

עבודה עם סכימות

כששולחים פקודת DDL (למשל CREATE TABLE), הDDL compiler מקבלת את הפקודה בתור קלט, והפלט שלה הוא DB description tables. הקלט הוא ברמה התפיסתית - מה צריך להיות בטבלה, איך היא צריכה להתנהג. הפלט הוא ברמה הפיזית - באיזה קובץ הטבלה תשב, מה יהיה ייצוג הנתונים וכו'.

עבודה עם נתונים - שאילתות

שאילתה יכולה להגיע ישר מהמסוף - user query - או מתוך תוכנית - application program.

השאילתה היא הקלט של הDML (QL) processor (זה יכול להיות interpreter, compiler, והפלט שלו DML processor מסתכל על הקלט שלו לפי DB description tables, והפלט שלו הוא גם שאילתה ברמה תפיסתית - אבל שאילתה יותר יעילה, והשאילתה הזו מגיעה לDB manager. DB manager מתרגם את השאילתה ברמה התפיסתית לשאילתה ברמה פיזית. השאילתה ברמה הפיזית מגיעה לFile manager. File manager מחליט לפי השאילתה הזו מה לעשות - ואת זה הוא מעביר לPhysical DB, שמבצע את הפעולה הפיזית. כל השלבים האלו צריכים לגשת לטבלות.

- הDML processor צריך לגשת לDB description tables - כדי שיוכל ליעל את השאילתה.
- הDML processor והDB manager צריכים לגשת לטבלת ההרשאות - Authorization tables - שם רשום למי מותר לעשות מה, כי לא כל המשתמשים יכולים לבצע את כל הפעולות על כל המידע.
- הDB manager והFile manager צריכים לגשת לconcurrent access tables. יכולים להיות מקרים ששני משתמשים רוצים לגשת לנתונים שונים בחלקים שונים של אותו קובץ (המשתמשים לא צריכים לדעת שהם ניגשים לאותו קובץ - זה נעשה מאחורי הקלעים). הconcurrent access מטפל במקרים כאלה.

Relational model

היתרון העיקרי של המודל הזה שהוא מבוסס על תורת הקבוצות Set Theory, ולכן אפשר לעשות בו הרבה דברים גם בצורה תיאורתית. במודל הרלציוני נתונים מאורגנים בטבלאות. יש קשרים בין הנתונים באותה טבלה - בכל שורה בטבלה, הנתונים קשורים אחד לשני.

דוגמה

נניח שיש לנו טבלה:

Stud		
Name	ID	Year
Cohen	123	2

איך נקראים המרכיבים בעולם מושגים של טבלאות ובעולם מושגים של קשרים?

Relational	Table	Example
Relation name	Table name	Stud
Attribute	Field name	Name, ID, Year
Tuple	Row	(Cohen, 123, 2)
Arity	Number of fields	3
Schema		$S = (\text{Name, ID, Year})$

Domain - יאָץ

Attribute (תכונה) לוקח ערכים מהתחום שלו. למשל:

$$D_{\text{Name}} = \{\dots\}$$

- A - set of attributes. $A = \{A_1, A_2, \dots, A_n\}$
- S - set of domains. $D = \{D_1, D_2, \dots, D_n\}$

ניתן לקחת מכפלה קרטזית של דומיינים - $\bar{D} = D_1 \times D_2 \times \dots \times D_n$. יחס (=טבלה) זה תת קבוצה במכפלה קרטזית כזו - $D \subseteq \bar{D}$.

דוגמה

$$D_{\text{Name}} = \{\text{cohen, lavan}\}$$

$$D_{\text{ID}} = \{111, 112, 121, 122, 211, 212, 222\}$$

$$D_{\text{Year}} = \{1, 2, 3, 4\}$$

$$\bar{D} = \begin{array}{l} \text{cohen} \quad 111 \quad 1 \\ \text{cohen} \quad 111 \quad 2 \\ \text{cohen} \quad 111 \quad 3 \\ \text{cohen} \quad 112 \quad 1 \\ \vdots \\ \text{lavan} \quad 222 \quad 4 \end{array}$$

ניקח תת קבוצה במכפלה הקרטזית:

Stud		
Name	ID	Year
cohen	122	2
lavan	112	3
cohen	112	2

אמנם פורמלית זה חלק מהמכפלה הקרטזית, אבל בחיים יש אילוצים - למשל לא יכולים להיות שני אנשים עם אותו מספר זהות. כלומר לא כל תת קבוצה במכפלה הקרטזית היא לגיטימית, בגלל תלויות בין נתונים.

מפתח Key

$$R(S)$$

הסכימה היא תת קבוצה של קבוצת התכונות האפשרויות:

$$S \subseteq A$$

ניקח תת קבוצה של הסכימה:

$$K \subseteq S$$

אומרים ש K מפתח של R (key of R) אם:

$$1. \text{ מפתח הוא ייחודי (Unique) - } \forall t_1, t_2 \in R (t_1(R) \neq t_2(R) \Rightarrow t_1 \neq t_2)$$

$$2. \text{ } \#_{K' \subseteq K} K' - \text{unique} - \text{כלומר לא קיימת תת קבוצה ממש של } K \text{ שגם היא ייחודית.}$$

כלומר המפתח צריך להיות קבוצה מינימלית של תכונות שיכולה להגדיר כל שורה באופן יחודי.

