

Информационные подходы статистической обработки многомерных данных

Таблица 1. Сравнение методов по показателям в количественной деятельности и в качестве эффективности

Показатель	Методы по количественной деятельности, тыс. руб.	Методы по качественной деятельности, тыс. руб.
1	1000	1000
2	2000	2000
3	3000	3000
4	4000	4000
5	5000	5000
6	6000	6000
7	7000	7000
8	8000	8000
9	9000	9000
10	10000	10000
11	11000	11000
12	12000	12000
13	13000	13000
14	14000	14000
15	15000	15000
16	16000	16000
17	17000	17000
18	18000	18000
19	19000	19000
20	20000	20000
21	21000	21000
22	22000	22000
23	23000	23000
24	24000	24000
25	25000	25000
26	26000	26000
27	27000	27000
28	28000	28000
29	29000	29000
30	30000	30000
31	31000	31000
32	32000	32000
33	33000	33000
34	34000	34000
35	35000	35000
36	36000	36000
37	37000	37000
38	38000	38000
39	39000	39000
40	40000	40000
41	41000	41000
42	42000	42000
43	43000	43000
44	44000	44000
45	45000	45000
46	46000	46000
47	47000	47000
48	48000	48000
49	49000	49000
50	50000	50000

Учебное пособие

Астана 2019

Кульмамиров С. А.,
Кубаев К. Е.,
Байшоланова К. С.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	7
1 ОСНОВЫ OLAP ТЕХНОЛОГИИ	11
1.1 OLAP - удобный инструмент анализа.....	12
1.2 OLAP - Многомерное представление: куб.....	13
1.3 Архитектура OLAP-приложений.....	16
1.4 Технические аспекты хранения данных.....	17
2 OLAP-ТЕХНОЛОГИИ	20
2.1 Основное содержание.....	20
2.2 Функции OLAP.....	29
3 ПРОПУСКИ, ОШИБКИ И АНОМАЛЬНЫЕ ЗНАЧЕНИЯ ДАННЫХ	37
3.1 Противоречивость информации.....	37
3.2 Аномальные значения данных.....	38
3.3 Ошибки ввода данных.....	38
4 ТАБЛИЧНЫЕ ФОРМАТЫ ДАННЫХ	40
4.1 Таблица фактов.....	40
4.2 Таблицы измерений.....	43
ЛЕКЦИЯ 5. ДАННЫЕ В ВИДЕ КУБА, МЕТКИ, ИЕРАРХИИ И УРОВНИ	52
5.1 Разрезание куба.....	52
5.2 Метки, иерархии и уровни.....	53
6 ХРАНИЛИЩА ДАННЫХ И АНАЛИТИЧЕСКИЕ СИСТЕМЫ	58
6.1 MRP и ERP.....	58
6.2 Сходство и различия двух подходов к архитектуре хранилищ данных.....	64
6.3 Аналитические системы и хранилища данных.....	67

**7 ИНФОРМАЦИОННАЯ СИСТЕМА OLAP –
АНАЛИТИЧЕСКАЯ ОБРАБОТКА
МНОГОМЕРНЫХ ДАННЫХ**

7.1 Технология аналитической обработки многомерных данных	74
7.2 Сравнение технологий оперативной аналитической обработки данных	78

**8 РАЗМЕРНОСТЬ МНОГОМЕРНОГО
ПРОСТРАНСТВА И МЕТОДЫ
ЕЕ СНИЖЕНИЯ**

8.1 Задачи снижения размерности	84
8.2 Компонентный анализ	84
8.3 Факторный анализ	86
8.4 Использование методов снижения размерности	88
8.5 Классификация многомерных наблюдений и кластерный анализ	89
8.6 Дискриминантный анализ	101

**9 ПРИМЕНЕНИЕ КЛАССИФИКАЦИИ
МНОГОМЕРНЫХ НАБЛЮДЕНИЙ**

9.1 Использование ППП в классификации	109
9.2 Кластерный анализ торговых предприятий	109
9.3 Кластерный анализ мировой демографической статистики	115
9.4 Дискриминантный анализ надежности банков	119
9.5 Дискриминантный анализ мировой демографической статистики	121

