

---

---

# МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ В ТЕОРИИ УПРАВЛЕНИЯ СЛОЖНЫХ ПРОЦЕССОВ

---

---

## МОДЕЛЬ ПРЕДСТАВЛЕНИЯ ДАННЫХ В РАСПРЕДЕЛЕННЫХ ИНФОРМАЦИОННЫХ ХРАНИЛИЩАХ

**А.Ю. Иванов, доктор технических наук, доцент.**

**В.С. Горшков.**

**Санкт-Петербургский университет ГПС МЧС России**

Рассмотрен многоуровневый подход к представлению данных в распределенных информационных хранилищах. В качестве базового структурного компонента модели принят многомерно-реляционный набор. Показаны особенности представления этого компонента на различных уровнях модели.

*Ключевые слова:* распределенное информационное хранилище, модель представления данных, многомерно-реляционный набор

## DATA MODEL IN DISTRIBUTED DATA WAREHOUSES

**A.Y. Ivanov; V.S. Gorshkov.**

**Saint-Petersburg university of State fire service of EMERCOM of Russia.**

The article discusses multilevel approach to data representation in distributed data warehouses. As a base structure component model a multidimensional relational set is received. Representation features of this component are shown at model different levels.

*Key words:* distributed data warehouse, data model, multidimensional relational set

Общий подход к представлению данных в распределенном информационном хранилище (РИХ) соответствует традиционному подходу, применяемому для классических баз данных (БД). При этом под моделью представления данных (моделью данных) понимают фиксированную систему понятий и правил, задающих спецификацию основного структурного элемента, порядок формирования из названных элементов самой базы данных и порядок манипулирования этими элементами с целью получения ответа на запрос, адресуемый базе данных.

Модель данных состоит из трех компонентов: структурного, манипуляционного и целостного. Структурный компонент признается основным и отражает видение данных пользователем. От характера структурного компонента зависит сущность остальных. Манипуляционный компонент определяет допустимые операции, выполняемые на соответствующей структуре данных. Целостный компонент содержит ограничения для контроля целостности данных.

В настоящее время наибольшее распространение получила реляционная модель данных, основным структурным элементом которой является реляционное отношение (в дальнейшем – отношение). Реляционную модель поддерживают практически все известные системы управления базами данных (СУБД).

Структурными составляющими реляционной модели данных выступают: атрибут, домен, кортеж, отношение [1].

Понятие *атрибута* соответствует некоторой характеристике объекта предметной области, сведения о котором подлежат учету в БД. Область возможных значений атрибута определяется как *домен*. При таком подходе отдельное значение  $a$ , которое может принимать в БД атрибут  $A$ , обязательно должно отвечать условию  $a \in \text{Dom}(A)$ , где  $\text{Dom}(A)$  – *домен* (область допустимых значений), атрибута  $A$ .

Понятие атрибута используется для построения понятий реляционного отношения и его кортежа.

Имеется совокупность  $n$  атрибутов  $\mathbf{R} = \langle A_1, A_2, \dots, A_n \rangle$  с соответствующими доменами  $\text{Dom}(A_1), \text{Dom}(A_2), \dots, \text{Dom}(A_n)$ . Совокупность  $\mathbf{R}$  называют *схемой отношения*. *Кортежем*  $r$  называется упорядоченная совокупность значений  $r = \langle a_1, a_2, \dots, a_n \rangle$  такая, что  $a_i \in \text{Dom}(A_i)$ ,  $i = 1, \dots, n$ . При этом значения атрибутов в кортеже соблюдают порядок следования атрибутов, задаваемый схемой  $\mathbf{R}$ .

*Реляционным отношением*  $R$  называется множество кортежей  $R = \{r_j\}$ , имеющих одинаковую схему  $\mathbf{R}$ . Поскольку  $R$  определяется как множество, в его составе не может быть одинаковых кортежей.

Реляционное представление данных, относящихся к некоторой предметной области, означает выявление в ней всех возможных атрибутов и формирование из них по некоторым правилам реляционных отношений. При этом один и тот же атрибут может повторяться (дублироваться) в различных отношениях. Совокупность реляционных отношений, на которые разбивается множество атрибутов, определяет *реляционную структуру* БД заданной предметной области.

