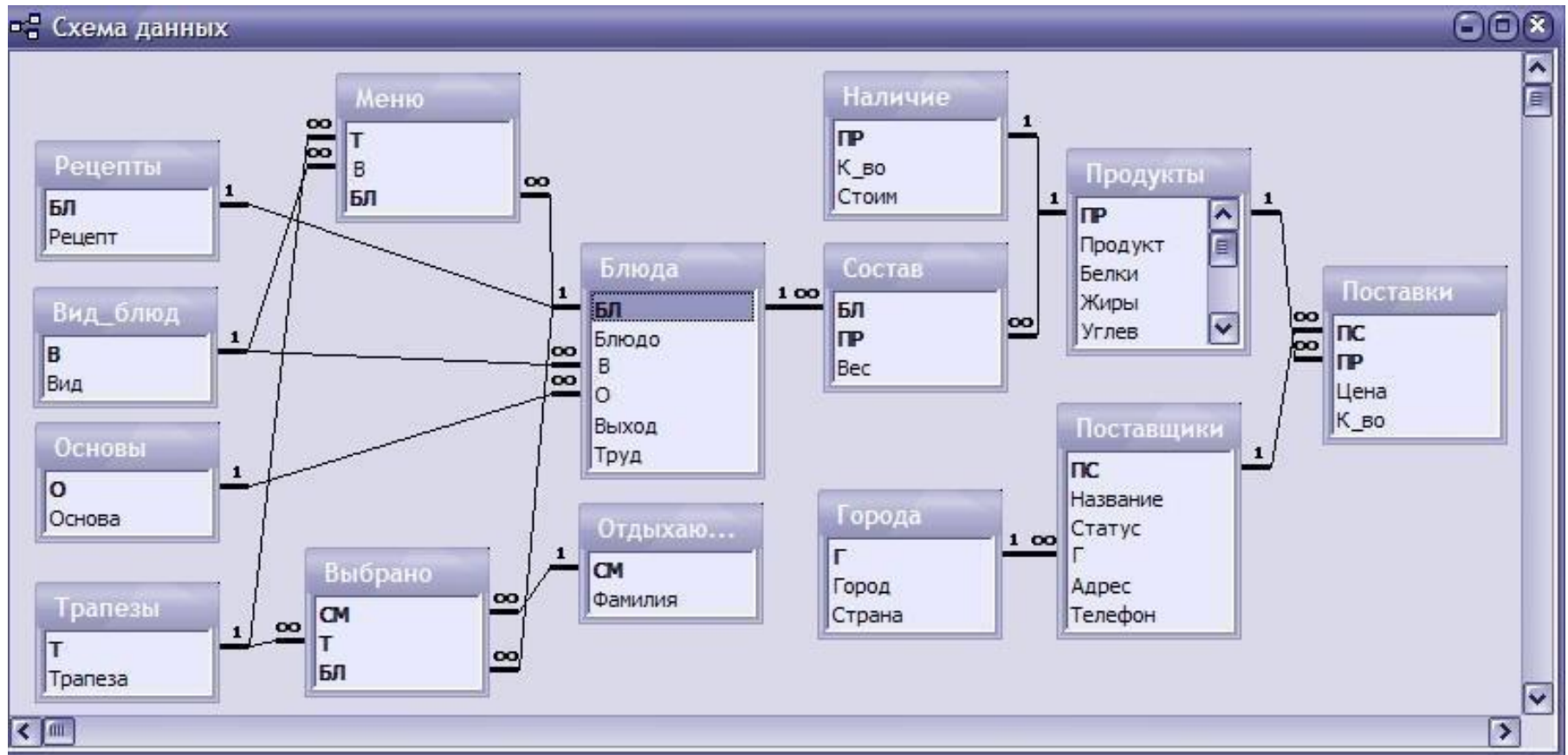


5. Технология проектирования реляционных баз данных, основанная на нормализации

Схема базы данных ПАНСИОН



. Цели проектирования БД

Основная цель проектирования БД – это **сокращение избыточности хранимых данных**, а следовательно, экономия объема используемой памяти, уменьшение затрат на многократные операции обновления избыточных копий и устранение возможности возникновения противоречий из-за хранения в разных местах сведений об одном и том же объекте.

Так называемый, "чистый" проект БД (**“Один факт в одном месте”**) можно создать, используя методологию нормализации отношений.

Нормализация и функциональные зависимости

Нормализация – это разбиение таблицы на две или более, обладающих лучшими свойствами при включении, изменении и удалении данных.

