

О.В. ТАРАКАНОВ
**СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ СПОСОБОВ ХРАНЕНИЯ
БОЛЬШИХ ДВОИЧНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ В БАЗАХ
ДАННЫХ ORACLE**

Тараканов О.В. Сравнительное исследование способов хранения больших двоичных последовательностей в базах данных ORACLE.

Аннотация. Работа выполнена с целью построения математического описания условий пригодности основных способов хранения больших двоичных последовательностей (BLOB и BFILE) в базах данных на платформе ORACLE. Исследование выполнено путем восстановления уравнений множественной линейной регрессии, учитывающих значимые на наш взгляд условия применения заявленных типов данных. Сформулированы некоторые практические рекомендации для обоснования выбора типа данных для хранения больших двоичных последовательностей.

Ключевые слова: база данных, большая двоичная последовательность, большой двоичный объект, BLOB, BFILE, уравнение множественной линейной регрессии.

Tarakanov O.V. A Comparative Research into Ways to Store Binary Large Objects in the ORACLE Databases.

Abstract. The research purpose was to describe the conditions of formation of mathematical suitability of main ways to store binary large sequences (BLOB or BFILE) in ORACLE databases. The solution is obtained by restoring multiple linear regression equations taking into account significant conditions of use of these datatypes as BLOB and BFILE. Some practical recommendations are formulated to justify the selection of the datatype to store binary large sequences.

Keywords: database; binary large sequence; binary large object; BLOB; BFILE; multiple linear regression equation.

1. Введение. Ведение двоичных данных в фактографических базах данных сопряжено с рядом трудностей и противоречий: необходимо обеспечить их связь с соответствующими идентификаторами, добиться выровненности строк, соблюдения эргономичности или хотя бы одинаковость обращения с ними пользователя. При этом не снизить эффективность системы баз данных в целом. Противоречивость требований к ведению двоичных последовательностей в базах данных обуславливает актуальность сравнительного исследования основных способов хранения двоичных последовательностей.

Современный уровень развития технологий баз данных предлагает два основных способа хранения больших двоичных последовательностей – внутренний и внешний. В первом случае последовательность помещается в ячейку реляционной таблицы с соответствующим типом данных. Второй способ предусматривает хранение в таблице базы данных ссылку (обычно объемом в 25 кБайт) на внешний файл, содержащий двоичную последовательность. Внутреннее хранение

двоичной последовательности обеспечивает защищенность объекта на уровне безопасности собственно базы данных, сильную связанность с соответствующими идентификаторами, следовательно, достаточную целостность, согласованность и непротиворечивость (обеспечивается базовыми механизмами реляционной модели данных) [1, 7, 8, 10].

Внешнее хранение обеспечивает сильную связь с сервером приложений, породившим последовательность, не нарушает выровненность строк таблицы, но не реализует связь с индексами, не гарантирует уровень защищенности не ниже, чем в базе данных [1, 6-8, 10]. Наивысшая опасность для надежности хранения обусловлена допустимостью размещения двоичной последовательности вне системной директории (каталога) системы баз данных.

Вариативность способов хранения двоичных последовательностей не обеспечивает однозначность в выборе наилучшего варианта. Свойства каждого из них являются достоинствами или недостатками только в определенных условиях применения. На основании данных положений можно считать задачу исследования взаимосвязи свойств способов хранения двоичных последовательностей в системах баз данных достаточно актуальной. Исследование проведено для систем баз данных на платформе ORACLE.

2. Постановка задачи. Имеется тестовый набор двоичных последовательностей объемом 14.2, 923 и 0.957 мБайт, предназначенных для хранения в базе данных ORACLE. Последовательности представляют собой аудио, видео записи и оцифрованную фотографию (статистические межсимвольные зависимости и преобладания символов одного вида не устранялись). Для их хранения имеется база данных, содержащая два односхемных фрагмента, отличающихся только типом данных атрибута, предназначенного для хранения больших двоичных последовательностей. В первом случае применен тип BLOB, а во втором – BFILE. Инфологическая модель базы данных предусматривает некоторую семантическую окраску, предназначенную для получения связанных таблиц (выполнение запросов с вложениями – рекурсивного типа).

