תיאוריית תכנון סכמות למסדי נתונים יחסיים חלק 4

Design Theory for Relational Databases

Part 4

1

צורה נורמלית של Boyce-Codd

Boyce-Codd Normal Form (BCNF)

2

דוגמה לסכמה בעייתית

- היחס

Student
Course
Teacher

S	C	T
Cohen	DB	Smith
Levy	OS	Jones
Levy	DB	Smith

- הסכמה היא SCT ומתקיימת הת"פ $C \rightarrow T$
- C אינו מפתח-על של הסכמה

3

אפיון סכמה בעייתית

- נתונה לנו סכמה R (קבוצת אטריביוטים)
- ונתונה קבוצה של ת"פ F שהיחסים עבור R צריכים לקיים
- הסכמה בעייתית אם קימת ת"פ $Y \rightarrow X$ לא טריוויאלית (קרי, $X \subsetneq Y$) כך ש-
  - $F \models X \rightarrow Y$
  - ו- X אינו מפתח-על

4

צורה נורמלית של Boyce-Codd

Boyce-Codd Normal Form (BCNF)

- סכמה R עם קבוצת ת"פ F היא *בצורה נורמלית של Boyce-Codd* (BCNF) אם לכל $F^+ \ni A \rightarrow X$ מתקיים
  - $A \rightarrow X$ היא טריוויאלית, כלומר $A \in X$, או
  - X הוא מפתח-על, כלומר $F \models X \rightarrow R$

5

אלגוריתם למציאת פירוק ב-BCNF

6

המטרה

- בהינתן סכמה R עם קבוצת ת"פ F, צריך למצוא פירוק $R_1, R_2, \ldots, R_n$ כך שמתקיימים הדברים הבאים:
  - הפירוק חסר אובדן
  - הפירוק משמר את התלויות
  - כל סכמה R_i הנה בצורה נורמלית

7

תחילה צריך לבדוק האם יש הפרה של BCNF

- מתחילים עם סכמה R וקבוצת ת"פ F
- מחפשים הפרה של BCNF, כלומר ת"פ $X \to A \in F^+$, כך ש-
  - $A \notin X$ וגם
  - $F \nvDash X \to R$
- טענה: אם יש הפרה ב- F^+, אז יש הפרה ב- F; לכן ניתן למצוא הפרה (אם יש כזאת) בזמן פולינומיאלי

8

דוגמה לפירוק המתבצע כשיש הפרה

- נתונה הסכמה ABC עם $F = \{A \to B, B \to C\}$
- $B \to C$ היא הפרה של BCNF, כי $F \nvDash B \to ABC$
- לכן נפרק לשתי סכמות:
  - סכמה אחת היא BC
    - סכמה זו כוללת את האטריביוטים של הת"פ המפרה
  - סכמה שנייה היא AB
    - סכמה זו כוללת את כל האיטריביוטים של הסכמה המקורית, פרט לאטריביוט בצד ימין של הת"פ המפרה

9

במקרה הכללי

- אם הת"פ $X \to A$ היא הפרה של BCNF בסכמה R, אז מפרקים את R לשתי סכמות:
  - סכמה XA (כל האטריביוטים של הת"פ המפרה)
  - סכמה R-A (כל האטריביוטים של R פרט לאטריביוט בצד ימין של הת"פ המפרה)
- בכל אחת מהסכמות החדשות, $X \to A$ היא לא הפרה של BCNF
  - לכן, בכל אחת מהסכמות החדשות יש פחות הפרות של BCNF מאשר בסכמה המקורית

10

המשך התהליך של מציאת פירוק BCNF

- עבור כל אחת מהסכמות החדשות נחשב כיסוי של הת"פ, כלומר אם S סכמה חדשה נחשב את $\pi_S(F)$
- נמשיך באופן רקורסיבי עם כל אחת מהסכמות הקיימות
- תמיד מחליפים סכמה בשתי סכמות, שבכל אחת פחות אטריביוטים
- האלגוריתם מסתיים כי סכמה של שני אטריביוטים תמיד מקיימת BCNF

11

דוגמה

- R=ABCD ו- $F = \{A \to B, B \to C, C \to D\}$
- $B \to C$ היא הפרה של BCNF
- לכן נפרק את R לשתי סכמות:
  - $R_1=BC$ ו- $F_1 = \{B \to C\}$
  - $R_2=ABD$ ו- $F_2 = \{A \to B, B \to D\}$
- הת"פ $B \to D$ מפרה את BCNF ב- R_2 ולכן נפרק את R_2 ל-
  - $R_2=BD$ ו- $F_2 = \{B \to D\}$
  - $R_3=AB$ ו- $F_3 = \{A \to B\}$

