

Лекция #3

Базы данных

Михаил Моисеев

Теория нормальных форм

Функциональные зависимости

РБД содержит как структурную, так и семантическую информацию.

Семантическая часть представляет множество функциональных зависимостей между атрибутами отношений.

Если даны два атрибута X и Y некоторого отношения, то говорят, что Y **функционально зависит** от X , если в любой момент времени каждому значению X соответствует ровно одно значение Y .

Обозначение: $X \rightarrow Y$ (Y функционально зависит от X).

Функциональные зависимости #2

Студенты: (Имя, Фамилия, Отчество, Дата рождения, Номер группы,...).

Группы: (Номер группы, Специальность, Номер курса, ...).

Специальности: (Название, Перечень предметов ...).

Атрибуты функционально зависят от РК.

Нормализация

Избыточная функциональная зависимость – может быть получена на основе других зависимостей БД.

Корректной считается такая схема базы данных, в которой **отсутствуют избыточные функциональные зависимости.**

Нормализация - обратимый пошаговый процесс замены данной совокупности отношений другой схемой с устранением избыточных функциональных зависимостей.

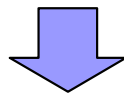
Обычно нормализация представляет собой декомпозицию отношений.

Нормализация #2

Условие обратимости требует, чтобы при нормализации сохранялась эквивалентность схем БД:

- не должны появляться ранее отсутствовавшие сущности (кортежи);
- на отношениях новой схемы должно выполняться исходное множество функциональных зависимостей.

Студент: (Id, ФИО , Номер группы, ФИО старосты группы, Специальность ...)



Студент: (Id, ФИО, Дата рождения, Ссылка на группу, ...).

Группа: (Номер группы, Ссылка на старосту, Ссылка на специальн., ...).

Наличие избыточных зависимостей приводит к дублированию данных и аномалиям при выполнении операций

Первая нормальная форма (1NF)

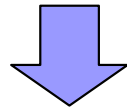
Простой атрибут - атрибут, значения которого атомарны (неделимы).

Сложный атрибут - получается соединением нескольких атомарных атрибутов, которые могут быть определены на одном или разных доменах.

Отношение находится в **1NF** если значения всех его атрибутов атомарны.

1NF #2

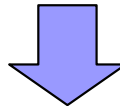
Студенты: (Имя, Фамилия, Отчество, Выполненные работы, ...).
Выполненные работы – может содержать от 0 до N элементов.



Студенты: (Имя, Фамилия, Отчество, Предмет, ...).
Выполненные работы: (Название работы, Ссылка на студента (FK) ...).

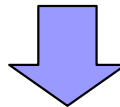
1NF #3

Предметы: (Название, Рекомендуемая литература, ...).



Предметы: (Название, ...).

Рекомендуемая литература : (Название, Ссылка на предмет, ...).



Предметы: (Название, ...).

Отношение : (Ссылка на предмет, Ссылка на литературу).

Рекомендуемая литература : (Название, Автор, ...).

Вторая нормальная форма (2NF)

PK состоящий из несколько атрибутов называется **составным**.

Атрибут **функционально полно** зависит от составного PK, если он зависит от всего PK в целом и не зависит от какого-либо атрибута PK.

Отношение находится в **2NF**, если оно находится в **1NF** и каждый неключевой атрибут функционально полно зависит от PK.

Для простого PK **1NF \equiv 2NF**

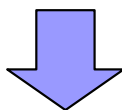
2NF #2

Предметы: (Название, Семестр, Учебная программа, ...).

РК: (Название, Семестр);

Ф3: (Название, Семестр → Учебная программа;
Название → Учебная программа)

При изменении учебной программы для некоторого предмета необходимо изменить записи для всех семестров, атрибут учебная программа дублируется.



Предметы: (Название, Учебная программа, ...).

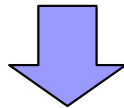
Расписание: (Ссылка на предмет, Семестр, ...).

2NF #3

Книги: (Название, Переиздание, Издательство ...).