Предполагается, что основными показателями эффективности использования больших двоичных последовательностей являются время их загрузки, выгрузки и удаления (t_3 , t_B , t_U соответственно), а также затраты дискового пространства для размещения на сервере – V . Это обусловлено значимостью соответствующих свойств – быстродействия и ресурсоемкости, компонента информационной системы (результативность не рассматривается) [2, 9]. Ожидается, что влияние на значения показателей будут оказывать: характеристики аппаратно-программной платформы сервера баз данных – x_1 ; размер большой

двоичной последовательности – x_2 ; тип запроса в базу данных по манипулированию последовательностью – x_3 . Другие факторы, влияющие на изменение значимых свойств исследуемой системы, отнесены в ограничения вследствие роста сложности исследовательской задачи.

Для обеспечения применимости линейного регрессионного анализа необходимо доказать, что значения показателей t_3 , t_B , t_Y и V распределены нормально [3]. Доказательство может быть построено в соответствии с логикой центральной предельной теоремы, когда изыскивается совокупность случайных составляющих исследуемого параметра и оценивается их вклад в текущее значение. Для данного случая имеется достаточная статистика, позволяющая получить доверительные оценки параметров распределения случайных величин t_3 , t_B , t_Y и V , на основании которых можно утверждать, что они распределены нормально при неизменных условиях реализации (постоянных значениях x_1, x_2, x_3).

Таким образом, задача состоит в определении вида зависимости значений временных и ресурсных показателей вариантов хранения двоичных последовательностей в базах данных ORACLE от характеристик (типа) аппаратно-программной платформы сервера, собственно размера двоичной последовательности и типа запроса в базу данных на манипулирование соответствующим объектом хранения двоичной последовательности. Формально необходимо восстановить уравнение множественной линейной регрессии вида:

$$y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + \varepsilon. \quad (1)$$

Здесь y обозначает одну из заявленных характеристик основных способов хранения двоичных последовательностей (среднее время выполнения запроса на выгрузку, загрузку или удаление двоичной последовательности); b_1 – коэффициенты линейной регрессии; b_0 – постоянная составляющая (необъяснимая) регрессии; ε – случайная погрешность (возмущение).

Влияние возмущения обычно оценить не возможно, поэтому перейдем к оценкам истинных значений параметров y_j , и в уравнении регрессии (1) не будем его учитывать. Тогда задачей исследования будет установление вида функции регрессии на двух выборках случайных величин показателей для двух вариантов хранения двоичных последовательностей, а именно:

$$\widehat{y}_1 = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3, \quad (2)$$

$$\widehat{y}_2 = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3. \quad (3)$$

Ограничения, при которых решается задача, сформулированы выше, а именно, случайное возмущение считается не соизмеримо малым – $\varepsilon = 0$; иные кроме x_1, x_2, x_3 факторы считаются незначимыми, а их влияние учтено постоянной составляющей b_0 .

3. Проведение эксперимента и восстановление уравнения регрессии. Исследование вида зависимости эффективности хранения больших двоичных последовательностей от совокупности разнородных факторов (в том числе – способа хранения последовательности) предполагает получение состоятельной статистики на правдоподобных операциях. Для получения необходимых исходных данных выполнен эксперимент с использованием системы баз данных ORACLE версии 11g R2. Целевая установка эксперимента состоит в определении степени зависимости времени выполнения запроса (вставка – загрузка, выборка – выгрузка, удаление) и затрачиваемого дискового ресурса от варианта организации хранения двоичной последовательности, производительности аппаратно-программной платформы сервера баз данных.

Время модификации последовательности не исследовалось ввиду не востребованности данной информационной операции. Действительно, при хранении последовательности во внешнем файле модификация данных осуществляется приложением-сервером, с помощью которого она была создана. В случае внутреннего хранения (домен BLOB) модификация возможна только после выгрузки последовательности во внешнюю прикладную программу с последующей загрузкой результата в файл данных. Данные положения позволяют обосновать отсутствие актуальности исследования времени модификации данных непосредственно в базе данных (типовая операция UPDATE).

