תיאוריית תכנון סכמות
למסדי נתונים יחסיים
חלק 3

Design Theory for
Relational Databases
Part 3

1

סכמה בעייתית

2

סכמה המאפשרת שכפול מידע הנה סכמה בעייתית

- דוגמה

Student
Department
Head

S	D	H
Levy	CS	Rubin
Cohen	Math	Bush
Barak	CS	Rubin

- למחלקה יש רק מנהל אחד אבל יש הרבה סטודנטים
- לכן המידע על מנהל המחלקה נשמר עבור כל סטודנט השייך למחלקה

3

מדוע קיים שכפול מידע?

Student
Department
Head

S	D	H
Levy	CS	Rubin
Cohen	Math	Bush
Barak	CS	Rubin

- העובדה שלמחלקה יש מנהל אחד וסטודנטים רבים נובעת מהאילוצים של הסכמה, כלומר
  - הת"פ $D \rightarrow H$ מתקיימת
  - אבל הת"פ $D \rightarrow S$ אינה מתקיימת

4

במילים אחרות...

- בדוגמה הקודמת מקור הבעיה הוא הת"פ $D \rightarrow H$
- זוהי ת"פ לא טריוויאלית שאמורה להתקיים ביחס, אבל צידה השמאלי איננו מפתח-על של הסכמה (כלומר, D אינו קובע את כל שאר האטריביוטים, אלא רק את H)

5

אפיון פורמלי של סכמה בעייתית

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

6

דוגמה של סכמה בעייתית

Student
Department
Head

S	D	H
Levy	CS	Rubin
Cohen	Math	Bush
Barak	CS	Rubin

- לסכמה SDH יש שתי ת"פ: $S \rightarrow D$ ו- $D \rightarrow H$
- $S \rightarrow D, D \rightarrow H \not\models D \rightarrow S$
- לכן D אינו מפתח-על של הסכמה
  - הסכמה בעייתית לפי האפיון הפורמלי

7

דוגמה נוספת של סכמה בעייתית

Student
Course
Teacher

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

- עבור הסכמה SCT קיימת הת"פ $C \rightarrow T$
- C אינו מפתח-על, כי $C \rightarrow T \not\models C \rightarrow S$

8

פתרון אפשרי לבעייה
שכפול המידע

9

פתרון אפשרי לבעייה
שכפול המידע בדוגמה הקודמת

- נפרק את הסכמה SCT לשתי סכמות
  - הסכמה SC ללא שום ת"פ
  - הסכמה CT עם הת"פ $C \rightarrow T$
- בשתי הסכמות החדשות אין ת"פ לא טריוויאלית שצידה השמאלי אינו מפתח-על
  - בעיית שכפול המידע נפתרה
- אבל האם שתי הסכמות החדשות הן תחליף נאות לסכמה המקורית?

10

היחסים עבור הסכמות החדשות הם ההטלות של היחס המקורי

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

S	C
Cohen	DB
Levy	OS
Levy	DB

C	T
OS	Jones
DB	Smith

11

היחס המקורי מתקבל ע"י צירוף של היחסים החדשים

S	C
Cohen	DB
Levy	OS
Levy	DB

$\bowtie$

C	T
OS	Jones
DB	Smith

=

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

12

האם כל פירוק הוא טוב?

S	C
Cohen	DB
Levy	OS
Levy	DB

$\bowtie$

S	T
Cohen	Smith
Levy	Jones
Levy	Smith

=

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

איבדנו מידע, כי עכשיו יש גם רשומות שלא היו ביחס המקורי

13

פירוק ללא אובדן

Lossless-Join Decomposition

14

הגדרה של פירוק

- נתונה סכמה R (קרי, קבוצת אטריביוטים) עם קבוצת ת"פ F
- הסכמות $R_1, R_2,\ldots,R_n$ נקראות *פירוק* של R אם מתקיימים שני הדברים הבאים:
  - לכל i, $R_i \subseteq R$
  - $\bigcup_{i=1}^{n} R_i = R$

15

דוגמה

- נתונה הסכמה R=ABCD
- $R_1=AB$ ו- $R_2=BCD$ הוא פירוק של R
- פירוק נוסף של R הוא
  - $R_1=AB$
  - $R_2=BC$
  - $R_3=CD$
- אבל $R_1=AB$ ו- $R_2=BC$ אינו פירוק של R

