

УДК 681.3.016:681.324

А.Б. Кунгурцев, канд. техн. наук, Ю.Н. Возовиков**ПОДДЕРЖКА ЭФФЕКТИВНОСТИ МЕХАНИЗМА УПРАВЛЕНИЯ
МАТЕРИАЛИЗОВАННЫМИ ПРЕДСТАВЛЕНИЯМИ**

Использование материализованных представлений (хранимых результатов запросов к базам данных) предусматривает предварительный анализ информационных систем. Однако с течением времени интенсивность и характер решаемых задач в любой системе изменяется и, следовательно, эффективность применения материализованных представлений снижается. Рассматривается метод поддержания эффективности использования материализованных представлений на основе непрерывного анализа потока запросов к системе.

Ключевые слова: информационные системы, материализованные представления, запрос.

A.B. Kungurtsev, Ph.D., Y.N. Vozovikov**SUPPORTING THE EFFECTIVENESS OF THE CONTROL
MATERIALIZED REPRESENTATIONS**

Using materialized representation (stored results of database queries) provide a preliminary analysis of information systems. However, over time the intensity and nature of tasks in any system changes and, consequently, the effectiveness of the materialized representation is reduced. In this paper we consider a method of maintaining the effectiveness of using materialized representation based on continuous flow analysis requests to the system.

Keywords: Information system, materialized representations, queries.

О.Б. Кунгурцев, канд. техн. наук, Ю.М. Возовиков**ПІДТРИМКА ЕФЕКТИВНОСТІ МЕХАНІЗМУ УПРАВЛІННЯ
МАТЕРІАЛІЗОВАНИМИ УЯВЛЕННЯМИ**

Використання матеріалізованих уявлень (збережених результатів запитів до баз даних) передбачає попередній аналіз інформаційних систем. Однак з часом інтенсивність і характер вирішуваних завдань у будь-якій системі змінюється і, отже, ефективність застосування матеріалізованих уявлень знижується. Розглядається метод підтримки ефективності використання матеріалізованих уявлень на основі безперервно-го аналізу потоку запитів до системи.

Ключові слова: Інформаційні системи, матеріалізовані уявлення, запит.

Введение

Использование материализованных представлений (МП), представляющих собой сохраненные результаты запросов – хорошее средство повышения производительности информационных систем (ИС), использующих реляционные базы данных. Повышения эффективности МП можно добиться, если учитывать периодичность появления различных запросов в ИС, выявленную на этапе её исследования [1,2].

Однако с течением времени в ИС происходят изменения. Некоторые из них затрагивают структуру системы, например, добавление/удаление таблиц в БД. Другие – могут быть связаны с появлением новых задач или изменением интенсивности решения существующих задач. Таким образом, с течением времени предварительно рассчитанная эффективность применения МП может уменьшиться. Снижение эффективности может

быть вызвано, например, изменением периодичности появления определенных запросов, изменением количества запросов, для которых не созданы МП, может отражать появление новых задач в ИС, что в свою очередь влияет на эффективность использования существующих МП.

Постановка задачи

Для поддержания эффективности применения МП на высоком уровне предлагается проводить корректировку ранее принятых решений на основании продолжения наблюдения за системой. Это также позволит определить периодичность в распределении запросов, которую ранее не представлялось возможности обнаружить. Например, перед внедрением механизма МП в некоторую ИС проводились наблюдения за её работой в течение двух месяцев. Однако набор и интенсивность задач, решаемых в ИС, зависит от сезона года. Очевидно, что такой период не мог быть определён за период наблюдения за

© Кунгурцев А.Б., Возовиков Ю.Н., 2011

работой системы на стадии проектирования МП.

На основании сказанного корректировка механизма управлением МП заключается в решении трёх задач:

- 1) получение текущего значения эффективности каждого из используемых МП;
- 2) сравнительной оценки количеств запросов, для которых созданы и не созданы МП;
- 3) накопления и анализа данных для возможности создания МП с большими периодами работы.

Расчетные значения за предыдущий период наблюдения

Пусть механизм МП был включен после предварительного наблюдения за системой начиная с момента времени t_{mo} в течение времени t_0 .

Будем считать, что в результате наблюдения за системой в течение времени t_0 получено множество всех запросов Q_o , множество разных запросов Q_r , множества запросов Q_i , для которых были созданы МП.

Каждый запрос представлен кортежем

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

где $T_q = \{T_i\}, i = \overline{1, N_{Tq}}$ – множество таблиц БД, участвующих в запросе q ; N_{Tq} – количество таких таблиц; T_y – тип запроса; t_{st} – время запуска

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

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, используемых в запросе.

