תיאוריית תכנון סכמות
למסדי נתונים יחסיים
חלק 5

Design Theory for
Relational Databases
Part 5

1

מגבלות התיאוריה לתכנון סכמות של מסדי נתונים יחסיים

2

תחילה, היתרונות של פירוק

3

תזכורת: סכמה בעייתית היא סכמה שגורמת לשכפול מידע

Student
Course
Teacher

- היחס

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

- הסכמה היא SCT ומתקיימת התי"פ C → T
- מפרקים לשתי סכמות: SC ו- CT

4

הפירוק מונע שכפול מידע וגם מאפשר לשמור מידע חלקי

- כיצד נשמר מידע חלקי?
  - בסכמה SC אפשר לשמור מידע על סטודנט שלוקח קורס גם אם המורה של הקורס טרם נקבע
  - בסכמה CT אפשר לשמור מידע על מורה של קורס לפני שהחלה הרשמת התלמידים

5

דוגמה למידע חלקי שנשמר במסד

- ל- PLAB אין עדיין סטודנטים

C	T
PLAB	Jones
DB	Smith

- ל- OS אין עדיין מורה

S	C
Cohen	DB
Levy	OS
Levy	DB

6

מה היה קורה אילו במסד הייתה הסכמה המקורית (במקום הפירוק)?

- ביחס עבור הסכמה המקורית SCT צריך להשתמש בערכי NULL כדי לשמור מידע חלקי

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

7

הבעיות של שימוש בערכי NULL בסכמה המקורית SCT

- ניסוח שאילתות עשוי להשתנות כאשר יש ערכי NULL
  - מקשה על המשתמש
- אי אפשר לשמור ערכי NULL במפתח
  - S הוא חלק מהפתח של הסכמה SCT ולפיכך אי אפשר לשמור מידע חלקי על קורס ומורה לפני שהחלה הרשמת התלמידים

8

לסיכום

- לפירוק SC ו- CT יש שני יתרונות בהשוואה לסכמה המקורית SCT
  - הפירוק מונע כפילויות
  - הפירוק מאפשר שמירת מידע חלקי ללא ערכי NULL
- אבל שמירת מידע חלקי גורמת לכך שהיחסים במסד אינם הטלות של יחס אחד מעל הסכמה המקורית

9

האם התיאוריה מתאימה למצב שבו היחסים במסד אינם הטלות של יחס אחד מעל הסכמה המקורית?

10

דוגמה למצב שבו היחס r עבור SCT חייב להכיל ערכי NULL, כדי שהיחסים עבור SC ו-CT יהיו ההטלות של r

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

S	C
Cohen	DB
Levy	OS
Levy	DB

C	T
PLAB	Jones
DB	Smith

11

במילים אחרות: הצירוף הטבעי של כל היחסים במסד אינו מכיל את כל המידע שיש ביחסים אלה

S	C
Cohen	DB
Levy	OS
Levy	DB

⋈

C	T
PLAB	Jones
DB	Smith

=

S	C	T
Cohen	DB	Smith
Levy	DB	Smith

12

תזכורת

- תכונת הצירוף ללא אובדן נועדה להבטיח שהצירוף של כל היחסים במסד אינו כולל רשומות מיותרות (שלא היו ביחס המקורי)
- אבל כעת נוצר מצב שבו יש מעט מדי רשומות בצירוף של הסכמות SC ו- CT

13

הפתרון

- למעשה, כדי לקבל ביחס אחד (עבור הסכמה המקורית) את כל המידע הקיים במסד, צריך לקחת איחוד של כל הצירופים הטבעיים ללא אובדן
- במקרה של **הסכמות SC ו- CT, יש שלושה צירופים טבעיים שהם ללא אובדן**
  - $SC \bowtie CT$
  - SC
  - CT

כ"א משלושת הצירופים הוא מעל קבוצת אטריביוטים אחרת – לכן אי אפשר לבצע איחוד

14

לפיכך, מוסיפים תחילה עמודות שיש בהן NULL עבור האטריביוטים החסרים

- עבור הצירוף SC מקבלים:

