

УДК 681.3

И.А. Телков, А.В. Бакулев, М.А. Бакулева

АЛГОРИТМ АВТОМАТИЗАЦИИ ПРОЕКТИРОВАНИЯ ХРАНИЛИЩ ДАННЫХ

Предлагается на рассмотрение алгоритм автоматизации проектирования хранилищ данных. Особенность алгоритма заключается в том, что его основу составляет тензорная модель исходной базы данных. Данное математическое представление позволяет упростить процедуру описания сложных взаимосвязей сущностей и их проекций в хранилище данных.

Введение. Одной из основных задач, решаемых в крупных информационных системах, является предоставление аналитической информации, необходимой для принятия решений. В процессе анализа данные, необходимые для принятия решений, должны поступать к потребителю в режиме реального времени. Если же данные собираются из разных источников, то, во-первых, отчет готовится недопустимо медленно, а, во-вторых, другие приложения, работающие с этими же реляционными системами управления базами данных (СУБД) во время выполнения отчета, вероятнее всего, будут работать значительно медленнее [1]. Решением проблемы производительности является создание специализированной базы данных (БД) — хранилища данных (ХД), предназначенной исключительно для обработки и анализа информации.

Актуальность. Переход от реляционной БД к ХД в системах масштаба предприятия должен происходить логично, грамотно и в приемлемые сроки, что невозможно без автоматизации этого процесса и моделирования соответствующих объектов (БД и ХД) единым непротиворечивым математическим представлением. Отсутствие интегрированных математических моделей приводит к значительным осложнениям при проектировании [4]. В связи с этим работа по созданию модели ХД, объединяющей все этапы его проектирования и создание алгоритмов автоматизации на основе этой модели, является актуальной и востребованной.

Описание математической модели. Для решения поставленных задач было решено использовать аппарат тензорной алгебры [3]. Как правило, большинство тензоров описываются с помощью многомерных массивов [2], в то время как это же представление имеют таблицы БД

(ТБД). Поэтому сущности исходной БД можно описать при помощи тензоров.

Рассмотрим пример. Пусть имеется сущность «Списание товара», которой соответствует ТБД (рисунок 1), где каждому атрибуту сущности присвоен произвольный символ.

№ товара	j
№ накладной	l
Количество	k

Рисунок 1 – ТБД «Списание товара»

Приведенная ТБД описывается следующей матрицей:

$$\Pi = \left\| k_{ij} \right\| = \begin{bmatrix} k_{11} & k_{12} & k_{13} & \dots \\ k_{21} & k_{22} & k_{23} & \dots \\ k_{31} & k_{32} & k_{33} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

или тензором Π_{jl}^k [5].

Составим тензорную модель для БД, представленную на рисунке 2, где переменные атрибутов приведены рядом с соответствующей ТБД. [1]

Сущности «Сотрудники» будет соответствовать тензор $F_r^{nx} = F_r^n * F_r^x$ [5], сущности «Накладные-расход» - тензор N_l^{drb} , «Продажа»- P_{jl}^k , «Товар»- T_j^{acv} , «Категория товара-склад»- W_v^{hs} , «Покупатели»- Q_b^{fy} , «Поставщики»- L_p^{mz} , «Накладные-приход»- M_g^{pp} , «Поставка» — E_{gj}^i .

Связь между сущностями БД, которая в реляционной модели осуществляется через первичные ключи, в данном случае будет определяться наличием у тензоров индексов.

