

УДК 004.658.2

А.Б. Кунгурцев, канд. техн. наук, проф.,
С.Л. Зиноватная, канд. техн. наук,
Мунзер Аль Абдо, магистр,
Одес. нац. политехн. ун-т

ИМИТАЦИОННАЯ МОДЕЛЬ ТАБЛИЦ РЕЛЯЦИОННОЙ БАЗЫ ДАННЫХ

О.Б. Кунгурцев, С.Л. Зиноватная, Мунзер Аль Абдо. Імітаційна модель таблиц реляційної бази даних. Представлено алгоритм формування груп таблиц і груп запитів для дослідження поведінки інформаційної системи. Введено класифікацію типів розвитку системи з погляду зміни змісту таблиц і складу запитів до бази даних.

А.Б. Кунгурцев, С.Л. Зиноватная, Мунзер Аль Абдо. Имитационная модель таблиц реляционной базы данных. Представлен алгоритм формирования групп таблиц и групп запросов для исследования поведения информационной системы. Введена классификация типов развития системы с точки зрения изменения содержимого таблиц и состава запросов к базе данных.

A.B. Kungurtsev, S.L. Zinovatnaya, Munzer Al Abdo. Simulation model of relational database tables. The algorithm of forming tables groups and queries groups for studying the behaviour of an information system is presented. Classification of types of the system development is introduced in terms of changing the tables content and database queries composition.

В последние годы информационные системы (ИС) используются повсеместно. Во многих случаях в связи с ростом количества данных, увеличением числа пользователей, появлением новых задач начинают возникать проблемы, связанные с недостаточной производительностью ИС, которая, в первую очередь, определяется временем выполнения запросов к реляционной базе данных (РБД). Существуют различные способы повышения производительности, не связанные с заменой аппаратуры, например, реструктуризация РБД, использование материализованных представлений, репликация данных, настройка индексной структуры и т.п. Применение указанного предусматривает предварительное исследование не только структуры РБД, но и ее работы в течение определенного периода времени.

© А.Б. Кунгурцев, С.Л. Зиноватная, Мунзер Аль Абдо, 2011

Помимо рабочей среды необходима тестовая среда, предназначенная для обнаружения узких мест в производительности разрабатываемого приложения, которая также используется для [1]:

— проверки на масштабируемость к требуемой нагрузке (количеству пользователей); тестирования производительности приложения, т.е. способности обработать нагрузку в допустимых пределах времени ожидания ответа;

— проверки работоспособности настроек, в частности, того, что они приводят к повышению, а не к снижению производительности ИС.

В процессе разработки тестовой системы необходимо использовать данные, являющиеся репрезентативными относительно реально работающей системы. В различных СУБД существуют средства отладки и оптимизации запросов. Однако, даже физическое размещение данных в таблице имеет значение для разработки оптимизатором плана запроса [2].

Приведенные доводы являются основанием для разработки специальных имитационных моделей РБД для проверки эффективности ее модернизации и настройки.

Обычно имитационные модели строятся для поиска оптимального решения в условиях ограничения ресурсов, когда другие модели оказываются слишком сложными. Имитационное моделирование — конструирование модели и ее аналитическое применение для изучения некоторой проблемы.

Имитационная модель таблиц РБД может использоваться самостоятельно, например, для анализа эффективности предлагаемой индексной структуры, анализа влияния количества записей и размеров полей на производительность работы РБД. Кроме этого она может быть элементом при построении других имитационных моделей РБД [3].

Предлагается методика определения параметров имитационной модели РБД для прогнозирования поведения ИС при различных вариантах развития, позволяющая формализовать исследование системы.

Пусть имеется РБД, состоящая из множества T таблиц. Предварительно будем считать, что отсутствуют требования секретности структуры и хранимых данных в РБД при моделировании.

Для того, чтобы условия испытания модели отвечали реальным, необходимо проанализировать запросы к РБД в течение определенного периода времени P с целью выяснения частоты использования таблиц, полей и операций над ними и получения другой информации.

