

УДК 004.652.42

ИССЛЕДОВАНИЕ «ПАРАДОКСА НОРМАЛИЗАЦИИ» В ФАКТОГРАФИЧЕСКИХ РЕЛЯЦИОННЫХ БАЗАХ ДАННЫХ НА ПЛАТФОРМЕ ORACLE

О.В. Тараканов^а, Н.Б. Пикалова^а, В.О. Тараканова^а

^а Академия ФСО России, Орел, 302034, Российская Федерация

Адрес для переписки: ole_g66@list.ru

Информация о статье

Поступила в редакцию 12.04.18, принята к печати 17.05.18

doi: 10.17586/2226-1494-2018-18-4-639-645

Язык статьи – русский

Ссылка для цитирования: Тараканов О.В., Пикалова Н.Б., Тараканова В.О. Исследование «парадокса нормализации» в фактографических реляционных базах данных на платформе ORACLE // Научно-технический вестник информационных технологий, механики и оптики. 2018. Т. 18. № 4. С. 639–645. doi: 10.17586/2226-1494-2018-18-4-639-645

Аннотация

Предмет исследования. Приведены результаты экспериментальной оценки силы взаимосвязи между степенью нормализованности базы данных и временем формирования ею выдачи по запросу. Целью исследования является поиск объективной оценки силы корреляции между степенью нормализованности базы данных и временем выполнения вложенных запросов к ней в зависимости от требуемой глубины рекурсии, реализуемой в запросе (модель «парадокса нормализации»). **Метод.** Применяемый метод основан на выполнении корреляционного анализа по результатам численного эксперимента для идентификации «парадокса нормализации». **Основные результаты.** Установлено, что в базах данных на платформе ORACLE достаточно слабая корреляция между глубиной рекурсии, обусловленной достижением определенной нормальной формы таблиц, и временем выполнения запроса. Взаимосвязь между объемом базы данных, определяемой числом учетных объектов, и временем формирования выдачи также слабая. Сформулирован и обоснован вывод об отсутствии оснований по проведению денормализации базы данных или не достижении высоких нормальных форм в угоду повышения ее быстродействия при выполнении запросов на выборку записей. Все значимые выводы сформулированы по отношению к фактографическим базам данных на платформе ORACLE. Транспонирование результатов на другие современные системы баз данных не производилось. **Практическая значимость.** Практическая значимость работы состоит в объективном обосновании необходимости (но не достаточности) приведения отношений реляционной базы данных под управлением ORACLE Database 11g R2 к высшим нормальным формам для обеспечения ее целостности. Новизна результата исследования обусловлена экспериментальным подтверждением истинности гипотезы о зависимости времени выборки кортежей от степени нормализованности базы данных при рекурсивной процедуре отбора и формулировкой модели «парадокса нормализации» для баз данных на платформе ORACLE.

Ключевые слова

фактографическая реляционная база данных, нормализация базы данных, время формирования выдачи по запросу, «парадокс нормализации», первичный ключ, суррогатный ключ

"NORMALIZATION PARADOX" RESEARCH IN FACTUAL RELATIONAL DATABASES ON ORACLE PLATFORM

O.V. Tarakanov^a, N.B. Pikalova^a, V.O. Tarakanova^a

^aAcademy of Federal Security Guard Service of the Russian Federation, Oryol, 302034, Russian Federation

Corresponding author: ole_g66@list.ru

Article info

Received 22.03.18, accepted 14.05.18

doi: 10.17586/2226-1494-2018-18-4-639-645

Article in Russian

For citation: Tarakanov O.V., Pikalova N.B., Tarakanova V.O. "Normalization paradox" research in factual relational databases on ORACLE platform. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2018, vol. 18, no. 4, pp. 639–645 (in Russian). doi: 10.17586/2226-1494-2018-18-4-639-645

