



# БАЗЫ ДАННЫХ И ЭКСПЕРТНЫЕ СИСТЕМЫ

ФИО преподавателя: Тараканов О.В., канд. техн. наук, доцент  
e-mail: [tarakanov@mirea.ru](mailto:tarakanov@mirea.ru)

*Канд. техн. наук, доцент ТАРАКАНОВ Олег Викторович*

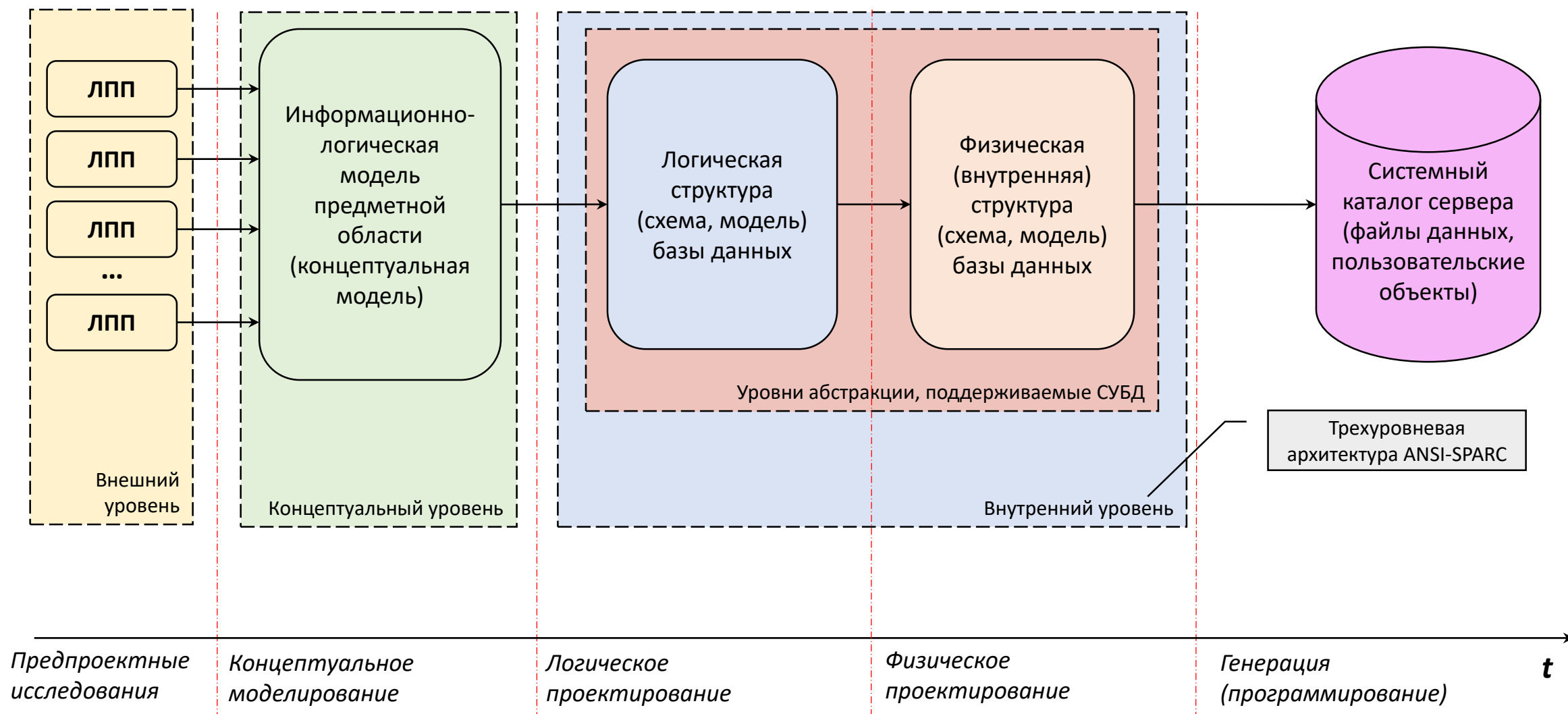
Базы данных и экспертные системы

# ПРОЕКТИРОВАНИЕ БАЗЫ ДАННЫХ

Учебные вопросы:

1. Общие вопросы проектирования базы данных
2. Проектирование нерегулярных структур
3. Языки информационного моделирования и проектирования баз данных

