נניח שיש לנו יחס עם סכימה $R\left(S\right)$, ואוסף של תלויות פונקציונאליות F.
כדי לנרמל את R, פירקנו אותו:

$$\begin{tikzcd}
 & & R \arrow[dl] \arrow[dr] & \\
 & R1\left(S1\right) \arrow[dl] \arrow[dr] & & R2\left(S2\right) \\
R3\left(S3\right) & & R4\left(S4\right) &
\end{tikzcd}$$

אבל האם ניתן לחזור מזה ליחס המקורי?
ישנן כמה תכונות של פירוק נכון:

Decomposition Properties

1) Elimination of anomalies $\surd$

האם חיסלנו את האנומליות כשפירקנו את R?

2) Recoverrablity of info $\surd$

האם ניתן לשחזר את המידע מהפירוק?

3) Preservation of dependencies χ

זה לא מתקיים ב$BCNF$.

נוכיח את התכונות:

1. למה אם יחס הוא ב$BCNF$ אז אין אנומליות ואין כפילות?

נחזור להגדרה של $BCNF$: יחס הוא ב$BCNF$ אם כל תלות לא טריוויאלית $X \to Y$ היא חלק ממפתח על. זה אומר שכל הtupleים שונים על X - ולכן אין אנומליות ואין כפילות.

2. למה ניתן לשחזר את המידע מהפירוק?

כשמצרפים את כל היחסים בפירוק, מקבלים איזהו יחס. יש גם פירוק של S לk סכימות:

$$S = S_1 \cup S_2 \cup \cdots \cup S_k$$

ניקח הטלות של R על הסכימות שמרכיבות את הפירוק הזה:

$$\pi_{S_1}R, \pi_{S_2}R, \ldots \pi_{S_k}R$$

איך אנחנו יכולים לחזור לR? נצרף אותם! אם מצרפים את ההטלות האלה, מקבלים יחס R':

$$R' = \pi_{S_1} R \bowtie \pi_{S_2} R \bowtie \ldots \bowtie \pi_{S_k} R$$

האם $R' \stackrel{?}{=} R$?

- אם $R' = R$, אז הפירוק נקרא loseless-join decomposition
- אם $R' \neq R$, אז הפירוק נקרא lossy-join decomposition

משפט

נתון יחס עם סכימה $R(S)$, $S = X \cup Y \cup Z$, וקבוצת תלויות פונקציונאליות $F = \{Y \to Z\}$.
אזי הפירוק

$$\pi_{X \cup Y} R \bowtie \pi_{y \cup Z} R = R$$

הוא loseless.

דוגמה

X	Y	Z
x_1	y	z
x_2	y	z

X	Y
x_1	y
x_2	y

Y	Z
y	z

X	Y	Z
x_1	y	z
x_2	y	z

לעומת זאת, אם ניקח פירוק שלא מקיים את התנאי:

$$R(X, Y, Z), Y \nrightarrow X, Y \nrightarrow Z$$

ומפרקים אותו

X	Y	Z
x_1	y	z_1
x_2	y	z_2

X	Y
x_1	y
x_2	y

Y	Z
y	z_1
y	z_2

X	Y	Z
x_1	y	z
x_2	y	z

אם יש לנו פירוק $BCNF$, אנחנו יודעים לפי המשפט שהוא loseless. אבל מה עושים אם נותנים לנו סתם פירוק?

בדיקה אם פירוק הוא loseless - Chase test

טענות:

1. $\bowtie$(צירוף טבעי)הוא אסוציאטיבית וקומוטטיבית.
   - אסוציאטיבית:

$$((((R1 \bowtie R2) \bowtie R3) \dots) \bowtie Rn)$$

   - קומוטטיבית:

$$R1 \bowtie R2 = R2 \bowtie R1$$

2. $R \subseteq R'$ - תמיד נכון
3. $R' \subseteq R$

לכן מה שצריך לבדוק את זה תכונה 3 - שאם $t \in R'$ אז $t \in R$. זהו הChase test.

Tableau

Chase test עובד בעזרת מבנה שנקרא "Tableau".

נתון יחס עם סכימה $R\left(S\right)$ ופירוק $S = S1 \cup \cdots \cup Sk$. נבנה Tableau:

לכל Si בונים tuple t_i, כך שהשדות המופיעים בסכימה יופיעו בלי אינדקס, ואלו שלא בסכימה מופיעים בלי אינדקס:

R(A, B, C, D)	A	B	C	D
S1=(A, D)	a	b_1	c_1	d
S2=(A, B)	a	b_2	c	d_2
S3=(B, C, D)	a_3	b	c	d

זהו הTableau בשביל הפירוק - כל איבר עם אינדקס מופיע רק פעם אחת(שכן בשורות אחרות יהיה אינדקס אחר)

הרעיון הוא לקחת tuple מR', $t \in R'$ ולבדוק האם $t \in R$.

