

УДК 336.767

РЕШЕНИЕ ЗАДАЧ ОБЕСПЕЧЕНИЯ ЦЕЛОСТНОСТИ ДАННЫХ

Орлова Марина Вячеславовна, канд. экон. наук, доц.

Воронежский государственный университет, Университетская пл., 1, Воронеж, Россия, 394018; e-mail: orlova-it@mail.ru

Цель: рассматриваются проблемы обеспечения целостности данных в процессе функционирования прикладных информационных систем, основанных на технологиях баз данных. *Обсуждение:* применение технологии баз данных при создании информационных систем, имеющих сложную организационную, функциональную и информационную структуры, использующих информационную модель взаимосвязанных экономических объектов в совокупности с актуальными, полными и достоверными данными об их функционировании, позволяет обеспечить информацией процесс разработки алгоритмов, реализующих необходимые экономико-математические методы, а также более простые информационно-поисковые функции на основе системы запросов и отчетов. При этом конечной целью является оперативное представление точных данных в полном объеме всем специалистам предприятия, связанным с обработкой данных. Сложность рассматриваемых технологий требует создания специальных средств, которые позволят обеспечить целостность и непротиворечивость информации в процессе эксплуатации информационной системы. *Результаты:* предложены рекомендации по включению в систему обработки данных универсальных ограничений целостности, обеспечивающих непротиворечивость, полноту данных и сокращающих трудоемкость и стоимость разработки и эксплуатации информационно-экономических систем.

Ключевые слова: базы данных, автоматизированные информационные системы, непротиворечивость данных.

DOI: 10.17308/meps.2017.8/1743

Введение

База данных (БД) обладает свойством целостности, если она удовлетворяет совокупности ограничений на значения данных, описание данных и сохраняет это свойство при любых модификациях логической и физической структуры БД.

Основными критериями, определяющими способ решения проблемы целостности БД, являются:

- Способ описания или задания ограничений целостности.
- Определение правил применения ограничений.
- Определение правил обработки данных при нарушении ограничений целостности.
- Минимизация затрат на проверку ограничений целостности.

Наиболее важные ограничения целостности (ОЦ) задаются при описании модели данных. Действительно, уже на уровне модели задаются отношения между объектами и их множествами, которые должны выполняться в каждый момент времени. Эти отношения выражают наиболее устойчивые семантические связи между объектами – функциональные. В реляционных моделях функциональные связи между отношениями задаются неявно, путем дублирования ключевых атрибутов в различных отношениях базы данных. В более сложных моделях связи могут указываться явно и быть поименованными. На любой модели данных определена совокупность операций над типами данных, ограничивающая допустимые процедуры манипулирования данными. В реляционном подходе это, например, реляционная алгебра, реляционное исчисление, SQL и другие языки. Использование допустимых операций над типами данных позволяет описать дополнительные ограничения целостности, которые должны выполняться между результатами выполнения операций в любом состоянии и в любой момент времени.

Таким образом, проблема обеспечения целостности БД может быть сформулирована следующим образом.

Для некоторой модели данных (M) и совокупности операций над элементами модели данных (P) задать множество ограничений целостности (S), отражающих требования к состоянию соотношений между объектами, множествами объектов (классами объектов), которые должны выполняться в каждый момент времени функционирования БД.

Основная часть

Для анализа возможных подходов к описанию ограничений целостности проведем концептуальное проектирование фрагмента предметной области «Экономика добычи». Исходные данные, полученные в результате анализа предметной области, формулируются следующим образом.

Общая характеристика.

Нефтегазодобывающее предприятие ведет добычу нефти на четырех месторождениях. Общий фонд скважин составляет 2000 штук. Из них в текущий момент добывающие – 1200, закачивающие воду – 100, простаивающие – 700. Скважины объединены в кусты от 4 до 8 скважин, но есть и одиночные скважины. Основной целью информационной системы является оперативный учет и анализ добычи нефти, закачки воды, состояния запасов, режимов работы скважин и затрат на эксплуатацию, определение рентабельности скважин.

В результате обследования объекта определены функции управления, которым соответствуют следующие виды обработки.

