

УДК 004.658.2:004.051

А.Ю.Левченко

МЕТОДИКА ТЕСТИРОВАНИЯ ПРОИЗВОДИТЕЛЬНОСТИ СИСТЕМ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ

***Аннотация.** Рассмотрена методика нагрузочного тестирования баз данных на основе имитационной модели с целью повышения производительности системы управления. Представлен программный комплекс, реализующий предложенный подход. Приведены результаты экспериментов с использованием комплекса. Для проверки достоверности полученных результатов тестирования предложено использовать доверительный интервал и доверительную вероятность.*

***Ключевые слова:** системы управления базами данных, нагрузочное тестирование, параметры конфигурации.*

A.Y.Levchenko

METHODOLOGY FOR DATABASE MANAGEMENT SYSTEMS PRODUCTIVITY TESTING

***Abstract.** The article represents the methodology of load testing based on simulation for DBMS productivity raising. The testing software and the testing experiment results that were obtained with its application are represented. For validation of the characteristics that were received it is proposed to calculate the confidence interval with specified confidence probability.*

***Keywords:** database management system, load testing, configuration knobs.*

О.Ю.Левченко

МЕТОДИКА ТЕСТУВАННЯ ПРОДУКТИВНОСТІ СИСТЕМ КЕРУВАННЯ БАЗАМИ ДАНИХ

***Анотація.** Розглянуто методику навантажувального тестування баз даних із застосуванням імітаційної моделі для підвищення продуктивності системи керування. Надано програмний комплекс, що реалізує запропонований підхід. Наведено результати експериментів із його застосуванням. Для перевірки достовірності отриманих результатів тестування запропоновано використувувати довірчий інтервал та довірчу ймовірність.*

***Ключові слова:** системи керування базами даних, навантажне тестування, параметри конфігурації.*

Введение. Каждая система управления базами данных (СУБД) конфигурируется с помощью параметров настройки. Как правило, значения параметров по умолчанию не позволяют получить необходимую производительность работы программ. Таким образом, одним из методов повышения производительности корпоративных информационных систем (КИС) является настройка параметров СУБД на основе оценки эффективности использования различных наборов параметров [1]. Исследование производительности предполагает использование нагрузочного тестирования с целью фиксирования и последующего сравнения показателей производительности СУБД исследуемой КИС с заданной нагрузкой в разных диапазонах изменения конфигурационных параметров СУБД. Это позволит принять решение об эффективности заданного набора конфигурационных параметров и их значений и в

© Левченко А.Ю., 2011

нейшем использовать их в качестве рекомендаций к определенному классу КИС.

Методика тестирования. Тестирование базы данных (БД) исследуемой КИС помогает подобрать такие значения параметров настройки СУБД, которые обеспечат повышение производительности БД КИС в заданных реальных условиях функционирования системы. Для этого необходимо сформировать тестовую среду, репрезентативную относительно реальной БД КИС. Таким образом, при построении модели для тестирования необходимо учитывать структуру таблиц, их наполнение данными, запросы пользователей и их последовательности. Нагрузочное тестирование может выполняться с использованием специальной имитационной модели [2] для проверки эффективности настройки параметров СУБД (рис.1).

Целью тестирования является повышение производительности КИС, характери-

стика H которой может выражаться следующими показателями:

TPS – количеством транзакций, обрабатываемых СУБД в секунду;

$KBPS$ – объемом данных, передаваемый СУБД в секунду;

ART – средним временем ответа на запрос в секундах;

ATT – средним временем выполнения транзакции в секундах.

Представим параметры настройки в виде множества $P = \langle \{p_1, v_1\}, \{p_2, v_2\}, \dots, \{p_n, v_n\} \rangle$, где $p_i, i=1, n$ – наименование параметра настройки, v_i – текущее значение параметра p_i . Также представим запросы, перечень которых получен из журнала транзакций, в виде множества $Q = \{q_1, q_2, \dots, q_n\}$.

Методика тестирования СУБД включает следующие этапы:

- 1) построение имитационной модели БД;
- 2) моделирование с заданным набором параметров, сбор статистики и анализ полученных показателей производительности;
- 3) формирование рекомендованных наборов параметров настройки и их значений P_{opt} для исследуемого класса КИС.