Для этого вводится следующее представление множества Q запросов, в котором каждый запрос представляется кортежем

$$q = \langle T_q, W, \text{Typ}, \langle t_{st}, \tau \rangle \rangle,$$

где $T_q = \langle T_i, \langle F, R \rangle \rangle, i = \overline{1, N_{T_q}}$ — множество таблиц T_i РБД, участвующих в запросе q ;

N_{T_q} — количество таблиц T_i ;

$F = \{F_j\}, j = \overline{1, N_{T_iQ}}$ — множество полей таблицы T_i ;

N_{T_iQ} — количество задействованных в запросе q полей F_j таблицы T_i ;

$R = \{R_k\}, k = \overline{1, N_{F_jQ}}$ — множество ролей поля F_j в запросе q ;

N_{F_jQ} — количество ролей поля F_j ;

$R_k \in \{P_R, O_R, G_R, C_R\}$; P_R — операция выполнения в запросе q вывода поля F_j в результирующую таблицу T_i ; O_R — выполнение сортировки данных по полю F_j ; G_R — выполнение группировки данных по полю F_j ; C_R — использование поля F_j в условии отбора строк таблицы T_i ;

W — идентификатор рабочего места, от которого в РБД направлен запрос q ;

$\text{Typ} \in \{TD, TI, TU, TS\}$ — тип запроса q , TD , TI , TU , TS — запрос на удаление, добавление, обновление, выборку данных, соответственно;

t_{st} — время запуска запроса;

τ — продолжительность выполнения запроса, запущенного в момент времени t_{st} .

В результате анализа множества запросов $Q' \subset Q, \text{Type} \in \{TD, TI, TU\}$ на модификацию данных, поступивших в РБД в течение временного периода P , и с учетом состояния РБД в начале периода исследования, может быть получено представление отдельной таблицы

$$T_T = \langle f_{\text{Ins}}(t), f_{\text{Del}}(t), \langle F_j, f_a \rangle \rangle,$$

где $f_{\text{Ins}}(t)$ — функция, определяющая количество операторов, выполняющих добавление строк в таблицу, в единицу времени;

$f_{\text{Del}}(t)$ — функция, определяющая количество операторов, выполняющих удаление строк из таблицы, в единицу времени;

f_a — функция изменения распределения значений в каждом поле таблицы T_T .

На основании функций $f_{\text{Ins}}(t)$ и $f_{\text{Del}}(t)$ с учетом количества N_0 строк в таблице в начале периода P можно построить функцию $f_{N_T}(t)$, определяющую изменение количества N_T строк в таблице от времени.

Функция $f_{N_T}(t)$ может быть приведена к следующим видам в зависимости от характера изменения количества N_T строк в таблице (рис. 1, где t_I и t_D — момент времени начала выполнения запроса на добавление строки в таблицу T_T и удаление строки из T_T , соответственно):

— при линейном изменении $f_{N_T}(t) = kt + N_0$; k — коэффициент, определяющий скорость изменения N_T в течение периода P (рис. 1, а);

— при нелинейном изменении функция $f_{N_T}(t)$ может быть представлена в виде множества линейных отрезков $P = \{P_i; i = \overline{1, n}\}$; для каждого отрезка определяется собственный коэффициент k_i (рис. 1, б);

— при постоянном количестве строк в таблице $f_{N_T}(t) = N_0, k=0$ (рис. 1, в).