Окончательная цель нормализации сводится к получению такого проекта базы данных, в котором **каждый факт появляется лишь в одном месте**, т.е. исключена избыточность информации. Это делается не столько с целью экономии памяти, сколько для исключения возможной противоречивости хранимых данных.

Каждая таблица в реляционной БД удовлетворяет условию, в соответствии с которым в позиции на пересечении каждой строки и столбца таблицы всегда находится единственное атомарное значение, и никогда не может быть множества таких значений.

Любая таблица, удовлетворяющая этому условию, называется **нормализованной**.

Всякая нормализованная таблица автоматически считается таблицей в *первой нормальной форме*, сокращенно **1НФ**.

Таким образом, "нормализованная" и "находящаяся в 1НФ" означают одно и то же.

Однако на практике термин "нормализованная" часто используется в более узком смысле – "полностью нормализованная", который означает, что в проекте не нарушаются никакие принципы нормализации.

В дополнение к 1НФ можно определить дальнейшие уровни нормализации – *вторую нормальную форму (2НФ), третью нормальную форму (3НФ)* и т.д.

По существу, таблица находится в 2НФ, если она находится в 1НФ и удовлетворяет, кроме того, некоторому дополнительному условию.

Таблица находится в 3НФ, если она находится в 2НФ и, помимо этого, удовлетворяет еще другому дополнительному условию и т.д.

Нормализация и функциональные зависимости

За время развития технологии проектирования реляционных БД были выделены следующие нормальные формы:

- ❖- первая нормальная форма (1НФ);
- ❖- вторая нормальная форма (2НФ);
- ❖- третья нормальная форма (3НФ);
- ❖- нормальная форма Бойса-Кодда (НФБК);
- ❖- четвертая нормальная форма (4НФ);
- ❖- пятая нормальная форма, или нормальная форма проекции-соединения (5НФ).

Обычно на практике применение находят только первые четыре нормальные формы.

Нормализация, функциональные и многозначные зависимости

Теория нормализации основывается на наличии той или иной зависимости между полями таблицы.

Определены три вида таких зависимостей:

функциональные, полные функциональные и многозначные.

Нормализация и функциональные зависимости

Определение функциональной зависимости. Поле В таблицы R **функционально зависит** от поля А той же таблицы если для каждого из различных значений поля А обязательно существует только одно из различных значений поля В. Здесь допускается, что поля А и В могут быть составными.

$R.A \rightarrow R.B$ $R(A \rightarrow B)$ ФЗ($A \rightarrow B$)

Города		
Г	Город	Страна
1	Петербург	Россия
2	Резекне	Россия
3	Пушкин	Россия
4	Луга	Россия
5	Паневежис	Литва
6	Йыхви	Литва

R	
F	T
1	4
3	2
5	3
7	3
3	2
7	3

Нормализация и функциональные зависимости

ПРИМЕР

A	B
1	b
1	b
2	b
3	c
...	...

Нормализация и функциональные зависимости

Пример. Таблица **Поставщики** вида

<u>Поставщик</u>	Статус	<u>Город</u>	Страна	Адрес	Телефон
------------------	--------	--------------	--------	-------	---------

Здесь поле **Страна** функционально зависит от составного ключа (**Поставщик, Город**). Однако последняя зависимость не является функционально полной, так как **Страна** функционально зависит и от части ключа – поля **Город**.

Нормализация и функциональные зависимости

Определение полной функциональной зависимости. Поле В находится в полной функциональной зависимости от составного поля А, если оно функционально зависит от А и не зависит функционально от любого подмножества поля А.

R1		
A	B	C
1	1	1
1	2	4
2	1	2
3	2	4
3	1	2
3	4	4

R2		
A	B	C
1	1	1
1	2	4
1	1	2
3	3	4
3	3	2
3	4	4

Нормальные формы. 1НФ

Определение. Таблица находится в *первой нормальной форме (1НФ)* тогда и только тогда, когда ни одна из ее строк не содержит в любом своем поле более одного значения и ни одно из ее ключевых полей не пусто.

Нормальные формы. 2НФ

Определение. Таблица находится во *второй нормальной форме (2НФ)*, если она удовлетворяет определению 1НФ и все ее поля, не входящие в первичный ключ, связаны полной функциональной зависимостью с первичным ключом.