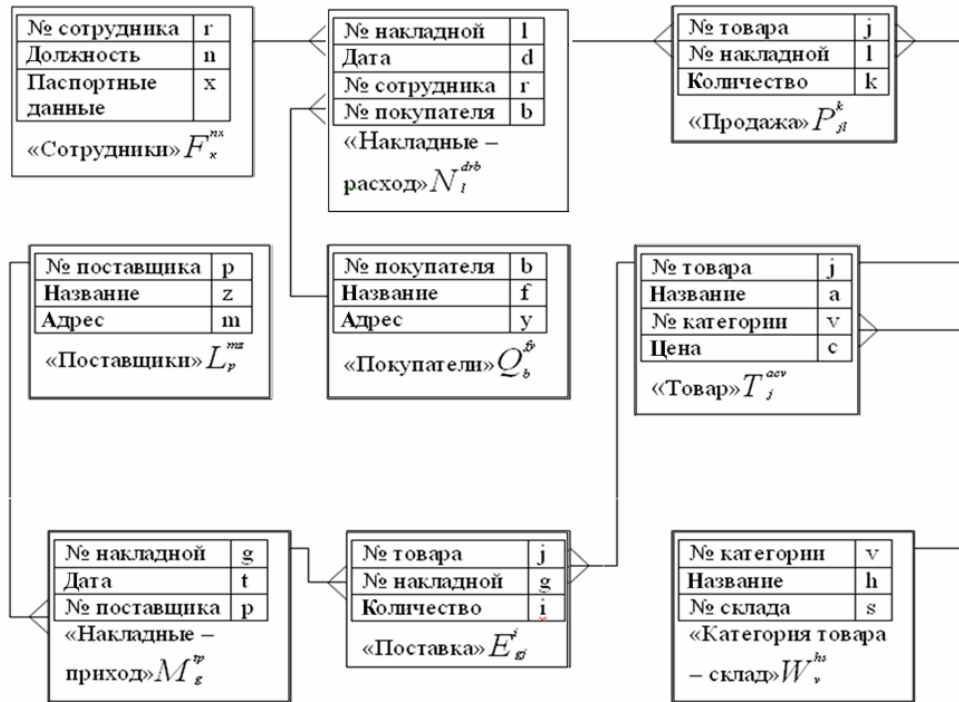


Рисунок 2 – «Структура БД»

Алгоритм автоматизации. Представленное математическое описание позволяет облегчить и ускорить процесс проектирования ХД. Для автоматизации этого процесса необходимо разработать алгоритм, основанный на предложенном математическом обеспечении.

Результатом работы алгоритма должны быть возможные структуры ХД, получаемые из исходной БД. Таким образом, исходные данные представляют собой структуру БД и атрибуты, соответствующие количественным показателям процессов, отраженных в БД. Исходная БД задается тензорной моделью, атрибуты – последовательностью контравариантных (верхних) индексов.

Вход:

$\|O_u\| = [F_r^{rx}, N_l^{drb}, P_{jl}^k, T_j^{acv}, W_v^{hs}, Q_b^y, L_p^{mz}, M_g^{fp}, E_{gj}^i]$ – массив тензоров, описывающих сущности БД; k, c, i – индексы атрибутов, отображающих количественные показатели, которые могут быть проанализированы.

Выход: тензорное описание возможных структур ХД.

Шаг 1. В массиве $\|O_u\|$ находим тензоры с контравариантными индексами k, c или i .

Найденные тензоры помещаем в массив $\|O_{u_1}\|$. С учетом входных данных

$$\|O_{u_1}^1\| = [P_{jl}^k, T_j^{acv}, E_{gj}^i].$$

Шаг 2. В массиве $\|O_{u_1}^1\|$ сравниваем ковариантные (нижние) индексы тензоров; если два и больше тензоров содержат хотя бы один одноименный ковариантный индекс, то атрибуты, обозначенные контравариантными индексами этих тензоров, будут входить в одну таблицу фактов в качестве мер.

В $\|O_{u_1}^1\|$ все тензоры содержат одноименный ковариантный индекс j , следовательно, в проектируемом ХД будет одна таблица фактов с мерами k, c, i . На данном этапе таблица фактов описывается тензором A_j^{kci} .

Шаг 3. Формирование таблицы фактов схемы «звезда»:

- 1) $u_1 = 0$;
- 2) $u_1 = u_1 + 1$;
- 3) $\|O_{u_2}^2\| = \|O_u\| \text{ minus } \|O_{u_1}^1\|$;
- 4) последовательно сравниваем $O_{u_1}^1$ с $\|O_{u_2}^2\|$.

Находим тензоры, имеющие одинаковые индексы с u_1 членом массива $\|O_{u_1}^1\|$, и помещаем их в массив $\|O_{u_3}^3\|$;

$$5) \|O_u\| = \|O_u\| \text{ minus } \|O_{u_3}^3\|;$$

6) проверка окончания цикла. Если рассмотрены все члены массива $\|O_{u_1}^1\|$, то переход к шагу 4, иначе — переход к 2).

Шаг 4. Составляется тензорное описание таблицы фактов схемы «звезда» проектируемого ХД на основе данных массива $\|O_{u_3}^3\|$. В данном случае массив $\|O_{u_3}^3\|$ примет вид: $\|O_{u_3}^3\| = [N_l^{drb}, W_v^{hs}, M_g^{tp}]$. Все ковариантные индексы тензоров из массива $\|O_{u_3}^3\|$ включаются в тензор A_j^{kci} , таким образом, получаем новое тензорное описание таблицы фактов: A_{lvg}^{kci} . Если в массиве $\|O_{u_2}^2\|$ остался хотя бы один неучтенный индекс (как, например, контравариантный индекс a тензора T_j^{acv}), то его переписывают в тензор таблицы фактов на ту же позицию (то есть индекс a останется ковариантным). В итоге тензор, описывающий таблицу фактов, примет вид: A_{lvg}^{kcia} .

