

Проектирование базы данных

Какое желание вами управляет при проектировании базы данных?

А что если сосредоточить все данные в одной таблице?

Нормализация - проектирование или
перестройка структуры таблиц

Нормальная форма — требование, предъявляемое к структуре таблиц в теории реляционных баз данных для устранения из базы избыточных функциональных зависимостей между атрибутами

Первая НФ требует, чтобы таблица содержала атомарные значения, то есть на пересечении строки и столбца находилось единственное значение, а не список либо вложенная таблица.

Атомарные значения - это атрибуты, которые хранят единственное значение и не являются ни списком, ни множеством значений.

Первая НФ

Ненормализованная сущность

Фирма	Модели
BMW	M5, X5M, M1
Nissan	GT-R

Первая НФ

Фирма	Модели
BMW	M5
BMW	X5M
BMW	M1
Nissan	GT-R

Бывает, что первая нормальная форма нарушена

пример из лёрна...

Y функционально зависит от X, если при фиксированных значениях атрибутов X однозначно определяются значения атрибутов Y.

$$F:X \rightarrow Y \text{ или } Y=F(X)$$

Таблица поставок R (S_n , $Scity$, $Cstatus$, P_n , Q)

- S_n – шифр поставщика
- $Scity$ – город проживания поставщика
- $Cstatus$ – статус города, выраженный целым числом в зависимости от его населения и административного значения
- P_n – шифр изделия
- Q – объем поставки изделия P_n поставщиком S_n .

Какие функциональные зависимости вы здесь выделите?

Функциональные зависимости:

- $S_n \rightarrow Scity$;
- $Scity \rightarrow Cstatus$;
- $S_n, P_n \rightarrow Q$. (Q не зависит функционально ни от S_n , ни от P_n по отдельности)
- $S_n, P_n \rightarrow Cstatus$ (P_n в этой зависимости не играет никакой роли)

Полной функциональной зависимостью

$F: X \rightarrow Y$ считают такую функциональную зависимость, в которой Y зависит от всех атрибутов множества X , а не от какой-то их части.

Полная функциональная зависимость

$S_n, P_n \rightarrow C_{status}$ является неполной, т.к. существует зависимость $S_n \rightarrow C_{status}$.

Таблица поставок R (S_n , $Scity$, $Cstatus$, P_n , Q)

- S_n – шифр поставщика
- $Scity$ – город проживания поставщика
- $Cstatus$ – статус города, выраженный целым числом в зависимости от его населения и административного значения
- P_n – шифр изделия
- Q – объем поставки изделия P_n поставщиком S_n .

Для таблицы R неудобны базовые операции работы с таблицами: включение новых записей, удаление записей, корректировка информации.

Основным способом приведения отношения ко второй НФ является декомпозиция исходного отношения на два или более новых отношений.

Отношение R находится во второй НФ,
если оно находится в первой НФ и каждый
неключевой атрибут функционально полно
зависит от любого возможного ключа.

Таблица поставок R (S_n , Scity, Cstatus, P_n , Q)

Так отношение R можно декомпозировать на отношения:

- S (S_n , Scity, Cstatus)
- SP (S_n , P_n , Q)

Проблемы возникают вследствие транзитивной функциональной зависимости Scity \rightarrow Cstatus, то есть последовательности зависимостей

$S_n \rightarrow \text{Scity} \rightarrow \text{Cstatus}$.

Отношение R находится в третьей НФ,
если оно находится во второй НФ, и
каждый неключевой атрибут зависит от
любого возможного ключа нетранзитивно.

Транзитивные зависимости между полями базы данных существует тогда, когда значения не ключевых полей зависят от значений других не ключевых полей.

Таблица поставок R (S_n , Scity, Cstatus, P_n , Q)

Так отношение R можно декомпозировать на отношения:

- S (S_n , Scity, Cstatus)
- SP (S_n , P_n , Q)

В отношении S имеется единственный ключ S_n , а атрибут Cstatus функционально зависит от неключевого атрибута Scity.

Выходом является декомпозиция отношения S на отношения S1 (S_n , Scity) и S2 (Scity, Cstatus).

Таблица сведения об успеваемости школьников

Таблица R (Sch, Adr, Cl, Teach, Subj, Sp_mark, Sp_date).

- Sch – идентификация школьника
- Adr – его адрес
- Cl – класс
- Teach – классный руководитель
- Subj – учебный предмет
- Sp_mark – список полученных оценок
- Sp_date – даты получения оценок.

Первая НФ

Таблица R (Sch, Adr, Cl, Teach, Subj, Sp_mark, Sp_date).

- Sch – идентификация школьника
- Adr – его адрес
- Cl – класс
- Teach – классный руководитель
- Subj – учебный предмет
- Mark – Полученная оценок
- Date – дата получения оценки.

