

УДК 004.054; 004.582

А.А. БЛАЖКО, А.Ю. ЛЕВЧЕНКО, А.С. ПРИГОЖЕВ

Одесский национальный политехнический университет, Украина

МОДЕЛИ ДЛЯ АВТОМАТИЗИРОВАННОЙ ОПТИМИЗАЦИИ ПРОИЗВОДИТЕЛЬНОСТИ СИСТЕМ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ

В работе рассматриваются методы повышения производительности информационных систем, построенных на базе СУБД. Предлагается использовать метод настройки параметров СУБД для сокращения времени ответа и уменьшения ресурсоемкости операций по настройке СУБД. Для выработки управляющих воздействий на СУБД предлагается структурная параметрическая модель. Для описания структуры базы данных и поведения пользователя в системе предлагается поведенческая модель. Для получения значений конфигурационных параметров для СУБД в заданных условиях функционирования предложено выполнять отображение параметрической модели на поведенческую модель. Рассмотрена архитектура и алгоритм построения операционного профиля БД.

Ключевые слова: система управления базой данных, настройка параметров.

Введение

Архитектура современных корпоративных информационных систем (ИС) является сложной и многоуровневой. Она может включать базы данных, обслуживаемые реляционными СУБД различных производителей (рис. 1). Эти СУБД могут различаться по быстрдействию, а также набору и количеству параметров, влияющих на него. Указанная особенность приводит к повышению времени настройки параметров ИС, снижению времени ответа и увеличению ресурсоемкости системы. Поэтому возникает задача повышения эффективности функционирования ИС. Эффективность ИС в данном случае понимается согласно стандарту ISO 9126 [1,2] как комбинация временных характеристик и ресурсоемкости.

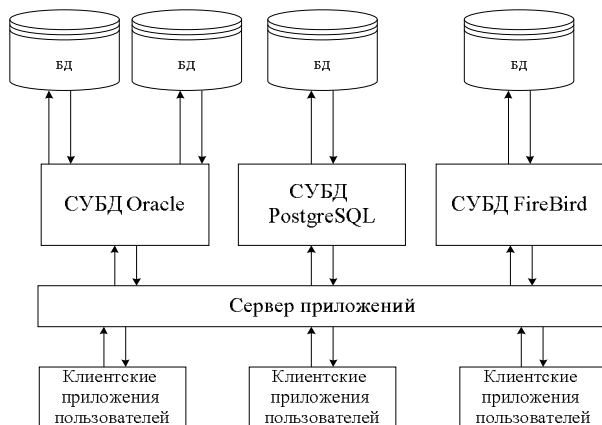


Рис. 1. Схема корпоративной ИС

В рамках временных характеристик функционирования ИС выделим два параметра: время на-

стройки и время ответа. Под ресурсоемкостью будем понимать объем выделяемой оперативной и дисковой памяти, а также процессорного времени.

Задачу повышения эффективности ИС можно рассматривать как задачу оптимизации, поскольку уменьшение времени ответа системы, как правило, вызывает увеличение объема выделяемых ресурсов: кэш-памяти, дисковой и оперативной памяти. И наоборот, сокращение объемов выделяемых ресурсов приводит к увеличению времени ответа.

Рассмотрим далее существующие методы повышения производительности программ ИС.

1. Существующие способы повышения эффективности программ ИС

На данный момент для корпоративных ИС наиболее популярными являются коммерческие СУБД Oracle, IBM DB2, Microsoft SQL Server. Учитывая аппаратные и программные особенности современных реляционных СУБД, можно выделить следующие методы улучшения эффективности функционирования программ ИС:

- изменение формы запросов к БД – логическая оптимизация, рефакторинг SQL-кода;
- включение избыточных структур данных (де-нормализация, индексация, материализованные представления, разделения, кластеризация) – физическая оптимизация [3-7];
- изменение в схеме базы данных, способствующее улучшению проекта базы данных и вместе с тем обеспечивающее сохранение функциональной и информационной семантики базы данных – рефакторинг БД;

– настройка параметров конфигурации СУБД. Логическая и физическая оптимизации требуют вмешательства в исходный код или структуру БД, что требует привлечения специалиста-программиста.

