Лекція 6
Реляційна алгебра

Реляційна модель даних

Реляційна модель даних (РМД) - логічна модель даних, прикладна теорія побудови баз даних, яка є додатком до завдань обробки даних таких розділів математики як теорії множин і логіка першого порядку.

На реляційної моделі даних будуються реляційні бази даних.

Реляційна модель даних включає такі компоненти:

- Структурний аспект (складова) - дані в базі даних є набором відносин.
- Аспект (складова) цілісності - відносини (таблиці) відповідають певним умовам цілісності. РМД підтримує декларативні обмеження цілісності рівня домену (типу даних), рівня відносини і рівня бази даних.
- Аспект (складова) обробки (маніпулювання) - РМД підтримує оператори маніпулювання відносинами (реляційна алгебра, реляційне числення).

Крім того, до складу реляційної моделі даних включають теорію нормалізації.

Термін "реляційний" означає, що теорія заснована на математичному понятті ставлення *(relation)*. Як неформального синоніма терміну "відношення" часто зустрічається слово таблиця. Необхідно пам'ятати, що "таблиця" є поняття нестроге і неформальне і часто означає не "ставлення" як абстрактне поняття, а візуальне уявлення відносини на папері або екрані. Некоректне і нестроге використання терміну "таблиця" замість терміна "ставлення" нерідко призводить до нерозуміння. Найбільш часта помилка полягає в міркуваннях про те, що РМД має справу з "плоскими", або "двовимірними" таблицями, тоді як такими можуть бути тільки візуальні представлення таблиць. Відносини ж є абстракціями, і не можуть бути ні "плоскими", ні "неплоским".

Для кращого розуміння РМД слід відзначити три важливі обставини:

- модель є логічною, тобто відносини є логічними (абстрактними), а не фізичними (збереженими) структурами;
- для реляційних баз даних вірний інформаційний принцип : все інформаційне наповнення бази даних представлено одним і тільки одним способом, а саме - явним завданням значень атрибутів у кортежі відносин; зокрема, немає ніяких покажчиків (адрес), що зв'язують одне значення з іншим;
- наявність реляційної алгебри дозволяє реалізувати декларативне програмування і декларативне опис обмежень цілісності, на додаток до навігаційного (процедурним) програмування і процедурної перевірці умов.

Принципи реляційної моделі були сформульовані в 1969 - 1970 роках Е. Ф. Коддом (EF Codd). Ідеї Кодда були вперше публічно викладені в статті "A Relational Model of Data for Large Shared Data Banks", що стала класичною.

Суворе виклад теорії реляційних баз даних (реляційної моделі даних) в сучасному розумінні можна знайти в книзі К. Дж. Дейта. "CJ Date. An Introduction to Database Systems" ("Дейт, К. Дж. Введення в системи баз даних").

Найбільш відомими альтернативами реляційної моделі є ієрархічна модель, і мережева модель. Деякі системи, що використовують ці старі архітектури, використовуються досі. Крім того, можна згадати про об'єктно-орієнтованої моделі, на якій будуються так звані об'єктно-орієнтовані СУБД, хоча однозначної і загальноприйнятого визначення такої моделі немає.

Операції реляційної алгебри

Реляційна алгебра - замкнута система операцій над відносинами в реляційної моделі даних, або **Реляці́йна алгебра** — відгалуження логіки першого порядку, множина відношень замкнених операторами. Оператори застосовуються до відношень, в результаті застосування отримується нове відношення.

В математиці, алгебра відношень є алгебраїчною структурою щодо математичної логіки та теорії множин.

РМД стала першою працездатною моделлю даних, оскільки мала ефективний інструментарій - операції реляційної алгебри. Основною одиницею обробки є відношення, а не його кортежі.

Реляційна алгебра включає дві групи операцій.

1. Традиційні операції над множинами (модифіковані з урахуванням того, що їх операндами є відношення) - об'єднання, перетин, різниця (віднімання), декартовій твір і розподіл.
2. Спеціальні реляційні операції - вибірка, проекція, з'єднання.

