**Лекція 2**

**Методи багатомірного розвідувального аналізу**

[**1. Багатовимірні моделі**](http://vns.lp.edu.ua/moodle/mod/page/view.php?id=21596)

Багатовимірна база даних – база даних, що підтримує багатовимірну модель даних на концептуальному рівні та призначена для інтерактивного аналітичного опрацювання аґреґованих історичних і проґнозованих даних.

Основними поняттями багатовимірної моделі даних є:

 гіперкуб даних rel,

 вимір V,

 атрибут A,

 комірка X,

 значення rel(V,A).

Гіперкуб даних містить один або більше вимірів і є впорядкованим набором комірок. Кожна комірка визначається одним і лише одним набором значень вимірів – атрибутів. Комірка може містити дані – значення або бути порожньою.

Під виміром розумітимемо множину атрибутів, що створюють одну з граней гіперкуба. Прикладом часового виміру є список днів, місяців, кварталів. Прикладом географічного виміру може бути перелік територіальних об'єктів: населених пунктів, районів, реґіонів, країн та ін.

Для отримання доступу до даних користувачу необхідно вказати одну або декілька комірок шляхом вибору значень вимірів, яким відповідають необхідні комірки. Процес вибору значень вимірів називатимемо фіксацією атрибутів, а множину вибраних значень вимірів – множиною фіксованих атрибутів.

Гіперкуб даних можна розглядати як множину відношень реляційної бази даних за значеннями кожного з вимірів. Отже, носієм багатовимірної моделі даних є відношення реляційної бази даних, зображені як зафіксовані виміри. Гіперкуб утворює багатовимірний простір даних, що складається з фіксованої кількості вимірів.

Операція зрізу здійснюється з метою отримання необхідної підмножини комірок і відсікання «непотрібних» значень шляхом послідовної фіксації атрибутів. Зріз, як правило, є двовимірним масивом (таблицею).

Зміна порядку подання (візуалізація) вимірів називається обертанням. Обертання забезпечує можливість візуалізації даних у формі, найбільш комфортній для їх сприйняття.

Операції згортки і деталізації здійснюються завдяки наявності ієрархічної структури вимірів. Значення вимірів (атрибутів) можуть об'єднуватися в ієрархії, що складаються з одного або декількох рівнів.

Багатовимірну модель даних (гіперкуб даних) можна подати за допомогою паралелотопу – геометричної фігури (множини точок), одержаної всіма можливими операціями згортки детальної точки, де значення міри в точках постійне. Паралелотоп є одним із підвидів моделі Версо.

Багатовимірна модель підтримує такі типи даних: класифікаційні правила (у термінах видобування знань називають ще асоціативними правилами); залежності між аґреґатами; залежності між елементами ієрархії.

Багатовимірна модель розширює реляційну: носій – вкладені (ненор-малізовані), залежні від часу дані, множина операцій, окрім традиційних реляційнх, містить операції зрізу, згортки, деталізації, обертання.

**2. OLAP-системи**

Основне призначення OLAP-систем - підтримка аналітичної діяльності, довільних запитів користува-чів-аналітиків. OLAP є аналітичним інструментом і спочатку ґрунтувався на багатовимірних базах даних (ВБД). Вони сконструйовані спеціально для підтримки аналізу кількісних даних з численною кількістю вимірювань, містять дані у багатовимірному вигляді.

***On-Line Analytical Processing*** (OLAP) - технологія оперативної аналітичної обробки даних, що використовує методи і засоби для збору, зберігання та аналізу багатовимірних даних з метою підтримки процесів прийняття рішень.

У1993 p. Е, Кодд опублікував статтю під назвою "OLAP для користувачів-аналітиків: яким він має бути". У цій роботі було запропоновано основні концепції оперативної аналітичної обробки і визначено вимоги, яким мають задовольняти продукти, що дають можливість виконувати оперативну аналітичну обробку.

