

ЕКОНОМІКО-МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ

УДК 004.652+338

Устенко С.В.

доктор економічних наук, професор

Дмитрів А.А.

аспірант кафедри інформаційних систем в економіці

ФОРМАЛЬНИЙ ОПИС АГРЕГАЦІЇ ДАНИХ

Find similar papers at core.ac.uk

provided by Institutional Repository of Vadym Hetma

Ustenko S.

Doctor of Economics, professor

Dmitryv A.

PhD student Information Systems in economy Division

FORMAL DESCRIPTION OF DATA AGGREGATION IN MULTIDIMENSIONAL ORACLE-SYSTEMS ONLINE ANALYTICAL DATA PROCESSING FOR SOLVING ECONOMIC TASKS

АНОТАЦІЯ. У статті вивчено можливості багатовимірних Oracle-систем аналітичної обробки даних на прикладі побудови формальної схеми моделі даних для вирішення економічних розрахунків. Аналізуються методи агрегації даних як засіб подальшого аналізу економічної інформації за допомогою такої системи. Технологія аналітичної обробки даних надає можливість їх представлення у різноманітних розрізах для полегшеного аналізу великих об'ємів інформації.

КЛЮЧОВІ СЛОВА: обчислювальна операція, багатовимірнамодель, агрегація даних, вимір, ієрархія даних, Oracle-система.

ANNOTATION. The article examines multidimensional Oracle analytical data processing systems possibilities as an example based on formal data model scheme for solving economic calculations. Data aggregation methods as means of the further economic information analysis is being examined using the above-mentioned system. Technology of analytical processing allows to represent data in different contexts for easier analysis of large amounts of information.

KEY WORDS: calculation, multidimensional model, data aggregation, dimension, data hierarchy, Oracle system.

Постановка проблеми. Багатовимірні Oracle-системи являються ефективним засобом для аналізу великих об'ємів інформації та представлення даних. Стрімкий розвиток управлінських інформа-

ційних технологій спонукає робити постійний моніторинг нових продуктів, технологій і рішень. Навіть перебуваючи в курсі останніх ринкових тенденцій, часом нелегко оцінити, який економічний ефект можна отримати від технологій, які переховуються за загадковими аббревіатурами OLAP, OLTP, DW, DM. Для сучасного підприємства характерна ситуація, коли є кілька систем автоматизації (для вирішення різних завдань), дані зберігаються розрізнено і, як наслідок, відсутній єдиний погляд на управлінську інформацію. Отже, одним із найактуальніших завдань, що стоять перед компаніями, є перетворення розрізнених даних у структуровану інформацію, описану в економічних термінах і дозволяє приймати обґрунтовані управлінські рішення. Саме для цього призначені спеціальні системи аналітичної обробки даних у режимі реального часу, наприклад провідним продуктом такої системи є Oracle Hyperion. Щоб моделювати діяльність компанії та оперативно складати точні звіти, потрібно вміти вибирати дані з багатьох джерел одночасно. Тому багатовимірні системи спеціально налаштовуються на різні формати джерел даних і формують багатомірний масив корпоративної інформації. Такі системи підключені до багатовимірної БД, де елементи відображені в ієрархічній структурі. Для реалізації такої системи необхідне існування об'єктно-реляційної системи управління базами даних (RDBMS), в якій зберігається інтегрований опис кожного з елементів, отриманих від зовнішніх джерел. Дані такої СУБД побудовані на базі таблиць, полів і записів. Кожна таблиця складається з рядків бази даних. Кожна таблиця рядка бази даних складається з одного або кількох полів таблиці бази даних. Кожний елемент окрім коду (Name, є унікальним) та назви (Alias, в БД існує один чи більше Alias для одного такого елемента) містить інформацію про приналежність до певного батьківського класу (Parent), багатовимірного представлення даних або кількох таких (кожна схема багатовимірного представлення даних створюється в outline, в одному outline можна створити два чи більше таких представлень), типу агрегації для кожного представлення окремо (+, -, *, /, %, ~), формату збереження даних (наприклад: постійне зберігання, динамічно-розрахунковий, динамічний зі зберіганням розрахунку, ніколи не поділяючий, формат тільки назва), типу аккаунта (Account, якщо це стаття, яка відноситься до економічних процесів, то відповідно формат — витратний, актив, пасив, валютний, невалютний і т.п.). Тим самим, кожне багатовимірне представлення містить виміри (Dimensions), наприклад такі: період (рік, квартал, місяць і т.п.), рік (2014, 2015...2020...), версія планування (використовується для планування бюджету) та подібні залежно від предметної області.

