



БАЗЫ ДАННЫХ И ЭКСПЕРТНЫЕ СИСТЕМЫ

ФИО преподавателя: Тараканов О.В., канд. техн. наук, доцент
e-mail: tarakanov@mirea.ru

НОРМАЛИЗАЦИЯ ОТНОШЕНИЙ

Учебные вопросы:

1. Функциональные, многозначные и зависимости соединения
2. Нормальные формы отношений
3. Доменно-ключевая нормальная форма

Функциональные, многозначные и зависимости соединения

Понятие функциональной зависимости

Функция – есть закон f однозначного соответствия одного числа y (значения функции) другому числу x (аргументу функции)

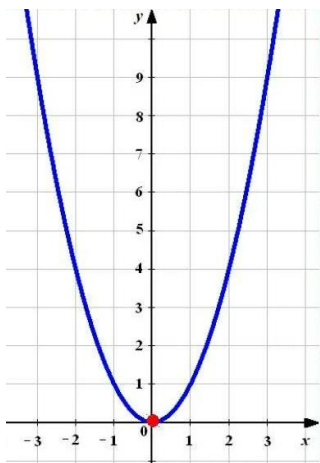
$y = f(x)$

$y = x^2$

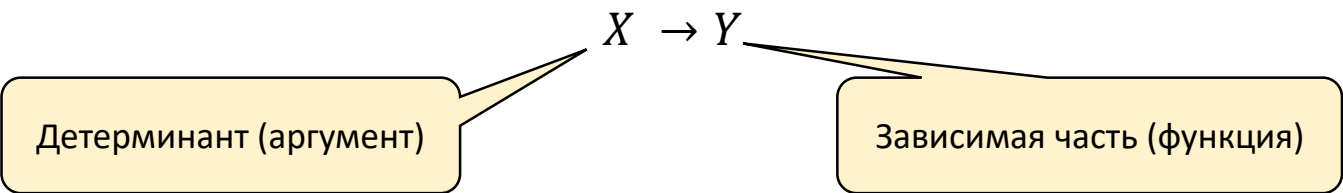
$y = \sin x$

$y = a \cdot x + b$

$y = \log_2 x$



Если переменные x и y определить на любых стандартных доменах (STRING, NUMBER, DATETIME, BLOB), то определение будет справедливым для **функциональной зависимости** зависимой части Y от детерминанта X .



Фамилия, имя, отчество → Дата рождения
Наименование университета → Почтовый адрес

Функциональная зависимость – есть закон однозначного соответствия факта – порции данных Y (зависимая часть) другому факту – порции данных X (детерминант)

Теорема о функциональной зависимости

Если исследуется атрибут X отношения R на предмет функциональной зависимости от него атрибута Y , то справедливо следующее утверждение – при изменении значения атрибута X значение атрибута Y может измениться или останется прежним; если изменить значение атрибута Y , то при существовании функциональной зависимости его от детерминанта X , значение последнего обязательно изменяется, иначе функциональная зависимость отсутствует.

Ф.И.О.	Дата рождения
Иванов М.П.	21.03.1982
Петров В.А.	21.03.1982
Сидоров А.И.	16.09.1990

Ф. И. О. → Дата рождения

Ф.И.О.	Должность	Дата назначения
Петров В.А.	механик	20.05.2011
Петров В.А.	инженер	18.03.2018

Ф. И. О. ✗ Должность

Функциональные, многозначные и зависимости соединения

Основные свойства функциональных зависимостей

Нетривиальность

Если существует функциональная зависимость $X \rightarrow (Y \cap Z)$, то вытекающие из нее ФЗ $X \rightarrow Y$ и $X \rightarrow Z$ **тривиальные**

Если существует функциональная зависимость $X \rightarrow (Y \cap X)$, то она **тривиальная**

Если существует функциональная зависимость $(X \cap Y) \rightarrow X$, то она **тривиальная**

Функциональная зависимость $X \rightarrow X$ **тривиальная**

Полнота

Если существует функциональная зависимость $(X \cap Y) \rightarrow Z$ и существует хотя бы одна из вытекающих из нее ФЗ $X \rightarrow Z$ и $Y \rightarrow Z$, то она **не полная**

Никакое подмножество атрибутов детерминанта не может обеспечивать функциональную зависимость с той же зависимой частью!

Транзитивность

Если существует функциональная зависимость $X \rightarrow Y$ и существует ФЗ $Y \rightarrow Z$, то функциональная зависимость $X \rightarrow Z$ является **транзитивной**

Инверсность (обратность)

Если существует функциональная зависимость $X \rightarrow Y$ и существует ФЗ $Y \rightarrow X$, то она является **инверсной (обратной)**

Функциональные, многозначные и зависимости соединения

Основные свойства функциональных зависимостей

Полнота

Если существует функциональная зависимость $X \rightarrow Y$, то существует и ФЗ, где детерминант и зависимая часть соединены с одним и тем же третьим атрибутом: $(X \cup Z) \rightarrow (Y \cup Z)$

Соединение атрибутов представляет собой их декартово (прямое) произведение, следовательно ФЗ **дистрибутивна (полна)**!