# Общие вопросы проектирования базы данных



# Общие вопросы проектирования базы данных

## Предпроектные исследования

Схема подготовки исходных данных	С независимым экспертом	С «зависимым» экспертом	Прогнозирование требований
Требуемая численность экспертов	10 – 15	5 – 8	1
Объективность результатов	высокая	средняя	
Трудозатраты на проведение	высокие	средние	
Приоритетность использования	наивысшая	при невозможности выполнения требований схемы с независимым экспертом	при отсутствии экспертов, в «пионерских» проектах

## Способы проектирования

Если перечень запросов неизменный, все они известны на этапе предпроектных исследований, изменение перечня запросов не предполагается, то способ проектирования **«от запроса»** – ЛПП представляют собой списки потребностей пользователей, перечень атрибутов извлекается из текстов запросов.

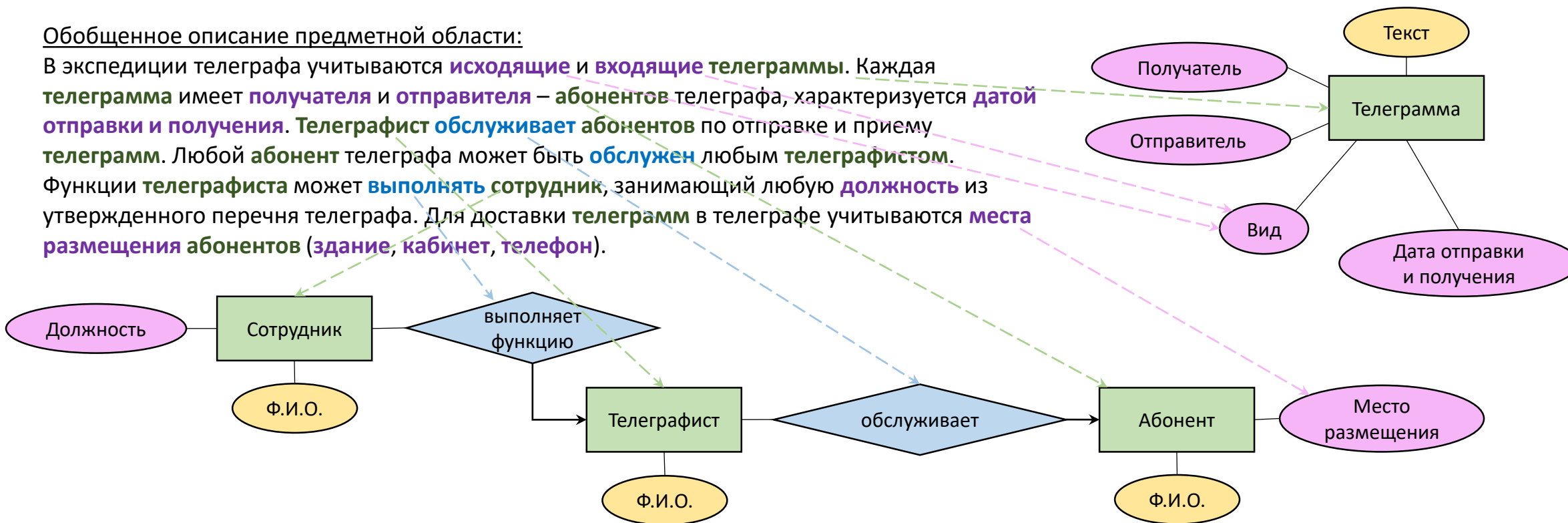
Если перечень запросов изменяемый, они не известны на этапе предпроектных исследований, то способ проектирования **«от предметной области»** – ЛПП представляют собой описание предметной области в понимании пользователей, перечень атрибутов извлекается из текстов описаний.

# Общие вопросы проектирования базы данных

## Концептуальное моделирование (проектирование)

Обобщенное описание предметной области:

В экспедиции телеграфа учитываются **исходящие** и **входящие телеграммы**. Каждая **телеграмма** имеет **получателя** и **отправителя** – **абонентов** телеграфа, характеризуется **датой отправки и получения**. **Телеграфист обслуживает абонентов** по отправке и приему **телеграмм**. Любой **абонент** телеграфа может быть **обслужен** любым **телеграфистом**. Функции **телеграфиста** может **выполнять сотрудник**, занимающий любую **должность** из утвержденного перечня телеграфа. Для доставки **телеграмм** в телеграфе учитываются **места размещения абонентов** (**здание, кабинет, телефон**).