Основными операциям манипулирования данными, характерными для реляционных отношений, являются: проекция, селекция и соединение. *Проекция* означает выборку из отношения значений только тех атрибутов, список которых задается в операции. *Селекция* означает выборку из отношения только тех кортежей, значения атрибутов которых удовлетворяют критериям, задаваемым в операции. *Соединение* позволяет из двух отношений, имеющих общие атрибуты, сформировать новый объект, конкатенируя кортежи исходных отношений, обладающие одинаковыми значениями общих атрибутов.

Информация в БД не должна быть противоречивой. Это достигается выполнением различных правил, называемых *ограничениями целостности*. Такие правила проявляются в различных формах.

*Контроль значений атрибутов отношения*. Здесь могут выполняться различные виды проверок, такие как проверка типа значения атрибута, проверка интервала значений атрибута, проверка значений одного атрибута по значению другого.

*Целостность ключа*. Означает требование, чтобы поля первичного ключа не были пустыми. Кроме того, значения первичного ключа в отношении не должны повторяться.

*Ссылочная целостность*. Требуется, чтобы значению внешнего ключа некоторого кортежа некоторого отношения обязательно соответствовало бы такое же значение первичного ключа другого отношения.

Информационные хранилища являются более сложными образованиями относительно традиционных баз данных. Этим определяются отличительные особенности модели данных. В первую очередь следует отметить, что *концептуальный уровень* представления данных в информационном хранилище предполагает использование многомерно-реляционной модели [2].

*Структурный компонент* модели данных РИХ на концептуальном уровне характеризуется тем, что ее базовым элементом является многомерно-реляционный набор (МРН).

Для многомерного отношения  $M^{(n)}$  степени  $n$  многомерно-реляционным набором  $S^{(n)}$  степени  $n$  является совокупность реляционных отношений  $S^{(n)} = \langle \{R_i^f\}, \{R_{1j}^d\}, \dots, \{R_{nj}^d\} \rangle$ , которая отображает многомерное отношение  $M^{(n)}$ .

В состав МРН  $S^{(n)}$  входят два типа реляционных отношений, называемых отношениями (таблицами) факторов  $(R_i^f)$  и отношениями (таблицами) измерений  $(R_{ij}^d)$ . Количество отношений факторов не регламентировано, но как минимум в МРН должна присутствовать одна такая таблица. Каждая из этих отношений имеет составной ключ, образованный из  $n$  простых внешних ключей отношений измерений. Если МРН содержит несколько отношений факторов, то каждое из этих отношений содержит одинаковый состав неключевых атрибутов  $\{A_i^f\}$ , являющихся атрибутами-факторами отображаемого многомерного отношения  $M^{(n)}$ .

Количество отношений измерений также строго не оговаривается, но они подразделяются по группам в соответствии со степенью МРН, определяющей число информационных измерений. Если элементы информационного измерения характеризуются двумя и более атрибутами, то для каждого измерения в МРН должна присутствовать как минимум одна соответствующая таблица. Если элементы информационного измерения имеют единственный атрибут, то соответствующее отношений измерений может отсутствовать.

Между отношениями факторов и измерений устанавливаются связи по принятому в реляционных базах данных принципу «общего атрибута» и имеют степень «1:1» (один к одному) или «1:М» (один ко многим). Между отношениями измерений разных информационных измерениям связи отсутствуют. При этом между таблицами одного измерения связи возможны.

*Пример структурного построения многомерно-реляционного набора.*

Для обоснования принимаемого решения на построение оптимальной группировки подчиненных сил и средств должностным лицам регионального центра МЧС требуется провести ретроспективный анализ результативности ликвидации последствий чрезвычайных ситуаций (ЧС) за некоторый предшествующий период, для чего требуется получить значения таких атрибутов, как «Численность личного состава» и «Период ликвидации» с учетом следующих информационных измерений:

1. «**Объект защиты**», содержащего атрибуты «Наименование объекта», «Тип объекта», «Координаты объекта», «Площадь объекта».
2. «**Чрезвычайная ситуация**» с атрибутами «Наименование ЧС», «Тип ЧС», «Масштаб ЧС».
3. «**Время**» – атрибуты «Дата», «Месяц», «Год».
4. «**Силы МЧС**» – атрибуты «Наименование подразделения», «Принадлежность», «Пункт дислокации».