Chase Test

מסתמכים על תלויות פונקציונליות:

$$\begin{matrix} X \to Y \\ t_1(X) = t_2(X) \end{matrix} \Rightarrow t_1(Y) = t_2(Y)$$

בכל פעם לוקחים שני tupleים שבהם יש לנו ערכים שווים:

1. $t_1(x) = t_2(x)$
   x-unsubscribed $\Rightarrow$ y-unsubscribed
2. $t_1(x) = t_2(x)$
   x-subscribed
   y-subscribed $\Rightarrow$ one of indexes

דוגמה

$$F = \{A \to B, B \to C, CD \to A\}$$

R(A,B,C,D)	A	B	C	D
S1=(A,D)	a	b_1	c_1	d
S2=(A,B)	a	b_2	c	d_2
S3=(B,C,D)	a_3	b	c	d

נתון$A \to B$, ויש לנו שני tupleים, $(a, b_1, \dots), (a, b_2, \dots)$ - לכן זהו אותו b:

R(A,B,C,D)	A	B	C	D
S1=(A,D)	a	b_1	c_1	d
S2=(A,B)	a	b_1	c	d_2
S3=(B,C,D)	a_3	b	c	d

עכשיו, נתון $B \to C$, ויש לנו שני tupleים עם b_1, אחד עם c ואחד עם c_1 - לכן נוריד את האינדקס מc_1:

R(A,B,C,D)	A	B	C	D
S1=(A,D)	a	b_1	c	d
S2=(A,B)	a	b_1	c	d_2
S3=(B,C,D)	a_3	b	c	d

כמו כן $CD \to A$, לכן נשנה את a_3 לa(שכן בשורה הראשונה והשלישית c, d זהים):

R(A,B,C,D)	A	B	C	D
S1=(A,D)	a	b_1	c	d
S2=(A,B)	a	b_1	c	d_2
S3=(B,C,D)	a	b	c	d

קיבלנו tuple (a, b, c, d) - וקיבלנו tuple ששייך למקור, לכן עברנו Chase test.

דוגמה לפירוק loosy

$$R\left(X,Y,Z\right)$$

$$S_1=(X,Y) \qquad S_2=(Y,Z)$$

R(X,Y,Z)	X	Y	Z
S1=(X,Y)	x	y	z_1
S2=(Y,Z)	x_2	y	z

$$F=\{X\to Y,Z\to Y\}$$

התלות הראשונה היא $X \to Y$, אבל אין שתי שורות עם X זהה, לכן זה לא עוזר לנו.
התלות השנייה היא $Z \to Y$, אבל אין שתי שורות עם Z זהה, לכן זה לא עוזר לנו.
מכיוון שנתקענו בלי למצוא (x,y,z), לא עברנו Chase Test, והפירוק הוא loosy-join decomposition.

3) Dependency Preservation χ

נתון יחס עם סכימה

$$R\left(\text{stud,course,lect}\right)$$

ותלויות פונקציונליות

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

נפרק לפי התלות lect $\to$ course:

$$S1=(\text{lect,course}) \qquad S2=(\text{lect,stud})$$

$$F1=\{\text{lect}\to\text{course}\} \qquad S2=\emptyset$$

בעיצוב המקורי היו לנו שני תלויות, אבל נשארנו עם תלות אחת.
פירוק BCNF **<u>לא</u>** שומר על תלויות פונקציונאליות!!! בעיצוב לעיל, המצב הבא הוא חוקי:

lect	course
Lahol	DB
Lavan	DB

$\bowtie$

lect	stud
Kahol	Avi
Lavan	Avi

=

lect	course	stud
Kahol	DB	Avi
Lavan	DB	Avi

יש כאן tradeoff - אנחנו מחסלים אנומליות, אבל מאבדים תלויות פונקציונאליות.
אפשר קצת להחליש את התנאים של BCNF, כדי לשמור על תלויות פונקציונאליות - על חשבון השארת כפילויות. כך מגיעים לצורה נורמלית שלישית

3NF - Third Normal Form

הגדרה

אומרים שיחס הוא 3NF אם לכל תלות לא טריוויאלית $X \to Y$, או שX הוא superkey, או שלכל $y_i \in Y \setminus X$, $y_i \in \text{some key}$(לאו בהכרח כולם באותו מפתח).

במקרה השני אומרים ש$y_i \in \text{key}$ הוא prime.

לכן ניתן לכתוב את התנאי ל3NF בתור

$$\forall_{X \to Y \text{ non-trivial}} \bigvee \begin{matrix} X \text{ is superkey} \\ \forall_{y_i \in X \setminus X} y_i \text{ is prime} \end{matrix}$$