**10 ИНФОРМАЦИОННАЯ
ИНФРАСТРУКТУРА ПРЕДПРИЯТИЯ**

10.1 Аналитическая пирамида	124
10.2 Уровень транзакционных систем	124
10.3 Системы бизнес-интеллекта	125

10.4 Аналитические приложения	130
10.5 Соотношение аналитических приложений и систем бизнес-интеллекта	132
11 АНАЛИТИЧЕСКИЕ БАНКОВСКИЕ СИСТЕМЫ	134
11.1 Назначение аналитической обработки данных в АБС	134
11.2 Архитектура хранилищ данных	135
11.3 Принципы обработки данных в хранилищах	141
11.4 Классификация продуктов OLAP по способ	142
11.5 Системы интеллектуального анализа данных у представления данных	144
12 СИСТЕМЫ БИЗНЕС ИНТЕЛЛЕКТА И АНАЛИТИЧЕСКИЕ ПРИЛОЖЕНИЯ	148
12.1 Проблема «единого взгляда» на управленческую информацию	148
12.2 Хранилища данных	149
12.3 Характеристики OLAP-систем	151
12.4 Разновидности многомерного хранения данных	153
12.5 Средства формирования запросов и построения отчетов	155
12.6 Аналитические приложения и BSC-системы	155
12.7 Системы корпоративного Планирования и бюджетирования	158
12.8 Системы консолидации финансовой отчетности	160
12.9 Другие аналитические приложения	162

13 УПРАВЛЕНИЕ ЭФФЕКТИВНОСТЬЮ БИЗНЕСА	169
13.1 История возникновения и сущность концепции ВРМ.....	169
13.2 Функциональность ВРМ-систем.....	171
13.3 Архитектура ВРМ-систем.....	176
14 ТЕХНОЛОГИИ DATA MINING	179
14.1 Статистические пакеты.....	183
14.2 Нейронные сети и экспертные системы.....	183
14.3 Информационные системы поддержки деятельности руководителя.....	187
15 КАТЕГОРИИ ИНФОРМАЦИОННЫХ СИСТЕМ	191
15.1 Типы данных в организации.....	191
15.2 От переработки данных к анализу.....	193
15.3 Системы диалоговой обработки транзакций.....	198
15.4 Управляющие информационные системы.....	202
15.5 Системы поддержки принятия решений.....	208
ЛИТЕРАТУРА	213
ПРИЛОЖЕНИЕ А. Краткий терминологический словарь (гlossарий).....	217

ВВЕДЕНИЕ

Развитие экономической отрасли нашей страны сейчас связано с ускоренным применением достижений новых технологий и цифровых информационных систем. Внедрение экономических информационных систем (ИС) является элементом управления экономикой страны и зависит от способности оперативно принимать управленческие решения и адаптироваться к изменениям внешней экономической среды страны и информационных потребностей общества.

В пособии излагаются математические методы и модели, позволяющие систематизировать обработку статистических данных большой размерности (Big Data) для их удобного представления и интерпретации. Они позволяют решать задачу снижения размерности исследуемого признакового пространства расположения данных и информации, осуществлять анализ структуры, тесноты и конкретного вида статистических связей между исследуемыми признаками различной природы – количественными, ординарными, номинальными, а также проводить структуризацию и классификацию объектов, имеющих многомерность в природе.

В более детальном виде в исследованиях авторов осуществлялся поиск ответов на следующие виды задач:

- освоение теоретических основ исследования сложных социальных, экономических явлений и процессов с учетом многообразия количественных и качественных взаимосвязей и закономерностей их развития;
- расширение и углубление математических знаний в области статистической обработки многомерных данных в составе функционирования информационной системы;
- овладение принципами применения многомерных статистических методов для комплексного анализа и моделирования социально-экономических процессов;
- изучение многомерных статистических моделей и приобретение практических навыков по их применению;

- ознакомление и овладение навыками работы с современной цифровой технологией и программным обеспечением многомерных статистических методов.