Вторая НФ

Таблица R (Sch, Adr, Cl, Teach, Subj, Sp_mark, Sp_date) декомпозируется на:

- S (Sch, Adr, Cl, Teach)
- R (Sch, Subj, Mark, Date)

В этой таблице имеется транзитивная зависимость $Sch \rightarrow Cl \rightarrow Teach$

третья НФ

Таблица S (Sch, Adr, Cl, Teach) декомпозируется на:

- S (Sch, Adr, Cl)
- C (Cl, Teach)

Результаты:

- R (Sch, Subj, Mark, Date)

Нормальная форма Бойса-Кодда (НФБК)

Таблицы поставок:

- S (Sn, Scity, Cstatus)
- SP (Sn, Sname, Pn, Q)

Пусть в определенном выше отношении SP присутствует еще и имя поставщика Sname. Имеются два возможных ключа

- Sn, Pn;
- Sname, Pn.

Определим детерминант отношения как набор атрибутов, от которого функционально полно зависит какой-либо другой атрибут.

Отношение R находится в
НФ Бойса-Кодда (НФБК),
если каждый детерминант является
ВОЗМОЖНЫМ КЛЮЧОМ ОТНОШЕНИЯ.

Приведение к НФБК подразумевает декомпозицию таблицы

Таблицы поставок:

- S (Sn, Scity, Cstatus)
- S3 (Sn, Sname),
- SP (Sn, Pn, Q)

Аномалии НФБК

отношение R (City, Adr, Ind) - сведения о почтовых адресах:

- City – название города
- Adr – адрес без указания города
- Ind – шестизначный почтовый индекс

В городе каждое почтовое отделение имеет свой индекс.

Имеются функциональные зависимости:

- City, Adr \rightarrow Ind;
- Ind \rightarrow City.

Вариант декомпозиции в НФБК

- R1 (Ind, City)
- R2 (Ind, Adr)

Пусть требуется по заданным значениям атрибутов City и Adr определить Ind

Четвертая нормальная форма

Рассмотрим таблицу R (Subj, Teach, Book)

- Subj – учебный предмет
- Teach– преподаватель по этому предмету
- Book – книга, рекомендуемая преподавателем Teach для изучения предмета Subj.

Subj	Teach	Book
Математика	Иванов	Учебник
Математика	Иванов	Задачник
Математика	Иванов	Методичка
Математика	Петров	Учебник
Математика	Петров	Задачник
Математика	Петров	Методичка

Четвертая нормальная форма

Декомпозируем таблицу R (Subj, Teach, Book)

- R1 (Subj, Teach)
- R2 (Subj, Book)

При многозначной зависимости
совокупностей атрибутов X и Y ($X \twoheadrightarrow Y$)
каждое значение X определяет множество
значений Y , которое не зависит от других
атрибутов отношения.

Subj \twoheadrightarrow Teach и Subj \twoheadrightarrow Book

Другой вариант разложения отношения R

Декомпозируем таблицу R (Subj, Teach, Book)

- R1 (Subj, Teach)
- R3 (Teach, Book)

Четвертая нормальная форма

Таблица R (City, Day, Model) содержит сведения о самолетных рейсах


City	Day	Model
Йошкар-Ола	Понедельник	747
Йошкар-Ола	Четверг	747
Йошкар-Ола	Понедельник	A380
Йошкар-Ола	Четверг	A380
Казань	Среда	707
Казань	Среда	747

Четвертая нормальная форма

Отношение R разлагается без потерь на два отношения

- R1 (City, Day)
- R2 (City, Model)

City	Day
Йошкар-Ола	Понедельник
Йошкар-Ола	Четверг
Казань	Среда

City	Model
Йошкар-Ола	747
Йошкар-Ола	A380
Казань	707 
Казань	747

Пусть в отношении R имеются два множества атрибутов X и Y . Обозначим через Z оставшиеся атрибуты отношения R , то есть $Z = R - (X, Y)$. Множество атрибутов Y отношения R многозначно зависит от множества атрибутов X ($X \twoheadrightarrow Y$), если для любых двух кортежей (x_1, y_1, z_1) и (x_1, y_2, z_2) в отношении существуют также кортежи (x_1, y_1, z_2) и (x_1, y_2, z_1) .

Многозначную зависимость называют тривиальной, если множество атрибутов X включает в себя множество атрибутов Y либо X и Y охватывают все атрибуты отношения.

Четвертая нормальная форма

Рассмотрим таблицу R (Subj, Teach, Book).

Многозначные зависимости Subj -» Teach и Subj -» Book.

Subj	Teach	Book
Математика	Иванов	Учебник
Математика	Иванов	Задачник
Математика	Иванов	Методичка
Математика	Петров	Учебник
Математика	Петров	Задачник
Математика	Петров	Методичка

Четвертая нормальная форма

Таблица R (City, Day, Model) содержит сведения о самолетных рейсах

City	Day	Model
Йошкар-Ола	Понедельник	747
Йошкар-Ола	Четверг	747
Йошкар-Ола	Понедельник	A380
Йошкар-Ола	Четверг	A380
Казань	Среда	707
Казань	Среда	747

Отношение находится в четвертой НФ, если оно находится в НФБК, и отсутствуют нетривиальные многозначные зависимости, то есть все многозначные зависимости фактически представляют собой функциональные зависимости от ключей отношения.