## Логическое проектирование. Уровни, задачи

- **Диаграмма «сущность-связь» (ERD)**: уточнение перечня сущностей и атрибутов, связывание сущностей, разрешение связей, типа M : N
- **Модель данных, основанная на ключах (КВ)**: нормализация отношений до ДКНФ, оптимизация, миграция ключа отношения
- **Полная атрибутивная модель (ФА)**: уточнение доменов, разработка правил ссылочной целостности, устранение ловушек соединения

# Общие вопросы проектирования базы данных

## Физическое проектирование. Уровни, задачи

→ **Обоснование выбора СУБД:** Мажоритарная выборка

Матрица соответствия возможностей СУБД потребностям проекта

Сравнительный признак	MS Access 2015	MS SQL Server 2019	ORACLE Database 11g R2	ORACLE Personal 12C	MongoDB v. 5.0	PostgreSQL 14	Вес признака
Возможность реализации архитектуры «клиент-сервер»	0	1	1	0	0	1	0.14
Реализация реляционной модели данных	0	1	1	1	0	1	0.45
Наличие RDAC-метода доступа к данным	0	1	1	0	0	1	0.10
Не требовательность к ресурсам вычислительной платформы	1	0	0	1	1	1	0.05
Наличие встроенного расширения языка манипулирования данными	0	1	1	1	0	1	0.01
Поддержка синтаксиса SQL-3	1	1	1	1	0	1	0.03
Нормированная стоимость (абсолютная в у.е.)	0.93 (80)	0.88 (135)	0.00 (1000)	0.82 (120)	0.46 (355)	1.00 (0)	0.16
Отсутствие необходимости дополнительного обучения	1	1	0	1	0	1	0.01
Свободная лицензия	0	0	0	0	0	1	0.05
<b>ОБЩАЯ ОЦЕНКА P</b>	0.24	0.88	0.73	0.68	0.12	1.00	

Свертка

$$P_o = \sum_n k_n \cdot I_n^{(H)}$$

Кодирование

$$I_n = \begin{cases} 1, & \text{если по свойству СУБД пригодна} \\ 0, & \text{если по свойству СУБД не пригодна} \end{cases}$$

