

Использование колоночных СУБД для формирования многомерных структур данных

Д. С. Куницкий*, М. Б. Фомин

ФГАОУ ВО «Российский университет дружбы народов имени Патриса Лумумбы», г. Москва, Российская Федерация

Адрес: 117198, Российская Федерация, г. Москва, ул. Миклухо-Маклая, д. 6

* 1142221031@rudn.ru

Аннотация

Работа посвящена анализу формирования многомерных информационно-аналитических систем с использованием данных, накопленных в колоночных СУБД. В последнее время такие СУБД получают распространение в системах, нацеленных на хранение и анализ больших объемов информации, генерируемой в результате функционирования интенсивных бизнес-процессов. В процессе анализа многомерных данных требуется производить выборку данных фактов, ассоциированных с сочетаниями значений нескольких измерений. В случае использования колоночных СУБД запросы с участием данных нескольких колонок могут быть организованы таким образом, что каждая колонка в значительной степени может рассматриваться как обособленный источник данных. Это обстоятельство позволяет облегчить переход к параллельной обработке информации. Учитывая, в дополнение к указанным свойствам, высокую эффективность чтения в колоночных СУБД, при выборе указанного подхода можно рассчитывать на эффективное решение задачи формирования многомерных структур данных и упрощение работы по обслуживанию существующих кубов данных в случае необходимости их модификации. Использование колоночных СУБД в многомерном анализе данных может оказаться эффективным в следующих случаях. Когда хранилище данных состоит из широких, многоаспектных таблиц данных с большим количеством колонок, над которыми выполняются сложные операции (агрегации, фильтрация, сортировки). В ситуации, когда бизнес-процесс генерирует большой поток входящих в информационную систему данных (миллиарды событий день, генерируемые в реальном времени).

Ключевые слова: хранилище данных, многомерная модель данных, витрина данных, OLAP, колоночная СУБД

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Для цитирования: Куницкий Д. С., Фомин М. Б. Использование колоночных СУБД для формирования многомерных структур данных // Современные информационные технологии и ИТ-образование. Современные информационные технологии и ИТ-образование. 2023. Т. 19, № 3. С. 607-613. <https://doi.org/10.25559/SITITO.019.202303.607-613>

© Куницкий Д. С., Фомин М. Б., 2023



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.



The Use of Column-Based DBMSS to Form Multidimensional Data Structures

D. S. Kunitskiy*, M. B. Fomin

Peoples' Friendship University of Russia named after Patrice Lumumba, Moscow, Russian Federation
Address: 6 Miklukho-Maklaya St., Moscow 117198, Russian Federation

* 1142221031@rudn.ru

Abstract

The work is devoted to the analysis of the method of designing information systems using the "Data vault" modeling technology. The described approach can be used when building a data warehouse within a multidimensional information system based on the classical approach of data warehouse design – a 3-level data representation architecture, which includes data preparation area, or operational data warehouse, data warehouse, and thematic data marts. This approach allows you to organize data storage within a data warehouse using a metadata repository based on the multidimensional approach. The metadata repository is responsible for the process of collecting data, storing data, and presenting data for analysis. The "Data vault" approach allows you to design the data warehouse of an information system based on a metamodel that is semantically related to the domain of the system and is easily rebuilt in the event of changes in the business model of the domain. When describing a metamodel, the first-order predicate calculus language is used, which makes it possible to control the metamodel using a declarative programming style – the "Prolog" language. The key point in the structure of the information system is the way of transition from the "Data vault" model to a multidimensional model of data representation based on associative rules of dependence between information objects.

Keywords: data warehouse, multidimensional data model, data mart, OLAP, column-based DBMS

Conflict of interests: The authors declare no conflict of interest.

