

## ИНСТРУМЕНТАРИИ МНОГОМЕРНЫХ БАЗ ДАННЫХ

## MULTI-DIMENSIONAL DATABASE TOOLS

**A. Nazarov**  
**A. Khakhina**  
**Yu. Vlasov**

*Summary:* The article substantiates the basics and definitions of the concepts of the context of a multidimensional data model, a comparative analysis of the technological processes for developing multidimensional data is carried out: their advantages and benefits are revealed. Based on the results of the research, the most optimal technical means and technologies for the implementation of a multidimensional database were identified.

*Keywords:* multidimensional database, unstructured data, analysis, data storage technologies.

**Назаров Артем Александрович**

Санкт-Петербургский политехнический  
 университет Петра Великого  
 nazarov2.aa@edu.spbstu.ru

**Хахина Анна Михайловна**

д.т.н., профессор, Санкт-Петербургский  
 политехнический университет Петра Великого  
 anna-hahina@mail.ru

**Власов Юрий Николаевич**

д.т.н., профессор, Российская таможенная академия  
 pobeda-872vlasov@yandex.ru

*Аннотация.* В статье приведены основные понятия и определения из контекста многомерной модели данных, проведен сравнительный анализ существующих технологий разработки многомерной базы данных: выявлены их преимущества и недостатки. В результате исследования выявлены наиболее оптимальные технические средства и технологии для реализации многомерной базы данных.

*Ключевые слова:* многомерная база данных, неструктурированные данные, анализ, технологии хранения данных.

## Введение

В связи с постоянным расширением информационной базы и формированием больших данных в недрах ФТС России появляется потребность в хранении, анализе и обработке все большего числа данных. При этом данные зачастую являются неструктурированными. Многомерная база данных позволяет работать с неструктурированными данными единым образом, проводить качественный анализ и сравнение, а также выявлять зависимости между показателями.

Актуальность исследования обусловлена масштабированием анализа большого количества разнородных данных практически во все сферы жизни общества, а также заполнением освободившегося места после ухода большого числа иностранных программных средств, которые используются для построения многомерной модели данных.

## Основные понятия многомерной модели данных

Рассмотрение проблем проведения многомерного анализа данных предполагает обращение к понятию Online Analytical Processing (OLAP). Этот термин служит для того, чтобы описать модель представления данных и технологий их обработки. Условия, которым должны удовлетворять системы OLAP: многомерное представление данных, прозрачность, доступность, неограниченная размерность и число уровней агрегации [4].

В соответствие с многомерным подходом исследуемые данные представляются внутри многомерного аналитического пространства. Зачастую такое представление называют OLAP-кубом или гиперкубом (пример изображен на рис. 1). Измерения (размерности) принято визуализировать в виде ребра многомерного куба, а ячейка — атомарная структура куба, которая соответствует полному набору значений [7].

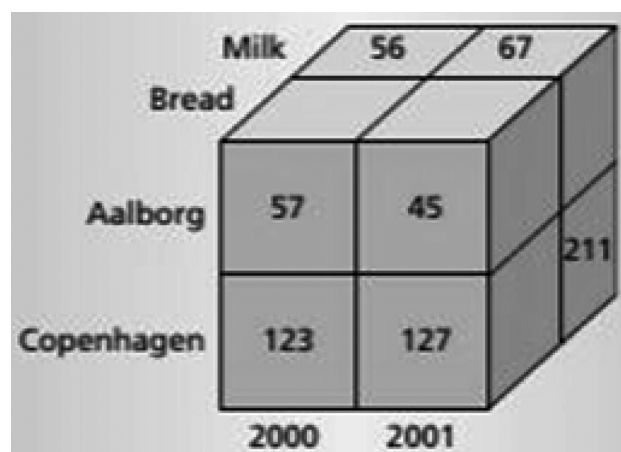


Рис. 1. Пример OLAP-куба

Для визуализации показателей в удобном для восприятия виде в многомерной модели вводятся следующие операции:

1. Поворот.
2. Проекция. При проекции значения в ячейках, лежащих на оси проекции, суммируются по определенному закону.

3. Раскрытие. Одно из значений измерения заменяется совокупностью из следующего уровня иерархии (например, год разбивается на 12 ячеек-месяцев).
4. Свертка. Это операция, обратная раскрытию.
5. Сечение.