S	C	T
Cohen	DB	NULL
Levy	OS	NULL
Levy	DB	NULL

- עבור הצירוף CT מקבלים:

S	C	T
NULL	PLAB	Jones
NULL	DB	Smith

15

היחס עבור הצירוף $SC \bowtie CT$ נשאר ללא שינוי, כי הוא כולל את כל העמודות

S	C	T
Cohen	DB	Smith
Levy	DB	Smith

16

עכשיו נבצע איחוד

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

אבל יש רשומות שכל המידע הכלול בהן מופיע גם ברשומות אחרות

17

נמחק רשומות שהן למעשה חלק מרשומות אחרות

- התוצאה הסופית של איחוד כל הצירופים הטבעיים ללא אובדן היא:

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

18

איחוד מורחב

- פעולת האיחוד המורחב מורכבת מהשלבים הבאים:
  - הוספת עמודות עם ערכי NULL עבור האטריביוטים החסרים
  - ביצוע פעולת איחוד רגילה
  - מחיקת רשומות שנכללות ברשומות אחרות

19

לסיכום

- היחס עבור הסכמה המקורית הוא תוצאת האיחוד המורחב של כל הצירופים הטבעיים ללא אובדן

20

צירוף מינימלי ללא אובדן הכולל קבוצת אטריביוטים נתונה

21

דוגמה

Employee, Department, Project, Manager

- הסכמה היא R=EDPM וקבוצת הת"פ היא
  $F = \{E \rightarrow D,\ E \rightarrow P,\ D \rightarrow M,\ P \rightarrow M\}$
- מקבלים את הפירוק
  - R_1=ED ו- $F_1 = \{E \rightarrow D\}$
  - R_2=EP ו- $F_2 = \{E \rightarrow P\}$
  - R_3=DM ו- $F_3 = \{D \rightarrow M\}$
  - R_4=PM ו- $F_4 = \{P \rightarrow M\}$
- זהו פירוק ללא אובדן, שמשמר את הת"פ ומקיים BCNF

22

דוגמה למסד נתונים עבור ארבעת הסכמות של הפירוק

E	D
Levy	CS

D	M
CS	Dolev

E	P
Levy	OS

P	M
OS	Barak

- נניח שרוצים למצוא לכל עובד את המנהלים שלו
- הצירוף הטבעי של כל היחסים הנו ריק
- בעיקרון, צריך לחשב הטלה על EM של האיחוד המורחב של כל הצירופים הטבעיים ללא אובדן

23

מהם כל הצירופים ללא אובדן בדוגמה זו?

- $ED \bowtie DM \bowtie EP \bowtie PM$
- $ED \bowtie DM \bowtie EP$
- $ED \bowtie EP \bowtie PM$
- $ED \bowtie DM$
- $EP \bowtie PM$
- ויש עוד מספר צירופים ללא אובדן שאינם מכילים את קבוצת האטריביוטים EM

24

צירוף מינימלי ללא אובדן הכולל קבוצת אטריביוטים נתונה

- למעשה מספיק לחשב רק את שני הצירופים הבאים:
  - $ED \bowtie DM$ ▶
  - $EP \bowtie PM$ ▶
- כ"א משני הצירופים האלה הנו *מינימלי* במובן שאי אפשר להשמיט אף אחד מהיחסים שמשתתפים בצירוף ועדיין לקבל צירוף ללא אובדן שמכיל את את האטריביוטים המבוקשים (קרי, EM)

25

איך נקבל את כל המידע על EM?