Декомпозируемость

Если существует функциональная зависимость $X \rightarrow Y$, а атрибут Y не атомарен и является композицией атрибутов $K \cup L$, то ФЗ $X \rightarrow Y$ может быть приведена к виду $(X \rightarrow K) \cap (X \rightarrow L)$ – композиции функциональных зависимостей, следовательно, исходная функциональная зависимость **декомпозируемая**

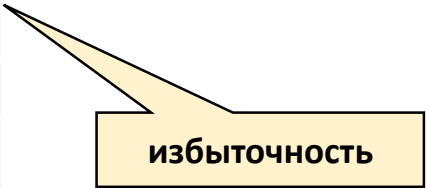
Процедура декомпозиции обратима, что означает существование возможности композиции и декомпозиции отношений на уровне атрибутов с учетом имеющихся декомпозируемых функциональных зависимостей!

Многозначность

Если атрибут X зависит от атрибута Y ТОЛЬКО на множестве значений, но не зависит функционально (на единичных значениях), то это означает, что между X и Y имеется **многозначная** зависимость: $Y \twoheadrightarrow X$

ЛЕКТОР → УЧЕБНАЯ ДИСЦИПЛИНА

Лектор	Учебная дисциплина
Самохвалов В.В.	Базы данных и экспертные системы
Васильев В.С.	Базы данных и экспертные системы
Самохвалов В.В.	Методы искусственного интеллекта
Константиновский М.А.	Методы искусственного интеллекта
Афанасьева А.М.	Высшая математика



Функциональные, многозначные и зависимости соединения

Понятие многозначной зависимости

Преподаватель

Ф.И.О.

Адрес

Учебная дисциплина

Потенциальный ключ

Ф.И.О. = {'Иванов И.П.'}

Адрес = {'ул. Гоголя, 12, кв. 6', 'ул. Красная, 32'}

Учебная дисциплина = {'Информатика', 'Базы данных', 'Программирование'}

Ф.И.О.	Адрес	Учебная дисциплина
Иванов И.П.	ул. Гоголя,12, кв.6	Базы данных
Иванов И.П.	ул. Гоголя,12, кв.6	Информатика
Иванов И.П.	ул. Гоголя,12, кв.6	Программирование
Иванов И.П.	ул. Красная, 32,	Базы данных
Иванов И.П.	ул. Красная, 32,	Информатика
Иванов И.П.	ул. Красная, 32	Программирование

} |Ф. И. О| × |Адрес| × |Учебная дисциплина| = 6

Наблюдаемые зависимости: Адрес, Учебная дисциплина → Ф.И.О.
Ф.И.О., Адрес → Учебная дисциплина
Ф.И.О., Учебная дисциплина → Адрес

Многозначным зависимостям присущи тривиальность, транзитивность и дополненность

Тривиальные многозначные зависимости – зависящая часть является подмножеством детерминанта или совпадает с ним по составу атрибутов

$A, B, C \twoheadrightarrow A, C$
 $A, B, C \twoheadrightarrow A, B, C$

Транзитивная многозначная зависимость

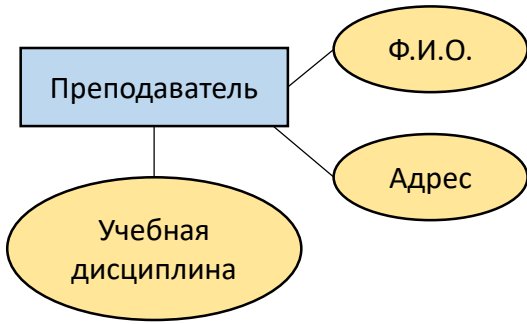
$A_1, A_2, \dots, A_n \twoheadrightarrow C_1, C_2, \dots, C_k$
если
 $\exists A_1, A_2, \dots, A_n \twoheadrightarrow B_1, B_2, \dots, B_m$
 $\exists B_1, B_2, \dots, B_m \twoheadrightarrow C_1, C_2, \dots, C_k$

Дополнение C многозначной зависимости A до B

$\exists A_1, A_2, \dots, A_n \twoheadrightarrow B_1, B_2, \dots, B_m$
 $A_1, A_2, \dots, A_n \twoheadrightarrow C_1, C_2, \dots, C_k \text{ true}$
 $\{B\} \cap \{C\} = \emptyset \text{ и } \{A\} \cap \{C\} = \emptyset$

