

УДК 681.3.016:681.324

Ю.Н. Возовиков, аспирант**А.А. Мунзер**, магистр

Одесский национальный политехнический университет, г. Одесса, Украина

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ МАТЕРИАЛИЗОВАННЫХ ПРЕДСТАВЛЕНИЙ

Предлагается метод анализа механизма реализации и эффективности применения материализованных представлений, основанный на использовании имитационной модели реляционных баз данных.

Вступление

Использование материализованных представлений (МП), то есть – сохраненных результатов запросов к реляционной базе данных (РБД) – эффективное средство повышения производительности информационной системы (ИС) [1]. Однако целесообразность применения такого решения определяется рядом факторов, которые аналитически трудно оценить. Кроме этого, сам механизм реализации МП достаточно сложен и требует экспериментальной проверки. Указанные доводы определяют целесообразность применения имитационного моделирования как для определения работоспособности и эффективности средств работы с МП, так и для целесообразности применения МП для конкретной ИС.

Формулирование целей статьи

Целью данной работы является создание имитационной модели для исследования различных механизмов реализации материализованных представлений, создаваемых для конкретной ИС.

Изложение основного материала статьи

В работе [2] изложены общие положения моделирования РБД. В нашем случае также необходимо сохранить исходное состояние базы данных (БД) или её фрагментов на момент tI начала испытаний модели, сохранить последовательность запросов с момента tI , а также позаботиться о сокращении количества таблиц, используемых в модели.

Для этого воспользуемся аппаратом множеств [3] и введем следующее представление множества запросов Q , в котором каждый запрос представляется кортежем:

$$q = \langle T_q, T_y, \langle t_{st}, \tau \rangle \rangle, \quad (1)$$

где $T_q = \{T_i\}, i = \overline{1, N_{Tq}}$ – множество таблиц БД, участвующих в запросе q ; N_{Tq} – количество таких таблиц.

Каждая таблица $T_i = \langle Nm_i, F_i \rangle$ представлена именем Nm и множеством полей $F_i = \{F_{ij}\}, j = \overline{1, N_{Fi}}$, где N_{Fi} – количество полей таблицы T_i .

T_y – тип запроса, $T_y \in \{TD, TI, TU, TS\}$, где TD – запрос на удаление строки (строк) из таблицы, TI – запрос на добавление данных; TU – запрос на обновление; TS – запрос на выборку;

t_{st} – время запуска запроса;

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

Введём представление результатов выполнения запросов в виде множества:

$$R = \{r_i\}, i = \overline{1, Ns},$$

где Ns – количество запросов типа SELECT.

Каждый элемент множества $r_i = \langle T_i, t_i \rangle$, где T_i – таблица результата, а t_i – момент времени получения результата.

I. Проверка работоспособности механизмов фильтрации запросов, выборки и обновления МП

Последовательность шагов:

1) Устанавливается длительность периода наблюдения t ИС. Для каждой таблицы БД сохраняется её состояние. В течение периода t (механизмы МП отключены) фиксируются все запросы к БД в соответствии с (1). Будем считать, что эти запросы составляют множество Q . Также фиксируются все результаты запросов типа SELECT (упорядоченное по времени множество R).

Предполагается, что к моменту начала испытания модели установлено множество всех запросов $Qmps$, для которых создаются МП. Поскольку за время наблюдения t не обязательно все запросы из $Qmps$ будут использованы, формируем множество запросов Qmp , имеющих МП и попадающих в нашу выборку:

$$Qmp = \{q_i \mid q_i \in Q \wedge q_i \in Qmps\}, i = \overline{1, |Q|} \quad (2)$$

2) Для сокращения объёма модели и времени её испытания из Qmp выбирается незначительное количество запросов m ($m \geq 2$, поскольку необходимо проверить механизм фильтрации запросов).

Минимизации количества таблиц, используемых в модели, можно добиться применением функции $fmint(Qmp, m)$, которая из множества запросов Qmp выбирает такие m запросов, для которых общее число используемых таблиц будет минимальным. В результате формируется множество запросов $Qmp1$.

$$Qmp1 = \{q_i \mid q_i \in Qmp, \sum_{i=1}^m N_{Tq_i} = \min\}, \quad (3)$$

Для проверки обработки запроса, который не принадлежит множеству Qmp , но имеет тип SELECT, отыскиваем такой запрос также с помощью функции $fmint$

$$fmint((Qmp1 \cup q_i \mid q_i \in Q \wedge q_i \notin Qmp \wedge Ty_{q_i} = Ts), 1), i = \overline{1, |Q|}, \quad (4)$$