Исходные данные получены в результате дробнофакторного эксперимента со значимостью $\alpha = 0.05$ и $\beta = 0.02$. Вероятность ошибки второго рода вычислена по семейству оперативных кривых [4], описывающих изменение мощности эксперимента при фиксированном числе степеней свободы случайного отклика и переменном числе реплик. Принадлежность экспериментальных данных генеральной совокупности определена на основании критерия Диксона [5] для текущих условий экспериментирования.

Экспериментальный стенд представляет собой совокупность трех независимых друг от друга серверов баз данных с различной аппаратно-программной платформой. Конкретные показатели платформы инвариантны результатам эксперимента [9, 11]. Основное условие – существенное различие вариантов платформ по производительности на операциях ввода/вывода и специфических вычислениях при выполнении запросов в базу данных на языке SQL (интерпретация текста

запроса, планирование процедуры выполнения, обмен с памятью). В эксперименте использованы имеющиеся у исследователей платформы. Вариант А: двухпроцессорная архитектура AMD Phenom™ II X4 925 2,81 ГГц /2048 мБайт/160 гБайт; вариант Б: 2x Intel(R) Core(TM) i5 CPU 750 2,67 ГГц /8192 мБайт/1 тБайт; вариант В: 2xIntel Xeon E5-2680V2/3x8192 мБайт/2x1 тБайт SATA RAID0.

Во всех вариантах применена локальная вычислительная сеть 100 мб/с Ethernet без внешней нагрузки (задержки в сети минимальны и одинаковы для всех вариантов аппаратно-программной платформы). Клиентское программное обеспечение ORACLE развернуто на персональном компьютере на базе архитектуры Intel Core i5 2390T/16384 мБайт/750 мБайт, одним и тем же для всех трех серверных платформ (реплики выполнялись последовательно).

Логическая структура базы данных для способов хранения отличалась только типом данных атрибута, предназначенного для ведения двоичной последовательности. В первом случае применен тип BLOB, во втором – BFILE. Мощность контрольного примера одинакова – 9 кортежей. Логическая структура предусматривает четыре связанных между собой отношения, все связи идентифицирующие. Степень нормализации до нормальной формы Бойса-Кодда. Для измерения времени отклика базы данных использован встроенный хронометр утилиты ORACLE SQLDeveloper.

В обеспечение применимости методов регрессионного анализа выполнено исследование закона распределения случайной величины отклика. Сформулировано предположение о нормальном распределении и выполнена проверка принадлежности случайной величины (времени выполнения запроса) нормально распределенной генеральной совокупности. Для всех уровней фактора (вид запроса в комбинации с вариантом аппаратно-программной платформы и размером двоичной последовательности) получены вероятности того, что время выполнения запроса отклоняется от своего математического ожидания на величину, меньшую, чем утроенное среднеквадратичное отклонение. Они принадлежат диапазону от 0.70628 до 0.98826. В соответствии с данными результатами сформулирован вывод о состоятельности гипотезы о смещенном нормальном распределении времени отклика базы данных на запрос по всем уровням фактора. Для построения уравнения регрессии выполнена спецификация – обоснование набора независимых переменных (факторов влияния), существенно влияющих на время выполнения запроса. По сути, доказана применимость сформулированного при постановке задачи набора факторов, влияющих на

время выполнения запроса по загрузке, выгрузке или удалению большого двоичного объекта.

Факторы влияния в исследуемой области качественные. Для построения математической модели применен способ оцифровки качественных параметров, варьирующихся в трех уровнях. Использовано серийно-порядковое кодирование. Данный метод предполагает формирование всех комбинаций уровней фактора, например, три варианта применяемой аппаратно-программной платформы сервера баз данных, с последующей нумерацией по порядку внутри серий – вариативных условий экспериментирования. Сформированный код играет роль индекса при соответствующей независимой переменной x_{ij} . Результаты кодировки факторов представлены в таблице. Здесь значение "1" указывает на факт использования данного параметра в опыте.