РК: (Название, Переиздание);

Ф3: (Название, Переиздание → Издательство;
Название → Издательство)



Книги: (Название, Издательство, ...).

Переиздания: (Переиздание, Ссылка на книгу (FK), Год выпуска, ...).

РК: (Переиздание, Ссылка на книгу)

Вопросы

- Что такое функциональная зависимость ?
- Сколько функциональных зависимостей может быть в отношении ?
- Между чем устанавливаются функциональные зависимости ?
- Какие функциональные зависимости нежелательны и почему ?
- Для чего выполняется нормализация ?
- Какие приемы используют при нормализации ?
- Какие условия должны выполняться при нормализации ?
- Определение 1NF.
- Приведите пример нарушения 1NF как они устраняются ?
- Что такое функционально полная зависимость ?
- Определение 2NF.
- Приведите пример нарушения 2NF как они устраняются ?
- В каких случаях 1NF эквивалента 2NF ?
- К чему приводят избыточные функциональные зависимости ?

Транзитивные зависимости

Если для атрибутов X , Y и Z некоторого отношения, выполняется $X \rightarrow Y$ и $Y \rightarrow Z$, а обратные ФЗ отсутствуют то Z **транзитивно зависит** от X .

обратные ФЗ: $Y \rightarrow X$ и $Z \rightarrow Y$ – отсутствуют

Студент: (Имя, Фамилия, Отчество, Число отличных оценок, Получен красный диплом, ...)

ФЗ: Имя, Фамилия, Отчество \rightarrow Число отличных оценок;
Число отличных оценок \rightarrow Получен красный диплом

Третья нормальная форма (3NF)

Отношение находится в **3NF** если оно находится в **2NF** и все неключевые атрибуты нетранзитивно зависят от первичного ключа.

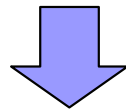
Сумма требований к отношению в 3NF:

- Атомарность атрибутов (из 1NF);
- Зависимость неключевых атрибутов от всего ключа в целом (из 2NF), подразумевает зависимость неключевых атрибутов от ключа;
- Нетранзитивность этих зависимостей.

3NF #2

Студент: (Имя, Фамилия, Отчество, Число отличных оценок,
Получен красный диплом, ...)

Аномалией является дублирование данных о получении
красного диплома



Студент: (Имя, Фамилия, Отчество, Число отличных оценок ...)
Получение красного диплома: (Число отличных оценок,
Получение диплома, ...)

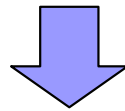
Сколько кортежей в отношении Получение красного диплома ?

3NF #3

Предмет: (Название, Учебная программа, Количество часов, ...)

Ф3: Название → Учебная программа

Учебная программа → Количество часов



Предмет: (Название, Ссылка на учебную программу, ...)

Учебная программа: (Название, Количество часов, ...)

Практика применения **3NF**

PK должен **однозначно определять** каждый кортеж в отношении.

Все атрибуты отношения должны быть **атомарны** до той степени, которая обеспечивается СУБД.

Каждый неключевой атрибут должен **непосредственно** зависеть от PK.

Каждый неключевой атрибут должен зависеть от **всего** PK.

Каждый неключевой атрибут не должен зависеть **ни от чего** другого кроме PK.

Возможные ключи

Возможный(альтернативный) ключ (**АК**) - это атрибут или несколько атрибутов, которые отвечают требованиям, предъявляемым к **РК**:

- Обеспечивает уникальность кортежей;
- Является не избыточным (не существует подмножества атрибутов РК, которые обеспечивают уникальность кортежей).

Отношение может иметь несколько возможных ключей - **АК**.

В качестве **РК** выбирается один из **АК** – обычно тот, который включает наименьшее число полей либо в качестве **РК** используется суррогатный ключ.

Форма Бойса-Кодда (BCNF)

Отношение находится в **BCNF**, если оно находится в **3NF** и в ней отсутствуют зависимости атрибутов первичного ключа от неключевых атрибутов.