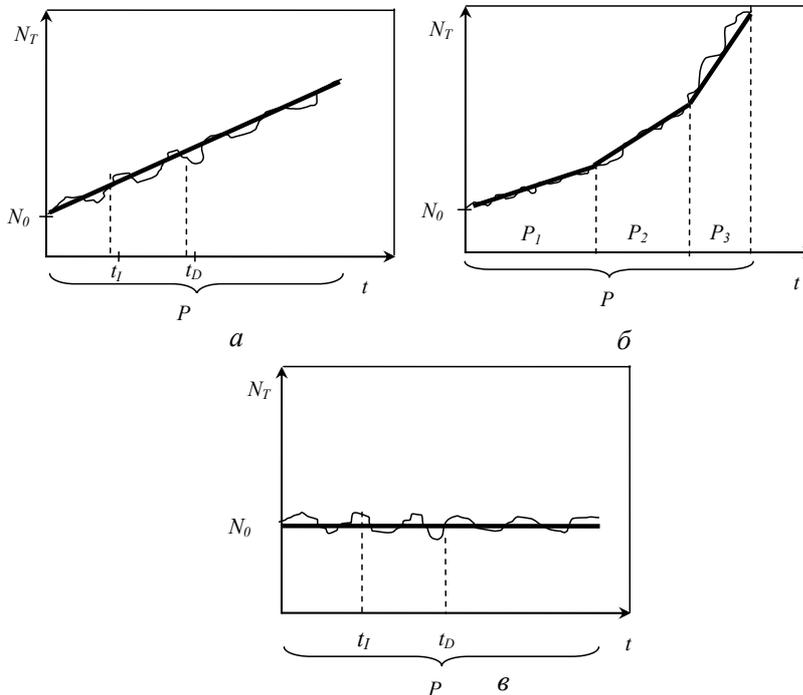


Рис. 1. Вид функции $f_{N_T}(t)$ в зависимости от характера изменения количества строк в таблице: линейное изменение (а); нелинейное изменение (б); не изменяется (в)

Построение имитационной модели таблиц РБД включает следующие этапы (рис. 2):

- сбор информации о запросах, поступающих в ИС в течение периода P ;
- аналитическая обработка собранной информации;
- формирование множества ГТ групп таблиц РБД и множества ГЗ групп запросов.

Группировка таблиц и запросов выполняется по следующему алгоритму.

1. Создать копию РБД в начале процесса анализа запросов для получения в дальнейшем результатов экспериментов, максимально приближенных к реальным условиям, и повторения с использованием модели реальной последовательности запросов в условиях некоторых изменений, например, при модернизации индексной структуры РБД.

2. Определить множество $T' \subseteq T$ таблиц, использовавшихся во множестве Q запросов; множество T' состоит из кортежей вида

$$\langle T_i, v_I, v_D, \langle F_j, v_G, v_O, v_C, v_U \rangle \rangle,$$

где v_I — относительная частота выполнения запросов на добавление строк в таблицу T_i за весь период наблюдения P ;

v_D — относительная частота выполнения запросов на удаление строк из таблицы T_i за весь период наблюдения P ;

v_G — относительная частота выполнения в запросах из множества Q группировки по значениям поля F_j ;

v_O — относительная частота выполнения в запросах из множества Q сортировки по значениям поля F_j ;

v_C — относительная частота выполнения операции выбора строк таблицы T_i по условию с участием значения поля F_j ;

v_U — относительная частота выполнения запросов на обновление поля F_j .