Обычно социальные и экономические объекты характеризуются достаточно большим числом параметров, образующих многомерные векторы, и особое значение в экономических и социальных исследованиях приобретают задачи изучения взаимосвязей между компонентами этих векторов, причем эти взаимосвязи необходимо выявлять на основании ограниченного числа многомерных наблюдений.

Многомерным статистическим анализом называется раздел математической статистики, изучающий методы сбора и обработки многомерных статистических данных, их систематизации и обработки с целью выявления характера и структуры взаимосвязей между компонентами исследуемого многомерного признака, получения практических выводов.

Предположим, что рассматривается некоторая совокупность, состоящая из n стран, для каждой из которых известны макроэкономические показатели: X_1 - валовой внутренний продукт, X_2 - площадь территории, X_3 - средняя продолжительность жизни населения. В результате получен набор из n наблюдений над k -мерным случайным вектором $X = (X_1, X_2, \dots, X_k)^*T$. Похожая ситуация возникает, когда изучается совокупность предприятий, для каждого из которых рассчитаны показатели X_1 - валовая прибыль, X_2 - численность работников, X_3 - стоимость основных производственных фондов.

Типичные задачи, которые можно решить методами многомерного статистического анализа, таковы:

- по наблюдавшимся значениям случайного вектора $X = (X_1, X_2, \dots, X_k)^*T$ может понадобиться изучить связь между его компонентами X_1, X_2, \dots, X_k ;

- может понадобиться определить, какие из (большого числа) рассчитанных показателей X_1, X_2, \dots, X_k в наибольшей степени влияют на валовой внутренний продукт или на продолжительность жизни населения;

- может понадобиться классифицировать страны по какому-либо признаку.

При этом способы сбора данных могут различаться. Так, если исследуется мировая экономика, то естественно взять в качестве объектов, на которых наблюдаются значения вектора X , страны, если же изучается национальная экономическая система, то естественно наблюдать значения вектора X на одной и той же (интересующей пользователя ИС) стране в различные моменты времени.

Такие статистические методы, как множественный корреляционный и регрессионный анализ, традиционно изучаются в курсах теории вероятностей и математической статистики [3, 14], рассмотрению прикладных аспектов регрессионного анализа (дисциплина «Эконометрика») [3, 18].

Именно таким методам исследования многомерных генеральных совокупностей на основании статистических данных посвящено данное учебное пособие. Методы снижения размерности многомерного пространства позволяют без существенной потери информации перейти от первоначальной системы большого числа наблюдаемых взаимосвязанных факторов к системе существенно меньшего числа скрытых (ненаблюдаемых) факторов, определяющих вариацию первоначальных признаков.

Методы многомерной классификации предназначены для разделения совокупностей объектов (характеризующиеся большим числом признаков) на классы, в каждый из которых должны входить однородные ОУ. Такую классификацию на основании статистических данных о значениях признаков на объектах можно провести методами кластерного и дискриминантного анализа.

Развитие вычислительной техники и программного обеспечения способствует широкому внедрению методов многомерного статистического анализа в практику. Пакеты прикладных программ с удобным пользовательским интерфейсом, такие как SPSS, Statistica, SAS, снимают трудности в применении указанных методов, заключающиеся в сложно-

сти математического аппарата, опирающегося на линейную алгебру, теорию вероятностей и математическую статистику, и громоздкости вычислений.

Однако применение программ без понимания математической сущности используемых алгоритмов способствует развитию у пользователя экономической ИС иллюзии простоты применения многомерных статистических методов, что может привести к неверным или необоснованным результатам. Значимые практические результаты могут быть получены только на основе профессиональных знаний в предметной области, подкрепленных владением математическими методами и пакетами прикладных программ, в которых эти методы реализованы.

Поэтому для методов статической обработки данных приводятся основные теоретические сведения, в том числе алгоритмы; обсуждается реализация этих методов и алгоритмов в пакетах прикладных программ. Рассматриваемые методы иллюстрируются примерами их практического применения в экономике с использованием пакетов моделирования статистики данных.