Определяем множество таблиц $T1$, используемое запросами из $Qmp1$

$$T1 = \{T_i \mid q_i \in Qmp1\}, i = \overline{1, m+1}, \quad (5)$$

1) Для проверки механизма обновления МП также необходимо выбрать все запросы, модифицирующие данные, используемые в МП. Для этого составляем множество запросов Qmd :

$$Qmd = \{q_i \mid q_i \in Q \wedge Ty_{q_i} \neq Ts \wedge F_{q_i} = F_{q_j} \mid q_j \in Qmp1\}, i = \overline{1, |Q|}, j = \overline{1, m+1} \quad (6)$$

Поскольку в условиях выборки запросов на модификацию могут присутствовать поля, которые не принадлежат таблицам из $T1$, окончательное множество таблиц, используемое в модели Tm , должно учитывать все таблицы, используемые в Qmd :

$$Tm = \{T_i \mid q_i \in Qmd\} \cup T1, i = \overline{1, |Qmd|}, \quad (7)$$

2) На основе множества Tm создаётся модель БД1 реальной БД. Все таблицы Tm заполняются данными, сохранёнными к моменту испытания модели.

3) Производится упорядочение множества Qmd по параметру t_{st} . Механизм МП не включается. Запросы из Qmd последовательно подаются в модель БД1. Результаты выполнения запросов сохраняются в множестве $R1$.

4) Восстанавливается исходное состояние модели БД1. Подключается механизм МП.

Запросы из Qmd повторно подаются в модель БД1. Результаты выполнения запросов сохраняются в множестве $R2$.

5) Сравниваются результаты двух испытаний модели БД1. Если выполняются равенства $T_{R1i} = T_{R2i}$ для всех $T_{R1i} \in R1$ и $T_{R2i} \in R2$ при $i = \overline{1, m+1}$, то можно сделать заключение о корректности работы механизмов МП и завершить испытания. В противном случае следует перейти к шагу 6.

б) Прежде всего следует проверить совпадение результатов выполнения запросов типа SELECT в реальной базе данных (БД) и модели (БД1) без включения механизмов МП.

Если $\forall r_i \in R1 \exists r_j \in R \mid T_{ri} = T_{rj}$ при $i=1, m+1$ и $j=1, |R|$, то можно считать, что модель БД1 построена правильно и ошибки вызваны применением механизма обслуживания МП.

Для дальнейшей локализации ошибок рекомендуется проверить работу механизма фильтрации запросов путем сравнения мощностей множеств R1 и R2. Если фильтрация выполняется правильно, то следует детально проверить механизм обновления путём демонстрации каждого элемента $r_i \in R1$ после очередного обновления данных в его таблице.

II. Анализ эффективности применения МП

Последовательность шагов:

1) Устанавливается длительность периода наблюдения t ИС. Для каждой таблицы БД сохраняется её состояние в модели БД1. В течение периода t (механизмы МП отключены) фиксируются все запросы к БД в соответствии с (1). Будем считать, что эти запросы составляют множество Q . С

Предполагается, что к моменту начала испытания модели установлено множество всех запросов Qmp , для которых создаются МП.

2) Определяется суммарное время выполнения всех запросов q_i из множества Q :

$$S0 = \sum_{i=1}^{|Q|} \tau_{q_i} \quad (8)$$

Формируется множество запросов Qmp , имеющих МП и попадающих в нашу выборку (множество Q):

$$Qmp = \{q_i \mid q_i \in Q \wedge q_i \in \overline{Qmp}\}, i = \overline{1, |Q|} \quad (9)$$

Определяется суммарное время выполнения для каждого запроса, входящего в множество Qmp :

$$s_j = \sum_{i=1}^{|Q|} (\tau_{q_i} \mid q_i \in Qmp \wedge q_i \in Q) \quad (10)$$

Определяется суммарное время выполнения всех запросов из множества Qmp :

$$S1 = \sum_{j=1}^{|Qmp|} s_j, \quad (11)$$

3) Восстанавливается исходное состояние модели БД1. Включается механизм МП. На вход модели подаётся последовательность запросов из Q . Определяется суммарное время выполнения всех запросов q_i из множества Q :

$$S0_m = \sum_{i=1}^{|Q|} \tau_{q_i} \quad (12)$$

Определяется суммарное время выполнения для каждого запроса, входящего в множество Qmp :

$$s_{mj} = \sum_{i=1}^{|Q|} (\tau_{q_i} \mid q_i \in Qmp \wedge q_i \in Q) \quad (13)$$

Определяется суммарное время выполнения всех запросов из множества Qmp .

$$S1m = \sum_{j=1}^{|Qmp|} s_j \quad (14)$$

4) Производится оценка полученных результатов по следующим формулам. Для оценки общей эффективности применения МП в ИС:

$$\lambda = 1 - \frac{S0 - Sm}{S0} \quad (15)$$

Для оценки эффективности применения МП для запросов из множества Qmp :

$$\lambda_{mp} = 1 - \frac{S1 - S1_m}{S1} \quad (16)$$

Для оценки эффективности применения каждого МП в отдельности:

$$\lambda_{mpj} = 1 - \frac{s_j - s_{mj}}{s_j} \quad (17)$$

Последняя оценка позволяет корректировать состав элементов множества Qmp .