Операції реляційної алгебри також називають *реляційними операціями.* Початковий набір з 8 операцій був запропонований Е. Коддом в 1970-і роки і включав як операції, які до цих пір використовуються (проекція, об'єднання і т.д.), так і операції, які не увійшли до вживання (наприклад, поділ відношень).

У процесі розвитку реляційної теорії і практики було запропоновано декілька нових реляційних операцій, наприклад полусоедіненіе *(SEMI-JOIN)* і полуразность, або анти-полусоедіненіе *(ANTI-SEMI-JOIN), CROSS APPLY* і *OUTER APPLY,* транзитивне замикання *(TCLOSE)* та ін

Оскільки багато операцій виразність один через одного, у складі реляційної алгебри можна виділити кілька варіантів базису (набору операцій, через який виразність всі інші). Найбільш відомий і строго певний базис (алгебра A) запропонований Крістофером Дейта і Х'ю Дарвеном.
Доведено, що реляційна алгебра і реляційне обчислення взаємно еквіваленти.

Замкнутість реляційної алгебри

Реляційна алгебра являє собою набір таких операцій над відносинами, що результат кожної з операцій також є відносинам. Це властивість алгебри називається замкнутістю.

Операції над одним відношенням називаються *унарний,* над двома відношеннями - *бінарними,* над трьома - *тернарних* (такі практично невідомі).

Приклад унарний операції - проекція, приклад бінарної операції - об'єднання.

Реляційну операцію *f* можна представити функцією з відносинами в якості аргументів: $R = f(R_1, R_2, \ldots, R_n)$

Оскільки реляційна алгебра є замкнутою, в якості операндів в реляційні операції можна підставляти інші вирази реляційної алгебри (підходящі за типом):

$R = f(f_1(R_{11}, R_{12}, \ldots), f_2(R_{21}, R_{22}, \ldots), \ldots)$

У реляційних виразах можна використовувати вкладені вирази як завгодно складної структури.

Обмеження на операції

Деякі реляційні операції, зокрема, операції *об'єднання, перетину* і *віднімання,* вимагають, щоб відносини мали збігаються (однакові) заголовки (схеми). Це означає, що збігаються кількість атрибутів, назви атрибутів і тип (домен) однойменних атрибутів.

Деякі відносини формально не є сумісними з-за відмінності в назвах атрибутів, але стають такими після застосування операції перейменування атрибутів.

Операції реляційної алгебри

Далі перераховані деякі операції реляційної алгебри, які представляють або історичний, або практичний інтерес. Всі операції перерахувати неможливо, оскільки будь-яка операція, що задовольняє визначенню реляційної, є частиною реляційної алгебри.

1. Перейменування

Перейменування є унарним оператором, що записується як $\rho_{a/b}(R)$. Результат застосування оператора ідентичний Rза винятком того, що поле bв усіх кортежах перейменовується на поле a. Цей оператор застосовується для простого перейменування атрибута відношення, або самого відношення.

В результаті застосування операції перейменування отримуємо нове ставлення, зі зміненими іменами атрибутів.

Синтаксис:

R RENAME $Atr_1, Atr_2, \ldots$ AS $NewAtr_1, NewAtr_2, \ldots$

де *R* - відношення $Atr_1, Atr_2, \ldots$ - вихідні імена атрибутів $NewAtr_1, NewAtr_2, \ldots$ - нові імена атрибутів

2. Об'єднання

Відношення з тим же заголовком, що і у сумісних за типом відношень *A* і *B,* і тілом, що складається з кортежів, які належать або *A,* або *B,* або обом відносинам.

Синтаксис:

A UNION *B*

3. Перетин

Відношення з тим же заголовком, що й у відносин *A* і *B,* і тілом, що складається з кортежів, які належать одночасно обом відносин *A* і *B.* Синтаксис:
A INTERSECT *B*

4. Віднімання

Відношення з тим же заголовком, що і у сумісних за типом відносин *A* і *B,* і тілом, що складається з кортежів, що належать відношенню *A* і не належать відношенню *B.*

Синтаксис:
A MINUS *B*

5. Декартовій твір

Відношення $(A_1, A_2, \ldots, A_m, B_1, B_2, \ldots, B_m)$, заголовок якого є зчепленням заголовків відносин *A* $(A_1, A_2, \ldots, A_m)$ і *B* $(B_1, B_2, \ldots, B_m)$, а тіло складається з кортежів, які є зчепленням кортежів відносин *A* і *B:*

$(A_1, a_2, \ldots, a_m, b_1, b_2, \ldots, b_m)$

таких, що $(a_1, a_2, \ldots, a_m) \in A$, $(b_1, b_2, \ldots, b_m) \in B$.

