



Международный журнал информационных технологий и энергоэффективности

Сайт журнала:

<http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004.65

ОПТИМИЗАЦИЯ АРХИТЕКТУРЫ БАЗ ДАННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ ПРИ ОБРАБОТКЕ БОЛЬШИХ ОБЪЕМОВ ИНФОРМАЦИИ

¹Гунько А.С., Мусихин А.Г.

ФГБОУ ВО «МИРЭА - РОССИЙСКИЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ», г. Москва, Россия (119454, г. Москва, Пр-т Вернадского, д. 78, стр.4), e-mail: ¹gunart2000@mail.ru

В данной статье рассматриваются вопросы оптимизации архитектуры баз данных информационных систем при обработке больших объемов информации. Описаны технологии возможной оптимизации с идентификацией узких мест и проблемных запросов, оказывающих значительное влияние на производительность базы данных.

Ключевые слова: Оптимизация архитектуры БД, нормализация, проблемные запросы, индексирование, нереляционные БД.

OPTIMIZATION OF THE DATABASE ARCHITECTURE OF INFORMATION SYSTEMS WHEN PROCESSING LARGE AMOUNTS OF INFORMATION

¹Gunko A.S., Musikhin A.G.

MIREA - RUSSIAN TECHNOLOGICAL UNIVERSITY, Moscow, Russia (119454, Moscow, avenue. Vernadsky, 78, b. 4), e-mail: ¹gunart2000@mail.ru

This article discusses the issues of optimizing the architecture of information systems databases when processing large amounts of information. Possible optimization technologies with identification of bottlenecks and problematic queries that have a significant impact on database performance are described.

Keywords: Database architecture optimization, normalization, problematic queries, indexing, non-relational databases.

Поддержка большинства бизнес-процессов по всех сферах деятельности ведется сегодня с использованием как минимум учетных систем, а при необходимости проведения анализа с применением разного рода систем принятия решений. Структуры данных, которые используются для ведения учета и анализа данных, формируются на результате сбора разного рода структурированных и неструктурированных данных. При этом ведение учета возможно только на основе работы со структурированными особым образом данными.

Построенная стройная теория организации обработки данных с использованием принципов реляционной алгебры до сих пор является оптимальным решением для получения оперативной учетной информации и извлечения необходимых для ведения деятельности данных.

Однако расширение спектра систем и компаний, вышедших на интернет-площадки привело к резкому росту объемов данных, которые используются для организации работы даже в средней компании. В результате уже хорошо зарекомендовавшие себя средства обработки и хранения данных в виде популярных СУБД не полной мере справляются с поставленными задачами.

Вариантом решения данной проблемы может стать применение новых видов нереляционных баз данных, которые работают с частично неструктурированной информацией. Однако такие базы не совсем удачно использовать для решения учетных задач и получения оперативных данных. Второй вариант решения связан с оптимизацией структуры существующей базы данных.

Исследованию вопросов оптимизация архитектуры баз данных информационных систем при обработке больших объемов информации и посвящена данная работа.

Целью данной работы является оптимизация архитектуры баз данных информационных систем при обработке больших объемов информации.

Изучению данной проблемы посвящено множество работ специалистов. Все исследования можно разделить по нескольким направлениям. К первому направлению относятся работы, связанные с определением методов тестирования производительности баз данных, к ним можно отнести работы А.С. Басова [1], М.А. Маркелова, Д.А. Коростелёва [5].

Большой спектр работ касается поднятия вопроса о проблемах, возникающих в ходе работы с большими объемами данных в различных прикладных областях. Например, проблемам в сфере обработки педагогических данных посвящена работа С.Н. Вачковой, Э.М. Кагана, С.В. Козина [3]. Отдельное направление исследований связано с проведением сравнительного анализа различных СУБД по производительности и скорости обработки данных. Например, работы М.В. Богдановой, В.В. Ворониной [2], Т.А. Шихвеледовой, А.Г. Саидбеговой, З.К. Абдулаевой [8]. Здесь также выделяются работы, описывающие сравнение традиционных и облачных СУБД (В.А. Воронин, О.Н. Ромашкова [4]). Непосредственно к исследованию путей оптимизации базы данных можно отнести следующие работы И.Н. Набродовой, Г.А. Кузнецова [6]. Отдельные аспекты оптимизации структуры базы данных для повышения их производительности отмечены в работах П.В. Пилькевича, М.А. Масловой [7].