12

דוגמה (המשך)

- הפירוק הסופי הוא:
  - $R_1=BC$ ו- $F_1 = \{B \to C\}$
  - $R_2=BD$ ו- $F_2 = \{B \to D\}$
  - $R_3=AB$ ו- $F_3 = \{A \to B\}$

$\boxed{F = \{A \to B, B \to C, C \to D\}}$

- הערה: לכל i, הקבוצה F_i היא כיסוי חסר כפילויות של $\pi_{R_i}(F)$ והחישוב שלה דורש (במקרה הכללי) זמן אקספוננציאלי
- איחוד ה- F_i אינו שקול ל- F המקורי, ולפיכך אין שימור של הת"פ

13

פירוק בדרך אחרת של הדוגמה הקודמת

- $R=ABCD$ ו- $F = \{A \to B, B \to C, C \to D\}$
- נפרק תחילה לפי $C \to D$ (ולא לפי $B \to C$)
- לכן נפרק את R לשתי סכמות:
  - $R_1=CD$ ו- $F_1 = \{C \to D\}$
  - $R_2=ABC$ ו- $F_2 = \{A \to B, B \to C\}$
- הת"פ $B \to C$ מפרה את BCNF ב- R_2 ולכן נפרק את R_2 ל-
  - $R_2=BC$ ו- $F_2 = \{B \to C\}$
  - $R_3=AB$ ו- $F_3 = \{A \to B\}$

14

פירוק בדרך אחרת (המשך)

- הפירוק הסופי הוא:
  - $R_1=CD$ ו- $F_1 = \{C \to D\}$
  - $R_2=BC$ ו- $F_2 = \{B \to C\}$
  - $R_3=AB$ ו- $F_3 = \{A \to B\}$
- הפעם איחוד ה- F_i שקול ל- F המקורי ולכן יש שימור של הת"פ

15

למה לגבי פירוק ללא אובדן

- תהי R סכמה עם קבוצת ת"פ F
- *למה:* פירוק של R לשתי סכמות R_1 ו- R_2 הנו ללא אובדן אם ורק אם מתקיים אחד משני התנאים הבאים:
  - $F \models R_1 \cap R_2 \to R_1$ או
  - $F \models R_1 \cap R_2 \to R_2$
- *מסקנה:* פירוק לשתי סכמות לפי הפרה של BCNF הוא תמיד ללא אובדן

16

תכונות הפירוק ל- BCNF

- הפירוק שהאלגוריתם מייצר הנו בעל התכונות הבאות:
  - כל סמכה בפירוק הנה BCNF
  - הפירוק הנו ללא אובדן
  - הפירוק אינו בהכרח משמר את הת"פ
- זמן הריצה של האלגוריתם הנו אקספוננציאלי בגלל חישוב $\pi_{R_i}(F)$ לכל i
- יש אלגוריתם פולינומיאלי (ספר של אולמן)

17

אולי תמיד אפשר למצוא לפחות פירוק אחד ל- BCNF שמשמר את הת"פ?

18

דוגמה

- נתונה הסכמה ABC עם
  $F = \{AB \to C, C \to A\}$
- $C \to A$ היא ההפרה היחידה של BCNF, ולכן צריך לפרק לפי ת"פ זו ומקבלים:
  - סכמה CA עם הת"פ $C \to A$
  - וסכמה BC ללא ת"פ
- *מסקנה:* אין שום פירוק ל- BCNF של הסכמה הנ"ל שמשמר את הת"פ

19

מה גרם לאי-שימור הת"פ?

- כשפרקנו את ABC, פיצלנו את המפתח AB של הסכמה המקורית בין שתי הסכמות החדשות
  - A בסכמה אחת
  - B בסכמה השניה
- כתוצאה מכך הת"פ $AB \to C$ הלכה לאיבוד
- מסקנה: אסור לפצל מפתחות אם רוצים לשמר את הת"פ

20

צורה נורמלית שלישית

Third Normal Form (3NF)

21

צורה נורמלית שלישית
Third Normal Form (3NF)

- סכמה R עם קבוצת ת"פ F היא *בצורה נורמלית שלישית* (3NF) אם לכל
  $X \to A \in F^+$ מתקיים אחד הדברים הבאים:
  - $X \to A$ היא טריוויאלית, כלומר $A \in X$,
  - X הוא מפתח-על, כלומר $F \models X \to R$, או
  - האטריביוט A שייך למפתח כלשהו של R

22

אלגוריתם למציאת פירוק המקיים 3NF

- מתחילים עם הסכמה R וקבוצת הת"פ F
- מוצאים כיסוי חסר כפילויות G של F
- מכל ת"פ $X \to A \in G$ מייצרים סכמה XA (שהמפתח שלה הוא X)

23

תכונות הפירוק שקבלנו

- הפירוק שהאלגוריתם מייצר הנו בעל התכונות הבאות:
  - כל סמכה בפירוק הנה 3NF (צריך להוכיח)
  - הפירוק משמר את הת"פ (ברור)
  - האם הפירוק בעל תכונת הצירוף ללא אובדן?

