

Реляционная теория – назад к основам

Широкопетлева Мария



#377 | KIEV 2015

Аудитория

- Разработчики:
 - Почему вся информация не в одной таблице?
- Администраторы:
 - Как найти компромисс между количеством таблиц и скоростью выполнения запросов?

Критерии оценки качества логической модели данных



- Адекватность базы данных предметной области
- Легкость разработки и сопровождения базы данных
- Скорость выполнения операций обновления данных (вставка, обновление, удаление)
- Скорость выполнения операций выборки данных

Адекватность БД предметной области



- Состояние базы данных в каждый момент времени должно соответствовать состоянию предметной области.
- Изменение состояния предметной области должно приводить к соответствующему изменению состояния базы данных
- Ограничения предметной области, отраженные в модели предметной области, должны некоторым образом отражаться и учитываться в базе данных.

Легкость разработки и сопровождения БД



- Использование хранимых процедур и триггеров

! Чем больше программного кода в виде триггеров и хранимых процедур содержит БД, тем сложнее ее разработка и дальнейшее сопровождение.

Скорость операций обновления данных



- скорость выполнения операции **вставки, обновления и удаления уменьшается** при увеличении количества индексов таблицы и мало зависит от числа строк в таблице
- чем больше атрибутов имеют отношения, тем медленнее будут выполняться операции обновления данных (затраты времени на перестройку большего количества индексов)

Скорость операций выборки данных



- Увеличение количества отношений приводит к замедлению выполнения операций выборки данных, особенно, если запросы заранее неизвестны



Аномалии БД

- Аномалия обновления
- Аномалии удаления
- Аномалии вставки

- Причины:
 - наличие в одном отношении разнородной информации
 - избыточность данных

Пример аномалии БД



Шифр_спец	Н_гр	Н_студ	ФИО	Оценка	Дата_Экз	Предмет	Преподаватель
122	ПМ-1	1	Альтов	С	11.09.15	ПМ	Андреев
122	ПМ-1	1	Альтов	В	23.12.15	ОП	Борисов
121	КИ-1	2	Петров	В	11.02.15	ВМ	Волков
122	ПМ-1	3	Басов	С	04.06.15	ОП	Борисов
122	ПМ-1	3	Басов	А	11.02.15	Ин.яз	Шутов
121	ПИ-5	7	Вовк		!	!	

Основные нормальные формы



- первая нормальная форма (1NF);
- вторая нормальная форма (2NF);
- третья нормальная форма (3NF);
- нормальная форма Бойса-Кодда (BCNF);
- четвертая нормальная форма (4 NF);
- пятая нормальная форма или нормальная форма проекции-соединения

1НФ

(Первая Нормальная Форма)



Первая нормальная форма (1НФ) - это обычное отношение. Обязательно: на пересечении любого столбца и любой записи находится атомарное значение.

Группы - Студенты – Оценки - Предметы

Шифр_спец	Н_гр	Н_студ	ФИО	Оценка	Дата_Экз	Предмет	Преподаватель
122	ПМ-1	1	Альтов	С	11.02.16	ВМ	Андреев
				В	23.12.15	ОП	Борисов
		3	Басов	В	11.02.16	ВМ	Андреев
				С	04.01.16	ОП	Борисов
121	КН-1	2	Петров	А	11.02.16	Ин.яз	Шутов

1НФ

(Первая Нормальная Форма)



Шифр_спец	Н_гр	Н_студ	ФИО	Оценка	Дата_Экз	Предмет	Преподаватель
122	ПМ-1	1	Альтов	С	11.02.16	ВМ	Андреев
122	ПМ-1	1	Альтов	В	23.12.15	ОП	Борисов
121	КН-1	2	Петров	В	11.02.16	ВМ	Андреев
122	ПМ-1	3	Басов	С	04.01.16	ОП	Борисов
122	ПМ-1	3	Басов	А	11.02.16	Ин.яз	Шутов