Функциональные, многозначные и зависимости соединения

Зависимость соединения



Ф.И.О.	Должность	Учебная дисциплина
Иванов И.П.	профессор	Информатика
Орлова А.В.	старший преподаватель	Информатика

Декомпозиция

$$R = R_1 \cup R_2 \cup \dots \cup R_n$$

Ф.И.О.	Учебная дисциплина
Иванов И.П.	Информатика
Орлова А.В.	Информатика

Должность	Учебная дисциплина
профессор	Информатика
старший преподаватель	Информатика

Композиция – соединение

$$R_1 \text{ JOIN } R_2 \text{ JOIN } \dots \text{ JOIN } R_n = R$$

Ф.И.О.	Должность	Учебная дисциплина
Иванов И.П.	профессор	Информатика
Иванов И.П.	старший преподаватель	Информатика
Орлова А.В.	профессор	Информатика
Орлова А.В.	старший преподаватель	Информатика

Ложные кортежи – потеря целостности

Нормальные формы отношений

Цель нормализации – преобразовать структуру отношения (совокупность атрибутов) таким образом, чтобы избавить его от избыточности и условий возникновения аномалий обновления. Нормализация – основное средство обеспечения целостности базы данных на уровне каждого отдельно взятого отношения.

Способ нормализации – последовательное применение к отношению (совокупности его атрибутов) правил (требований нормальной формы), обеспечивающих организацию эффективных функциональных зависимостей между атрибутами, и налагающих на необходимые из них специальные ограничения (CONSTRAINTS – ограничения целостности).

Процесс нормализации – последовательное преобразование отношения через первую нормальную форму (1НФ), 2НФ, 3НФ, нормальную форму Бойса-Кодда (НФБК), 4НФ, 5НФ (проекционно-соединительную нормальную форму) к доменно-ключевой (ДКНФ).

1НФ → 2НФ → 3НФ → НФБК → 4НФ → 5НФ → ДКНФ

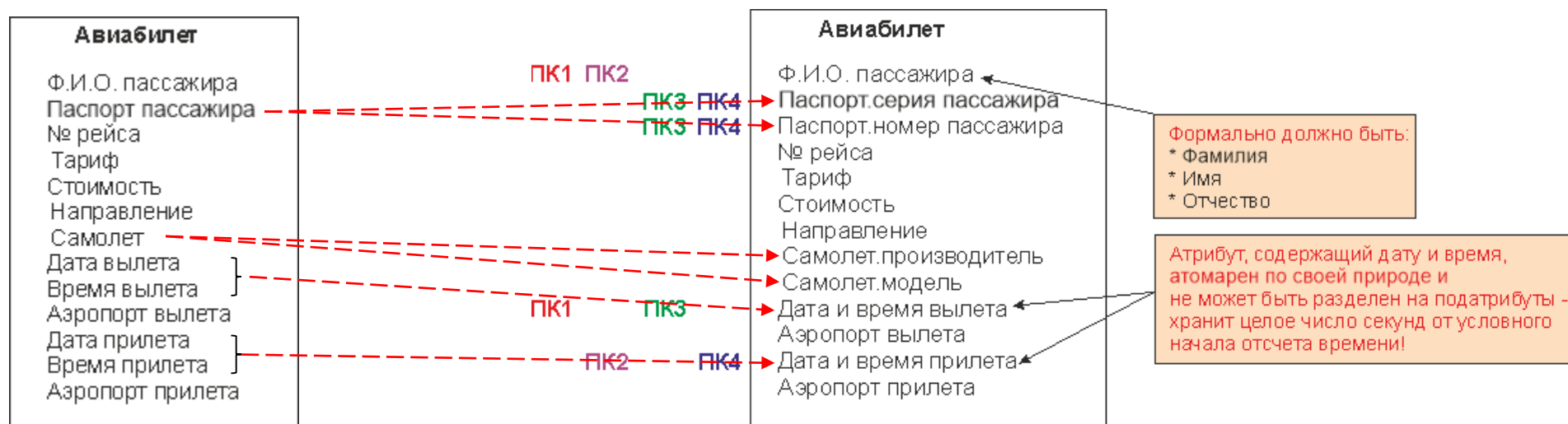
Нормальные формы отношений

Первая нормальная форма – 1НФ

Отношение (сущность) называется нормализованным или приведенным к **первой нормальной форме (1НФ)**, если все его атрибуты простые (атомарные) и определены все возможные потенциальные ключи, однозначно идентифицирующие каждый экземпляр.

{ Каждый атрибут атомарен
{ \exists {совокупность потенциальных ключей} – полная

⟨Атрибут⟩ – **атомарный** (простой) атрибут, не может быть разделен на атрибуты без потери смысла.



Потенциальный ключ (возможный ключ, *candidate key*) – атрибут или совокупность атрибутов, которые своими **значениями** однозначно отличают кортеж от остальных. Именно этот тип ключа однозначно идентифицирует каждый кортеж отношения.