Таблица. Кодировка переменных уравнения регрессии

Наименование уровня фактора	Возможные значения	Значение фиктивной переменной		Математическое обозначение в уравнении
Тип аппаратно-программной платформы	Вариант А	0	0	x_{11}
	Вариант Б	1	0	x_{12}
	Вариант В	0	1	x_{13}
Размер хранимого объекта	14,2 мБайт	0	0	x_{22}
	923 мБайт	1	0	x_{21}
	0,957 мБайт	0	1	x_{23}
Тип запроса	Удаление	0	0	x_{31}
	Выгрузка	1	0	x_{32}
	Загрузка	0	1	x_{33}
Время выполнения запроса при способе хранения	BLOB			y_1
	BFILE			y_2

Для определения коэффициента множественной корреляции измеренные в ходе опытов величины времени выполнения запросов (получены с помощью утилиты SQLDeveloper) введены в лист данных пакета прикладных программ "Statistika 6.0". С помощью данного программного средства выполнена статистическая обработка. Результаты для способа хранения BLOB получены на 270 наблюдениях. Точечная диаграмма распределения времени манипулирования тремя большими двоичными объектами представлена на рисунке 1. Здесь же приведена линейная аппроксимация роста задержки выполнения запроса в зависимости от объема обрабатываемого большого двоичного объекта.

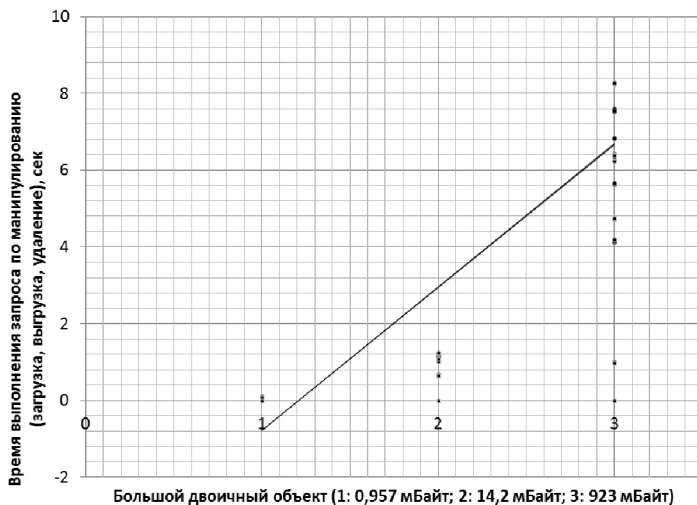


Рис. 1. Время выполнения операции манипулирования для трех разных по объему больших двоичных объектов

На графике рисунке 2 представлены результаты измерения времени выполнения основных операций по загрузке, выгрузке и удалению по основным способам хранения большой двоичной последовательности.

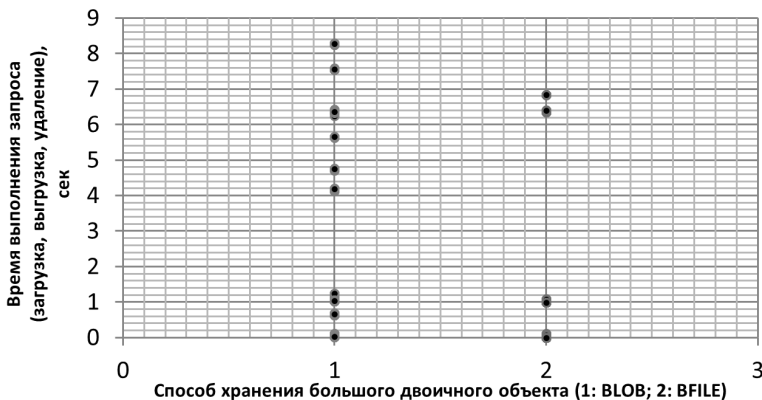


Рис. 2. Время манипулирования данными большого двоичного объекта для двух вариантов организации хранения

Отклики времени манипулирования большими двоичными объектами на различных аппаратно-программных платформах сервера баз данных приведены на рисунке 3.