For citation: Kunitskiy D. S., Fomin M. B. The Use of Column-Based DBMSS to Form Multidimensional Data Structures. *Modern Information Technologies and IT-Education*. 2023;19(3):607-613. <https://doi.org/10.25559/SITITO.019.202303.607-613>



Введение

В недавнем прошлом, когда говорилось о стремительном росте числа реализаций информационных систем, прежде всего имелись в виду системы, предназначенные для оперативной обработки данных (Online Transaction Processing, OLTP) [1]. Для построения таких систем используются реляционные системы управления базами данных (СУБД). Такие СУБД оказались абсолютными победителями и заняли доминирующие позиции в компьютерной индустрии, позволяя успешно решать задачи автоматизации отдельных процессов. Структуры данных и методы их организации в реляционных базах данных были оптимальны для накопления и обработки информации, они оказались далеки от оптимальных для анализа накопленной информации. Большие объемы данных сложно хранить и еще сложнее анализировать, т. к. в процессе поиска решений возникает необходимость в построении зависимостей между различными параметрами. Кроме того, число таких параметров может варьироваться в широких пределах. Так возникла потребность в системах, призванных обеспечить надежный и быстрый доступ к большому объему накопленной информации в целях проведения интерактивной аналитической обработки (Online Analytic Processing, OLAP) [2-7].

Модели многомерного представления данных

Реляционный подход к проектированию баз данных стал широко известен благодаря работам Эдгара Ф. Кодда, впервые опубликованным в 1970 году. Концепция OLAP была также предложена Коддом в 1993 году. В статье [1] он рассмотрел недостатки реляционной модели, указав в первую очередь на невозможность «объединять, просматривать и анализировать данные с точки зрения множественности измерений», а также определил 12 требований, которым должны удовлетворять система OLAP. В 1995 году свод правил Кода был переработан Н. Пендсом и Р. Критом в так называемый тест FASMI (Fast Analysis of Shared Multidimensional Information).

При работе с реляционными базами данных используется двумерное пространство — таблица с записями (строками) и полями (колонками). Информационная модель OLAP-системы описывается через многомерное пространство, которое может иметь любое количество измерений, и эти измерения не обязаны быть одинакового (или даже похожего) размера. Для определения многомерного пространства используется термин куб (cube). Множество всех измерений куба образует систему координат представляемого пространства данных. Многомерный куб — это многомерная структура, состоящая из множества ячеек и хранящая взаимосвязанные данные, описывающие предметную область или ее составляющую. Ячейка (cell) является атомарной структурой куба и определяются набором значений измерений куба [8].

С момента появления технологии OLAP она подвергалась постоянному развитию и совершенствованию. Поскольку исторически системы анализа данных возникли из реляционных СУБД, они изначально унаследовали подходы к проектированию, принятые в этой области. Для построения многомерного куба стала использоваться реляционная схема «Звезда», удобная для хранения многомерных данных. При этом таблица фактов (центр звезды) содержит агрегированные данные для составления отчетов, а денормализованные таблицы измерений (лучи) описывают хранимые данные. Модификацию звездной схемы, когда отдельные таблицы измерений нормализованы с целью сокращения избыточности и определения всех зависимостей, называют «Снежинкой». Она использует меньше дискового пространства и лучше сохраняет целостность данных, но ей свойственен такой недостаток, как сложность запросов из-за увеличенного числа соединений таблиц, что, в свою очередь, ведет к росту издержек производительности¹.

Большинство СУБД специализируются на каком-то одном виде обработки данных, но в то же время они имеют инструменты для выполнения и других операций, если это необходимо. В случае, если СУБД сначала развивались исключительно как OLAP или как OLTP, перед разработчиками может появиться задача объединения подходов OLTP и OLAP в рамках СУБД, подразумевающей одинаково эффективное выполнение обоих видов операций по обработке данных [9]. Такая задача может быть решена путем замены строково-ориентированной реляционной базы данных как способа построения многомерного пространства на колоночно-ориентированную базу данных, содержащую широкие, многоаспектные таблицы данных с большим количеством колонок, над которыми выполняются сложные операции (агрегации, фильтрация, сортировки). При этом осуществление такой замены не подразумевает отказ от использования многомерного пространства для реализации естественного для инженера данных способа анализа данных и описания запросов. Многомерное пространство в этом случае строится через виртуальный слой над небольшим набором широких таблиц [10-18].

