

УДК 004.032.2:378(045)

МОДЕЛЬ ОБРОБКИ РЕЗУЛЬТАТІВ ЕКЗАМЕНАЦІЙНИХ СЕССІЙ ВИЩИХ НАВЧАЛЬНИХ ЗАКЛАДІВ ЗАСОБАМИ OLAP

Труш О. І., канд. техн. наук, доц., Чабан Т. О.

Національний авіаційний університет

tochaban@gmail.com

У статті проведено порівняльний аналіз зручності використання системи аналітичної обробки даних у режимі реального часу з метою зберігання інформації про успішність студентів університету та аналізу даних залежно від об'єктивних чинників.

Ключові слова: OLAP-додаток, тест FASMI, критерії порівняння, база даних, університет.

The article is dedicated to a comparative analysis of the OLAP's (Online Analytical Processing) usability to store information about the result of university students and analyze data based on objective factors

Keywords: OLAP, test FASMI, criterion of comparison, data base, university.

Постановка проблеми

Сучасна система освіти у вищих навчальних закладах потребує автоматизації управління навчальним процесом з використанням новітніх технологій. Зокрема більш гнучкої та багатофункціональної системи обробки даних для автоматизації учбового процесу, ніж використовуємо зараз СУБД Access. Таким програмним інструментарієм можна вважати систему аналітичної обробки даних у режимі реального часу — *Online Analytical Processing (OLAP)*.

Враховуючи те, що за наказом Міністерства освіти та науки України № 1/9-817 від 27.11.09 уся статистична звітність ВНЗ повинна складатися з використанням СУБД Access, то повний перехід ведення баз даних університету на засоби OLAP не може бути виконаним. Але використовувати системи OLAP як додатковий інструмент аналізу даних, на думку авторів, було б доцільним. Адже виявлення тенденцій в успішності студентів дозволяє встановити вирішальні в освітянському процесі чинники та навчитися впливати на процес навчання. Окрім того, OLAP-додатки дають змогу обробляти дані в режимі реального часу, що іноді є необхідним при роботі з БД університету.

Новина роботи полягає в тому, що порівняння OLAP-продуктів у повному обсязі раніше майже не проводилося. Окрім того, багатовимірні моделі даних для аналізу освітянського процесу в Україні практично не застосовувалися.

Аналіз останніх досліджень і публікацій

Сьогодні OLAP-додатки є досить поширеними під час аналізу діяльності підприємств. Існує досить багато літератури, що висвітлює користування та проектування OLAP-додатків. Але треба зауважити, що аналізу програмних продуктів ринку OLAP у цих виданнях не приділялось значної уваги, зокрема, використанню OLAP-забезпечення при аналізі освітянського процесу.

Постановка завдання

Метою даного проведення дослідження було порівняння OLAP-продуктів задля оптимального вибору програмного середовища для створення системи керування даними та їх оперативного аналізу в режимі реального часу в навчальному процесі університету. Адже керівництво навчального закладу повинно приймати безліч рішень стосовно оптимальної діяльності закладу. Окрім того, приймати рішення стосовно того, як організувати навчальний процес та які методи викладання застосовувати викладацькому складу. Чим більш обґрунтовані рішення будуть прийняті, тим більших успіхів у викладанні досягне навчальний заклад.

Для багатьох співробітників університету, що беруть участь у прийнятті рішень закладу, здатність швидше і ефективніше аналізувати навчальні процеси означає прийняття оптимальних рішень задля досягнення більш ефективної системи навчання студентів.

У наведених випадках OLAP-додатки дозволяють виконувати швидкий і ефективний аналіз великих обсягів даних. Результати зберігаються в багатовимірному вигляді (наприклад, у кубах), що точно відображає природний стан реальних даних навчального процесу.

Виклад основного матеріалу дослідження

Передусім було визначено основні властивості системи, що дозволяють вважати конкретний продукт OLAP-додатком. У 1993 р. британський розробник теорії реляційних баз даних С. Ф. Кодд опублікував статтю «OLAP для користувачів-аналітиків: інформаційно-технологічний мандат» [1], де сформулював 12 основоположних функціональних принципів OLAP.

У 1995 р. до них було додано ще шість. Доктор Кодд розбив на чотири групи ці правила, назвавши їх особливостями.

Основні функціональні особливості

F1. Багатовимірне концептуальне представлення даних.

F2. Інтуїтивне маніпулювання даними. Інструмент повинен пропонувати вибір найбільш зручного режиму роботи з даними.

F3. Доступність (OLAP виступає як посередник).