P1. Составить отчет о выполнении плана добычи нефти и закачки воды за сутки (месяц и накопленные) по месторождениям.

P2. Составить отчет о режимах работы скважин за сутки (месяц и накопленные).

P3. Составить отчет о затратах по каждой скважине за месяц.

P4. Составить отчет о рентабельности каждой скважины, исходя из того, что цена определена для каждого месторождения.

P5. Составить отчет об изменении запасов по каждому месторождению за месяц.

P6. Составить отчет о дебете скважин за сутки (объем добычи – объем закачки воды).

На основании совокупности входных документов и сообщений определены следующие возможные процедуры модификации БД.

P7. Поступление данных о суточной добыче нефти, закачке воды по скважине.

P8. Поступление данных об изменении режима работы скважины.

P9. Ежемесячное изменение цены на нефть и стоимости по каждому виду затрат.

P10. Ввод в действие новой скважины с указанием ее номера, даты, режима работы.

P11. Исключение скважины из информационной системы (консервация) – перенос сведений о ней в БД – «История разработки месторождения».

P12. Ежедневная корректировка объемов запасов, исходя из объемов добычи.

P13. Ежемесячная корректировка планов добычи.

P14. Поступление данных о фактических затратах на эксплуатацию скважины с указанием даты произведенных затрат.

Исходя из классификации рассматриваемых объектов и процедур модификации БД, сформулированы следующие ограничения целостности, которые должны выполняться в каждый момент функционирования системы.

S1. Скважина однозначно определяет месторождение (может входить только в одно месторождение).

S2. Скважина в конкретный момент времени может работать только в одном режиме. Режим работы скважины однозначно определяется номером скважины и кодом режима работы.

S3. Цена нефти определяется для всех скважин месторождения. Для различных месторождений она различна.

S4. Затраты ведутся по каждой скважине, по каждому виду затрат.

S5. При консервации скважин из БД исключаются ее режимы работы и затраты.

S6. Объем накопленной добычи не может превышать объем запасов по месторождению.

S7. Объем добычи (сутки) может существовать только для добывающих скважин

S8. Объем закачки (сутки) может существовать только для нагнетательных скважин.

S9. Для простаивающих скважин не может существовать объем добычи и объем закачки.

S10. Объем дневной добычи не может превышать объем накопленной добычи.

S11. Жизненный цикл скважины всегда меньше, чем жизненный цикл месторождения.

S12. Время работы скважины в заданном режиме всегда меньше, чем жизненный цикл скважины.

Определив исходные данные для концептуального проектирования модели рассматриваемой предметной области, проведем собственно ее моделирование с использованием одной из возможных моделей. В качестве примера используем модель сущностей и связей, или ER-модель.

В ER-модели базовыми понятиями являются «тип объектов» и «связь между объектами». Типы считаются множествами, а связи – кортежами (O_1, O_2, \dots, O_n), состоящими из объектов (сущностей). Связи также объединяются в множество однотипных связей. Таким образом, ER-модель – это множество E_1, \dots, E_n однотипных объектов и R_1, \dots, R_m однотипных связей (отношений) между $E_1 \dots E_n$, имеющих различную размерность.

Третье базовое понятие – это свойство, или атрибут, множеств однотипных объектов E_i и однотипных связей R_j . С каждым атрибутом связывается его область значений, которая является либо атомарным множеством значений $V_k, k=1, \dots, l$, либо произведением некоторых атомарных множеств. Атрибут является функцией из E_i или R_j в соответствующую область значений.

Для графического изображения предметной области строится ER-диаграмма, в которой сущности изображаются прямоугольником, а связи – ромбом, который соединяется дугами с прямоугольниками. Дуги могут помечаться характеристикой связи или именами ролей, в которых сущность выступает в данной связи.

ER-диаграммы строятся на основе анализа предметной области, описываемого на естественном языке. Рекомендуемые формы предложений заключаются в соответствии понятий – сущностям, а глаголов – связям.