Кожний із вимірів може мати елементи з певною для цього виміру інформацією про них.

Аналіз останніх досліджень і публікацій. Питання розвитку багатовимірних oracle-систем аналітичної обробки даних стало об'єктом дослідження команди розробників інформаційно-аналітичного центру управління ефективністю діяльності підприємства (ERP Information Development Team). Також проаналізовано публікації, пов'язані зі структурою, математичною моделлю OLAP-систем, агрегацією показників за допомогою них, таких авторів: В.П. Кулагін, В.Т. Матчин, В. В. Миронов, Е. С. Макарова, П. Мальцев, С.Д. Кузнецов, док. Анжана Госейн (Dr. Anjana Gosain).

Формування цілей статті. Метою статті є формально, за допомогою математичних формул та операторів, описати процедури розрахунку даних та їх агрегації у багатовимірних Oracle-системах. Процедура агрегації даних дозволяє перетворити групи значень елементів у значення, що містять агреговану інформацію по відповідній групі, і створювати новий — агрегований файл даних або створювати нові змінні в активному наборі даних, що містять агреговану інформацію. Агрегація — це взаємозв'язок одного батьківського класу сутностей з кількома чи більше дочірніми класами сутностей. У даній статті наведено схематичну модель за умови агрегації та розрахунку даних при їх аналітичній обробці за допомогою багатовимірних систем.

Викладання основного матеріалу. Нехай маємо елементи багатовимірної інформаційної моделі, які представляються у вигляді множини:

$$T = \{D_1, D_2, \dots, D_n\}, \quad (1)$$

де D_i — i -ий вимір у даній схемі представлення БД. Вимір — це список значень, що належать до одного і того ж типу даних з точки зору користувача. Наприклад, дні, місяці, квартали і роки з точки зору користувача відносяться до одного і того ж типу «Час». Поняття вимірів дозволяє інтуїтивно зрозуміло організувати роботу людину з даними. Відповідно, якщо взяти до уваги деяку множину елементів виміру в ієрархії, то представлення буде мати вигляд:

$$\Psi = \Psi_1 \cup \Psi_2 \cup \dots \cup \Psi_q, \quad (2)$$

де Ψ_i — множина елементів рівнів i -ого виміру, q — кількість вимірів. Графічно ієрархічна структура може бути представлена у вигляді дерева. Під генерацією або поколінням (generation) елемента треба розуміти ранг елемента від «старшого» до «молодшого» починаючи з одиниці (не плутати з рівнем ієрархії, ранг елемента якого визначається навпаки: від «молодшого» дочірнього до «старшо-

го» батьківського). Позначимо через DH_i^g — g -е покоління ієрархії i -ого виміру, тобто кожна множина ψ_i буде мати вигляд:

$$\begin{aligned} \psi_1 &= \{DH_1^1, DH_1^2, \dots, DH_1^m\} \text{ — множина поколінь 1-ого виміру,} \\ \psi_2 &= \{DH_2^1, DH_2^2, \dots, DH_2^m\} \text{ — множина поколінь 2-ого виміру,} \\ &\dots \\ \psi_q &= \{DH_q^1, DH_q^2, \dots, DH_q^m\} \text{ — множина поколінь } q\text{-ого виміру.} \end{aligned}$$

Для кожного виміру існує множина значень, які належать йому. Нехай Y_i — множина значень усіх елементів i -ого виміру:

$$Y_i = \{C_i^1, C_i^2, \dots, C_i^k\}, \quad (3)$$

де C_i^j — j -е значення елемента i -ого виміру, k — кількість елементів виміру[1].

На прикладі схеми (рис. 1) організації ієрархії на прикладі розміщення філій по регіонах можна детально зрозуміти відношення кожного елемента до свого значення, виміру, генерації ієрархії. У даному випадку вимір позначено деяким значенням D_6 .