№ зачетной книжки	ФИО студента	Место Рождения	Дата Рождения	Курс	Средний балл
1992412-11	Карасев А.А.	г. Чита	27.08.75	1	4,5
1992412-11	Карасев А.А.	г. Чита	27.08.75	2	4,1
1992412-11	Карасев А.А.	г. Чита	27.08.75	3	5
1992412-11	Карасев А.А.	г. Чита	27.08.75	4	4,8
1992432-11	Данилов О. В.	г. Алма-Ата	27.08.75	1	4,6
1992432-11	Данилов О. В.	г. Алма-Ата	27.08.75	2	4,4
1992432-11	Данилов О. В.	г. Алма-Ата	27.08.75	3	4,2
1992432-11	Данилов О. В.	г. Алма-Ата	27.08.75	4	4,3
1992432-12	Раевский А. И.	г. Бишкек	20.05.75	1	3,8
.
1992432-12	Раевский А. И.	г. Бишкек	20.05.75	4	5

УСПЕВАЕМОСТЬ (№ ЗАЧЕТНОЙ КНИЖКИ, ФИО СТУДЕНТА, МЕСТО РОЖДЕНИЯ, ДАТА РОЖДЕНИЯ, КУРС, СРЕДНИЙ БАЛЛ)

Нормальные формы. 2НФ

Функциональные зависимости:

№ ЗАЧЕТНОЙ КНИЖКИ -> ФИО СТУДЕНТА;

№ ЗАЧЕТНОЙ КНИЖКИ -> МЕСТО РОЖДЕНИЯ;

№ ЗАЧЕТНОЙ КНИЖКИ -> ДАТА РОЖДЕНИЯ;

№ ЗАЧЕТНОЙ КНИЖКИ, КУРС -> СРЕДНИЙ БАЛЛ.

Атрибуты ФИО СТУДЕНТА, МЕСТО РОЖДЕНИЯ и ДАТА РОЖДЕНИЯ функционально зависят только от части первичного ключа - № ЗАЧЕТНОЙ КНИЖКИ.

При работе с таким ненормализованным отношением невозможно обеспечить корректную работу при выполнении операций **вставки, удаления и обновления** строк.

Эти недостатки реляционных отношений устраняются путем нормализации, поскольку отношение УСПЕВАЕМОСТЬ не удовлетворяет требованиям 2НФ.

Нормальные формы. 2НФ

Приведение данного отношения к 2НФ заключается в его разбиении (*декомпозиции*) на два отношения, удовлетворяющих соответствующим требованиям нормализации. Можно произвести следующую декомпозицию отношения УСПЕВАЕМОСТЬ в два отношения СТУДЕНТЫ и УСПЕВАЕМОСТЬ СТУДЕНТОВ:

СТУДЕНТЫ (№ ЗАЧЕТНОЙ КНИЖКИ, ФИО СТУДЕНТА, МЕСТО РОЖДЕНИЯ, ДАТА РОЖДЕНИЯ)

УСПЕВАЕМОСТЬ (№ ЗАЧЕТНОЙ КНИЖКИ, КУРС, СРЕДНИЙ БАЛЛ)

Функциональные зависимости в данных отношениях?

Каждое из этих двух отношений находится в 2НФ, и в них устранены отмеченные выше недостатки.

Нормальные формы. 3НФ

Добавим в отношении СТУДЕНТЫ два атрибута: № ГРУППЫ и ФИО КУРАТОРА:

СТУДЕНТЫ (№ ЗАЧЕТНОЙ КНИЖКИ, ФИО СТУДЕНТА, МЕСТО РОЖДЕНИЯ, ДАТА РОЖДЕНИЯ, № ГРУППЫ, ФИО КУРАТОРА)

Здесь появилась новая функциональная зависимость:
№ ГРУППЫ -> ФИО КУРАТОРА

В результате возникли проблемы при выполнении операций **вставки, удаления и обновления** строк.

Их можно устранить путем дальнейшей нормализации.

Нормальные формы. 3НФ

Определение. Таблица находится в *третьей нормальной форме (3НФ)*, если она удовлетворяет определению 2НФ и не одно из ее неключевых полей не зависит функционально от любого другого неключевого поля.