Abstract

Subject of Research. The paper presents the results of experimental assessment of relationship force between database normalization degree and database response formation time on a query. The research aim lies in the search of objective assessment of correlation force between database normalization degree and the run time of the enclosed inquiries to it depending on the required depth of the recursion realized in the inquiry ("normalization paradox" model). **Method.** The applied method is based on implementation of the correlation analysis according to the results of a numerical experiment for "normalization paradox" identification. **Main Results.** It is established that correlation between the recursion depth caused by achievement of a certain normal form of tables and inquiry performance time is rather weak in databases on the ORACLE platform. Interrelation between the database volume determined by the number of considered objects and delivery formation time is also weak. The valid conclusion is formulated about the absence of grounds for carrying out the database denormalization or not achievement of high normal forms for the good of the increase in its speed when performing requests for selection of records. All significant conclusions are drawn on the relation of factual databases on the ORACLE platform. Transposition of results on the other up-to-date database systems was not performed. **Practical Relevance.** The practical relevance of work consists in objective justification of the need (but not sufficiency) for reductions of the relations of the relational database under control of ORACLE Database 11g R2 to the highest normal forms for ensuring its integrity. The research result novelty is caused by experimental confirmation of the hypothesis validity about selection time dependence of tuples on database normalization degree at the recursive selection procedure and formulation of "normalization paradox" model for databases on the ORACLE platform.

Keywords

factual relational database, database normalization, response formation time on query, "normalization paradox", primary key, surrogate key

Введение

В теории и практике реляционных баз данных широко известен «парадокс нормализации». Сущность парадокса сводится к декларации зависимости быстродействия выполнения поисковых запросов от числа используемых для формирования выдачи связанных таблиц [1–3]. Тожественные по информационному наполнению (по числу объектов учета и по числу учитываемых характеристик каждого из них) базы данных могут быть существенно различными по числу таблиц, по примененным инженерным подходам к проектированию (от запроса или от предметной области) или по форме нормализации.

Первый случай в работе не рассматривается. В результате качественной нормализации, действительно, возможен существенный рост числа таблиц базы данных. Особенно это заметно при формировании второй и третьей нормальных форм [4, 5]. Так как нормализация базы данных проводится путем проектирования таблиц (разделения множества атрибутов отношения на совокупность подмножеств, удовлетворяющим требованиям соответствующей нормальной формы), следует ожидать увеличения числа таблиц в нормализованной базе данных. В процессе выполнения запроса к совокупности связанных таблиц реализуется рекурсивная процедура с глубиной, прямо зависящей от числа таблиц, связанных в цепочку [2, 6]. Отсюда следует собственно «парадокс нормализации» – обеспечение целостности базы данных приводит к снижению ее быстродействия.

Постановка задачи

Сущность «парадокса нормализации» обусловлена рекурсивностью алгоритма формирования выдачи на запрос к нескольким таблицам. Очевидно, что глубина рекурсии оказывает влияние на время выполнения запроса, определяемое задержками на загрузку дочерних таблиц в глобальную системную область (кеш-буфер данных) [4]. Современные алгоритмы выборки строк в многотабличных базах данных обладают существенной кортежной избыточностью, определяемой физической организацией хранения записей. Здесь актуальны механизмы поблочной записи (на уровне блоков файловой системы операционной системы и блоков базы данных для высокоэффективных систем управления базами данных), механизмы формирования непрерывных цепочек экстенгов, настройки методов доступа для табличных пространств (локализация и дефрагментация файлов данных), квоты на использование ресурсов сервера для профиля учетной записи пользователя, быстродействие аппаратно-программной платформы сервера и тому подобное [7, 8]. Следовательно, уравнение времени выборки вида

$$t = \sum_i t_i,$$

где t_i – время выполнения i -го подзапроса к i -ой дочерней таблице рекурсии глубины i , не актуально и не может обосновывать допущение, что ненормализованная база данных обладает более высоким быстродействием по сравнению с нормализованной.

Данная гипотеза принимается в качестве рабочей для настоящего исследования. Анализ «парадокса нормализации» может быть выполнен путем экспериментальной проверки, которая обладает рядом ограничений [9]. К ним в первую очередь следует отнести структуру и наполненность базы данных, используемую систему управления базой данных (СУБД), технические характеристики аппаратно-программной платформы сервера.

В настоящей работе пространство вариаций объекта наблюдения ограничено классом фактографических баз данных, применяемых в так называемых OLTP (On-Line Transaction Processing) системах.

Воспроизводимость, интерпретируемость, адекватность и достоверность исследования обеспечена за счет корректной постановки и проведения отдельных экспериментов, статистической обработки экспериментальных данных, статистического обоснования принимаемых гипотез.