Узкая специализация автоматизированных средств повышения производительности методом настройки параметров. Перечисленные СУБД включают службы, позволяющие автоматически реконфигурировать параметры настройки, отвечающие за распределение областей памяти.

По логической и физической оптимизации существует много работ и исследований, инструментальные средства, реализующие эти методы, значительно развиты. Например, компания Microsoft Research в СУБД MS SQL предлагает в области физической оптимизации БД проект AutoAdmin, который реализует поисковые алгоритмы (например, «what if») для формирования множества интегрированных рекомендаций для индексов, индексированных представлений, материализованных представлений, разделений.

Аналогичным инструментом в СУБД IBM DB2 является DB2 Design Advisor, в СУБД Oracle – SQL Access Advisor. В области логической оптимизации помимо встроенных средств оптимизации и рефакторинга SQL запросов (например, LEarning Optimizer в DB2) существуют приложения сторонних разработчиков. Среди них следует отметить приложение SQL Expert фирмы LECCO, ориентированное на работу с СУБД различных производителей, в их числе Oracle, MS SQL Server, DB2, Sybase.

Автоматизированные средства оптимизации СУБД методом настройки параметров слабо развиты

Но из всех представленных этот метод наиболее быстро реализуем. Для его использования достаточно администратора БД.

Вместе с тем, существует значительное количество параметров настройки СУБД (в среднем более 160 параметров на 1 СУБД), они разнотипны, ими сложно управлять. Таким образом, администратору БД сложно подобрать оптимальные значения параметров конфигурации СУБД, влияющих как на время реакции системы, так и на ресурсоемкость. Существующих экспертных рекомендаций, предлагаемых в литературе и технической документации к СУБД, также недостаточно для подбора оптимальных значений параметров.

Возникает задача повысить производительность программ ИС путем сокращения времени настройки СУБД за счет автоматизации настройки конфигурационных параметров. При этом необходимо учитывать специфику базы данных и внешней среды (ИС) в которой она функционирует. Для решения поставленной задачи необходимо формализовать модели для описания множества конфигурационных параметров СУБД и поведения пользователей ИС, включающей БД, что является целью представляемого исследования.

2. Модели для автоматизации настройки параметров

Для синтеза формальных моделей представим проблему настройки СУБД в рамках ИС в виде усовершенствованной схемы взаимодействия пользователя и информационной системы [8] (рис. 2).

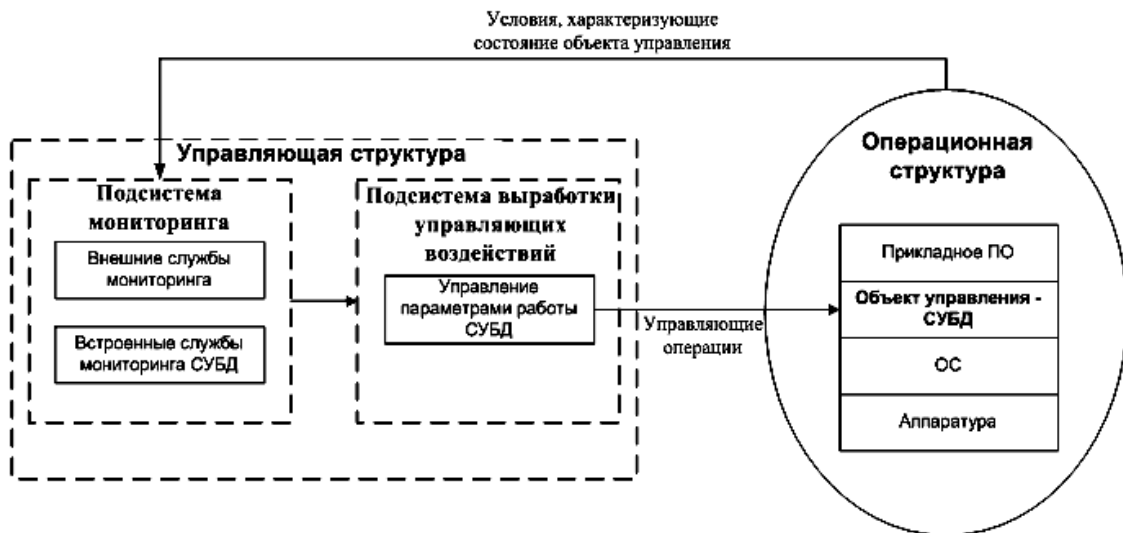


Рис. 2. Интерпретация абстрактной схемы управления Глушкова для задачи настройки параметров