Каждый потенциальный ключ должен быть минимальным (нетривиальным). Это означает, что если из него исключить любой, входящий в его состав атрибут, то он утрачивает свойство идентификации кортежа.

Нормальные формы отношений

Вторая нормальная форма – 2НФ

Отношение находится во **второй нормальной форме (2НФ)**, если оно приведено к 1НФ и каждый не ключевой атрибут **функционально полно** зависит от каждого потенциального ключа.

Не ключевой атрибут - это атрибут, который не входит в состав никакого потенциального ключа (не входит в суперключ отношения).

Атрибут функционально полно зависит от потенциального ключа тогда, когда он, вообще от него зависит, и не зависит ни от какой его составной части (любого подмножества атрибутов потенциального ключа). Полная функциональная зависимость предопределяет, что при исключении из потенциального ключа любого атрибута, не ключевой атрибут более не зависит от преобразованного таким образом потенциального ключа.

Теорема о функциональной зависимости

Если изменить значение детерминанта (идентификатора экземпляра объекта учета), то зависимая часть может измениться или нет. При изменении значения зависимой части, при наличии функциональной зависимости, значение детерминанта должно также измениться. Если существует ситуация, при которой значение детерминанта не изменяется, то это означает, что искомая функциональная зависимость не существует.

$$\begin{cases} \exists 1НФ \\ \forall НКА = ФПЗ(\text{каждого ПК}) \end{cases}$$

ПК1 ПК2

ПК3 ПК4
ПК3 ПК4

ПК1

ПК3

ПК2

ПК4

Авиабилет

Ф.И.О. пассажира
Паспорт.серия пассажира
Паспорт.номер пассажира
№ рейса
Тариф
Стоимость
Направление
~~Самолет.производитель~~
~~Самолет.модель~~
Дата и время вылета
Аэропорт вылета
Дата и время прилета
Аэропорт прилета

Нормальные формы отношений

Третья нормальная форма – 3НФ

Отношение считается приведенным к третьей нормальной форме (3НФ) тогда и только тогда, когда оно уже приведено ко второй (2НФ) и все не ключевые (non-key) атрибуты функционально нетранзитивно зависят от всех потенциальных ключей.

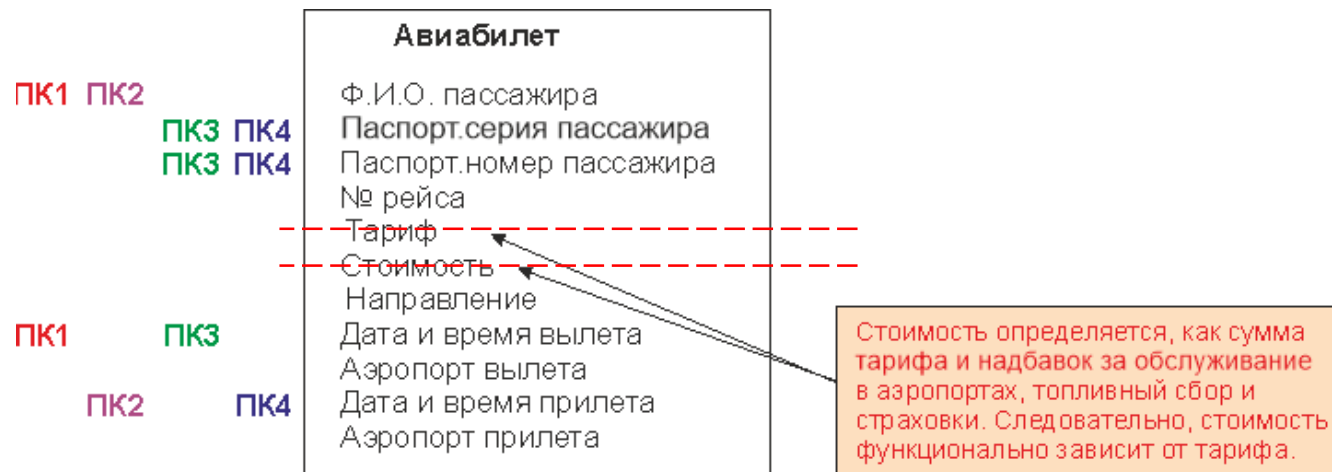
$$\begin{cases} \exists 2НФ \\ \forall НКА = ФНТЗ(\text{каждого ПК}) \end{cases}$$

Если не ключевой атрибут функционально зависит от потенциального ключа, а другой не ключевой атрибут функционально зависит от первого не ключевого атрибута, то это означает, что второй не ключевой атрибут функционально зависит от того же потенциального ключа. Атрибут **функционально нетранзитивно** зависит от потенциального ключа тогда, когда он, вообще от него зависит, и не зависит ни от какого другого не ключевого атрибута.