Пусть имеется база данных, состоящая из единственной таблицы, приведенной к первой нормальной форме – все атрибуты атомарны, определены все потенциальные ключи. Таблица содержит n атрибутов и описывает некоторую предметную область. В процессе нормализации – приведения совокупности n атрибутов к соответствующим нормальным формам – будут сформированы x_j таблиц, где j – соответствующая нормальная форма, по которым будут распределены все n атрибутов исходного отношения.

Каждая из x_j таблиц должна содержать подмножество кортежей, семантически адекватное набору из исходной таблицы. Таблицы размещаются на сервере в вариациях числа записей k . Для получения соизмеримых значений времени отклика на семантически идентичный запрос обеспечена равнонаполненность различных таблиц. Для получения характеристических оценок исключено влияние на время отклика параметров, отличных от предположительно управляющего. Исследование ограничено единственной СУБД – ORACLE Database 11g R2. Обоснованный выбор СУБД не производится, используется имеющаяся система, транспонирование результатов исследования на иные СУБД не предполагается.

Для измерения времени отклика системы на поисковый запрос (запрос на выборку кортежей) формируется множество $\{m\}$ семантически тождественных запросов, направляемых к базе данных в соответствующей нормальной форме. Для учета статистической погрешности по каждому j, k -измерению (вариации нормальной формы базы данных и числа учтенных в ней объектов) предполагается выполнить по l реплик. Таким образом, будет сформирован массив измерений для дисперсионного анализа и определения корреляции между параметром j и временем t ожидания выборки кортежей (строк):

$$t = \varphi(j, k).$$

Задача исследования ставится следующим образом. Имеется ненормализованная база данных, представленная единственным отношением, описывающим некоторую предметную область. В таблице базы данных хранится k кортежей, учитывающих k гипотетических объектов учета. Применена СУБД ORACLE Database 11g R2. Для экспериментальной оценки зависимости времени выполнения выборки от степени нормализованности базы данных используются версии базы данных, приведенные к первой, второй и третьей нормальным формам.

Выдвигаются следующие гипотезы исследования относительно вида функции φ : линейная – H_1 , экспоненциальная – H_2 , обратная экспоненциальная – H_3 , постоянная – H_4 . Для проверки гипотез H_m , $m \in \overline{1, 4}$, необходимо набрать статистику в виде совокупности реплик $t_{j,k}$, поочередно зафиксировав параметр j , а затем параметр k . Полученная матрица измерений подлежит проверке на аномальность крайних значений ранжированного ряда измерений по каждому фактору j и k . Воспроизводимость эксперимента проверяется путем оценки однородности дисперсий по критерию Кохрена. Достаточность количества реплик обосновывается по оперативным характеристикам при обеспечении мощности эксперимента на уровне $\alpha \leq 0,05$ и $\beta \leq 0,05$ [9, 10].

Задача состоит в том, чтобы оценить истинность m -й гипотезы и определить силу взаимосвязи времени выполнения выборки кортежей из базы данных от j -й степени ее нормализованности на популярной аппаратно-программной платформе сервера базы данных на примере СУБД ORACLE Database 11g R2.

Описание метода

Структура базы данных сформирована в виде шести схем, отличающихся друг от друга степенью нормализованности (первая, вторая, третья нормальные формы), а также использованием или неиспользованием суррогатных первичных ключей. Это усложняет матрицу измерений, но позволяет оценить эффективность применения в базах данных суррогатных первичных ключей. Использование только первых трех степеней нормализованности j оправдано, так как на практике приведение базы данных в высшие нормальные формы (нормальная форма Бойса-Кодда и выше) заключается в проверке выполнения условий этих форм без преобразования структуры отношений [2, 11]. Основной рост числа таблиц в ходе нормализации наблюдается при достижении второй и третьей нормальных форм.

Результаты исследования предполагаются получить в форме подтверждения истинности одной из H_m гипотез, обуславливающих вид зависимости и статистическую точность ее аппроксимации известным математическим уравнением (функцией), что может быть использовано как математическая модель «парадокса нормализации». При этом подтверждение гипотезы H_4 , по сути, опровергает существование парадокса, что является сильным обоснованием в пользу проведения нормализации баз данных.

