

Дисциплина: Базы данных

Тема курса: Манипулирование данными:
хранение, вставка, редактирование и удаление
данных, навигация по набору данных

Тема урока: Формирование: первая, вторая и
третья нормальные формы. Нормальная форма
Бойс-кода

Нормальные формы отношений

1. Нормализация отношений
2. Первая нормальная форма
3. Вторая нормальная форма
4. Третья нормальная форма
5. Нормальная форма Бойса-Кодда
6. Четвертая нормальная форма
7. Пятая нормальная форма

Нормализация отношений

Нормализация отношений информационной модели предметной области является механизмом создания логической модели реляционной базы данных.

С математической точки зрения задача построения как информационной модели предметной области, так и логической модели реляционной базы данных является результатом решения следующих комбинаторных задач:

- группировка атрибутов в отношении предметной области;
- распределение атрибутов по отношениям базы данных.

Нормализация отношений

Перечень наиболее важных требований при нормализации:

- Первичные ключи отношений должны быть минимальными (требование минимальности первичных ключей).
- Число отношений базы данных должно по возможности давать наименьшую избыточность данных (требование надежности данных).
- Число отношений базы данных не должно приводить к потере производительности системы (требование производительности системы).

Нормализация отношений

- Данные не должны быть противоречивыми, т.е. при выполнении операций включения, удаления и обновления данных их потенциальная противоречивость должна быть сведена к минимуму (требования непротиворечивости данных).
- Схема отношений базы данных должна быть устойчивой, способной адаптироваться к изменениям при ее расширении дополнительными атрибутами (требование гибкости структуры базы данных).

Нормализация отношений

- Разброс времени реакции на различные запросы к базе данных не должен быть большим (требование производительности системы).
- Данные должны правильно отражать состояние предметной области базы данных в каждый конкретный момент времени (требование актуальности данных).

Многие из требований находятся в противоречии друг к другу: Требование производительности находится в противоречии к требованию гибкости. Требование минимизировать число отношений в базе данных находится в противоречии к требованию надежности данных.

Нормализация отношений

В процессе эксплуатации *реляционной базы данных* происходит модификация данных и данные в базе данных с течением времени становятся *несогласованными* и *противоречивыми*.

Потенциальная противоречивость (или аномалии) при выполнении операций манипулирования данными в базе данных зависит как от типа ФЗ между атрибутами, так и от группировки этих атрибутов по отношениям.

Нормализация отношений

Пример: Аномалии операций над базой данных

ПОСТАВКИ (Поставщик, Адрес, Товар, Количество, Стоимость)

Обновление. Если Поставщик поставляет несколько видов товара, то значение атрибута Адрес повторяется для каждого кортежа и, следовательно, при изменении адреса поставщика необходимо изменить значение этого атрибута во всех этих кортежах. Иначе будет нарушена целостность данных базы данных.

Нормализация отношений

Удаление. Если Поставщик прекращает поставку товаров на некоторое время, то кортежи со всеми его поставками удаляются. При этом происходит потеря реквизитов поставщика - Адреса и Наименования.

Вставка. Если договор на поставку уже заключен, а поставка еще не произведена, то невозможно включить в рассматриваемое отношение значения атрибутов Поставщик и Адрес, поскольку нет сведений о поставках (проблема интерпретации нуль-значений).

Нормализация отношений

Теория функциональных зависимостей позволяет установить определенные требования к схемам отношений в реляционной базе данных.

Эти требования формулируются в терминах свойств отношений и называются *нормальными формами схем отношений*.

Каждая нормальная форма отношений связана с определенным классом ФЗ, которые представлены в отношениях.

Нормализация отношений

Теория ФЗ утверждает, что существует минимальное покрытие множества ФЗ базы данных, т.е. минимальный набор ФЗ, из которых можно вывести все остальные.

Каждой ФЗ из минимального покрытия можно отвести по самостоятельному отношению.

Однако,

во-первых, для заданного множества ФЗ может существовать несколько минимальных покрытий (выбор среди возможных альтернатив),

во-вторых, для практики важно иметь наглядный способ построения минимального покрытия.

Нормализация отношений

Процесс устранения потенциальной противоречивости и избыточности данных в отношениях реляционной базы данных называется нормализацией исходных схем отношений.