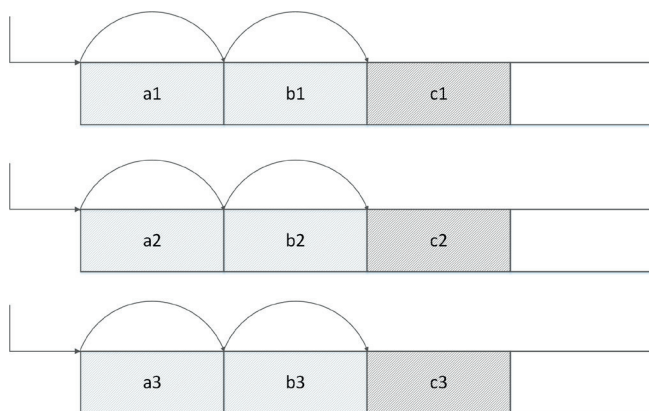
Появление и популяризация колоночных СУБД (Google BigQuery, запущенная в 2010 году, Amazon Redshift, 2012 год, и Yandex ClickHouse, 2014 год) стало технологическим прорывом, сильно отразившимся на технологиях обработки данных. Чтобы охарактеризовать отличие колоночных баз данных от строковых с точки зрения практического использования, рассмотрим, как данные хранятся в долговременной памяти и как они считываются. В долговременной памяти данные на самом низком уровне организованы в блоки, которые представляют собой наименьшие объемы информации, обрабатываемые за один такт. Если искомые данные хранятся в меньшем количестве блоков, СУБД данных будет работать намного быстрее.

¹ Бежитская Е. А., Сорокина А. В., Черкашин Е. А. Возможности OLAP-технологий в многомерном анализе данных // Решетневские чтения : Материалы XXVII Международной научно-практической конференции, посвященной памяти генерального конструктора ракетно-космических систем академика М. Ф. Решетнева, Красноярск, 08-10 ноября 2023 г. Красноярск : СибГУ, 2023. С. 201-202. EDN: UUZXYI; Моделирование OLAP-куба: реализация гиперкуба / М. А. Азарян [др.] // Аллея науки. 2022. № 1(64). С. 1278-1284. EDN: VQIGFT



Хранение данных, ориентированное на строки

В базах данных, ориентированных на строки, запись выполняется построчно, т. е. блоки долговременной памяти будут заполняться разрозненными значениями множества колонок записываемых строк [19]. Чтение записанных таким образом данных будет состоять из следующих шагов: перейти на первую строку; найти колонку с требуемым именем и ее значение; отобразить строку, если требуемое значение найдено; перейти к следующей строке. Последовательности перечисленных действий для нескольких строк можно наглядно проследить на рис. 1.



Р и с. 1. Чтение данных в строчной базе данных
F i g. 1. Reading data in a row database

Источник: здесь и далее в статье все рисунки составлены авторами.
Source: Hereinafter in this article all figures were made by the authors.

Использование СУБД, ориентированных на строки, эффективно, когда необходимо получить доступ к большому числу колонок одновременно. Как следствие, не рекомендуется формировать очень широкие таблицы, т. к. маловероятно, что к ним потребуется постоянный доступ. Если же одновременно будут нужны данные всего нескольких колонок, то ориентированное на строки хранение приводит к необходимости считывания избыточной информации.

