

## Анализ избыточности хранения темпоральных данных средствами реляционных СУБД

© А.В. Балдин, С.А. Тоноян, Д.В. Елисеев

МГТУ им. Н.Э. Баумана, Москва, 105005, Россия

*Рассмотрены различные состояния базы данных (снимки), которые отражают различные состояния объектов реального мира, изменяющихся со временем под влиянием внешних факторов. Показано, что в таких системах все хранимые в базе данные являются актуальными в момент выполнения соответствующих запросов и хранят единственное (текущее) состояние объектов предметной области. Последнее свойственно большинству современных реляционных баз данных. Указано существование множества предметных областей, в которых необходимо хранить историю изменения структуры базы данных и самих данных. В таких системах сложно выполнить запрос к изменяющимся данным и определить ограничения целостности, которые содержатся в нескольких состояниях базы данных. Одним из путей решения подобных задач является применение нетемпоральной модели базы данных и расширение функциональности языков запросов. Предложено провести преобразование нетемпоральной модели базы данных к темпоральной посредством добавления временных отметок к данным для хранения их истории. Это достигается введением в базу данных новых кортежей, в которых может изменяться значение только одного атрибута, что приводит к избыточности хранения информации. В статье дан анализ и оценка избыточности хранения информации в темпоральных реляционных базах данных.*

**Ключевые слова:** структура данных, операции изменения, ограничения целостности, темпоральные данные, темпоральные модели СУБД, нетемпоральные модели СУБД, реляционная база данных, кортеж, многомерное пространство, избыточность хранения данных.

**Темпоральные данные и темпоральные базы данных.** Система управления базами данных (СУБД) основывается на модели, которая определяет конструкции и формализмы, доступные для определения, изменения и доступа к данным, и состоит из трех компонент: структуры данных, операций и ограничений целостности [1]. Модель данных позволяет описывать объекты предметной области и выполнять над ними соответствующие операции.

Совокупность значений, описывающих объект предметной области в определённый период времени, называется состоянием объекта предметной области, а совокупность состояний объектов предметной области в определённый период времени — состоянием базы данных.

Современные СУБД хранят единственное состояние объектов предметной области, как правило, текущее. Такие СУБД и соответствующие им модели данных называются нетемпоральными. Они поддерживают операции изменения, которые переводят базу данных из од-

ного состояния в другое, тем самым заменяя старые значения новыми. В таких системах считается, что все хранимые данные являются действительными в момент выполнения соответствующих запросов.

Существует множество предметных областей, в которых необходимо получать доступ не только к самому последнему состоянию базы данных, но и к прошлым и будущим состояниям. Это, например, система управления предприятием [2, 3], система управления персоналом [4], финансовые приложения, страховые приложения и ряд других. Одной из причин, по которой управление изменяющимися во времени данными не рассматривается для большинства приложений, является отсутствие соответствующей поддержки современными СУБД.

*Темпоральные данные* — произвольные данные, которые связаны с определёнными датами или промежутками времени [5]. В вышеуказанных предметных областях различные состояния базы данных сохраняются как темпоральные данные. Темпоральные модели данных позволяют хранить информацию об эволюции объектов: для любого объекта, который был создан в момент времени  $t_1$  и закончил свое существование в момент времени  $t_2$ , в базе данных будут сохранены все его состояния на временном интервале  $[t_1, t_2]$ .

*Темпоральная СУБД* — это СУБД для хранения и обработки темпоральных данных [5]. Такие системы и содержащиеся в них данные можно рассматривать как темпоральные только в том случае, если известно правило интерпретации временных меток и интервалов. В категорию темпоральных СУБД не попадают обычные реляционные СУБД, в которых поддерживаются связанные со временем типы данных, но интерпретацией и связью данных между собой с учетом времени занимается разработчик приложения. В темпоральной СУБД учитывается изменчивость данных с течением времени.

**Темпоральность на уровне базы данных** означает, что все отношения в базе данных и связанные с ними кортежи и значения атрибутов имеют одинаковые периоды действительного времени. База данных фиксируется отметкой времени и принимается однородность всех этих уровней.