OLAP дає змогу організувати вимірювання у вигляді ієрархії. Дані представлені у вигляді гіперкубів (кубів) - логічних і фізичних моделей показників, що спільно використовують вимірювання, а також ієрархії у цих вимірюваннях. Деякі дані заздалегідь агреговані в БД, інші розраховуються відразу.

OLAP-куб містить базові дані та інформацію про вимірювання (агрегати). Куб потенційно містить всю інформацію, потрібну для відповідей на будь-які запити.

Засоби OLAP дають змогу досліджувати дані за різними вимірюваннями. Користувачі можуть вибирати, які показники аналізувати, які вимірювання і як відображати в крос-таб-лиці, поміняти рядки і стовпці pivoting, потім робити зрізи, щоб концентруватися на певній комбінації розмірностей. Можна змінювати деталізацію даних, рухаючись рівнями за допомогою деталізації та збільшення, а також крос-деталізацію через інші вимірювання.

Для підтримки ББД використовуються OLAP-сервери, оп-тимізовані для багатовимірного аналізу і які поставляються з аналітичними можливостями.

Нині реляційні СУБД застосовуються для емуляції ББД, вони підтримують багатовимірний аналіз.

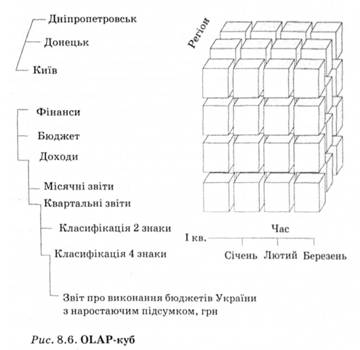
Причина виникнення OLAP для обробки запитів - це передусім забезпечення швидкості дій. Реляційні БД зберігають сутності в окремих нормалізованих таблицях. Ця структура зручна для операційних БД (систем OLTP), але складні багато-табличні запити в ній виконуються повільно. Зручнішою моделлю для запитів є просторові БД.

Основне призначення OLAP-систем - підтримка інформаційно-аналітичної діяльності, тобто довільних запитів користувач і в-а на літикі в. Якщо системи регламентованої звітності відповідають на питання типу "Які сумарні витрати до-могосподарств у І кварталі 2007 р. на споживання товарів довготривалого користування?", то OLAP покликаний дати відповіді, скажімо, на питання: ''наскільки треба забезпечити зростання темпів підвищення кінцевих споживчих витрат державного сектору на індивідуальне споживання, щоб перевершити темпи його витрат на колективне споживання у півтора раза?" Мета OLAP-аналізу - перевірка потенційних гіпотез.

Аналізуючи макропоказник, наприклад, виконання місцевих бюджетів, генерується звіт, що визначає показники конкретного району, регіону у певний момент часу. Таким чином, обсяг даних може бути представлений у вигляді тривимірного куба (рис. 8.6), грані якого - це масиви даних за показниками, регіонами і часом.

У процесі аналізу кожний факт розглядають як функцію від його характеристик. Сукупність цих характеристик може бути подана у вигляді моделі даних - багатовимірного куба. Осями багатовимірної системи координат слугують основні атрибути бізнес-процесу, що досліджується. Кожний з масивів даних може містити не просто перелік значень, а набір дерев, або ієрархію значень, де верхнє значення ієрархії розкривається даними, що знаходяться нижче.

За визначенням, запропонованим Б. Коддом, багатовимірне концептуальне представлення (multi-dimensional conceptual view) - це множинна перспектива, що складається з кількох незалежних вимірювань, уздовж яких можуть бути проаналізовані визначені сукупності даних. Одночасний аналіз за кількома вимірюваннями визначається як багатовимірний аналіз.



Кожне вимірювання може бути представлене у вигляді ієрархічної структури, а деякі вимірювання можуть мати декілька видів ієрархічного представлення. На перетинах осей вимірювань (Dimensions) розташовуються дані, що кількісно характеризують аналізовані факти, - міри (Measures).

У кожному кубі обов'язково присутня ієрархія часу. На верхньому рівні розташовані роки, потім - квартали, місяці, Дні.