Для исследования сформирована логическая структура базы данных в предметной области учета субъектов хозяйственной деятельности регионов. Тематика базы данных в исследовании не имеет значения, декларированная предметная область выбрана из прагматических соображений (структура разрабатывалась для решения других задач).

Для получения согласованного наполнения исследовательских баз данных (идентичности объектов

учета по количеству учитываемых объектов и значениям их характеристик) разработана и применена прикладная программа генерации правдоподобных записей (Свидетельство ФИПС о государственной регистрации программы для ЭВМ № 2016616123 от 07.06.2016). С ее помощью сформированы варианты согласованного наполнения исследовательских баз данных фактами о 1000, 10000 и 100000 объектов учета. Данное положение позволяет считать, что используется одна база данных в шести вариантах реализации при трех ее объемах.

В качестве факторов влияния параметров базы данных на время выполнения алгоритма формирования выдачи выбраны: уровень рекурсии (количество вложений запроса в нормальных формах – от 1 до 3) – фактор A (вариация параметра j исследовательской задачи); количество записей в базе данных – фактор B (вариация параметра k исследовательской задачи). Уровень рекурсии определяется числом операций связывания отдельных таблиц для выполнения запроса. В запросах к единственной таблице $A=0$, при связывании двух таблиц $A=1$, а при n таблицах потребуется $n - 1$ рекурсия. Для ограничения пространства исследования применены запросы со связыванием не более чем трех таблиц. Фактор A варьируется в диапазоне 0–3 ступенчато. Фактор B также варьируется ступенчато из множества значений {1000, 10000, 100000}, представляющих число хранящихся в базе данных объектов учета.

В качестве отклика использовано время реакции базы данных на запрос. При этом сформирована гомогенная по прагматике выдачи смесь запросов, синтаксически изменяющаяся в соответствии с конкретной парой значений $\{j, k\}$. При равном количестве объектов учета (фиксированном значении фактора B) во всех вариантах исследовательской базы данных (перебор всех значений фактора A) должны быть получены одинаковые по количеству и значениям записей выдачи. Данное положение проверено на этапе подготовки эксперимента. При вариациях фактора B применяется та же смесь запросов, но выдачи различаются по количеству получаемых записей.

Для исследования выбран многофакторный несимметричный план эксперимента с двумя факторами. Матрица измерений представляет собой совокупность двумерных таблиц, для каждой из которых зафиксировано значение фактора B . В ходе эксперимента установлено, что в случае с первичным ключом строкового типа и суррогатным первичным ключом изменение времени отклика лежит в пределах погрешности измерений. Данное явление позволяет утверждать, что тип данных первичного ключа не оказывает влияния на время реакции базы данных на запрос, что обусловило возможность исключения из корреляционного анализа результатов, полученных для клонов вариантов базы данных с суррогатным первичным ключом. Этот вывод не избавляет от необходимости выбирать эффективный первичный ключ при достижении доменно-ключевой нормальной формы (в том числе путем формирования суррогатного), так как, кроме быстродействия, следует обеспечивать и максимально достижимую целостность базы данных.

Число реплик по каждой комбинации уровней факторов определено посредством вычисления их минимально допустимого количества, обеспечивающего требуемую мощность эксперимента. При допустимых значениях вероятности ошибки первого и второго рода $\alpha \leq 0,05$ и $\beta \leq 0,05$ (мощность эксперимента $1 - \beta \geq 0,95$) по критерию Фишера определено минимальное число реплик – не менее девяти. В ходе эксперимента выполнено по десять реплик по каждой комбинации значений уровней факторов.

На корректность экспериментальных данных существенное влияние оказывает алгоритм кэширования данных сервера баз данных. Для исключения влияния указанного эффекта перед каждым измерением производился перезапуск сеанса, гарантирующий незаполненность кэш-буфера данных глобальной системной области [7, 8]. В таких условиях гарантируется измерение собственной скорости работы алгоритма выборки (без коррекции со стороны механизмов кэширования).

Результаты экспериментирования и оценка корреляции

Экспериментальные данные, полученные для значений фактора B , равного 1000, 10000 и 100000, представлены на рис. 1–3 соответственно в виде семейств откликов в координатной плоскости времени реакции базы данных на запрос. В одной координатной плоскости собраны все отклики при вариациях значений фактора A и фиксированном значении фактора B .