Ситуация когда **3NF** не эквивалента **BCNF** возникает когда имеется несколько **АК**, которые являются составными имеют общий атрибут.

BCNF #2

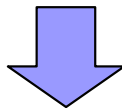
Студент_НИР: (Имя, Фамилия, Отчество, Номер паспорта, Серия паспорта, Название НИР, ...)

АК: Имя, Фамилия, Отчество, Название НИР - **PK**

Номер паспорта, Серия паспорта, Название НИР

ФЗ: Номер паспорта, Серия паспорта → Имя, Фамилия, Отчество

Аномалия – дублирование данных.



Студент: (Имя, Фамилия, Отчество, Номер паспорта, Серия паспорта, ...)

Студент_НИР: (Ссылка на исполнителя, Ссылка на НИР, ...)

НИР: (Название работы, Заказчик, Ссылка на руководителя, ...)

Четвертая нормальная форма (4NF)

Многозначная ФЗ – зависимость между множествами значений атрибутов.

Преподаватель: (ФИО, Курс, Учебное пособие, ...)

Аномалия из-за дублирования данных, преподаватель читает несколько курсов и имеет несколько учебных пособий. Для изучения каждого курса предлагается использовать все учебные пособия преподавателя.

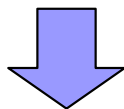
Отношение находится в **4NF** в случае, если оно находится в **BCNF** и все многозначные зависимости являются функциональными.

4NF #2

Преподаватель: (ФИО, Курс, Учебное пособие, ...)

Для изучения каждого курса предлагается использовать **все** учебные пособия преподавателя.

ФЗ: ФИО → Курс;
ФИО → Учебное пособие;
Курс → Учебное пособие;
Учебное пособие → Курс;



Преподаватель_Курс: (ФИО, Курс,...)

Преподаватель_Пособие: (ФИО, Учебное пособие,...)

Пятая нормальная форма (5NF)

Обычно при нормализации производится декомпозиция одного отношения в два. Иногда это сделать не удастся, но возможна декомпозиция в большее число отношений, каждое из которых обладает лучшими свойствами. Такие ФЗ называются **зависимостями по соединению**.

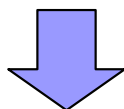
Отношение находится в **5NF** когда оно находится в **4NF** и проведена декомпозиция зависимостей по соединению.

Студент-Группа-НИР(Ссылка на студента, Ссылка на рабочую группу, Ссылка на НИР)

Предположим, что студент может работать в разных рабочих группах над разными НИР.

5NF #2

Студент-Группа-НИР(Ссылка на студента, Ссылка на рабочую группу, Ссылка на НИР)



Студент-Группа(Ссылка на студента, Ссылка на рабочую группу)

Студент-НИР(Ссылка на студента, Ссылка на НИР)

Группа-НИР(Ссылка на рабочую группу, Ссылка на НИР)

Практика применения НФ

Приведение схемы БД (всех отношений) к **3NF** считается необходимым и достаточным.

Приведение к более высоким нормальным формам требуется лишь в некоторых случаях (при наличии нежелательных аномалий).

Работа с ненормализованной БД иногда используется для повышения производительности запросов.

