**Лекція 4. OLAP**

**OLAP** ([англ.](http://uk.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D1%96%D0%B9%D1%81%D1%8C%D0%BA%D0%B0_%D0%BC%D0%BE%D0%B2%D0%B0) *online analytical processing*, аналітична обробка в реальному часі) — це технологія обробки інформації, що дозволяє швидко отримувати відповіді на багатовимірні аналітичні запити. OLAP є частиною такого ширшого поняття, як бізнес-аналітика, що також включає такі дисципліни як [реляційна звітність](http://uk.wikipedia.org/w/index.php?title=%D0%A0%D0%B5%D0%BB%D1%8F%D1%86%D1%96%D0%B9%D0%BD%D0%B0_%D0%B7%D0%B2%D1%96%D1%82%D0%BD%D1%96%D1%81%D1%82%D1%8C&action=edit&redlink=1) та [добування данних](http://uk.wikipedia.org/wiki/%D0%94%D0%BE%D0%B1%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F_%D0%B4%D0%B0%D0%BD%D0%B8%D1%85) (спосіб аналізу інформації в базі даних з метою відшукання аномалій та трендів без з'ясування смислового значення записів). Служить для підготовки бізнес-звітів з продажів, маркетингу, в цілях управління, для прогнозування, фінансової звітності та в схожих областях.

Бази даних, сконфігуровані для OLAP, використовують багатовимірні моделі даних, що дозволяє виконувати складні аналітичні та спеціалізовані запити за коротокий проміжок часу. Вони запозичують окремі аспекти навігаційних та ієрархічних баз даних, які є швидшими за реляційні БД.

Зазвичай результати OLAP-запитів представляють у формі матриць, де *виміри* складають рядки та колонки, а значеннями матриці є *розміри*.

Головна причина використання OLAP для обробки запитів — це швидкість. Реляційні БД зберігають сутності в окремих таблицях, які зазвичай добре нормалізовані. Ця структура зручна для операційних БД (системи [OLTP](http://uk.wikipedia.org/wiki/OLTP)), але складні багатотабличні запити в ній виконуються відносно повільно. Зручнішою моделлю для виконання запитів (але не для внесення змін) є [просторова БД](http://uk.wikipedia.org/w/index.php?title=%D0%9F%D1%80%D0%BE%D1%81%D1%82%D0%BE%D1%80%D0%BE%D0%B2%D0%B0-%D0%91%D0%94&action=edit&redlink=1). OLAP робить миттєвий знімок реляційної БД і структурує її в просторову модель для запитів. Заявлений час обробки запитів в OLAP становить близько 0,1% від аналогічних запитів до реляційної БД.

**Концепція OLAP**

*Докладніше:* [*OLAP-куб*](http://uk.wikipedia.org/wiki/OLAP-%D0%BA%D1%83%D0%B1)

Ядром будь-якої OLAP-системи є ідея [OLAP-куба](http://uk.wikipedia.org/wiki/OLAP-%D0%BA%D1%83%D0%B1) (*багатовимірний куб*, або [*гіперкуб*](http://uk.wikipedia.org/wiki/%D0%A2%D0%B5%D1%81%D0%B5%D1%80%D0%B0%D0%BA%D1%82)). OLAP-структура, створена з робочих даних, називається OLAP-кубом. Він складається з чисельних фактів (*розмірів*), розподіленних за *вимірами*. Зазвичай куб створюється за допомогою з'єднання таблиць із застосуванням схеми «зірка», або схеми «сніжинка». В центрі «зірки» знаходиться таблиця, яка містить ключові факти, за якими робляться запити. Множинні таблиці з вимірами приєднані до таблиці фактів. Ці таблиці показують, як можуть аналізуватися агреговані реляційні дані. Кількість можливих агрегацій визначається кількістю способів, якими первинні дані можуть бути ієрархічно відображені. Наприклад, всі клієнти можуть бути згруповані за містами, або за регіонами країни (Захід, Схід, Північ і т. д.), таким чином, 50 міст, 8 регіонів і 2 країни складуть 3 рівні ієрархії з 60 членами. Також клієнти можуть бути об'єднані за відношенням до продукції; якщо існують 250 продуктів у двох категоріях, 3 групи продукції і 3 виробничих підрозділи, то кількість агрегатів складе 16560. При додаванні вимірів в схему, кількість можливих варіантів швидко досягає десятків мільйонів і більше.