У цьому правилі Кодд особливо підкреслює роль OLAP як прошарку між гетерогенними джерелами даних і їх поданням для кінцевого користувача.

F4. Пакетний доступ проти інтерпретації.

Це правило потребує, щоб продукт у рівній мірі ефективно забезпечував доступ як до власного сховища даних, так і до зовнішніх даних.

F5. Моделі аналізу OLAP.

OLAP продукти повинні підтримували чотири моделі аналізу: категоріальний, тлумачний, умовляндний та стереотипний.

F6. Архітектура «клієнт-сервер».

Продукт має бути не тільки клієнт-серверним, але й щоб серверний компонент був би досить інтелектуальним для того, щоб різні клієнти могли підключатися з мінімумом зусиль. Це означає, що OLAP доступний з робочого столу користувача.

F7. Прозорість.

Повна відповідність цій вимозі означає, що користувач електронної таблиці здатний отримати всі необхідні дані з OLAP-машини, навіть якщо невідомо звідки вони отримуються.

F8. Забезпечення багатоклієнтської підтримки.

Інструменти OLAP повинні забезпечувати односторонній доступ (зчитування і запис), інтеграцію та конфіденційність.

Спеціалізовані особливості

F9. Обробка ненормалізованих даних.

Необхідність інтеграції між OLAP-сервером і ненормалізованими джерелами даних.

F10. Збереження результатів OLAP за принципом розміщення їх у пам'яті окремо від вихідних даних. OLAP-програми, працюючі в режимі читання-запису, повинні не впливати безпосередньо на оброблювані дані, і дані, модифіковані в OLAP.

F11. Виключення відсутніх значень.

Відсутні значення повинні відрізнятися від нульових значень.

F12. Обробка відсутніх значень. Усі відсутні значення повинні ігноруватися OLAP-аналізатором без урахування їхнього джерела.

Особливості подання звітності

F13. Гнучкість формування звітів, тобто вимірювання повинні бути розміщені у звіті так, як це потрібно користувачу.

F14: Стандартна продуктивність звітів.

Продуктивність формування звітів не повинна істотно падати зі зростанням кількості вимірювань і розмірів бази даних.

F15. Автоматичне налаштування фізичного рівня.

OLAP-додатки повинні автоматично налаштувати свою фізичну схему залежно від типу моделі, обсягів даних і розрідженості бази даних.

F16. Універсальність вимірювань.

Усі вимірювання мають бути рівноправні, кожний вимір — еквівалентним і в структурі, і в операційних можливостях.

F17. Необмежена кількість вимірювань і рівнів агрегації.

Технічно немає продукту, який міг би відповідати цій вимозі, тому що немає такого режиму, як необмежений об'єкт на обмеженому комп'ютері, а в разі ухвалення деякого максимуму, він повинен забезпечувати принаймні два десятки вимірювань.

F18. Необмежені операції між просторами.

Всі види операцій повинні бути дозволені для будь-яких вимірів, а не тільки для вимірювань типу «показник» — міра.

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 Multidimensional 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].

У практиці застосування OLAP-додатка в організації навчального процесу дає можливість оцінити вплив як об'єктивних, так і суб'єктивних факторів на успішність студентів.

Наприклад, зі збережених в OLAP-додатку даних можна побудувати куб (див. рисунок).

У результаті аналізу цього куба можна оцінити залежність успішності студентів від пори року (зимова та літня сесії).

Як видно з рисунку, студенти мають кращу успішність за результатами зимової сесії.

Тому буде доцільним при складанні навчального плану поставити деякі «важкі» дисципліни на 1 семестр навчального року.

	2009–2010	2008–2009	2007–2008
	Зимова сесія	Літня сесія	
Група 401 ФКН	4,7	4,4	
Група 402 ФКН	4,3	4,1	
Група 403 ФКН	4,6	4,7	

Куб даних успішності студентів за результатами зимової та літньої сесій

Завдяки такому підходу можна оцінити вплив на успішність студентів безлічі чинників: пори року, статі студента, віку студента та викладача, середнього балу атестату про загальну середню освіту та ін.

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

Досліджувані OLAP-засоби

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

Критерії порівняння OLAP-додатків

Враховуючи специфіку освітньої системи в Україні в умовах фінансової та технічної обмеженості, основними критеріями під час порівняння OLAP-засобів були надійність і дешевизна продукту. Крім того, продукт не повинен вимагати великої кількості системних ресурсів. Розглянемо перелік обраних для порівняння критеріїв.

1. Підтримка багатокористувацького режиму роботи з OLAP-сервером.