Невід’ємною частиною багатовимірної системи є виконання функцій агрегації. Оскільки метою цієї системи є створення багаторівневої моделі аналізу, дані на рівнях, відмінних від фактичного, повинні бути відповідним чином агреговані. Важливо відзначити, що по кожному виміру можна задавати власну (і не одну) функцію агрегації. Дистрибутивні функції дозволяють розбивати вхідні дані і обчислювати окремі підсумки, які потім можна об’єднувати.

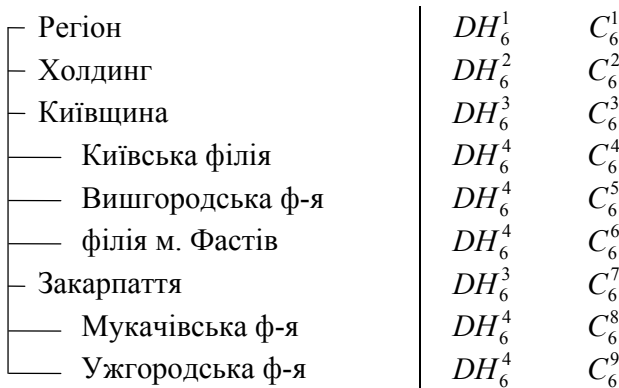


Рис. 1. Ієрархічна схема організації виміру «Регіон» D_6

Алгебраїчні функції можна представити комбінацією з дистрибутивних функцій (наприклад, $Average()$ можна представити як $\frac{sum()}{count()}$). Холістичні функції неможливо обчислювати на часткових даних або представляти будь-яким чином [4].

При виконанні обчислювальної операції у багатовимірній системі необхідно звертатися до певних блоків, де знаходяться дані. Блок — найменша одиниця виділення простору в Oracle. У блоках зберігаються рядки даних, індексів або проміжні результати сортувань. Саме блоками сервер Oracle зазвичай виконує читання і запис на диск. Сегмент складається з одного або більше екстентів, а екстенст — це група розташованих поспіль блоків. Розмір блоку в базі даних з моменту її створення — величина постійна, тому всі блоки в базі даних одного розміру.

Міжвимірний оператор (\rightarrow) вказує на значення даних конкретних комбінацій членів. Міжвимірний оператор створюється з дефісом ($-$) і правою кутовою дужкою ($>$), без пробілу між ними. Міжвимірний оператор необхідний для звернення до конкретних блоків даних у БД при виконанні операцій, створюючи зріз даних [2].

Нехай маємо у системі створені ще три виміри D_1, D_3, D_4 . Перший стосується показників (D_1), тобто вимір, у якому створено економічні показники, які пов'язані між собою формулами чи певним економічним змістом; третій — «Період» $[(D)_3]$, — це вимір, за допомогою якого дані можна заповнити чи відобразити у певному розрізі часу; четвертий — «Продукт», тобто вимір ієрархічно представляє продукцію/товари, щов даному випадку реалізуються $[(D)_4]$:

Показник	DH_1^1	C_1^1
— Реалізація, од.	DH_1^2	C_1^2
— Ціна реалізації, грн/од.	DH_1^2	C_1^3
— Собівартість, грн/од.	DH_1^2	C_1^4
— Дохід, грн	DH_1^2	C_1^5
...	DH_1^2	...