В связи с этим в учебном пособии рассматриваются вопросы теории и практики проведения статистической обработки многомерных данных в среде функционирования интегрированных экономических информационных систем. Они предназначены для использования на всех уровнях управления экономическими объектами, а также организации и управления процессом проектирования цифровых ИС с использованием различных методов и инструментальных средств обработки больших данных.

Предлагаемые главы пособия способствуют формированию у обучающихся вузов знаний о современных экономических ИС и перспективах развития статистической обработки многомерных данных в среде цифровых систем и технологий.

1 ОСНОВЫ OLAP ТЕХНОЛОГИИ

Термин OLAP неразрывно связан с термином «хранилище данных» (Data Warehouse). Данные в хранилище попадают из оперативных систем (OLTP-систем), которые предназначены для автоматизации бизнес-процессов. Кроме того, хранилище может пополняться за счет внешних источников, например статистических отчетов.

Зачем строить хранилища данных - ведь они содержат заведомо избыточную информацию, которая и так “живет” в базах или файлах оперативных систем? Ответить можно кратко: анализировать данные оперативных систем напрямую невозможно или очень затруднительно. Это объясняется различными причинами, в том числе разрозненностью данных, хранением их в форматах различных СУБД и в разных “уголках” корпоративной сети. Но даже если на предприятии все данные хранятся на центральном сервере БД (что бывает крайне редко), аналитик почти наверняка не разберется в их сложных, подчас запутанных структурах. Автор имеет достаточно печальный опыт попыток “накормить” голодных аналитиков “сырыми” данными из оперативных систем - им это оказалось “не по зубам”.

Таким образом, задача хранилища - предоставить “сырье” для анализа в одном месте и в простой, понятной структуре. Ральф Кимбалл в предисловии к своей книге “The Data Warehouse Toolkit” пишет, что если по прочтении всей книги читатель поймет только одну вещь, а именно: структура хранилища должна быть простой, - автор будет считать свою задачу выполненной.

Есть и еще одна причина, оправдывающая появление отдельного хранилища - сложные аналитические запросы к оперативной информации тормозят текущую работу компании, надолго блокируя таблицы и захватывая ресурсы сервера. Под хранилищем можно понимать не обязательно гигантское скопление данных - главное, чтобы оно было удобно для анализа. Вообще говоря, для маленьких хранилищ предна-

значается отдельный термин - Data Marts (киоски данных), но в нашей российской практике его не часто услышишь.

1.1 OLAP - удобный инструмент анализа

Централизация и удобное структурирование - это далеко не все, что нужно аналитику. Ему ведь еще требуется инструмент для просмотра, визуализации информации. Традиционные отчеты, даже построенные на основе единого хранилища, лишены одного - гибкости. Их нельзя "покрутить", "развернуть" или "свернуть", чтобы получить желаемое представление данных. Конечно, можно вызвать программиста (если он захочет прийти), и он (если не занят) сделает новый отчет достаточно быстро - скажем, в течение часа (пишу и сам не верю - так быстро в жизни не бывает; давайте дадим ему часа три). Получается, что аналитик может проверить за день не более двух идей. А ему (если он хороший аналитик) таких идей может приходиться в голову по нескольку в час. И чем больше "срезов" и "разрезов" данных аналитик видит, тем больше у него идей, которые, в свою очередь, для проверки требуют все новых и новых "срезов". Вот бы ему такой инструмент, который позволил бы разворачивать и сворачивать данные просто и удобно! В качестве такого инструмента и выступает OLAP.

Хотя OLAP и не представляет собой необходимый атрибут хранилища данных, он все чаще и чаще применяется для анализа накопленных в этом хранилище сведений.

Компоненты, входящие в типичное хранилище, представлены на рисунке 1.

Оперативные данные собираются из различных источников, очищаются, интегрируются и складываются в реляционное хранилище. При этом они уже доступны для анализа при помощи различных средств построения отчетов. Затем данные (полностью или частично) подготавливаются для OLAP-анализа. Они могут быть загружены в специальную БД OLAP или оставлены в реляционном хранилище. Важнейшим его

элементом являются метаданные, т. е. информация о структуре, размещении и трансформации данных. Благодаря им обеспечивается эффективное взаимодействие различных компонентов хранилища.