Користувач може розрізати куб за різними напрямками і отримувати зведені відомості. Значення, щофіксуються вздовж вимірів, називаються членами виміри. Члени виміру використовуються як для розрізування куба, так і для фільтрації даних. Значення членів виміру відображаються у двовимірному представленні куба як заголовки рядків і стовпців.

Над гіперкубом можуть виконуватися такі операції:

***o зріз*** (Slice) - формується підмножина багатовимірного масиву даних. Якщо розглядати термін "зріз" з позиції кінцевого користувача, то найчастіше його роль виконує двовимірна проекція куба;

***o обертання*** (Rotate) - зміна розташування вимірювань, представлених у звіті або на сторінці, що відображається. Операція обертання може полягати в перестановці місцями рядків і стовпців таблиці або переміщенні вимірювань у стовпці чи рядків створюваного звіту, що дозволяє надавати йому бажаного вигляду. Крім того, обертанням куба даних є переміщення вимірювань, які не входять до складу таблиць, на місце вимірювань, представлених на сторінці, що відображається, і навпаки (вимірювання, що не є табличним, стає новим вимірюванням рядка чи вимірюванням стовпця);

***o консолідація*** (Drill Up) і ***деталізація*** (Drill Down) - операції, які визначають перехід вгору за напрямом від детального (down) представлення даних до агрегованого (up) і навпаки, відповідно. Напрям деталізації (узагальнення) може бути задано як за ієрархією окремих вимірювань, так і згідно з іншими відношеннями, встановленими у межах вимірювань.

Є 12 правил, що визначають OLAP, згідно із концепцією Б. Кодда.

1. ***Багато вимірність*** - OLAP-система на концептуальному рівні має представляти дані у вигляді багатовимірної моделі, що спрощує процеси аналізу і сприйняття інформації.

2. ***Прозорість*** - це спосіб організації даних, джерела, засоби обробки і зберігання.

3. ***Доступність*** - OLAP-система має надавати користувачу єдину, узгоджену і цілісну модель даних, забезпечуючи доступ до даних незалежно від того, як і де вони зберігаються.

4. ***Постійна продуктивність при розробленні звітів -*** продуктивність OLAP-систем не має дуже зменшуватися при збільшені кількості вимірювань, за якими виконується аналіз.

5. ***Клієнт-серверна архітектура*** - OLAP-система має бути здатна функціонувати у клієнт-серверному середовищі, оскільки більшість даних, які потрібно обробляти, зберігається децентралізовано. Серверний компонент інструменту OLAP має бути достатньо інтелектуальним і дозволяти будувати загальну концептуальну схему на основі узагальнення й консолідації різних логічних і фізичних схем корпоративних БД для забезпечення ефекту прозорості.

6. ***Рівноправність вимірювань*** - OLAP-система має підтримувати багатовимірну модель, в якій усі вимірювання рівноправні. За необхідності додаткові характеристики можуть бути надані окремим вимірюванням, але така можливість має бути надана будь-якому вимірюванню.

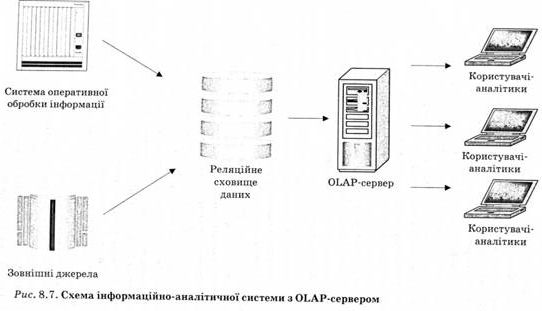
7. ***Динамічне управління розрідженими матрицями -*** OLAP-система має забезпечувати оптимальну обробку розріджених матриць. Швидкість доступу повинна зберігатися незалежно від розташування осередків даних і бути постійною величиною для моделей, що мають різну кількість вимірювань і різний ступінь розрідженості даних.

8. ***Підтримка розподіленого режиму доступу*** - OLAP-система має надавати можливість працювати кільком користувачам спільно з однією аналітичною моделлю або створювати для них різні моделі з єдиних даних. При цьому можливі усі операції.