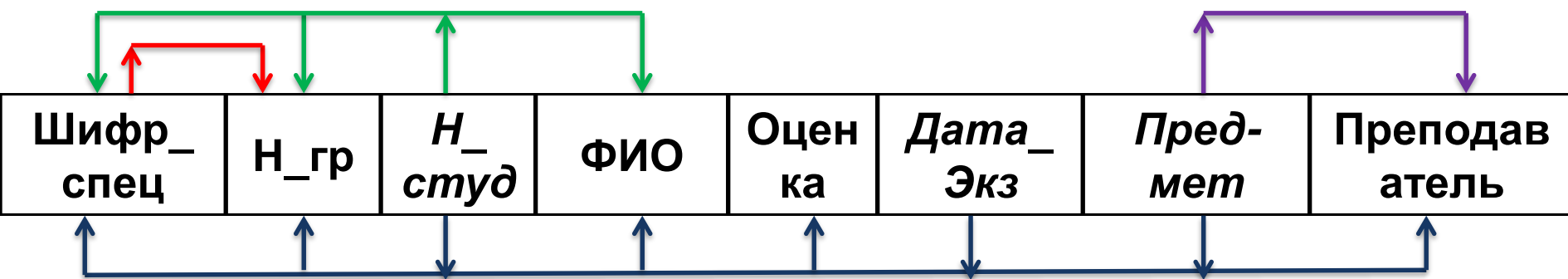
2NF



(Вторая Нормальная Форма)

Отношение находится во *второй нормальной форме (2NF)* в том и только в том случае, когда оно находится в 1NF, и каждый неключевой атрибут полностью функционально зависит от первичного ключа отношения.

Группы - Студенты – Оценки - Предметы



Процедура приведения ко 2НФ



Группы - Студенты – Оценки - Предметы

Шифр_спец	* Н_гр	Н_студ	ФИО	Оценка	*Дата_Экз	*Предмет	Преподаватель
-----------	--------	--------	-----	--------	-----------	----------	---------------

Группы - Студенты

Шифр_спец	Н_гр	*Н_студ	ФИО
122	ПМ-1	1	Альтов
121	КН-1	2	Петров
122	ПМ-1	3	Басов

Предметы

*Предмет	Преподаватель
ВМ	Андреев
ОП	Борисов
Ин.яз	Шутов

Оценки

*Н_студ	Оценка	*Дата_Экз	*Предмет
1	С	11.02.16	ВМ
1	В	23.12.15	ОП
2	В	11.02.16	ВМ
3	С	04.01.16	ОП
3	А	11.02.16	Ин.яз

3NF

(Третья Нормальная Форма)



Отношение находится в третьей нормальной форме (3NF) в том и только в том случае, если находится в 2NF и каждый неключевой атрибут нетранзитивно зависит от первичного ключа.