Взвешивание

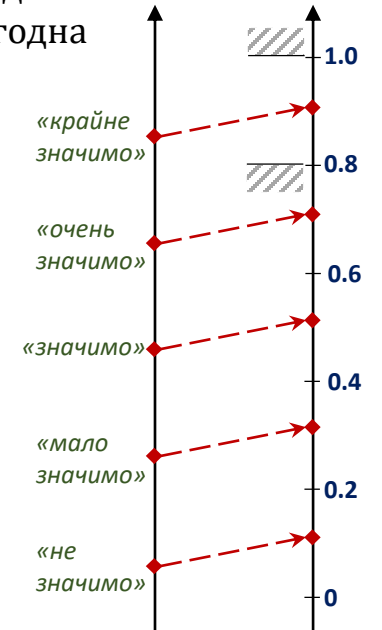
Нормировка

$$I_n^{(H)} = a_n \cdot I_n + b_n$$

$$a_n = \frac{1}{I_{\max} - I_{\min}}$$

$$b_n = -\frac{I_{\min}}{I_{\max} - I_{\min}}$$

Оцифровка мнений экспертов о весах признаков

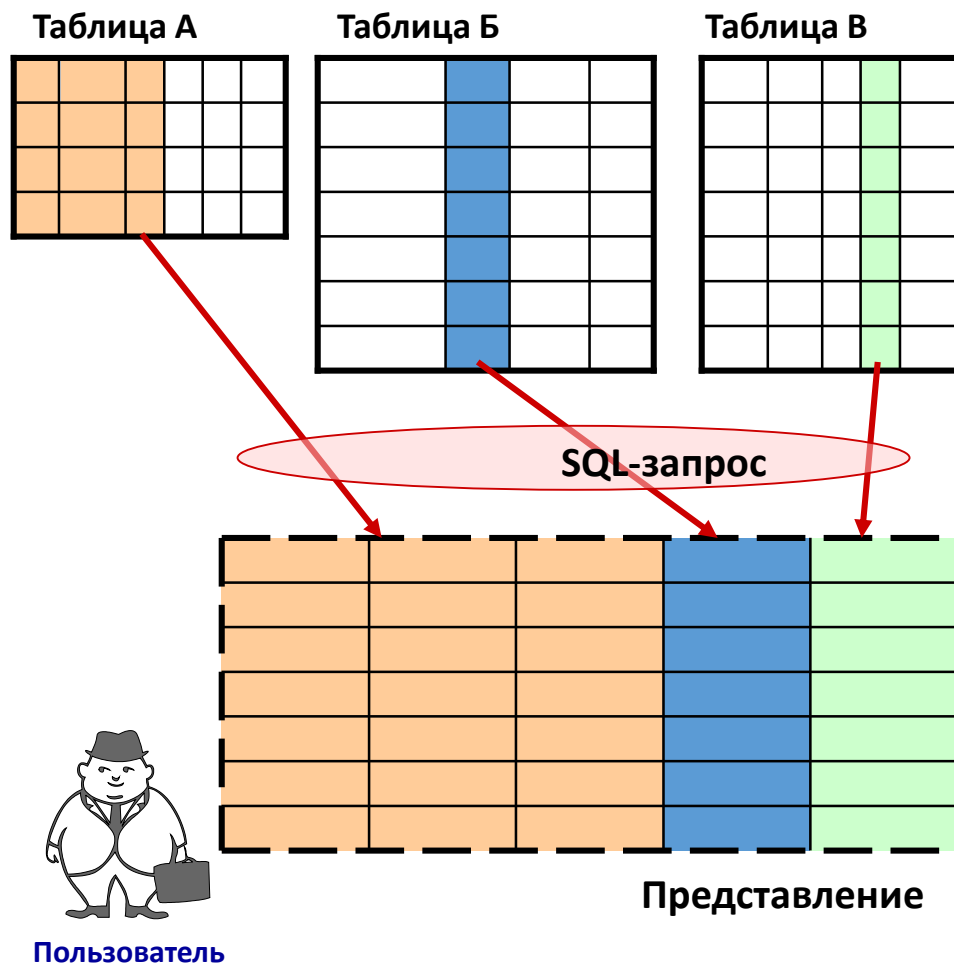


# Общие вопросы проектирования базы данных

→ **Задание типов данных по умолчанию:** СТРОКА >> CHAR() / VARCHAR()    ЧИСЛО >> INTEGER / FLOAT / DOUBLE    ДАТАВРЕМЯ >> DATE / TIME

→ **Описание (уточнение) атрибутов:**    VARCHAR(20) ≠ VARCHAR(150)    REAL ≠ DOUBLE    DATE ≠ TIMESTAMP

→ **Создание представлений**



# Общие вопросы проектирования базы данных

## Оценка объема базы данных

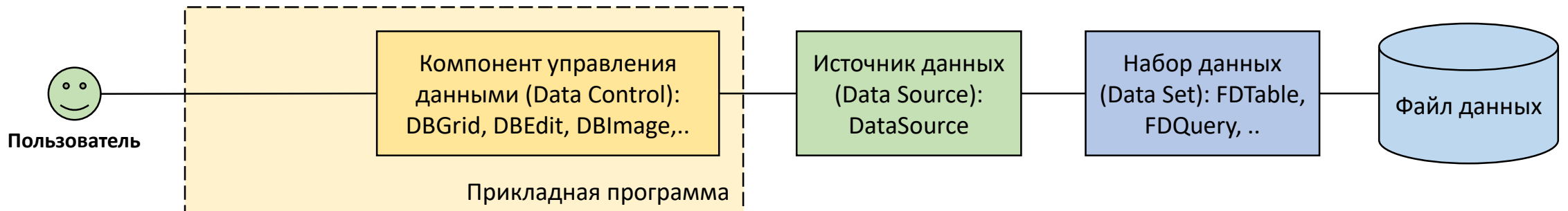
$n$			
Студент	Предмет	Дата	Оценка
Иванов И.В.	ТОАУ	21.11.2010	3
Иванов И.В.	БД	14.04.2012	4
Петров В.А.	ТОАУ	30.11.2010	3
Петров В.А.	ПЯВУ	13.05.2011	5
...			
Шабанов С.Ю.	ПЯВУ	10.02.2011	3
Шабанов С.Ю.	ПЯВУ	15.04.2011	2
Шабанов С.Ю.	БД	21.05.2012	5
$b$			