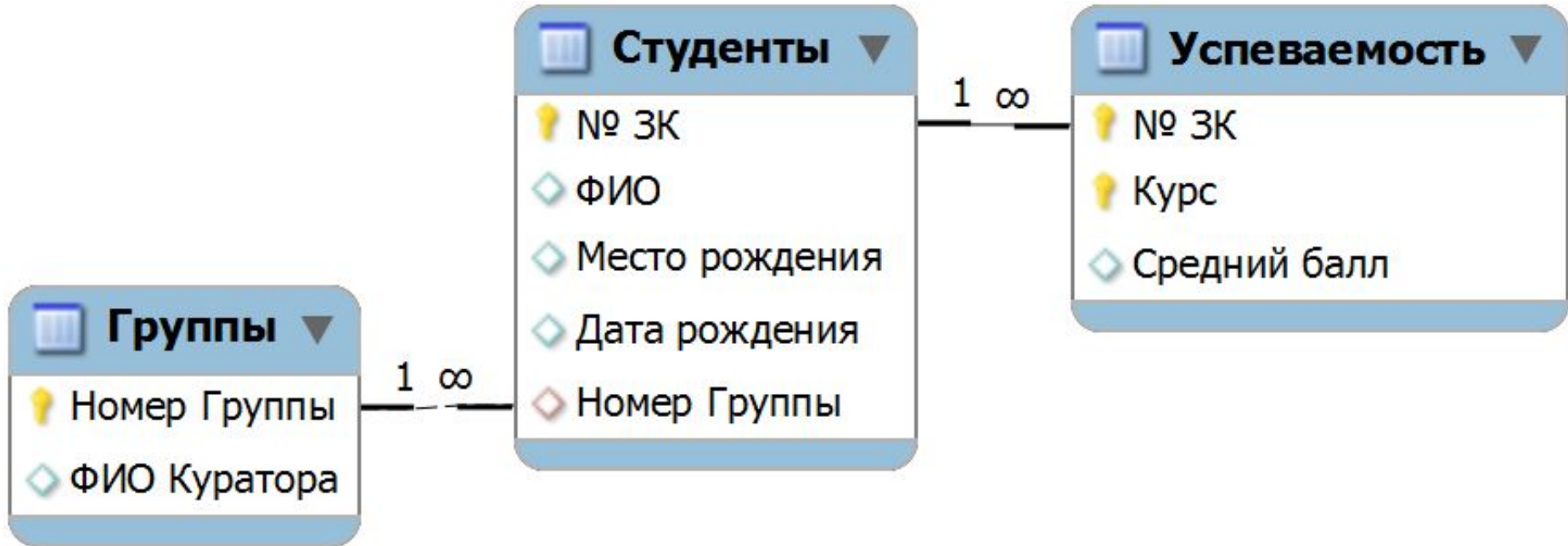
Произведем декомпозицию отношения СТУДЕНТЫ в два отношения СТУДЕНТЫ и ГРУППЫ:

СТУДЕНТЫ (№ ЗАЧЕТНОЙ КНИЖКИ, ФИО СТУДЕНТА, МЕСТО РОЖДЕНИЯ, ДАТА РОЖДЕНИЯ, № ГРУППЫ)

ГРУППЫ (№ ГРУППЫ, ФИО КУРАТОРА)

Результирующие отношения находятся в 3НФ и свободны от вышеуказанных недостатков.

Нормальные формы. 3НФ



Указать типы сущностей самостоятельно.

Нормальные формы. НФБК

База данных ПАНСИОН:

Блюда (БЛ, Блюдо, В, О, Выход, Труд)

Функциональные зависимости:

БЛ->Блюдо

Блюдо->В, Блюдо->О, Блюдо->Выход, Блюдо->Труд

Бл->В, Бл->О, Бл->Выход, Бл->Труд

Т.е. таблица **Блюда** **не находится** в 3НФ. Следовательно, необходима декомпозиция таблиц вида:

Блюда(БЛ, Блюдо),

Вид_блюд(БЛ, В);

и т.д.

Нормальные формы. НФБК

Определение. Таблица находится в **нормальной форме Бойса-Кодда** (НФБК), если любая функциональная зависимость между его полями сводится к полной функциональной зависимости от **возможного** ключа.

Для упрощения нормализации таблиц с искусственными цифровыми ключами целесообразно использовать следующую рекомендацию.

Рекомендация. При проведении нормализации таблиц, в которые введены цифровые (или другие) заменители первичных и внешних ключей, следует хотя бы мысленно подменять их на исходные ключи, а после окончания нормализации снова восстанавливать.

Процедура нормализации

Эта процедура **основывается** на том, что единственными функциональными зависимостями в любой таблице должны быть зависимости вида $K \rightarrow F$, где K – первичный ключ, а F – некоторое другое поле.

Цель нормализации состоит в том, чтобы избавиться от всех этих "других" функциональных зависимостей, т.е. таких, которые имеют иной вид, чем $K \rightarrow F$.

Если подменить на время нормализации коды первичных (внешних) ключей на исходные ключи, то, по существу, следует рассмотреть лишь два случая:

Процедура нормализации

1. Таблица имеет составной первичный ключ вида, скажем, (K1, K2), и включает также поле F, которое функционально зависит от части этого ключа, например, от K2, но не от полного ключа.