Эффективность работы базы данных напрямую зависит от структуры построенной модели. Чем более корректно разделены данные, тем меньше операций необходимо совершать для обеспечения целостности данных, а также возможно использование встроенных инструментов СУБД для поддержки созданных связей между сущностями.

Наиболее популярным методом построения структур данных до сих пор является использование реляционной модели несмотря на развитие и объектного подхода. Реляционная модель ориентирована на максимальное снижение вариантов дублирования данных, что обеспечивает меньшее число операций по отслеживанию возможных аномалий при обновлении.

Практически все современные СУБД поддерживают сопровождение реляционной модели с сохранением необходимых связей. Однако не всегда возникает необходимость поддержки конкретной связи при решении определенной прикладной задачи. Если связь существует в реальном мире, однако не используется информационной системой, то нет необходимости вводить данную связь в модель данных, так как она будет использовать ресурсы базы данных для поддержания целостности, в которой нет необходимости.

Примером такой связи могут служить данные об абитуриентах как физических лицах и студентах. После проведения приемной комиссии данные о бывших абитуриентах могут архивироваться, так как они не становятся частью процесса обучения в таком случае наличие

связи будет мешать удалению данных, так как они связаны с информацией о части абитуриентов, ставших студентами.

Решение такой проблемы может быть выполнено двумя путями: разделением данных систем приема и обучения, что не всегда удобно в крупном учебном заведении, либо путем разрыва сформированных связей после формирования приказов о зачислении.

Аналогичные проблемы возникают в процессе обработки данных кандидатов на должности в компании, которые также как физические лица связаны с понятием сотрудника на конкретной должности.

Ключевой функцией базы данных в информационной системе является предоставление необходимой информации в виде запросов. Запросы формируют как необходимую информацию как для ввода согласованных первичных данных, так и участвуют в получении дополнительной агрегированной информации для выполнения анализа.

Технология выполнения индексирования позволяет регулировать скорость получения данных в виде сокращения объемов обрабатываемых записей в базе данных. Формирование запроса предполагает чтение, поиск и сведение необходимых данных, которые были разнесены в процессе нормализации для исключения дублирования.

Индексированию могут быть подвергнуты практически все атрибуты записей, исключая специфические типы данных, что связано с конкретной реализацией и возможностями используемой СУБД. По умолчанию индексируются первичные ключи и внешние ключи как объекты связи. Дополнительное индексирование направлено на повышение эффективности выполнения транзакций.

Варианты проведения дополнительного индексирования связаны с конкретными потребностями пользователей и описываемыми структурами данных.

Наиболее часто реализуется выборочное дополнительное индексирование, которое связано с реализацией поиска данных более привычных для человека, например, поиск по названию компании вместо ее регистрационного номера.

Введение таких индексов в значительной мере позволяет повысить скорость выполнения запросов с использованием включений, упорядочивания и по параметру с участием анализируемых атрибутов записи. Однако применение дополнительной индексации слишком широко может вызвать повышение уровня накладных расходов, связанных с обеспечением хранения и обработки данных в целом.

Используемые в ходе разработки структуры базы данных правила реляционной алгебры не всегда отвечают потребностям пользователей конкретных базы данных. Сформированные теоретические обоснования необходимые для разделения объектов анализируемой предметной области предполагают использования принципов выделения общих объектов с единими свойствами. Однако в таком случае может возникнуть таблица базы данных, хранящая описание множества похожих субъектов обработка данных которых происходит частично раздельно.

Положенная в основу крупной информационной системы типа обеспечения учета пациентов для платформы «Здоровье Петербуржца» (<https://gorzdrav.spb.ru/>) или более масштабный проект записи к врачу через ГОСУСЛУГИ (<https://esia.gosuslugi.ru/>) приведет к необходимости организации поиска данных по всем записям таблицы «Человек». В результате обработка запросов, связанных с поиском данных о специальности врача будет включать проход по записям таблицы «Человек».

Очевидно, что использование такой структуры для решения конкретной задачи управления данными пациентов и записи их к врачу является неоптимальной и требуется перенесение данных сущности «Человек» в уже существующие сущности «Врач» и «Пациент».

Приведенные примеры говорят об особенностях применения данной технологии оптимизации, которая связана с конкретными аспектами решаемой прикладной задачи. Универсальный подход в определении факторов разделения некоторых таблиц для повышения производительности базы данных отсутствует.

