

Модели и технологии оперативного анализа данных

Лекция 2

Технологии оперативного анализа данных

Гедранович Ольга Брониславовна,
старший преподаватель кафедры ИТ, МИУ
volha.b.k@gmail.com

24.01.2017 2


Вопросы лекции

- Многомерная модель данных.
- Определение OLAP-систем.
- Архитектура OLAP-систем: ROLAP, MOLAP, HOLAP.
- Основные операции OLAP-систем.

24.01.2017 3

Многомерная модель данных

В 1993 году Э. Кодд — основоположник реляционной модели БД — рассмотрел ее недостатки, указав в первую очередь на **невозможность «объединять, просматривать и анализировать данные с точки зрения множественности измерений, т. е. самым понятным для аналитиков способом».**



24.01.2017 4


Многомерная модель данных

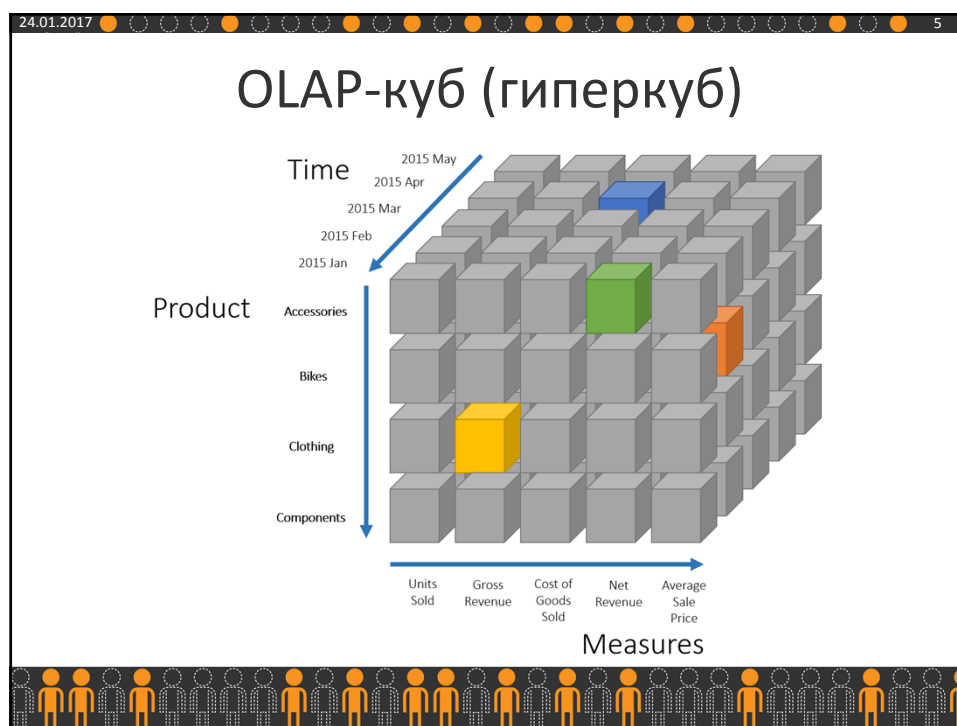
Измерение (Dimension) — это последовательность значений одного из анализируемых параметров.
 Например, для параметра "время" это последовательность календарных дней, для параметра "регион" это может быть список городов.

Каждое измерение может быть представлено в виде иерархической структуры. Например, измерение "Исполнитель" может иметь следующие иерархические уровни: "предприятие — подразделение — отдел — служащий". Более того, некоторые измерения могут иметь несколько видов иерархического представления. Например, измерение "Время" может включать две иерархии со следующими уровнями: "год — квартал — месяц — день" и "неделя — день".

На пересечениях осей измерений располагаются данные, количественно характеризующие анализируемые факты, — **меры (Measures)**.

Это могут быть объемы продаж, выраженные в единицах продукции или в денежном выражении, остатки на складе, издержки и т. п.