24

דוגמה

מה הקשר בין G_i לבין $\pi_{R_i}(F)$?

- נתונה הסכמה ABC עם
  $F = \{A \rightarrow C, B \rightarrow C\}$
- F הוא כבר חסר כפילויות, לכן מקבלים:
  - $G_1 = \{A \rightarrow C\}$ ו- $R_1=AC$
  - $G_2 = \{B \rightarrow C\}$ ו- $R_2=BC$
- G_i מציין את קבוצת הת"פ שמהן נוצרה הסכמה R_i

25

נבדוק האם הפירוק הוא ללא אובדן

- $F = \{A \rightarrow C, B \rightarrow C\}$
- הפירוק הוא: AC ו- BC

A	B	C
a_1	$b_{1,3}$	a_3
$b_{2,1}$	a_2	a_3

הפירוק אינו מקיים את תכונת הצירוף ללא אובדן

26

אלגוריתם למציאת פירוק 3NF (המשך)

- אם בין הסכמות שקבלנו מהכיסוי חסר הכפילויות, אין אף סכמה שקבוצת האטריביוטים שלה היא מפתח-על של הסכמה המקורית, אז
  - מוסיפים מפתח של הסכמה המקורית כסכמה נוספת של הפירוק

27

הדוגמה הקודמת (שנית)

- נתונה הסכמה R=ABC עם
  $F = \{A \rightarrow C, B \rightarrow C\}$
- F הוא כבר חסר כפילויות, לכן מקבלים:
  - $G_1 = \{A \rightarrow C\}$ ו- $R_1=AC$
  - $G_2 = \{B \rightarrow C\}$ ו- $R_2=BC$
- אף סכמה בפירוק אינה מפתח-על של ABC
- לכן, מוסיפים לפירוק את הסכמה AB, שהיא מפתח של הסכמה המקורית ABC
  - $R_3=AB$ ו- $G_3 = \emptyset$ (האם גם $\pi_{R_3}(F)$ ריקה?)

28

דוגמה נוספת של מציאת פירוק ב- 3NF

- $R=ABCD$ ו- $F = \{A \rightarrow B, B \rightarrow C, C \rightarrow D\}$
- F כיסוי חסר כפילויות, לכן נייצר את הסכמות:
  - $G_1 = \{A \rightarrow B\}$ ו- $R_1=AB$
  - $G_2 = \{B \rightarrow C\}$ ו- $R_2=BC$
  - $G_3 = \{C \rightarrow D\}$ ו- $R_3=CD$
- $(R_1)^+_F=R$ ולכן R_1 מפתח-על של R ואין צורך להוסיף אף סכמה נוספת
- *הערה:* במקרה זה הפירוק גם מקיים BCNF

29

עוד דוגמה

G_i היא קבוצת הת"פ שמהן נוצרה R_i, אבל G_i לא בהכרח שקולה ל- $\pi_{R_i}(F)$ (קבוצת כל הת"פ המתקיימות ב- R_i)

- נתונה הסכמה R=ABC עם
  $F = \{AB \rightarrow C, C \rightarrow A\}$
- נייצר את הסכמות:
  - סכמה $R_1=ABC$ עם $G_1 = \{AB \rightarrow C\}$
  - סכמה $R_2=AC$ עם $G_2 = \{C \rightarrow A\}$
- למעשה ב- R_1 מתקיימת גם הת"פ $C \rightarrow A$ אבל היא לא ניתנת באופן מפורש ע"י האלגוריתם, ואין צורך לדעת זאת כדי להראות שימור הת"פ

באופן מעשי צריך לוודא שהת"פ $C \rightarrow A$ מתקיימת ביחס עבור הסכמה R_1

30

שיפור נוסף לאלגוריתם למציאת פירוק 3NF

- אפשר לאחד סכמות עם אותו המפתח, קרי סכמות שהתקבלו מת"פ עם אותו צד שמאל
- דוגמה: R=ABC, $F = \{A \rightarrow B, A \rightarrow C\}$
- מקבלים:
  - $R_1=AB$ ו- $G_1 = \{A \rightarrow B\}$
  - $R_2=AC$ ו- $G_2 = \{A \rightarrow C\}$
- אפשר לאחד לסכמה אחת:
  - $R_1=ABC$ ו- $G_1 = \{A \rightarrow B, A \rightarrow C\}$

31

הוכחה שהאלגוריתם אומנם מייצר פירוק בצורה נורמלית שלישית

32

מה צריך להוכיח?