Данные при этом организованы в хранилищах соответствующим образом согласно основным технологиям хранения данных.

### Основные технологии хранения данных

В научной литературе и практике выделяют 3 технологии хранения данных:

1. Multidimensional OLAP — детальные и агрегированные данные хранятся в многомерной базе данных. Такой способ позволяет работать с данными как с многомерным массивом, благодаря чему скорость вычисления одинакова для любого из измерений. Пример архитектуры приведен на рис. 2.
2. Relational OLAP — классическое хранение данных в виде реляционных таблиц. Трансформация в многомерный вид возможна благодаря исполь-

зованию промежуточного слоя метаданных. Пример архитектуры приведен на рис. 3.

3. Hybrid OLAP — попытка соединить преимущества технологий ROLAP и MOLAP. Исходные данные, по которым проводятся исследования, остаются в реляционной базе данных аналогично ROLAP. Агрегаты (получившиеся результаты) хранятся в многомерной базе данных подобно тому, как это реализовано в технологии MOLAP [6].

### Сравнительный анализ технологий хранения данных

1. MOLAP.

К преимуществам технологии MOLAP можно отнести:

- оптимизация запросов благодаря эффективному индексированию;
- возможность анализировать слабо структурированные данные без приведения к единообразному реляционному виду;
- простота отображения — не требуется сложная агрегация и множественные соединения таблиц для получения результирующего набора.

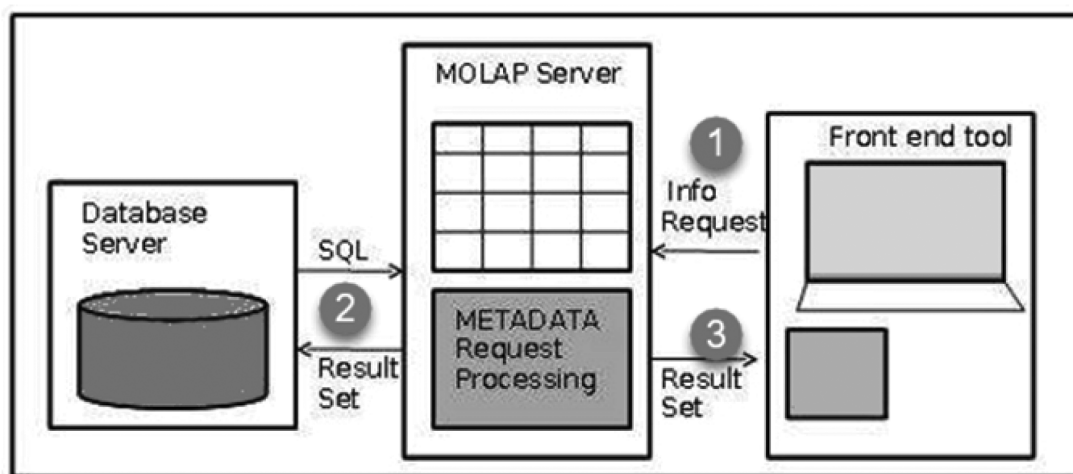


Рис. 2. Пример архитектуры Multidimensional OLAP

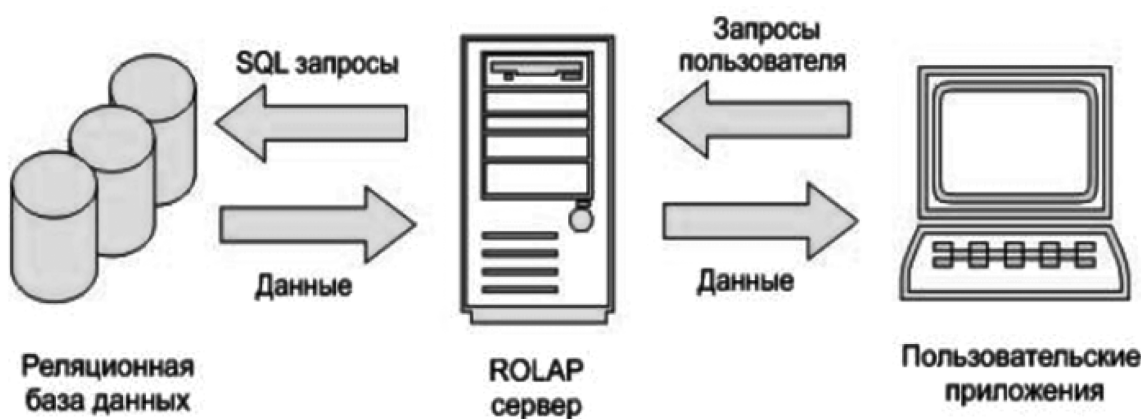


Рис. 3. Пример архитектуры Relational OLAP

К недостаткам можно отнести:

- низкая масштабируемость — со значительным ростом данных значительно усложняется агрегация;
- слабая поддержка разграничения прав.