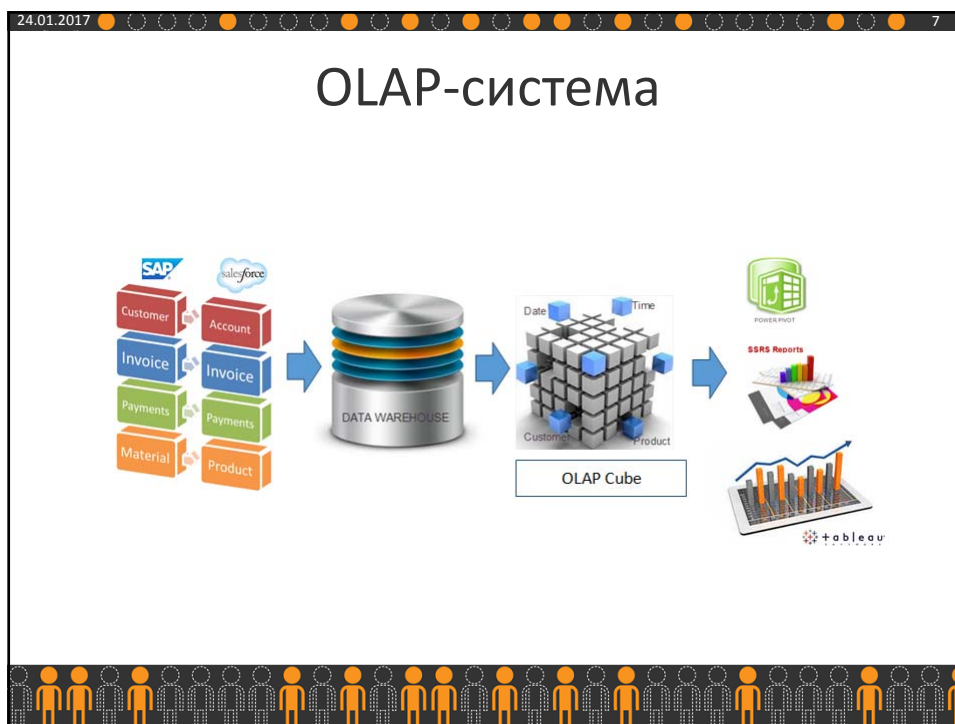


24.01.2017 6

OLAP-система

OLAP (OnLine Analytical Processing — оперативная аналитическая обработка данных) — подход к аналитической обработке данных, базирующийся на их многомерном иерархическом представлении, являющийся частью более широкой области информационных технологий — бизнес-аналитики (BI — Business Intelligence).

Системы, построенные на основе технологии OLAP, предоставляют возможности по составлению отчетов, выполнению сложных аналитических расчетов, построению прогнозов и сценариев, разработке множества вариантов планов.



24.01.2017 8

Требования к OLAP-системам

- Э. Кодд в 1993 году опубликовал статью «OLAP для пользователей-аналитиков: каким он должен быть», где изложил основные концепции оперативной аналитической обработки и определил 12 требований, которым должны удовлетворять продукты, позволяющие выполнять оперативную аналитическую обработку.
- В 1995 году Э. Кодд к приведенному перечню добавил 6 правил и разбил все 18 правил на четыре группы, назвав их особенностями. Эти группы получили названия B, S, R и D.


24.01.2017 8

24.01.2017 9

Требования к OLAP-системам

Основные особенности (B):

1. Многомерность (Multidimensional conceptual view).
2. Прозрачность (Transparency).
3. Доступность (Accessibility).
5. Клиент-серверная архитектура (Client/server architecture).
8. Поддержка многопользовательского режима (Multi-user support).
10. Интуитивная манипуляция данными (Intuitive data manipulation).
13. Пакетное извлечение против интерпретации.
14. Поддержка всех моделей OLAP-анализа.



24.01.2017 10


Требования к OLAP-системам

Специальные особенности (S):

15. Обработка ненормализованных данных.
16. Сохранение результатов OLAP: хранение их отдельно от исходных данных.
17. Исключение отсутствующих значений.
18. Обработка отсутствующих значений.