Рисунок 1. Структура хранилища данных

Подытоживая, можно определить OLAP как совокупность средств многомерного анализа данных, накопленных в хранилище. Теоретически средства OLAP можно применять и непосредственно к оперативным данным или их точным копиям (чтобы не мешать оперативным пользователям). Но тут есть риск наступить на уже описанные выше грабли, т. е. начать анализировать оперативные данные, которые напрямую для анализа непригодны.

1.2 OLAP - Многомерное представление: куб

OLAP предоставляет удобные быстродействующие средства доступа, просмотра и анализа деловой информации. Пользователь получает естественную, интуитивно понятную модель данных, организуя их в виде многомерных кубов (Cubes). Осями многомерной системы координат служат основные атрибуты анализируемого бизнес-процесса. Например, для продаж это могут быть товар, регион, тип покупателя. В качестве одного из измерений используется время. На пересечениях осей - измерений (Dimensions) - находятся данные, количественно характеризующие процесс - меры (Measures). Это могут быть объемы продаж в штуках или в

денежном выражении, остатки на складе, издержки и т. п. Пользователь, анализирующий информацию, может "разрезать" куб по разным направлениям, получать сводные сведения и осуществлять прочие манипуляции, которые ему придут в голову в процессе анализа.

В качестве мер в трехмерном кубе, изображенном на рисунке 2, использованы суммы продаж, а в качестве измерений - время, товар и магазин. Измерения представлены на определенных уровнях группировки: товары группируются по категориям, магазины - по странам, а данные о времени совершения операций - по месяцам. Чуть позже мы рассмотрим уровни группировки (иерархии) подробнее.

	США	Канада	Мексика
Напитки	10 000	2000	1 000
Продукты питания	5000	500	250
Прочие товары	5000	500	250

Рисунок 2. Пример куба

"Разрезание" куба. Даже трехмерный куб сложно отобразить на экране компьютера так, чтобы были видны значения интересующих мер. Что уж говорить о кубах с количеством измерений, большим трех? Для визуализации данных, хранящихся в кубе, применяются, как правило, привычные двумерные, т. е. табличные, представления, имеющие сложные иерархические заголовки строк и столбцов.

Двумерное представление куба можно получить, "разрезав" его поперек одной или нескольких осей (измерений): мы фиксируем значения всех измерений, кроме двух, - и получаем обычную двумерную таблицу. В горизонтальной оси таблицы (заголовки столбцов) представлено одно измерение, в вертикальной (заголовки строк) - другое, а в ячейках таблицы - значения мер. При этом набор мер фактически рассматрива-

ется как одно из измерений - мы либо выбираем для показа одну меру (и тогда можем разместить в заголовках строк и столбцов два измерения), либо показываем несколько мер (и тогда одну из осей таблицы займут названия мер, а другую - значения единственного "неразрезанного" измерения).

Взгляните на рисунок 3: здесь изображен 2-мерный срез куба для одной меры - Unit Sales (продано штук) и двух "неразрезанных" измерений - Store (Магазин) и Время (Time).

	США	Канада	Мексика
Январь	20 000	4000	2000
Февраль	30 000	6000	3000
Март	50 000	10 000	5000

Рисунок 3. Двумерный срез куба для одной меры

На рисунке 4 представлено лишь одно "неразрезанное" измерение - Store, но зато здесь отображаются значения нескольких мер - Unit Sales (продано штук), Store Sales (сумма продажи) и Store Cost (расходы магазина).

	США	Канада	Мексика
Unit Sales	2000	400	200
Store Sales	30 000	6000	3000
Store Cost	10 000	2000	1000

Рисунок 4. Двумерный срез куба для нескольких мер

Двумерное представление куба возможно и тогда, когда "неразрезанными" остаются и более двух измерений. При этом на осях среза (строках и столбцах) будут размещены два или более измерений "разрезаемого" куба (см. рисунок 5).