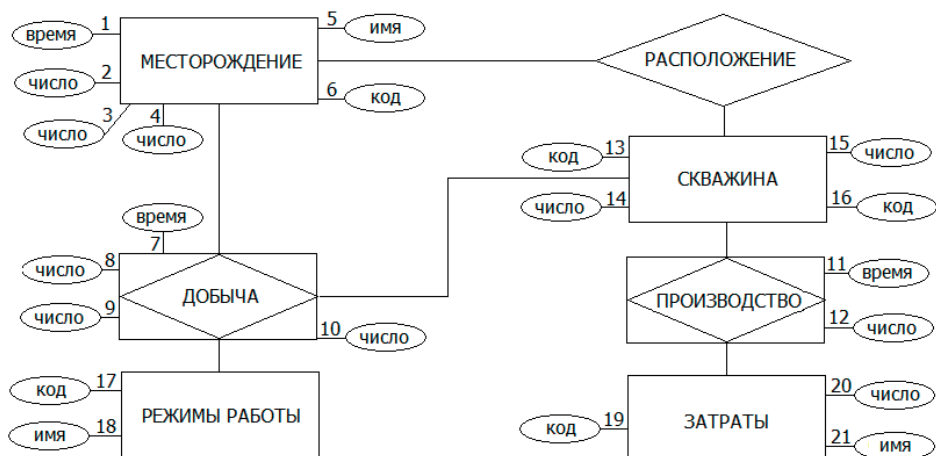


Рис. ER-диаграмма с указанием характеристик объектов и связей

Значение ссылок для рисунка 1: 1 – дата разработки; 2 – накопленная добыча; 3 – объем запасов; 4 – цена за тонну; 5 – наименование; 6 – код; 7 – дата; 8 – добыча за сутки фактическая; 9 – закачка воды за сутки; 10 – добыча за сутки по плану; 11 – дата; 12 – сумма затрат; 13 – номер скважины; 14 – накопленная добыча; 15 – накопленная закачка воды; 16 – код куста; 17 – код режима; 18 – характеристика; 19 – код затрат; 20 – цена за единицу; 21 – наименование.

Существуют два различных подхода к описанию предметной области: структурный анализ и объектно-ориентированный подход. Этим подходам соответствуют два способа задания ОЦ: процедурный и декларативный.

Структурный анализ предполагает, что объектом описания являются процессы, происходящие в предметной области, а сами информационные объекты являются элементами этого процесса, например, в нашей предметной области процессами являются: добыча нефти, ремонт скважин, в них участвуют информационные объекты: скважина, режим работы, затраты.

В объектно-ориентированном подходе первоначально описываются информационные объекты разных классов (реальные и абстрактные), а затем на их основе разрабатываются процедуры организации и ведения баз данных (процессы).

В процедурном способе явным образом указывается последовательность операций для каждого процесса (операций работы с данными). В декларативном способе процесс функционирования БД задается неявно, путем описания соотношений структур данных. Например, скважина в конкретный момент времени может работать только в одном режиме. При этом ограничение S6 в декларативном виде формулируется следующим образом: объем запасов по месторождению больше, чем объем накопленной добычи.

В процедурной форме ему соответствует следующая формулировка. При корректировке значения характеристики «объем накопленной добычи»

по месторождению проверить, что она меньше, чем значение характеристики «объем запасов».

Из приведенного примера видно, что на концептуальном уровне декларативное описание ограничений целостности является более компактным, понятным пользователю и соответствующим лексике естественного языка.

Вместе с тем на уровне реализации при использовании конкретных СУБД они в ряде случаев дают возможность использования обоих подходов, поскольку наряду со способами задания ограничений целостности важным аспектом является минимизация затрат на проверку ограничений.

Рассмотрим структуру записи ограничения целостности. На физическом уровне ограничения целостности должны выполняться при модификации данных, а именно при вводе, удалении, обновлении физических записей. Таким образом, первой характеристикой ОЦ является его место применения, а именно ввод, удаление, обновление. Результат проверки ОЦ определяет состояние БД, а следовательно, завершающую часть проверки ОЦ.

Определим, что каждое ОЦ является некоторой транзакцией к БД, характеризующейся тремя параметрами:

A1 – местом применения (I – ввод, D – удалить, U – обновить);

A2 – состоянием результата проверки (0 – целостная БД, 1 – нарушение ограничений целостности);

A3 – применением правила обработки при нарушении целостности (0 – выдача сообщения и блокировка транзакций к БД, 1 – подключение процедуры обработки).