Для корректности корреляционного анализа экспериментальных данных оценена их аномальность по критерию Диксона, позволяющему выявить случайные выбросы в массиве данных [9]. Массивы экспериментальных данных ранжированы, и для крайних значений рассчитаны значения коэффициентов r_{11} при десяти измерениях для уровня значимости 0,05. Табулированное значение коэффициента Диксона, пригодное по условиям эксперимента, составляет $r_{11} = 0,477$. При всех вариациях факторов A и B расчетные значения коэффициента r_{11} не превышают табулированное значение как для наименьшего, так и для наибольшего значения отклика, что свидетельствует об отсутствии случайных выбросов в экспериментальных данных.

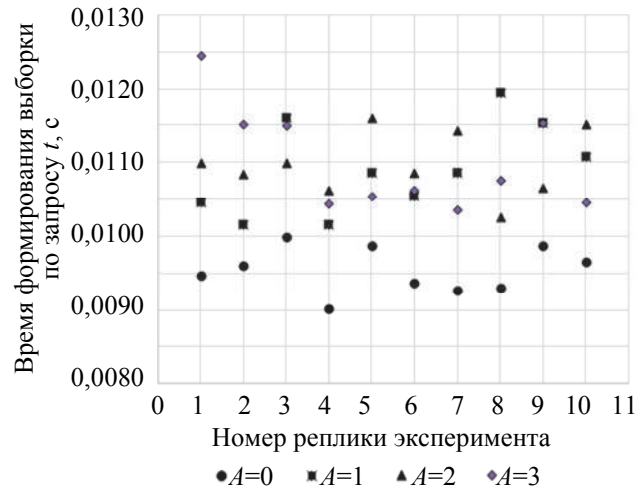


Рис. 1. Время формирования выборки по запросу к базе данных с объемом хранения 1000 объектов учета

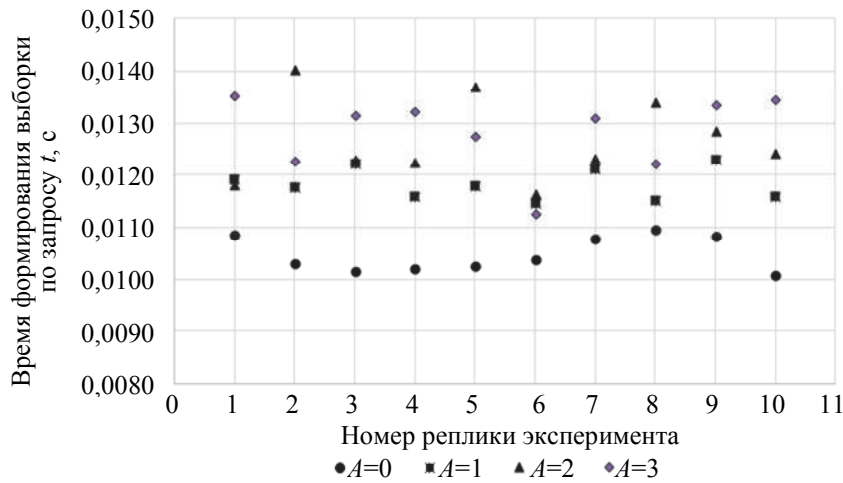


Рис. 2. Время формирования выборки по запросу к базе данных с объемом хранения 10000 объектов учета

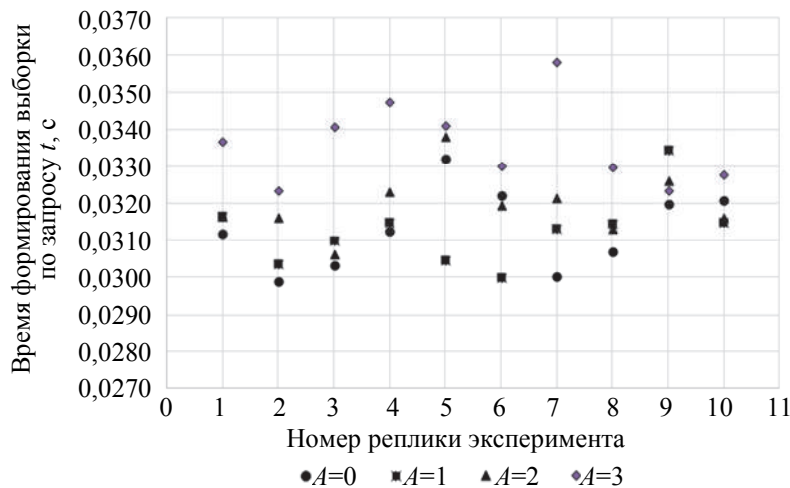


Рис. 3. Время формирования выборки по запросу к базе данных с объемом хранения 100000 объектов учета