Темпоральность на уровне базы данных подразумевает, что периодически сохраняются снимки ее состояний в определенные моменты времени. При таком подходе темпоральной модели свойственна избыточность данных, так как в темпоральных базах данных два смежных состояния отличаются незначительно, а сохраняются отдельно, что приводит к многократному дублированию одних и тех же данных.

**Темпоральность на уровне отношения** означает, что каждое отношение фиксируется отметкой времени. Таким образом, база дан-

ных состоит из набора исторических отношений, каждое из которых рассматривается как последовательность экземпляров отношений, причем каждый экземпляр представляет особое состояние исторического отношения. Таким образом, историческое отношение можно рассматривать как трехмерный объект, двумя измерениями которого являются атрибуты и кортежи, а третьим — время. Каждое сечение по временному измерению является статическим экземпляром отношения.

В этом случае предполагается, что все кортежи, сохраненные в каком-либо экземпляре рассматриваемого отношения, появляются во всех состояниях этого исторического отношения. Если кортеж не существует в каком-либо состоянии, он содержит пустое значение во всех атрибутах, не принадлежащих первичному ключу отношения. Это необходимо для выполнения запросов, которые сравнивают различные состояния исторического отношения.

**Темпоральность на уровне кортежа** означает, что каждый кортеж в отношении связывается с отметкой времени, т. е. содержит состояние объекта предметной области вместе с временной информацией. Состояние существует на интервале времени, в течение которого ни один из атрибутов кортежа не изменяет своего значения [1, 6].

Темпоральный кортеж обычно применяется в темпоральных реляционных моделях данных, поддерживающих только отношения в первой нормальной форме (1НФ). Темпоральность кортежа достигается за счёт расширения исходной реляционной модели данных, к которой добавляются специальные временные атрибуты в каждую нетемпоральную схему отношения. Отношение с темпоральными кортежами содержит инвариантный во времени ключ и изменяющиеся во времени атрибуты.

*Синхронными атрибутами отношения* называют атрибуты, которые изменяют свои значения всегда в одно и то же время, а *асинхронными* — атрибуты, которые изменяют свои значения независимо от других атрибутов в отношении [1].

Главным недостатком темпорального кортежа является то, что информация об объекте предметной области распространена на несколько кортежей, каждый из которых представляет состояние этого объекта, существовавшее в течение определенного периода времени. Темпоральность кортежа также приводит к избыточности хранения данных. При обновлении значений кортежа в отношении создается новый кортеж, где все значения атрибутов, не затронутых модификацией, повторяются.

*Вертикальная темпоральная аномалия* — это свойство реляционного отношения с темпоральными кортежами, обозначающее факт, при котором история объекта реального мира распространяется на несколько кортежей [1, 6]. Не существует способа исключить верти-

кальную темпоральную аномалию в темпоральных реляционных моделях данных, остающихся в 1НФ и использующих темпоральность кортежа.

*Темпоральная нормальная форма* — это отношение с темпоральными кортежами, в котором продолжительность действия кортежа и значений его атрибутов являются одинаковыми либо все атрибуты в отношении являются синхронными [1].

Темпоральная нормальная форма отношения достигается посредством нормализации отношений, где все атрибуты изменяют свои значения одновременно. Однако в случае, когда отношение не имеет синхронных атрибутов, темпоральная нормальная форма приводит к тому, что все отношения состоят из ключевого атрибута, который является постоянным во времени; атрибута, изменяющегося во времени, и отметки времени.

*Горизонтальная темпоральная аномалия* — это свойство реляционного отношения с темпоральными кортежами, обозначающее факт нормализации исходного отношения к отношению в темпоральной нормальной форме [1].

