

Тема 6 СППР на основі сховищ даних та OLAP-технологій

6.1 Розвиток та застосування СППР на основі сховищ даних та OLAP-систем

6.2 Зародження й розвиток OLAP-систем. 12 правил Кода

6.1 Розвиток та застосування СППР на основі сховищ даних та OLAP-систем

Системи підтримки прийняття рішень на основі сховищ даних та OLAP-систем, як і самі сховища даних (Data Warehouses) та системи аналітичного онлайнового оброблення даних, належать до типу орієнтованих на дані СППР. У загальному вигляді орієнтовану на дані систему підтримки прийняття рішень (ОДСППР) можна визначити як інтерактивну комп'ютеризовану систему, що допомагає ОПР використовувати дуже велику базу даних із внутрішніх даних компанії та деякі зовнішні дані з навколишнього середовища системи з метою прийняття обґрунтованих рішень. Наприклад, система може надавати дані щодо збуту продукції як самої компанії, так і її конкурентів. Деякі дані можуть бути деталізованими даними транзакцій, а деякі - агрегованими. У більшості реалізованих нині ОДСППР користувачі можуть виконувати незаплановані або в режимі на даний випадок (ad hoc) аналізи даних та формулювати запити. За допомогою таких систем менеджери обробляють дані для ідентифікації фактів та отримання висновків у вигляді графічних зображень (діаграм, графіків, трендів).

Орієнтовані на дані СППР, зокрема системи аналітичного онлайнового оброблення, інколи називають бізнес-інформаційними (*Business intelligence*). Цей широко відомий термін запропонував 1989 року аналітик Говард Дрезнер (Howard Dresner) з групи «Gartner», що консультує фірми «Fortune 1000» з питань інформаційних технологій, для описання низки понять та методів, призначених для вдосконалення бізнесових рішень за допомогою

використання оснований на фактах систем підтримки прийняття рішень (власне замість терміна «програмне забезпечення» («decision support») був запропонований термін «Business intelligence»).

Зокрема, у 1994 році Говард Дрезнер писав: «До 1996 року використання розв'язків Business intelligence різко переміститься від спеціалізованих аналітиків до всіх менеджерів та професіоналів як найкращий шлях розуміння бізнесу... Замість вузького кола аналітиків, які витрачають 100 відсотків свого часу на аналізування даних, усі менеджери та професіонали витратять 10 відсотків свого часу для цього, використовуючи програмне забезпечення Business intelligence» [13]. Хоча ні 1996, ні навіть 2001 року повною мірою цей прогноз не справдився, проте багато менеджерів освоїли й продовжують вивчати цей тип програмних продуктів. Така обставина пояснюється тим, що програмне забезпечення орієнтованих на дані СППР надає змогу користувачам досягати як загальної для всієї корпорації мети, тобто підвищення рівня конкурентоспроможності за рахунок своєчасного отримання важливої для успіху інформації, недосяжної раніше іншими шляхами, так і поліпшувати особисту управлінську продуктивність, тобто різко скорочувати завдяки таким СППР час на отримання таких самих результатів.

Термін «Business intelligence» використовується як синонім «OLAP» в інструкціях щодо використання, в описаннях виконавчих інформаційних систем. «Business intelligence» є терміном з маркетингу. Він вживається також для позначення орієнтованих на дані СППР. Цей тип СППР допомагає користувачам реалізувати практичне оброблення вниз (drill down) для детальнішого огляду інформації, оброблення вгору (drill up), щоб оглянути ширшу, інтегровану інформацію, «розрізати базу даних тонкими скибочками й нарізувати кубиками» («slice and dice»), щоб змінити вимірність, яку вони переглядають. Результати «drilling» і «slicing та dicing» подаються в таблицях та схемах [1]. Чому саме такі та подібні їм можливості стали конче потрібними ОПР та менеджерам?

