

УДК 004.9

А.В. Балдин, Д.В. Елисеев

ОБЗОР СПОСОБОВ ПОСТРОЕНИЯ ТЕМПОРАЛЬНЫХ СИСТЕМ НА ОСНОВЕ РЕЛЯЦИОННОЙ БАЗЫ ДАННЫХ

Проведен анализ подходов к реализации темпоральных систем на основе реляционной базы данных, даны определения «темпоральные данные» и «темпоральная база данных». Представлены подходы к моделированию времени, на основе которых описаны типы временных меток, используемые в темпоральных базах данных. Приведены два способа реализации темпоральных систем, в которых применяются современные реляционные базы данных: реализация темпоральной поддержки на уровне приложения и расширение нетемпоральной реляционной модели до темпоральной модели. Для каждого подхода выявлены соответствующие достоинства и недостатки.

E-mail: iu5vmch@ramler.ru

Ключевые слова: темпоральная база данных, темпоральные данные, реляционная модель данных, моделирование времени, момент времени, временной интервал.

Введение. Нетемпоральные модели данных и их реализация в системах управления базами данных (СУБД) сохраняют единственное состояние объектов предметной области. СУБД поддерживают операции изменения, которые переводят базу данных из одного состояния в другое, тем самым заменяя старые значения новыми. В таких системах все хранимые в базе данные являются актуальными в момент выполнения соответствующих запросов.

Однако существует множество информационных систем, для которых необходимо хранить прошлые состояния базы данных и возможно будущие, например, система управления предприятием [1, 2], система управления персоналом [3, 4], финансовые приложения, страховые приложения и другие системы, которые обладают свойствами:

- 1) обрабатывать темпоральные (изменяющиеся во времени) данные;
- 2) накапливать историю изменения темпоральных данных.

В настоящее время реляционная модель наиболее распространена в базах данных. В разработку методов хранения и обработки темпоральных данных на базе реляционной модели внесли вклад такие ученые, как Р. Снодграс [5, 10], К. Дженсен [5, 11], Дж. Бен-Зви [5], С. Гадия [5], Е. МакКензи [5, 13], А. Стейнер [12]. Исследования темпоральных баз данных привели к многообразию соответствующих темпоральных моделей. В статье приведены подходы к модели-

рованию времени в темпоральных моделях, а также подходы к их реализации на базе реляционных СУБД.

Темпоральные данные и темпоральные базы данных. Система управления базой данных основана на модели данных, которая определяет конструкции и формализмы, доступные для описания данных, их изменения и доступа к ним. Модель данных

$$M = (DS, OP, C),$$

где DS — структура данных; OP — структура операций; C — структура ограничений целостности [8, 9, 12].

Совокупность значений, описывающих объект предметной области в определенный период времени, называется *состоянием объекта предметной области*. Совокупность состояний объектов предметной области в определенный период времени — это *состояние базы данных*.

В современных СУБД обрабатываются постоянные данные, т. е. данные существуют дольше, чем приложения выполняют задачи. Одной из причин, по которой управление изменяющихся во времени данных не рассматривается для большинства приложений, является отсутствие соответствующей поддержки современными СУБД.

Темпоральные данные — произвольные данные, которые связаны с определенными датами или периодами времени [6].

В перечисленных выше информационных системах различные состояния базы данных сохраняются как темпоральные данные. Темпоральные модели данных хранят информацию об эволюции объектов: для любого объекта, созданного в момент времени t_1 и закончившего свое существование в момент времени t_2 , база данных запомнит все его состояния на временном интервале $[t_1, t_2)$ [7].

Темпоральная СУБД — это СУБД для хранения и обработки темпоральных данных [6].

Система управления базами данных и содержащиеся в них данные могут рассматриваться как темпоральные данные только в том случае, если известно правило интерпретации временных меток и интервалов. В категорию темпоральных СУБД не входят обычные реляционные СУБД, в которых поддерживаются связанные со временем типы данных. Однако интерпретацией и связью данных между собой с учетом времени приходится заниматься разработчику приложения [6]. В темпоральной СУБД учитывается изменчивость данных с течением времени.