Критерий Кохрена (однородность дисперсий) позволяет интерпретировать эксперимент как воспроизводимый при практическом достижении его табулированного значения [10]. Установлено, что по критерию Кохрена дисперсии при всех вариациях значений уровней факторов A и B однородны, что позволяет считать эксперимент воспроизводимым.

При использовании критериев Диксона и Фишера предполагается, что генеральная совокупность полученных экспериментальных значений распределена по нормальному закону. Проверка соответствия массива экспериментальных данных нормальному распределению проведена на основании правила «трех сигм». В соответствии с правилом «трех сигм» все экспериментальные данные можно считать распределенными нормально.

Достаточность десяти реплик по каждой комбинации значений уровней факторов проверена по операционным характеристикам при значении $\Phi = 0,55$ и $\alpha = 0,05$ с учетом значений степеней свободы $\nu_1 = 3$ и $\nu_2 = 36$. Вероятность ошибки второго рода $\beta \leq 0,05$ позволяет считать эксперимент состоятельным и воспроизводимым. Истинность m -ой гипотезы и сила взаимосвязи времени выполнения выборки кортежей из базы данных от j -ой степени ее нормализованности на популярных аппаратно-программных платформах сервера базы данных на примере СУБД ORACLE Database 11g R2 определена путем вычисления выборочного коэффициента корреляции Пирсона [10]:

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{j=1}^n (Y_j - \bar{Y})^2}}$$

где $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$, $\bar{Y} = \sum_{j=1}^n \frac{Y_j}{n}$, X_i и Y_i – величины, корреляция между которыми подлежит оценке. Интерпретация значений коэффициента R выполнена в соответствии с [10]. Выборочные коэффициенты корреляции для факторов A и B составляют $R_A = 0,31$ и $R_B = 0,44$. Для фактора A (число рекурсий при выполнении запроса, зависящее от нормализованности базы данных) подтверждена гипотеза H_A , позволяющая считать, что время выполнения выдачи на запрос в базах данных на платформе ORACLE Database 11g R2 слабо зависит от числа таблиц, связанных для выполнения алгоритма запроса, что свидетельствует о несостоятельности «парадокса нормализации» в заявленных базах данных.

Выполненное экспериментальное исследование подтверждает очевидное предположение, что с ростом числа записей в таблицах базы данных соответственно увеличивается время работы алгоритма выборки. Установлено, что для баз данных на платформе ORACLE Database 11g R2 указанная взаимосвязь слаба. Это отличает выбранную платформу от СУБД Access компании Microsoft, для которой известны иные результаты, подтверждающие существование «парадокса нормализации» [12]. Для СУБД MySQL и NoSQL платформы MapReduce/Hadoop4 опубликованы аналогичные исследования [13, 14], в которых оценена средняя корреляция между временем выполнения запроса к связанным таблицам и числом таблиц, участвующих в запросе (зависит от степени нормализованности базы данных). Известны результаты экспериментальных исследований для СУБД PostgreSQL 9.4 под управлением операционной системы Linux [15], где, судя по приведенным данным, корреляция между временем выполнения запроса на выборку и глубиной рекурсии в связке согласованных таблиц слаба.

Результаты упомянутых исследований показывают, что «парадокс нормализации» есть свойство конкретного программного продукта. Данное положение свидетельствует о том, что увеличение времени выборки по запросу при росте числа участвующих в нем связанных таблиц не позволяет обосновать денормализацию или отказ от приведения таблиц базы данных к высшим нормальным формам.

Таким образом, «парадокс нормализации» в системах баз данных на платформе ORACLE Database 11g R2 проявляется незначительно на базах с числом объектов учета до 100000. Это означает, что приведение базы данных к высшим нормальным формам является необходимым (но не достаточным) условием обеспечения их целостности.