16

פירוק בעל תכונת הצירוף ללא אובדן
Lossless-Join Decomposition

- נתונה סכמה R (קרי, קבוצת אטריביוטים) עם קבוצת ת"פ F
- הפירוק $R_1, R_2,\ldots,R_n$ הנו *בעל תכונת הצירוף ללא אובדן* (או בקיצור *פירוק ללא אובדן*) אם לכל יחס r עבור הסכמה R, שמקיים את כל הת"פ של F, מתקיים השיוויון הבא:

$$\pi_{R_1}(r) \bowtie \pi_{R_2}(r) \bowtie \cdots \bowtie \pi_{R_n}(r) = \bowtie_{i=1}^{n} \pi_{R_i}(r) = r$$

זכור שפעולת הצירוף היא אסוציאטיבית

17

דוגמה

- יהי r היחס

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

- ויהי SC, CT פירוק של SCT
- במקרה זה קיים $\pi_{SC}(r) \bowtie \pi_{CT}(r) = r$

18

$\pi_{SC}(r) \bowtie \pi_{CT}(r) = r$

S	C
Cohen	DB
Levy	OS
Levy	DB

$\bowtie$

C	T
OS	Jones
DB	Smith

=

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

דוגמה נוספת

- היחס r כמו קודם

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

- ונתון הפירוק SC ו- ST של SCT
- במקרה זה $\pi_{SC}(r) \bowtie \pi_{ST}(r) \supsetneq r$

$\pi_{SC}(r) \bowtie \pi_{ST}(r) \supsetneq r$

S	C
Cohen	DB
Levy	OS
Levy	DB

$\bowtie$

S	T
Cohen	Smith
Levy	Jones
Levy	Smith

=

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

$\supsetneq$

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

עובדה חשובה

- נתונה סכמה R (קרי, קבוצת אטריביוטים) עם קבוצת ת"פ F
- יהי $R_1, R_2, \ldots, R_n$ פירוק של R
- יהי r יחס כלשהו עבור R
- **תמיד קיים $\bowtie_i \pi_{R_i}(r) \supseteq r$**
- השאלה היא מתי קיים גם שיוויון לכל r שמקיים את F?

עובדה חשובה נוספת

- נתונה סכמה R (קרי, קבוצת אטריביוטים)
- יהי $R_1, R_2, \ldots, R_n$ פירוק של R
- יהי r יחס כלשהו עבור R
- יהי $s = \bowtie_i \pi_{R_i}(r)$
- איזה יחס (קרי, שיוויון או הכלה) מתקיים בין s לבין $\bowtie_i \pi_{R_i}(s)$?

אלגוריתם לבדיקה האם פירוק הוא ללא אובדן

איך נבדוק האם פירוק הוא ללא אובדן?

- נתונה הסכמה SCT עם $F = \{C \rightarrow T\}$
- האם הפירוק SC, CT הוא ללא אובדן?
- ננסה לבנות דוגמה נגדית

S	C	T
a_1	a_2	$b_{1,3}$
$b_{2,1}$	a_2	a_3

25

מדוע זאת דוגמה נגדית?

S	C
a_1	a_2
$b_{2,1}$	a_2

$\bowtie$

C	T
a_2	$b_{1,3}$
a_2	a_3

=

S	C	T
a_1	a_2	$b_{1,3}$
$b_{2,1}$	a_2	a_3
a_1	a_2	a_3
$b_{2,1}$	a_2	$b_{1,3}$

26

התכונה החשובה

- היחס

S	C	T
a_1	a_2	$b_{1,3}$
$b_{2,1}$	a_2	a_3

מקיים רק את השוויונות שהנם הכרחיים, כדי ליצור רשומה שכולה a בצירוף של ההטלות

$$\bowtie_{i=1}^{n} \pi_{R_i}(r)$$

27

כלומר

- אם יש דוגמה נגדית, אז היחס הזה הוא דוגמה נגדית, כי כל דוגמה נגדית אחרת חייבת להכיל לפחות שתי רשומות שיש ביניהן אותם שוויונות, כדי ליצור (בצירוף של ההטלות) רשומה חדשה (קרי, הרשומה שכולה a) שאיננה ביחס המקורי

S	C	T
a_1	a_2	$b_{1,3}$
$b_{2,1}$	a_2	a_3

28

אבל האם זאת באמת דוגמה נגדית?