Нормализация отношений заключается в выполнении *декомпозиции* или *синтеза отношений*, назначении *ключей отношений* в соответствии с определенными правилами, гарантирующими целостность отношений базы данных.

Требование минимальности числа отношений, т.е. построения минимального покрытия ФЗ обычно является *опциональным*.

На минимальном покрытии, как правило, не может быть достигнута высокая производительность обработки запросов.

Нормализация отношений

Нормализация схем отношений дает наилучший результат при моделировании предметной области с использованием реляционной модели данных; при этом не вводится большого числа ограничений, не искажаются данные.

Таким образом, нормализация отношений является методом удаления из отношения ФЗ, которые приводят к аномалиям модификации данных.

Нормализация отношений помогает проектировать реляционную базу данных, которая не содержит избыточных данных и гарантирует их целостность.

Первая нормальная форма

Отношение находится в первой нормальной форме (1НФ), если все атрибуты отношения являются простыми (требование атомарности атрибутов в реляционной модели), т.е. не имеют компонентов.

Домен атрибута должен состоять из неделимых значений и не может включать в себя множество значений из более элементарных доменов.

Так, например, дата не может считаться простым атрибутом.

Первая нормальная форма

Пример: Приведение отношения к 1НФ

Отношение SHIPMENT (ОТГРУЗКА) является ненормализованным. Оно содержит повторяющиеся группы, представляющие массив значений, 1st Consignment, 2st Consignment, 3st Consignment (партии грузов).

Shipment
<u>Ship Registration Number</u>
<u>Departure Data</u>
Ship Name
Origin
Destination
1st Consignment
2st Consignment
3st Consignment
Customs Value
Ship Capacity
Customs Declaration

Для такого отношения следует ввести бизнес-правило, требующее, чтобы груз состоял не более чем из трех партий!

Первая нормальная форма

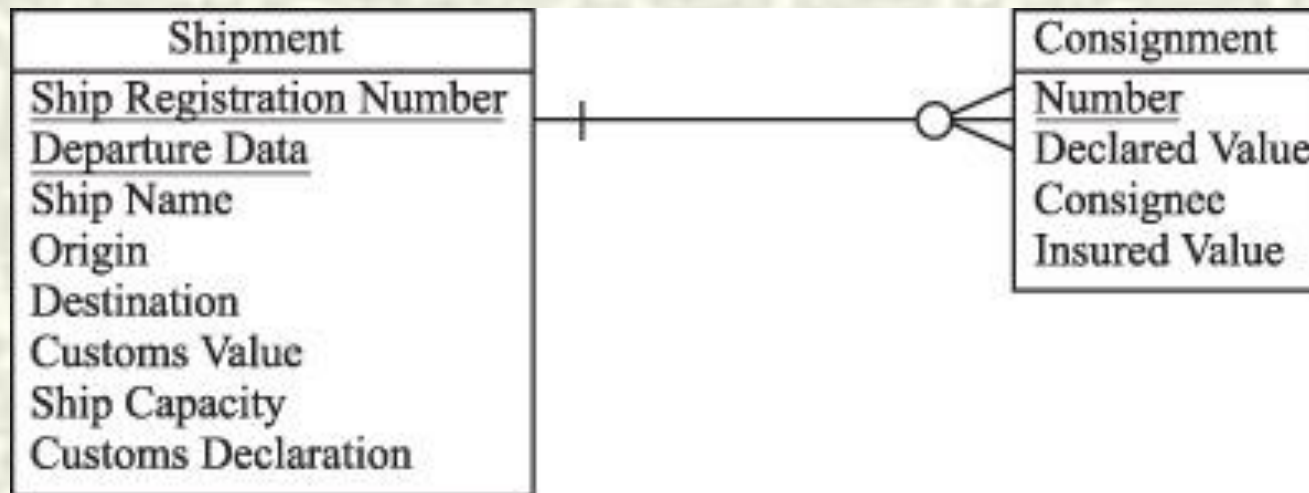
Использование отношения, представленного не в 1НФ, может породить следующие проблемы:

- если груз аннулируется и строка, связанная с грузом, удаляется из отношения, то вместе с ней удаляются все сведения о партиях груза на борту судна;
- если на склад прибывает новая партия груза, и она еще не включена в состав груза, подлежащего отправке, то сведения о партии заносить некуда;
- необходимо вводить ограничение: в грузе не может быть более трех партий.