9. ***Підтримка перехресних операцій*** - OLAP-система мас забезпечувати збереження функціональних відношень, описаних за допомогою певної формальної мови між осередками гі-перкуба при виконанні будь-яких операцій зрізу, обертання, консолідації або деталізації. Система має автоматично виконувати перетворення встановлених відношень, не вимагаючи від користувача їх перевизначення.

10. ***Інтуїтивна маніпуляція даними*** - OLAP-система має надавати спосіб виконання операцій зрізу, обертання, консолідації і деталізації над гіперкубом без необхідності для користувачів здійснювати дії з інтерфейсом. Вимірювання, визначені в аналітичній моделі, мають містити всю необхідну інформацію для виконання операцій.

11. ***Гнучкі можливості отримання звітів*** - OLAP-система має підтримувати різні способи візуалізації даних, тобто звіти мають подаватися у будь-якому можливому вигляді.



подання даних, що завантажуються з реляційної або багатовимірної БД на ПК клієнта.

JOLAP - колективна OLAP-API-ініціатива, що базується на Java, призначена для створення й управління даними і мета-даними на серверах OLAP.

***MOLAP-cepeepu*** використовують для зберігання та управління даними багатовимірними БД. MOLAP використовує БД, що показує результуючі дані, спеціальний варіант процесора просторових БД. Дані зберігаються у вигляді впорядкованих багатовимірних масивів. Такі масиви поділяються на гіперку-би і по ліку би.

У гіперкубі всі комірки, що зберігаються в БД, мають однакову розмірність, тобто знаходяться у якнайбільшому базисі вимірювань.

У полікубі кожна комірка зберігається із власним набором вимірювань, і пов'язані з цим труднощі обробки перекладаються на внутрішні механізми системи.

Фізичні дані, подані у багатовимірному вигляді, зберігаються у двовимірних файлах. Куб представляється у вигляді однієї пласкої таблиці, в яку за рядками вписуються усі комбінації членів усіх вимірювань з відповідними їх значенням мір.

Переваги використання багатовимірних БД в OLAP-системах:

o пошук і вибірка даних здійснюються значно швидше, ніж при багатовимірному концептуальному підході на реляційну БД, оскільки багатовимірна база даних денормалізована і містить апріорі агреговані показники, забезпечуючи оптимі-зований доступ до комірок запитів і не вимагаючи додаткових перетворень при переході від пов'язаних таблиць до багатовимірної моделі;

o багатовимірні БД підтримують задачі включення в інформаційну модель різноманітних функцій.

Недоліки використання багатовимірних БД в OLAP-системах:

o за рахунок денормалізації і заздалегідь виконаної агрегації обсяг даних у багатовимірній БД відповідає за оцінюван-

12. ***Необмежена розмірність і кількість рівнів агрегації*** - одночасно може використовуватися близько 20 вимірювань.

У 1995 p. Е. Кодд до цих правил додав ще такі шість:

1. ***Пакетне здобуття даних*** - OLAP-система має ефективно забезпечувати доступ до внутрішніх і зовнішніх даних.

2. ***Підтримка всіх моделей OLAP-аналізу*** - OLAP-система має підтримувати всі чотири моделі аналізу даних, визначені Е. Коддом: категоріальну, тлумачну, умоглядну і стереотипну.

3. ***Обробка ненормалізованих даних*** - OLAP-система має бути інтегрована з ненормалізованими джерелами даних. Модифікації даних, виконані в середовищі OLAP, не мають призводити до змін даних, що зберігаються у зовнішніх системах.

4. ***Збереження результатів OLAP:*** зберігання їх окремо від початкових даних - OLAP-система, що функціонує в режимі читання-запису, після модифікації початкових даних має зберігати результати окремо. Забезпечується безпека початкових даних.

5. ***Виключення відсутніх значень*** - OLAP-система має виключати усі відсутні значення. Відсутні значення мають відрізнятися від нульових значень.

6. ***Обробка відсутніх значень*** - OLAP-система має ігнорувати всі відсутні значення без урахування їх джерела.