Она состоит из двух структур: управляющей и операционной. Управляющая структура включает подсистему мониторинга и подсистему выработки управляющих воздействий. Подсистема мониторинга состоит из внешних служб мониторинга и встроенных служб мониторинга СУБД.

Встроенные службы мониторинга – это инструменты для сбора статистики, входящие в состав СУБД.

Внешние службы мониторинга – программное обеспечение сторонних производителей, которые предоставляют возможность собирать информацию о БД под управлением СУБД различных производителей.

Операционная структура описывает объект управления (СУБД) и т.н. внешнюю среду в которой функционирует СУБД – аппаратура, операционная система (ОС), прикладное программное обеспечение (ПО), используемое конечным пользователем.

Прикладное ПО посылает множество запросов к СУБД.

Подсистема мониторинга собирает информацию о поведении ИС, в которой функционирует СУБД, и поведении самого объекта управления.

На основе полученных данных, подсистема выработки управляющих воздействий формирует оптимальное или близкое к оптимальному решение и путем изменения параметров настройки воздействует на СУБД.

Для определения пространства выработки управляющих воздействий на СУБД, как объекта управления, проведен анализ и классификация параметров настройки 4-х промышленных СУБД (Oracle 10g, DB2 IBM 9, Sybase ASE 15 и opensource-СУБД PostgreSQL 8.3) и предложена модель параметров СУБД для управляющей структуры [9, 10]:

$$P = \{ \langle db, nm, tp, dl, un, rg, vl, df, mn \rangle \}, \quad (1)$$

где $db \in \{ \text{'Oracle'}, \text{'DB2'}, \text{'Sybase'}, \text{'PostgreSQL'}, \dots \}$ – наименование СУБД;

nm – название параметра в СУБД db типа.

$tp \in \{ c, d, o, l, t \}$ – тип параметра,

где c – управление кэш-областью;

d – управление обменом с диском;

o – оптимизация выполнения запросов;

l – управление резервным копированием и восстановлением;

t – управление транзакциями.

$dl \in \{ 1, 2, 3, 4 \}$ – уровень динамичности изменения значений параметра,

где l – изменение параметра вступает в силу внутри сессии;

2 – после перезапуска сессии;

3 – после перезапуска сервера;

4 – после повторной установки БД.

un – единица измерения параметра;

rg – тип представления значений параметра (в виде диапазона значений или в виде множества допустимых значений);

vl – текущее значение параметра (диапазон значений);

df – значение по умолчанию;

mn – способ управления параметром (с помощью SQL-операторов или через редактирование файлов конфигурации).

Структурная параметрическая модель (1) позволит унифицировать механизм управления СУБД, который не ограничен представлениями об этом процессе конкретным производителем СУБД. Параметрическая модель позволяет сформировать базу знаний для параметров настройки разных СУБД. В рамках модели проведена классификация параметров настройки СУБД по признакам типа параметра и динамичности его изменения. Для классов c, d, o, l, t проведен статистический анализ мощности этих классов [11].

Для описания операционной структуры, включающей объект управления, необходимо описать структуру и связи БД, поведение пользователя. Для представления структуры БД воспользуемся определением и фундаментальными понятиями реляционной алгебры [12]:

$$S = \langle DB, \{ \text{СИГН} \} \rangle, \quad (2)$$

где $DB = \langle R, L \rangle$ – структура базы данных, включающая подмножество отношений, атрибутов и связей между отношениями;

СИГН – традиционные операции над множествами (объединение, пересечение, разность, декартово произведение) и специальные реляционные операции (выборка, проекция, соединение, деление).