Графическое представление многомерно-реляционного набора рассматриваемой предметной области представлено на рисунке.



Рис. Структура многомерно-реляционного набора

*Манипуляционный компонент* модели данных РИХ характеризуется следующим. Дуальность многомерно-реляционного набора как совокупности реляционных отношений, с одной стороны, и как многомерного отношения, с другой стороны, определяет, что к нему могут быть применены операции двух классов. К первому относятся классические операции над реляционными отношениями: проекция, селекция и соединение [3]. Вторым класс образуют специфические операции над многомерными данными: срез (вертикальный и горизонтальный), подъем и спуск.

*Замечание 1.* Относится к содержательной части операций, применяемых к МРН. Поскольку многомерно-реляционные СУБД в настоящее время отсутствуют, то ведение информационных хранилищ на основе МРН предполагается вести средствами традиционных систем управления. В связи с этим любая операция над МРН рассматривается как последовательность традиционных, операций над реляционными отношениями, приводящая к требуемому результату. Это так называемая эмуляция операций над МРН операциями реляционной алгебры [2].

*Замечание 2.* Касается специфики классических операций применительно к МРН. Обусловлено тем, что объектами манипулирования в традиционных БД операндами являются одиночные отношения. Поскольку при выполнении операций над МРН в качестве операндов выступают совокупности отношений, то реализация операций осуществляется в соответствии со следующей процедурой. На начальном этапе МРН преобразуется в универсальное отношение [1, 2]. Затем выполняется собственно операция, но не над МРН как таковым, а над полученным универсальным отношением. На заключительном этапе результат операции подвергается обратному преобразованию путем его декомпозиции к МРН.

Сущность операций над МРН может быть описана следующим образом.

*Проекция* предназначается для сокращения числа учитываемых в МРН атрибутов и выдачи таблиц только с теми атрибутами, которые интересуют пользователя.

*Селекция* предназначается для выборки из таблиц МРН кортежей, значения атрибутов которых удовлетворяют логическому условию. *Прямая селекция* соответствует случаю, когда по фиксированным значениям атрибутов-измерений необходимо отыскать значения атрибутов-факторов. *Обратная селекция* ставит противоположную цель: определить значения атрибутов-измерений по заданным в условии операции значениям атрибутов-факторов. *Смешанная селекция* имеет место при совместном вхождении в условие операции атрибутов-измерений и атрибутов-факторов.

*Соединение* МРН позволяет формировать новый набор на основе двух исходных, имеющих одно или несколько общих информационных измерений.

С помощью операции *среза* МРН возможно отображение значений атрибутов-факторов по фиксированным измерениям и значениям атрибутов-измерений. *Вертикальный срез* МРН фиксирует некоторые из имеющихся в наборе измерений и возвращает для них значения атрибутов-факторов, обобщенные по выпадающим из рассмотрения измерениям. *Горизонтальный срез* МРН выдает значения атрибутов-факторов для фиксированных значений одного или нескольких информационных измерений набора.

Операции *подъема* и *спуска* предназначены для отображения атрибутов-факторов на уровнях иерархии, соседних по отношению к заданному. Обе операции являются условными, содержащими элементарное условие поиска, и сохраняют степень набора. Операция *подъема* выдает атрибуты-факторы для вышестоящего уровня иерархии, а *спуска* – для нижестоящего.

*Целостный компонент* модели представления данных выражается в совокупности правил, называемых ограничениями целостности и обеспечивающих поддержание в непротиворечивом состоянии базовых элементов модели (МРН) и адекватность манипулирования ими при обработке запросов.

В зависимости от назначения все ограничения целостности можно разделить на три группы [2].

*Интенциональные* ограничения целостности (характеризуют структурные свойства

МРН).

1. Обязательность наличия первичного ключа во всех таблицах, образующих МРН.
2. Ссылочная целостность таблиц соседних уровней иерархии одного измерения.
3. Ссылочная целостность таблиц факторов и соответствующих им таблиц измерений.
4. Идентичность неключевых атрибутов в таблицах факторов МРН.

*Экстенциональные* ограничения целостности (относятся к вопросам заполнения таблиц данных МРН).