Крім того, ці правила поділяються на чотири групи: В, S, R і D. Основні особливості В включають такі правила:

o багатовимірне концептуальне представлення даних (правило 1);

o інтуїтивне маніпулювання даними (правило 10);

o доступність (пра вило 3) ;

o пакетне здобуття проти інтерпретації (правило 13);

o підтримка всіх моделей OLAP-аналізу (правило 14);

o архітектура клієнт - сервер (правило 5);

o прозорість (правило 2);

o багатокористувацький доступ (правило 8). Спеціальні особливості S:

o обробка ненормалізованих даних (правило 15);

o збереження результатів OLAP: зберігання їх окремо від початкових даних (правило 16);

o виключення відсутніх значень (правило 17);

o обробка відсутніх значень (правило 18). Особливості представлення звітів R:

o гнучкість формування звітів (правило 11);

o стандартна продуктивність звітів (правило 4);

o автоматичне налагодження фізичного рівня (правило 7). Управління вимірюваннями D:

o універсальність вимірювань (правило 6);

o необмежена кількість вимірювань і рівнів агрегації (правило 12);

o необмежені операції між розмінностями (правило 9).

Виробники OLAP-систем забезпечують швидкість виконання запитів у межах 1-5 сек. ***Архітектура OLAP систем*** включає два основні компоненти:

1. OLAP-сервер - забезпечує зберігання даних, виконання над ними необхідних операцій і формування багатовимірної моделі на концептуальному рівні. Нині OLAP-сервери об'єднують зі сховищами даних або вітринами даних.

2. OLAP-клієнт - надає користувачу інтерфейс до багатовимірної моделі даних, забезпечуючи його можливістю зручно маніпулювати даними для виконання задач аналізу.

OLAP-сервери приховують від кінцевого користувача спосіб реалізації багатовимірної моделі. Вони формують гіперкуб, з яким користувачі за допомогою OLAP-клієнта виконують усі необхідні маніпуляції, аналізуючи дані (рис. 8.7).

Усі компоненти системи, починаючи з джерел даних до автоматизованих робочих місць аналітиків, традиційно розподіляються в корпоративній мережі.

Виокремлюють і використовують три основні способи реалізації OLAP-сервера для реалізації багатовимірної моделі:

1) MOLAP - багатовимірні БД;

2) ROLAP - реляційні БД;

3) HOLAP - багатовимірні і реляційні БД.

У літературі також зустрічаються абревіатури DOLAP і JOLAP. DOLAP - настільний (desktop) OLAP. Проста у використанні OLAP-система, призначена для локального аналізу і ням Кодда у 2,5-100 разів меншому обсягу початкових деталізованих даних;

o інформаційний гіперкуб є дуже розрідженим, оскільки дані зберігаються у впорядкованому вигляді, невизначені значення вдається видалити тільки за рахунок вибору оптимального порядку сортування, який дає змогу організувати дані у великі безперервні групи. Доводиться знаходити компроміс між швидкодією і надмірністю дискового простору, що займається базою даних;

o багатовимірні БД чутливі до змін у багатовимірній моделі. При додаванні нового вимірювання доводиться змінювати структуру всієї БД, що призводить до великих витрат часу.

Швидкість виконання запитів у MOLAP визначається тільки правильністю побудови багатовимірної бази. Проте в цьому випадку багатовимірна база надмірна, оскільки копіює всі дані зі сховища.

На основі аналізу переваг і недоліків багатовимірних БД можна передбачити умови, за яких їх використання є ефективним:

o обсяг початкових даних для аналізу не більше, ніж кілька гігабайт - високий рівень агрегації даних;

o набір інформаційних вимірювань стабільний;

o час відповіді системи на нерегламентовані запити є найбільш критичним параметром;

o застосовують складні вбудовані функції для виконання кросмірних обчислень над комірками гіперкуба, у тому числі функцій користувача.