Группы - Студенты

Шифр_спец	Н_гр *	Н_студ	ФИО
122	ПМ-1	1	Альтов
121	КН-1	2	Петров
122	ПМ-1	3	Басов

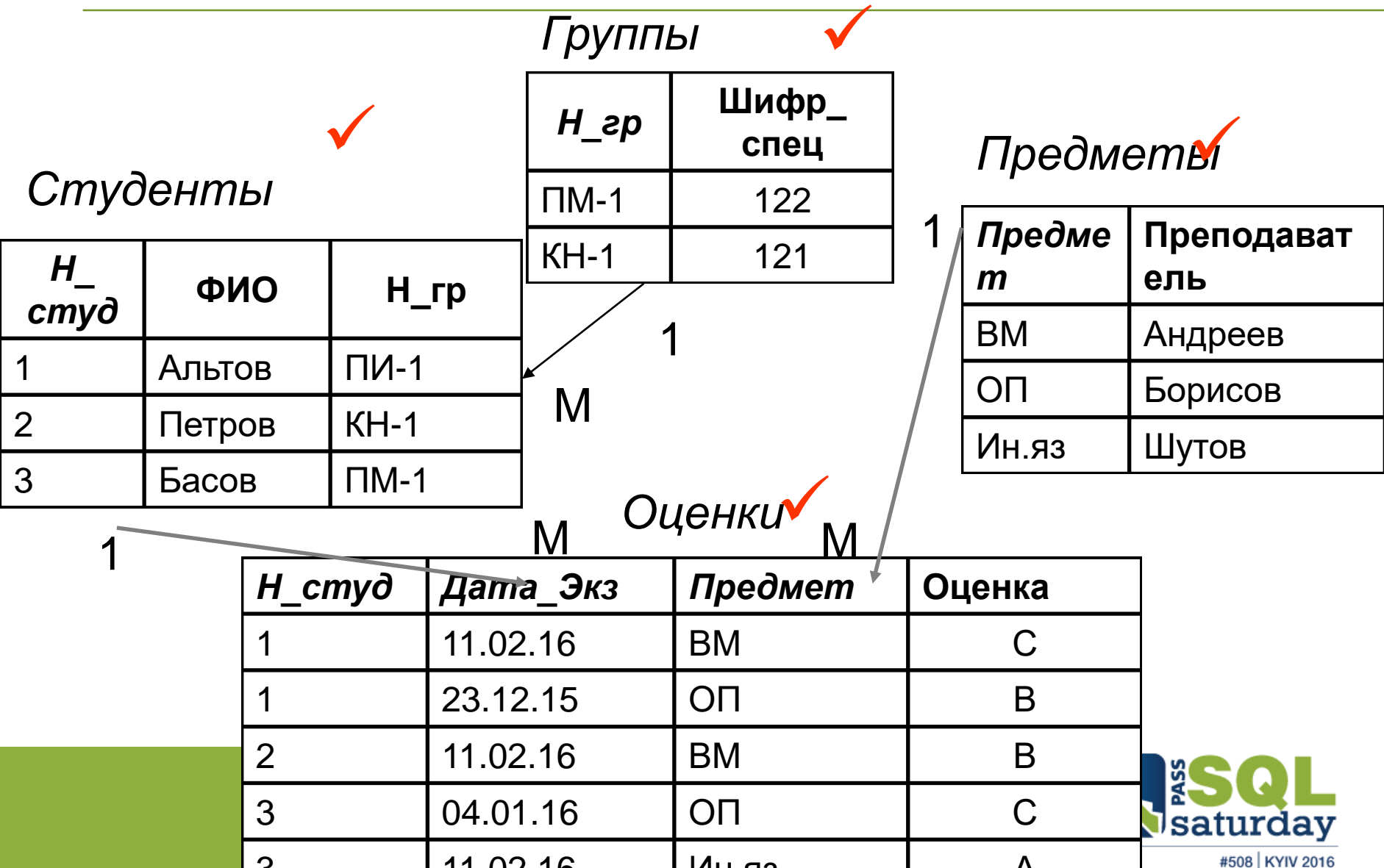
Группы

Н_гр	Шифр_спец
ПМ-1	122
КН-1	121

Студенты

Н_студ	ФИО	Н_гр
1	Альтов	ПМ-1
2	Петров	КН-1
3	Сидоров	ПМ-1

Пример схемы БД в 3НФ



Нормальная форма Бойса-Кодда (BCNF)



Отношение находится в BCNF тогда и только тогда, когда каждая его нетривиальная и неприводимая слева функциональная зависимость имеет в качестве своего детерминанта некоторый потенциальный ключ.

Чаще всего 3NF и BCNF эквивалентны.

Ситуация, когда отношение будет находиться в 3NF, но не в НФБК, возникает, например, при наличии в отношении двух или более потенциальных ключей, которые являются составными и имеют общий атрибут

Приведение к BCNF

<i>H_студ</i>	Оценка	<i>Дата_Экз</i>	<i>Предмет</i>	Преподаватель
1	С	11.02.16	ВМ	Андреев
1	В	23.12.15	ОП	Борисов
2	В	11.02.16	ВМ	Андреев
3	С	04.01.16	ОП	Борисов
3	А	11.02.16	Ин.яз	Шутов

Считаем, что только 1 преподаватель ведет 1 предмет и наоборот

Приведение к BCNF



Предметы

1)

<i>*Предмет</i>	<i>Преподаватель</i>
ВМ	Андреев
ОП	Борисов
Ин.яз	Шутов

Оценки

<i>Н_студ</i>	<i>Оценка</i>	<i>Дата_Экз</i>	<i>Предмет</i>
1	87	11.02.16	ВМ
1	52	23.12.15	ОП
2	77	11.02.16	ВМ
3	99	04.01.16	ОП
3	64	11.02.16	Ин.яз

Преподаватели

2)