Традиційна технологія підготовки інтегрованої інформації на основі запитів та звітів у звичайних системах підтримки прийняття рішень стала неефективною через різке зростання кількості й збільшення різноманітності вихідних даних, що знаходяться в численних оперативних і виробничих системах організацій. Ця обставина стала гальмувати виконання функцій управління у разі необхідності швидко створювати й приймати рішення. Крім того, постійне накопичення даних у корпоративній базі даних для прийняття рішень й подальший їх аналіз гальмують оперативну роботу з даними.

У типовій організації доступні оперативні бази даних розроблялися для здійснення регулярних процедур (транзакцій). Такі потреби можуть включати, наприклад, підготовку нового або поновлення попереднього замовлення. Регулярні процедури рідко пов'язані з запитам, тому, зазвичай, оперативні бази даних містять тільки поточну інформацію. Вони не можуть задовольнити інформаційні потреби користувачів СППР, тому що їм не вистачає архівних даних та їх стабільності, що стає вирішальним у разі необхідного аналізування. Наприклад, ОПР може звернутися до системи з запитанням: «Як проведені компанією останні заходи щодо розповсюдження товару впливали на його продаж за останній квартал у порівнянні з таким самим просуванням товару за останній рік?» Швидкої та вичерпної відповіді на подібні запитання традиційні СППР забезпечити не могли.

Крім того, часто дані знаходяться в різних операційних системах, які мають суттєві відмінності в організації даних. Фактично поєднання даних з систем DB2, Oracle та COBOL з даними бази даних Sybase чи Informix може бути нелегким та ускладнюючим аналіз цих даних. Оперативні бази даних за структурою та розміром є недостатніми для загального аналізу. Як результат з цього випливає, що коли потреби оперативних баз даних повністю відрізняються від потреб СППР, то вони не є оптимальними для використання в СППР, що призводить до неефективного виконання запитів до СППР.

В історичному контексті це означає, що СППР не були такими корисними, як могли б бути, оскільки необхідні дані не були вчасно доступними. Для компенсації цього недоліку часто адаптовували окремі «заморожені» фрагменти виходів інформаційної системи для наступного аналізу. Ці фрагменти забезпечують лише інформацією про відібрані об'єкти в окремі моменти часу. Хоча такі фрагменти ефективніші, ніж безпосереднє використання оперативної бази даних, але їм недостає широкого спектру інформації для повного її аналізу або гнучких специфічних запитів. Інакше кажучи, враховуючи всі переваги такого вирішення проблеми, воно загалом не створює сприятливого середовища для використання СППР.

Альтернативним та популярнішим підходом є створення *сховища даних* для підтримки потреб СППР, а також застосування інструментальних засобів *OLAP-систем*, які забезпечують доступ та оброблення накопичених за достатній період внутрішніх даних організації, а також у деяких випадках і зовнішніх.

У загальному випадку дані ОДСППР є операційними даними щодо фіксації щоденних бізнесових операцій компанії, тобто дані про транзакції (ведення справ, угод) та бізнесові випадки (явища). Вони формуються, щоб забезпечити тактичні й стратегічні бізнесові дії на основі операційних та доречних зовнішніх даних. Відмінність у цілях застосування операційних баз даних та баз даних СППР полягає в тому, що формати й структури цих видів даних відрізняються принаймні за шістьма основними показниками: *структурою даних, діапазоном часу, підсумовуванням даних, мінливістю даних, вимірністю даних та метаданими*. Зупинимось на них докладніше.

Структури даних. Розглянемо діапазон і суть відмінностей операційних даних та даних ОДСППР за форматом та структурою. Почнемо з операційних даних, які часто упорядковуються системою керування реляційної бази даних. Реляційні системи оброблення запитів (транзакційні системи) мають структури даних, які називаються таблицями, що повинні бути надзвичайно нормалізованими. Таблиці нормалізують, щоб уникнути аномалій у даних,

коли виконуються, наприклад, такі операції як оновлення, додавання чи вилучення записів. Нормалізація є процесом переведення складної структури даних у найпростішу, найстійкішу структуру. За нормалізації вилучаються зайві атрибути, ключі й відношення у концептуальній моделі даних [5].