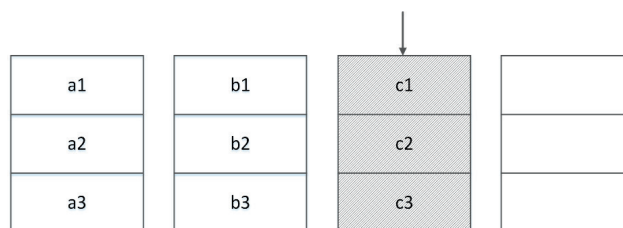
Хранение данных, ориентированное на колонки

Можно выделить три основных преимущества [19-23], за счет которых достигается высокая производительность в случае использования метода хранения данных в колонках:

1. Блоки долговременной памяти будут заполняться данными одной колонки. Соответственно и чтение будет выполняться блоками, заполненными значениями одной колонки. Это означает, что все значения одного атрибута сущности будут сгруппированы вместе. Такой принцип хранения данных позволяет существенно повышать выполнение запросов, затрагивающих только подмножество колонок, доступных для чтения.

2. Лучшее сжатие по сравнению с реляционными базами данных на основе строк. Когда данные одного типа хранятся вместе, можно сжать их намного лучше, чем в случае хранения разнотипных фрагментов информации, даже если это происходит за счет вычислений по распаковке во время выполнения чтения [11]. Как следствие, при выполнении запроса агрегации в память может быть загружено больше данных, что, в свою очередь, приводит к более быстрому выполнению общего запроса.

3. Последнее преимущество заключается в том, что сжатие в колоночных базах данных позволяет вводить разреженные индексы, позволяющие быстрее выполнять запросы с большими диапазонами фильтрации. Для таблицы из представленного выше примера хранение и чтение в случае использования колоночной СУБД изображено на рисунке 2.



Р и с. 2. Чтение данных в колоночной базе данных
F i g. 2. Reading data in a columnar database

Однако есть и обратная сторона: низкая производительность обновления колоночной базы данных (чтобы обновить одну «строку», требуется выполнить операцию в каждой из колонок). В результате многие современные колоночные базы данных ограничивают возможность обновления данных после их сохранения [24, 25].

Заключение

В работе был рассмотрен подход формирования многомерных информационно-аналитических систем путем замены строково-ориентированной реляционной базы данных на колоночно-ориентированную базу данных. Данные в колоночной базе хранятся не по строкам, а по столбцам, что позволяет сократить количество блоков, необходимых для обработки данных, и ускорить выполнение запросов. В итоге колоночные базы данных могут быть использованы для обработки данных в реальном времени и быстрого выполнения аналитических запросов на больших объемах данных. Также колоночные базы данных позволяют создавать многомерные пространства данных, для которых обеспечивается одинаковая эффективность выполнения операций чтения и записи данных. Таким образом, замена реляционной базы данных на колоночно-ориентированную базу данных является важным шагом в развитии СУБД, позволяющим обеспечить эффективное выполнение широкого спектра операций по аналитической обработке данных.



