

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 ההנחות על הנתונים, ולפי זה לבנות תלויות פונקציונאליות. בדוגמה שלנו:

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

יש תלויות:

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

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

$$\text{Flight} \rightarrow \text{Airpl}$$

$$\text{Airpl} \rightarrow \text{Num}$$

נמצא מפתח ראשי:

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

ואז

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

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

$$K - \text{key}$$

$$K' \subset K$$

$$K' \rightarrow X \quad K \rightarrow X - \text{PFD}$$

כלומר מפרקים למפתח, וכל מה שתלוי במפתח בצורה מלאה:

- 1) $R1 \quad (K, S - K - X)$
- 2) $R2 \quad (K', X)$

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

R1			R2		
Flight	Class	Price	Flight	Airpl	Num

עוד דוגמה

a	b	c	d

$a \rightarrow b$
 $c \rightarrow d$

הפעם צריך לפרק ל3 יחסים(המפתח בכל יחס מסומן בקו תחתון):

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

עוד דוגמה

a	b	c	d	e

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

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

נחזור לדוגמה המקורית

נשים לב שביחס $R2$ יש תלות טרנזיטיבית - $Flight \rightarrow Airpl \rightarrow Num$. 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 $X \rightarrow Y \in R$, X הוא superkey.

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

בדוגמה שלנו

התלויות הן

Flight, Class \rightarrow Price
Flight, Price \rightarrow Class
Flight \rightarrow Airpl
Airpl \rightarrow Num

האם היחס הראשון ב-BCNF?

Flight | Class | Price | Airpl | Num

נבדוק את התלויות:

✓ Flight, Class \rightarrow Price
✓ Flight, Price \rightarrow Class
✗ Flight \rightarrow Airpl
✗ Airpl \rightarrow Num

לכן היחס אינו ב-BCNF.
נבדוק את הפירוק השני

Flight | Class | Price Flight | Airpl | Num

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

✓ Flight \rightarrow Airpl
✗ Airpl \rightarrow Num

גם לא ב-BCNF. נבדוק את השלישי:

Flight | Class | Price Flight | Airpl Flight | Num

✓ Airpl \rightarrow Num

לכן היחס השלישי ב-BCNF!
נשים לב שכדי להגיע ל-BCNF, היינו צריכים לפרק את היחס שוב ושוב.

משפט

אם יחס R הוא בינארי, אזי $R \in BCNF$

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 $R \in BCNF$. במקרה הזה $K = A, B$ אבל זה לא נחוץ להוכחה.
2. $A \rightarrow B, B \not\rightarrow A$. אזי $K = A$, ולכן זה ממלא את התנאי של BCNF.
3. $B \rightarrow A, B \not\rightarrow A$. אזי $K = B$, ולכן זה ממלא את התנאי של BCNF.
4. $A \rightarrow B, B \rightarrow A$. 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 שתייהן 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 $R_0(S_0)$ וקבוצת תלויות פונקציונליות F_0 . קוראים לפונציה

$dec_BCNF(S_0, F_0)$;

המוגדרת ע"י

```
dec_BCNF(S,F)
  if(S∈BCNF) return S;
  else
  if( $X \rightarrow Y \in F$  violates BCNF)
  {
    compute  $X^+$ ;
     $S_1=X^+; S_2=(X, S - X^+)$ ;
    compute F1; compute F2;
    return( $dec\_BCNF(S_1, F_1) \cup dev\_BCNF(S_2, F_2)$ );
  }
```

הרעיון הוא שכל פעם מפרקים למה שתלוי ב $X(S_1)$, ולמה שלא תלוי ב $X(S_2)$ (שכולל גם את X עצמו, אבל לא שדות שתלויים בו ולא נמצאים בו)

דוגמה 1

נתון יחס

St 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

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

התלויות הפונקציונאליות הן

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

צד שמאל לא מפתח. נחשב

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

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

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

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

$$F2 = \emptyset$$

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

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

והגענו לתוצאה תוך שלב אחד.

דוגמה 2

St2 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

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

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

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

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

נפרק את היחס

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

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

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

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

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

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

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

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

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

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

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

וסיימנו