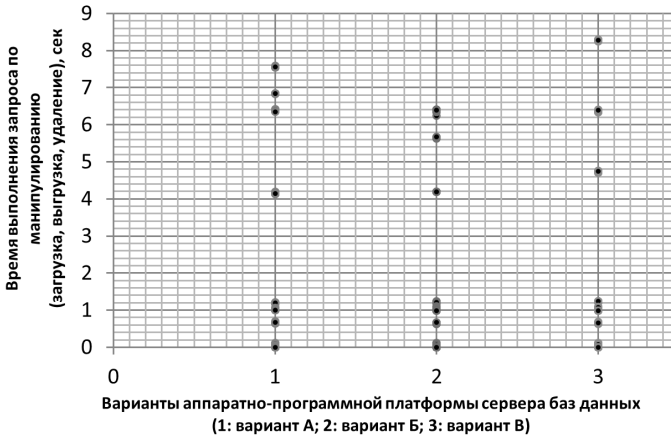


Рис. 3. Время выполнения единичных операций манипулирования большими двоичными последовательностями на различных аппаратно-программных платформах

Коэффициент множественной корреляции для способа хранения BLOB составляет 0.972, что позволяет считать зависимость между факторами и откликом как сильную. Для способа BFILe статистика рассчитана также на основании 270 измерений и показывает сильную множественную корреляцию между фактором и откликом: $R = 0.714$. Методом наименьших квадратов восстановлены уравнения множественной линейной регрессии для основных способов хранения:

$$y_1 = 55.0203 + 0.2382 \cdot x_{12} + 0.4793 \cdot x_{13} - 61.8185 \cdot x_{22} - 63.0109 \cdot x_{23} + 8.566 \cdot x_{32} + 12.1851 \cdot x_{33}, \quad (4)$$

$$y_2 = 15.0655 + 0.0009 \cdot x_{12} - 0.0013 \cdot x_{13} - 21.659 \cdot x_{22} - 22.7687 \cdot x_{23} + 23.4377 \cdot x_{32} - 0.0254 \cdot x_{33}. \quad (5)$$

Детальное рассмотрение уравнений (4), (5) и результатов статистической обработки данных эксперимента позволяет сформулировать ряд утверждений. Очевидно, что производительность аппаратно-программной платформы оказывает прямое влияние на время выполнения запроса. При этом существенно более производительная платформа по варианту В не дает подавляющего преимущества в сравнении с показателями платформы по варианту А. Преимущество внешнего способа хранения (BFILe) по времени загрузки, выгрузки и удаления объекта определяется отсутствием дополнительной ступени преобразования двоичной последовательности при взаимодействии с файлом данных и словарем данных. Не зависимо от размера двоичной последовательности этот выигрыш приблизительно втроекратный в среднем. Иллюстрация приведена на рисунке 4.

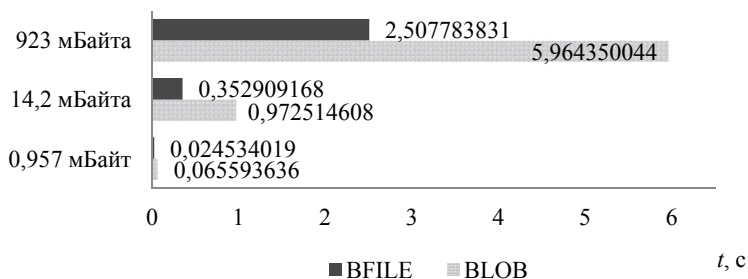


Рис. 4. Сравнительная диаграмма отличия временных задержек при манипулировании двоичными последовательностями разных объемов

Еще одним очевидным результатом является тот факт, что среднее время манипулирования двоичной последовательностью прямо зависит от ее размера. При этом для последовательностей размером до 1 мБайта преимущества способа хранения не наблюдается – среднее время манипулирования отличается в рамках статистической погрешности измерений. С увеличением размера двоичной последовательности наблюдается расхождение производительности системы хранения по манипулированию данными: как для размера порядка 15 мБайт, так и для последовательностей размера около 1 гБайта – задержки для способа внешнего хранения меньше в пять – восемь раз, чем для способа внутреннего хранения. Сравнительная диаграмма представлена на рисунке 5 (здесь показано изменение времени загрузки последовательности для разных вариантов аппаратно-программной платформы). Задача восстановления вида зависимости времени манипулирования двоичной последовательностью от ее размера при разных способах хранения может быть исследована отдельно.