III. Анализ эффективности механизма периодического включения МП [4]

Последовательность шагов:

1) Выполнение этого шага аналогично выполнению шагов 1 и 2 предыдущего анализа. В результате получаем суммарное время выполнения для каждого запроса, входящего в множество Qmp :

$$s_{mj} = \sum_{i=1}^{|Q|} (\tau_{qi} \mid q_i \in Qmp \wedge q_i \in Q), \quad (18)$$

$$Sm = \{s_{mj}\}, j = \overline{1, |Qmp|} \quad (19)$$

2) Восстанавливается исходное состояние модели БД1. Включается механизм МП и устанавливается период включения/выключения МП для каждого запроса $q_j \in Qmp$.

Управление запросами осуществляется множеством функций:

$$Fc = \{fc_j\}, j = \overline{1, |Qmp|} \quad (20)$$

Каждая функция использует в качестве аргументов запрос q_j и заданный определённым образом период рабочего состояния МП. Конкретные моменты времени включения/выключения рассчитываются функцией на основании текущей даты календаря. Более подробное описание функции изложено в [4].

3) Аналогично пункту 3 предыдущего исследования на вход модели подаётся последовательность запросов из Q . Определяется суммарное время выполнения всех запросов q_i из множества Q .

$$SO_{mc} = \sum_{i=1}^{|Q|} \tau_{qi} \quad (21)$$

Определяется суммарное время выполнения для каждого запроса, входящего в множество Qmp :

$$s_{mcj} = \sum_{i=1}^{|Q|} (\tau_{qi} \mid q_i \in Qmp \wedge q_i \in Q) \quad (22)$$

Определяется суммарное время выполнения всех запросов из множества:

$$S1mc = \sum_{j=1}^{|Qmp|} s_{mcj} \quad (23)$$

4) Выполняется анализ эффективности механизма управления запросами. Величина

$$\gamma_s = 1 - \frac{SO_m - SO_{mc}}{SO_m} \quad (24)$$

показывает относительное сокращение времени выполнения всех запросов $q_i \in Q$ за счет использования механизма управления.

Величина

$$\gamma_{mps} = 1 - \frac{S1m - S1mc}{S1m} \quad (25)$$

показывает насколько механизм управления влияет на эффективность выполнения запросов $q_i \in Q_{mp}$.

5) Выполняется анализ эффективности управления каждым запросом $q_i \in Q_{mp}$ по отдельности. Для этого рассчитываются значения:

$$\gamma_{mpj} = 1 - \frac{s_{mj} - s_{mcj}}{s_{mj}} \quad (26)$$

6) На основании анализа, проведенного в пунктах 4 и 5 может быть принято решение о перенастройке функций из множества F_c . После этого процесс исследования повторяется, начиная с пункта 3 до фиксации оптимальных аргументов функций.

Если же принимается решение о добавлении или удалении элементов в множество Q_{mp} , то процесс моделирования повторяется с начала.

7) Результаты, полученные в пунктах 3 и 4, могут сохраняться на длительный период для оценки влияния изменений в ИС на эффективность используемого механизма МП. Что даёт основание для повторного исследования системы и новой настройки механизмов МП.

Выводы

Предложенный метод построения имитационной модели РБД для исследования применимости МП, позволяет выполнить комплексный и детальный анализ как различных механизмов реализации МП, так и эффективности выбора запросов – кандидатов для МП и управления их включением/выключением. Кроме этого, реализуется возможность оценки влияния текущих изменений в ИС на эффективность применения МП и принятия своевременных решений по перенастройке механизма МП.

Список использованных источников

1. Кунгурцев А. Б. Анализ возможности применения материализованных представлений в информационных системах / Кунгурцев А. Б., Куок Винь Нгуен Чан // Труды ОНПУ. – Одесса, 2003. – С. 102-106.
2. Кунгурцев А.Б. Задачи имитационного моделирования реляционных баз данных / Кунгурцев А.Б., Зиноватная С.Л., Мунзер аль Абдо // Труды десятой международной научно-практической конференции СИЭТ-2009. – Одесса, 2010. – С.210.
3. Певзнер Л.Д. Математические основы теории систем / Певзнер Л.Д., Чураков Е.П. – М.: Высш. шк., 2009. – 503с.
4. Кунгурцев А.Б. Управление материализованными представлениями в информационных системах / Кунгурцев А.Б., Возовиков Ю.Н // Восточно-европейский журнал передовых технологий. – 2010. – № 1/ 4(43). – С. 18-21.
5. Кунгурцев А.Б. Поиск закономерностей в распределении запросов для управления материализованными представлениями / Кунгурцев А.Б., Возовиков Ю.Н. // Труды Одесского политехн. ун-та. – Одесса, 2008. – № 2(30). – С. 135-140.