Формирование модели состоит из четырех этапов:

- 1) сбор информации о состоянии БД на момент времени P_0 . Сохраняются структура таблиц, связи между ними и наполнение таблиц. Для выполнения данной процедуры можно использовать встроенные функции резервного копирования исследуемой СУБД. Чтобы сократить ресурсы в процессе тести-

рования целесообразно заменить реальные имена и атрибуты таблиц некоторыми переменными, что позволит объединить таблицы с одинаковой структурой;

2) сбор информации о выполняемых запросах за период $\Delta P = P_i - P_0$, где P_i – текущий момент времени. Для этого предварительно в файле настроек СУБД указываются параметры зафиксированных в журнале действий пользователей (сохраненный журнал транзакций);

3) синтаксический разбор (парсинг) шаблонов запросов из журнала транзакций, их последующая классификация. Шаблон запроса – SQL-запрос пользователя к БД, в котором вместо имён таблиц и атрибутов используются некоторые переменные. Использование шаблонов запроса позволяет в значительной степени абстрагироваться от смыслового значения имён таблиц и объединять запросы с одинаковой структурой реляционных операций. Классификация включает отнесение шаблонов запросов анализируемой КИС к одному из классов, описываемых эталонами шаблонов;

4) создание тестовой БД на основе собранной информации о структуре таблиц и их наполнении.

Этап моделирования включает следующие шаги:

- 1) при исходном наборе параметров P_0 определение характеристик общей производительности БД (для всего множества Q) H^{P_0}

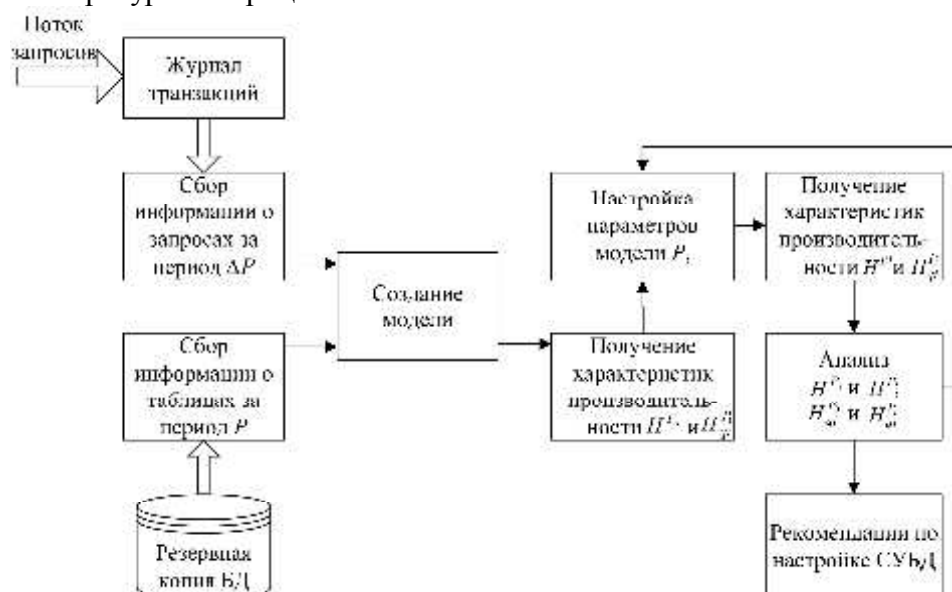


Рис. 1. Нагрузочное тестирование БД

- и отдельных запросов $H_{qi}^{P_0}$;
- 2) изменение значения v_k параметра настройки p_k в наборе P_i в заданном диапазоне значений (диапазон ограничен экспертными рекомендациями по настройке параметров);
- 3) получение новых характеристик H^{P_i} и $H_{qi}^{P_i}$ путем моделирования с заданным набором P_i ;
- 4) сравнение показателей общей производительности H^{P_0} и H^{P_i} и производительностей отдельных запросов $H_{qi}^{P_0}$ и $H_{qi}^{P_i}$;
- 5) повторение пп. 2-4, пока $H_{qi}^{P_i} > H_{qi}^{P_{i-1}}$;
- 6) сравнение показателей общей производительности при изменении всех параметров $H^{P_{opt}}$ и отдельного параметра H^{P_i} ;
- 7) поиск согласованных настроек, таких что $H^{P_i} \rightarrow \max$.

Результаты тестирования в виде *рекомендаций по изменению производительности* КИС – это знания, для хранения которых используется модель представления знаний, основанная на фактах и правилах. Она является эффективной для решения простых однородных задач. Данная модель по критериям уровня сложности (абстрактности) элемента знаний, с которыми работает модель, и универсальности представления знаний – возможности описать знания из различных предметных областей, в полной мере отвечает задаче хранения рекомендаций по настройке параметров конфигурации. Каждой модели класса КИС ставится в соответствие рекомендованный набор параметров настройки СУБД и их значений:

ЕСЛИ <класс КИС>, ТО <параметр; значение (диапазон)>.