2. ROLAP.

К основным преимуществам технологии ROLAP можно отнести:

- высокая масштабируемость;
- безопасность — сильная поддержка разграничения прав доступа.

Среди недостатков главный — меньшая производительность по сравнению с MOLAP. Однако, обосновано, что при определенной структуре реляционной базе данных (а именно, организации таблиц в схемы «звезда» или «снежинка») можно получить сравнимые результаты эффективности.

3. HOLAP.

Что касается HOLAP: с одной стороны, эта технология берет преимущество масштабируемости со стороны реляционной организации и оптимизации запросов за счет индексирования со стороны многомерной организации. Однако, подобная технология обладает существенностью как логической, так и технической сложностью, поэтому использование подобных технологий реализовано только самыми мощными программными средствами в области многомерности: продуктов компаний Oracle и Microsoft [5].

**Подготовка и трансформация данных**

В зависимости от технологии хранения данных (рассмотрены выше) существуют разные способы подготовки и трансформации данных в многомерный вид. В случае реализации технологии MOLAP данные уже хранятся в виде многомерного массива, что позволяет одинаково эффективно обращаться и агрегировать результаты по любым измерениям (как пример, СУБД Greenplum на основе PostgreSQL). Если же выбрать технологию ROLAP с целью добиться лучшей масштабируемости и лучшего разграничения прав доступа — то существуют специальные инструменты-трансформаторы, которые позволяют данные, хранящиеся в каком-либо хранилище, собрать и обрабатывать как многомерный гиперкуб.

На примере технологии ROLAP будет рассмотрен полноценный алгоритм реализации многомерной базы данных:

1. Создание модели данных. Формирование структуры реляционной базы данных.
2. Подготовка данных к обработке. Нормализация реляционной базы данных и приведение схемы к формату «звезда» или «снежинка».

3. Настройка связи между многомерным и реляционным представлением. Создание источника и настройка подключения к базе для создания уровней обобщения и агрегации.
4. Создание каркаса многомерной модели.
5. Визуализация результатов запросов. Визуализировать можно в виде диаграмм или отчетов.

Далее будут рассмотрены средства для реализации алгоритма: будет произведены сравнительные анализы СУБД для хранения и подготовки данных, а также специального программного обеспечения, позволяющего осуществить миграцию данных в многомерный вид.

**Сравнительный анализ СУБД для подготовки данных**

Для анализа СУБД требуется определить критерии сравнения. Именно по ним и будет происходить выявление преимуществ и недостатков средств друг перед другом. Из рассмотренных Хахиной А. и Бровченко Е. характеристик можно выделить следующие: быстродействие, доступность, отказоустойчивость, кроссплатформенность, наличие аудита и логирования (результаты сравнения приведены в таблице 1) [2].

Таблица 1.

Результаты сравнения БД

СУБД	Быстродействие	Доступность	Отказоустойчивость	Кроссплатформенность	Аудит
PostgreSQL	+/-	+/+	+	+	+
Oracle	+	+/-	+	+	+
MySQL	+/-	+/+	+	+	+
MSSQL	+	+/-	+	+/-	+

Под быстродействием подразумеваются временные характеристики при тестировании запросов UPDATE, SELECT, INSERT. Оценка основана на работе Мишкина А. и Крысина И., где авторы сравнивают скорости работы систем на основе выполнения основных запросов указанного выше формата и выявляют, что Oracle и MSSQL показывают лучшие результаты, чем PostgreSQL и MySQL [1]. Что касается доступности, результат основан на наличии качественной документации на официальном сайте, а также возможности скачать полноценную версию. Критерий является приоритетным, поскольку качественная разработка подразумевает достаточный уровень комплектации системы. По данному критерию лидируют PostgreSQL и MySQL. Все упомянутые СУБД обладают достаточной отказоустойчивостью, кроссплатформенностью (за исключением MSSQL, которая поддерживает только Windows и Linux), все обладают возможностью логирования и аудита.