Тип запроса

$$T_y \in \{TD, TI, TU, TS\},$$

где TD – запрос на удаление записи (записей) из таблицы; TI – запрос на добавление данных;

TU – запрос на обновление; TS – запрос на выборку.

Определим количества запросов в множествах: $n_o = |Q_o|$, $n_r = |Q_r|$ и $n_i = |Q_i|$.

В результате предварительного наблюдения за системой было определено суммарное время выполнения всех запросов

$$S_0 = \sum_{j=1}^{|n_o|} \tau_{qj} \quad (2)$$

и суммарное время выполнения всех запросов вида q_i с применением МП.

$$S_{mpi} = n \tau_{mpi} + \tau_s + \tau_{up}, \quad (3)$$

где n – количество запросов вида q_i , использующих MP_i ; τ_{mpi} – время выборки данных из таблицы MP_i ; τ_s – усредненное время выборки запроса, имеющего МП, из потока всех запросов (при этом следует учитывать в дальнейшем, что τ_s добавляется ко времени выполнения всех запросов и, таким образом, ухудшает общий показатель эффективности применения МП); $\tau_{up} = k_{up} \tau_{0up}$ – время обновления таблицы MP_i ; k_{up} – количество запросов q_u , которые изменяют данные в MP_i .

Оценка текущей эффективности МП, созданных ранее

Для анализа работы МП предлагается метод скользящей оценки эффективности МП. Суть метода заключается в непрерывном отслеживании работы системы и периодическом расчете эффективности МП в заранее определённые моменты.

Прежде всего, следует определить минимальный интервал времени Δt , по истечении которого имеет смысл проводить анализ данных. Выбор Δt определяется следующими соображениями:

величина Δt не может быть слишком короткой, поскольку в этом случае изменения, происшедшие в системе, мало отразятся на оценке эффективности МП;

величина Δt не может быть слишком большой, так как радикальные изменения в ИС, происшедшие за этот период, могут не только понизить эффективность некоторых МП, но и сделать их применение совершенно бесполезным.

Исходя из приведенных соображений, максимальным значением может быть $\Delta t = t_0$, а минимальным – самый длинный период t_{max} некоторого реализованного МП. Например, t_0 – два месяца, а t_{max} – неделя, тогда $\Delta t = 7$ суток.

Оценка эффективности механизма МП может выполняться разными способами в зависимости от времени эксплуатации:

- 1) время эксплуатации не превышает времени наблюдения за ИС t_0 ;

2) время эксплуатации превышает время наблюдения за ИС t_0 .

В первом случае могут сравниваться расчетные эффективности с реальными эффективностями для некоторых МП. Во втором – имеется возможность проанализировать изменения реальных эффективностей.

Рассмотрим первый случай. Пусть для некоторого вида запроса q_i , для которого было создано MP_i , на этапе наблюдения был установлен период включения t_i .

Пусть $Q\delta$ – множество всех запросов, поступивших в ИС в течение времени Δt .

Определяем множество запросов $Q\delta_i$ вида q_i , выполненных в течение времени Δt ($q_i \in Q\delta$) во время периодов t_i .

Суммарное время выполнения этих запросов

$$S\delta_i = \sum_{j=1}^{|Q\delta_i|} \rho_{ij} . \quad (4)$$

Теперь можно сопоставить среднее время выполнения запроса q_i , полученное расчётным путём на этапе настройки $S_{si} = \frac{S_{mpi}}{n}$,

со средним временем выполнения этих же запросов в течение времени Δt $S_{s\delta i} = \frac{S\delta_i}{|Q\delta_i|}$.

Если значения $S_{s\delta i}$ и S_{si} существенно не отличаются, то следует перейти к сопоставлению относительных оценок.

Относительная эффективность применения MP_i для ИС, полученная на этапе наблюдения

$$\varepsilon_{oi} = 1 - \frac{S_0 - S_{mpi}}{S_0} , \quad (5)$$

где $S_0 = \sum_{j=1}^{|Q|} \tau_{qj}$ – суммарное время выполнения всех запросов q_i из множества Q .

Относительная эффективность применения MP_i , полученная в течение времени Δt ,

$$\varepsilon_{\delta i} = 1 - \frac{S\delta - S\delta_i}{S\delta} , \quad (6)$$

где $S\delta = \sum_{j=1}^{|Q\delta|} \tau_{qj}$ – суммарное время выполнения всех запросов q_i , поступивших в ИС за время Δt .

Если значения ε_{oi} и $\varepsilon_{\delta i}$ существенно не отличаются, то следует перейти к анализу других MP_i .

