

Введение в теорию БД

Новакова Н.Е.

Реляционная модель данных

- структурная часть;
- целостная часть;
- манипуляционная.

Структурная часть

- описывает, какие объекты рассматриваются **реляционной моделью**;
- **реляционной базой данных** называется база данных, состоящая из набора **отношений**;
- **схемой** реляционной базы данных называется набор заголовков отношений, входящих в базу данных

Целостная часть

- целостная часть описывает ограничения специального вида, которые должны выполняться для любых отношений в любых реляционных базах данных;
- различают целостность сущностей и целостность внешних ключей

Манипуляционная часть

- Манипуляционная часть описывает два эквивалентных способа манипулирования реляционными данными — реляционную алгебру и реляционное исчисление.
- С практической точки зрения, важным является вывод о реляционной полноте структурного языка запросов SQL в том или ином стандарте, который и реализует манипуляционную часть реляционной модели в реальных СУБД.

Структурная часть реляционной БД

- Основы реляционной модели данных были впервые изложены в статье Е. Кодда в 1970 г.;
- термин реляционное представление данных, введенный Коддом, происходит от латинского RELATIO, что означает "отношение" или "таблица«;
- Таблица является наиболее привычным и удобнымместилищем для хранения информации;
- Таблица имеет жестко оговоренное количество поименованных и упорядоченных столбцов (структуру), и может неограниченно расти по количеству строк.

Теоретические основы реляционных БД

- теория множеств и реляционная алгебра (алгебра отношений)
- фундаментальным понятием реляционной модели данных является теоретико-множественное определение понятия **отношения**
- Пусть даны N множеств D_1, D_2, \dots, D_N (не обязательно различных), называемых **доменами** отношения R . Тогда R есть отношение над этими множествами, если само множество R есть множество упорядоченных n -кортежей вида $\langle d_1, d_2, \dots, d_n \rangle$, где d_1 — элемент из D_1 , d_2 — элемент из D_2 , ..., d_n — элемент из D_N .
- Отношение R , определенное на множестве доменов D_1, D_2, \dots, D_N , имеет две части: заголовок и тело.

Заголовок отношения

- Заголовок отношения содержит фиксированное количество названий атрибутов (столбцов) отношения. Имена атрибутов должны быть уникальны в пределах отношения. Часто имена атрибутов отношения совпадают с именами соответствующих доменов.
- Заголовок статичен, он не меняется во время работы с базой данных.
- Если в отношении изменены, добавлены или удалены атрибуты, то в результате получим уже другое отношение (пусть даже с прежним именем).

Тело отношения

- Тело отношения представляет собой набор кортежей, т. е. подмножество декартового произведения доменов, что собственно и является отношением в математическом смысле слова.
- Тело отношения может модифицироваться во время работы с базой данных — кортежи могут изменяться, добавляться и
- удаляться.

Отношения

- Понятие отношения фактически лежит в основе теории реляционных баз данных.
- В практической реализации любой реляционной СУБД отношения являются математическим аналогом таблиц.

Сравнение терминологии

Реляционный термин	Соответствующий "табличный" термин, связанный с программной реализацией СУБД
Отношение	Таблица
Заголовок отношения	Заголовок таблицы
Тело отношения	Тело таблицы
Атрибут отношения	Наименование столбца (поля) таблицы
Кортеж отношения	Строка (запись) таблицы
Степень отношения	Количество столбцов таблицы
Мощность (кардинальность) отношения	Количество строк таблицы
Домен	Базовый или пользовательский тип данных

Свойства отношений

- Все элементы отношения есть однотипные **кортежи**
- За исключением крайнего случая отношение включает в себя не все возможные кортежи из декартового произведения доменов, на которых оно определено.
- Это значит, что для каждого отношения имеется *критерий*, позволяющий определить, какие кортежи входят в отношение, а какие нет.
- Этот критерий определяет смысл, или семантику, отношения. Он является логическим выражением и называется *предикатом отношения R* .

Свойства отношений (продолжения)

- В отношении нет одинаковых кортежей, и кортежи не упорядочены сверху вниз, т. к. само отношение есть множество кортежей, и, следовательно, неупорядочено.
- Кроме того, как всякое множество отношение не может содержать неразличимые элементы. Поэтому одно и то же отношение может быть изображено разными таблицами, в которых строки идут в различном порядке.