Особенности представления отчетов (R):

4. Постоянная производительность при разработке отчетов (Consistent reporting performance).
7. Динамическое управление разреженными матрицами (Dynamic sparse matrix handling).
11. Гибкие возможности получения отчетов (Flexible reporting).




24.01.2017 11

Требования к OLAP-системам

Управление измерениями (D):


- 6. Равноправие измерений (Generic Dimensionality).
- 9. Неограниченные перекрестные операции (Unrestricted cross-dimensional operations).
- 12. Неограниченная размерность и число уровней агрегации (Unlimited Dimensions and aggregation levels).



24.01.2017 12

Требования к OLAP-системам

Найджел Пендс (Nigel Pendse) предложил использовать взамен предложенных Э. Коддом правил OLAP так называемый **тест FASMI** (Fast Analysis of Shared Multidimensional Information — быстрый анализ доступной многомерной информации), более точно характеризующий требования к таким системам.




24.01.2017 13

Архитектура OLAP-систем

OLAP-система включает в себя два основных компонента:

- **OLAP-сервер** — обеспечивает хранение данных, выполнение над ними необходимых операций и формирование многомерной модели на концептуальном уровне. В настоящее время OLAP-серверы объединяют с ХД или ВД;
- **OLAP-клиент** — представляет пользователю интерфейс к многомерной модели данных, обеспечивая его возможностью удобно манипулировать данными для выполнения задач анализа.



24.01.2017 14


Архитектура OLAP-систем

Выделяют три основных способа реализации многомерной модели:

- **MOLAP** — многомерный (multivariate) OLAP — используют многомерные БД;
- **ROLAP** — реляционный (relational) OLAP — используют реляционные БД;
- **HOLAP** — гибридный (hybrid) OLAP — используют и многомерные, и реляционные БД.

Часто в литературе по OLAP-системам можно встретить аббревиатуры DOLAP и JOLAP:

- **DOLAP** — настольный (desktop) OLAP. Является недорогой и простой в использовании OLAP-системой, предназначенной для локального анализа и представления данных, которые загружаются из реляционной или многомерной БД на машину клиента;
- **JOLAP** — основанная на Java коллективная OLAP-API-инициатива, предназначенная для создания и управления данными и метаданными на серверах OLAP.



24.01.2017 15

Представление данных в MOLAP

Измерения				Меры	
Клиент	Время	Продавец	Продукт	Сумма сделки	Объем сделки
Школа №25	20.08.2016	Юрий Т.	Карандаши	690	30
Школа №25	20.08.2016	Юрий Т.	Ручки	830	40
Школа №25	20.08.2016	Юрий Т.	Тетради	500	25
Школа №25	20.08.2016	Юрий Т.	Фломастеры	700	35
Школа №25	20.08.2016	Юрий Т.	Краски	600	15
Школа №25	20.08.2016	Юрий Т.	Маркеры	1 500	100
Школа №25	20.08.2016	Дмитрий А.	Карандаши	690	30
Школа №25	20.08.2016	Дмитрий А.	Ручки	830	40
Школа №25	20.08.2016	Дмитрий А.	Тетради	500	25
Школа №25	20.08.2016	Дмитрий А.	Фломастеры	700	35
Школа №25	20.08.2016	Дмитрий А.	Краски	2 000	50
Школа №25	20.08.2016	Дмитрий А.	Маркеры	2 250	150
Школа №25	20.08.2016	Алексей Ш.	Карандаши	230	10
Школа №25	20.08.2016	Алексей Ш.	Ручки	1 000	0

24.01.2017 16

OLAP vs. OLTP

Data Warehouse (OLAP)	Operational Database (OLTP)
Involves historical processing of information.	Involves day-to-day processing.
OLAP systems are used by knowledge workers such as executives, managers and analysts.	OLTP systems are used by clerks, DBAs, or database professionals.
Useful in analyzing the business.	Useful in running the business.
It focuses on Information out.	It focuses on Data in.
Based on Star Schema, Snowflake, Schema and Fact Constellation Schema.	Based on Entity Relationship Model.
Contains historical data.	Contains current data.
Provides summarized and consolidated data.	Provides primitive and highly detailed data.
Provides summarized and multidimensional view of data.	Provides detailed and flat relational view of data.
Number of users is in hundreds.	Number of users is in thousands.
Number of records accessed is in millions.	Number of records accessed is in tens.
Database size is from 100 GB to 1 TB	Database size is from 100 MB to 1 GB.
Highly flexible.	Provides high performance.