Рассмотрим случай, когда значения $S_{s\delta i}$ и S_{si} существенно отличаются. Если считать, что τ_{mpi} – время выборки данных из таблицы MP_i , а также τ_s – усредненное время выборки запроса в течение длительного периода сохраняют свои значения (3), то существенное расхождение в значениях $S_{s\delta i}$ и S_{si} может определяться изменением соотношения количества запросов q_i и запросов q_{ui} , обновляющих MP_i . Если оказалось, что $S_{s\delta i} \ll S_{si}$, то эффективность применения MP_i , возможно, увеличилась. В противном случае она уменьшилась. Во всех случаях можно рекомендовать перейти к расчёту относительных оценок эффективности формулы (5) и (6).

Рассмотрим случай, когда относительная эффективность MP_i существенно снизилась ($\varepsilon_{oi} \gg \varepsilon_{\delta i}$). Это может произойти в силу следующих причин:

- 1) уменьшилось относительное количество запросов q_i ;
- 2) изменилась периодичность появления запросов q_i ;
- 3) увеличилось количество запросов обновления MP_i .

Пусть n_{Ompi} будет количеством запросов вида q_i , использующих MP_i на этапе наблюдения, а $n_{\delta mpi}$ – в течение времени Δt . Тогда при

$$\frac{n_{Ompi}}{n_o} \gg \frac{n_{\delta mpi}}{n_\delta} \quad (7)$$

и $\varepsilon_{\delta i} < \varepsilon_p$ (пороговое значение эффективности), можно ставить вопрос об удалении MP_i из ИС, если период включения $t_i \ll \Delta t$. В противном случае рекомендуется продлить анализ эффективности MP_i на несколько периодов Δt .

Если $\frac{n_{Ompi}}{n_o} \leq \frac{n_{\delta mpi}}{n_\delta}$, то причиной понижения

эффективности может быть изменение периодичности появления q_i либо увеличение количества запросов обновления q_{ui} . В любом случае следует поставить вопрос о проведении анализа периодичности и эффективности запроса q_i для его использования в МП. При этом следует учитывать соотношение длительности t_0 и Δt .

Если решено провести новый анализ периодичности [1], то рекомендуется использовать следующее множество запросов.

$$Qo\delta = \{q|q \in Qo \wedge \tau_{sq} \geq t_o + \Delta t\} \cup Q\delta.$$

Если время эксплуатации МП превышает время наблюдения за ИС, то во всех ранее приведенных оценках работы МП следует вместо периода наблюдения t_0 брать равный по длительности период работы ИС, что позволит получить более точные результаты.

Если отмечено снижение эффективности для многих МП, а также существенно увеличилось количество запросов, для которых нет МП, необходимо выполнить полный анализ запросов системы для нового периода наблюдения.

Анализ возможности применения новых МП

Если обеспечить накопление данных о запросах в течение времени эксплуатации МП, то появляется возможность проанализировать добавление новых МП с большими периодами включения/выключения. При отсутствии ориентировочной информации о возможной периодичности задач, решаемых в конкретной системе, можно рассмотреть периоды, основанные на общеизвестных календарных отметках времени таких, как месяц, квартал, сезон года. При этом предлагается связать с каждой временной отметкой, уточняющий атрибут «начало», «конец», «середина».

Для исследования новой периодичности следует решить две задачи:

- 1) определить запросы, для которых следует искать новый период повторения;
- 2) определить длительности нового периода повторения.

В качестве запросов – «кандидатов» для нового исследования – следует выбирать все виды запросов, для которых ранее не создавались МП, а также запросы с низкой эффективностью соответствующих МП.

При определении длительности периода повторения t_j предлагается руководствоваться отношением

$$2t_j \leq t_o + t_e,$$

где t_e – время эксплуатации механизма МП.

Метод определения периодичности запросов детально изложен в работах [1,2].

Настройка механизма управления МП после внесения существенных изменений в ИС

Если в ИС внесены существенные изменения конструктивного плана, например, удалены или добавлены новые таблицы, удалены или введены новые рабочие места и т. д., то эффективность многих МП может существенно измениться. Это может быть определено автоматически уже после следующего периода Δt . В этом случае администратору системы можно дать право исключения MP_i , если, например, запрос q_i связан с ликвидированным рабочим местом. Появление новых запросов автоматически вызовет определение для них периодов включения/выключения сначала на коротком интервале (Δt), а потом на всё более длинном по мере работы системы.

Аналогичные операции должны быть выполнены, если произошли существенные изменения в расписании работы организации (изменение начала, окончания, длительности рабочего дня; изменение расписания приёма клиентов, выдачи справок и отчётов и т. д.).