Список использованных источников

- [1] Bellatreche L., Cuzzocrea A., Song I.-Y. Advances in data warehousing and OLAP in the big Data Era // *Information Systems*. 2015. Vol. 53. P. 39-40. <https://doi.org/10.1016/j.is.2015.06.001>
- [2] Francia M., Gallinucci E., Golfarelli M. COOL: A framework for conversational OLAP // *Information Systems*. 2022. Vol. 104. Article number: 01752. <https://doi.org/10.1016/j.is.2021.101752>
- [3] Honcharenko T., Terentyev O., Gorbatyuk E. Mathematical Modeling of Information System Designing Master Plan of the Building Territory Based on OLAP Technology // *International scientific-practical conference*. 2022. P. 3-15. https://doi.org/10.1007/978-3-030-89902-8_1
- [4] Cuzzocrea A. Privacy-Preserving OLAP via Modeling and Analysis of Query Workloads: Innovative Theories and Theorems // *SSDBM '23: Proceedings of the 35th International Conference on Scientific and Statistical Database Management*. New York, USA: Association for Computing Machinery, 2023. P. 1-12. <https://doi.org/10.1145/3603719.3603735>
- [5] Fonseca R., de Carvalho Victorino M., Holanda M. ROLAP DW transformation proposal for OLAP architecture in NoSQL database // *EATIS '20: Proceedings of the 10th Euro-American Conference on Telematics and Information Systems*. New York, USA: Association for Computing Machinery, 2020. P. 1-7. <https://doi.org/10.1145/3401895.3401899>
- [6] Scalable real-time OLAP on cloud architectures / F. Dehne [et al.] // *Journal of Parallel and Distributed Computing*. 2015. Vol. 79-80. P. 31-41. <https://doi.org/10.1016/j.jpdc.2014.08.006>
- [7] Вычислительная производительность методов редукции гиперкубов многомерных данных аналитических OLAP-систем / А. А. Ахрем [и др.] // *Искусственный интеллект и принятие решений*. 2019. № 4. С. 23-28. <https://doi.org/10.14357/20718594190403>
- [8] Doka K., Tsoumakos D., Koziris N. Online querying of d-dimensional hierarchies // *Journal of Parallel and Distributed Computing*. 2011. Vol. 71, № 3. P. 424-437. <https://doi.org/10.1016/j.jpdc.2010.10.005>
- [9] Bimonte S., Coulibaly F. A., Rizzi S. An approach to on-demand extension of multidimensional cubes in multi-model settings: Application to IoT-based agro-ecology // *Data & Knowledge Engineering*. 2024. Vol. 150. Article number: 102267. <https://doi.org/10.1016/j.datak.2023.102267>
- [10] Aggregation and Exploration of High-Dimensional Data Using the Sudokube Data Cube Engine / S. Basil John [et al.] // *SIGMOD '23: Companion of the 2023 International Conference on Management of Data*. New York, USA: Association for Computing Machinery, 2023. P. 175-178. <https://doi.org/10.1145/3555041.3589729>
- [11] Columnar NoSQL CUBE: Agregation operator for columnar NoSQL data warehouse / K. Dehdouh [et al.] // *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. San Diego, CA, USA: IEEE Computer Society, 2014. P. 3828-3833. <https://doi.org/10.1109/SMC.2014.6974527>
- [12] Efficient querying of multidimensional RDF data with aggregates: Comparing NoSQL, RDF and relational data stores / F. Ravat [et al.] // *International Journal of Information Management*. 2020. Vol. 54. Article number: 102089. <https://doi.org/10.1016/j.ijinfomgt.2020.102089>
- [13] Агрегация показателей в OLAP-кубе / А. Б. Чемидова [и др.] // *Российский экономический интернет-журнал*. 2019. № 4. С. 138. EDN: XFTSUA
- [14] Попов С. Г., Лисенкова А. А. Алгоритмы динамической генерации MDX-запросов к многомерным OLAP-кубам // *Информатика. Телекоммуникации и управление*. 2018. Т. 11, № 4. С. 21-35. <https://doi.org/10.18721/JCSTCS.11402>
- [15] Анализ вычислительной сложности методов декомпозиции OLAP-гиперкубов многомерных данных / А. А. Ахрем [и др.] // *Математика и математическое моделирование*. 2020. № 4. С. 52-64. <https://doi.org/10.24108/mathm.0420.0000221>
- [16] Носов А. П., Ахрем А. А., Рахманкулов В. З. Анализ эффективности декомпозиции OLAP-гиперкубов данных для методов экспоненциальной вычислительной сложности // *Математика и математическое моделирование*. 2021. № 3. С. 29-45. <https://doi.org/10.24108/mathm.0321.0000258>
- [17] Ахрем А. А., Носов А. П., Рахманкулов В. З. Анализ эффективности методов полиномиальной степени сложности при декомпозиции OLAP-кубов многомерных данных // *Математика и математическое моделирование*. 2021. № 1. С. 27-42. <https://doi.org/10.24108/mathm.0121.0000244>
- [18] Бедняк С. Г., Головин А. В., Захарова О. И. Разработка математической модели OLAP-куба с помощью 1С и Pentaho Bi // *Информационные системы и технологии*. 2020. № 5(121). С. 44-49. EDN: VARFQQ
- [19] Cuzzocrea A., Simitsis A., Song I.-Y. Big Data Management: New Frontiers, New Paradigms // *Information Systems*. 2017. Vol. 63. P. 63-65. <https://doi.org/10.1016/j.is.2016.07.002>
- [20] Building a novel physical design of a distributed big data warehouse over a Hadoop cluster to enhance OLAP cube query performance / Y. Ramdane [et al.] // *Parallel Computing*. 2022. Vol. 111. Article number: 102918. <https://doi.org/10.1016/j.parco.2022.102918>
- [21] Cuzzocrea A., Hafsaoui A., Leung C. K. Machine-Learning-Based Multidimensional Big Data Analytics over Clouds via Multi-Columnar Big OLAP Data Cube Compression // *2023 IEEE International Conference on Big Data (BigData)*. Sorrento, Italy: IEEE Computer Society, 2023. P. 5206-5212. <https://doi.org/10.1109/BigData59044.2023.10386560>
- [22] Ordonez C., Garcia-Alvarado C., Song I.-Y. Special issue on DOLAP 2015: Evolving data warehousing and OLAP cubes to big data analytics // *Information Systems*. 2017. Vol. 68. P. 1-2. <https://doi.org/10.1016/j.is.2017.03.006>