Эксперимент. Для реализации предложенной методики разработан комплекс программных средств, реализующий предложенные алгоритмы тестирования параметров настройки СУБД. Рассмотрим работу методики тестирования на примере СУБД PostgreSQL, которая обслуживает исследуемую БД деканата студентов. В качестве модуля выполнения тестов используется ути-

та *pgbench*. Утилита позволяет выполнять тестирование как с заданным набором транзакций типа ТРС-В, так и создавать собственные сценарии транзакций, что позволяет тестировать параметры настройки для разных классов КИС.

Благодаря разработанному модулю тестирования (МТ) можно менять значения выбранных параметров настройки в заданном диапазоне и с заданным шагом и выполнять тестирование для определенной рабочей нагрузки (количество симулируемых клиентов, коэффициент масштабирования, количество транзакций и др.).

В ходе эксперимента протестированы 14 параметров настройки СУБД PostgreSQL, которые были заранее классифицированы по типам [1]. Каждый тип отвечает за настройку различных подсистем СУБД: КЭШ – управление кэш-областью; ДИСК – управление обменом с диском; ТРН – управление транзакциями; ВОСТ – управление резервным копированием и восстановлением; ОПТ – оптимизация выполнения запросов. Описание некоторых параметров из протестированного множества представлено в табл.1. Тестирование выполнялось посредством МТ на рабочей станции со следующими характеристиками: ОС – Windows 7 Профессиональная 64-разрядная; оперативная память – 4096МВ RAM; процессор – Intel® Core™ i5 CPU M430 2.27GHz (4CPUs), ~2,3GHz; файл подкачки – 3507 МБ использовано, 4389МБ свободно.

Параметры тестирования: тип транзакций – ТРС-В (sort of); коэффициент масштабирования – 1; протокол, используемый для представления запросов к серверу – простой (simple); количество рабочих потоков – 2; количество клиентов (количество одновременных сеансов БД) – 10; количество транзакций для каждого клиента – 1000.

Эксперименты выполнялись в соответствии с планом, представленным в табл. 2.

На рис. 2 представлены результаты тестирования параметра *commit_delay*, который задает задержку записи транзакций на диск, полученные средствами визуализации МТ. В процессе эксперимента значение параметра

варьировалось в диапазоне от 0 до 100000 мксек с шагом 10000 мксек. Для получения
1. Описание протестированных параметров СУБД PostgreSQL

Параметр	Тип	Диапазон значений	Значение по умолчанию	Описание
shared_buffers, блоки по 8 кБ	КЭШ	16 – 32768	64	Параметр устанавливает число блоков общего буфера сервера
work_mem, 1 кБ	КЭШ	64 – 32768	1024	Объем памяти для сортировки результата запроса
wal_buffers, блоки по 8 кБ	ДИК, ТРН	4 – 32768	8	Определяет размер буфера журнала транзакций, в котором накапливаются записи перед сбросом их на диск
checkpoint_segments, сегменты по 16 МБ	ВОСТ	1 – 32768	3	Количество сегментов журнала транзакций
checkpoint_timeout, сек	ВОСТ	30 – 3600	300	Максимальное время между автоматическими WAL контрольными точками
commit_delay, мксек	ДИСК	0-100000	0	Задержка между попаданием записи в буфер журнала транзакций и сбросом ее на диск
commit_siblings, мксек	ДИСК	1 -1000	5	
effective_cache_size, блоки по 8кБ	КЭШ, ОПТ	0 – 32768	1000	Примерный объем файлового кэша операционной системы, оптимизатор использует эту оценку для построения плана запроса

более достоверных результатов, при каждом новом значении исследуемого параметра выполнялись 10 повторов теста. В результате были получены графические зависимости показателей производительности H^k от значений параметра v_k . В рассматриваемом примере H – количество транзакций, обрабатываемых СУБД в секунду (TPS). Каждая точка графика соответствует среднему значению показателя производительности (ось ординат) для заданного значения исследуемого параметра (ось абсцисс).

Оценка точности статических характеристик. В процессе тестирования важно выполнять оценку достоверности полученных характеристик. Недостаточное количество экспериментов приводит к

снижению их достоверности, в то время как неограниченное увеличение количества экспериментов повышает время тестирования и требует большего количества ресурсов, что приводит к снижению производительности.