***ROLAP-cepвepu*** використовують реляційні БД. За висловлюванням Б. Кодда, реляційні БД є найбільш відповідною технологією для зберігання даних. ROLAP функціонує безпосередньо із реляційним сховищем, фактами і таблицями. Для зберігання агрегатів створюються додаткові реляційні таблиці. Нині поширені дві основні схеми реалізації багатовимірного подання даних за допомогою реляційних таблиць: "зірка" та "сніжинка". Основними складовими таких схем є денор-малізована таблиця фактів (Fact Table) і множина таблиць вимірювань (Dimension Tables).

Таблиця фактів, як правило, містить відомості про об'єкти або події, сукупність яких аналізуватиметься далі. Є чотири типи фактів, що найчастіше зустрічаються. Вони пов'язані:

o із транзакціями (Transaction facts) і засновані на окремих подіях;

o моментальними знімками (Snapshot facts) і ґрунтуються на стані об'єкта у певні моменти часу;

o елементами документа (Line-item facts) і засновані на документі й містять докладну інформацію про елементи цього документа;

o подіями або станом об'єкта (Event or state facts) і представляють виникнення події без деталей про неї.

Таблиця фактів, як правило, містить унікальний складений ключ, що об'єднує первинні ключі таблиць вимірювань. Як ключові, так і деякі неключові поля мають відповідати вимірюванням гіперкуба. Крім цього таблиця фактів містить одне або кілька числових полів, на підставі яких будуть одержані агрегатні дані.

Для багатовимірного аналізу придатні таблиці фактів, що містять якомога докладніші дані, які відповідають членам нижніх рівнів ієрархії відповідних вимірювань. У таблиці фактів немає ніяких відомостей про те, як групувати записи при обчисленні агрегатних даних.

Таблиці вимірювань містять незмінні або рідко змінювані дані. У більшості випадків цими даними є один запис для кожного члена нижнього рівня ієрархії у вимірюванні. Таблиці вимірювань також містять як мінімум одне описове поле і одне ключове поле з цілих чисел для однозначної ідентифікації члена вимірювання. Якщо вимірювання відповідної таблиці містить ієрархію, то така таблиця також може містити поля, вказуючи на "батька" даного члена в цій ієрархії. Кожна таблиця вимірювань має знаходитися відносно "один - до - багатьох" з таблицею фактів.

Швидкість збільшення таблиць вимірювань має бути незначною порівняно зі швидкістю зростання таблиці фактів.

У складних задачах з ієрархічними вимірюваннями використовують розширену схему "сніжинка" (Snowfiake Schйma). У цих випадках окремі таблиці фактів створюються для можливих поєднань рівнів узагальнення різних вимірювань. Це дає змогу підвищувати продуктивність, але часто призводить до надмірності даних і значних ускладнень у структурі бази даних.

Збільшення кількості таблиць фактів у базі даних визначається не тільки множинністю рівнів різних вимірювань, а й тією обставиною, що загалом факти мають різну множину вимірювань. При абстрагуванні від окремих вимірювань користувач має одержувати проекцію максимально повного гіперкуба, причому не завжди значення показників у ній повинні бути результатом елементарного підсумовування. Отже, за великої кількості незалежних вимірювань необхідно підтримувати множину таблиць фактів, що відповідають кожному можливому поєднанню вибраних у запиті вимірювань, що також призводить до неекономного використання зовнішньої пам'яті, збільшення часу завантаження даних у БД схеми "зірки" із зовнішніх джерел і труднощів адміністрування.

Переваги використання реляційних БД в OLAP-системах:

o корпоративні сховища даних реалізуються засобами реляційних СУБД, і інструменти ROLAP дають змогу здійснювати аналіз безпосередньо над ними;

o у разі змінної розмірності задачі ROLAP-системи з динамічним представленням розмірності є оптимальним рішенням, оскільки в них такі модифікації не вимагають фізичної реорганізації БД;

o реляційні СУБД забезпечують значно вищий рівень захисту даних і високі можливості розмежування прав доступу.

Недоліки використання реляційних БД в OLAP-системах.

Головний недолік ROLAP порівняно з багатовимірними СУБД - менша продуктивність.