- [23] Hose K., Romero O., Song I.-Y. Trends in Design, Optimization, Languages, and Analytical Processing of Big Data (DOLAP 2020) // *Information Systems*. 2022. Vol. 104. Article number: 101929. <https://doi.org/10.1016/j.is.2021.101929>
- [24] The Design and Implementation of Modern Column-Oriented Database Systems / D. Abadi [et al.] // *Foundations and Trends in Databases*. 2012. Vol. 5, no. 3. P. 197-280. <http://dx.doi.org/10.1561/1900000024>
- [25] Shioi T., Hatano K. Rule- and Cost-Based Optimization of OLAP Workloads on Distributed RDBMS with Column-Oriented Storage Function // 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW). Vienna, Austria: IEEE Computer Society, 2016. P. 165-170. <https://doi.org/10.1109/W-FiCloud.2016.44>

Поступила 10.08.2023; одобрена после рецензирования 25.09.2023; принята к публикации 06.10.2023.

Об авторах:

Куницкий Дмитрий Сергеевич, аспирант кафедры математического моделирования и искусственного интеллекта факультета физико-математических и естественных наук, ФГАОУ ВО «Российский университет дружбы народов имени Патриса Лумумбы» (117198, Российская Федерация, г. Москва, ул. Миклухо-Маклая, д. 6), ORCID: <https://orcid.org/0009-0005-8899-3055>, 1142221031@rudn.ru

Фомин Максим Борисович, доцент факультета физико-математических и естественных наук, кандидат физико-математических наук, ФГАОУ ВО «Российский университет дружбы народов имени Патриса Лумумбы» (117198, Российская Федерация, г. Москва, ул. Миклухо-Маклая, д. 6), кандидат физико-математических наук, ORCID: <https://orcid.org/0000-0002-7924-9743>, mfofin@sci.pfu.edu.ru

Все авторы прочитали и одобрили окончательный вариант рукописи.