**Сравнительный анализ инструментов для трансформации данных**

Для сравнительного анализа инструментов для трансформации данных были выбраны следующие критерии: доступность (возможность установить и подключить по крайней мере пробную версию с достаточным набором функций), скорость формирования гиперкуба для последующей аналитики, наличие стабильной поддержки иерархий и уровней агрегации, универсальность, а также объем места для установки на личный компьютер.

Для сравнения были выбраны следующие инструменты: Microsoft Analysis Services, Oracle Olap, а также российский аналог — ИнфоВизор, включающий в себя средство формирования гиперкуба ИнфоВизор Олар-дизайнер и средство визуализации результатов — ИнфоВизор Web-Аналитик. Кратко комплект и принцип работы ИнфоВизора изображен на рис. 4 [8].

Далее каждый инструмент будет рассмотрен подробнее с детализацией преимуществ и недостатков.

1. Analysis Services — входит в пакет MSSQL. На стороне сервера должна быть установлена СУБД MSSQL, на стороне клиента — VisualStudio и табличный процессор MS Excel. К достоинствам можно отнести сравнительно быстрое формирование гиперкуба. К значительным недостаткам можно отнести отсутствие универсальности в выборе источника данных (помимо баз данных источника Microsoft можно использовать только Oracle), а также необходимый большой объем свободного места на диске (помимо СУБД требуется ставить Visual Studio с доп. модулями).
2. Oracle Olap — является частью СУБД Oracle. Аналогично на сервере требуется наличие соответствующей СУБД, на стороне клиента — специализированного ПО для визуализации и Oracle Java — среды выполнения. Среди достоинств можно отметить также сравнительно быстрое формирование гиперкуба. Что касается недостатков, это все та же неуниверсальность (допускается только 1 источник), а также существует слабая поддержка иерархий (1–2 уровня) [2].