Первая нормальная форма

Приведение отношения SHIPMENT к 1НФ заключается в изъятии данных о партиях груза из отношения SHIPMENT и создании для них связанного подчиненного отношения CONSIGNMENT (ПАРТИЯ_ГРУЗА).

Результат приведения отношения SHIPMENT к 1НФ :



Вторая нормальная форма

Атрибут отношения считается ключевым если он является элементом какого-либо ключа отношения.

В противном случае атрибут будет считаться *неключевым атрибутом*.

Так в отношении (Город, Адрес, Почтовый_индекс) все атрибуты являются ключевыми, поскольку при заданных ФЗ (*город, адрес*) → *почтовый_индекс* и *почтовый_индекс* → *город* ключами являются пары (*город, адрес*) и (*адрес, почтовый_индекс*).

Вторая нормальная форма

Отношение находится во второй нормальной форме (2НФ), если оно находится в 1НФ, и все неключевые атрибуты отношения функционально полно зависят от составного ключа отношения.

Иными словами, 2НФ требует, чтобы отношение не содержало частичных ФЗ.

Вторая нормальная форма

Пример: Приведение отношения ко 2НФ.

Отношение SHIPMENT содержит частичную ФЗ: неключевой атрибут Ship Capacity (грузоподъемность корабля) не зависит от ключевого атрибута Departure Date (даты убытия), а зависит от ключевого атрибута Ship Registration Number (регистрационный номер корабля).

Shipment
<u>Ship Registration Number</u>
<u>Departure Date</u>
Ship Name
Origin
Destination
1st Consignment
2st Consignment
3st Consignment
Customs Value
Ship Capacity
Customs Declaration

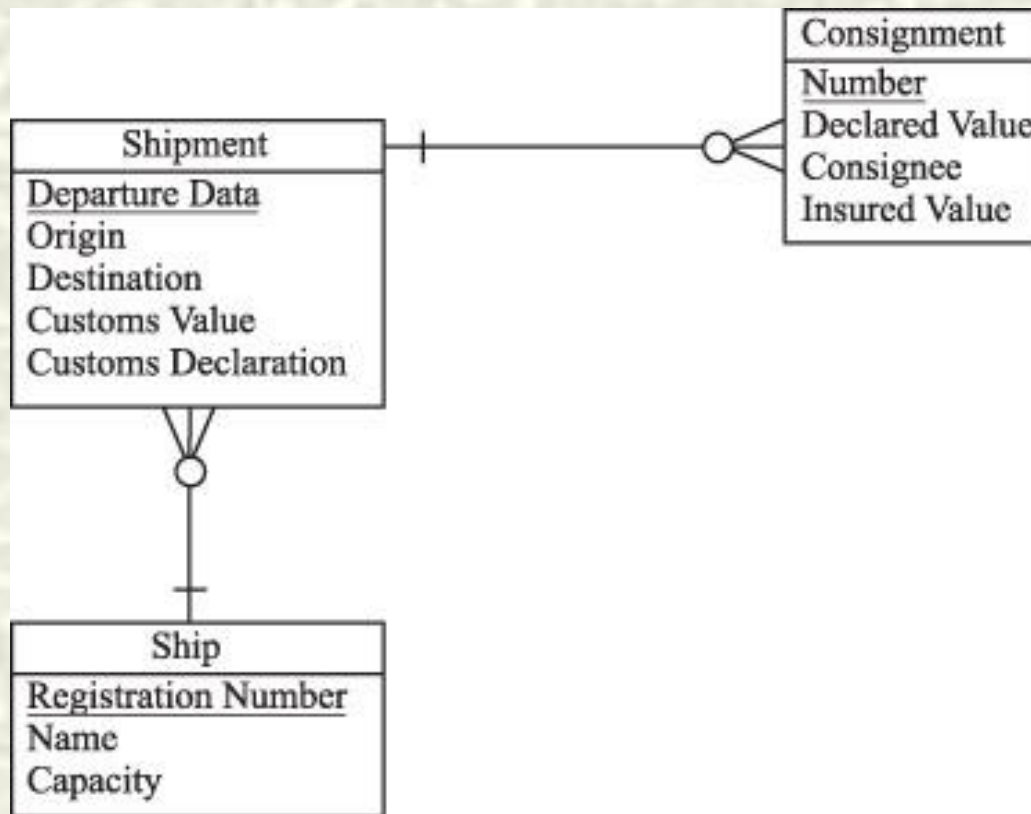
Вторая нормальная форма

Использование отношения, представленного не во 2НФ, может породить следующие проблемы:

- невозможно занести в базу данных название и грузоподъемность корабля, который не доставил еще ни одного груза, - можно только ввести для него фиктивный груз;
- если удалить кортеж из отношения Shipment после отправки груза, то потеряются все данные о кораблях, для которых в настоящее время нет груза;
- невозможно отразить факт переоборудования корабля и получения им новой грузоподъемности; если переписать все предыдущие кортежи об этом корабле, то получится, что он в прошлом плавал недогруженным или перегруженным.

Вторая нормальная форма

Приведение отношения SHIPMENT ко 2НФ заключается в изъятии атрибутов частичной ФЗ из отношения SHIPMENT и создании для нее связанного подчиненного отношения SHIP.



Третья нормальная форма

Отношение находится в третьей нормальной форме (3НФ), если оно находится во 2НФ, и все неключевые атрибуты отношения зависят только от первичного ключа.

Иными словами, 3НФ требует, чтобы отношение не содержало транзитивных ФЗ неключевых атрибутов от ключа.

Третья нормальная форма

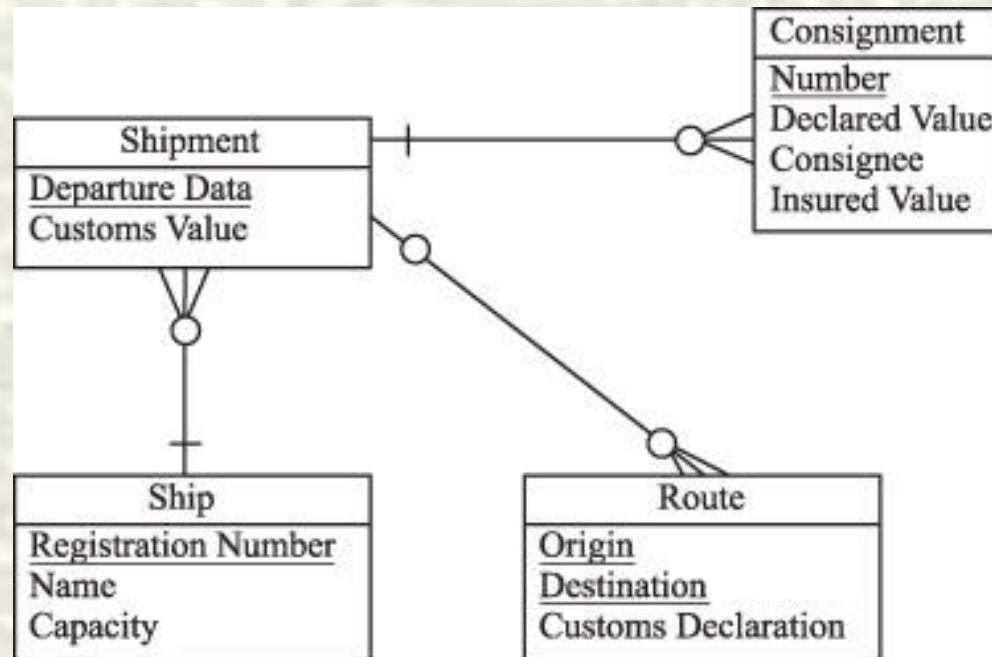
Возможна следующая проблема:

Наличие транзитивной зависимости $X \rightarrow Y \rightarrow A$ не позволяет связать значения Y и X , если не существует значения A , связанного со значением Y . Это затрудняет вставку и обновление данных, которые необходимо выполнить сразу для пары связей, а в случае удаления данных приводит к потере связи.

Третья нормальная форма

Пример: Приведение отношения к 3НФ

Отношение SHIPMENT содержит транзитивную ФЗ: атрибут Customs Declaration (таможенная декларация) является по своей сути свойством атрибутов Origin (пункт отправления) и Destination (пункт назначения).