Вертикальная и горизонтальная темпоральные аномалии возникают в темпоральных моделях данных, использующих темпоральность кортежей. При вертикальной аномалии не происходит дополнительная нормализация исходного отношения, но возникает избыточность хранения данных, а при горизонтальной аномалии избыточность хранения данных уменьшается, но необходима дополнительная нормализация исходного отношения, что приводит к многократному выполнению операции соединения для получения результата запроса. Таким образом, при увеличении вертикальной аномалии уменьшается горизонтальная и наоборот. На практике горизонтальной аномалии стараются избегать, увеличивая вертикальную аномалию.

Используя темпоральность на уровне кортежа, Р. Снодграсом [7], К. Дженсеном [7], Дж. Бен-Зви [1, 7] были разработаны темпоральные модели, основанные на обобщенной темпоральной модели, которая состоит из набора отношений, имеющих набор атрибутов  $A_1, \dots, A_n, T$ , где  $T$  — атрибут, описывающий момент времени.

**Оценка избыточности хранения данных в темпоральных реляционных моделях.** Поскольку в темпоральной базе данных хранится история изменений состояния объекта, то объем этой информации будет значительно превосходить объем информации в нетемпоральной системе. Если же обратиться к вопросу количественной оценки этих различий, то однозначный ответ дать сложно, так как объем информации будет сильно зависеть от уровня темпоральности данных и частоты изменения информации.

Рассмотрим темпоральность на уровне кортежа. В этом случае отношение состоит из темпоральных кортежей, которые определяют состояния объектов предметной области. График зависимости размера отношения с темпоральными кортежами при изменении схемы отношения представлен на рисунке.

При неизменной схеме отношения его размер растет линейно при увеличении числа кортежей:

$$V_{\text{отн}} = kn,$$

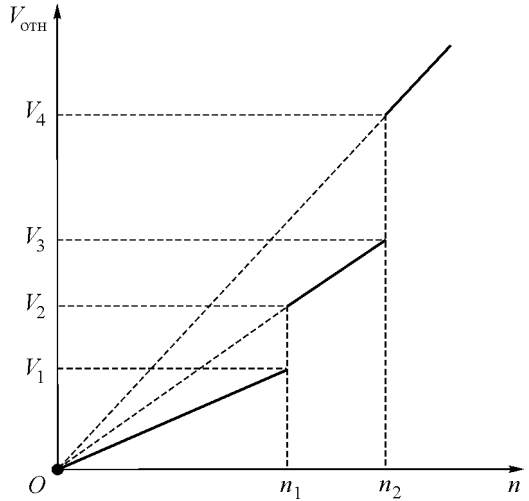
где  $k$  — число атрибутов в схеме отношения.

Изменение схемы отношения в этом случае состоит только в добавлении новых атрибутов, чтобы сохранить накопленную ранее историю. При таком изменении схемы отношения размер отношения увеличивается скачкообразно, так как место под новые атрибуты выделяется в уже существующих кортежах. Изменение схемы отношения изображено на графике при числе кортежей  $n_1$  и  $n_2$ . При изменении схемы отношения, когда  $n = n_1$ , размер отношения увеличивается от  $V_1$  до  $V_2$  и изменяется угол наклона графика. Аналогично происходит, когда  $n = n_2$ .

Математически зависимость размера отношения с темпоральными кортежами от числа кортежей в отношении при изменяющейся схеме отношения имеет вид

$$V_{\text{отн}} = \begin{cases} kn, & 0 \leq n < n_1; \\ (k + \Delta k_1)n, & n_1 \leq n \leq n_2; \\ (k + \Delta k_1 + \Delta k_2)n, & n_2 \leq n \leq n_3; \\ \dots \dots \dots \end{cases}$$

Таким образом, после каждого изменения схемы отношения увеличивается угол наклона прямой, т.е. размер отношения возрастает быстрее. Кроме того, устаревшие атрибуты из схемы отношения не удаляются, что приводит к разрастанию схемы, сложным запросам и уменьшению эффективности их выполнения.



Зависимость размера отношения  $V_{\text{отн}}$  с темпоральными кортежами от их числа  $n$  в отношении при изменяющейся схеме отношения

**Выводы.** Проведенный анализ показал, что хранение темпоральных данных в реляционных СУБД приводит к избыточности, связанной с дублированием информации, которая не изменяется в последовательных состояниях объектов предметной области.