- כזכור, מייצרים סכמות מכיסוי חסר כפילויות G
- תהי $S=XA_1\ldots A_n$ סכמה שהתקבלה מכל הת"פ מהצורה $X \rightarrow A_i$, קרי כל הת"פ שצד שמאל שלהן הוא X
- צריך להוכיח שב- S אין הפרה של 3NF

33

הת"פ המפרה $V \rightarrow A_k$

- **נניח שהסכמה $S=XA_1\ldots A_n$ מפרה 3NF בגלל ת"פ $V \rightarrow B \in G^+$**
  - לכן $B \notin X$, כי X מפתח של הסכמה ואם $B \in X$ אז אין זו הפרה של 3NF
  - לפיכך, B הוא אחד ה- A_i ונניח ש- $B=A_k$
  - נסמן את הת"פ המפרה ע"י $V \rightarrow A_k$

34

המשך ההוכחה: $X \not\subseteq V$

- $V \rightarrow A_k$ היא הת"פ המפרה 3NF
- $X \not\subseteq V$ (אחרת V מכיל מפתח ולכן V הינו מפתח-על של הסכמה S והת"פ $V \rightarrow A_k$ איננה הפרה של 3NF)
- המשך ההוכחה דן בשני מקרים:
  - $V \subseteq X$
  - $V \not\subseteq X$

35

מקרה 1: $V \subseteq X$

- $V \rightarrow A_k \in G^+$ היא הת"פ המפרה 3NF
- כאמור $X \not\subseteq V$ ולכן *V חלקית ממש ל- X*
- $X \rightarrow A_k \in G$ כי לפי ההנחה יצרנו סכמה מת"פ זו
- נובע שהאטריביוטים של $X - V$ הנם מיותרים בצד שמאל של הת"פ $X \rightarrow A_k$, בסתירה לכך ש- G כיסוי חסר כפילויות

36

מקרה 2: $X \not\subseteq V$

- $V \to A_k \in G^+$ היא הת"פ המפרה 3NF
- יש גזירה של $V \to A_k$ מ- G, שחייבת להשתמש בת"פ של תת-קבוצה $G_1 \subseteq G$
- ב- G_1 אין אף ת"פ שצידה השמאלי הוא X (אחרת $X \subseteq V^+$ ולכן V מפתח-על של S והת"פ $V \to A_k$ אינה הפרה של 3NF)
- *מסקנה 1:* $G_1 \vdash V \to A_k$ וגם $X \to A_k \notin G_1$

37

מקרה 2: $X \not\subseteq V$ (המשך)

- תהי $G_2 \subseteq G$ קב' כל הת"פ $X \to A_i$ עבור $A_i \in V$
- מ- G_2 אפשר לגזור את $X \to V$ (בעזרת אקסיומת האיחוד)
- $X \to A_k \notin G_2$ (אחרת $A_k \in V$ ולכן $V \to A_k$ אינה הפרה של 3NF)
- *מסקנה 2:* $G_2 \vdash X \to V$ וגם $X \to A_k \notin G_2$

38

מקרה 2: $X \not\subseteq V$ (סיכום)

- *מסקנה 1:* $G_1 \vdash V \to A_k$ וגם $X \to A_k \notin G_1$
- *מסקנה 2:* $G_2 \vdash X \to V$ וגם $X \to A_k \notin G_2$
- לכן מ- $G_1 \cup G_2$ אפשר לגזור את $X \to A_k$ ולפיכך $X \to A_k$ הנה ת"פ מיותרת ב- G, בסתירה לכך ש- G חסר כפילויות
- בכל אחד משני המקרים, הראנו שהסכמה $S=XA_1\ldots A_n$ אינה מפרה 3NF

39

סיום ההוכחה

- נשאר להוכיח שאם מוסיפים סכמה S שהיא מפתח של הסכמה המקורית R, אז S מקיימת 3NF
- בין האטריביוטים של מפתח של הסכמה המקורית R לא יכולות להתקיים שום ת"פ (אחרת זה לא מפתח)
- לכן בין האטריביוטים של מפתח אין שום הפרה של 3NF

40

תכונות הפירוק שהאלגוריתם מייצר

- הפירוק שהאלגוריתם מייצר הנו בעל התכונות הבאות:
  - כל סמכה בפירוק הנה 3NF (הוכחנו)
  - הפירוק משמר את הת"פ (ברור)
  - הפירוק הנו ללא אובדן (צריך להוכיח – אפשר למצוא הוכחה בספר של אולמן)
- לאלגוריתם זמן ריצה פולינומיאלי

41