

Тенденции и проблемы OLAP

Кудрявцев Юрий

Oracle BI Пиво 2008

- 1 Определение OLAP
- 2 Проблемы обработки и хранения OLAP-данных
- 3 Тенденции
 - MDX как стандартный язык запросов к OLAP данным
 - Тенденции развития OLAP

OLAP (OnLine Analytical Processing) – инструмент интерактивного анализа данных пользователем. Задачи OLAP-средств:

- Предоставление больших объемов данных пользователю для анализа
- Большая скорость работы (время обработки запроса не должно превышать 3 секунд)
- Удобная модель представления данных (многомерные кубы). Интуитивно понятные интерфейсы работы (roll-up, drill-down)

Термин введен в статье Кодда “Providing OLAP for End-User Analysis” отосланной в IEEE Computer. В статье сформулированы 12 признаков OLAP-системы.

Последние 4 страницы статьи посвящены Essbase – проверка соответствия OLAP критериям.

Жена Кодда в это время работает в Arbor Software (разработчик Essbase). Arbor Software спонсировало написание статьи.

Журнал Computer после публикации официально изымает статью Кодда из своих архивов

Nigel Pendse – <http://www.olapreport.com>

FASMI

- FAST
- Analysis
- Shared
- Multidimensional
- Information

Система, удовлетворяющая вышеуказанным критериям – OLAP-система.

- Измерение. В SQL обычно аргументы запроса с group by.
- Иерархия в измерении
- Мера
- Агрегирующие функции – агрегирующие функции group by запросов.
- Куб – Cube by, group by по всем комбинациям измерений.

При формировании куба возникают ячейки, которых не было в исходных данных (за счет выбора координат) – это разреженность, и масса дублирующих друг-друга агрегатов – избыточность.

Простые задачи:

Нарастающий итог продаж (за квартал, running period). Window by запросы в SQL.

Кросс-таблица.

```
P: asc exec distinct p from t  
exec P!(p!v) P by k:k from t
```

Сложные задачи Break-back, back-solving (обратное вычисление уравнений в многомерном пространстве)

Имитационное, сценарное моделирование

- Categorical – простые запросы
- Exegetical – многомерный анализ, drill-up/down
- Contemplative – изменение расчетных результатов, при изменении входных параметров
- Formulaic – задание правил поведения системы и цели расчетов, сценарное моделирование

Форматы хранения:

- ROLAP – все данные куба (фактическая таблица + агрегаты) хранятся в реляционной таблице
- MOLAP – и фактические данные, и агрегаты хранятся в многомерной бд
- HOLAP – фактические данные хранятся в реляционной таблице, агрегаты – в виде многомерных структур (многомерной БД)

Проблемы:

- Хранение агрегатов (материализация) или вычисление на лету. Задача выбора оптимального набора материализованных представлений NP-полна.
- Моделирование измерений и вычислений
- Схемы хранения «снежинка» и «звезда» (Кимбалл и Инмон)

Расширения языка SQL:

- Grouping Set
- Rollup
- Cube
- Window By
- Oracle model by

Реляционные СУБД все еще плохо подходят для аналитических задач

Простой вопрос: посчитать в SQL тренд?

В Excel это просто $=Trend(X, Y)$

В Cognos $@Trend$

trend в Oracle Express.

И просто в Oracle

```
SELECT REG, X,
SUM(
CASE
WHEN ((SELECT REGR_SXX(Y,X) FROM data_t t1 WHERE T.REG = t1.
REG) = 0) THEN (Y)
ELSE X*(SELECT REGR_SXY(Y,X)/REGR_SXX(Y,X) FROM data_t t1
WHERE T.REG = t1.REG)+(SELECT REGR_AVGY(Y,X) - REGR_SXY(Y,
X)/REGR_SXX(Y,X)*REGR_AVGX(Y,X) FROM data_t t2 WHERE T.REG
=t2.REG )
END
)
FROM data_t T
GROUP BY reg, X
```

Реализуется в:

Microsoft UDM

Cognos DMR

Essbase 11 (возможность Rolap партиций)

Таким образом все запросы формулируются в OLAP-стиле, а где хранятся данные – не так уж и важно. В качестве языка запросов все чаще выступает MDX

Разработан в Microsoft

OLE DB for OLAP -> XML/A

Де-факто стандарт обращения к OLAP-движкам

Поддерживается:

- Microsoft Analysis Services
- Oracle Essbase
- SAP BW
- Pentaho Mondrian

Не является открытым стандартом, не ANSI, не ISO...

```
SELECT
{ [Measures].[Dollar Sales], [Measures].[Unit Sales] }
on columns,
{ [Time].[Q1, 2005], [Time].[Q2, 2005] }
on rows
FROM [Sales]
WHERE ([Customer].[MA])
```

В SQL

```
SELECT x,y FROM t;
```

по мощности больше, чем

```
SELECT x,y FROM t WHERE z = 1;
```

В MDX

```
SELECT x on axis(0),y on axis(1) FROM t; -- в этом случае по  
z выбирается DefaultMember
```

=

```
SELECT x on axis(0),y on axis(1) FROM t WHERE z = 1;
```

т.к. выбирается другой срез по z.

Запрос

```
SELECT
{ filter ([Stores] , [Measures].[Sales]>0) }
on columns ,
{ [Customer].[Groups] }
on rows
FROM [Sales]
WHERE ([Time].[Years].[2008])
```

Даст совершенно разные результаты (по колонкам) в MS AS и SAP BW. См. блог разработчиков Panorama.

<http://www.panorama.com/blog/?p=44>

- Консолидация скупленного и правильное позиционирование на рынке.
- Представление всей отчетности в многомерной форме
- Текстовые меры. Essbase 11.
- Real-time OLAP. Сокращение времени обновления кубов.


- Языки векторного программирования (APL, K)
- Column-Store databases
- In-memory базы данных (TimesTen, Applix, KX) как буфера для хранения агрегатов в СУРБД.
- OLAP in Clouds (Google + Panorama, Vertica)

- <http://www.ritman.net>
- <http://www.tim-tows-hyperion-blog.blogspot.com>
- <http://www.vlamis.com/blog>
- <http://www.looksmarter.blogspot.com>
- <http://www.oracleolap.blogspot.com>
- <http://www.oracledmt.blogspot.com>

 [Codd E.F.](#)
Providing OLAP for end-user analysis: An IT mandate.
[IEEE Computer](#),1993

 [Thomsen Eric](#)
OLAP Solutions: Building Multidimensional Information Systems. Second Edition.
[Wiley](#), 2002

 [Rafanelli M.](#)
Multidimensional Databases — Problems and Solutions.
[Idea Group Pub.](#), Hershey, London, Melbourne, Singapore, Beijing, 2003

 [Celko Joe.](#)
Analytics and OLAP in SQL.
[Morgan Kaufmann](#), 2006