Синтаксис:

A TIMES *B*

6. Вибірка (обмеження)

Узагальнена вибірка це унарний оператор, що записується як $\sigma_\varphi(R)$, де φ є формулою числення висловлень, що складається із атомів, дозволених у звичайній вибірці та логічних операторів $\wedge$(кон'юнкції), $\vee$(диз'юнкції) та $\neg$ (заперечення). Така вибірка вибирає всі кортежі із R для яких φ істина.

Або, вібірка – це відношення з тим же заголовком, що і у відносини *A,* і тілом, що складається з кортежів, значення атрибутів яких при підстановці в умову *с* дають значення ИСТИНА. *С* являє собою логічне вираження, до якого можуть входити атрибути відносини *A* і / або скалярні вирази. Синтаксис:
A WHERE *c*

7. Проекція

Проекція — це унарна операція, що записується як $\pi_{a_1,\ldots,a_n}(R)$, де $a_1, \ldots, a_n$ є множиною назв атрибутів. Результат проекції визначається як множина, що отримується із всіх кортежів із R, що обмежуються $\{a_1, \ldots, a_n\}$.

Інакше кажучи, Проекція – це відношення з заголовком *(X, Y, ..., Z)* і тілом, що містить безліч кортежів виду *(x, y, ..., z),* таких, для яких щодо *A* знайдуться кортежі зі значенням атрибута *X* рівним *x,* значенням атрибута *Y* рівним *y,* ..., значенням атрибута *Z* рівним *z.* При виконанні проекції виділяється "вертикальна" вирізка відносини-операнда з природним знищенням потенційно виникають кортежів-дублікатів. Синтаксис:
A [X, Y, ..., Z] або PROJECT *A {x, y, ..., z}*

8. Об'єднання

Операція об'єднання є результат послідовного застосування операцій декартового твору та вибірки. Якщо у відносинах і є атрибути з однаковими найменуваннями, то перед виконанням з'єднання такі атрибути необхідно перейменувати.

Операція об'єднання - збирає разом коржеті декількох відношень у один кортеж , проте у результуюче відношення включаються лише рядки, що задовільняють певним співвідношенням між атрибутами з''єднання відповідних відношень таблиць.

Синтаксис:

(A TIMES *B)* WHERE *c*

9. Розподіл

Відношення з заголовком $(X_1, X_2,..., X_n)$ і тілом, що містить безліч кортежів $(x_1, x_2,..., x_n)$, таких, що для всіх кортежів $(y_1, y_2,..., y_m) \in B$ щодо A $(X_1, X_2,..., X_n, Y_1, Y_2,..., Y_m)$ знайдеться кортеж $(x_1, x_2,..., x_n, y_1, y_2,..., y_m)$.

Синтаксис:

A DIVIDEBY *B*

Застосування реляційної алгебри в роботі з БД

Основні положення реляційної алгебри є основою для реалізації роботи СКБД безпосередньо з базою даних.

До відношень можна застосувати систему операцій, що дозволяють отримати одні відносини з інших. Виняток становлять операції створення та заповнення таблиць, а також операції опису та перейменування стовпців. Результатом запиту до реляційної БД може бути нове ставлення, обчислена на основі наявних відносин.

Операції з даними в реляційній базі даних включають операції над рядками (кортежами відношень) та операції над відношеннями.

Операції, що виконуються на рівні кортежів, - вставка (додається новий рядок), вилучення (знищується рядок), поновлення (здійснюються зміни значень атрибутів у рядках).

Основною одиницею обробки даних у реляційній моделі є відношення (файл).