$$V_{\text{строки}} = \sum_n [V(\text{ячейка}_n) + V(\text{индекс}_n)]$$

$$V_{\text{таблицы}} = \sum_b V_{\text{строки}_b}$$

$$V_{\text{БД}} = 1,2 \times \sum_i V_{\text{таблица}_i}$$

## Программирование базы данных. Интерфейс данных

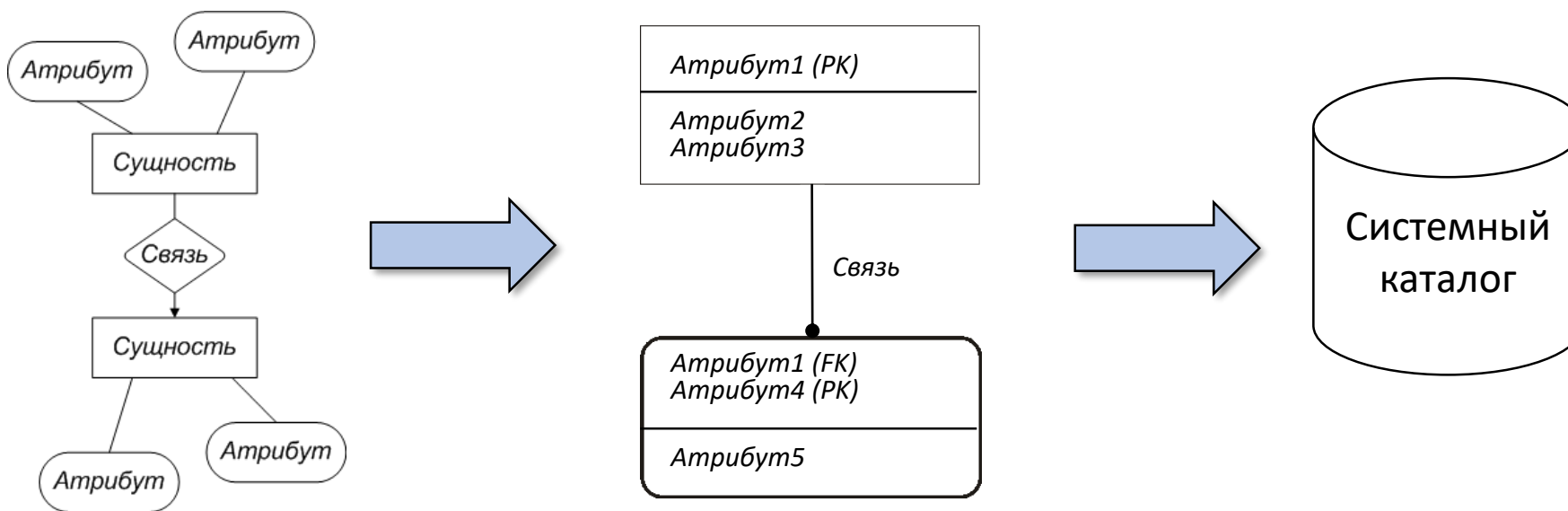




# Общие вопросы проектирования базы данных

## Обратное проектирование (реинжиниринг)

*Прямое проектирование (Forward Engineering)*



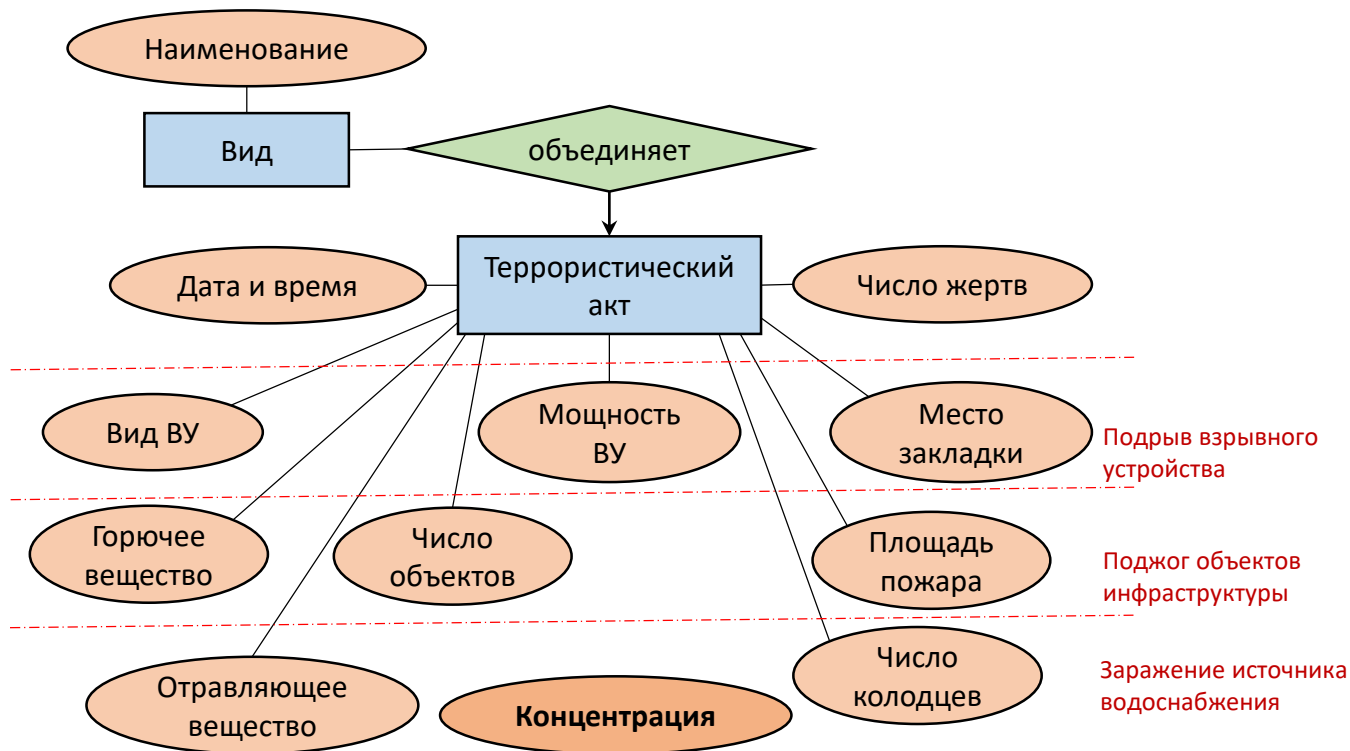
*Обратное проектирование (Reverse Engineering)*

# Проектирование нерегулярных структур

## Проектирование структур с переменным числом атрибутов

**Задача:** Необходимо хранить данные о характеристиках