- ניקח את כל הצירופים המינימליים ללא אובדן שמכילים את האטריביוטים EM
- נטיל כל אחד על EM
- וניקח את האיחוד, כלומר

$$\pi_{EM}(EP \bowtie PM) \cup \pi_{EM}(ED \bowtie DM)$$

- כך נענה על שאילתה לגבי E ו- M

26

שימור התלויות הפונקציונליות

27

נחזור לדוגמה הקודמת

Employee, Department, Project, Manager

- הסכמה היא R=EDPM וקבוצת הת"פ היא

$$F = \{E \to D,\ E \to P,\ D \to M,\ P \to M\}$$

- מקבלים את הפירוק
  - R_1=ED ▶ ו- $F_1 = \{E \to D\}$
  - R_2=EP ▶ ו- $F_2 = \{E \to P\}$
  - R_3=DM ▶ ו- $F_3 = \{D \to M\}$
  - R_4=PM ▶ ו- $F_4 = \{P \to M\}$
- זהו פירוק ללא אובדן, שמשמר את הת"פ ומקיים BCNF
- הת"פ $E \to M$ נגררת מ- F

28

האם הת"פ $E \to M$ מתקיימת במסד המורכב מהיחסים הבאים?

E	D
Levy	CS

D	M
CS	Dolev

E	P
Levy	OS

P	M
OS	Barak

- כל יחס מקיים את הת"פ שלו
- הפירוק משמר את הת"פ ומשמעות הדבר שהצירוף הטבעי של כל היחסים מקיים את כל הת"פ
- הצירוף הטבעי של כל היחסים ריק – חלקי המידע הסותרים את הת"פ נעלמים כשלוקחים את הצירוף

29

צריך לוודא שהת"פ $E \to M$ מתקיימת כפי שעונים על שאילתות מעל EM

- ניקח את כל הצירופים המינימליים ללא אובדן שכוללים את האטריביוטים EM
- נטיל כל אחד על EM
- וניקח את האיחוד, כלומר

$$\pi_{EM}(EP \bowtie PM) \cup \pi_{EM}(ED \bowtie DM)$$

- בתוצאה הנ"ל בודקים אם אין הפרה של הת"פ
- כזכור, כך גם עונים על שאילתות לגבי E ו- M

30

אבל יתכן שהת"פ בכלל אינן מבטאות את המציאות בצורה נכונה...

31

מגבלות השימוש בת"פ

- לכל הגזירות של אותה ת"פ חייבת להיות אותה משמעות
- אי אפשר לקבוע שהת"פ E# → E# מתקיימת במובן שמספר עובד קובע את מספר העובד של המנהל שלו, כי כבר קיימת הת"פ הטריוויאלית E# → E# (קרי, מספר עובד קובע את עצמו)

32

נחזור שוב לדוגמה הקודמת

Employee, Department, Project, Manager

- R=EDPM ו-
  F = {E → D, E → P, D → M, P → M}
- אפשר לגזור את E → M מהת"פ
  - E → D ו- D → M
  - המשמעות היא "מנהל המחלקה של העובד"
- אפשר לגזור את E → M מהת"פ
  - E → P ו- P → M
  - המשמעות היא "מנהל הפרויקט של העובד"

33

מסקנה

- R=EDPM ו-
  F = {E → D, E → P, D → M, P → M}
- לכן, אם אלה אומנם הת"פ הנכונות, אז חייב להתקיים שאותו אדם הוא גם מנהל המחלקה וגם מנהל הפרויקט של העובד
- אם אין הדבר כך, אז חייבים לבצע שינוי שם של האטריביוט M

34

שינוי שם

- נחליף את M בשני אטריביוטים חדשים
  - M_D מנהל מחלקה
  - M_P מנהל פרויקט
- $R=EDPM_DM_P$ ו-
  $F = \{E \to D, E \to P, D \to M_D, P \to M_P\}$

35

אבל מה עם ה- ISA?

- עדיין יש צורך באטריביוט M כדי לייצג את קבוצת כל המנהלים
  - M_D isa M
  - M_P isa M
- $R=EDPM_DM_PM$ ו-
  $F = \{E \to D, E \to P, D \to M_D, P \to M_P, M_D \to M, M_P \to M\}$

שוב יש שתי גזירות שונות של E → M והבעיה לא נפתרה

36

פתרון אפשרי

- נתעלם מהת"פ הנובעות מקשרי ISA, כלומר מהת"פ $M_D \to M$ ו- $M_P \to M$
- ניתן להתעלם מהת"פ האלה, כי אין צורך ביחס מיוחד בין M ל- M_D או בין M ל- M_P
- אין צורך ביחס מיוחד, כי ל- M, M_D ו- M_P ערכים מאותו התחום (למשל, מספר עובד)