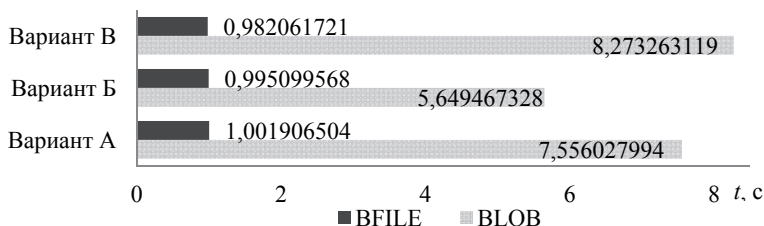


Рис. 5. Сравнительная диаграмма зависимости времени манипулирования данными при разных объемах и способах хранения большой двоичной последовательности

Эффект от применения различных способов хранения больших двоичных последовательностей иллюстрируется диаграммой на рисунке 6. Здесь представлены результаты статистического исследования времени, затрачиваемого на загрузку, выгрузку и удаление одинаковых по объему двоичных последовательностей. Получены ожидаемые результаты при учете алгоритмов обработки последовательностей в базе данных. Практическое значение сведений на рисунке 6 состоит в том, что становится очевидной инвариантность выгрузки двоичной последовательности и сильная зависимость времени загрузки от способа хранения. Данный результат является основанием для выбора способа хранения для систем с преобладающей операцией загрузки, например, видео контента.

4. Заключение. Экспериментальные исследования в целом подтверждают противоречивую природу основных способов хранения больших двоичных последовательностей в базах данных на примере систем на основе решений ORACLE. Нельзя утверждать, что какой-либо основной способ хранения подобных объектов лучше. При выборе способа хранения больших двоичных последовательностей основным фактором является предполагаемый режим манипулирования ими (загрузка, выгрузка или удаление). Для систем, накапливающих данные (автоматизированные системы контроля, автоматизированные системы научных исследований, системы автоматизированного проектирования и им подобные), предпочтительнее использовать внешний способ хранения больших двоичных последовательностей (тип данных BFILE). В системах управления и регулирования, где высока доля операций по выгрузке больших двоичных последовательностей, предпочтительнее применение внутреннего способа хранения (тип данных BLOB).

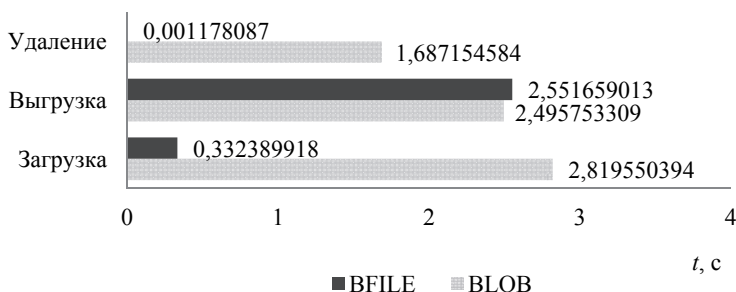


Рис. 6. Сравнительная оценка времени манипулирования двоичной последовательностью при использовании разных способов хранения

Данные утверждения обоснованы для случаев хранения объектов размером в десятки мБайт и более. Двоичные объекты не значительных размеров (фотографии, аудио записи в форматах сжатия с потерями и им подобные) инвариантны к применяемым способам хранения с точки зрения временных задержек при манипулировании. В данном случае обоснование выбора способа хранения целесообразно основывать на функциональных показателях и свойствах, обеспечивающих защищенность.