Рис. 2. Ієрархічна схема виміру «Показник» D_1

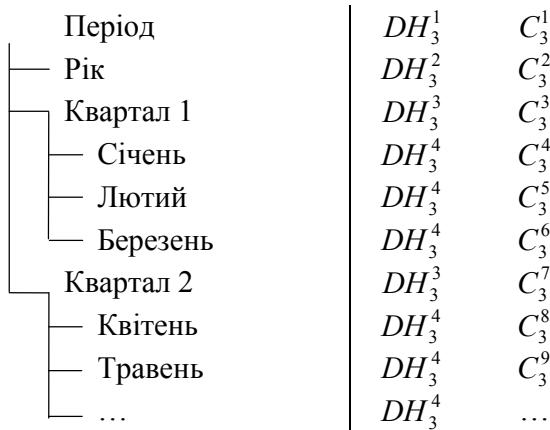


Рис. 3. Ієрархічна схема виміру «Період» D_3

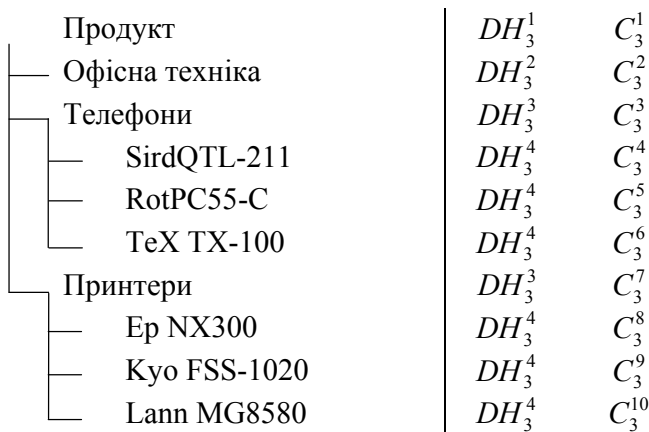


Рис. 4. Ієрархічна схема виміру «Продукт» D_4

Зведення за вимірами — процедура, що передбачає певну послідовність агрегування дочірніх елементів. А саме, спочатку елементи агрегуються за координатами одного виміру, потім отриманий агрегат, не залежний від цього виміру, агрегуються за координатами іншого виміру, далі — по координатам третього виміру, і т.д. до останнього виміру, при агрегації за координатами якого виходить шуканий зведений показник. Якщо необхідне агрегування допускає зведення за вимірами по ієрархії, то це створює зручності на практиці. Формально можна позначити фу-

нкцією *Agg* об'єднання значень бази даних. По суті ця функція означає виконання зведення елементів за вимірами:

$$Agg(C_i^j | DH_i^g). \quad (4)$$

— запис означає, що система виконає агрегацію для *j*-ого значення елемента усіх його дочірніх елементів, починаючи операцію від *g*-ого покоління певного *i*-ого виміру. Наприклад, $Agg(C_3^7 | DH_3^3)$ означає агрегацію по елементам усіх вимірів до першого покоління, окрім вказаного обмеження по виміру «Період» елемента «Квартал 2», тобто у розрізі «Рік» або «Квартал 1» значення не будуть зведені та не міститимуть значень.

Класична функція агрегації означає зведення даних до батьківського елемента від дочірніх. При плануванні цільового бюджету, де на будь-який батьківський рівень вказується необхідне цільове число, відбувається підрахунок дочірніх за принципом розподілу батьківського елемента у певних пропорціях чи формульних відношеннях. При зведенні значень від найнижчого рівня чи найбільшого покоління елементів для показників типу «Собівартість», «Ціна реалізації» виконується їх перерахунок на батьківському рівні, тому що втрачається сенс виконання агрегації таких показників ієрархії. Скасування виконання такої операції відбудеться при віднесенні його до неагрегованих. Такі елементи позначаються знаком ~ у RDBMS.

У багатовимірних Oracle-систем можливе виконання команди *AGG (DimList)*. Ця команда відмінна від формального запису функції *Agg*. Не вникаючи сутність мови програмування, для таких багатовимірних систем використовується дана формальна функція. Команда *AGG* виконує агрегацію тільки при вказанні виміру чи списку вимірів через кому.

Підрахуємо кількість реалізованих усіх типів принтерів за перший квартал Ужгородської філії за допомогою функції агрегації:

$$Agg(C_1^2 | DH_1^2 \rightarrow C_3^7 | DH_3^3 \rightarrow C_6^9 | DH_6^4 \rightarrow C_3^3 | DH_3^3). \quad (5)$$

Елементи функції пов'язані міжвимірним оператором, вказуючи обмеження за яким необхідно виконати розрахунок у системі. Опис такої операції полягає у зв'язку даних по ієрархії:

«Реалізація, од.» → «Принтери» → «Ужгородська ф-я» → «Квартал 1».

У даному випадку функція агрегації виконує операцію у розрізі записаних вище елементів, піднімаючи їх дочірніх за ієрархічним зв'язком, якщо такий існує. Записи створюються у блоках багатовимірної бази даних. Існує варіант перегляду таких даних у різних розрізах для аналізу отриманої інформації, наприклад, значення по

показнику «Реалізація, од.» можна також побачити у розрізі будь-якого місяця першого кварталу для реалізованих принтерів:

«Реалізація, од.» → «Принтери» → «Ужгородська ф-я» → «Січень».

До виконання функції агрегації блоки в даному розрізі були закриті і не створені системою (#MISSING).