Заключение

Результаты проведенного исследования не обеспечивают полноту описания исследовательской области. В этом случае необходимо более полное исследование процедуры нормализации отношения с приведением его вплоть до доменно-ключевой нормальной формы (включая переносимость результатов на случаи применения других популярных платформ систем управления базами данных). Однако полученные результаты могут использоваться при проектировании исходных отношений информационно-логической модели предметной области. Новизна научного результата обусловлена экспериментальным подтверждением истинности H_m -й гипотезы зависимости времени выборки кортежей от степени нормализованности базы данных при рекурсивной процедуре отбора. Научность исследования обеспечена истинностью математической модели «парадокса нормализации», подтверждаемой результатами корректного экспериментального исследования, интересубъективностью, определяемой независимостью модели «парадокса нормализации» от моделируемой с помощью базы данных предметной области (для фактографической базы данных), и системностью получения научного результата, обеспеченной корректным применением положений теории системного анализа и теории статистического планирования экспериментов.

Литература

1. Карпук А., Краснопрошин В.В. Циклы в структурах функциональных зависимостей // *International Journal of Open Information Technologies*. 2017. Т. 5. № 7. С. 38–44.
2. Мартынов В.В., Прошин Е.Н. Подход к анализу характеристик производительности баз данных // *Вестник УГАТУ*. 2009. Т. 13. № 2(35). С. 198–205.
3. Мендкович Н.А., Кузнецов С.Д. Оценка эффективности минимизации ограничений запросов к СУБД // *Труды Института системного программирования РАН*. 2013. Т. 25. С. 113–130.
4. Брешенков А.В., Белоус В.В. Преобразование заполненных реляционных таблиц ко второй нормальной форме // *Наука и образование: научное издание МГТУ им. Н.Э. Баумана*. 2007. № 2. С. 4.
5. Панченко Б.Е. Исследования доменно-ключевой схемы реляционной базы данных // *Кибернетика и системный анализ*. 2012. № 6. С. 157–172.
6. Poolet M.A. Database Harmony [Электронный ресурс]. SQL Server PRO. 2002. URL: sqlmag.com/business-intelligence/database-harmony. Яз. англ. (дата обращения 19.05.2018).
7. Кайт Т. Oracle для профессионалов. СПб.: ДиаСофтЮП, 2003. 672 с.
8. Слепцова К.А., Комков А.Е., Кузовлев В.И. Представление и обработка информационных ресурсов для управления деятельностью предприятия // *Современные наукоемкие технологии*. 2016. № 4-1. С. 57–61.
9. Блохин В.Г., Глудкин О.П., Гуров А.И., Ханин М.А. Современный эксперимент: подготовка, проведение, анализ результатов. М.: Радио и связь, 1997. 232 с.
10. Гмурман В.Е. Теория вероятностей и математическая статистика. М.: Высшая школа, 1972. 367 с.
11. Саенко И.Б., Куваев В.О. Модели и методика оценки и обеспечения оперативной доступности ресурсов в едином информационном пространстве // *Математические методы в технике и технологиях*. 2016. № 6(88). С. 139–141.
12. Pavlo A., Paulson E., Rasin A., Abadi D.J., DeWitt D.J., Madden S., Stonebraker M. A comparison of approaches to large-scale data analysis // *Proc. Int. Conf. on Management of Data*. New York, 2009. P. 165–178. doi: 10.1145/1559845.1559865
13. Григорьев Ю.А., Плутенко А.Д. Анализ процесса выполнения запроса на соединение таблиц в строчной параллельной СУБД // *Информатика и системы управления*. 2013. № 4(38). С. 3–15.
14. Григорьев Ю.А., Устимов А.И. Сравнение времени выполнения запроса к хранилищу данных в среде MapReduce/Hadoop и СУБД MySQL // *Информатика и системы управления*. 2016. № 3(49). С. 3–12. doi: 10.22250/isu.2016.49.3-12
15. Leis V. et al. How good are query optimizers, really? // *Proceedings of the VLDB Endowment*. 2015. V. 9. N 3. P. 204–215.