Нормальная форма Бойса-Кодда

3НФ упрощает решение проблем контроля избыточности данных, интерпретации нуль-значений, контроля за операциями модификации данных, только если в отношениях отсутствуют какие-либо другие ФЗ, в частности *обратные ФЗ неключевого атрибута на один из атрибутов составного первичного ключа или многозначные ФЗ*.

В противном случае вышеперечисленные проблемы остаются неразрешенными. Для устранения таких проблем, связанных с существованием обратных ФЗ неключевых атрибутов на часть составного ключа, была предложена усиленная 3НФ или *НФ Бойса-Кодда*.

Нормальная форма Бойса-Кодда

Отношение находится в нормальной форме Бойса-Кодда (НФБК), если оно находится в ЗНФ, и в нем отсутствуют зависимости ключевых атрибутов от неключевых атрибутов.

Иными словами, НФБК допускает наличие только таких ФЗ, в которых ключ определяет один или более других атрибутов

Нормальная форма Бойса-Кодда

Схема отношения в НФБК обладает теми же достоинствами, что и схема в 3НФ, но устраняет некоторые дополнительные аномалии, не устраняемые 3НФ.

Пример: в отношении (Город, Адрес, Почтовый_индекс), находящееся в 3НФ, невозможно записать кортеж для города с известным почтовым индексом, если не известен адрес с этим почтовым индексом.

Данное отношение не находится в НФБК, так как имеет место ФЗ *Почтовый_индекс*→*Город*, а атрибут почтовый_индекс не является ключом этого отношения.

Нормальная форма Бойса-Кодда

Важное замечание: В отличие от ЗНФ, исходные отношения не всегда могут быть приведены в НФБК.

Пример: Приведение отношения к НФБК

Допустим, что экипаж судна разделен на команды, каждая из которых отвечает за разные виды работ. Члены экипажа могут входить в разные команды, но в каждую команду входит только один руководитель. Команда может иметь несколько руководителей. Каждый член экипажа может руководить только одной командой.

Нормальная форма Бойса-Кодда

Член экипажа	Отношение КОМАНДА	
	команда	Руководитель
Иванов	Наблюдение	Прохоров
Иванов	Питание	Макаров
Петров	Наблюдение	Леонтьев
Модин	Наблюдение	Прохоров
Васин	Питание	Лазарев
Фролов	Обслуживание	Сидоров
Ивлев	Обслуживание	Сидоров

Отношение находится в 3НФ, однако содержит аномалию удаления. Если Петрова удалить из команды наблюдения, то будет потеряна информация о том, что Леонтьев является руководителем команды наблюдения, и при назначении нового члена экипажа в команду наблюдения не будет известно, что у нее есть еще один руководитель, кроме Прохорова.

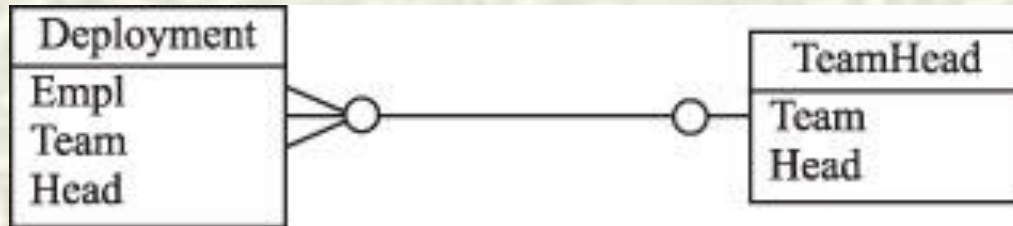
Нормальная форма Бойса-Кодда

Приведение отношения к НФБК заключается в создании дополнительного отношения, содержащего сведения о руководителях команд:

Отношение РУКОВОДИТЕЛЬ_КОМАНДЫ	
Команда	Руководитель
Наблюдение	Прохоров
Питание	Макаров
Наблюдение	Леонтьев
Питание	Лазарев
Обслуживание	Сидоров

Нормальная форма Бойса-Кодда

Результат приведения отношения КОМАНДА к НФБК:



Следует обратить внимание на возникновение избыточности данных

Четвертая нормальная форма

Отношение находится в четвертой нормальной форме (4НФ), если оно находится в 3НФ или НФБК и все независимые многозначные ФЗ разнесены в отдельные отношения с одним и тем же ключом.