Целостная часть реляционной БД

- Домены;
- Целостность сущностей.

Домены

- домен — это семантическое понятие, которое несет определенную смысловую нагрузку.
- Домен имеет уникальное имя в пределах базы данных, определен на простом типе данных
- Возможно наличие некоторого логического условия, позволяющего описать подмножество данных, допустимых для данного домена или на другом домене.

Домены (продолжение)

- для реляционной модели данных тип используемых данных не важен;
- Требование того, чтобы тип данных был простым, нужно понимать так, что в реляционных операциях не должна учитываться внутренняя структура данных;
- должны быть описаны действия, которые можно производить с данными как с единым целым.
- Основное назначение доменов состоит в том, что они ограничивают сравнения.

Целостность сущностей

- Под **сущностью** понимается некоторый объект, представляющий интерес для описания предметной области при создании модели базы данных.
- Понятие сущности возникает из методологии семантического моделирования данных.
- Финальной целью этого процесса является представление сущностей в виде **отношений**, поэтому для введения понятий, связанных с описанием целостной части реляционной базы данных, достаточно понимания сущности как синонима понятия отношения или таблицы.

Ключи

- **Ключ** (возможный ключ, потенциальный ключ) отношения — это минимальный набор атрибутов, который может быть использован для однозначной идентификации любого кортежа отношения.

Атомарные и составные ключи

- Ключи отношений бывают атомарными и составными
- *Атомарный ключ* состоит из единственного атрибута
- *Составной ключ* представляет собой набор атрибутов
- Ключ не должен содержать лишних атрибутов. Это означает, что если любой атрибут исключить из ключевого набора, то оставшихся атрибутов будет уже недостаточно для идентификации кортежа.

Потенциальные ключи

- Любое отношение имеет по крайней мере один потенциальный ключ
- Если никакой атрибут или группа атрибутов не являются потенциальным ключом, то в силу уникальности кортежей все атрибуты вместе образуют потенциальный ключ.
- Отношение может иметь несколько потенциальных ключей.

Первичный и альтернативный ключи

- один из потенциальных ключей объявляется **первичным ключом** (PK, Primary Key), а остальные — альтернативными
- различия между первичными и альтернативными ключами могут быть важны в конкретной реализации реляционной СУБД

Отношение

Табельный номер	Фамилия	Зарплата
1002	Иванов	1000
1003	Петров	2000
1004	Сидоров	3000

- Понятие потенциального ключа является *семантическим* и отражает введенную определенным образом трактовку понятий из конкретной предметной области.
- в этом отношении имеются три потенциальных ключа — в каждой колонке таблицы содержатся уникальные данные.
- Однако среди сотрудников могут быть однофамильцы и сотрудники с одинаковой зарплатой.
- Табельный же номер по сути своей уникален для каждого сотрудника. Именно понимание семантики данных привело к утверждению, что в данном отношении имеется только один потенциальный ключ — "Табельный номер".

Потенциальные ключи

- Потенциальные ключи служат единственным средством адресации на уровне кортежей в отношении.
- Только знание значения потенциального ключа кортежа позволяет точно указать этот кортеж
- С точки зрения *семантического моделирования данных*, потенциальные ключи служат средством идентификации объектов предметной области — экземпляров сущностей, данные о которых хранятся в отношении.
- Поскольку эти экземпляры должны быть различимы по определению, их идентификаторы не могут содержать неизвестные значения. воляет точно указать этот кортеж.

NULL-значения

- Обычно для ситуации наличия неизвестных или неполных данных используются типы данных, пополненные так называемым NULL-значением.
- NULL-значение — это, собственно, не значение, а некий маркер, показывающий, что значение неизвестно. За кажущейся естественностью такого подхода скрываются весьма неоднозначные проблемы, наиболее явной из которых является необходимость использования трехзначной логики при оперировании NULL-значениями.

Проблема использования NULL-значения

- Проблема использования NULL-значения в теории реляционных баз данных окончательно не решена.
- Практически все реализации современных реляционных СУБД позволяют использовать NULL-значения, несмотря на их недостаточную теоретическую обоснованность.
- Безусловно, следует избегать появления NULL-значений, обусловленных некорректным проектированием базы данных, но в реальной ситуации, когда какие-то данные об экземпляре сущности просто неизвестны, без NULL-значений не обойтись.

