

Карпук Анатолий Алексеевич, к.т.н., доцент,

Расчетный центр Национального банка Республики Беларусь, Anatoly_Karpuk@bisc.by

Аннотация: изложена разработанная автором методология проектирования баз данных сложных систем, основанная на построении канонической модели предметной области сложной системы и ее отображении в логическую и физическую структуру базы данных.

Ключевые слова: проектирование баз данных, информационно–логическая модель предметной области, логическая структура базы данных, физическая структура базы данных

При проектировании баз данных (БД) сложных автоматизированных информационных систем (АИС) до настоящего времени применяется методология IDEF1X, разработанная по заказу департамента Военно–воздушных сил США в 1981 г., расширенная компанией DACOM и в окончательном виде опубликованная в конце 1985 г. [1]. В основе методологии IDEF1 лежат расширенная реляционная модель данных Э. Кодда и модель «сущность–связь» П. Чена. Специалисты, применяющие методологию IDEF1X для построения информационно–логической (инфологической) модели предметной области, должны хорошо знать предметную область сложной системы и методы проектирования реляционных баз данных. Они должны уметь выделить в предметной области сущности линейной структуры, собственные либо наследуемые атрибуты сущностей с непустыми неделимыми значениями, ключи сущностей и отношения между сущностями. Хотя понятие нормализации в методологии IDEF1X прямо не используется, правила выделения сущностей и отношений таковы, что выделенные сущности должны находиться в третьей нормальной форме. В реальных предметных областях сложных систем сущности имеют иерархическую структуру, в качестве значений атрибутов могут выступать векторы и повторяющиеся группы, значения некоторых атрибутов могут быть пустыми. Более того, в реальных предметных областях можно обнаружить сущности, состав атрибутов которых может изменяться в процессе функционирования АИС. Все это создает значительные трудности разработчикам БД, использующим методологию IDEF1X. Эти трудности становятся практически непреодолимыми, если нет специалиста, знающего предметную область всей сложной системы, а есть несколько специалистов, каждый из которых знает свой фрагмент предметной области АИС.

В работе представлена разработанная автором методология проектирования БД сложных АИС, возможности которой во многом превосходят возможности методологии IDEF1X, и при использовании которой отсутствуют перечисленные выше трудности. Методология была успешно апробирована при проектировании БД ряда сложных систем, в частности, при проектировании БД архива межбанковских расчетов.

Классическая концепция БД в соответствии со стандартом ANSI/SPARC предусматривает три уровня описания данных: внешний, концептуальный и внутренний. На каждом уровне используется соответствующая модель данных и схема БД. Для сложных предметных областей практически невозможно сразу построить концептуальную и внешние схемы БД, поэтому в современных методологиях и CASE –средствах проектирования БД концептуальный уровень описания данных разбит на два: инфологический и даталогический. На инфологическом уровне БД описывается в виде, не зависящем от используемой СУБД, а на даталогическом уровне БД описывается на языке описания данных конкретной СУБД (рисунок 1).



Рисунок 1 – Традиционная методология проектирования БД

Инфо-логическая модель данных, применяемая для описания предметной области сложной АИС, должна удовлетворять следующим требованиям:

- 1) средствами модели должны описываться все объекты, предметы, явления предметной области и все отношения между ними, используемые при решении прикладных задач (ПЗ) в АИС;
- 2) средствами модели должны описываться все ограничения целостности данных, имеющиеся в предметной области;
- 3) средствами модели должны задаваться все объемные характеристики данных, необходимые для проектирования логической и физической структуры БД;
- 4) базис понятий модели должен быть максимально приближенным к базису понятий постановок и алгоритмов ПЗ, а также к лексике специалистов по предметной области;
- 5) должны существовать методы объединения описаний фрагментов предметной области, соответствующих отдельным ПЗ, в глобальное описание всей предметной области;
- 6) средствами модели должен описываться процесс решения каждой ПЗ с указанием, какие операции над какими данными и с какой частотой будут выполняться;
- 7) должны существовать методы и методики отображения глобального описания предметной области в логические, внешние и физические модели данных применяемых СУБД.

Несмотря на многообразие существующих моделей данных, практически невозможно построить инфологическую модель данных для описания предметной области сложной АИС, удовлетворяющую всем перечисленным требованиям. В работе [2] было предложено выделить среди инфологических моделей данных одну модель, названную канонической и удовлетворяющую следующим условиям:

- для каждой инфологической модели данных, используемой для описания предметной области ПЗ, существует отображение в каноническую модель данных;
- для канонической модели данных существует эффективный метод объединения канонических моделей фрагментов предметной области;
- существует отображение канонической модели предметной области в логическую, внешние и физическую структуры БД для выбранной СУБД.

Таким образом, для построения инфологической модели предметной области сложной АИС будем использовать две модели данных. Для построения инфологической модели фрагментов предметной области для ПЗ (одной или нескольких) будем использовать инфологическую модель данных, удовлетворяющих требованиям 1–4, а для построения канонической модели фрагментов предметной области и глобальной канонической модели предметной области всей АИС будем использовать каноническую модель данных, удовлетворяющей требованиям 5–7 (рисунок 2).

Множество допустимых структурных компонентов инфологической модели данных имеет вид $\{A, F, NR, Q, NQ\}$, где A – множество атрибутов, которые могут быть простыми или составными; F – множество функциональных зависимостей (ФЗ) между атрибутами, существующих в предметной области; NR – множество нефункциональных отношений (НО) между атрибутами, существующих в предметной области; Q – множество типов объектов, причем для каждого типа объектов задаются имена атрибутов основного ключа, имена обязательных и возможных атрибутов, каждый тип объектов может иметь элементарную, простую или сложную иерархическую структуру; NQ – множество отношений между типами объектов, которые определяются на атрибутах их основных ключей, либо сводятся к НО между атрибутами типов объектов.