За зберігання операційних даних або даних транзакційних систем програмне забезпечення й апаратні засоби оптимізовані так, щоб підтримувати транзакції стосовно щоденних операцій компанії. Наприклад, кожного разу, коли продається виріб, то все, що з цим актом пов'язане, мусить бути записане й враховане, тобто обчислене у відповідній таблиці транзакцій. Також пов'язані з цією операцією дані - дані щодо замовників та запасів матеріалів, змінюються в системах оброблення транзакцій. Для того, щоб забезпечити ефективну актуалізацію БД, системи транзакцій зберігають дані в багатьох малих таблицях, кожна з яких має мінімальну кількість полів. Так, наприклад, простій транзакції з оброблення запиту щодо збуту продукції потрібно мати дані, елементи яких записуються в п'яти або більше різних таблицях, оскільки потрібно добавляти або змінювати запис у таблиці накладної, в таблиці рядків накладної, таблиці дисконтів, таблиці запасів та таблиці департаменту.

Хоч такий структурований підхід до створення багатьох малих таблиць розглядається як ефективний для бази даних транзакцій, однак він не призначений для організації даних в ОДСППР. За такого підходу запити будуть виконуватися повільно, оскільки потрібно з'єднати багато таблиць, щоб завершити опрацювання запиту, на що витрачається багато часу та використовуються обшивні ресурси системи.

Операційні дані, зазвичай, зберігаються в багатьох різних таблицях і містять інформацію про специфічні транзакції, а дані ОДСППР - у значно меншій кількості таблиць, причому в них не завжди можна відшукати детальні відомості щодо кожної транзакції, бо вони є переважно підсумковими даними транзакцій. Дані з численних операційних баз даних інтегровані,

агреговані й підсумовані в базі даних ОДСПР, щоб задовольняти наперед невизначені потреби щодо підтримки прийняття рішень.

Орієнтовані на дані СПР можуть містити надлишкові дані, якщо це сприяє прискоренню оброблення запитів. Компонентами даних ОДСПР на основі сховища даних є: метадані, поточні деталізовані дані, давніші докладні дані, підсумовані дані тощо. Загальна нормалізація не ефективна для даних ОДСПР і навіть деяка часткова нормалізація може реально зменшити ефективність оброблення запитів у орієнтованих на дані СПР.

Діапазон часу. Операційні дані є поточними даними, оскільки вони відображають теперішній стан бізнесових транзакцій. Дані ОДСПР - це миттєві знімки станів у задані моменти часу, тобто вони являють собою упорядковану за часом сукупність або серію операційних даних. Фактично в ОДСПР зберігаються численні «часові зрізи» операційних даних.

Підсумовування. Дані ОДСПР можуть підсумовуватися за допомогою програмного забезпечення аналітичного оброблення. Можна відправити деякі дані з бази даних ОДСПР у мультимірний куб даних для прискореного аналізу. Деякі бази даних ОДСПР утворюються виключно з підсумків або (як їх часто називають) похідних чи вторинних даних. Наприклад, скоріше, ніж зберігати кожен з 10 000 транзакцій зі збуту в окремих елементах пам'яті на даний період, база даних ОДСПР може містити загальну кількість проданих одиниць та обсяг збуту. Дані ОДСПР можна було б подати у такий спосіб, щоб спостерігати обсяги збуту в грошовому еквіваленті для кожного магазину або збут у різних одиницях виміру для кожного типу продукту. Метою підсумовування є визначення й оцінювання трендів продажу або порівняння збутів різних типів продукції. Користувач може захотіти, наприклад, дізнатися: який тренд продукту А? Чи доцільно припинити продаж деякого продукту? Чи були ефективними витрати на рекламу для створення сприятливих змін у збуті? На всі ці запитання можна відповідати, використовуючи інтегровані дані. Операційні дані не підсумовуються в базах даних.

Мінливість даних. Тільки два види дій відбуваються в сховищі даних або базі даних ОДСППР: завантаження даних та організація доступу до них. Можна додавати дані в пакетах, але це вже не буде інтерактивною актуалізацією даних. Отже, дані СППР не змінюються з часом. Операційні дані мають непостійний характер, оскільки вони змінюються, як тільки відбуваються нові транзакції.