Иными словами, 4НФ применяется при наличии в отношении более чем одной многозначной ФЗ и требует, чтобы отношение не содержало независимых многозначных ФЗ.

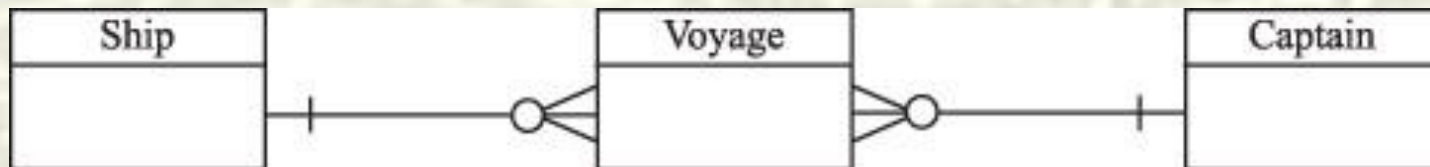
Четвертая нормальная форма

Пример: Приведение к 4НФ

Рассмотрим отношение, содержащее сведения о кораблях (Ship), совершаемых ими рейсах (Voyage) и капитанах (Captain)

Отношение КАПИТАН_КОРАБЛЬ_РЕЙС		
Акбар	Иванов	Санкт-Петербург - Калининград
Акбар	Петров	Санкт-Петербург - Калининград
Акбар	Ивлев	Санкт-Петербург - Калининград
Акбар	Прохоров	Санкт-Петербург - Калининград
Акбар	Лазарев	Санкт-Петербург- Лондон
Акбар	Прохоров	Санкт-Петербург- Лондон
Жучка	Петров	Санкт-Петербург - Марсель
Жучка	Фролов	Санкт-Петербург - Стокгольм
Жучка	Ивлев	Санкт-Петербург - Стокгольм

Четвертая нормальная форма



Отношение находится в НФБК и содержит только многозначные ФЗ.

Однако имеет место аномалия удаления: если капитан Петров уйдет в отставку и все кортежи о нем будут удалены, то будут потеряны сведения о том, что корабль Жучка совершает рейсы Санкт-Петербург - Марсель. Если добавить новый рейс, то, возможно, придется ввести несколько кортежей в наше отношение.

Четвертая нормальная форма

Приведение отношения к 4НФ заключается в выделении для каждой многозначной ФЗ своего отношения:



Пятая нормальная форма

Нормализация отношений выполнялась путем разложения (декомпозиции) схем отношений.

Очевидно, что при таком подходе должен соблюдаться принцип обратимости: соединение проекций должно приводить к исходным отношениям.

Это предполагает отсутствие потери кортежей; появление ранее не существовавших кортежей; сохранение ФЗ (семантика взаимосвязей между данными не должна нарушаться).

Пятая нормальная форма

Декомпозиция схем отношений не всегда гарантирует обратимость. Это обстоятельство связано с существованием класса ФЗ по соединению.

Если отношение удовлетворяет ФЗ по соединению, то оно может быть восстановлено по своим проекциям.

Отношения, содержащие более трех МФЗ, требуют особого внимания при построении *логической модели реляционной базы данных*. Также 4НФ не устраняет избыточность данных полностью, поэтому требуется дальнейшая декомпозиция схем отношений.