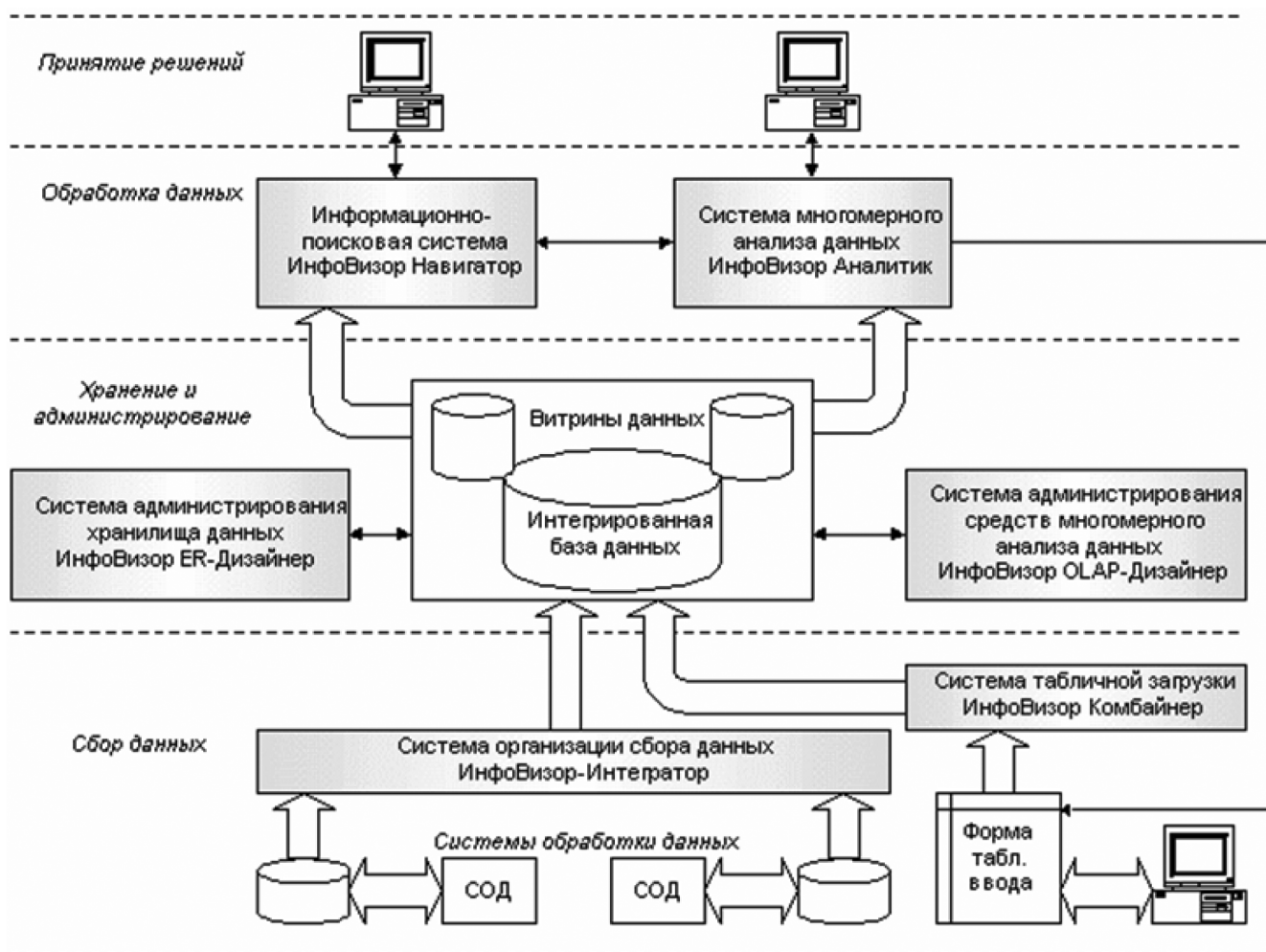


Рис. 4. Комплект и принцип работы ИнфоВизора



3. ИнфоВизор Оlap-дизайнер — средство для трансформации данных из реляционного формата в многомерный с последующей визуализацией компонентами ИнфоВизора. На сервере требуется наличие СУБД, на стороне клиента — достаточно настроенного ИнфоВизора, дополнительных сред или программ устанавливать не требуется. К достоинствам безусловно относится универсальность, отсутствие привязки к конкретной СУБД, на текущий момент уже реализовано подключение к MySQL, PostgreSQL, возможно настроить драйвер и подключиться к другим СУБД (например, Oracle, MsSQL), а также небольшой объем необходимого места на диске для установки. К недостаткам можно отнести меньшую производительность (ориентируюсь на данные из документации, сравнительного анализа с при-

веденными выше инструментами в открытом доступе не было).

Сравнение приводится в таблице 2.

Таблица 2.

Сравнение БД

ПО	Доступность	Скорость	Поддержка	Универсальность	Место
Analysiservices	–	+	+	–	+-
OracleOlap	–	+	+-	–	+-
ИнфоВизор	+	+-	+	+	+

Очевидно, что по всем основным параметрам опережает другие базы данных ИнфоВизор.

#### ЛИТЕРАТУРА

1. Мишкин А., Крысин И. Сравнительный анализ СУБД при работе с большим объемом информации // Электронный журнал: наука, техника и образование. 2020. Т. 1, № 3. С. 21–26.
2. Мосин С.В. Методы и алгоритмы формирования многомерных данных с использованием промежуточных представлений / Диссертация на соискание ученой степени кандидата физико-математических наук.
3. Хахина А., Бровченко Е. Сравнительный анализ СУБД для разработки базы данных предприятия // Заметки ученого. 2022. Т. 1, № 5. С. 127–131.
4. Codd E., Codd S. Providing OLAP to User-Analysts: An IT Mandate. — Associates, 1993. 24 p.
5. Park C.-S., Kim M.-H., Lee Y.-J. Usability-based caching of query results in OLAP systems // Journal of Systems and Software. 2003. Vol. 68, no. 2. P. 103–119.
6. Vassiliadis P., Sellis T.A. Survey of Logical Models for OLAP Databases // SIGMOD Record. 1999. Vol. 28, no. 4. P. 64–69.
7. Сайт Национального Открытого Университета. — URL: <https://intuit.ru> (дата обращения: 13.11.2022).
8. Сайт Программного комплекса «ИнфоВизор». — URL: <http://www.infovisor.ivanovo.ru/> (дата обращения: 15.11.2022).

© Назаров Артем Александрович (nazarov2.aa@edu.spbstu.ru); Хахина Анна Михайловна (anna-hahina@mail.ru); Власов Юрий Николаевич (pobeda-872vlasov@yandex.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»