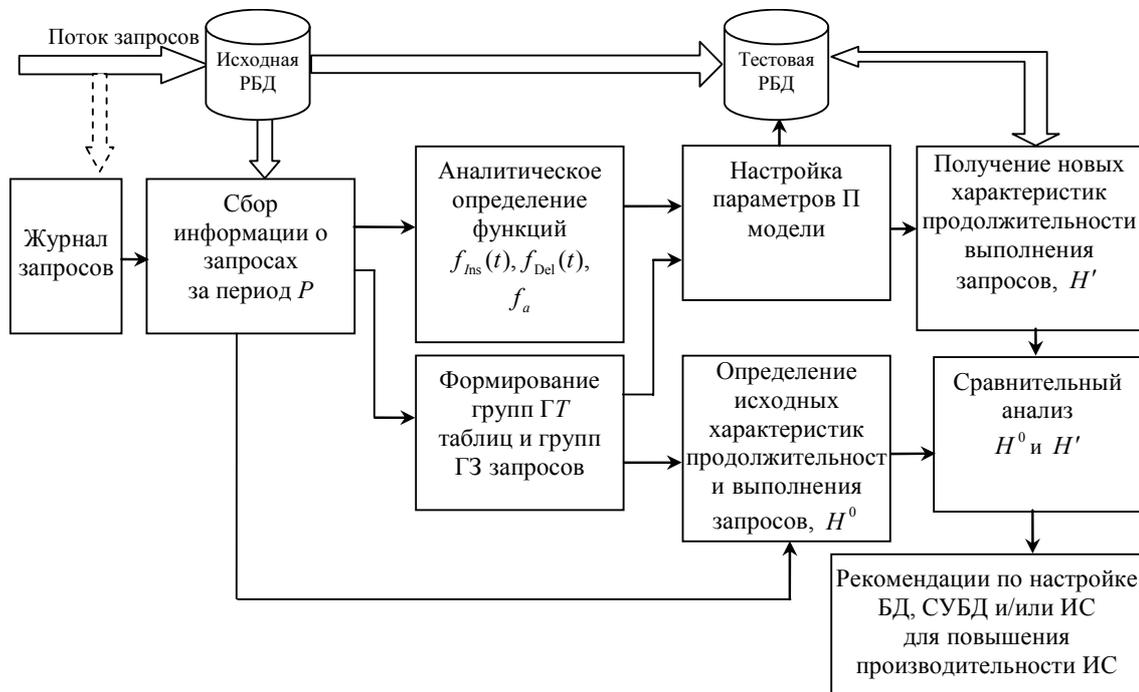


Рис. 2. Построение имитационной модели РБД и моделирование таблиц РБД

Аналогичные характеристики следует связать со временем, т.е. рассматривать не только частоту появления запросов, но и относительную продолжительность их выполнения с использованием перечисленных операций.

3. Ввести пороговое значение ε_T для исключения из рассмотрения редко используемых таблиц T_i , поскольку изменения, относящиеся к ним, очевидно, не принесут ощутимого увеличения производительности. Тогда сумма относительных частот запросов различного типа к таблице

$$v_T = \sum v_r \ll \varepsilon_T,$$

где $r \in \{I, D, G, O, C, U\}$. В результате получено множество таблиц $T'' \subseteq T'$.

4. Для каждой таблицы $T_i \in T''$ составить группу ΓT_{T_i} таблиц, использованных вместе с данной в запросах к РБД, а также составить группу соответствующих запросов ΓZ_{T_i} (рис. 3).

$$T_j \in \Gamma T_{T_i}, \text{ если } \exists q, T_i \in T_q, T_j \in T_q.$$

Тогда $q \in \Gamma Z_{T_i}$, если $T_q \cap \Gamma T_{T_i} \neq \emptyset$.

Моделирование таблиц РБД включает следующие этапы (см. рисунок 2):

— настройка необходимых параметров Π модели;

— определение для каждой группы таблиц $\Gamma T_i, i = \overline{1, N_\Gamma}$, где N_Γ — количество сформированных групп таблиц, исходных характеристик H^0 продолжительности выполнения запросов из группы ΓZ_i , соответствующей ΓT_i ;

— получение новых характеристик H' с помощью моделирования с введенными параметрами Π ;

— сравнительный анализ характеристик H^0 и H' ;

— формулировка рекомендаций по настройке БД, СУБД и/или ИС для повышения производительности ИС.

Для каждой группы таблиц ΓT_i можно выполнить моделирование поведения ИС в период P' , следующий за периодом P , с учетом возможного ее развития. Рассмотрим 4 варианта развития системы (см. таблицу).