Представление времени в темпоральных моделях данных. Темпоральная модель данных на базе реляционной модели расширяет конструкции последней, чтобы описывать и обрабатывать темпоральные данные. Принципиальные отличия между темпоральными моделями данных на базе реляционной модели зависят от интерпретации ключевых понятий, используемых в этих моделях. К ключевым понятиям относятся, в частности представление времени и тип временной метки [12].

Время в темпоральных моделях данных — набор чисел, которые упорядочены относительно предикатов сравнения. Таким образом, время может быть представлено как ось в системе координат. В темпоральных моделях данных время является непрерывным или дискретным [12].

В непрерывной модели время рассматривается как изоморфное к вещественным числам. Каждое вещественное число соответствует точке на оси времени. Следовательно, в непрерывной модели времени между двумя временными точками существует другая временная точка

$$\forall t_1, t_2, t_1 < t_2; \exists t_3 : t_1 < t_3 < t_2.$$

Поскольку темпоральные модели используются в базах данных, то невозможно преобразовать эту модель времени к существующим типам данных без потерь.

Дискретная модель преобразует время к целым числам [12]. В такой модели можно определить функцию поиска следующего момента времени δ . Между точками на оси времени и других временных точек не существует

$$\forall t_1, t_2, \text{ если } t_2 = \delta(t_1), \text{ то не } \exists t_3 : t_1 < t_3 < t_2.$$

Ось времени — набор временных точек. Наименьшая неразложимая временная единица на оси времени называется *квантом времени* [12]. Таким образом, отдельный квант времени является подынтервалом фиксированной продолжительности на оси времени. Наименьшая неразложимая временная единица необходима, так как цифровые компьютеры поддерживают только ограниченную степень детализации вещественных чисел. Для непрерывной модели времени квант времени охватывает небольшую часть оси времени, тогда как временная точка представляет единственную точку на этой оси. Для дискретной модели кванты времени и временные точки определяются аналогично.

Пусть задан момент времени t , соответствующий точке на оси времени. Тогда событие — мгновенный факт, который выполняется в момент времени t [12]. При рассмотрении более укрупненной степени детализации временных единиц, соответствующих квантам времени, если событие происходит в квант времени qt , то оно происходит в любой момент времени t в течение кванта времени qt : $t \in qt$.

Темпоральная база данных хранит факты и время, когда они были действительны в предметной области. С теоретической точки зрения это означает, что все моменты времени, когда факт является истинным, должны быть сохранены в темпоральной базе данных. Когда это множество моментов времени образует плотное подмножество, оно может быть записано как временной интервал. Временные интервалы используются для моделирования периода времени, в течение которого факт был истинным в предметной области [12].

Интервал времени представляет собой период времени с начальными моментом времени t_1 и конечными моментом времени t_2 , т. е. наименьшими и наибольшими значениями временного интервала. Наименьшее и наибольшее значения интервала могут сравниваться с каждым другим моментом времени предикатами сравнения. Временной интервал $[t_1, t_2)$ рассматривается как интервал, в который входит наименьшее значение (t_1) и не входит наибольшее значение (t_2) [12].

Для интервалов определен набор теоретических операций объединения, пересечения и разности [12]. Эти операции необходимы, чтобы делать выводы о темпоральных фактах. Однако рассматриваемые операции не замкнуты относительно временных интервалов. Результатом объединения двух непересекающихся интервалов является набор интервалов, результатом разности двух интервалов — пустое множество, один или два интервала.

Пример. Пусть даны три интервала:

$$I_1 = [2000, 2006), I_2 = [2008, 2011), I_3 = [1995, 2012).$$

Тогда

$$I_1 \cup I_2 = \{[2000, 2006), [2008, 2011)\};$$

$$I_1 \setminus I_3 = \emptyset;$$

$$I_3 \setminus I_2 = \{[1995, 2008), [2011, 2012)\}.$$

Для интервалов также можно определить логические операции включения и сравнения: предшествование интервалов; пересечение интервалов; один интервал непосредственно следует за другим; интервал содержит другой интервал; равенство интервалов и т. д.