Исследование подтвердило физическую природу механизмов и алгоритмов обработки и хранения больших двоичных последовательностей в базах данных, основанных на решениях ORACLE. Статистическая независимость исследования позволяет воспроизвести его для других популярных платформ. При этом следует ожидать результатов, объясняемых заложенными в программное обеспечение серверов баз данных программно-техническими решениями. Воспроизведение представленного исследования может быть полезно при оценивании возможностей по хранению больших двоичных последовательностей в так называемых «проприетарных» системах с неизвестными или неясными механизмами манипулирования данными.

Литература

1. *Гарсиа-Молина Г., Ульман Дж., Уидом Дж.* Системы баз данных. Полный курс: пер. с англ. // М.: Издательский дом "Вильямс". 2003. 1088 с.
2. *Анфилатов В.С., и др.* Системный анализ в управлении: Учеб. пособие. / Под ред. А.А. Емельянова. // М.: Финансы и статистика. 2002. 368 с.
3. *Бронштейн И.Н., Семендяев К.А.* Справочник по математике для инженеров и учащихся втузов. – 13-е изд., исправленное. // М.: Наука, Гл. ред. физ.-мат. лит. 1986. 544 с.
4. *Монтгомери Д.К.* Планирование эксперимента и анализ данных: пер. с англ. // Л.: Судостроение. 1980. 384 с.
5. *Блохин В.Г., Глудкин О.П., Гуров А.И., Ханин М.А.* Современный эксперимент: подготовка, проведение, анализ результатов / Под ред. О.П. Глудкина // М.: Радио и связь. 1997. 232 с.
6. *Алапати С.Р.* Oracle Database 11g: руководство администратора баз данных: пер. с англ. // М.: ООО "И.Д. Вильямс". 2010. 1440 с.
7. *Kaim T.* Oracle для профессионалов: пер. с англ. // Спб.: ООО "ДиаСофтЮП". 2003. 672 с.
8. *Панченко Б.Е.* Каркасное проектирование доменно-ключевой схемы реляционной базы данных // Кибернетика и системный анализ. 2012. № 3. С. 174–187.
9. *Стукалов В.А., Савотченко С.Е.* Моделирование процесса загрузки данных в автоматизированные библиотечные информационные системы с открытым кодом // Фундаментальные исследования. 2013. № 8–1. С. 69–72.
10. *Григорьев Ю.А.* Алгоритм синтеза частично оптимальной схемы реляционной базы данных // Наука и образование: электронное научно-техническое издание. 2012. № 1. С. 31–35.

11. *Сорокин А.А., Королев С.П., Смагин С.И., Поляков А.Н.* Макет отказоустойчивой информационной системы для облачного хранения наборов научных данных // Вычислительные технологии. 2013. Том 18. № 1. С. 87–95.