Правило целостности сущностей

- Каждая сущность, реализованная тем или иным отношением, должна иметь по крайней мере один потенциальный ключ, входящие в состав которого атрибуты не могут принимать NULL-значений.
- Этот потенциальный ключ лучше всего объявлять первичным ключом таблицы, соответствующей данной сущности.
- Следует отметить, что большинство СУБД вполне позволяют создавать таблицы и без первичных ключей. Однако нарушение правила целостности сущностей на практике сразу дает о себе знать.

Целостность внешних ключей

- Различные объекты предметной области, информация о которых хранится в базе данных, всегда взаимосвязаны.
- Наиболее типичный способ подобной связи между отношениями описывается ограничением **внешнего ключа** (FK, Foreign Key).
- **Определение.** Пусть дано отношение R . Подмножество атрибутов FK отношения R будем называть внешним ключом, если существует отношение S (S может совпадать с R) с потенциальным ключом K , и каждое значение FK в отношении R всегда совпадает со значением K для некоторого кортежа из S либо является NULL-значением.

Целостность внешних ключей (продолжение)

- Отношение S называется родительским отношением, а отношение R — дочерним отношением. Атрибут K часто называют вторичным ключом, а пару атрибутов K и FK — полями связи отношений S и R.
- Другими словами, для каждого отличного от NULL-значения поля, входящего во внешний ключ дочернего отношения, существует одно и только одно значение поля связи в родительском отношении.
- На уровне реализации СУБД поле связи в родительском отношении — это чаще всего его первичный ключ, а если нет, то оно должно быть объявлено альтернативным ключом отношения.

Внешний ключ

- Внешний ключ, так же как и потенциальный, может быть простым и составным.
- Внешний ключ должен быть определен на тех же доменах, что и соответствующий первичный ключ родительского отношения.
- Внешний ключ, как правило, не обладает свойством уникальности. Так и должно быть, поскольку в дочернем отношении может быть несколько кортежей, ссылающихся на один и тот же кортеж родительского отношения, что, собственно, и дает тип связи между отношениями "один-ко-многим".

Внешний ключ

- Это стандартный тип связи с сохранением ссылочной целостности. Если внешний ключ все-таки обладает свойством уникальности, то связь между отношениями имеет тип "один-к-одному".
- Хотя каждое значение внешнего ключа обязано совпадать со значениями потенциального ключа в некотором кортеже родительского отношения, обратное, вообще говоря, неверно. В поле связи родительской таблицы могут присутствовать значения, на которые не ссылается ни одно из значений внешнего ключа.

NULL-значения для атрибутов внешнего ключа

- NULL-значения для атрибутов внешнего ключа допустимы только в том случае, когда атрибуты внешнего ключа не входят в состав никакого потенциального ключа.
- Поскольку внешние ключи фактически служат ссылками на кортежи в другом (или в том же самом) отношении, то эти ссылки не должны указывать на несуществующие объекты.

Правило целостности внешних ключей или ссылочной целостности реляционной базы данных

- внешние ключи не должны быть несогласованными, т. е. для каждого значения внешнего ключа должно существовать соответствующее значение в поле связи в родительском отношении.

Манипуляционная часть реляционной базы данных

- Манипуляционная часть реляционной базы данных описывает средства, с помощью которых, во-первых, из данных, хранящихся в базе данных, можно получать какие-либо выборки или сводные результаты, а во-вторых, изменять сами эти данные или их структуру.
- Манипуляционной частью современных промышленных СУБД является специальный язык запросов, называемый SQL, который стал фактическим стандартом доступа к реляционным базам данных.

Операторы языка SQL

- Операторы этого языка по своей сути представляют собой требование к базе данных на выполнение каких-то действий.
- Язык SQL является реляционно полным. Это означает, что любой оператор реляционной алгебры может быть выражен подходящим оператором языка SQL
- Все СУБД, претендующие на название "реляционные", реализуют тот или иной диалект SQL.
- Многие нереляционные системы также имеют средства доступа к реляционным данным. Целью стандартизации является возможность переноса приложений между различными СУБД. Однако в настоящее время ни одна из систем не реализует стандарт SQL в полном объеме.