Для описания физической структуры БД представим отношение (таблицу) в виде множества двоек вида:

$$R = \langle \langle a_1 : d_1 \rangle, \langle a_2 : d_2 \rangle, \dots, \langle a_n : d_n \rangle \rangle,$$

где a_n – атрибут отношения, для описания столбцов таблиц;

d_n – домен, определяющий тип данных заданного атрибута.

Для описания связей между отношениями воспользуемся следующей структурой:

$$L = \left\{ \begin{array}{l} \langle (R_1.a_1 \langle R_1.a_2, \dots, R_1.a_n \rangle) \rangle, \\ \langle (R_2.a_3 \langle R_2.a_4, \dots, R_2.a_n \rangle) \rangle \end{array} \right\},$$

Связь может быть установлена между одним или несколькими атрибутами одного отношения, и одним или несколькими атрибутами другого отношения.

Для описания поведения пользователя усовершенствуем модель (2) путем ввода множества

V – операционный профиль, который специфицирует использование СУБД в терминах операций в системе и вероятностей их появления [13]:

$$S = \langle DB, Q, V, \{СИГН\} \rangle, \quad (3)$$

где Q – множество запросов, отсылаемых СУБД прикладным ПО;

V – операционный профиль, представляет собой множество пар:

$$V = \left\{ (q_i, p_{q_i}), i = 1, \dots, n \right\},$$

где n – количество SQL-запросов;

q_i – i -ый запрос;

p_{q_i} – вероятность появления запроса q_i в процессе взаимодействия пользователя и БД.

СИГН – сигнатура, состоящая из объединения операций реляционной алгебры и операций вычисления вероятностей запроса к БД. Операция ВЕРОЯТНОСТЬ (PR) определяется как выполнение отображения:

$$PR : Q \rightarrow \{0;1\},$$

Вероятность для каждого запроса рассчитывается на основе классического определения вероятности, в нашем случае

$$P(q_i) = \frac{n(q_i)}{M(Q)}, \quad (4)$$

где $M(Q)$ – мощность множества;

$n(q_i)$ – количество запросов i -го типа.

Адаптация параметрической модели для каждой конкретной системы выполняется отображением модели информационной системы на параметрическую модель.

$$F = S \rightarrow P',$$

где $P' \subseteq P$ – описывает параметры настройки СУБД, используемых пользователем ИС.

Под отображением понимаются алгоритмы экспертного определения значений компонента v_i для каждого элемента из P' на основе анализа поведенческой модели. Это позволяет осуществлять процедуру настройки параметров СУБД в заданных условиях функционирования ИС.

3. Построение операционного профиля для поведенческой модели

Одним из входных данных системы автоматизации настройки параметров является время отклика базы данных на конкретный запрос при конкретных значениях параметров. Для получения значений времени отклика на запрос предлагается реализовать систему, основанную на использовании программных агентов. Это позволит осуществлять процесс тестирования времени отклика СУБД с использованием реально существующей архитектуры ИС.

Рассмотрим архитектуру и принципы функционирования такой системы.

Как указывалось ранее, современные ИС имеют сложную неоднородную распределённую архитектуру. Анализируя структуру корпоративной ИС (рис. 1), можно сделать вывод, что время ответа такой ИС складывается из двух составляющих: времени ответа сервера приложений и времени ответа СУБД, настроенной с использованием конкретного множества значений параметров.

Процесс определения времени отклика на запрос состоит из двух этапов: расчёта операционного профиля для каждого клиента ИС и отправки запросов в соответствии с данным операционным профилем. Такой подход значительно эффективнее, чем использование известных синтетических тестов для СУБД, поскольку позволяет использовать для определения времени ответа существующую структуру БД.

Для расчёта операционного профиля каждого клиента ИС используется файл журнала базы данных, который содержит историю запросов в базу со стороны каждого клиента ИС. Поскольку данный файл содержит сведения о клиенте ИС и запрос клиента к СУБД, становится возможным рассчитать операционный профиль для каждого клиента ИС.