Существуют точечные оценки неизвестного параметра – математическое ожидание (МО) и дисперсия, однако при малом числе наблюдений актуальна оценка попадания параметра в заданный интервал. Чтобы дать представление о точности и надежности вычисленной характеристики в математической статистике используют доверительные интервалы и доверительные вероятности [3].

Согласно центральной предельной теореме значения вычисленной статистической характеристики будут распределены по нормальному закону. Эти числа тем лучше описывают

нормальный закон, чем больше число экспериментов.

2. План эксперимента

Параметр номера опыта	p_1	p_2	...	p_n	H_1	...	H_m
P_1	v_1	v_1		v_1	$H_1^{P_1}$		$H_m^{P_1}$
P_2	v_2	v_2		v_2	$H_1^{P_2}$		$H_m^{P_2}$
P_3	v_3	v_3		v_3	$H_1^{P_3}$		$H_m^{P_3}$
P_4	v_4	v_4		v_4	$H_1^{P_4}$		$H_m^{P_4}$
P_5	v_5	v_5		v_5	$H_1^{P_5}$		$H_m^{P_5}$
...
P_k	v_k	v_k		v_k	$H_1^{P_k}$		$H_m^{P_k}$

Пусть произведено n независимых опытов над случайной величиной X , характеристики которой – МО m , дисперсия D и среднее квадратическое отклонение σ оценки m :

$$m = \frac{\sum_{i=1}^n X_i}{n}; D = \frac{\sum_{i=1}^n (X_i - m)^2}{n-1}; \sigma_m = \sqrt{D}. \quad (1)$$

Тогда доверительный интервал выражается в виде

$$l_\beta = (m - t_\beta \sigma_m; m + t_\beta \sigma_m), \quad (2)$$

где t_β – определяет для нормального закона число среднее квадратических отклонений, которое нужно отложить вправо и влево от центра рассеивания для того, чтобы вероятность попадания в полученный участок была равна β :

$$t_\beta = F^{-1}\left(\frac{1+\beta}{2}\right), \quad (3)$$

где $F^{-1}(x)$ – обратная функция $F(x)$, т.е. такое значение аргумента, при котором нормальная функция распределения равна x (интегральная функция Лапласа). Вычисляется по заранее рассчитанным таблицам, например табл. 3. В случаях, когда распределение результатов наблюдений нормально, но их дисперсия неизвестна, т.е. при малом числе наблюдений n , коэффициент t_β (t -критерий) определяется значением задаваемой доверительной вероятности β и числом степеней свободы $n-1$ по таблицам распределений Стьюдента.

Рассчитаем доверительный интервал на примере выборки, полученной в процессе тестирования параметра

`commit_delay` при заданном значении 50000 мксек (табл. 4), соответствующий доверительной вероятности $\beta = 0,95$.

3. Соответствие значений t -критерия при заданной доверительной вероятности β

β	t_β	β	t_β
0,80	1,282	0,91	1,694
0,81	1,310	0,92	1,750
0,82	1,340	0,93	1,810
0,83	1,371	0,94	1,880
0,84	1,404	0,95	1,960
0,85	1,439	0,96	2,053
0,86	1,475	0,97	2,169
0,87	1,513	0,98	2,325
0,88	1,554	0,99	2,576
0,89	1,597	0,9973	3,000
0,90	1,643	0,999	3,290

4. Тестовая выборка

i	x_i	i	x_i
1	302,53	6	328,29
2	514,80	7	507,34
3	502,37	8	543,67
4	502,64	9	318,13
5	337,97	10	427,12

Вычисляем МО, дисперсию, среднее квадратическое отклонение (1):

$$m = 428,49; D = 9352,19; \sigma = 96,71.$$

По табл. 3 находим $t_\beta = 1,960$. Вычисляем доверительные границы (2):

$$x_{min} = 428,49 - 1,960 \cdot 96,71 = 238,94;$$

$$x_{max} = 428,49 + 1,960 \cdot 96,71 = 618,04.$$

Доверительный интервал:

$$l_\beta = (238,94; 618,04).$$

Вывод. Предложенная методика позволит определить наборы параметров настройки и их значений, использование которых повышает производительность программ КИС. Благодаря методике появляется возможность проанализировать степень влияния каждого отдельного параметра на общую производительность системы и учесть ситуацию, когда изменение отдельной настройки приводит к большему повышению производительности, чем изменение всех значений параметров. Кроме того, анализ влияния на производительность отдельных групп запросов (шаблонов) позволит дополнительно сделать вывод о необходимости дальнейшего их использования в имитационной модели БД.