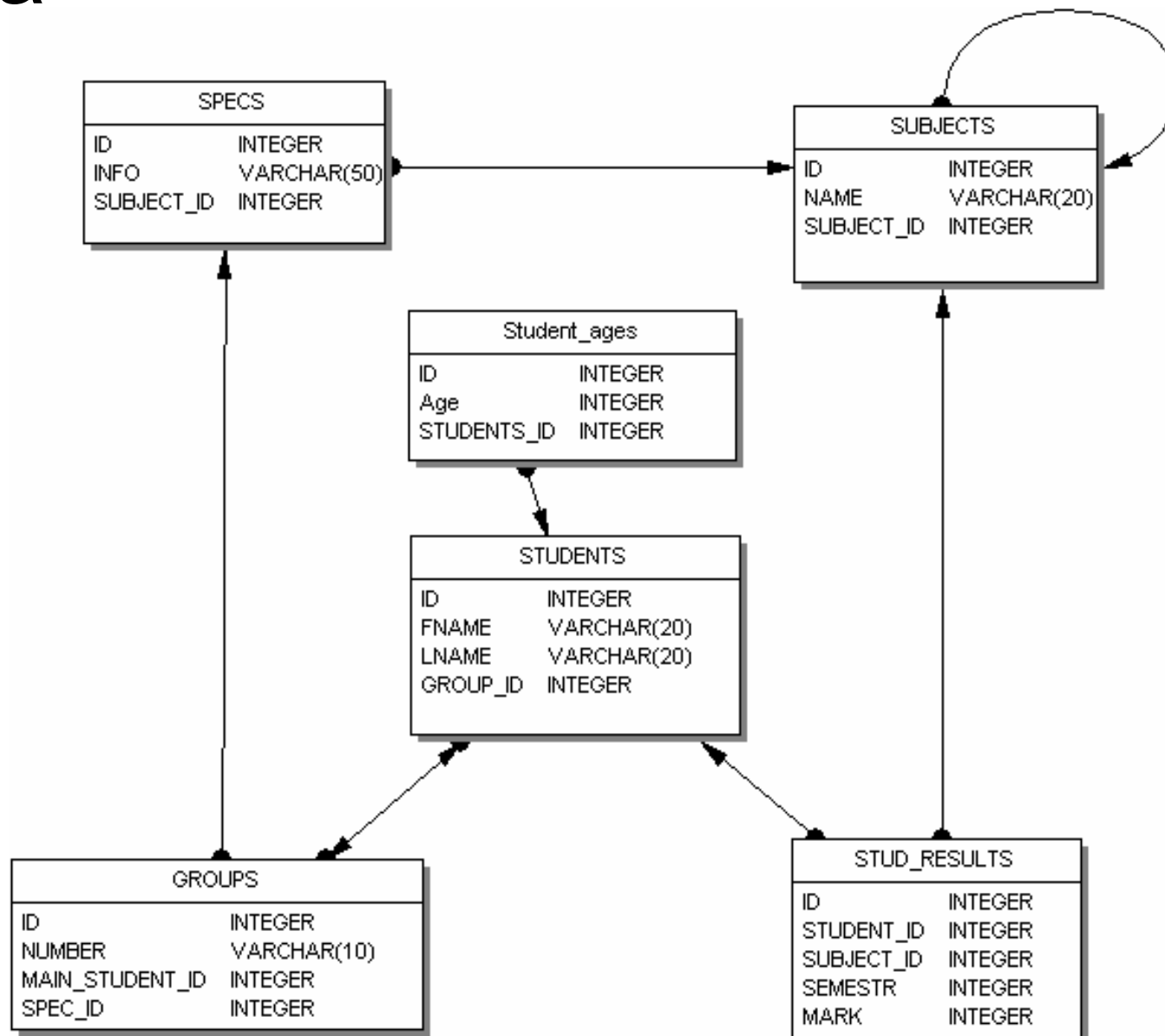
Ненормализованная БД - некоторые отношения не удовлетворяют 3NF.

Существуют другие нормальные формы.

Вопросы

- Перечислите основные требования 3NF.
- От чего должен зависеть неключевой атрибут ?
- От чего не должен зависеть неключевой атрибут ?
- Сколько АК может быть в одном отношении ?
- Какие требования предъявляются к РК ?
- Перечислите достоинства и недостатки применения суррогатного РК.
- В чем отличия BCNF от 3NF ?
- Что такое множественная ФЗ ?
- В чем отличия 4NF от BCNF ?
- К каким аномалиям приводят несоответствие требованиям BCNF и 4NF ?
- Чем отличается 5NF от 4NF ?
- Какие из НФ применяются на практике и почему ?
- Всегда ли следует добиваться выполнения требований НФ ?
- Какими основными свойствами обладает 6NF ?

Задача



Решение