Каждая пара не ключевых атрибутов, состоящих в функциональной зависимости друг от друга (не обязательно взаимно), должна быть исключена из отношения. Все исключаемые из отношения атрибуты помещаются в новые отношения, подлежащие нормализации.

Лайфхак (инженерный прием):

Сформировать все возможные пары из не ключевых атрибутов. Выявить функциональную зависимость между атрибутами каждой из пар, по очереди считая детерминантом то один из них, то другой. Все пары, в которых обнаружена функциональная зависимость удалить из отношения.



Нормальные формы отношений

Нормальная форма Бойса-Кодда – НФБК

Отношение считается приведенным к нормальной форме Бойса - Кодда (НФБК) тогда и только тогда, когда оно уже приведено к 3НФ и каждый его детерминант является только потенциальным ключом.

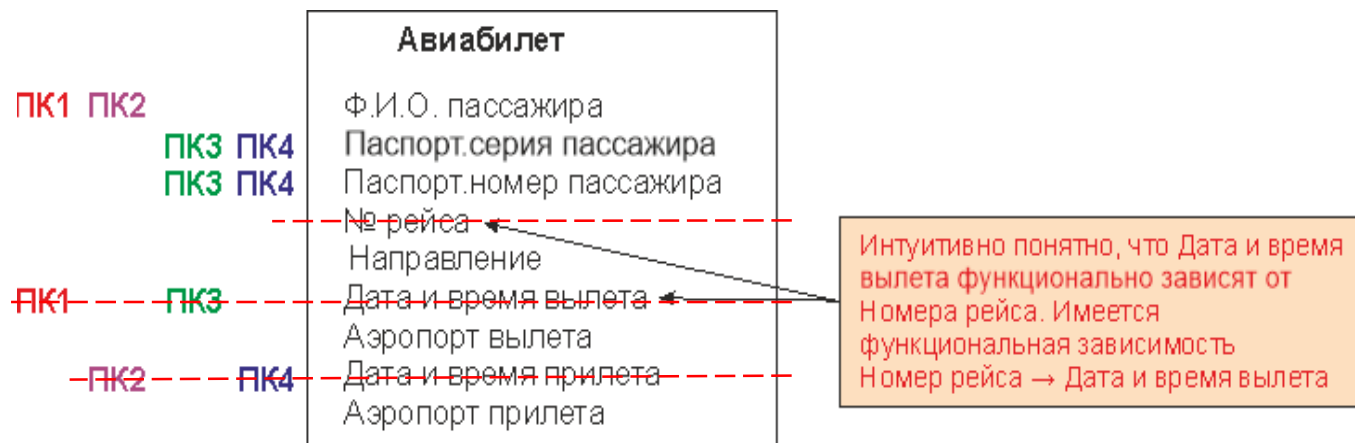
$$\begin{cases} \exists 3НФ \\ \forall \text{ детерминант} \in \{ПК_i\} \end{cases}$$

Если атрибут составного потенциального ключа функционально зависит от не ключевого атрибута, то такая зависимость называется инверсной (обратной) и она должна быть исключена из отношения, приведенного в НФБК.

Каждая пара не ключевой атрибут – потенциальный ключ, где детерминантом является не ключевой атрибут, должна быть исключена из отношения. Все исключаемые из отношения атрибуты помещаются в новые отношения, подлежащие нормализации.

Лайфхак (инженерный прием):

Для достижения НФБК необходимо составить все возможные пары атрибутов, где в качестве детерминанта рассматривается не ключевой атрибут, а в качестве зависимой части атрибут составного потенциального ключа. Проверить наличие функциональной зависимости во всех парах. Пары, в которых будет обнаружена функциональная зависимость исключить из отношения.



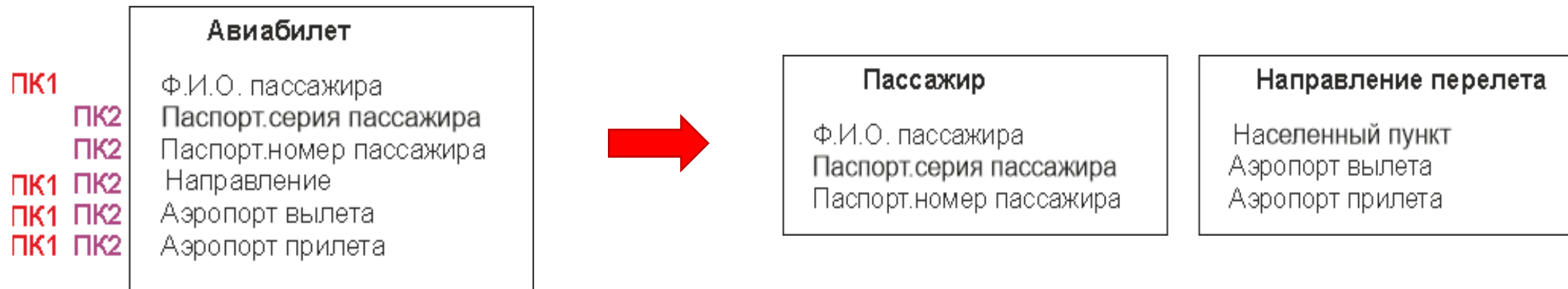
Нормальные формы отношений

Четвертая нормальная форма – 4НФ

Отношение считается приведенным к четвертой нормальной форме тогда и только тогда, когда оно уже приведено к НФБК и не содержит нетривиальных многозначных зависимостей.