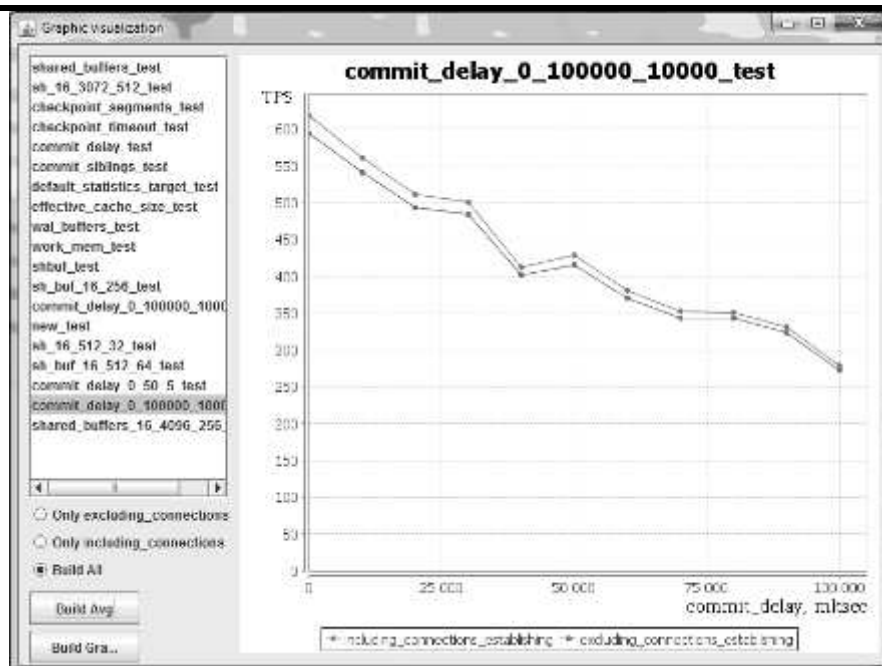


Рис.2. Значения показателя производительности (*tps*) при 10-ти прогонах

Для реализации предложенной методики разработан комплекс программных средств, реализующий алгоритмы тестирования. Модуль тестирования позволяет выполнять тестирования различных параметров конфигурации СУБД PostgreSQL с использованием имитационной модели исследуемой КИС. Был проведен эксперимент по тестированию 14 параметров конфигурации СУБД и получены рекомендации для настройки параметров настройки СУБД КИС деканата студентов ВУЗа. Для оценки достоверности полученных характеристик предлагается рассчитывать доверительный интервал по заданной доверительной вероятности.

Список использованной литературы

1. Блажко А.А. Модели для автоматизированной оптимизации производительности СУБД / А.А. Блажко, А.Ю. Левченко, А.С. Пригожев // *Радиоелектронні і комп'ютерні системи*. – ХАІ, Харків: – 2010. – № 7(48). – С. 24–29.

2. Зиноватная С.Л., Задачи имитационного моделирования реляционных баз данных / С.Л. Зиноватная, А.Б. Кунгурцев, А.А. Мунзер // *Труды XI междуна. науч.-практ. конф. «Сучасні інформаційні та електронні технології» (СИЭТ-2010)*, 24–28 мая. 2010 г., Одесса: Т.1. Одеск.нац.політехн.ун-та, 2010. – 118 с.

3. Вентцель Е.С. Теория вероятностей: [учеб. для высш. уч. зав.] / Вентцель Е.С. – 6-е изд. стер. – М.: Высш. шк., 1999. – 576 с.

Получено 21.10.2011

References

1. Blazhko A.A., Levchenko A.J., Prigogine A.A. Model for automated performance optimization-centered database / *Radioelektronni i komp'yuterni system*. – HAI – Kharkov: 2010. – № 7 (48). – P.24-29 [in Russian].

2. Zinovatnaya S.L., Kungurtsev A.B., Munzer A.A. Target simulation and simulation of relational databases of data / *Proceedings of the XI IU. Scientific-Practical. Conf. "Suchasni informatsiyini ta elektronni tehnologii"* (SIET 2010), 24-28 May. 2010, Odessa: V.1. – ONPU, 2010. –118 p. [in Russian].

3. Wentzel E.S. .Probability theory [Text-book. for higher. uch. head.] – 6th ed. ster. – Moscow: Higher.wk., 1999. – 576 p. [in Russian].



Левченко
Александра Юрьевна,
ас. каф. системного программн.
обеспечения Одесск.
нац. политехн. ун-та,
т.р.7348-566, e-mail:
Aleksadra.Levchenko@gmail.com