- לא, כי הת"פ של F לא מתקיימות, קרי $C \rightarrow T$ איננה מתקיימת

S	C	T
a_1	a_2	$b_{1,3}$
$b_{2,1}$	a_2	a_3

29

נשנה את היחס כך שיקיים את F

S	C	T
a_1	a_2	$b_{1,3}$
$b_{2,1}$	a_2	a_3

- נחליף את $b_{1,3}$ ב- a_3 כדי ש- $C \rightarrow T$ תתקיים

S	C	T
a_1	a_2	a_3
$b_{2,1}$	a_2	a_3

30

אבל כעת זאת כבר לא דוגמה נגדית

S	C	T
a_1	a_2	a_3
$b_{2,1}$	a_2	a_3

- כעת היחס הוא כבר לא דוגמה נגדית, כי יש בו שורה שכולה a – וזאת השורה שאמורה הייתה להיות בצירוף של ההטלות, אבל לא ביחס המקורי

31

אכן זאת לא דוגמה נגדית

S	C
a_1	a_2
$b_{2,1}$	a_2

$\bowtie$

C	T
a_2	a_3

=

S	C	T
a_1	a_2	a_3
$b_{2,1}$	a_2	a_3

32

האלגוריתם הכללי

- נתונה סכמה $R = A_1 A_2 \ldots A_k$ וקבוצת ת"פ F
- ונתון פירוק $R_1, R_2, \ldots, R_n$
- רוצים לבדוק האם הפירוק חסר אובדן
- בונים טבלה שיש בה שורה לכל R_j, שכוללת
  - a_i בעמודה של A_i אם $A_i \in R_j$
  - $b_{j,i}$ בעמודה של A_i אם $A_i \notin R_j$

33

האלגוריתם הכללי (המשך)

- אם ב- F יש ת"פ $X \rightarrow A$ ויש שתי שורות ששוות על X ושונות על A, אז השווה את השורות גם בעמודה A
- משווים שתי שורות ע"י כך שבוחרים את אחד משני המשתנים, המופיעים בשורות אלה בעמודה A, ומחליפים את <u>כל</u> המופעים שלו ע"י המשתנה השני
- a תמיד מחליף b ולא להיפך

34

בסיום האלגוריתם

- משווים משתנים כל זמן שיש הפרות של ת"פ
- בסיום, האלגוריתם קובע שהפירוק ללא אובדן אם יש שורה שכולה a; אחרת, הפירוק עם אובדן
- זמן ריצה פולינומיאלי
  - כדי לבדוק אם יש הפרה של ת"פ דרוש זמן פולינומיאלי
  - לכל היותר ניתן להשוות מספר פולינומיאלי של פעמים (כמספר המשתנים)

35

דוגמה

- $R = ABCD$ ו- $F = \{B \rightarrow C, C \rightarrow D\}$
- הפירוק הוא AB, BC, CD
- הטבלה

A	B	C	D
a_1	a_2	$b_{1,3}$	$b_{1,4}$
$b_{2,1}$	a_2	a_3	$b_{2,4}$
$b_{3,1}$	$b_{3,2}$	a_3	a_4

36

דוגמה (המשך)

- R = ABCD ו- F = {B → C, C → D}
- נחליף את $b_{1,3}$ ב- a_3 בגלל B → C

A	B	C	D
a_1	a_2	$b_{1,3}$	$b_{1,4}$
$b_{2,1}$	a_2	a_3	$b_{2,4}$
$b_{3,1}$	$b_{3,2}$	a_3	a_4

⇒

A	B	C	D
a_1	a_2	a_3	$b_{1,4}$
$b_{2,1}$	a_2	a_3	$b_{2,4}$
$b_{3,1}$	$b_{3,2}$	a_3	a_4

דוגמה (המשך)

- R = ABCD ו- F = {B → C, C → D}
- נחליף את $b_{1,4}$ ב- $b_{2,4}$ בגלל C → D ושתי השורות הראשונות

