

ПРАКТИЧНІ АСПЕКТИ ЗАСТОСУВАННЯ ВІКОННИХ ФУНКЦІЙ ПРИ РОБОТІ З БАЗАМИ ДАНИХ

*Доц., канд. ф-м. наук І.В. Рассоха, доц., канд. пед. наук С.П. Рендюк
Національний університет «Полтавська політехніка
імені Юрія Кондратюка», м. Полтава, Україна*

*Ст. інженер-програміст О.В. Рассоха
GRID DYNAMICS POLAND SP. Z O.O., м. Вроцлав, Польща*

Як відомо, сучасна робота з базами даних передбачає достатньо високий рівень володіння математичним апаратом, що забезпечує ефективний аналіз з метою оптимізації, прогнозування, встановлення трендів тощо. Серед розділів математики, що активно використовуються при роботі із СУБД можна назвати алгебру, теорію ймовірностей та математичну статистику, комбінаторику тощо. При цьому одним із ключових понять є поняття функції [1].

Найчастіше використовуються агрегатні функції. З точки зору математики агрегатною функцією є скалярна функція, що залежить від скінченної множини значень: вектор, список, назви стовпчиків, функції, будь які їхні комбінації тощо [2]. Такі функції володіють рядом математичних характеристик: комутативність, асоціативність, інваріантність до порядку, редукція розмірності, інваріантність до повторів.

Іншим класом функцій, що широко застосовуються на практиці при роботі з СУБД, є віконні функції. Зупинимося детально на їх особливостях та принципах використання.

У загальному вигляді під поняттям віконної функції розуміють оператор, що застосовується до частини (вікна) послідовності даних та повертає результат, який може змінюватися для кожного елемента. Таким чином, віконна функція має своїми аргументами не весь набір даних, а лише його локальну підмножину («вікно»). Таке «вікно» визначається за певними правилами в залежності від контекста задач (задачами сортування, умовами на позицію елементів в послідовності). Таким чином, математично віконну функцію можна записати наступним чином.

$$y_i = f(W_i), \quad (1.1)$$

де W_i – локальне вікно, підмножина впорядкованої послідовності $X = \{x_1, x_2, \dots, x_n\}$. В данному випадку функція f може бути агрегатою (наприклад, ковзне середнє значення) або аналітичною функцією. Віконна функція має певні властивості: локальність, сталість або зростання розмірності та інші. Звісно, що віконна агрегатна функція додатково володіє властивостями агрегатних функцій, аналітична функція – властивостями аналітичних функцій данного класу. Віконна функція є функціоналом, тому працює не з окремими елементами, а з множиною значень.

З точки зору дискретного функціонального аналізу віконні функції можуть бути розділені на наступні класи: лінійні оператори, нелінійні функції впорядкування та оператори локального контексту.

Наведемо деякі практичні приклади застосування віконних функцій для побудови запиту на SQL.

- 1) Знаходження середнього по підмножині (*Partitioned average*).

$$\text{Локальне вікно: } \{x_i, \dots, x_{i+j}\}, \text{ середнє: } y_i = \frac{1}{j}(x_i + \dots + x_{i+j})$$

Приклад. Середня зарплата по відділу.

SQL: AVG(salary) OVER (PARTITION BY department_id)

Таблиця 1 – Вхідна таблиця

employee_id	department_id	salary
1	10	4000
2	10	5000
3	20	6000

Таблиця 2 – Результат

employee_id	department_id	salary	avg_dept_salary
1	10	4000	4500
2	10	5000	4500
3	20	6000	6000

Середнє значення по певній множині широко використовується для вирішення економічних завдань. Таких, як знаходження середньої заробітної плати, середньої ціни товарів, середнього доходу та інших економічних показників.

- 2) Знаходження ковзного середнього (*Moving average*).

$$\text{Локальне вікно: } \{x_i, \dots, x_{i-k+1}\}, \text{ ковзне середнє: } y_i = \frac{1}{k}(x_i + \dots + x_{i-k+1})$$

Приклад. Ковзне середнє температури

SQL: AVG(temp) OVER (ORDER BY day ROWS BETWEEN 1 PRECEDING AND CURRENT ROW)

Таблиця 3 – Вхідна таблиця

day	temp
Monday	10
Tuesday	14
Wednesday	12

Таблиця 4 – Результат

day	temp	moving_avg
Monday	10	10.0
Tuesday	14	12.0
Wednesday	12	13.0

Ковзне середнє широко використовується в економіці для згладжування коливань в часових рядів, що дозволяє виявити довгострокові тенденції [3].