В этом случае рекомендуется сформировать другую таблицу, содержащую K2 и F (первичный ключ – K2), и удалить F из первоначальной таблицы:

Заменить T(K1,K2,F), первичный ключ (K1,K2), ФЗ(K2->F)
на ???

Процедура нормализации

Заменить $T(K1, K2, F)$, первичный ключ $(K1, K2)$,
 $\Phi_3(K2 \rightarrow F)$

на $T1(K1, K2)$, первичный ключ $(K1, K2)$,

и $T2(K2, F)$, первичный ключ $K2$.

Процедура нормализации

2. Таблица имеет первичный (возможный) ключ K , не являющееся возможным ключом поле $F1$, которое, конечно, функционально зависит от K , и другое неключевое поле $F2$, которое функционально зависит от $F1$.

Решение здесь, по существу, то же самое, что и прежде – формируется другая таблица, содержащая $F1$ и $F2$, с первичным ключом $F1$, и $F2$ удаляется из первоначальной таблицы:

Заменить $T(K, F1, F2)$, первичный ключ K , $\Phi_3(F1 \rightarrow F2)$
на ???

Процедура нормализации

Заменить $T(K, F1, F2)$, первичный ключ K ,
 $\Phi_3(F1 \rightarrow F2)$

на $T1(K, F1)$, первичный ключ K ,

и $T2(F1, F2)$, первичный ключ $F1$.

Примеры нормализации

Постановка задачи. Дано отношение.

- 1) определить все функциональные зависимости отношения и первичный ключ отношения;
- 2) привести отношение к 3НФ, указать первичные и внешние ключи полученных отношений, построить схему "Таблица-Связь".

Пример 1

НаименованиеВуза	Команда	Приз
КрасГУ	МатФак	Торт
КрасГУ	ЭконФак	Торт
НГУ	ФилФак	Арбуз
НГУ	МатФак	Бананы

Пример 2

Наименование эмитента	Тип ЦБ	Дата Эмиссии	Номинальная стоимость
ОАО "КрАЗ"	акция привилегированная	23.06.2009	200 руб.
ОАО "КрАЗ"	акция обыкновенная	23.06.2009	200 руб.
ТОО "Искра"	акция привилегированная	23.06.2009	200 руб.
ТОО "Искра"	акция обыкновенная	20.06.2009	200 руб.

Пример 3

Наименование Эмитента	Тип ЦБ	Дата Эмиссии	Номинальная Стоимость
ОАО “КрАЗ”	акция обыкновенная	23.06.2009	100 руб.
ОАО “КрАЗ”	акция обыкновенная	23.06.2009	200 руб.
ТОО “Искра”	акция привилегированная	20.06.2009	500 руб.
ТОО “Искра”	акция привилегированная	23.06.2009	500 руб.

Нормальные формы более высоких порядков

Пример. Требуется учитывать данные об абитуриентах, поступающих в ВУЗ. При анализе предметной области были выделены следующие требования:

1. Каждый абитуриент имеет право сдавать экзамены на несколько факультетов одновременно.
2. Каждый факультет имеет свой список сдаваемых предметов.
3. Один и тот же предмет может сдаваться на нескольких факультетах.
4. Абитуриент обязан сдавать все предметы, указанные для факультета, на который он поступает, несмотря на то, что он, может быть, уже сдавал такие же предметы на другом факультете.

Попытаемся хранить данные в одном отношении "Абитуриенты – Факультеты - Предметы":

Нормальные формы более высоких порядков

Отношение “Абитуриенты – Факультеты – Предметы”

Абитуриент	Факультет	Предмет
Иванов	Математический	Математика
Иванов	Математический	Информатика
Иванов	Физический	Математика
Иванов	Физический	Физика
Петров	Математический	Математика
Петров	Математический	Информатика

АФП		
НА	НФ	НП
1	1	1
1	1	2
1	2	1
1	2	3
2	1	1
2	1	2

АБИТУРИЕНТЫ	
НА	Абитуриент
1	Иванов
2	Петров

ФАКУЛЬТЕТЫ	
НФ	Факультет
1	Математический
2	Физический

ПРЕДМЕТЫ	
НП	Предмет
1	Математика
2	Информатика
3	Физика

Нормальные формы более высоких порядков

Определение многозначной зависимости. Поле А **многозначно определяет** поле В той же таблицы, если для каждого значения поля А существует конечное множество соответствующих значений В.

ФЗ(Факультет ->> Абитуриент)

ФЗ(Факультет ->> Предмет)

НФ	НА
1	1
2	1
1	2

НФ	НП
1	1
1	2
2	1
2	3

Недостатки нормализации