1. Уникальность значений первичного ключа.
2. Обязательность присутствия всех значений первичного ключа таблицы измерения в записях соседствующей с ней таблицы нижестоящего уровня иерархии того же измерения.
3. Обязательность присутствия всех значений первичного ключа таблицы измерения в значениях составного первичного ключа соответствующей ей таблицы факторов.
4. Отсутствие в таблицах факторов записей, в которых атрибуты-факторы принимают пустые значения (*null*-значения).

*Манипуляционные* ограничения целостности (определяют порядок выполнения операций манипулирования МРН).

1. Обязательность включения всех измерений в условие проекции МРН.
2. Обязательность выборки значений атрибутов нижестоящих уровней иерархии информационного измерения при селекции МРН.
3. Идентичность составов таблиц общего измерения в обоих МРН, являющихся операндами операции соединения двух МРН.

Поскольку распределенные информационные хранилища относятся к классу географически рассредоточенных систем, то по аналогии с распределенными базами данных необходимо введение в рассмотрение дополнительных уровней представления данных [1, 4].

**Фрагментный уровень** представления данных используется для формирования непесекающиеся подмножеств РИХ, называемых логическими фрагментами. В случае использования классической реляционной модели возможны следующие виды фрагментации отношений [5]:

- а) вертикальная;
- б) горизонтальная;
- в) смешанная.

*Вертикальный фрагмент* состоит из подмножества атрибутов исходного отношения и определяется с помощью операции *проекции* над исходным отношением  $F=R[A_1 \in A]$ , где  $A_1$  – атрибут (список атрибутов), включаемых во фрагмент  $F$ ;  $A$  – список атрибутов исходного отношения  $R$ .

*Горизонтальный фрагмент* формируется посредством определения предиката, с помощью которого осуществляется отбор кортежей из исходного отношения. Фрагменты выделяются из исходного отношения при помощи операции *селекции* следующим образом:

$$F=(R[A\theta v])\vee(R[A\theta B]),$$

где  $v$  – константа;  $A$  – некоторый атрибут (список атрибутов) отношения  $R$ ;  $B$  – атрибут (список атрибутов) отношения  $R$ , отличный от  $A$ ;  $\theta$  – одна из операций ( $>$ ,  $<$ ,  $=$ ,  $\geq$ ,  $\leq$ ,  $\neq$ ).

*Смешанный фрагмент* образуется посредством дополнительной вертикальной фрагментации созданных ранее горизонтальных фрагментов или путем вторичной горизонтальной фрагментации предварительно выделенных горизонтальных фрагментов. Этот тип фрагментации также выполняется с помощью операций селекции и проекции.

С переходом от одиночных реляционных отношений к многомерно-реляционным наборам сущность фрагментации сохраняется, однако уточняются механизмы ее реализации.

*Вертикальная фрагментация первого порядка* возможна применением к МРН операции вертикального среза

$$S2 = VSlice(S1|k_{slice}), k_{slice} = \{k_1, k_2, \dots, k_r\},$$

где  $S1$  – исходный МРН;  $S2$  – результирующий МРН;  $k_{slice}$  – вектор индексов  $r$  информационных измерений, по которым проводится операция. Степень получаемого при выполнении операции вертикального среза фрагмента (результирующего МРН) понижается на  $(n-r)$ , где  $n$  – степень исходного набора;  $r$  – количество фиксируемых измерений.

Для осуществления *вертикальной фрагментации второго порядка* необходимо задействовать операцию проекции МРН

$$S2 = Project(S1 | \mathbf{A}_\pi),$$

где  $S1$  – исходный МРН;  $S2$  – результирующий МРН;  $\mathbf{A}_\pi$  – список атрибутов исходного МРН, выступающий условием проекции.

В отличие от операции вертикального среза, которая понижает степень исходного набора, проецирование МРН не обязательно приводит к такому же результату. Если  $\mathbf{A}_\pi = \mathbf{A}_{\pi 1}^d \cup \mathbf{A}_{\pi 2}^d \cup \dots \cup \mathbf{A}_{\pi n}^d \cup \mathbf{A}_\pi^f$ , где  $\mathbf{A}_{\pi i}^d$  и  $\mathbf{A}_\pi^f$  – произвольные совокупности атрибутов  $i$ -го измерения и атрибутов-факторов МРН соответственно, степень результирующего набора не меняется относительно степени исходного МРН. Если же  $\mathbf{A}_\pi = \mathbf{A}_{\pi 1}^d \cup \mathbf{A}_{\pi 2}^d \cup \dots \cup \mathbf{A}_{\pi n}^d$ , исходный МРН вырождается и преобразуется в совокупность независимых реляционных отношений.