24.01.2017 17

Основные операции OLAP-систем

Над гиперкубом могут выполняться следующие операции:

- **Slice** (двумерный (плоскостной) срез);
- **Dice** (многомерный подкуб);
- **Roll-Up** (консолидация, агрегация, обобщение);
- **Drill down** (детализация);
- **Pivot** (вращение).

24.01.2017 17

24.01.2017 18

Slice

The diagram illustrates the Slice operation. A 3D cube is shown with dimensions: Locations (cities) with values Chicago, New York, Toronto, Vancouver; Time (Quarter) with values Q1, Q2, Q3, Q4; and item(types) with values Mobile, Modem, Phone, Security. The Q1 slice is highlighted. Below, a 2D table shows the resulting slice for Q1.

Locations (cities)	Mobile	Modem	Phone	Security
Chicago	605	825	14	400
New York				
Toronto				
Vancouver				

- The slice operation selects one particular dimension from a given cube and provides a new sub-cube.
- Here Slice is performed for the dimension "time" using the criterion time = "Q1".
- It will form a new sub-cube by selecting one or more dimensions.

24.01.2017 18

24.01.2017 19

Dice

Dice for (location = "Toronto" or "Vancouver") and (time = "Q1" or "Q2") and (item = "Mobile" or "Modem")

- Dice selects two or more dimensions from a given cube and provides a new sub-cube.
- The dice operation on the cube based on the following selection criteria involves three dimensions:
 (location = "Toronto" or "Vancouver")
 (time = "Q1" or "Q2")
 (item = "Mobile" or "Modem")

24.01.2017 20

Roll-up

roll-up on location (from cities to countries)

- Roll-up performs aggregation on a data cube in any of the following ways:
 - By climbing up a concept hierarchy for a dimension;
 - By dimension reduction.
- Roll-up is performed by climbing up a concept hierarchy for the dimension location.
- Initially the concept hierarchy was "street < city < province < country".
- On rolling up, the data is aggregated by ascending the location hierarchy from the level of city to the level of country.
- The data is grouped into cities rather than countries. But when roll-up is performed, one or more dimensions from the data cube are removed.

24.01.2017 21

Drill-down

- Drill-down is the reverse operation of roll-up. It is performed by either of the following ways:
 - By stepping down a concept hierarchy for a dimension;
 - By introducing a new dimension.
- Drill-down is performed by stepping down a concept hierarchy for the dimension time.
- Initially the concept hierarchy was "day < month < quarter < year".
- On drilling down, the time dimension is descended from the level of quarter to the level of month.
- When drill-down is performed, one or more dimensions from the data cube are added.
- It navigates the data from less detailed data to highly detailed data.

24.01.2017 22

Pivot

- The pivot operation is also known as rotation. It rotates the data axes in view in order to provide an alternative presentation of data.
- In this the item and location axes in 2-D slice are rotated.

24.01.2017 23

Основная литература

- Анализ данных и процессов : учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. — 3-е изд., перераб. и доп. — СПб : БХВ-Петербург, 2009. — 512 с.
- Data Warehousing – OLAP [Electronic resource] / TutorialsPoint. – Mode of access: https://www.tutorialspoint.com/dwh/dwh_olap.htm. – Date of access: 20.01.2017.
- OLAP-системы [Electronic resource] / TAdviser. – Mode of access: <http://www.tadviser.ru/index.php/Статья:OLAP-системы>. – Date of access: 20.01.2017.