На рис.1 представлен алгоритм корректировки настроек механизма МП.

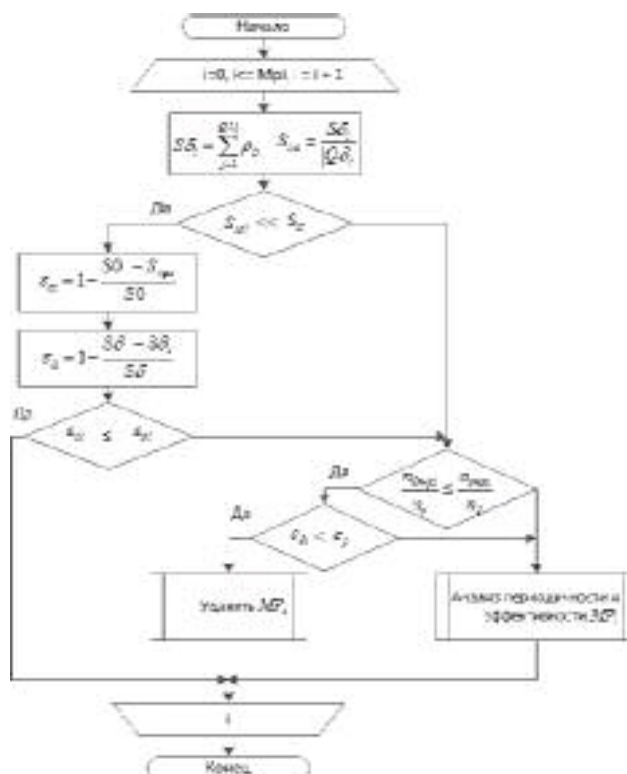


Рис.1. Алгоритм управления работой МП

Выводы

Предложенный в работе метод корректировки механизма управления МП в процессе эксплуатации ИС позволяет поддерживать эффективность МП на высоком уровне, даёт возможность создавать новые МП с большими периодами включения / выключения, а также перенастраивать механизм управления МП при внесении в ИС существенных изменений.

Список использованной литературы

1. Кунгурцев А.Б. Анализ возможности применения материализованных представлений в информационных системах /А.Б Кунгурцев, Куок Винь Нгуен Чан – Одесса: Труды ОНПУ.– 2003. – С. 102-106.

2. Кунгурцев А.Б. Сравнение запросов в реляционных базах данных для построения материализованных представлений / А.Б. Кунгурцев, Куок Винь Нгуен Чан, А.А. Бражко. – Одесса: Праці УНДІРТ. № (39). –2004. – С. 35-38.

3. Кунгурцев А.Б. Поиск закономерностей в распределении запросов для управления материальными представлениями / А.Б. Кунгурцев, Ю.Н. Возовиков, Као Ван Ань. – Одесса: Труды ОПУ. – Вып. 2 (30). – 2008. – С. 135-140.

4. Кунгурцев А.Б. Управление материализованными представлениями в информационных системах / А.Б. Кунгурцев, Ю.Н. Возовиков. – Харьков: Восточно-европейский журнал передовых технологий. – № 1/4(43). – 2010.– С. 18-21.

Получено 23.10.2011

References

1. Kungurtsev A., Nguyen Quoc Vinh. Tran Analysis of the possibility of using materialized representations in information systems. – Odessa: Proceedings ONPU – 2003. – P. 102–106 [in Russian].

2. Kungurtsev A., Nguyen Quoc Vinh Tran, Brazhko A.A. Comparison of queries in relational databases for of constructing materialized representation-making – Odessa: Pratsi UNDIRT. –№ (39). 2004. – P. 35–38. [in Russian].

3. Kungurtsev A., Vozovikov Y., Cao Van Anh. Search for regularities in the distribution of requests for materials management concepts. – Odessa: Proceedings of the OPU.– Vol. 2 (30). – 2008.– P. 135–140 [in Russian].

4. Kungurtsev A., Vozovikov Y. Management materialistic polarized views in information systems. – Kharkov: East European Journal of advanced technologies. – 2010.– No 1/4(43). – P.18–21 [in Russian].



Кунгурцев
Алексей Борисович,
проф.каф.систем.програм.
обеспеч. Одесск. нац. поли-
техн. ун-та.
Тел.(048) 779-74-76
e-mail: abkun@te.net.ua



Возовиков
Юрий Николаевич,
аспирант каф.систем
программ.обеспеч.Одесск.нац.
политехн.ун-та.
Тел. +38067 784 2891
e-mail: yuri_email@mail.ru.