	Январь			Февраль		
	США	Канада	Мексика	США	Канада	Мексика
Unit Sales	500	100	50	800	100	50
Store Sales	7500	1500	750	7500	1500	750
Store Cost	2500	500	250	2500	500	250

Рисунок 5. Двумерный срез куба с несколькими измерениями на одной оси

Метки. Значения, откладываемые вдоль измерений, называются членами или метками (members). Метки используются как для разделения куба, так и для ограничения выби-
раемых данных - когда в измерении, остающемся "нераз-
резанным", интересуют не все значения, а их подмножество.
Например 3 города из нескольких десятков. Значения меток
отображаются в двумерном представлении куба как заголов-
ки строк и столбцов.

Иерархии и уровни. Метки могут объединяться в
иерархии, состоящие из одного или нескольких уровней
(levels). Например, метки измерения "Магазин" (Store) есте-
ственно объединяются в иерархию с уровнями: All (Мир),
Country (Страна), State (Штат), City (Город), Store (Магазин).

В соответствии с уровнями иерархии вычисляются агре-
гатные значения, например объем продаж для USA (уровень
"Country") или для штата California (уровень "State"). В од-
ном измерении можно реализовать более одной иерархии -
скажем, для времени: {Год, Квартал, Месяц, День} и {Год,
Неделя, День}.

1.3 Архитектура OLAP-приложений

Все, что говорилось выше про OLAP, по сути, относи-
лось к многомерному представлению данных. То, как данные
хранятся, грубо говоря, не волнует ни конечного пользовате-
ля, ни разработчиков инструмента, которым клиент пользует-
ся.

Многомерность в OLAP-приложениях может быть раз-
делена на три уровня:

- Многомерное представление данных - средства конеч-
ного пользователя, обеспечивающие многомерную визуали-
зацию и манипулирование данными; слой многомерного
представления абстрагирован от физической структуры дан-
ных и воспринимает данные как многомерные.

- Многомерная обработка - средство (язык) формулиро-
вания многомерных запросов (традиционный реляционный

язык SQL здесь оказывается непригодным) и процессор, умеющий обработать и выполнить такой запрос.

- Многомерное хранение - средства физической организации данных, обеспечивающие эффективное выполнение многомерных запросов.

Первые два уровня в обязательном порядке присутствуют во всех OLAP-средствах. Третий уровень, хотя и является широко распространенным, не обязателен, так как данные для многомерного представления могут извлекаться и из обычных реляционных структур; процессор многомерных запросов в этом случае транслирует многомерные запросы в SQL-запросы, которые выполняются реляционной СУБД.

Конкретные OLAP-продукты, как правило, представляют собой либо средство многомерного представления данных, OLAP-клиент (например, Pivot Tables в Excel 2000 фирмы Microsoft или ProClarity фирмы Knosys), либо многомерную серверную СУБД, OLAP-сервер (например, Oracle Express Server или Microsoft OLAP Services).

Слой многомерной обработки обычно бывает встроен в OLAP-клиент и/или в OLAP-сервер, но может быть выделен в чистом виде, как, например, компонент Pivot Table Service фирмы Microsoft.

1.4 Технические аспекты хранения данных

Как уже говорилось выше, средства OLAP-анализа могут извлекать данные и непосредственно из реляционных систем. Такой подход был более привлекательным в те времена, когда OLAP-серверы отсутствовали в прайс-листах ведущих производителей СУБД. Но сегодня и Oracle, и Informix, и Microsoft предлагают полноценные OLAP-серверы, и даже те IT-менеджеры, которые не любят разводить в своих сетях "зоопарк" из ПО разных производителей, могут купить (точнее, обратиться с соответствующей просьбой к руководству компании) OLAP-сервер той же марки, что и основной сервер баз данных.