Вимірність даних. Багатовимірність даних є, можливо, найхарактернішою особливістю даних ОДСППР. З погляду менеджера чи аналітика дані ОДСППР завжди пов'язані між собою багатьма різними способами. Наприклад, коли аналізується збут продукту окремому споживачеві протягом певного проміжку часу, то можна зробити такий запит: «Скільки виробів типу X було продано споживачеві Y протягом останніх шести місяців?» Дані ОДСППР можуть досліджуватися в різних аспектах, наприклад, за видами продуктів, за регіонами і за часом. Здатність аналізувати, виділяти та подавати дані як інформацію в зручному вигляді є однією із головних позитивних характеристик ОДСППР. На противагу ним, операційні дані мають тільки одну вимірність.

Метадані. В орієнтованій на дані СППР важливо розробити й підтримувати *метадані* про дані СППР. Словники баз даних можуть бути і для систем оброблення транзакцій, але через те, що дані ОДСППР можуть надходити від багатьох джерел, створення словників і метаданих є особливо важливим для СППР. *Метадані* - це «інформація про дані» в базі даних СППР. До ресурсів метаданих належать каталоги та словники бази даних, а також імена змінних, довжини полів, допустимі значення змінних та описи елементів даних. Семантичні дані часто зберігаються в словнику бази даних. Метадані зберігають інформацію про зміни у схемі початкових джерел сховища даних або бази даних.

Орієнтовані на дані СППР часто відносять до типу *аналітичних систем (АС)*, тобто інформаційних систем, метою яких є лише аналіз даних. Інколи терміни «АС» і «ОДСППР» вживають як синоніми. Зауважимо, що стосовно інформаційних процесів аналітичні системи є вторинними по відношенню до

операційних транзакційних систем OLTP (On-line transaction processing), оскільки всі дані, що використовуються для аналізу, необхідно спочатку нагромадити і, за можливості, частково обробити, чим і займаються різні транзакційні системи, а лише потім їх проаналізувати. Аналітичні завдання залежно від концепції аналізу можна поділити на дві групи: завдання статичного та завдання оперативного аналізу. Ці дві групи аналітичних завдань суттєво відрізняються між собою.

Перша група завдань характеризується тим, що вони реалізуються на основі традиційної технології автоматизації розв'язання. За цієї технології спочатку формулюється технічне завдання, яке передається програмісту для програмування. Програміст складає програму та тестує її і лише після цього отримує результат у вигляді регламентованих, тобто чітко визначених форм. За такого підходу виникає велика затримка в часі між моментом виникнення потреби в аналізі та отриманням відповідного результату. Дуже часто результат аналізу, який був потрібний аналітику, отримують пізно і рішення приймається без його врахування. Тому для прийняття оперативних рішень такий вид аналізу не підходить.

Саме потреба в оперативному багатоаспектному бізнес-аналізі привела до виникнення нової OLAP-технології розв'язання аналітичних завдань. Ця технологія призначена забезпечувати аналітиків динамічним багатовимірним аналізом консолідованих даних. Як уже зазначалося, розв'язання аналітичних завдань не може обмежуватись лише даними транзакційних систем. Для порівняльного аналізу та виявлення тенденцій потрібно мати великі обсяги зовнішніх даних з різних статистичних збірників, з електронних та інших джерел. Зручним способом зберігання даних для виконання оперативних аналітичних завдань є сховища даних, що утворюють основу аналітичних інформаційних систем.

Орієнтовані на дані СППР мусять мати дані найвищої якості, інакше дані можуть призвести до невдач у вирішенні проблем. Дані найвищої якості - це *точні, своєчасні, значимі* (важливі) та *повні* (комплектні) дані. Оцінювання

або вимірювання якості джерел даних є попереднім завданням, яке пов'язане з оцінюванням технічної здійсності проекту орієнтованої на дані СППР.