37

הדוגמה ביתר פירוט

- $R=EDPM_DM_PMS$
  - S מציין מזכיר של המנהל
- הת"פ הנן
$F = \{E \to D, E \to P, D \to M_D, P \to M_P, M \to S\}$
- F איננה כוללת את הת"פ הנובעות מקשרי ה- ISA
  - M_D isa M
  - M_P isa M
  - M isa E
  - S isa E

38

נפעיל את האלגוריתם למציאת פירוק ב- 3NF

- מקבלים את הפירוק
  - $R_1=ED$ ו- $F_1 = \{E \to D\}$
  - $R_2=EP$ ו- $F_2 = \{E \to P\}$
  - $R_3=DM_D$ ו- $F_3 = \{D \to M_D\}$
  - $R_4=PM_P$ ו- $F_4 = \{P \to M_P\}$
  - $R_5=MS$ ו- $F_5 = \{M \to S\}$
- אין צורך ביחסים עבור קשרי ה- ISA, כי הערכים בעמודות S, M, M_P, M_D ו- E באים כולם מהתחום של מספרי עובדים – לכן קשרי ה- ISA הם קשרי הזהות

39

מיזוג סכמות

40

נחזור לדוגמה קודמת של יצירת פירוק ב- 3NF

G_i היא קבוצת הת"פ שמהן נוצרה הסכמה R_i, אבל אינה בהכרח כיסוי של כל הת"פ המתקיימות ב- R_i

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

41

עדיף לשלב את שתי הסכמות לסכמה אחת

- נניח ששומרים את שתי סכמות
  - סכמה R_1=ABC עם $G_1 = \{AB \to C\}$
  - סכמה R_2=AC עם $G_2 = \{C \to A\}$
- עדיין צריך לוודא שהתלות $C \to A$ מתקיימת ב- R_1
- לפיכך, עדיף לחזור לסכמה המקורית
- R=ABC ו- $F = \{AB \to C, C \to A\}$
- במקרה הכללי, אם סכמה אחת מכילה את כל האטריביוטים של סכמה שניה, אז עדיף למזג את שתי הסכמות לסכמה אחת, שיש לה את הת"פ של שתי הסכמות

42

אבל צריך לזכור שמאבדים את היכולת לשמור מידע חלקי

- אם ממזגים את שתי הסכמות, אז באופן מעשי מאבדים את היכולת לשמור מידע חלקי עבור AC, כי B הוא אטריביוט ששייך למפתח של הסכמה ABC (ובמערכות מסדי נתונים קיימות, כגון אורקל, אי אפשר לשמור ערכי NULL באטריביוטים של מפתח)

43

תלויות רב-ערכיות ותלויות צירוף

Multivalued Dependencies
and Join Dependencies

44

דוגמה לסכמה שניתנת לפירוק שמקיים את תכונת הצירוף ללא אובדן

- היחס

Student
Course
Teacher

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

- הסכמה היא SCT ומתקיימת הת"פ $C \rightarrow T$
- ניתן לפרק ל- SC ול- CT

45

האם ייתכן שהסכמה ניתנת לפירוק גם אם אין אף ת"פ?

- בין קורסים לסטודנטים קיים יחס של רבים לרבים
- בין קורסים למורים תיתכנה מספר אפשרויות
  - לקורס יש תמיד יש מורה יחיד, כלומר מתקיימת הת"פ $C \rightarrow T$ ויש פירוק ללא אובדן
  - לקורס יש מספר מורים – נדון במקרה הזה בשקפים הבאים

46

אפשרות א': כל תלמיד לומד אצל כל המורים

- מספר מורים (אחד או יותר) מתחלקים ביניהם בהעברת הקורס
- אם זה תמיד המצב, אז מהו בעצם ייצוג נכון של הנתונים ביחס עבור הסכמה SCT?
- קל יותר לראות מה הייצוג הנכון ב- SC ו- CT