A	B	C	D
a_1	a_2	a_3	$b_{1,4}$
$b_{2,1}$	a_2	a_3	$b_{2,4}$
$b_{3,1}$	$b_{3,2}$	a_3	a_4

⇒

A	B	C	D
a_1	a_2	a_3	$b_{2,4}$
$b_{2,1}$	a_2	a_3	$b_{2,4}$
$b_{3,1}$	$b_{3,2}$	a_3	a_4

דוגמה (המשך)

- R = ABCD ו- F = {B → C, C → D}
- נחליף את <u>כל</u> מופעי $b_{2,4}$ ב- a_4 בגלל C → D ושתי השורות האחרונות

A	B	C	D
a_1	a_2	a_3	$b_{2,4}$
$b_{2,1}$	a_2	a_3	$b_{2,4}$
$b_{3,1}$	$b_{3,2}$	a_3	a_4

⇒

A	B	C	D
a_1	a_2	a_3	a_4
$b_{2,1}$	a_2	a_3	a_4
$b_{3,1}$	$b_{3,2}$	a_3	a_4

דוגמה (המשך)

- R = ABCD ו- F = {B → C, C → D}
- אין יותר הפרות של ת"פ וקבלנו שורה שכולה a ולכן הפירוק חסר אובדן

A	B	C	D
a_1	a_2	a_3	a_4
$b_{2,1}$	a_2	a_3	a_4
$b_{3,1}$	$b_{3,2}$	a_3	a_4

הטלה של קבוצת ת"פ

Projection of a Set of FDs

איזה ת"פ מתקיימות בפירוק?

- נתונה הסכמה SCT עם F = {C → T}
- מפרקים את SCT ל- SC ו- CT
- ל- SC אין ת"פ
- ל- CT יש ת"פ אחת: C → T

דוגמה נוספת

- נתונה הסכמה ABC עם
  $F = \{A \to B, B \to C\}$
- מפרקים את ABC ל- AB ו- AC
- ל- AB יש ת"פ אחת: $A \to B$
- ל- AC יש ת"פ אחת: $A \to C$
  - הת"פ $A \to C$ נמצאת ב- F^+ אבל לא ב- F

43

הגדרה: הטלה של קבוצת ת"פ

- תהי R סכמה עם קבוצת ת"פ F
- תהי $R \subseteq S$ תת-סכמה של R
- *ההטלה* של F על S, המסומנת ע"י $\pi_S(F)$, היא הקבוצה
  $\{X \to Y \mid X \to Y \in F^+ \text{ and } XY \subseteq S\}$
- כלומר, כל הת"פ הנגררות מ- F וכוללות רק אטריביוטים של S

44

חישוב ההטלה של קבוצת ת"פ

- כדי לחשב את ההטלה של F על S, חייבים לחשב את F^+
- מכיוון שהגודל של F^+ יכול להיות אקספוננציאלי בהשוואה ל- F, החישוב של $\pi_S(F)$ דורש זמן אקספוננציאלי
- האם הגודל של $\pi_S(F)$ יכול להיות אקספוננציאלי בגודל של F?

45

הת"פ של תת-סכמה

The FDs of a Sub-Schema

46

הת"פ של תת-סכמה

- תהי R סכמה עם קבוצת ת"פ F
- תהי $R \subseteq S$ תת-סכמה של R
- $\pi_S(F)$ היא קבוצת הת"פ של תת-הסכמה S
  - ההצדקה לכך ניתנת ע"י שתי הלמות הבאות

47

למה ראשונה: ההטלה של יחס r מקיימת את $\pi_S(F)$ אם r מקיים את F

- תהי R סכמה עם קבוצת ת"פ F
- תהי $R \subseteq S$ תת-סכמה של R
- *למה:* יהי r יחס מעל הסכמה R המקיים את כל הת"פ של F. ההטלה $\pi_S(r)$ מקיימת את כל הת"פ של $\pi_S(F)$.

48

למה שנייה: אם $\pi_S(F) \not\models X \to Y$, אז יש יחס r שמקיים את F אבל ההטלה שלו אינה מקיימת את $X \to Y$

- תהי R סכמה עם קבוצת ת"פ F
- תהי $S \subseteq R$ תת-סכמה של R
- *למה:* תהי $X \to Y$ $(XY \subseteq S)$ ת"פ שאיננה נגררת מ- $\pi_S(F)$. קיים יחס r עבור R שמקיים את F, כך שההטלה $\pi_S(r)$ אינה מקיימת את $X \to Y$.