Шаг 5. Тензорное описание таблиц измерений будет совпадать с тензорами, содержащимися в массиве $\|O_{u_3}^3\|$. Их связь с таблицей фактов описывается следующей системой тензорных уравнений:

$$\begin{cases} e^j * (A_{lvg}^{kcia})^T * N_l^{drb} = e^{drb} \\ e^j * (A_{lvg}^{kcia})^T * W_v^{hs} = e^{hs} \\ e^j * (A_{lvg}^{kcia})^T * M_g^{tp} = e^{tp} \end{cases}$$

Шаг 6. Полученная структура ХД в тензорном представлении примет вид:

$$A_{lvg}^{kcia} \Rightarrow \begin{Bmatrix} N_l^{drb} \\ M_g^{tp} \\ W_v^{hs} \end{Bmatrix}$$

Примечание: знак « \Rightarrow » обозначает связь «один – ко – многим».

Шаг 7. Проверка окончания алгоритма. Если массив $\|O_u\|$ пустой, то переход к шагу 14, иначе — переход к шагу 8.

Шаг 8. Проектирование схемы «снежинка»:

1) $u_3 = 0$;

2) $u_3 = u_3 + 1$;

3) последовательно сравниваем $O_{u_3}^3$ с $\|O_{u_2}^2\|$.

Находим тензоры, имеющие одинаковые индексы с u_3 членом массива $\|O_{u_3}^3\|$, и помещаем их в массив $\|O_{u_4}^4\|$;

4) $\|O_{u_2}^2\| = \|O_{u_2}^2\| \text{ minus } \|O_{u_4}^4\|$;

5) $\|O_{u_3}^3\| = \|O_{u_3}^3\| - O_{u_3}^3$

6) проверка окончания цикла. Если массив $\|O_{u_3}^3\|$ пустой, то переход к шагу 9, иначе — переход к 2).

Шаг 9. Тензорное описание таблиц измерений схемы «снежинка» проектируемого ХД будет совпадать с тензорами, содержащимися в массиве $\|O_{u_4}^4\| = [F_r^{nx}, Q_b^{fy}, L_p^{mz}]$. Их связь с таблицами измерений схемы «звезда» описывается следующей системой тензорных уравнений:

$$\begin{cases} e^l * N_l^{drb} = e^r * e^b \\ e^r * F_r^{nx} = e^{nx} \\ e^b * Q_b^{fy} = e^{fy} \\ e^g * M_g^{tp} = e^{tp} \\ e^p * L_p^{mz} = e^{mz} \end{cases}$$

Шаг 10. Полученная структура ХД в тензорном представлении примет вид:

$$\begin{aligned} N_l^{drb} &\Rightarrow \begin{Bmatrix} F_r^{nx} \\ Q_b^{fy} \end{Bmatrix} \\ A_{lvg}^{kcia} &\Rightarrow M_g^{tp} \Rightarrow \begin{Bmatrix} L_p^{mz} \end{Bmatrix} \\ &W_v^{hs} \end{aligned}$$

Шаг 11. Проверка окончания алгоритма. Если массив $\|O_{u_2}^2\|$ пустой, то переход к шагу 14, иначе — переход к шагу 12.

Шаг 12. Переписываем данные из массива $\|O_{u_4}^4\|$ в пустой массив $\|O_{u_3}^3\|$.

Шаг 13. Переход к шагу 8.

Шаг 14. Конец.

Выводы. Программная реализация предложенного алгоритма позволяет значительно ускорить процесс проектирования ХД и избежать ошибок, связанных с выбором нерациональной структуры формируемого ХД.

Библиографический список

1. Маклаков С. Проектирование хранилищ данных//Компьютер пресс. – 2001. – № 1.
2. Корячко В.П., Скворцов С.В., Телков И.А. Архитектуры многопроцессорных систем и параллельные вычисления.: учеб. пособие. М.:Высш.шк. 1999. 235 с.
3. Арменский А.Е. Тензорные методы построения информационных систем. М.: Наука, 1989. 152 с.
4. Пржияковский В.В. Абстракции в проектировании БД. СУБД №1 /1998.
5. Бакулева М.А. Тензорная модель работы реляционной СУБД// Информационные технологии в проектировании. Межвуз. сб. научн. трудов. Рязань: РГРТА, 2004. С. 39-43.