דוגמאות

$$A = \text{airpl, num, flight, class, price, ...}$$

טבלה של מספר המושבים בסוגי מטוסים שונים:

R1	
airpl	num
boeing	300
Airbus	400

$$S1 = (\text{airpl, num}) \subset A$$

$$R1 \subseteq D_{\text{airpl}} \times D_{\text{num}}$$

מה יהיה המפתח? בטבלה הרשומה כאן, גם airpl וגם num ייחודיים. אבל צריך לבחור מפתח לא לפי טבלה ספציפית, אלא לפי הבנה של מה צריך להיות - ולכן המפתח לא יכול להיות num, כי לוגית יכולים להיות שני סוגי מטוסים עם אותו מספר מושבים.

מצד שני, aripl יכול להיות מפתח - כי לכל סוג מטוס יש שם משלו, ולכן לא יכולות להיות שני סוגי מטוסים עם אותו aripl, ולכן מובטח לנו ש-airpl יהיה יחודי. לכן airpl הוא הבחירה היחידה למפתח:

$$K = \text{airpl}$$

נוסיף עוד טבלה - של מחירי טיסות:

R2

flight	class	price
123	busin	500
123	econ	400
124	busin	400

כאן, price לא יהיה המפתח(כי יכולות להיות טיסות שונות עם אותו מחיר). נשארנו עם flight וclass. נשים לב שאף אחד מהם לא יחודי, אבל שניהם ביחד ייחודיים, ולכן נבחר

$$K = \text{flight, class}$$

מצד שני, במקרה הזה סביר להניח שמחלקת עסקים תמיד תהיה יותר יקרה ממחלקת תיירים באותה טיסה, ולכן ניתן גם לבחור

$$K = \text{flight, price}$$

כלומר, יש לנו שני מועמדים:

$$\text{Possible Keys: } \begin{cases} K_1 = \text{flight, class} \\ K_2 = \text{flight, price} \end{cases}$$

את Possible Keys קיבלנו בצורה פורמלית. מתוכם צריך לבחור מפתח ראשי - K=Primary Key. מכיוון שאי אפשר להמשיך את הסינון בצורה פורמלית, הבחירה ביניהם צריכה להתבצע על ידי הגיון אנושי.

הערה אם יש שני מועמדים למפתחות - $K_1 = A, B$ - אז שניהם מינימלים, למרות ש K_1 יותר ארוך מ K_2 , כי כל אחד מהם אם מורידים ממנו איבר הוא מפסיק להיות יחודי. אנחנו לא יכולים לבחור אוטומטית את K_1 רק בגלל שהוא יותר קצר, כי יכול להיות שדווקא K_2 יותר מתאים.

נוסיף עוד טבלה:

R3

flight	class
123	busin
123	econ
124	busin

במקרה הזה, המפתח שווה לסכימה - $K = S$ - כי אי אפשר לוותר על אף שדה. נוסיף עוד טבלה:

R4

flight	class	price	airpl	num
123	busin			
123	econ			
124	busin			

המפתח כאן הוא $K = \text{flight, class}$ - הוא ייחודי ומינימלי.
ניקח קבוצה שמכילה את K :

$$K'' = \text{flight, class, price}$$

$$K'' \supset K$$

K'' לא מינימלי, אבל הוא ייחודי (כי K ייחודי). קבוצה שהיא ייחודית אבל לא מינימלית נקראת "מפתח על" - "superkey".

בעיות

נוסיף עוד שורות לR4:

flight	class	price	airpl	num
123	busin	500	boeing	300
123	econ	400	boeing	300
124	busin	400	airbus	400
125	busin	300	boeing	300
125	econ	250	boeing	300

(1) Redundancy כפילות

יש ערכים שחוזרים על עצמם. יש ערכים שחייבים לחזור על עצמם - למשל חייבים לציין את המחלקה ואת מספר הטיסה וכו'. אבל מספר המושבים הוא תמיד 300 בכל טיסה של בוינג. זה מבזבז מקום, וזה מוביל לבעיה נוספת:

(2) Update anomaly

אם משנים את מספר המושבים במטוס בוינג, אז צריך לעדכן בכל המופעים שלו בטבלה.

(3) Insertion anomaly

נגיד שמוסיפים טיסה חדשה למערכת - טיסה 126, מחלקת עסקים, עולה 600. קיבלנו $(126, \text{busin}, 600)$ - זה מידע חלקי (tuple קצר מדי) - ולכן אנחנו לא יכולים להכניס אותו לטבלה.

את זה אפשר לפתור על ידי הוספת ערכים ריקים - $(126, \text{busin}, 600, \text{NULL}, \text{NULL})$. אבל, אם רוצים להוסיף סוג חדש של מטוס - חייבים שורה שלמה, גם אם אין לו טיסות, כי שם המטוס הוא לא חלק מהמפתח.

Deletion anomaly (4)

- נניח ששדה התעופה מוכר את מטוס airbus שלו. מוחקים את השורה השלישית - אבל זה מוחק גם את טיסה 124.
- נניח שמבטלים את טיסה 124 - מוחקים את השורה השלישית, אבל זה מוחק גם את המידע שלנו על מטוס aribus.

בשני המקרים האלה מחקנו מידע שלא היינו רוצים למחוק.