;
        return(dec_BCNF(S1,F1)∪dev_BCNF(S2,F2));
    }
```

הרעיון הוא שכל פעם מפרקים למה שתלוי ב$X$($S1$), ולמה שלא תלוי ב$X$($S2$)(שכולל גם את $X$ עצמו, אבל לא שדות שתלויים בו ולא נמצאים בו)

## דוגמה 1

נתון יחס

| St<br>stud | course | sem | lect | time |
|---|---|---|---|---|
| Avi | 89-01 | b | Lavan | Mon 17-19 |
| Avi | 89-01 | b | Lavan | Wed 12-14 |
| Beni | 89-02 | b | Kahoe | Wed 12-14 |

הסכימה היא

$$S = (\text{stud,course,sem,lect,time})$$

התלויות הפונקציונאליות הן

$$F = \left\{ \begin{array}{c} \text{stud,course} \rightarrow \text{sem} \\ \text{stud,course} \rightarrow \text{lect} \end{array} \right\}$$

צד שמאל לא מפתח. נחשב

$$(\text{stud,course})^+ = \text{stud,course,sem,lect}$$

$$S1 = (\underline{\text{stud}}, \underline{\text{course}}, \text{sem,lect})$$

$$S2 = (\underline{\text{stud}}, \underline{\text{course}}, \underline{\text{time}})$$

$$F1 = \left\{ \begin{array}{c} \text{stud,course} \rightarrow \text{sem} \\ \text{stud,course} \rightarrow \text{lect} \end{array} \right\}$$

$$F2 = \emptyset$$

$$\text{dec_BCNF}(S1, F1) \Rightarrow \text{BCNF}$$

$$\text{dec_BCNF}(S2, F2) \Rightarrow \text{BCNF}$$

והגענו לתוצאה תוך שלב אחד.

## דוגמה 2

| St2<br>Stud | Course | Lect | L_city | L_region |
|---|---|---|---|---|
| Avi | 89-01 | Lavan | RG | Center |
| Avi | 89-01 | Lavan | RG | Center |
| Beni | 89-02 | Kahoe | Haifa | North |

$$S = (\text{Stud,Course,Lect,L_city,L_region})$$

$$F = \left\{ \begin{array}{c} \text{Stud,Course} \rightarrow \text{Lect} \\ \text{Lect} \rightarrow \text{L_city} \\ \text{L_city} \rightarrow \text{L_region} \end{array} \right\}$$

התלות הראשונה בסדר, אבל השניה לא. לא צריך לבדוק את השלישית - אפשר ישר להתחיל לעבוד על השנייה.

$$\text{Lect}^+ = \text{Lect,L_city,L_region}$$

נפרק את היחס

$$S1 = \text{Lect,L_city,L_region}$$

$$S2 = \text{Lect,Stud,Course}$$

$$F1 = \left\{ \begin{array}{c} \text{Lect} \rightarrow \text{L_city} \\ \text{City} \rightarrow \text{L_region} \end{array} \right\}$$

$$F2 = \{\text{Stud,Course} \rightarrow \text{Lect}\}$$

נתחיל ב$F1$. התלות הראשונה בסדר, אבל השנייה לא. צריך להפעיל את האלגוריתם:

$$\text{L_city} \rightarrow \text{L_region}$$

$$\text{L_city}^+ = \text{L_city,L_region}$$

$$S1' = \text{L_city,L_region}$$

$$S2' = \text{L_city,Lect}$$

$$F1' = \text{L_city} \rightarrow \text{L_region}$$

$$F2' = \text{Lect} \rightarrow \text{L_city}$$

וסיימנו