Для забезпечення продуктивності, порівнянної з MOLAP, реляційні системи вимагають ретельного опрацьовування схеми бази даних і настройки індексів, тобто зусиль з боку адміністраторів БД. Тільки за умови використання схем типу "зірка" продуктивність добре набудованих реляційних систем може наблизитися до продуктивності систем на основі багатовимірних баз даних.

Особливим випадком ROLAP є ROLAP реального часу (Realtime ROLAP). У R-ROLAP для зберігання агрегатів не створюються додаткові реляційні таблиці, а агрегати розраховуються у момент здійснення запиту. Багатовимірний запит до OLAP-системи автоматично перетворюється в SQL-запит до реляційних даних.

***HOL АР-сервери,*** використовують гібридну архітектуру, що об'єднує технології ROLAP і MOLAP. На відміну від MOLAP, що працює краще, коли дані є щільними, сервери ROLAP показують кращі параметри у тих випадках, коли дані розріджені. Сервери HOLAP застосовують підхід ROLAP для розріджених областей багатовимірного простору і підхід MOLAP - для щільних областей. Сервери HOL АР поділяють запит на кілька підзапитів, направляють їх до відповідних фрагментів даних, комбінують результати, а потім надають результат користувачу.

Отже, HOLAP використовує реляційні таблиці для зберігання базових даних і багатовимірні таблиці для агрегатів. На практиці саме гібридна модель застосовується найчастіше, оскільки за рахунок розподілу навантаження на багатовимірний і реляційний сервери вдається досягти оптимальної продуктивності при виконанні аналітичних запитів.

MOLAP найкраще підходить для невеликих наборів даних, він швидко розраховує агрегати і повертає відповіді, але при цьому генеруються величезні обсяги даних. ROLAP оцінюється як більш масштабоване рішення, що використовує невеликий простір. Швидкість обробки значно знижується.

DOLAP-застосунки також успішно розвиваються. Сюди можна віднести, наприклад, продукти від Brio Software і Business Objects. Настільні системи призначені для аналітичної обробки невеликих обсягів даних без виконання складних аналітичних досліджень. Функціональність настільних застосувань, як правило, обмежується можливостями візуального представлення даних, обертання і виконання зрізів куба.

**3. Тест FASMI**

Зазначені особливості Кодд поклав в основу тесту FASMI, як основного критерію для визначення, чи належить СУБД до OLAP-додатків.

Це означає, що OLAP-додатки повинні забезпечувати «швидкий аналіз розподіленої багатовимірної інформації» (*Fast Analysis of Shared Multidimen-sional Information*).

*Fast* — OLAP-додаток повинен забезпечувати швидкість видачі відповіді на запит у межах 5 с.

*Analysis* — OLAP-додаток повинен виконувати будь-який статичний чи логічний аналіз.

*Shared* — OLAP-додаток повинен виконувати усі вимоги до захисту конфіденційності до рівня комірки даних.

*Multidimensional* — OLAP-додаток повинен забезпечувати багатовимірне концептуальне представлення даних.

*Information* — OLAP-додаток повинен забезпечувати отримання необхідної інформації в умовах реального процесу [2].

Ринок OLAP-додатків на сьогодні є досить розвиненим. Тому розробників цих продуктів досить багато: IBM, Microsoft, Oracle, Sybase, Hyperion Solutions, Informix, Cognos, Cristal Decisions, Cartesis, Knosys та ін. Враховуючи кількість розробок на ринку, додатки було обмежено відомими та надійними розробниками: Microsoft, Sybase, IBM, Oracle, Informix (табл. 1). Ці компанії внесли великий внесок у розвиток технологій СУБД та мають OLAP-засоби, що широко використовуються.

*Таблиця 1*

3

|  |  |
| --- | --- |
| **Досліджувані OLAP-засоби** Компанія-розробник | OLAP-додаток |
| Microsoft | OLAP Services |
| Sybase | Adaptive Server IQ Multiplex |
| IBM | DB2 OLAP Server |
| Oracle | Express |
| Informix | MetaCube |

**4. Критерії порівняння OLAP-додатків**

1. *Підтримка багатокористувацького режиму роботи з OLAP-сервером*.