[OLAP-куб](http://uk.wikipedia.org/wiki/OLAP-%D0%BA%D1%83%D0%B1) містить в собі базові дані і інформацію про вимірювання (агрегати). Куб потенційно містить всю інформацію, яка може виявитися необхідною для відповідей на будь-які запити. Через величезну кількість агрегатів, часто повний розрахунок відбувається тільки для деяких вимірювань, для останніх же проводиться «на вимогу».

**Типи**

Традиційно OLAP-системи поділяють на такі види:

* OLAP з багатьма вимірюваннями (Multidimensional OLAP), MOLAP;
* реляційна OLAP (Relational OLAP), ROLAP;
* гібридна OLAP (Hybrid OLAP), HOLAP.

MOLAP це класична форма OLAP, так що її часто називають просто OLAP. Вона використовує підсумовуючу БД, спеціальний варіант процесора просторових БД і створює необхідну просторову схему даних зі збереженням як базових даних, так і агрегатів. ROLAP працює безпосередньо з реляційним сховищем, факти і таблиці з вимірюваннями зберігаються в реляційних таблицях, і для зберігання агрегатів створюються додаткові реляційні таблиці. HOLAP використовує реляційні таблиці для зберігання базових даних і багатовимірні таблиці для агрегатів. Особливим випадком ROLAP є ROLAP реального часу (Real-time ROLAP, або R-ROLAP). На відміну від ROLAP, в R-ROLAP для зберігання агрегатів не створюються додаткові реляційні таблиці, а агрегати розраховуються у момент запиту. При цьому багатовимірний запит до OLAP-системи автоматично перетвориться в SQL-запит до реляційних даних.

Кожен тип зберігання має певні переваги, хоча є розбіжності в їх оцінці у різних виробників. MOLAP краще всього підходить для невеликих наборів даних, він швидко розраховує агрегати і дає відповіді, але при цьому генеруються величезні обсяги даних. ROLAP оцінюється як більш масштабоване рішення, яке до того ж використовує найменший можливий простір. При цьому швидкість обробки значно знижується. HOLAP знаходиться між цими двома підходами, він досить добре масштабується і швидко обробляється. Архітектура R-ROLAP дозволяє проводити багатовимірний аналіз OLTP-даних в режимі реального часу.

Складність в застосуванні OLAP полягає в створенні запитів, виборі базових даних і розробці схеми, внаслідок чого більшість сучасних продуктів OLAP поставляються разом з величезною кількістю заздалегідь сконфігурованих запитів. Інша проблема полягає в базових даних. Вони повинні бути повними і несуперечливими.

**Реалізації OLAP**

Першим продуктом, що виконував OLAP-запити, був Express (компанія IRI). Проте сам термін OLAP був запропонований «батьком реляційних БД» Едгаром Коддом. А робота Кодда фінансувалася Arbor, компанією, що випустила свій власний OLAP-продукт Essbase роком раніше (пізніше куплений Hyperion, яка в 2007 р. була поглинена компанією Oracle). Як результат, «OLAP» Кодда з'явився в їх описі Essbase.

Інші добре відомі OLAP-продукти включають [Microsoft](http://uk.wikipedia.org/wiki/Microsoft) Analysis Services (що раніше називалися OLAP Services, частина SQL Server), DB2 OLAP Server від [IBM](http://uk.wikipedia.org/wiki/IBM) (фактично, EssBase з доповненнями від IBM), продукти MicroStrategy і інших виробників.

З технічної точки зору, представлені на ринку продукти діляться на «**фізичний OLAP**» і «**віртуальний**».

У першому випадку наявна програма, що виконує попередній розрахунок агрегатів, які потім зберігаються в спеціальній багатовимірній БД, що забезпечує швидкий доступ. Приклади таких продуктів: Microsoft Analysis Services, Oracle OLAP Option, Oracle/Hyperion EssBase, Cognos PowerPlay.

У другому випадку дані зберігаються у реляційних СУБД, а агрегати можуть не існувати взагалі або створюватися за першим запитом у СУБД або кеші аналітичного ПО. Приклади таких продуктів: SAP BW, BusinessObjects, Microstrategy.

Системи, що мають в своїй основі «фізичний OLAP» забезпечують стабільно кращий час відгуку на запити, ніж системи «віртуальний OLAP». Постачальники систем «віртуальний OLAP» заявляють про більшу масштабованість їх продуктів в плані підтримки дуже великих обсягів даних.

З погляду користувача обидва варіанти виглядають схожими за можливостями.

Найбільше застосування OLAP знаходить в продуктах для бізнес-планування і сховищах даних.