*Горизонтальная фрагментация первого порядка* реализуется операцией горизонтального среза, которая не меняет степень исходного набора

$$S2 = HSlice(S1 | P_\alpha(\mathbf{A}^d)).$$

где  $S1$  – исходный МРН;  $S2$  – результирующий МРН;  $P_\alpha(\mathbf{A}^d)$  – условие поиска на атрибутах измерений.

При выполнении *горизонтальной фрагментации второго порядка* требуется использовать операцию селекции

$$S2 = Select(S1 | P_\sigma(\mathbf{A}_\sigma)),$$

где  $S1$  – исходный МРН;  $S2$  – результирующий МРН,  $P_\sigma = P_\sigma(\mathbf{A}_\sigma)$  – условие селекции.

Отличием селекции от горизонтального среза является то, что при селектировании степень результирующего набора может понижаться в зависимости от условия, заданного в операции. Например, если в условии отсутствуют атрибуты некоторых таблиц измерений, в МРН остаются только те таблицы измерений и связанные с ними таблицы факторов, атрибуты которых входят в условие операции.

**На уровне распределения** копии логических фрагментов, называемые физическими фрагментами, размещаются в узлах сети. Известны три варианта (стратегии) размещения фрагментов данных: разделение; дублирование; комбинирование.

При *разделении* каждый физический фрагмент существует в единственном экземпляре и располагается в определенном при проектировании узле сети, который способен обеспечить наиболее эффективное использование данных. В случае *дублирования* каждый фрагмент многократно копируется и размещается в тех узлах сети, которые назначены для размещения данных. При *комбинированном* размещении предполагается размещение в узлах сети не всех физических фрагментов, а только их определенного набора, возможно индивидуального для каждого узла. Общность последнего варианта размещения состоит в том, что любая часть базы может быть дублирована произвольное число раз, при этом в каждом узле может храниться желаемое подмножество данных.

В любом случае представление данных на уровне распределения устанавливает отношение между узлами сети и логическими фрагментами распределенного информационного хранилища. Это описание может задаваться в виде матрицы соответствия  $X = \|x_{ij}\|$ , где  $x_{ij}$  – булева переменная, такая что

$$x_{ij} = \begin{cases} 1, & \text{если } j\text{-й фрагмент подлежит размещению в } i\text{-м узле сети;} \\ 0, & \text{если } j\text{-й фрагмент не подлежит размещению в } i\text{-м узле сети,} \end{cases}$$

$j = 1, \dots, N$ ;  $i = 1, \dots, M$ ;  $N$  – число логических фрагментов,  $M$  – число узлов.

На **физическом уровне** данные представляются в виде совокупности реляционных таблиц, хранящихся в каждом узле сети. При этом многомерные наборы, полученные при фрагментации исходного МРН, преобразуются к табличной форме представления. Это позволяет считать представление данных на физическом уровне соответствующим классическому реляционному подходу.

### **Литература**

1. Иванов А.Ю., Саенко И.Б. Основы построения и проектирования реляционных баз данных. СПб.: ВАС, 1997. 68 с.
2. Саенко И.Б. Теоретические основы многомерно-реляционного представления данных и их применение для построения баз данных АСУ связью. СПб.: ВУС, 2001. 176 с.
3. Грэй П. Логика, алгебра и базы данных / пер. с англ. Х.И. Килова, Г.Е. Минца; под ред. Г.В. Орловского, А.И. Слисенко. М.: Машиностроение, 1989. 386 с.
4. Иванов А.Ю. Мобильные распределенные базы данных автоматизированных информационно-управляющих систем МЧС России: монография / под ред. В.С. Артамонова СПб.: Санкт-Петербургский ун-т ГПС МЧС России, 2008. 152 с.
5. Голенищев Э.П., Клиненко И.В. Информационное обеспечение систем управления. Сер. Учебники и учебные пособия. Ростов н/Д: Феникс, 2003. 285 с.