Пятая нормальная форма

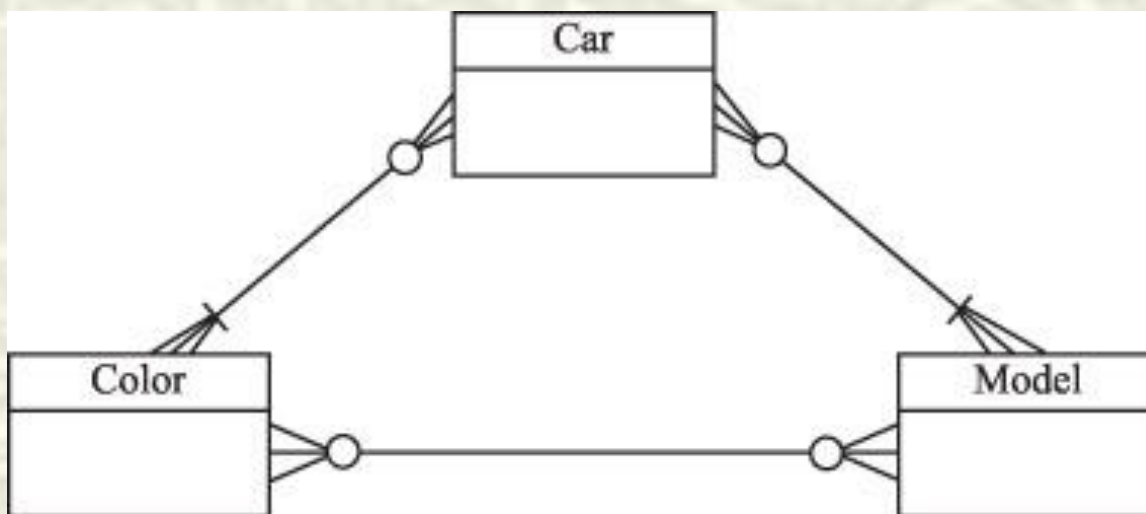
Отношение находится в пятой нормальной форме (5НФ), если оно находится в 4НФ и удовлетворяет зависимости по соединению относительно своих проекций.

5НФ называют также нормальной формой с проецированием соединений. Она используется для разрешения трех и более отношений, которые связаны более чем тремя ФЗ по типу «многие-ко-многим»

Пятая нормальная форма

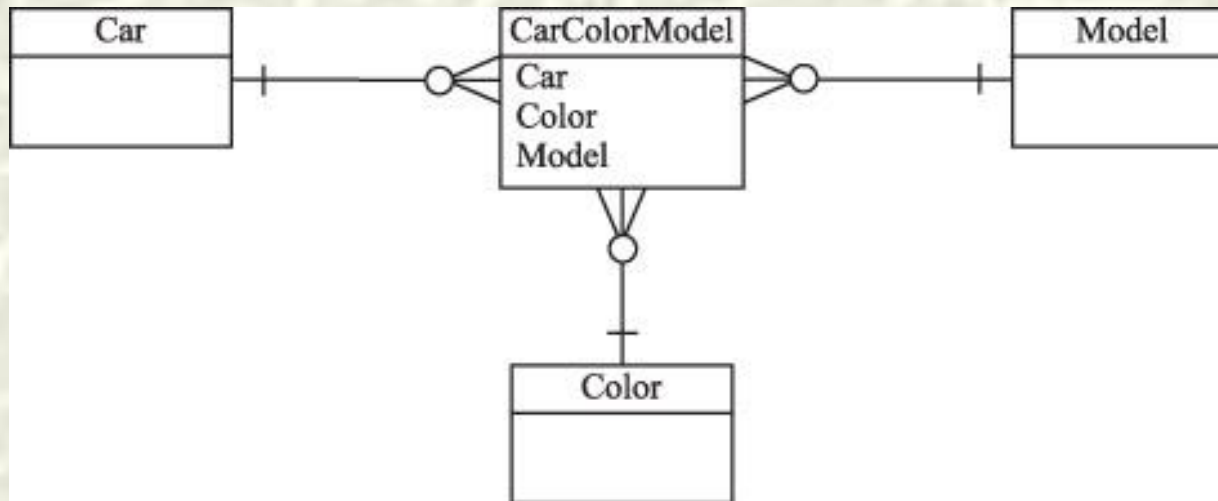
Пример: Приведение к 5НФ

Рассмотрим отношение с несколькими многозначными зависимостями:



Пятая нормальная форма

Приведение отношения к 5НФ заключается во введении еще одного отношения, связывающего три исходных отношения:



Процедура приведения отношения, содержащего многозначные ФЗ, к 5НФ состоит в построении связывающего отношения, позволяющего исключить появление в соединениях ложных кортежей.

Выводы (свойства нормальных форм)

- 1НФ - все атрибуты отношения простые;
- 2НФ - отношение находится в 1НФ и не содержит частичных ФЗ;
- 3НФ - отношение находится во 2НФ и не содержит транзитивных ФЗ от ключа;
- НФБК - отношение находится в 3НФ и не содержит ФЗ ключей от неключевых атрибутов;
- 4НФ, применяется при наличии более чем одной многозначной ФЗ - отношение находится в НФБК или 3НФ и не содержит независимых многозначных ФЗ;
- 5НФ - отношение находится в 4НФ и не содержит ФЗ по соединению.