3) Знаходження кумулятивної суми (*Cumulative sum*).

Локальне вікно: $\{x_1, x_2, \dots, x_i\}$. Кумулятивна сума $y_i = \sum_{j=1}^i x_j$

Приклад. Кумулятивні продажі.

SQL: SUM(sales) OVER (PARTITION BY region ORDER BY date)

Таблиця 5 – Вхідна таблиця

region	date	sales
A	2024-01-01	100
A	2024-01-02	200
A	2024-01-03	150
A	2024-01-03	120

Таблиця 6 – Результат

region	date	sales	cum_sales
A	2024-01-01	100	100
A	2024-01-02	200	300
A	2024-01-03	150	450
A	2024-01-03	120	570

Приклад. Сума продажів за останні три дні.

SQL: SUM(sales) OVER (ORDER BY date ROWS BETWEEN 2 PRECEDING AND CURRENT ROW)

Таблиця 7 – Вхідна таблиця

date	sales
2024-01-01	50
2024-01-02	60
2024-01-03	40
2024-01-04	100

Таблиця 8 – Результат

date	sales	rolling_sum
2024-01-01	50	50
2024-01-02	60	110
2024-01-03	40	150
2024-01-04	100	200

Кумулятивна сума часто використовується для аналізу тенденцій, зміни обсягів продажів, виробництва та іншого.

4) Лаг (затримка) та лід (випередження).

Локальне вікно: x_{i-j} , лаг: $y_i = x_{i-j}$. Локальне вікно: x_{i+j} , лід: $y_i = x_{i+j}$

Приклад. Зміна продажів порівняно з попереднім та наступним днями.

Таблиця 9 – Вхідна таблиця

date	sales
2024-01-01	100
2024-01-02	150
2024-01-03	120

SQL: sales – LAG(sales) OVER (ORDER BY date)

Таблиця 10 – Результат

date	sales	diff_prev_day
2024-01-01	100	NULL
2024-01-02	150	50
2024-01-03	120	-30

SQL: sales – LEAD(sales) OVER (ORDER BY date)

Таблиця 11 – Результат

date	sales	diff_prev_day
2024-01-01	100	-50
2024-01-02	150	30
2024-01-03	120	NULL

Функції лагу та ліду застосовуються в економіці для опису та аналізу динаміки економічних процесів. Наприклад, відхилень від запланованих показників.

5) Знаходження рангу (*Rank*).

Локальне вікно: $\{x_i, x_i \in X, \text{впорядковано}\}$, ранг y_i : позиція x_i у впорядкованій множині X .

Приклад. Рейтинг товарів за ціною.

SQL: RANK() OVER (PARTITION BY category ORDER BY price DESC)

Таблиця 12 – Вхідна таблиця

product	category	price
A	Phones	300
B	Phones	500
C	Phones	300

Таблиця 13 – Результат

product	category	price	price_rank
B	Phones	500	1
A	Phones	300	2
C	Phones	300	2

В економіці, ранжування застосовується для оцінки, порівняння та відбору різноманітних об'єктів, таких як компанії, проекти, інвестиції, або економічні показники.

Наведені приклади ілюструють застосування елементарних віконних функцій. Часто в практичних задачах зустрічаються їх композиції. Тобто результат однієї віконної функції є аргументом або локальним вікном іншої.

Висновки

Таким чином, у роботі з базами даних віконні функції (*window functions*) стали невід'ємним інструментом для аналітики, обчислень і статистичної обробки. Одночасно віконні функції мають чітке математичне підґрунтя: вони працюють як оператори над послідовностями, подібно до згорток, інтегралів, порядкових відображень і статистичних функцій у вищій математиці.

Посилання

1. Claudio Fratarcangeli. Technique for Universal Quantification in SQL. SIGMOD RECORD, Vol. 20, No. 3, September 1991. – pp. 16 – 24.
2. Dejan Sarka, Greg Low, Roger Wolter, Ed Katibah, Isaac Kunen, Itzik Ben-Gan, Inside Microsoft SQL Server 2008: T-SQL Programming (Developer Reference) 1st Edition Microsoft Press, 2010, P.640.
3. Zafar, RF, Khoo, MBC, You, HW, Saha, S., & Yeong, WC (2024). Distributional control chart of a weighted moving average. Journal of Applied Statistics, 52 (3), 744–777.