Как правило, на основе реляционной модели могут описывать темпоральные данные либо с помощью момента времени, либо с помощью временного интервала. Темпоральные реляционные модели данных, остающиеся в первой нормальной форме (1НФ), ограничены использованием временных меток, которые можно преобразовать к дополнительным скалярным значениям атрибутов кортежа.

При описании темпоральных данных с помощью момента времени предполагается, что данные истинны только в заданный момент времени. Отношения, содержащие темпоральные данные с моментом времени, называют **таблицами событий**. Темпоральная реляционная модель с моментом времени может быть смоделирована путем расширения схемы с неявным атрибутом T типа «дата». При этом отношение остается в 1НФ. Если $R = (A_1, A_2, \dots, A_n)$ — схема нетемпорального отношения, то таблица событий для него имеет следующую схему: $R_t = (T, A_1, A_2, \dots, A_n)$.

При описании темпоральных данных с помощью временного интервала предполагается, что данные истинны в течение заданного временного интервала. Отношения, содержащие темпоральные данные с временным интервалом, называют **таблицами состояний** [12]. Тем-

поральность с временными интервалами может использоваться с отношениями в 1НФ, если наименьшее и наибольшее значения временного интервала преобразуются в два дополнительных атрибута T_1 и T_2 типа «дата». Если $R = (A_1, A_2, \dots, A_n)$ — схема нетемпорального отношения, то таблица состояний для него будет иметь следующую схему: $R_t = (T_1, T_2, A_1, A_2, \dots, A_n)$.

Записи в отношении, которые являются истинными в течение нескольких неперекрывающихся периодов времени, хранятся отдельно для каждого периода времени. Временные интервалы не замкнуты относительно ряда теоретико-множественных операций (разность и объединение), то история объектов предметной области может распространяться на несколько записей как при хранении отношений, так и при вычислении результата запроса. Рассмотрим пример таблицы состояний для отношения «Работник», в которой хранятся данные о двух работниках, действительные в разные неперекрывающиеся временные интервалы.

Пример таблицы состояний отношения «Работник»

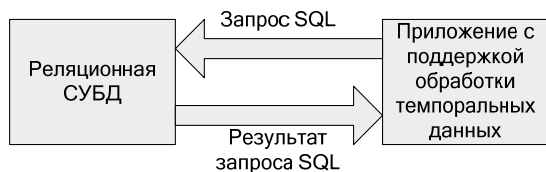
Id	Имя	Зарплата	Подразделение	T_1	T_2
1	Иван	20 000	Отдел 1	01.10.2000	30.11.2005
1	Иван	25 000	Отдел 2	01.12.2005	31.08.2010
1	Иван	30 000	Отдел 1	01.09.2010	—
2	Петр	20 000	Отдел 2	02.03.2004	06.07.2008
2	Петр	20 000	Отдел 2	07.07.2008	31.12.2011

Способы реализации темпоральных баз данных. Изменение или расширение модели данных для работы с темпоральными объектами означает, что реализация модели, т. е. СУБД, также должна быть изменена. Поскольку большинство СУБД рассматриваются как черные ящики, то соответствующее изменение программного обеспечения может быть выполнено только разработчиками СУБД. Еще один подход обработки темпоральных данных заключается во встраивании темпоральной функциональности в приложение или использование расширяемости, присущей некоторым СУБД. Фактически, можно выделить два подхода к реализации темпоральных реляционных баз данных:

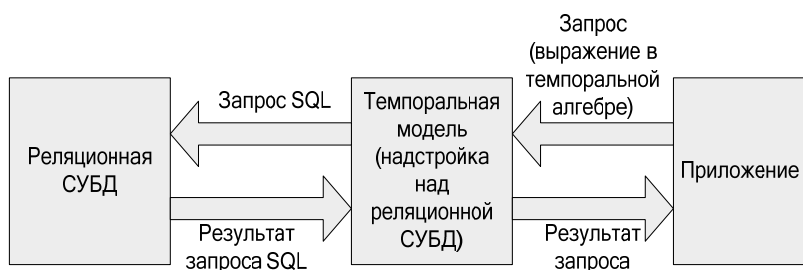
- 1) реализация темпоральной поддержки на уровне приложения;
- 2) расширение нетемпоральной модели данных до темпоральной модели [5].

Первый подход не подразумевает каких-либо изменений СУБД. Система берется как есть, а работа с темпоральными данными реализуется программистами приложения. Второй подход связан с доработкой реляционной СУБД.

При реализации темпоральной поддержки на уровне приложения (часть *a* рисунка) единственным инструментом является использование атрибута типа «дата», поддерживаемого непосредственно СУБД, с помощью которого описываются временные метки хранящихся данных. Любая темпоральная семантика должна быть встроена в приложение программистом самостоятельно. Темпоральные операции OP^T и темпоральные ограничения целостности C^T также должны поддерживаться в самом приложении [5, 12].



a



б

Схемы реализации темпоральной поддержки на уровне приложения (*a*) и расширения реляционной модели данных до темпоральной модели (*б*)

Данный подход означает, что дополнительная нагрузка ложится на программиста, так как отсутствует специальная поддержка со стороны СУБД. Это ненадежно и требует много времени для дополнительной реализации темпоральной оболочки, используемой приложением. Темпоральная оболочка обычно заново разрабатывается каждой компанией, имеющей дело с темпоральными данными. В этом случае отсутствует единый стандартизированный подход к реализации обработки темпоральных данных, что приводит к неэффективным, сложным в администрировании приложениям и проблемам при их доработке.

Расширение нетемпоральной модели данных до темпоральной модели заключается в использовании понятий, уже поддерживаемых нетемпоральными моделями, для определения темпорального расширения. Новые понятия вводятся, когда трудно или невозможно выразить их в нетемпоральной модели [5, 12].

Так, темпоральные данные можно описать, расширив нетемпоральную схему или типы специальными временными атрибутами.

Языки запросов и алгебры расширяют дополнительными операциями для обработки темпоральных данных. В результате для темпоральных моделей данных определяют темпоральную алгебру, которой обрабатываются специальные временные атрибуты для темпоральных вычислений. Это расширение схем и языков запросов можно применять для реляционной СУБД (часть *б* рисунка).

В схеме, приведенной на части *б* рисунка, темпоральность реляционных структур данных достигается за счет связывания их с соответствующими временными метками. При этом темпоральная модель представляет собой некоторую надстройку над реляционной СУБД, в которой используются понятия и конструкции последней.

Фактически, такой подход является самым используемым на практике для реализации темпоральных моделей данных. Преимущество заключается в том, что только часть нетемпоральной модели должна быть изменена, например, язык запросов и ограничения целостности. Относительно существующей СУБД это означает, что часть ее должна быть также изменена. Так, методы доступа и структуры хранения данных не затрагиваются.

К недостатку этого подхода относится автоматическое наследование темпоральной моделью ограничений исходной модели вследствие повторного использования существующих понятий нетемпоральной модели данных. Добавляя дополнительные атрибуты к отношениям, только специальные структуры данных, например кортежи, могут быть темпоральными. Конструкции нетемпоральной модели данных могут также изменяться со временем, например набор отношений и схем отношений, поэтому хранение и обработку истории развития каждой из этих конструкций необходимо выполнять в темпоральных системах, что при рассматриваемом подходе реализовать невозможно.