Авторы

Тараканов Олег Викторович – кандидат технических наук, доцент, сотрудник, Академия ФСО России, Орел, 302034, Российская Федерация, ORCID ID: 0000-0001-6095-4154, ole_g66@list.ru

Пикалова Наталья Борисовна – сотрудник, Академия ФСО России, Орел, 302034, Российская Федерация, ORCID ID: 0000-0002-9081-7764, borisovna.311@mail.ru

Тараканова Валерия Олеговна – студент, Академия ФСО России, Орел, 302034, Российская Федерация, ORCID ID: 0000-0002-3663-8044, talero@list.ru

References

1. Karpuk A.A., Krasnoproschin V.V. Cycles in structures of functional dependencies. *International Journal of Open Information Technologies*, 2017, vol. 5, no. 7, pp. 38–44. (in Russian)
2. Martynov V.V., Proshin E.N. Approach to analyzing database performance characteristics. *Vestnik UGATU*, 2009, vol. 13, no. 2, pp. 198–205. (in Russian)
3. Mendkovich N.A., Kuznetsov S.D. Minimization of data base query"s conditions: evolution of efficiency. *Proceedings of ISP RAS*, 20113, vol. 25, pp. 113–130. (in Russian)
4. Breshenkov A.V., Belous V.V. Converting the completed relational tables to the second normal form. *Science and Education of Bauman MSTU*, 2007, no. 2, p. 4. (in Russian)
5. Panchenko B.E. Domain-key schema of the relational database. *Kibernetika i Sistemnyi Analiz*, 2012, no. 6, pp. 157–172. (in Russian)
6. Poolet M.A. *Database Harmony*. SQL Server PRO. 2002. Available at: sqlmag.com/business-intelligence/database-harmony (accessed 19.05.2018).
7. Kyte T. *Expert One-on-One Oracle*. Pearson, 2002.
8. Sleptsova K.A., Komkov A.E., Kuzovlev V.I. Presentation and processing of information resources management activities of the company. *Modern High Technologies*, 2016, no. 4-1, pp. 57–61. (in Russian)
9. Blokhin V.G., Gludkin O.P., Gurov A.I., Khanin M.A. *Modern Experiment: Preparation, Conduct, Analysis of Results*. Moscow, Radio i Svyaz' Publ., 1997, 232 p. (in Russian)
10. Gmurman V.E. *Theory of Probability and Mathematical Statistics*. Moscow, Vysshaya Shkola Publ., 1972, 367 p. (in Russian)
11. Saenko I.B., Kuvaev V.O. Models and methodology for assessing and ensuring the operational availability of resources in a single information space. *Mathematical Methods in Engineering and Technology*, 2016, no. 6, pp. 139–141. (in Russian)
12. Pavlo A., Paulson E., Rasin A., Abadi D.J., DeWitt D.J., Madden S., Stonebraker M. A comparison of approaches to large-scale data analysis. *Proc. Int. Conf. on Management of Data*. New York, 2009, pp. 165–178. doi: 10.1145/1559845.1559865
13. Grigorev U.A., Plutenko A.D. Analysis of query process for tables combinations in the parallel rows of DBMS. *Information Science and Control Systems*, 2013, no. 4, pp. 3–15. (in Russian)
14. Grigor'ev Yu.A., Ustimov A.I. Comparison of query execution time to data warehouse in MapReduce/Hadoop and RDBMS MySQL. *Information Science and Control Systems*, 2016, no. 3, pp. 3–12.
15. Leis V. et al. How good are query optimizers, really? // *Proceedings of the VLDB Endowment*, 2015, vol. 9, no. 3, pp. 204–215.

Authors

Oleg V. Tarakanov – PhD, Associate Professor, employee, Academy of Federal Security Guard Service of the Russian Federation, Oryol, 302034, Russian Federation, ORCID ID: 0000-0001-6095-4154, ole_g66@list.ru

Natalia B. Pikalova – employee, Academy of Federal Security Guard Service of the Russian Federation, Oryol, 302034, Russian Federation, ORCID ID: 0000-0002-9081-7764, borisovna.311@mail.ru

Valeriya O. Tarakanova – student, Academy of Federal Security Guard Service of the Russian Federation, Oryol, 302034, Russian Federation, ORCID ID: 0000-0002-3663-8044, talero@list.ru