Ефективність реляційної системи управління базою даних визначається здатністю виконувати над відношеннями операції алгебри відношень. Основними операціями над відношеннями РБД є традиційні операції над множинами: об'єднання, переріз, різниця (віднімання), декартів добуток.

Розглянемо основні оператори мови реляційної алгебри (для наочності відношення доповнено колонкою "Кортеж").

Об'єднання - операція виконується над двома сумісними відношеннями R1, R2 (з ідентичною структурою – d1, d2,dn) (табл. 2.19 і 2.20).

Таблиця 2.19

Науковий семінар 1

Кортеж	Прізвище та ініціали	Група
С11	Кубів І. П.	12

C12	Сомів М. В.	25
C13	Кріт І. В.	40
C14	Вовчук В. І.	31

Таблиця 2.20

Науковий семінар 2

Кортеж	Прізвище та ініціали	Група
C21	Кріт І. В.	40
C22	Кубів І. П.	12
C23	Якубів Н. З.	18

У результаті операції об'єднання будується нове відношення:

$$R = R_1 \cup R_2.$$

Відношення R має той самий склад атрибутів і сукупність кортежів вихідних відношень. Причому в цю сукупність не включаються дублікати (табл. 2.21).

Таблиця 2.21

Науковий семінар

Кортеж	Прізвище та ініціали	Група
C11 (C22)	Кубів І. П.	12
C12	Сомів М. В.	25
C13 (C21)	Кріт І. В.	40
C14	Вовчук В. І.	31
C23	Якубів Н. З.	18

У нове відношення не ввійшли кортежі C21 і C22, оскільки вони дублюють кортежі C11 і C13.

Переріз - операція, яка виконується над двома сумісними відношеннями R1,R2 (див. табл. 2.19, 2.20). Результуюче відношення $RP = R_1 \cap R_2$, містить однакові кортежі, які є в кожному з двох вихідних. Результат перетину має той самий склад атрибутів, що й у вихідних (табл. 2.22).

Таблиця 2.22 Перетин відношень

Кортеж	Прізвище та ініціали	Група
С11 (С22)	Кубів І. П.	12
С13 (С21)	Кріт І. В.	40

Різниця - операція виконується над двома сумісними відношеннями Rl9 R2 з ідентичним набором атрибутів. У результаті операції віднімання будується нове відношення RV = R] - R2 з ідентичним набором атрибутів, яке містить лише ті кортежі відношення Rj, які не повторюються в другому відношенні R (табл. 2.23).

Таблиця 2.23

Різниця відношень

Кортеж	Прізвище та ініціали	Група
С12	Сомів М. В.	25
С14	Вовчук В. І.	31

Декортів добуток виконується над двома відношеннями R1, R2, що мають різний склад атрибутів (d1, d2,…,d*n* і p1, p2,…, p*m*) відповідно (табл. 2.24, 2.25). У результаті операції декартового добутку утворюється нове відношення RD =R1 x R2, яке містить усі атрибути вихідних відношень (d1, d2,…,d*n* і p1, p2,…, p*m*). Результат відношення складається з можливих угруповань кортежів вихідних відношень R1 і R2.

Таблиця 2.24

Студент

Кортеж	Прізвище та ініціали	Група
С11	Кумів 0. Я.	15
С12	Васькін Л. М	23
С13	Мойсак Т. В.	3

Таблиця 2.25 Графік іспитів

Кортеж	Дисципліна	Дата
C21	АРМ бухгалтера	28.12.97
C22	Бухгалтерський облік	5.01.98

Кількість кортежів декартового добутку дорівнює добутку кількості кортежів у вихідних відношеннях (табл. 2.26).

Таблиця 2.26

Екзаменаційна відомість

Кортеж	Прізвище та ініціали	Група	Дисципліна	Дата
C11 (C21)	Кумів 0. Я.	15	АРМ бухгалтера	28.12.97
C12 (C21)	Васькін Л. М.	23	АРМ бухгалтера	28.12.97
C13 (C21)	Мойсак Т. В.	31	АРМ бухгалтера	28.12.97
C11 (C22)	Кумів 0. Я.	15	Бухгалтерський облік	5.01.98
C12 (C22)	Васькін Л. М.	23	Бухгалтерський облік	5.01.98
C13 (C22)	Мойсак Т. В.	31	Бухгалтерський облік	5.01.98

Наведений перелік операцій з даними в реляційній моделі демонструє простоту переформатування відношень залежно від конкретних ситуацій.