OLAP-додатки мають досить високу вартість, тому використання їх одним користувачем буде нераціональним.

2. *Можливість одночасної обробки декількох кубів даних*.

Куби даних можуть знаходитися на різних OLAP-серверах. Можливість одночасної обробки декількох кубів дозволяє працювати з даними, що належать до різних областей аналізу.

3. *Тимчасове зберігання багатовимірних даних на клієнтських комп’ютерах*.

Процес дає змогу виконувати аналіз даних без з’єднання з OLAP-сервером. Завдяки тимчасовому зберіганню даних на клієнтській машині зменшується обсяг передачі даних через інтернет-з’єднання та забезпечується автономія клієнтських вузлів.

4. *Забезпечення інформаційної безпеки на рівні записів*.

Відомо, що рівень безпеки в множинах записів у таблиці даних, тобто комірок, є більш надійним, ніж у множинах таблиць.

5. *Можливість попереднього обчислення агрегатів*.

Швидкодія OLAP-додатків, у першу чергу, визначається схемою створених агрегатів. Враховуючи те, що OLAP-продукти повинні працювати з максимальною швидкодією та продуктивністю, забезпечуючи обробку даних у режимі реального часу, то можливість попереднього обчислення агрегатів має ключове значення під час вибору системи.

6. *Оптимізація схеми агрегування*.

Збільшення кількості агрегатів сприяє підвищенню швидкодії системи та скорочення часу реакції на запит. Але збільшення чисельності агрегатів призводить до збільшення обсягу бази даних у пам’яті комп’ютера. Тому при роботі з OLAP-додатком необхідно залишати лише найчастіше використовувані агрегати. Це оптимізує схему агрегування [3].

7. *Розв’язання проблеми так званого вибухового зростання обсягу даних*.

Дана проблема виникає під час конструювання схеми агрегування даних. Якщо кількість агрегатів починає перевищувати кількість точок даних, то це призводить до катастрофічного збільшення обсягу даних, що зберігаються. Деякі OLAP-додатки вирішують цю проблему не дозволяючи базі даних розширюватися в обсязі [4].

8. *Можливість представлення даних у формі MOLAP і HOLAP*.

Більшість з OLAP-додатків сьогодення передбачують представлення даних тільки в реляційній формі — ROLAP (Relational OLAP) або реля-ційні OLAP.

Що визначає такий недолік. ROLAP-додатки мають більш низьку швидкість роботи, ніж MOLAP (Multidimensional OLAP) або багатовимірні OLAP.

До переваг ROLAP-забезпечення можна віднести можливість масштабування, що обмежується лише апаратними засобами. HOLAP (Hybrid OLAP) або гібридні OLAP мають переваги обох попередніх систем: краща масштабованість та продуктивність [5].

9. *Підтримка доступу до нереляційних джерел даних*.

OLAP-додаток може забезпечувати доступ не тільки до СУБД за допомогою ODBC (*Open Database Connectivity*, відкритий <інтерфейс> зв’язку з базою даних), але й до інших джерел даних. Ними можуть бути, наприклад, електронна пошта, текстові файли, файлові системи та ін.

Прикладом цього протоколу є OLEDB (*Object Linking and Embedding*, *Database*) зв’язування, впро-вадження об’єктів бази даних, що дає змогу здійснювати доступ до нереляційних джерел даних.

10. *Кросплатформеність*.

OLAP-продукт не повинен бути прив’язаним до тієї чи іншої операційної системи (ОС), адже цим вона обмежує адміністраторів у виборі ОС. Тому важливим є те, щоб OLAP-додаток був здатен працювати з різними ОС.

11. *Можливість створення обчислюваних позначок*.

Обчислювані позначки дозволяють додати до багатовимірної бази даних міри та виміри, що не існували в початковій базі даних.

12. *Масштабованість*.

Масштабованість дає можливість OLAP-додатку працювати з обсягами даних, що змінюються в досить широких межах.