Таким образом, ограничения целостности задаются пятеркой вида:

S (T, A1, A2, A3, PR), где PR-множество процедур обработки, выполняемых при нарушении семантической целостности БД (A3=1), T – требования к состоянию соотношений между объектами.

В условиях распределенной обработки данных проблема целостности приобретает дополнительный аспект, связанный с решением задачи одновременного доступа к данным. Такая ситуация может возникнуть, когда два или более процесса (пользователя) одновременно изменяют совместно используемые данные. В традиционных СУБД решение данной проблемы осуществляется блокировкой, т. е. запрещением двум процессам совместно использовать данные. Однако в условиях БД жесткие ограничения на совместное использование данных свели бы на нет одну из целей самого подхода к применению баз данных. Поэтому делается попытка максимальным образом ограничить монополярный доступ к базе данных. Наиболее оригинальный способ согласованности реализуется с помощью «сегментов отката». На общем уровне модификация не блокируется, но если программа (процесс), выполняя транзакцию, во время чтения БД обнаруживает измененную запись, она просматривает «сегмент отката» в поисках той вер-

сии данных, которая существовала в момент запуска запроса. Риск такого подхода состоит в том, что «сегмент отката» может переполниться, хотя на практике эта ситуация встречается редко. Администратор БД способен настраивать размер «сегмента отката».

В общем случае блокировка может иметь неприятное последствие – системный тупик. Системный тупик возникает в тех случаях, когда два (или более) одновременных процесса заблокировали некоторые ресурсы (например, запись БД) и каждому из них для полного завершения нужен ресурс, заблокированный другим процессом. Если такая ситуация не распознается и возникает системный тупик, то ни один процесс не завершится. Таким образом, ни одна из процедур в ситуации системного тупика не завершается. Очевидным решением данной проблемы является «откат» одной из транзакций и ее повторный рестарт после завершения альтернативной транзакции, когда необходимый ресурс будет освобожден.

Заключение

Установлено, что наиболее важные и постоянные ограничения целостности задаются уже на уровне описания модели, следовательно, модель данных, реализуемая СУБД, является главным фактором решения проблемы целостности баз данных.

Необходимо отметить, что игнорирование проблемы обеспечения целостности данных при разработке и эксплуатации информационных систем приводит к следующим негативным результатам:

- получению в ряде случаев неправильной информации по запросам пользователей к БД в связи с нарушением согласованности данных;
- росту количества ошибок и сбоев при работе информационной системы в связи с отсутствием системного контроля целостности данных.

Все это приводит к значительному увеличению затрат на эксплуатацию и сопровождение информационных систем. В предельном случае экономический эффект от использования информационной системы может оказаться меньше затрат на ее создание.

Список источников

1. Баронов В.В., Калянов Г.Н., Попов Ю.Н., Титовский И.Н. *Информационные технологии и управление предприятием*. Москва, Компания АйТи, 2004.
2. Зубец В.В. Разработка базы данных как универсальный проект // *Вестник Тамбовского университета. Серия: Естественные и технические науки*, 2013, no. 1.
3. Козырев А.А. *Информационные технологии в экономике и управлении*. Санкт-Петербург, Изд-во Михайлова В.А., 2000.
4. Коннолли Т., Бегг К. *Базы данных. Проектирование, реализация и сопровождение. Теория и практика*. Москва, Вильямс, 2003.
5. Малыхина М.П., Буянов М.В. Интеллектуальная система для проектирования баз данных эффективной структуры // *Научный журнал КубГАУ*, 2013, no. 89.
6. Новиков Б.А., Домбровская Г.Р. *Настройка приложений баз данных*. Санкт-Петербург, БХВ-Петербург, 2012.
7. Петрянин Д.Л., Горячев Н.В., Юрков Н.К. Анализ систем защиты информации в базах данных // *Ника*, 2015, т.2.
8. Попов Ф.А., Максимов А.В. Подходы к проектированию баз данных для

автоматизированных систем // *Известия АлтГУ*, 2003, no. 1.