6.2 Зародження й розвиток OLAP-систем. 12 правил Кода

OLAP (абревіатура від On-line Analytical Processing - інтерактивне аналітичне оброблення) фактично означає не окремі конкретні програмні продукти, а технологію багатовимірного аналізу даних, основу якої започаткувала опублікована 1993 року праця Е.Ф. Кода (E.F. Codd) «OLAP для користувачів-аналітиків: яким воно має бути», у якій він запропонував 12 правил, які виражали концепцію оперативного аналітичного оброблення даних і фактично послужили стандартом інструментальних засобів оперативного аналітичного оброблення.

Правила Кода для OLAP

1. Багатовимірне концептуальне зображення (Multidimensional conceptual view). Уможливорює користувачу перегляд даних, які можна аналізувати за регіонами, часом тощо.
2. Прозорість (Transparency). Робить базову (що є основою) аналітичну здатність цілком прозорою для користувача.
3. Доступність (Accessibility). Дає змогу створювати власну логічну схему для запам'ятовування неоднорідних фізичних даних.
4. Узгоджена продуктивність щодо створення повідомлень (Consistent reporting performance). Забезпечує надійну продуктивність підготовки звітів для будь-якої кількості вимірів, яку обирає користувач.
5. Архітектура клієнт/сервер (Client/server architecture). Забезпечує мінімум зусиль для використання цієї архітектури.

6. Генерування вимірів (Generic dimensionality). Має тільки одну логічну структуру для подання всіх вимірів.

7. Динамічне оброблення розріджених матриць (Dinamic sparse matrix handling). Ефективно поводитьсь з порожніми пропусками в матриці, що роблять матрицю розрідженою.

8. Багатокористувацька підтримка (Multiuser support). Уможлиблює одночасний доступ, захист та цілісність для багатьох користувачів.

9. Необмежені перехресно-вимірні операції (Unrestricted cross-dimensional operations). Виконує обчислення та інші операції над вимірами без втручання користувача.

10. Інтуїтивне маніпулювання даними (Intuitive data manipulation). Уможлиблює оброблення з деталізацією (drilling down), наближення/віддалення об'єкта, переорієнтацію та консолідоване подання даних і аналізів.

11. Гнучка організація створення звітів (Flexible reporting). Дає змогу користувачам легко й ефективно маніпулювати звітами даних.

12. Необмеженість вимірів та рівнів агрегації (Unlimited dimensoins and aggregation levels). Містить щонайменше 15, а то й 20 вимірів даних.

1995 року до них було додано ще кілька правил, що у своїй сукупності визначили основні сучасні вимоги до OLAP-систем. Ці правила були поділені на чотири групи.

Базові характеристики: багатовимірність моделі даних; інтуїтивні механізми маніпулювання даними; доступність; пакетне отримання даних; клієнт-серверна архітектура; прозорість (для користувача); багатокористувацька робота.

Спеціальні характеристики: оброблення неформалізованих даних; зберігання результатів окремо від вхідних даних; виокремлення даних, яких бракує (тобто вони мусять відрізнятися від нульового значення); оброблення значень, яких бракує (всі значення, яких бракує, мають бути проігноровані в разі аналізу).

Характеристики побудови звітів: гнучкі можливості одержання звітів; стабільна продуктивність підготовки звітів; автоматичне регулювання фізичного рівня.

Керування розмірністю: загальна функціональність; необмежена кількість вимірів та рівнів агрегування; необмежена кількість операцій над даними різних вимірів.

Ці концепції покладені в основу технології OLAP, хоча реально наявні OLAP-системи мають далеко не повний перелік описаних характеристик. OLAP-технологія, яку можна назвати також інтерактивним (діалоговим) аналітичним обробленням, дає змогу на основі багатовимірної (гіперкубічної) моделі даних (на відміну від плоскої реляційної моделі даних) моделювати реальні структури й зв'язки, що є виключно важливими для аналітичних систем. Вона призначена для створення багатопараметричних моделей з метою адекватнішого відображення реальних бізнес-процесів. Технологія OLAP дає змогу швидко змінювати погляди на дані залежно від вибраних параметрів і забезпечувати особу, що приймає рішення, повною картиною щодо ситуацій, які аналізуються.