Список литературы

1. Басов А.С. Обзор механизмов тестирования баз данных / А.С. Басов // Вестник науки. – 2020. – Т. 4. – № 7 (28). – С. 44-49.
2. Богданова М.В. Исследование быстродействия различных СУБД / М.В. Богданова, В.В. Воронина // В сборнике: Информационные технологии в образовательном процессе вуза и школы. Материалы XV Всероссийской научно-практической конференции. Редколлегия: Р.М. Чудинский (науч. ред.) [и др.]. Воронеж, 2021. – С. 85-88.
3. Вачкова С.Н. Большие данные для педагогических исследований: возможности, проблемы, ограничения / С.Н. Вачкова, Э.М. Каган, С.В. Козин // Вестник Православного Свято-Тихоновского гуманитарного университета. Серия 4: Педагогика. Психология. – 2021. – № 63. – С. 28-39.
4. Воронин В.А. Сравнительная оценка производительности облачной и традиционной баз данных по критерию времени отклика / В.А. Воронин, О.Н. Ромашкова // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. – 2022. – № 10-2. – С. 38-43.
5. Короткович, Д. SQL Server. Наладка и оптимизация для профессионалов/ Д. Короткович. – СПб.: Питер, 2022. – 512 с.
6. Набродова И.Н. Повышение производительности баз данных / И.Н. Набродова, Г.А. Кузнецов // Известия Тульского государственного университета. Технические науки. – 2021. – № 9. – С. 371-373.
7. Пилькевич П.В. Влияние индексов реляционных баз данных на производительность поиска / П.В. Пилькевич, М.А. Маслова // В сборнике: Угрозы и риски на Юге России в условиях геополитического кризиса. Достижения и перспективы научных исследований молодых ученых Юга России. Материалы научных мероприятий: Всероссийской конференции с международным участием; XIX Ежегодной молодежной научной конференции. Ростов-на-Дону, 2023. – С. 310.

8. Шихвеледова, Т. А. Анализ производительности и преимуществ разных подходов в управлении данными SQL и NoSQL / Т.А. Шихвеледова, А.Г. Саидбегова, З.К. Абдулаева // Молодой ученый. – 2022. – № 15 (410). – С. 35-38.

References

1. Basov A.S. Review of database testing mechanisms / A.S. Basov // Bulletin of Science. - 2020. – Т. 4. – № 7 (28). – pp. 44-49.
 2. Bogdanova M.V. Investigation of the performance of various DBMS / M.V. Bogdanova, V.V. Voronina // In the collection: Information technologies in the educational process of universities and schools. Materials of the XV All-Russian Scientific and Practical conference. Editorial board: R.M. Chudinsky (scientific editor) [and others]. Voronezh, 2021. – pp. 85-88.
 3. Vachkova S.N. Big data for pedagogical research: opportunities, problems, limitations / S.N. Vachkova, E.M. Kagan, S.V. Kozin // Bulletin of the Orthodox St. Tikhon's University for the Humanities. Series 4: Pedagogy. Psychology. - 2021. – No. 63. – pp. 28-39.
 4. Voronin V.A. Comparative assessment of the performance of cloud and traditional databases according to the response time criterion / V.A. Voronin, O.N. Romashkova // Modern science: actual problems of theory and practice. Series: Natural and Technical Sciences. - 2022. – No. 10-2. – pp. 38-43.
 5. Korotkovich, D. SQL Server. Commissioning and optimization for professionals/ D. Korotkovich. – St. Petersburg: St. Petersburg, 2022. – p.512
 6. Nabrodova I.N. Improving database performance / I.N. Nabrodova, G.A. Kuznetsov // Izvestiya Tula State University. Technical sciences. - 2021. – No. 9. – pp. 371-373.
 7. Pilkevich P.V. The influence of relational database indexes on search performance / P.V. Pilkevich, M.A. Maslova // In the collection: Threats and risks in the South of Russia in the context of the geopolitical crisis. Achievements and prospects of scientific research of young scientists of the South of Russia. Materials of scientific events: the All-Russian Conference with international participation; the XIX Annual Youth Scientific Conference. Rostov-on-Don, 2023. – p. 310.
 8. Shikhveledova, T. A. Performance analysis and advantages of different approaches in SQL and NoSQL data management / T.A. Shikhveledova, A.G. Saidbegova, Z.K. Abdulaeva // Young Scientist. – 2022. – № 15 (410). – pp. 35-38.
-