References

- [1] Bellatreche L., Cuzzocrea A., Song I.-Y. Advances in data warehousing and OLAP in the big Data Era. *Information Systems*. 2015;53:39-40. <https://doi.org/10.1016/j.is.2015.06.001>
- [2] Francia M., Gallinucci E., Golfarelli M. COOL: A framework for conversational OLAP. *Information Systems*. 2022;104:01752. <https://doi.org/10.1016/j.is.2021.101752>
- [3] Honcharenko T., Terentyev O., Gorbatyuk E. Mathematical Modeling of Information System Designing Master Plan of the Building Territory Based on OLAP Technology. *International scientific-practical conference*. 2022;3-15. https://doi.org/10.1007/978-3-030-89902-8_1
- [4] Cuzzocrea A. Privacy-Preserving OLAP via Modeling and Analysis of Query Workloads: Innovative Theories and Theorems. In: SSDBM '23: Proceedings of the 35th International Conference on Scientific and Statistical Database Management. New York, USA: Association for Computing Machinery; 2023. p. 1-12. <https://doi.org/10.1145/3603719.3603735>
- [5] Fonseca R., de Carvalho Victorino M., Holanda M. ROLAP DW transformation proposal for OLAP architecture in NoSQL database. In: EATIS '20: Proceedings of the 10th Euro-American Conference on Telematics and Information Systems. New York, USA: Association for Computing Machinery; 2020. p. 1-7. <https://doi.org/10.1145/3401895.3401899>
- [6] Dehne F., Kong Q., Rau-Chaplin A., Zaboli H., Zhou R. Scalable real-time OLAP on cloud architectures. *Journal of Parallel and Distributed Computing*. 2015;79-80:31-41. <https://doi.org/10.1016/j.jpdc.2014.08.006>
- [7] Akhrem A.A., Nosov A.P., Rakhmankulov V.Z., Yuzhanin K.V. Computational Performance of Hypercube Reduction Methods for Multidimensional Data of Analytical OLAP System. *Artificial Intelligence and Decision Making*. 2019;(4):23-28. (In Russ., abstract in Eng.) <https://doi.org/10.14357/20718594190403>
- [8] Doka K., Tsoumakos D., Koziris N. Online querying of d-dimensional hierarchies. *Journal of Parallel and Distributed Computing*. 2011;71(3):424-437. <https://doi.org/10.1016/j.jpdc.2010.10.005>
- [9] Bimonte S., Coulibaly F.A., Rizzi S. An approach to on-demand extension of multidimensional cubes in multi-model settings: Application to IoT-based agro-ecology. *Data & Knowledge Engineering*. 2024;150:102267. <https://doi.org/10.1016/j.datak.2023.102267>
- [10] Basil John S., Lindner P., Jiang Z., Koch C. Aggregation and Exploration of High-Dimensional Data Using the Sudokube Data Cube Engine. In: SIGMOD '23: Companion of the 2023 International Conference on Management of Data. New York, USA: Association for Computing Machinery; 2023. p. 175-178. <https://doi.org/10.1145/3555041.3589729>
- [11] Dehdouh K., Bentayeb F., Boussaid O., Kabachi N. Columnar NoSQL CUBE: Aggregation operator for columnar NoSQL data warehouse. In: 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC). San Diego, CA, USA: IEEE Computer Society; 2014. p. 3828-3833. <https://doi.org/10.1109/SMC.2014.6974527>
- [12] Ravat F., Song J., Teste O., Trojahn C. Efficient querying of multidimensional RDF data with aggregates: Comparing NoSQL, RDF and relational data stores. *International Journal of Information Management*. 2020;54:102089. <https://doi.org/10.1016/j.ijinfomgt.2020.102089>
- [13] Chemidova A.B., Suvorov S.V., Tsarkova N.I., Zhilyaeva I.A. Aggregation of Indicators in OLAP-Cube. *Russian economic online journal*. 2019;(4):138. (In Russ., abstract in Eng.) EDN: XFTSUA
- [14] Popov S.G., Lisenkova A.A. Algorithms for MDX-query generation in multidimensional OLAP-cubes. *Computing, Telecommunication*