Далее, на каждом клиенте ИС устанавливается клиентский агент, который отправляет запросы на сервер в соответствии с рассчитанным операционным профилем. При этом в клиентском агенте указывается, какое количество запросов отправляется к базе данных, а также закон распределения времени отправки запросов. На основании количества запросов рассчитывается точное количество запросов каждого вида по формуле (4). Количество запросов рассчитывается на основе лог файла БД.

Рассчитанное количество запросов n вместе с самими запросами хранится в библиотеке запросов клиента.

При отправке запросов клиент случайным образом выбирает запрос из библиотеки и уменьшает его счётчик на единицу. Если счётчик равен 0, клиент повторно выбирает запрос. Если выбраны все запросы, клиент посылает серверу команду о завершении работы.

Отправленные запросы принимает агент сервера, который фиксирует время их прихода и перенаправляет запросы непосредственно к СУБД, после чего ожидает от неё ответа. Время, за которое был получен ответ, также фиксируется, для чего в серверную часть встраивается программный код счётчика.

По результатам запросов строится таблица значений, содержащая текущие значения параметров,

запросы и время ответа БД на них, которая в дальнейшем будет использована в рамках экспертной настройки параметров.

Выводы

В ходе представленного исследования были проанализированы существующие средства оптимизации производительности СУБД. Рассмотренные средства оптимизации обеспечивают решение задач оптимизации производительности СУБД в области физической и логической оптимизации, а также рефакторинга SQL кода. Однако, использование этих средств связано с большими временными затратами.

Как альтернативный способ настройки СУБД был выбран метод настройки параметров. Данный метод позволяет значительно сократить время, необходимое на оптимизацию, требует только наличия администратора и может быть автоматизирован.

Получила дальнейшее развитие модель взаимодействия пользователя и автоматизированной информационной системы для задачи автоматизированной настройки параметров СУБД. Предложено рассматривать управляющую структуру как композицию двух подсистем: мониторинга и выработки управляющих воздействий. Предложенная модель позволяет формализовать задачу настройки параметров как задачу управления.

Проанализированы существенные характеристики параметров для современных СУБД и на их основе впервые предложена параметрическая модель СУБД, позволяющая формализовать в дальнейшем процесс настройки параметров.

Для описания условий, характеризующих состояние объекта управления (СУБД) и ИС, в которой он функционирует, в статье предложена поведенческая модель пользователя ИС. Она является усовершенствованием реляционной алгебры путем введения дополнительной основы – операционного профиля БД. Рассмотрены алгоритмы и архитектура системы для построения этого профиля.

Впервые предложено формальное описание процесса настройки параметров в виде отображения поведенческой модели на модель параметров СУБД конкретной ИС. Это позволит сократить время проведения эксперимента для заданной ИС путем минимизации количества настраиваемых параметров и уменьшения их диапазонов значений.

Построенные модели позволяют в дальнейшем разработать формальные методы анализа и оптимальной настройки параметров.