{
 ∃ НФБК
 ∄ нетривиальных ФЗ

Многократное повторение кортежей с одинаковым значением Ф.И.О. пассажира (а также номером и серией паспорта) является следствием многозначной зависимости атрибутов НАПРАВЛЕНИЕ, АЭРОПОРТ ВЫЛЕТА и АЭРОПОРТ ПРИЛЕТА от атрибутов Ф.И.О. и ПАСПОРТ.СЕРИЯ Пассажира совместно с ПАСПОРТ.НОМЕР Пассажира.



Лайфхак (инженерный прием):

Для достижения 4НФ необходимо по правилам методологии "сущность-связь" уточнить сколько объектов учета одновременно абстрагировано в нормализуемом отношении. Если найдено более одного объекта учета, то необходимо отношение декомпозировать на подотношения в соответствие с найденными объектами учета.

Нормальные формы отношений

Пятая нормальная форма – 5НФ

Отношение считается приведенным к пятой нормальной форме (5НФ) тогда и только тогда, когда оно уже приведено к 4НФ и не содержит зависимости соединения.

{
 ∃ НФБК
 ∄ зависимостей соединения

Направление перелета

Направление	Аэропорт вылета	Аэропорт прилета
Москва	ULV	DME
Москва	ULY	SVO

Вылет

Направление	Аэропорт вылета
Москва	ULV
Москва	ULY

JOIN

Прилет

Направление	Аэропорт прилета
Москва	DME
Москва	SVO

Вылет JOIN Прилет

Направление	Аэропорт вылета	Аэропорт прилета
Москва	ULV	DME
Москва	ULV	SVO
Москва	ULY	DME
Москва	ULY	SVO

Зависимость соединения порождает эффект декомпозиции с соединением (JOIN) с потерей целостности результата (возникновение ложных кортежей).

Решение задачи декомпозиции должно выполняться с учетом исключения в новых отношениях зависимости соединения. Это возможно в том случае, когда свойства одного объекта учета сосредоточены в единственном отношении, а не "размазаны" по нескольким.

Лайфхак (инженерный прием):

Для достижения 5НФ необходимо проверить, чтобы все атрибуты, абстрагирующие значимые свойства объекта учета были сосредоточены в рамках единственного отношения, абстрагирующего сам этот объект учета. Если будет обнаружено, что значимые свойства объекта учета абстрагируются атрибутами любых других отношений, то необходимо их перенести в "родное" отношение.

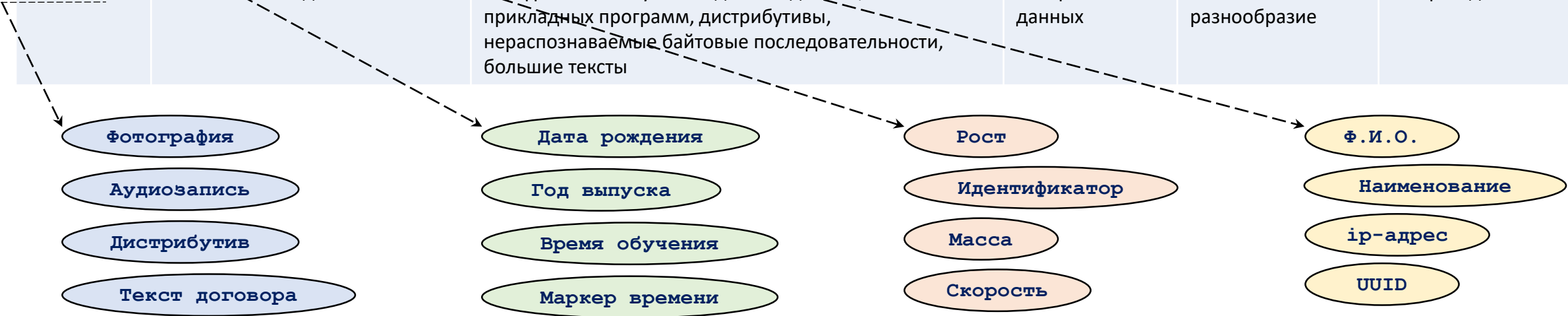
СОЕДИНЕНИЕ ДОЛЖНО ПРОИЗВОДИТЬСЯ НАД ОБЪЕКТАМИ УЧЕТА, А НЕ НА УРОВНЕ ОТДЕЛЬНЫХ ИХ СВОЙСТВ!