Принци Діріхле

При́нцип Діріхле́ (також *принцип коробок Діріхле*, *принцип голубів і кліток*) — комбінаторне твердження, сформульоване німецьким математиком Петером Діріхле.

Формулювання

Найчастіше в україномовній і російськомовній літературі використовується неформальне формулювання з кролями і клітками. В англомовній літературі частіше у формулюванні присутні голуби (звідси поширена назва **pigeonhole principle**).

Найбільш поширене наступне формулювання цього принципу:

Припустимо, деяке число кроликів розсаджені в клітках. Якщо число кроликів більше, ніж число кліток, то хоч би в одній з кліток буде більше одного кролика.

Більш загальне формулювання:

Припустимо, $\mathbf{m}$ *кроликів розсаджені в* $\mathbf{n}$ *клітках. Тоді, якщо* $\mathbf{m} > \mathbf{n}$*, то хоч би в одній клітці міститься не менше* $\lceil m/n \rceil$ *кроликів, а також хоч би в одній іншій клітці міститься не більше* $\lfloor n/m \rfloor$ *кроликів.*

У рамках більш абстрактних понять:

Нехай задана функція $f : A \to B$ *і потужність множини* A *більше потужності* B*. Тоді функція* f *не є ін'єктивною.*

Можливі також формулювання для окремих випадків, наприклад:

Якщо число кліток більше, ніж число кролів, то принаймні одна клітка порожня.

Приклади застосування

- 10 друзів відправили один одному святкові листівки. Кожний із них відправив 5 листівок. Довести, що якихось двоє друзів відправили листівки один одному.

  Доведення. кількість пар, що можна утворити з 10 друзів: $C^2_{10} = 45$. А всього відправлених листівок 5·10=50. Отже, згідно з принципом Діріхле, на деякі з пар друзів припадає дві листівки.
- Картки пронумеровані послідовно цілими числа ми від *1* до *2n +1*. Яку найбільшу кількість карток можна вибрати так, щоб жоден з номерів не дорівнював сумі якихось двох інших номерів карток?

  Розв'язання. Припустімо, що таких карток існує не менше ніж *n+2*. Розташувавши вибрані картки в порядку зростання їхніх номерів, віднімемо від усіх номерів найменший номер картки. Одержується *n + 1* різних чисел, відмінних від 0. Тоді, згідно з принципом Діріхле, отримана множина має принаймні один спільний елемент із початковою. Це число відповідно буде сумою двох чисел. Легко перевірити, що для *n + 1* картки з непарними номерами *{1,3,5,..., 2n +1}* умови задачі вже виконуються.
- З використанням принципу Діріхле можна довести, що для довільного ірраціонального a, множина $\{[na]: n \text{ ціле число}\}$ дробових частин є щільною в [0, 1]. Взявши M таке що $1/M < e$, згідно з принципом Діріхле серед чисел $n_1, n_2 \in \{1, 2, ..., M + 1\}$ such that n_1a and n_2a існують такі два n_1, n_2, що n_1a належить $(p + k/M, p + (k + 1)/M)$, і n_2a належить $(q + k/M, q + (k + 1)/M)$, для деяких цілих чисел p, q і деякому k з $\{0, 1, ..., M - 1\}$. Легко перевірити, що $(n_2 - n_1)a$ належить $(q - p - 1/M, q - p + 1/M)$. Тобто $[na] < 1/M < e$, де $n = n_2 - n_1$ або $n = n_1 - n_2$. Тобто 0 є граничною точкою множини $\{[na]\}$. З цього можна отримати твердження для довільного p з (0, 1]: нехай n таке що $[na] < 1/M < e$; тоді якщо $p \in (0, 1/M]$, одержується твердження. В іншому випадку p з $(j/M, (j + 1)/M]$, і взявши $k = \sup\{r \in N : r\{na\} < j/M\}$, одержуємо $|\{(k + 1)na\} - p| < 1/M < e$.