Усі OLAP-системи побудовані на двох базових принципах:

1) дані, необхідні для прийняття рішень, слід попередньо агрегувати на всіх відповідних рівнях і організувати так, щоб забезпечити максимально швидкий доступ до них;

2) мова маніпулювання даними основана на бізнес-поняттях. Дані параметруються кількома рівноправними вимірами, наприклад, дані стосовно продажу у великій торговельній компанії можна аналізувати в таких вимірах: «час» (день, тиждень, місяць, квартал, рік), «географія» (місто, область, країна), «товар» (фірма-виробник, вид товару), «покупець» (стать, вік).

Оскільки в основі OLAP-технології лежить концепція гіперкуба моделі даних, то залежно від відповіді на запитання, чи існує гіперкуб як окрема фізична структура? Чи це є лише віртуальна модель даних? Розрізняють два основних типи аналітичного оброблення даних: MOLAP та ROLAP.

MOLAP (Multidimensional OLAP) - це багатовимірна OLAP-система, в якій гіперкуб реалізується як окрема база даних нереляційної структури, яка забезпечує багатовимірне зберігання, оброблення й подання даних. Програмні продукти, що належать до цього типу OLAP-технології, як правило, мають сервер багатовимірних баз даних. Ця структура забезпечує максимально ефективний щодо швидкості доступ до даних, проте потребує додаткового ресурсу пам'яті. Крім того, велика розмірність моделі даних і розрідженість гіперкубів приводить до необхідності витрат великих апаратних ресурсів, що не завжди може бути доцільним. Є чотири провідні продавці серверів MOLAP: Arbor Software (Essbase), Kenan Technologies (Acumate Enterprise), Oracle/IRI (Express), D& B/Pilot Software (Lightship). Крім того, були розроблені нова версія сервера Cognos PowerPlay та новий мультिवимірний сервер інституту SAS.

У ROLAP (Relational OLAP) багатовимірна структура реалізується реляційними таблицями. Така структура забезпечує зберігання великих обсягів інформації, проте є менш продуктивною з погляду ефективності OLAP-операцій. Незважаючи на те, що продавці ROLAP увійшли в ринок програмних продуктів тільки протягом останніх кількох років, чотири з них стали досить відомими. Два з них знаходяться серед найшвидше зростаючих корпорацій у Сполучених Штатах. Цими чотирма компаніями є: Microstrategy, Information Advantage, Stanford Technology Group (недавно придбаний Informix), IQ Software. На сьогодні відкриті сайти ROLAP, що містять аналітичні механізми сервера, створення звітів та інструментальні засоби аналізу, інструментальні засоби проектування системи та програмне забезпечення.

Певні недоліки, характерні основним типам OLAP-технологій, привели до появи нового типу аналітичних інструментів - HOLAP-систем, що являють собою гібридне (hybrid) оперативне аналітичне оброблення даних, де реалізуються обидва підходи, тобто доступ може надаватися як до багатовимірних баз даних, так і до даних реляційного типу.

На сьогодні розроблено досить багато систем підтримки прийняття рішень, сконструйованих з використанням OLAP-технології (Hyperion OLAP, Elite OLAP, Oracle Express та багато інших). Ринок програмних OLAP-продуктів постійно розширюється. Сучасні системи оперативного аналітичного оброблення надають користувачам можливість виконувати ключові завдання управління бізнес-процесом, зокрема додатки Hyperion OLAP; аналіз прибутковості; аналіз напрямів розвитку продукції; аналіз продажу; аналіз стану на ринку; аналіз асортименту продуктів; аналіз ризику; аналіз конкурентоспроможності; складання звітів з продуктивності; моделювання сценарію; аналіз бюджету й прогнозів тощо. Слід ще раз підкреслити, що згідно з сучасними поглядами на створення інформаційних систем OLAP-системи мають базуватися на спеціальній базі даних - сховищі даних, але можуть використовуватися і автономно.