Рассматриваемая избыточность хранения темпоральных данных увеличивается при изменении структуры темпоральных объектов предметной области, так как изменяются структуры уже записанных неактуальных данных. Это требует разработки новых методов хранения и доступа к темпоральным данным с изменяющейся структурой.

## ЛИТЕРАТУРА

- [1] Steiner A. A generalisation approach to temporal data models and their implementations, 1998.  
URL: <http://www.timeconsult.com/Publications/diss.pdf> (дата обращения: 14.08.2011).
- [2] *Информационные технологии в инженерном образовании*. Коршунов С.В., Гузненков В.Н., ред. Москва, Изд-во МГТУ им. Н.Э. Баумана, 2007, 432 с.
- [3] Федоров И.Б., Черненький В.М. *Информационная управляющая система МГТУ им. Н.Э. Баумана «Электронный университет»: концепция и реализация*. ред. Москва, Изд-во МГТУ им. Н.Э. Баумана, 2009, 376 с.
- [4] *«Персонал – Кадры» – адаптивная функционально-развиваемая система информационной поддержки управления кадрами*.  
URL: <http://www.personal.bravosoft.ru/doc/doc002.htm> (дата обращения: 14.08.2011).
- [5] Костенко Б.Б., Кузнецов С.Д. *История и актуальные проблемы темпоральных баз данных*, 2007.  
URL: <http://citforum.ru/database/articles/temporal/4.shtml> (дата обращения: 14.08.2011).
- [6] Сафонов В.С. К вопросу о темпоральных расширениях реляционных баз данных. *«Проблемы развития территории»*, 2009.  
URL: <http://pdt.vscs.ac.ru/?module=Articles&action=view&aid=436> (дата обращения: 30.09.2011).
- [7] Базаркин А. Н. *Исследование и разработка темпоральной модели данных в рамках МИС Интернет PROMIS*, 2009. URL: [https://edu.botik.ru/proceedings/psta2009/v2/037-Bazarkin.Research\\_and.pdf](https://edu.botik.ru/proceedings/psta2009/v2/037-Bazarkin.Research_and.pdf) (дата обращения: 14.08.2011).

Статья поступила в редакцию 17.07.2014

Ссылку на эту статью просим оформлять следующим образом:

Балдин А.В., Тоноян С.А., Елисеев Д.В., Анализ избыточности хранения темпоральных данных средствами реляционных СУБД. *Инженерный журнал: наука и инновации*, 2014, вып. 4.

URL: <http://engjournal.ru/catalog/it/hidden/1273.html>

**Балдин Александр Викторович** родился в 1950 г., окончил МВТУ им. Н.Э. Баумана в 1972 г. Д-р техн. наук, директор НОЦ ЭУ МГТУ им. Н.Э. Баумана. Автор более 90 научных работ. Сфера научных интересов: информационные технологии. e-mail: bal@bmstu.ru

**Тоноян Славик Анушаванович** родился в 1955 г., окончил МВТУ им. Н.Э. Баумана в 1977 г. Канд. техн. наук, доцент кафедры «Системы обработки информации и управления» МГТУ им. Н.Э. Баумана. Автор более 30 научных работ. Область научных интересов: информационные технологии. e-mail: tonoyansl@mail.ru

**Елисеев Дмитрий Владимирович** родился в 1986 г., окончил МГТУ им. Н.Э. Баумана в 2008 г., Канд. техн. наук, программист НОЦ ЭУ МГТУ им. Н.Э. Баумана. Автор более 15 научных работ. Область научных интересов: информационные технологии. e-mail: d-eli@mail.ru