Литература

1. ISO/IEC 9126-2:2003. *Software Engineering – Product Quality – Part 2: External Metrics.*
2. ISO/IEC 9126-2:2003. *Software Engineering – Product Quality – Part 3: Internal Metrics.*
3. Chaudhuri S. *Self-Tuning Database Systems: A Decade of Progress / S. Chaudhuri, V.R. Narasayya // The Procs. of the 33rd International Conference on VLDB. – Vienna, Austria, 2007. – P. 3-14.*
4. Zilio et al. *DB2 Design Advisor: Integrated Automatic Physical Database Design // The Procs. of 30th International Conference on VLDB. – Toronto, Canada. – 2004. – P. 1087-1097.*
5. Кузнецов С.Д. *Методы оптимизации выполнения запросов в реляционных СУБД / С.Д. Кузнецов // Тем. изд. Итоги науки и техники. Вычислительные науки. – Т.1. – С. 76-153.*
6. *Query Optimization in Oracle 9i [Электрон. ресурс] / An Oracle White Paper. – Режим доступа к ресурсу: http://technet.oracle.com/products/bi/pdf/o9i_optimization_twp.pdf. – February, 2002.*
7. Chaudhuri S. *Automating Statistics Management for Query Optimizers / S. Chaudhuri, V. Narasayya // Proceedings of 16th International Conference on Data Engineering. – San Diego, USA, 2000.*
8. Пригожев. А.С. *Система поддержки пользователя на основе экспертной системы / А.С. Пригожев, Е.И. Алёшкина, М.В. Бабичева, А.В. Вовк, В.А. Писаренко // Проблемы программирования. – 2008. – № 2-3. – С. 605-612.*
9. Балл Г.А. *О применении понятий «процедура», «алгоритм», и «квазиалгоритм» для описания функционирования активных систем / Г.А. Балл // Теория задач и способов их решения. – К.: Ин-т кибернетики АН УССР, 1974. – С. 69-81.*
10. Балл Г.А. *О системе основных понятий теории задач / Г.А. Балл // Теория задач и способов их решения. – К.: Ин-т кибернетики АН УССР, 1974. – С. 57-69.*
11. Блажко А.А. *Унифицированная параметрическая модель настройки системы управления базами данных / А.А. Блажко, А.Ю. Левченко // Современные информационные и электронные технологии: тезисы докл. междунар. науч.-техн. конф. 18-21 мая, 2009 г. – Одесса, 2009. – Том 1. – С. 68.*
12. Дейт К. *Введение в системы баз данных пер. с англ., 7-е изд. / К. Дейт. – М: Вильямс, 2001. – 1072 с.*
13. Коваль Г.И. *Концепция профилей в инженерии надежности программных систем / Г.И. Коваль, Г.Б. Мороз, Т.М. Коротун // Математичні машини і системи. – 2004. – № 1. – С. 166-184.*

Поступила в редакцию 02.03.2010

Рецензент: д-р техн. наук, проф. С.А. Положаенко, Одеський національний політехнічний університет, Інститут комп'ютерних систем, Україна.

МОДЕЛІ ДЛЯ АВТОМАТИЗОВАНОЇ ОПТИМІЗАЦІЇ ПРОДУКТИВНОСТІ СИСТЕМ КЕРУВАННЯ БАЗАМИ ДАНИХ

О.А. Блажко, О.Ю. Левченко, О.С. Пригожев

У роботі розглядаються методи підвищення продуктивності інформаційних систем, побудованих на базі СКБД. Пропонується використовувати метод налаштування параметрів СКБД для скорочення часу відповіді й зменшення ресурсоемності операцій по налаштуванню СКБД. Для виробітку керуючих впливів на СКБД пропонується структурна параметрична модель. Для опису структури бази даних і поведінки користувача в системі пропонується поведінкова модель. Для одержання значень конфігураційних параметрів для СКБД у заданих умовах функціонування запропоновано виконувати відображення параметричної моделі на поведінкову модель. Розглянута архітектура й алгоритми побудови операційного профілю БД.

Ключові слова: Тестування баз даних, налагодження параметрів.

MODELS FOR AUTOMATED OPTIMIZATION DBMS PERFORMANCE

O.A. Blazhko, O.Yu. Levchenko, O.S. Prigozhev

We consider the methods to enhance the DBMS based information systems performance. It is proposed to use the database parameters setting method to reduce response time and to minimize the resource-intensive of the database configuring operations. The structural parametric model proposed to develop control actions on the database. The behavioral model is proposed to describe the database structure and user behavior. It is proposed to map the parametric model on behavioral model to obtain the database configuration parameters values in specified operating conditions. The architecture and algorithm for database operational profile constructing is considered.

Keywords: DBMS testing, parameters configuration.

Блажко Александр Анатольевич – канд. техн. наук, доц. кафедри системного програмного забезпечення, Одеський національний політехнічний університет, Одеса, Україна, e-mail: blazhko@ieee.org.

Левченко Александра Юрьевна – асистент кафедри системного програмного забезпечення, Одеський національний політехнічний університет, Одеса, Україна, e-mail: aleksandra.levchenko@gmail.com.

Пригожев Александр Сергеевич – канд. техн. наук, ст. преп. кафедри системного програмного забезпечення, Одеський національний політехнічний університет, Одеса, Україна, e-mail: prigozhev@rambler.ru.