Варианты развития системы

Варианты развития системы		Характер изменения размера таблиц $\Gamma T_i, R$	Характер изменения интенсивности поступления запросов на выборку в группе $\Gamma Z_i, S$
1	Стабильные данные/Стабильная выборка	Const	Const
2	Стабильные данные/Динамичная выборка	Const	Variable
3	Динамичные данные/Стабильная выборка	Variable	Const
4	Динамичные данные/Динамичная выборка	Variable	Variable

В столбце R "Const" означает, что функции $f_{ins}(t)$ и $f_{Del}(t)$ остаются постоянными, т.е. изменение количества строк в таблице происходит с той же интенсивностью, что и в исходном периоде P ; "Variable" означает, что изменятся частота запросов на добавление/удаление строк в таблице, соответственно изменяется коэффициент(ы) k функции $f_{NT}(t)$.

В столбце S "Const" означает, что частота поступления запросов на выборку к таблицам из группы ΓT_i не изменяется, и все поступившие запросы выполняются с той же и интенсивностью, что и в исходном периоде P ; "Variable" означает, что изменяется количество запросов на выборку к таблицам из группы ΓT_i в единицу времени.

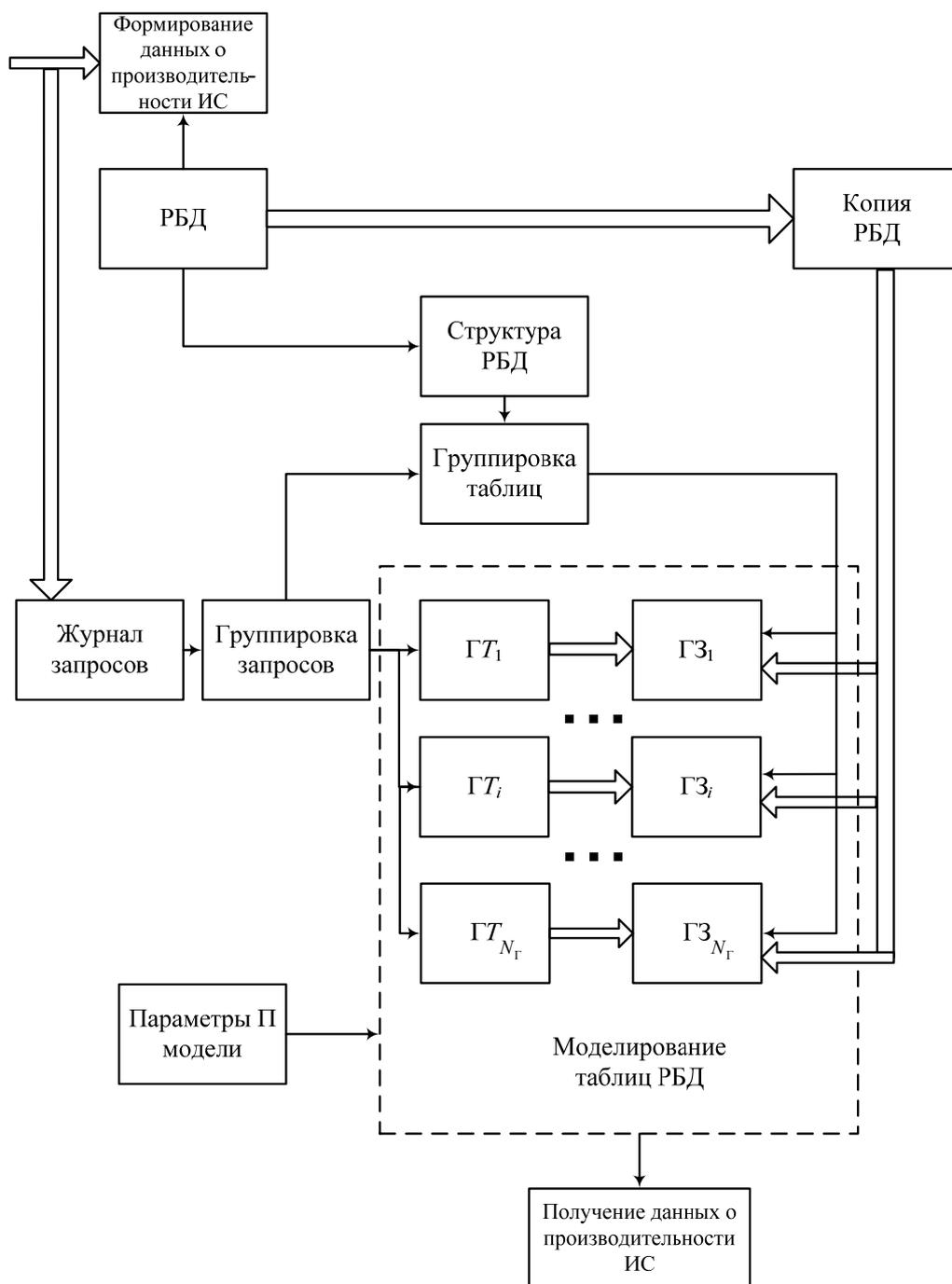


Рис. 3. Моделирование РБД на уровне отдельной таблицы

Во втором — при сохранении интенсивности изменения количества строк в исследуемых таблицах интенсивность запросов на выборку изменяется в течение периода P' .

В третьем — при стабильных видах выборки из таблиц исследуемой группы количество строк в используемых в запросах таблицах изменяется с другой скоростью относительно исходного периода наблюдения.

В четвертом, наиболее общем, предполагается такое развитие системы, при котором произвольно изменяются и количество строк в таблицах, и набор и/или частота выполнения запросов на выборку к этим таблицам.

Примеры развития системы для предметной области “Продажа товаров по предварительным заказам” по каждому из вариантов, соответственно:

— заказы на продажу товаров поступают с прежней интенсивностью, количество запросов к таблицам на выборку от рабочих мест ИС не изменяется;