Доменно-ключевая нормальная форма

Отношение находится в **доменно-ключевой нормальной форме (ДКНФ)** тогда и только тогда, когда оно приведено к 5НФ и каждое наложенное на неё ограничение является логическим следствием ограничений доменов и ограничений ключей [Wikipedia].

Отношение считается приведенным к **доменно-ключевой нормальной форме (ДКНФ)** тогда и только тогда, когда оно уже приведено к 5НФ (ПСНФ), каждый его атрибут определен на одном из стандартных доменов (STRING, NUMBER, DATETIME, BLOB) и назначен первичный ключ.

Домен	Характеристика	Предназначение	Объем памяти	Разночтение	Пригодность для РК
STRING	Статический или динамический символьный массив	Строки, символы, UUID, адреса, булевы значения	1 – 2 байта на один символ	Аномалии вставки и обновления	низкая
NUMBER	Помеченная область памяти	Числа	Обычно 4 байта	Практически невозможно	наивысшая
DATETIME	Большое целое число	Дата, время, часовой пояс, интервал дат, интервал времени	4 – 16 байт	Практически невозможно	высокая
BLOB	Байтовая последовательность	Оцифрованные мультимедийные данные, пакеты прикладных программ, дистрибутивы, нераспознаваемые байтовые последовательности, большие тексты	Вне файла данных	Наибольшее разнообразие	не пригодно



Доменно-ключевая нормальная форма

Первичный ключ (англ. *primary key*) — в реляционной модели данных один из потенциальных ключей отношения, выбранный в качестве основного ключа (или ключа по умолчанию).

$PK \in \{ПК_i\}$

Требования к первичному ключу:

компактность

защищенность от аномалий вставки

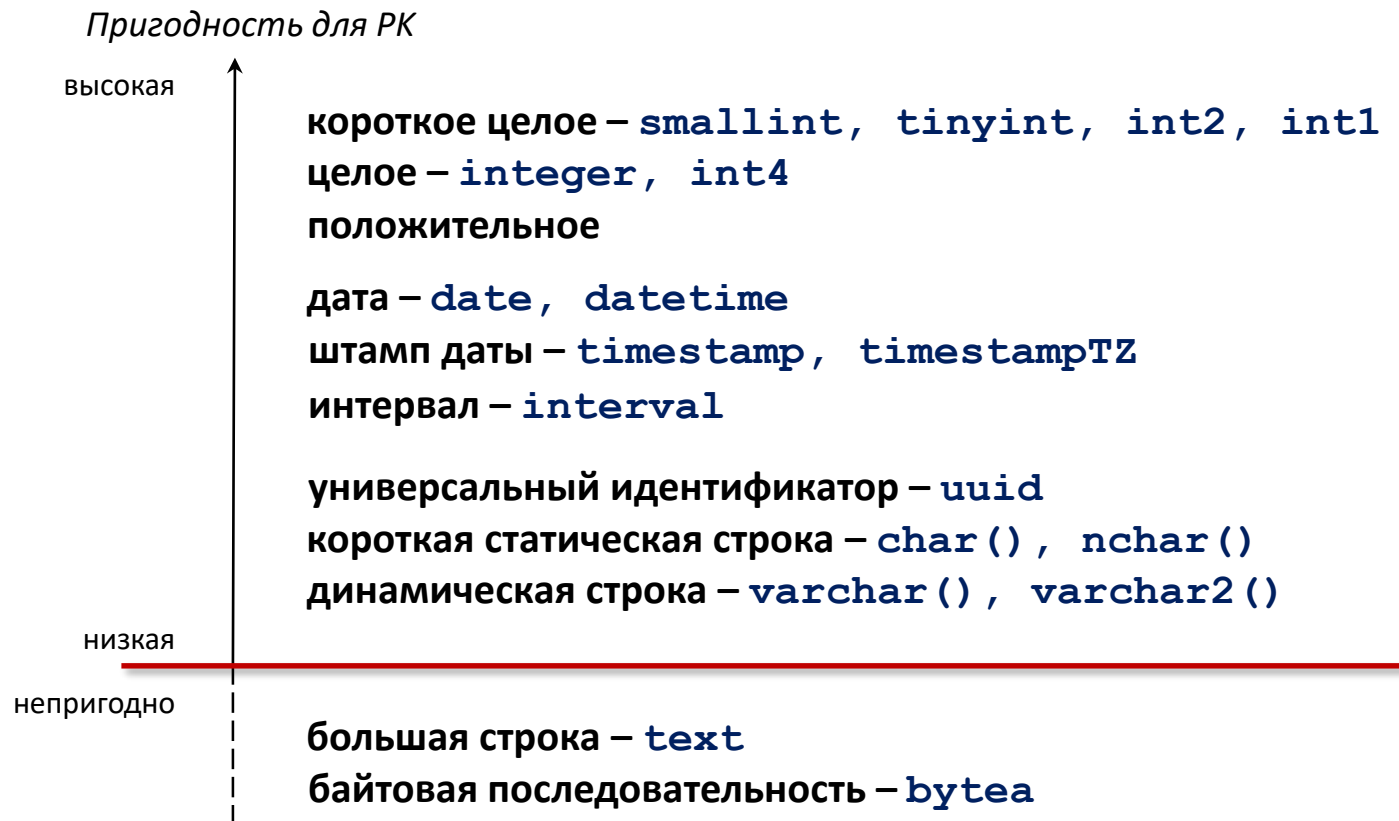
малая емкость

Лайфхак (инженерный прием):

в случае, если в множестве потенциальных ключей нет пригодного для объявления его первичным, то следует создать суррогатный атрибут со свойствами потенциального ключа и объявить его первичным, например ИДЕНТИФИКАТОР (ID).

Лайфхак (инженерный прием):

Для исключения сложностей нормализации, получения максимально качественной заготовки структуры базы данных, зачастую не требующей преобразований на уровне НФБК, 4НФ и 5НФ, необходимо выполнить качественное информационное моделирование, где безошибочно определить все объекты учета, их значимые свойства, абстрагировать их отношениями и атрибутами соответственно. Связать отношения между собой, разрешить связи типа "многие-ко-многим". Полученные отношения подвергнуть нормализации. Сформированная конструкция представляет собой логическую структуру базы данных, которую можно начинать преобразовывать в физическую на платформе выбранной СУБД.



Литература

1. **Кайт, Т., Кун, Д.** Oracle для профессионалов: архитектура и методики программирования, 3-е изд.: Пер. с англ. – Москва: ООО "ИД Вильямс", 2016. – 960 с.
2. **Гарсиа-Молина, Г.** Системы баз данных. Полный курс: пер. с англ. / Гарсиа-Молина – Москва : Издательский дом "Вильямс", 2003. – 1088 с.
3. **Конноли, Т.** Базы данных: проектирование, реализация и сопровождение. Теория и практика: учебное пособие / Т. Конноли, К. Бегг, А. Страчан –2-е изд. : пер. с англ. : – Москва : Издательский дом "Вильямс", 2000. – 1120 с.
4. **Мейер, Д.** Теория реляционных баз данных / Д. Мейер – Москва: Мир, 1987 г.
5. **Дейт, К. Дж.** Введение в системы баз данных: пер. с англ. – 7-е изд. / К. Дж. Дейт. – Москва: Издательский дом "Вильямс", 2001. – 1072 с.
6. **Кузнецов, С. Д.** Основы баз данных: курс лекций : учеб. пособие для студентов вузов, обучающихся по специальностям в области информационных технологий / С. Д. Кузнецов. – Москва: Интернет-ун-т ин-форм. технологий, 2005. – 488 с.
7. **Тараканов, О. В., Паршенкова, Ю. А., Дементьев, А. Н., Конышев, М. Ю., Смирнов, С. В.** Системы баз данных: организация, инженерия, ведение / Под ред. О. В. Тараканова. – Москва: РТУ – МИРЭА, 2023. – 335 с.
8. **Бойко, В. В.,** Проектирование баз данных информационных систем / В. В. Бойко, В. М. Савинков. – Москва: Финансы и статистика, 1989.
9. **Смирнов С. Н. Задворьев И.С.** Работаем с ORACLE.: Учебное пособие/2-е изд., испр. и доп. – М: Гелиос АРВ, 2002 г. – 496 с.
10. **Фейерштейн, С., Прибыл, Б.,** Oracle PL/SQL. Для профессионалов. 6-е изд. – Санкт-Петербург: Питер, 2015. – 1024 с.
11. **Маклаков, С. В.,** BPwin и ERwin. CASE-средства разработки информационных систем / С. В. Маклаков – Москва : ДИАЛОГ-МИФИ, 1999. – 256 с.
12. **ISO/TR 16044:** 2004 – Graphic technology – Database architecture model and control parameter coding for process control and workflow (Database AMPAC).

Электронные ресурсы образовательного портала [ACADEMY.ORACLE.COM](https://academy.oracle.com).

Электронные ресурсы образовательного портала [INTUIT.RU](https://intuit.ru).

Электронные ресурсы портала [HTTPS://ORACLEPLUSQL.RU](https://oracleplusql.ru).

Электронные ресурсы портала [POSTGRESPRO.RU](https://postgrespro.ru).