S	C
Cohen	DB
Levy	DB

C	T
DB	Jones
DB	Smith

47

כל תלמיד בקורס לומד אצל כל מורה של הקורס, לכן היחס מעל SCT שמייצג נכון את הנתונים הוא בדיוק הצירוף של SC ו- CT

S	C
Cohen	DB
Levy	DB

$\bowtie$

C	T
DB	Jones
DB	Smith

$=$

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

48

במקרה שתיארנו מתקיימת התלות הרב-ערכית $C \twoheadrightarrow T$

- התלות הרב-ערכית $C \twoheadrightarrow T$ פירושה שלכל קורס יש קבוצת מורים שאינה תלויה באטריביוטים האחרים, קרי בסטודנטים
- באופן שקול, התלות הרב-ערכית מתקיימת בכל יחס שמקיים

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

- לכן, אם היחסים האפשריים הם רק אלה שמקיימים את התלות הרב ערכית $C \twoheadrightarrow T$, אז הפירוק של SCT ל- SC ול- CT הנו ללא אובדן

אפשרות ב': אין חוקיות המקשרת בין סטודנט לבין קבוצת המורים שמלמדת אותו

- האם יתכן שהפירוק של SCT ל- SC, CT ו- ST הנו ללא אובדן?
- נתונות העובדות הבאות
  - משה מלמד את לוי (אפשרי שיש ללוי מורים נוספים)
  - לוי לומד ב- OS (וכמובן שהוא לומד גם בקורסים אחרים)
  - משה הוא אחד המורים של OS
- משלושת העובדות האלה לא בהכרח נובע שמשה הוא מורה של לוי ב- OS
- לכן, אין אפשרות לפרק את SCT

אפשרות ג': יש חוקיות המקשרת בין מורה לסטודנט, אבל היא מצריכה הוסמת אטריביוט

- הקורס מחולק לקבוצות, כאשר כל סטודנט שייך לקבוצה אחת ולכל קבוצה מורה יחיד
- צריך להוסיף אטריביוט G שמציין קבוצה
- לכן R=SGCT ו- $F = \{G \rightarrow C,\ G \rightarrow T\}$

סביר יותר ש- G היא ישות חלשה
(המפתח שלה מורכב ממספר קבוצה וממספר קורס)

- במקרה זה R=SGCT והתלויות הן
  $F = \{SC \rightarrow G,\ CG \rightarrow T\}$
- פירוק ל- 3NF נותן
- R_1=SCG ו- $F_1 = \{SC \rightarrow G\}$
- R_2=CGT ו- $F_2 = \{CG \rightarrow T\}$

תלויות צירוף

- אפשר להכליל את המושג של תלות רב-ערכית באופן הבא
- אם אפשר לפרק ללא אובדן את R למספר סכמות $R_1, R_2, \ldots, R_n$ ($n \geq 1$), אז אומרים שמתקיימת תלות הצירוף $\bowtie [R_1, R_2, \ldots, R_n]$
- מתי הדבר שימושי?

באלגוריתם לפירוק ל- 3NF

- כזכור, בשלב האחרון של האלגוריתם, אם הפירוק שנוצר מהכיסוי אינו חסר אובדן, אז מוסיפים סכמה R שהיא מפתח של הסמכה המקורית
- בסכמה R אין ת"פ, אבל ייתכן שניתן לפרק אותה למספר סכמות קטנות יותר, בגלל תלות רב-ערכית או תלות צירוף

לדוגמה

- נניח ש- S, C, ו- T הם חלק מהאטריביוטים, שכל אחד מייצג ישות אחרת ואין ביניהם ת"פ
- יש אטריביוטים נוספים שמתארים את התכונות של ישויות אלה, לדוגמה S → Name
- לכן, בפירוק ל- 3NF תהיינה שלוש סכמות, עבור שלושת הישויות, ובנוסף הסכמה SCT, שהיא מפתח
- ניתן לפרק את SCT לפי תלות רב-ערכית ל- SC ו- CT

55