49

הת"פ המתקיימות בצירוף

50

צירוף מקיים את כל הת"פ של תתי-הסכמות

- נתונה סכמה R עם קבוצת ת"פ F
- יהיו $R_1, R_2, \ldots, R_n$ פירוק של R, כך שלכל R_i קבוצת ת"פ F_i (כלומר, $F_i = \pi_{R_i}(F)$)
- *למה:* יהיו $r_1, r_2, \ldots, r_n$ יחסים עבור הסכמות $R_1, R_2, \ldots, R_n$, בהתאמה, כך שכל r_i מקיים את הת"פ של F_i. הצירוף $\bowtie_{i=1}^{n} r_i$ מקיים את כל הת"פ של $\cup_{i=1}^{n} F_i$
- הערה: $r_1, r_2, \ldots, r_n$ הם לאו דווקא הטלות של יחס אחד מעל הסכמה R

51

אם $\cup_i F_i \not\models X \to Y$ אז יש $r_1, r_2, \ldots, r_n$ כך שכל r_i מקיים את F_i אבל הצירוף $\bowtie_i r_i$ אינו מקיים את F

- נתונה סכמה R עם קבוצת ת"פ F
- יהיו $R_1, R_2, \ldots, R_n$ פירוק של R, כך שלכל R_i קבוצת ת"פ F_i (כלומר, $F_i = \pi_{R_i}(F)$)
- *למה:* אם יש ת"פ $X \to Y$ שאינה נגררת מ- $\cup_{i=1}^{n} F_i$ אז יש יחסים $r_1, r_2, \ldots, r_n$ כך שכל r_i מקיים את F_i אבל הצירוף $\bowtie_{i=1}^{n} r_i$ אינו מקיים את $X \to Y$

52

פירוק שמשמר ת"פ

Decomposition that Preserves FDs

53

משמעות המושג "שימור התלויות הפונקציונליות"

- מתחילים עם סכמה R שיש לה קבוצת ת"פ F
  - יחס r עבור R מקיים את F
- עוברים לפירוק $R_1, R_2, \ldots, R_n$
- יחס r_i עבור R_i מקיים את $\pi_{R_i}(F)$
- באיזה תנאי מובטח שהצירוף $\bowtie_i r_i$ מקיים את F?

54

הגדרה: פירוק שמשמר את הת"פ

- נתונה סכמה R עם קבוצת ת"פ F
- יהיו $R_1, R_2,\ldots,R_n$ פירוק של R
- הפירוק $R_1, R_2,\ldots,R_n$ *משמר* את הת"פ אם

$$\left(\bigcup_{i=1}^{n} \pi_{R_i}(F)\right)^+ = F^+$$

55

דוגמה

- נתונה הסכמה ABC עם $F = \{A \rightarrow B, B \rightarrow C\}$
- מפרקים את ABC ל- AB ו- AC
- ל- AB יש ת"פ אחת: $A \rightarrow B$
- ל- AC יש ת"פ אחת: $A \rightarrow C$
- הפירוק אינו משמר את הת"פ, כי $B \rightarrow C \notin \{A \rightarrow B, A \rightarrow C\}^+$

56

איך נחשב אם פירוק משמר את הת"פ?

- כדי לבדוק את התנאי $\left(\bigcup_{i=1}^{n} \pi_{R_i}(F)\right)^+ = F^+$

אפשר לחשב את $\pi_{R_i}(F)$ לכל i, אבל הדבר דורש זמן אקספוננציאלי

57

אפשר לבדוק בזמן פולינומיאלי

- כדי לוודא קיום התנאי $\left(\bigcup_{i=1}^{n} \pi_{R_i}(F)\right)^+ = F^+$
- מספיק לבדוק שלכל $X \rightarrow Y \in F$ מתקיים $X \rightarrow Y \in \left(\bigcup_{i=1}^{n} \pi_{R_i}(F)\right)^+$
- אפשר לבדוק זאת בזמן פולינומיאלי, מבלי לחשב אף $\pi_{R_i}(F)$

58