OLAP-серверы, или серверы многомерных БД, могут хранить свои многомерные данные по-разному. Прежде чем рассмотреть эти способы, нам нужно поговорить о таком важном аспекте, как хранение агрегатов. Дело в том, что в любом хранилище данных - и в обычном, и в многомерном - наряду с детальными данными, извлекаемыми из оперативных систем, хранятся и суммарные показатели (агрегированные показатели, агрегаты), такие, как суммы объемов продаж по месяцам, по категориям товаров и т. п. Агрегаты хранятся в явном виде с единственной целью - ускорить выполнение запросов. Ведь, с одной стороны, в хранилище накапливается, как правило, очень большой объем данных, а с другой - аналитиков в большинстве случаев интересуют не детали, а обобщенные показатели. И если каждый раз для вычисления суммы продаж за год пришлось бы суммировать миллионы индивидуальных продаж, скорость, скорее всего, была бы неприемлемой. Поэтому при загрузке данных в многомерную БД вычисляются и сохраняются все суммарные показатели или их часть.

Но, как известно, за все надо платить. И за скорость обработки запросов к суммарным данным приходится платить увеличением объемов данных и времени на их загрузку. Причем увеличение объема может стать буквально катастрофическим - в одном из опубликованных стандартных тестов полный подсчет агрегатов для 10 Мб исходных данных потребовал 2,4 Гб, т. е. данные выросли в 240 раз! Степень "разбухания" данных при вычислении агрегатов зависит от количества измерений куба и структуры этих измерений, т. е. соотношения количества "отцов" и "детей" на разных уровнях измерения. Для решения проблемы хранения агрегатов применяются подчас сложные схемы, позволяющие при вычислении далеко не всех возможных агрегатов достигать значительного повышения производительности выполнения запросов.

Различные варианты хранения информации. Как детальные данные, так и агрегаты могут храниться либо в рели-

ционных, либо в многомерных структурах. Многомерное хранение позволяет обращаться с данными как с многомерным массивом, благодаря чему обеспечиваются одинаково быстрые вычисления суммарных показателей и различные многомерные преобразования по любому из измерений. Некоторое время назад OLAP-продукты поддерживали либо реляционное, либо многомерное хранение. Сегодня, как правило, один и тот же продукт обеспечивает оба этих вида хранения, а также третий вид - смешанный.

Каждый из этих способов имеет свои преимущества и недостатки и должен применяться в зависимости от условий - объема данных, мощности реляционной СУБД.

При хранении данных в многомерных структурах возникает потенциальная проблема "разбухания" за счет хранения пустых значений. Ведь если в многомерном массиве зарезервировано место под все возможные комбинации меток измерений, а реально заполнена лишь малая часть (например, ряд продуктов продается только в небольшом числе регионов), то большая часть куба будет пустовать, хотя место будет занято. Современные OLAP-продукты умеют справляться с этой проблемой.

2.1 Основное содержание OLAP-технологии

В 1993 году основоположник реляционного подхода к построению баз данных Эдгар Кодд с партнерами (Edgar Codd, математик и стипендиат IBM), опубликовали статью, инициированную компанией "Arbor Software" (сегодня известнейшая компания "Hyperion Solutions"), озаглавленную "Обеспечение OLAP (оперативной аналитической обработки) для пользователей-аналитиков", в которой сформулированы 12 особенностей технологии OLAP, которые впоследствии были дополнены еще шестью. Эти положения стали основным содержанием новой и очень перспективной технологии.

Основные особенности технологии OLAP (Basic):

- многомерное концептуальное представление данных;
- интуитивное манипулирование данными;
- доступность и детализация данных;
- пакетное извлечение данных против интерпретации;
- модели анализа OLAP;
- архитектура "клиент-сервер" (здесь OLAP доступен с рабочего стола);
- прозрачность (прозрачный доступ к внешним данным);
- многопользовательская поддержка.

Специальные особенности (Special):

- обработка неформализованных данных;
 - сохранение результатов OLAP: хранение их отдельно от исходных данных;
 - исключение отсутствующих значений;
 - обработка отсутствующих значений.
- Особенности представления отчетов (Report):
- гибкость формирования отчетов;
 - стандартная производительность отчетов;
 - автоматическая настройка физического уровня хранения данных.

Управление измерениями (Dimension):