<i>*Преподаватель</i>	<i>Предмет</i>
Андреев	ВМ
Борисов	ОП
Шутов	Ин.яз

Оценки

<i>Н_студ</i>	<i>Оценка</i>	<i>Дата_Экз</i>	<i>Преподаватель</i>
1	87	11.02.16	Андреев
1	52	23.12.15	Борисов
2	77	11.02.16	Андреев
3	99	04.01.16	Борисов
3	64	11.02.16	Шутов

Введение суррогатного ключа



Н_студ	Оценка	Дата_Экз	Предмет	Преподаватель
1	С	11.02.16	ВМ	Андреев
1	В	23.12.15	ОП	Борисов
2	В	11.02.16	ВМ	Андреев
3	С	04.01.16	ОП	Борисов
3	А	11.02.16	Ин.яз	Шутов

Н_студ	Оценка	Дата_Экз	Н_пред_преп
1	С	11.02.16	11
1	В	23.12.15	12
2	В	11.02.16	11
3	С	04.01.16	12
3	А	11.02.16	13

Н_пред_преп	Предмет	Преподаватель
11	ВМ	Андреев
12	ОП	Борисов
13	Ин.яз	Шутов

4 нормальная форма



Отношение находится в 4NF, если оно находится в BCNF и не содержит нетривиальных многозначных зависимостей.

Т.е. все многозначные зависимости, по сути являются функциональными зависимостями от ключей отношения

4 нормальная форма



Функциональные зависимости:

Дисциплина -> Преподаватель

Дисциплина -> Учебник

Дисциплина	Преподаватель	Учебник
Информатика	Шипилов П.А.	MSDN
Информатика	Шипилов П.А.	Г.Буч. ООП
Информатика	Головко Г.Л.	MSDN
Информатика	Головко Г.Л.	Г.Буч. ООП

Преподаватель дисциплины

Дисциплина	Преподаватель
Информатика	Шипилов П.А.
Информатика	Головко Г.Л.

Учебники по дисциплине

Дисциплина	Учебник
Информатика	MSDN
Информатика	Г.Буч. ООП



5 нормальная форма

Отношение находится в пятой нормальной форме (иначе - в НФ проекции-соединения) тогда и только тогда, когда каждая нетривиальная зависимость соединения в нем определяется потенциальным ключом (ключами) этого отношения.



5 нормальная форма

5 НФ – конечная нормальная форма в контексте операций проекции – соединения.

5НФ всегда достигается.

Практически всегда, если отношение находится в 4НФ, то оно находится в 5 НФ.

Исключения: предметные области с возможностями ограничений, не выраженных в структуре отношений

Доменно-ключевая НФ (ДКНФ)

Отношение R находится в ДКНФ тогда и только тогда, когда каждое наложенное ограничение является логическим следствием ограничений доменов и ограничений ключей.

Отношение в ДКНФ не имеет аномалий модификации.

Т.е. чтобы не изменялось – ничего не потеряется, если выполнены все ограничения по отношению к ключам и доменам.

Доменно-ключевая НФ (ДКНФ)



Ограничение домена - это ограничение, которое предлагает использование для определенного атрибута значений только из некоторого заданного домена.

Ограничение ключа - это ограничение, которое утверждает, что некоторый атрибут или комбинация атрибутов является потенциальным ключом.

Эффективность нормализации



Критерий	Отношения слабо нормализованы (1НФ, 2НФ)	Отношения нормализованы (3НФ)	Отношения сильно нормализованы (4НФ, 5НФ)
Адекватность базы данных предметной области	ХУЖЕ (-)	ЛУЧШЕ (+)	ЛУЧШЕ (+)
Легкость разработки и сопровождения базы данных	СЛОЖНЕЕ (-)	ЛЕГЧЕ (+)	СЛОЖНЕЕ (-)
Скорость выполнения вставки, обновления, удаления	МЕДЛЕННЕЕ (-)	БЫСТРЕЕ (+)	МЕДЛЕННЕЕ (-)
Скорость выполнения выборки данных	БЫСТРЕЕ (+)	МЕДЛЕННЕЕ (-)	МЕДЛЕННЕЕ (-)



DEMO

Спасибо за внимание!