Заключение. Проведен анализ подходов к моделированию времени и способов описания времени на основе реляционных баз данных. Поскольку темпоральные базы данных, в отличие от нетемпоральных, сохраняют историю изменения состояний объектов предметной области, с которой связано выполнение темпоральных запросов, то были рассмотрены подходы и способы реализации таких темпоральных систем. Наиболее используемым на практике является подход расширения реляционной модели для обработки темпоральных данных. Существенное достоинство данного подхода — использование хорошо проработанных механизмов современных СУБД, которые не требуется заново реализовывать для темпоральной СУБД. Таким образом, рассмотренные подходы к моделированию времени и способы описания темпоральных данных на основе реляционной модели позволяют использовать современные реляционные СУБД при разработке темпоральных систем.

СПИСОК ЛИТЕРАТУРЫ

1. Информационная управляющая система МГТУ им. Н.Э. Баумана «Электронный университет»: концепция и реализация / Т.И. Агеева, А.В. Балдин, В.А. Барышников и др.; Под ред. И.Б. Федорова, В.М. Черненко. — М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. — 376 с.

2. Информационные технологии в инженерном образовании / Под ред. С.В. Коршунова, В.Н. Гузненкова. М.: Изд-во МГТУ им. Н.Э. Баумана, 2007. — 432 с.
3. Куприянова С.В. «1С: Зарплата и Кадры Бюджетного Учреждения 8» — мощный инструмент по автоматизации расчета заработной платы в учебных заведениях // Новые информационные технологии в образовании: Десятая Междунар. науч.-практич. конф.: Сб. науч. тр. М., 2010. Ч. 1. — С. 311—315.
4. «Персонал — Кадры» — адаптивная функционально-развиваемая система информационной поддержки управления кадрами [Электронный ресурс] — Режим доступа: <http://www.personal.bravosoft.ru/doc/doc002.htm> (дата обращения: 09.04.2012).
5. Исследование и разработка темпоральной модели данных в рамках МИС Интернет PROMIS [Электронный ресурс]/ А.Н. Базаркин, 2009. — Режим доступа: https://edu.botik.ru/proceedings/psta2009/v2/037Bazarkin.Research_and.pdf (дата обращения: 09.04.2012).
6. История и актуальные проблемы темпоральных баз данных [Электронный ресурс] / Б.Б. Костенко, С.Д. Кузнецов, 2007. — Режим доступа: <http://citforum.ru/database/articles/temporal/4.shtml> (дата обращения: 14.08.2011).
7. Варламов О.О. Эволюционные базы данных и знаний для адаптивного синтеза интеллектуальных систем. Миварное информационное пространство. — М.: Радио и связь, 2002. — 286 с.
8. Дейт К. Введение в системы баз данных. Пер. с англ. — М.: Вильямс, 2006. — 1328 с.
9. Гарсия-Молина Г., Ульман Дж. Д., Уидом Дж. Системы баз данных. Полный курс. Пер. с англ. — М.: Вильямс, 2003. — 1088 с.
10. Snodgrass R.T. Developing Time-Oriented Database Applications in SQL. — San Francisco: Morgan Kaufmann Publishers, Inc., 1999. — 504 p.
11. Jensen C. S., Snodgrass R. T. Semantics of Time-Varying Information // Information Systems, Vol. 21. No. 4. 1996. p. 311–352. [Электронный ресурс]. Режим доступа: <http://www.cs.arizona.edu/~rts/pubs/ISJune96.pdf>, свободный (дата обращения: 09.04.2012).
12. Steiner A. A. Generalisation Approach to Temporal Data Models and their Implementations [Электронный ресурс], 1998. — Режим доступа: <http://www.timeconsult.com/Publications/diss.pdf> (дата обращения: 09.04.2012).
13. McKenzie E., Snodgrass R.T. An evaluation of relational algebras incorporating the time dimension in databases [Электронный ресурс] // ACM computing surveys. — 1991. — 23(4). — P. 501—543. Режим доступа: <http://www.cs.arizona.edu/~rts/pubs/CompSurv91.pdf> (дата обращения: 09.04.2012).

Статья поступила в редакцию 4.07.2012