9. Рыженко Л.И. Подход к смысловой организации информационных баз данных // *Вестник ОмГУ*, 2010, no. 4.

10. Утебов Д.Р., Белов С.В. Классификация угроз в системах управления базами данных // *Вестник Астраханского*

государственного технического университета, 2008, no. 1(42), с. 87-92.

11. Хансен Г., Хансен Д. *Базы данных: разработка и управление*. Москва, БИНОМ, 2001.

12. Чернышев А.Н. Методы сжатия баз данных // *Математика и информационные технологии в нефтегазовом комплексе*, 2015, no. 2.

SOLVING THE PROBLEMS OF ENSURING THE INTEGRITY OF DATA

Orlova Marina Viacheslavovna, Cand. Sc. (Econ.), Assoc. Prof.

Voronezh State University, University sq., 1, Voronezh, Russia, 394018;

e-mail: orlova-it@mail.ru

Purpose: ensuring data integrity during the operation of application information systems based on database technologies are considered.

Discussion: application of database technology in the creation of information systems with complex organizational, functional and information structures that use the information model of interrelated economic entities in conjunction with relevant, complete and reliable data on their functioning allows information to be provided by the process of developing algorithms implemented the necessary economic-mathematical methods, as well as more simple information retrieval functions based on the system of requests and reports. At the same time, the ultimate goal is the prompt presentation of accurate data in full to all specialists of the enterprise related to data processing. The complexity of the technologies in question requires the creation of special tools that will ensure the integrity and non-inconsistency of information during the operation of the information system. *Results:* we proposed some recommendations to include universal integrity constraints in the data processing system, ensuring consistency, completeness of data and reducing the complexity and cost of developing and operating information and economic systems.

Keywords: databases, automated information systems, consistency of data.

References

1. Baronov V.V., Kaljanov G.N., Popov Ju.N., Titovskij I.N. *Informacionnye tehnologii i upravlenie predpriatijem*. Moscow, Kompanija AjTi, 2004. (In Russ.)
2. Zubec V.V. Razrabotka bazy dannyh kak universal'nyj proekt. *Vestnik Tambovskogo universiteta. Serija: Estestvennye i tehniczeskie nauki*, 2013, no. 1. (In Russ.)
3. Kozyrev A.A. *Informacionnye tehnologii v jekonomike i upravlenii*. Sankt-Peterburg, Izd-vo Mihajlova V. A., 2000. (In Russ.)
4. Konnolli T., Begg K. *Bazy dannyh. Proektirovanie, realizacija i soprovozhdenie. Teorija i praktika*. Moscow, Vil'jams, 2003. (In Russ.)
5. Malyhina M.P., Bujanov M.V. Intellektual'naja sistema dlja proektirovanija baz dannyh jeffektivnoj struktury. *Nauchnyj zhurnal KubGAU*, 2013, no. 89. (In Russ.)
6. Novikov B.A., Dombrovskaja G.R. *Nastrojka prilozhenij baz dannyh*. Sankt-Peterburg, BHV-Peterburg, 2012. (In Russ.)
7. Petrjanin D.L., Gorjachev N.V., Jurkov N.K. Analiz sistem zashhity informacii v bazah dannyh. *NiKa*, 2015, vol.2. (In Russ.)
8. Popov F.A., Maksimov A.V. Podhody k proektirovaniju baz dannyh dlja avtomatizirovannyh system. *Izvestija AltGU*, 2003, no. 1. (In Russ.)
9. Ryzhenko L.I. Podhod k smyslovoj

organizacii informacionnyh baz dannyh. *Vestnik OmGU*, 2010, no. 4. (In Russ.)

10. Utebov D.R., Belov S.V. Klassifikacija ugroz v sistemah upravlenija bazami dannyh. *Vestnik Astrahanskogo gosudarstvennogo tehničeskogo universiteta*, 2008, no. 1(42), pp. 87-92. (In Russ.)

11. Hansen G., Hansen D. *Bazy dannyh: razrabotka i upravlenie*. Moscow, BINOM, 2001. (In Russ.)

12. Chernyshev A.N. Metody szhatija baz dannyh. *Matematika i informacionnye tehnologii v neftegazovom komplekse*, 2015, no. 2.