Висновок. Технологія багатовимірних баз даних також застосовується до нових типів даних, які сучасні технології часто не в змозі адекватно аналізувати. Наприклад, класичні методики, такі як агрегація, не можуть гарантувати невеликий час відповіді на запити, якщо дані постійно змінюються, як це відбувається, наприклад, коли інформація надходить з датчиків або від рухомих об'єктів, таких як автомобілі, оснащені засобами глобального позиціонування.

Технологія багатовимірних баз даних усе більше буде застосовуватися там, де результати аналізу безпосередньо передаються в інші системи, тим самим, виключаючи участь людини в цьому процесі. Цей контекст у сукупності з необхідністю постійного оновлення пред'являє жорсткіші вимоги до продуктивності, яким не задовольняє сучасна технологія. Деякі Oracle-системи призначені для побудови аналітичних систем, заснованих на принципах багатовимірної технології. Такі системи дозволяють зберігати й обробляти в одній і тій же базі даних реляційну і багатовимірну інформацію для вирішення економічних та інших завдань.

Література

1. Семченков С.Ю. Формальное представление структуры систем аналитической обработки данных, основанных на OLAP-технологии. – // Рязанский государственный радиотехнический университет. — 2010.

2. Oracle ESSBASE Technical reference /Authors: EPM Information Development Team // Oracle Enterprise performance management system documentation release 11.1.2.2. — 2015. — [Електронний ресурс]. — https://docs.oracle.com/cd/E17236_01/epm.1112/esb_tech_ref.pdf

3. Database Administrator's Guide / Authors: EPM Information Development Team //Oracle Enterprise performance management system documentation release 11.1.2.2. — 2015.– [Електронний ресурс]. — https://docs.oracle.com/cd/E17236_01/epm.1112/aps_admin.pdf

4. Sanjay Goil and Alok Choudhary. High performance OLAP and Data Mining on parallel computers. — Center of Parallel and Distributed Computing Technical Report TR-97-05, — 1997.

References

1. Semchenkov S.YU. Formal'noye predstavleniye struktury sistem analiticheskoy obrabotki dannykh, osnovannykh na OLAP-tekhnologii[The

formal representation of systems analytical processing structure, based on OLAP-technologies.].Ryazan State Radio Engineering University. 2010. [In Russian].

2. Oracle ESSBASE Technical reference / Authors: EPM Information Development Team. Oracle Enterprise performance management system documentation release 11.1.2.2.2015. [E-resource]. https://docs.oracle.com/cd/E17236_01/epm.1112/esb_tech_ref.pdf

3. Database Administrator's Guide / Authors: EPM Information Development Team. Oracle Enterprise performance management system documentation release 11.1.2.2. 2015. [E-resource]. https://docs.oracle.com/cd/E17236_01/epm.1112/aps_admin.pdf.

4. Sanjay Goil and Alok Choudhary. High performance OLAP and Data Mining on parallel computers.Center of Parallel and Distributed Computing Technical Report TR-97-05,1997.

УДК [347.238.8+658.727+658.847.7.]:519.876.5:004.

Кікоть О. Ю.,

здобувач, кафедра менеджменту банківської діяльності ДВНЗ «КНЕУ імені Вадима Гетьмана», викладач Дніпродзержинського технологічного коледжу Дніпродзержинського державного технічного університету

КОНЦЕПТУАЛЬНІ ЗАСАДИ МОДЕЛЮВАННЯ АУКЦІОНІВ

Kikot O.Y.,

Postgraduate, Department of Banking Management, Kyiv National Economic University named after Vadym Hetman Lector of Economics, Dneprodzerzhinsk College of Technologies, Dneprodzerzhinsk State Technical University

CONCEPTUAL FOUNDATIONS OF MODELING AUCTIONS

АНОТАЦІЯ. В статті запропоновано загальний підхід до моделювання аукціонів. Також розглянуто характеристику елементів моделі аукціону, а саме систему припущень, які дозволяють моделювати економічне середовище аукціонних торгів, правила проведення аукціону та алгоритм визначення цінових пропозицій його учасниками.

КЛЮЧОВІ СЛОВА: аукціон, економічна модель, економічне середовище, механізм, формат аукціонних торгів, підхід до вирішення проблеми, аукціон першої ціни.