Рисунок 2 – Предлагаемая методология проектирования БД

Множество допустимых структурных компонентов канонической модели данных имеет вид $\{D, A, R, RE\}$, где D – множество доменов; A – множество атрибутов; R – множество отношений в четвертой нормальной форме (4НФ); RE – множество ФЗ между отношениями. Каждая ФЗ между отношениями классифицируется как полная, частичная сильная, частичная слабая или частичная слабая с обратной связью.

Алгоритм отображения инфологической модели фрагмента предметной области в каноническую модель этого фрагмента предметной области состоит из следующих шагов: выделение отношений между атрибутами и построение системы образующих структуры ФЗ между атрибутами; построение минимального элементарного базиса структуры ФЗ между атрибутами [3]; синтез отношений в третьей нормальной форме [4]; приведение отношений к 4НФ и выделение возможных атрибутов; выделение и классификация ФЗ между отношениями; удаление транзитивных ФЗ между отношениями.

Алгоритм объединения канонических моделей фрагментов предметной области в каноническую модель предметной области всей системы состоит из следующих шагов:

анализ атрибутов канонической модели фрагмента предметной области, по результатам которого новые атрибуты добавляются в каноническую модель предметной области АИС и определяются атрибуты, имеющие аналог в канонической модели предметной области АИС;

анализ ФЗ между атрибутами канонической модели фрагмента предметной области, по результатам которого новые ФЗ добавляются в каноническую модель предметной области АИС, определяются аналоги ФЗ в канонической модели предметной области АИС, производится замена транзитивной ФЗ на цепочку ФЗ или удаление ФЗ из канонической модели фрагмента предметной области;

анализ нефункциональных отношений между атрибутами канонической модели фрагмента предметной области, по результатам которого новые НО добавляются в каноническую модель предметной области АИС и определяются НО, имеющие аналог в канонической модели предметной области АИС;

уточнение структуры ФЗ между отношениями канонической модели предметной области всей АИС.

Для отображения канонической модели предметной области АИС в логическую структуру БД для выбранной СУБД, обеспечивающего заданное максимальное и среднее время решения ПЗ при наличии минимальной избыточности данных, применяется итерационный процесс, состоящий из следующих шагов:

1) строится начальная логическая структура БД, обеспечивающая минимизацию объемов памяти для хранения данных для любого фиксированного метода физической организации БД;

2) для имеющейся логической структуры БД варьируются методы физической организации БД и выбираются те из них, которые обеспечивают заданное максимальное и среднее время решения ПЗ с минимальными затратами памяти для хранения данных;

3) если не удалось найти методов физической организации БД, удовлетворяющих условиям шага 2, то производится модификация имеющейся логической структуры БД, и снова выполняется шаг 2.

Выбор физической структуры БД производится на основе унифицированной математической модели задачи проектирования физической структуры БД для любых СУБД [5]. В этой модели критерий производительности представляется целевой функцией и строится система ограничений, учитывающих максимальные допустимые временные и объемные характеристики, а также отражающих связи между параметрами структуры хранения.

Процесс построения физической структуры БД, обеспечивающей заданное максимальное и среднее время решения ПЗ при наличии минимальной избыточности данных, можно представить как совокупность процессов субоптимизации по локальным критериям оптимальности с использованием эвристических методов. В качестве начальной выбирается физическая структура БД с минимальными затратами внешней памяти для хранения данных. Производится оценка максимального и среднего времени работы СУБД при решении каждой ПЗ. Если для некоторой совокупности ПЗ получены неудовлетворительные оценки, то выделяется множество физических файлов, используемых этими ПЗ, и строится множество допустимых модификаций каждого файла, не увеличивающих максимальное и среднее время решения остальных ПЗ. Выбираются допустимые модификации, дающие удовлетворительные оценки максимального и среднего времени решения рассматриваемых ПЗ с минимальными дополнительными затратами внешней памяти.

Список использованных источников:

1. Integrated Information Support System. Information Modeling Manual IDEF1 –Extended (IDEF1X) / D. Appleton Company Inc. – Manhattan Beach, California, December 31, 1985. – 108 p.

2. Карпук, А. А. Информационное моделирование предметной области автоматизированных информационных систем / А. А. Карпук // Технологии информатизации и управления: сб. научн. ст., вып. 2 / Белорус. гос. ун –т; редкол.: А. М. Кадан (отв. ред.) [и др.]. – Минск, 2011. – С. 24–30.

3. Карпук, А.А. О построении элементарного базиса системы функциональных зависимостей в базе данных / А. А. Карпук // Информационные технологии и программные средства: проектирование, разработка и применение: сб. научн. ст. / Гродн. гос. ун –т; редкол.: Л. В. Рудикова (отв. ред.) [и др.]. – Гродно, 2011. – С. 185–190.

4. Карпук, А. А. Об алгоритмах нормализации отношений в реляционных базах данных / А. А. Карпук // Международный конгресс по информатике: информационные системы и технологии: материалы междунар. науч. конгр., Минск, 31 окт.–3 нояб. 2011 г.: в 2 ч. / Белорус. гос. ун –т; редкол.: С. В. Абламейко (отв. ред.) [и др.]. – Минск, 2011. – Ч. 2. – С. 283–288.

5. Карпук, А.А. Математическое моделирование выбора физической организации баз данных / А. А.

Карпук // Устойчивое развитие экономики: состояние, проблемы, перспективы: сб. трудов VI междунар. науч. –практ. конф., г. Пинск, 26–27 апреля 2012 г. / Полесский гос. ун –т; редкол.: К. К. Шебеко [и др.]. – Пинск, 2012. – С. 202–204.