Система університету, як і будь-якого навчального закладу, передбачає деяку кількість персоналу, що відповідає за прийняття рішень та керування організацією. Крім того, аналіз ефективності навчання студентів є необхідним і для викладачів, щоб вони могли приймати обґрунтовані рішення. Тому реалізація багатокористувацького режиму в OLAP-додатку є одним з ключових моментів при виборі програмного забезпечення. Окрім того, 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. Наприклад, якщо в одному кубі знаходяться дані про викладача, а в іншому — про успішність групи, то можна провести аналіз впливу навіть віку викладача на успішність студентів.

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.

База даних університету чи іншого навчального закладу має конфіденційний характер і не повинна бути доступна неавторизованим користувачам. 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 швидкістю 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 реляційні 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, впровадження 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. За визначеним переліком критеріїв результати оцінювання наведено в табл. 2.

Програмні продукти було оцінено за такою шкалою: якщо OLAP-додаток підтримує визначену функцію — 1 бал, якщо ні — 0 балів.

Таблиця 2

Порівняння OLAP-додатків

Критерій\продукт	Microsoft OLAP Services	Sybase Adaptive Server IQ Multiplex	IBM DB2 OLAP Server	Oracle Express	Informix MetaCube
Підтримка багатокористувацького режиму роботи з OLAP-сервером	1	1	1	1	1
Можливість одночасної обробки кількох кубів даних	1	0	0	1	0
Тимчасове зберігання багатовимірних даних на клієнтських комп'ютерах	1	0	0	0	0
Забезпечення інформаційної безпеки на рівні записів	0	0	0	1	1

Закінчення табл. 2

Критерій\продукт	Microsoft OLAP Services	Sybase Adaptive Server IQ Multiplex	IBM DB2 OLAP Server	Oracle Express	Informix MetaCube
Можливість попереднього обчислення агрегатів	1	0	1	0	1
Оптимізація схеми агрегування	1	0	0	0	1
Розв'язання проблеми «вибуху даних»	1	1	1	1	1
Можливість представлення даних у формі MOLAP і HOLAP	1	0	0	1	0
Підтримка доступу до нереляційних джерел даних	1	0	0	0	0
Кросплатформеність	0	1	1	1	1
Можливість створення обчислюваних позначок	1	1	1	1	1
Масштабованість	1	1	1	1	1
Сума балів	10	5	6	8	9

Таким чином, як видно з табл. 2, найбільш підходящим OLAP-додатком виявилось програмне забезпечення від фірми Microsoft — OLAP Services. OLAP Services здобув найбільшу кількість балів.

Головною перевагою OLAP-дodatка від Microsoft є найнижча на ринку ціна. Оскільки цей продукт входить до пакету прикладних програм MS SQL Server.

Отже, не потрібно проводити закупівлю додаткового програмного забезпечення для підтримки функціоналу OLAP-дodatка.

Інші продукти мають завищену вартість, враховуючи їх функціональну обмеженість.

Користування OLAP-дodatками Informix, IBM, Sybase чи Oracle може бути обгрунтованим у випадку, якщо IT-інфраструктура побудована на продуктах від того ж постачальника.

Висновки

1. Запропоновано модель зберігання, обробки та керування даними навчального процесу за допомогою OLAP-дodatка.

2. Наведено приклад реалізації даної моделі у вищих навчальних закладах.

3. Визначено основні параметри, котрі повинен задовольняти OLAP-дodatок, з огляду на специфіку його використання в університеті.

4. Проведено порівняльний аналіз існуючих OLAP-дodatків на предмет їх відповідності вимогам та специфіці університету.

5. Визначено OLAP-дodatок, який у найбільш повному обсязі задовольняє потреби вищого навчального закладу при організації системи керування даними навчального процесу.

ЛІТЕРАТУРА

1. Баргенян А. Технология анализа данных: Data Mining, Visual Mining, Text Mining, OLAP /А. Баргенян, М. Куприянов, В. Степаненко, И. Холод. — СПб. : БХВ-Петербург, 2007. — С. 48—49.

2. Хрусталёв Е. Агрегация данных в OLAP-кубах- <http://www.thalion.kiev.ua/idx.php/4/305/article/>

3. Codd E. «Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate», E. Codd, S. B. Codd, C. T. Salley. — E. F. Codd & Associates, 1993).

4. Потгитер Й. Масштабируемость OLAP-данных. Режим доступа: <http://citcity.ru/11158/>

5. Чубикова И. Курс лекций по Data Mining / И. А. Чубикова. Интернет-университет INTUIT: 2006. — С. 99—100.

Стаття надійшла до редакції 28.04.2011.