— заказы продолжают поступать с прежней интенсивностью, но изменилось количество рабочих мест в ИС, которым требуется получать отчеты, сформированные на основе данных из таблиц РБД, что приводит к изменению интенсивности запросов на выборку данных;

— возросло количество клиентов и заказов, следовательно, увеличилась скорость добавления строк в таблицы РБД, но не изменилось количество рабочих мест ИС, которым требуются отчеты, формируемые в результате выполнения запросов на выборку к таблицам РБД;

— возросло количество клиентов и заказов, а также количество рабочих мест, и, следовательно, частота появления запросов на выборку увеличилась.

При настройке параметров модели должен быть учтен вариант развития ИС и соответствующим образом реализовано множество запросов, поступающих в исследуемую РБД. В частности, возможно, потребуется выполнить преобразование заданной пользователем функции $f_{N_T}(t)$ в последовательность запросов на добавление и удаление строк в таблице для периода

$$P': f_{N_T}(t) \Rightarrow \begin{cases} f_{\text{Ins}}(t) \\ f_{\text{Del}}(t) \end{cases}.$$

Предложенное представление запросов к РБД от ИС позволяет формализовать преобразование поведения системы в характеристики таблиц, в частности, в количество строк.

Введенная классификация вариантов развития системы позволяет вести настройку параметров модели для исследования поведения в условиях “что, если...”, задаваемых пользователем.

Литература

1. Кайт, Т. Oracle: Эффективное проектирование приложений / Т. Кайт — М. : Лори, 2008. — 637 с.
2. МакДональд, К. Oracle PL/SQL для профессионалов: практические решения / К. МакДональд, Х. Кац, К. Бек и др. — К. : ДияСофт, 2005. — 560 с.
3. Кунгурцев, А.Б. Задачи имитационного моделирования реляционных баз данных / А.Б. Кунгурцев, С.Л. Зиноватная, Мунзер Аль Абдо // Тр. Одиннадцатой междунар. науч.-практ. конф. “Современные информ. и электрон. технологии”. — Одесса, 2010. — С. 118.

Рецензент д-р техн. наук, проф. Одес. нац. политехн. унта Крисилев В.А.

Поступила в редакцию 28 октября 2010 г.