- and Control (St. Petersburg Polytechnical University Journal. Computer Science. Telecommunication and Control Systems. 2018;11(4):21-35. (In Russ., abstract in Eng.) <https://doi.org/10.18721/JCSTCS.11402>
- [15] Akhrem A.A., Nosov A.P., Rakhmankulov V.Z., Yuzhanin K.V. Computational complexity analysis of decomposition methods of OLAP hyper-cubes of multidimensional data. *Mathematics and Mathematical Modelling*. 2020;(4):52-64. (In Russ., abstract in Eng.) <https://doi.org/10.24108/mathm.0420.0000221>
- [16] Nosov A.P., Akhrem A.A., Rakhmankulov V.Z. Efficiency Analysis of OLAP-data Hypercube Decomposition for Exponential Computational Complexity Methods. *Mathematics and Mathematical Modelling*. 2021;(3):29-45. (In Russ., abstract in Eng.) <https://doi.org/10.24108/mathm.0321.0000258>
- [17] Akhrem A.A., Nosov A.P., Rakhmankulov V.Z. Analysing Efficiency Methods of Polynomial Complexity Degree in Multidimensional OLAP Cube Data Decomposition. *Mathematics and Mathematical Modelling*. 2021;(1):27-42. (In Russ., abstract in Eng.) <https://doi.org/10.24108/mathm.0121.0000244>
- [18] Bednyak S.G., Golovin A.V., Zaxarova O.I. Development of a Mathematical Model of an OLAP Cube Using 1C and Pentaho BI. *Information Systems and Technologies*. 2020;(5):44-49. (In Russ., abstract in Eng.) EDN: VARFQQ
- [19] Cuzzocrea A., Simitsis A., Song I.-Y. Big Data Management: New Frontiers, New Paradigms. *Information Systems*. 2017;63:63-65. <https://doi.org/10.1016/j.is.2016.07.002>
- [20] Ramdane Y., Boussaid O., Boukraà D., Kabachi N., Bentayeb F. Building a novel physical design of a distributed big data warehouse over a Hadoop cluster to enhance OLAP cube query performance. *Parallel Computing*. 2022;111:102918. <https://doi.org/10.1016/j.parco.2022.102918>
- [21] Cuzzocrea A., Hafsaoui A., Leung C.K. Machine-Learning-Based Multidimensional Big Data Analytics over Clouds via Multi-Columnar Big OLAP Data Cube Compression. In: 2023 IEEE International Conference on Big Data (BigData). Sorrento, Italy: IEEE Computer Society; 2023. p. 5206-5212. <https://doi.org/10.1109/BigData59044.2023.10386560>
- [22] Ordóñez C., García-Alvarado C., Song I.-Y. Special issue on DOLAP 2015: Evolving data warehousing and OLAP cubes to big data analytics. *Information Systems*. 2017;68:1-2. <https://doi.org/10.1016/j.is.2017.03.006>
- [23] Hose K., Romero O., Song I.-Y. Trends in Design, Optimization, Languages, and Analytical Processing of Big Data (DOLAP 2020). *Information Systems*. 2022;104:101929. <https://doi.org/10.1016/j.is.2021.101929>
- [24] Abadi D., Boncz P., Harizopoulos S., Idreos S., Madden S. The Design and Implementation of Modern Column-Oriented Database Systems. *Foundations and Trends in Databases*. 2013;5(3):197-280. <http://dx.doi.org/10.1561/19000000024>
- [25] Shioi T., Hatano K. Rule- and Cost-Based Optimization of OLAP Workloads on Distributed RDBMS with Column-Oriented Storage Function. In: 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW). Vienna, Austria: IEEE Computer Society; 2016. p. 165-170. <https://doi.org/10.1109/W-FiCloud.2016.44>

Submitted 10.08.2023; approved after reviewing 25.09.2023; accepted for publication 06.10.2023.

About the authors:

Dmitriy S. Kunitskiy, Postgraduate Student of the Department of Mathematical Modeling and Artificial Intelligence, Faculty of Science, Peoples' Friendship University of Russia named after Patrice Lumumba (6 Miklukho-Maklaya St., Moscow 117198, Russian Federation), **ORCID:** <https://orcid.org/0009-0005-8899-3055>, 1142221031@rudn.ru

Maxim B. Fomin, Associate Professor of Faculty of Science, Peoples' Friendship University of Russia named after Patrice Lumumba (6 Miklukho-Maklaya St., Moscow 117198, Russian Federation), Cand. Sci. (Phys.-Math.), **ORCID:** <https://orcid.org/0000-0002-7924-9743>, mfomin@sci.pfu.edu.ru

All authors have read and approved the final manuscript.