References

1. Garsia-Molina G., Ul'man Dzh., Uidom Dzh. *Sistemy baz dannyh. Polnyj kurs*. [The database system. Full course]; Per. s angl. M.: Izdatel'skij dom "Vil'jams", 2003. 1088 p. (In Russ.).
2. Anfilatov V.S., et al. *Sistemnyj analiz v upravlenii* [System analysis in management]; Ucheb. posobie. Ed. A.A. Emel'janova. M.: Finansy i statistika [Finance and Statistics], 2002. 368 p. (In Russ.).
3. Bronshtejn I.N., Semendjaev K.A. *Spravochnik po matematike dlja inzhenerov i uchashhihsja vtuzov* [Handbook of mathematics for engineers and students of technical colleges]. M.: Nauka, Gl. red. fiz.-mat. lit., 1986. 544 p. (In Russ.).
4. Montgomeri D.K. *Planirovanie jeksperimenta i analiz dannyh*. [Planning the experiment and analysis of data]; Per. s angl. L.: Sudostroenie, 1980. 384 p. (In Russ.).
5. Blohin V.G., Gludkin O.P., Gurov A.I., Hanin M.A. *Sovremennyj jeksperiment: podgotovka, provedenie, analiz rezul'tatov* [Modern experiment preparation, conduct, analysis of the results]. Ed. O.P. Gludkina. M.: Radio i svjaz', 1997. 232 p. (In Russ.).
6. Alapati S.R. *Oracle Database 11g: rukovodstvo administratora baz dannyh* [Oracle Database 11g: Administrator's Guide database]; Per. s angl. M.: OOO "I.D. Vil'jams", 2010. 1440 p. (In Russ.).
7. Kajt T. *Oracle dlja professionalov* [Oracle professionals.]. Spb.: OOO "DiaSoftJuP", 2003. 672 p. (In Russ.).
8. Panchenko B.E. [Framework design of a domain-key scheme of a relational database]. *Kibernetika i sistemnyj analiz – Cybernetics and system analysis*. 2012. no. 3. pp. 174–187. (In Russ.).
9. Stukalov V.A., Savotchenko S.E. [Simulation of loading data into the automated library information systems with open source]. *Fundamental'nye issledovanija – Fundamental research*. 2013. no. 8-1. pp. 69–72. (In Russ.).
10. Grigor'ev Ju.A. [The algorithm of synthesis partially optimal relational database schema]. *Nauka i obrazovanie: jelektronnoe nauchno-tehnicheskoe izdanie – Sciences and Educational: electronic science-technical issue*. 2012. no. 1. pp. 31–35. (In Russ.).
11. Sorokin A.A., Korolev S.P., Smagin S.I., Poljakov A.N. [Layout fault-tolerant information system for the cloud storage of scientific data sets]. *Vychislitel'nye tehnologii – Computational technologies*. 2013. vol. 18. no. 1. pp. 87–95. (In Russ.).

Тараканов Олег Викторович — к-т техн. наук, доцент, сотрудник, Академия ФСО России. Область научных интересов: Базы данных, структуры данных, информационные технологии, обработка данных. Число научных публикаций — 100. ole_g66@list.ru; 302040, г. Орел, ул. 8 Марта, д. 66, кв. 222; р.т.: +79208019202.

Tarakanov Oleg — Ph.D., assistant professor, collaborator, Academy of the Federal Security Service of the Russian Federation. Research interests: Database, data model, information technology, data processing. The number of publications — 100. ole_g66@list.ru; 8 Marta, 66, fl. 222, Oryol, 302040; office phone: +79208019202.

РЕФЕРАТ

Тараканов О.В. Сравнительное исследование способов хранения больших двоичных последовательностей в базах данных ORACLE.

В статье рассматриваются вопросы сравнения основных способов хранения больших двоичных последовательностей в базах данных ORACLE по их отдельным прагматическим свойствам. Исследованы типы данных BLOB и BFILE. Статья содержит отчет о результатах эксперимента, проведенного над некоторой базой данных, представленной двумя версиями. В одной применен тип данных BLOB, а в другой – BFILE. В результате статистической обработки полученных экспериментальных данных восстановлены уравнения линейной множественной регрессии. Они являются математическими описаниями корреляции между временем выполнения запроса и факторов, обусловленных типом аппаратно-программной платформы сервера баз данных, размером двоичной последовательности и собственно типом выполняемого запроса (выгрузка, загрузка, удаление). Восстановленные уравнения линейной множественной регрессии интерпретированы для применения по обоснованию выбора способа хранения больших двоичных последовательностей в базах данных ORACLE.

SUMMARY

Tarakanov O.V. A Comparative Research into Ways to Store Binary Large Objects in the ORACLE Databases.

The article deals with the comparison of the individual pragmatic properties of the main ways to store binary large sequences in databases ORACLE. The research into data types BLOB and BFILE is conducted. The article contains a report on the results of an experiment conducted on some database, presented in two versions. One type of data applies BLOB, and the other - BFILE. As a result of statistical processing of the experimental data, linear multiple regression equations are restored. They are mathematical descriptions of the correlation between the query execution time and factors arising from the type of hardware and software platform of the database server, the size of the binary sequence and the actual type of the query executed (uploading, modification, deletion). Recovered linear multiple regression equations are interpreted for use in justifying the choice of method of storing large binary sequences in the ORACLE databases.