## **Analysis of temporal data storage redundancy by means of RDBMS**

A.V. Baldin, S.A. Tonoyan, D.V. Eliseev

Bauman Moscow State Technical University, Moscow, 105005, Russia

*Real-world objects change with time, and this factor sets the database from one state to another, replacing the current values by new data. We show that in such systems all stored data is up-t- date at the time of execution of the corresponding requests. The majority of modern database management systems (DBMS) hold the only current domain objects. However, there are many subject areas in which you need to keep a history of changes to the database and data. The challenge is to query time-varying data and identify constraints that are contained in several of the states in the database, using the non-temporal data models and query languages. The conversion of the non-temporal database model for the temporal can be done by adding periods of time to the data to store their history. This is accomplished by adding the DB, new tuples, which can change only the value of one attribute, and this leads to redundancy of information.*

*The article provides an analysis and assessment of redundancy of information in the temporal relational databases.*

**Key words:** *data structure, modification operations, integrity constraints, temporal data, temporal model DBMS, not temporal model DBMS, relational database, tuple, multidimensional space, redundancy of data storage.*

### REFERENCES

- [1] Steiner A. *A generalisation approach to temporal data models and their implementations*. 1998. Available at: <http://www.timeconsult.com/Publications/diss.pdf> (accessed 14 August 2011).
- [2] *Informatsionnye tekhnologii v inzhenernom obrazovanii* [Information technology in engineering education]. Ed. S.V. Korshunov, V.N. Guznenkov. Moscow, BMSTU Publ., 2007, 432 p.
- [3] *Informatsionnaia upravliaiushchaia sistema MGTU im. N.E. Baumana «Elektronnyi universitet»: kontseptsiiia i realizatsiia* [Information management system of BMSTU "Electronic University": the concept and realization]. S.A. Tonoyan, A.V. Baldin, V.A. Baryshnikov et al. Ed. I.B. Fedorov, V.M. Chernen'ky. Moscow, BMSTU Publ., 2009, 376 p.
- [4] «*Personal – Kadry*» – *adaptivnaia funktsional'no-razvivaemaia sistema informatsionnoi podderzhki upravleniia kadrami* ["Personnel" - adaptive functional-developing system of personnel management information support]. Available at: <http://www.personal.bravosoft.ru/doc/doc002.htm> (accessed at: 14 August 2011).
- [5] Kostenko B.B., Kuznetsov S.D. *Istoriia i aktual'nye problemy temporal'nykh baz dannykh* [History and Challenges of temporal databases]. 2007. Available at: <http://citforum.ru/database/articles/temporal/4.shtml> (accessed at: 14 August 2011).
- [6] Safonov V.S. Elec. journal. *Problemy razvitiia territorii* [Problems of the territory development]. 2009. Available at: <http://pdt.vsc.ac.ru/?module=Articles&action=view&aid=436> (accessed at: 30 September 2011).



- [7] *Issledovanie i razrabotka temporal'noi modeli dannykh v ramkakh MIS Interin PROMIS* [Research and development of the temporal data model IIA Interin PROMIS]. Bazarkin A. N., 2009. Available at: [https://edu.botik.ru/proceedings/psta2009/v2/037-Bazarkin.Research\\_and.pdf](https://edu.botik.ru/proceedings/psta2009/v2/037-Bazarkin.Research_and.pdf) (accessed at: 14 August 2011).

**Baldin A.V.**, Dr. Sci. (Eng.), director of the Scientific Educational Center “Electronic University” at Bauman Moscow State Technical University. Author of more than 90 publications in the field of information technologies. e-mail [bal@bmstu.ru](mailto:bal@bmstu.ru)

**Tonoyan S.A.**, Ph.D. (Eng.), lecturer of the Information Processing and Control Systems Department at Bauman Moscow State Technical University. Author of more than 30 publications in the field of computer and information technologies. e-mail: [tonoyansl@mail.ru](mailto:tonoyansl@mail.ru)

**Eliseev D.V.**, Ph.D. (Eng.), programmer of the Scientific Educational Center “Electronic University” at Bauman Moscow State Technical University. Author of more than 10 publications in the field of computer and information technologies. e-mail [d-eli@mail.ru](mailto:d-eli@mail.ru)