террористических актов. Набор характеристик зависит от вида совершенного террористического акта. Число характеристик и единицы их измерения различаются для разных видов.

### Решение 1

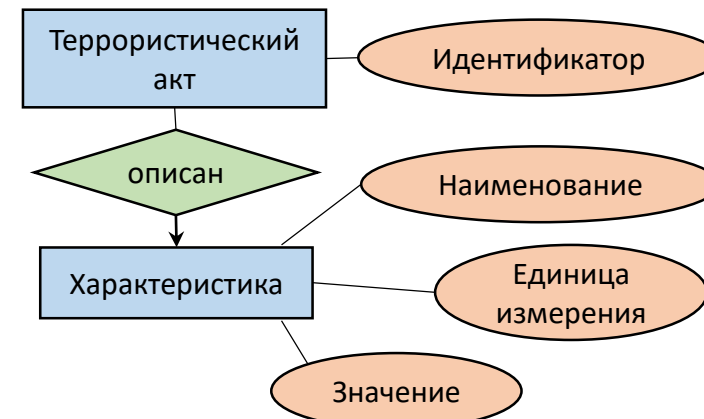


Террористический акт

Дата и время	Число жертв	Вид ВУ	Мощность ВУ	Место закладки	Горючее вещество	Число объектов	Площадь пожара	Отравляющее вещество	Число колодцев
26.09.22 03:18	0	С4	1200 кг	1 и 2 нитка	NULL	NULL	NULL	NULL	NULL
21.12.22 23:44	0	NULL	NULL	NULL	бензин	2	4 кв. м.	NULL	NULL
15.03.22 06:22	10	NULL	NULL	NULL	NULL	NULL	NULL	Стирол	2

`ALTER TABLE "Террористический акт" ADD ("Концентрация" ...);`

### Решение 2

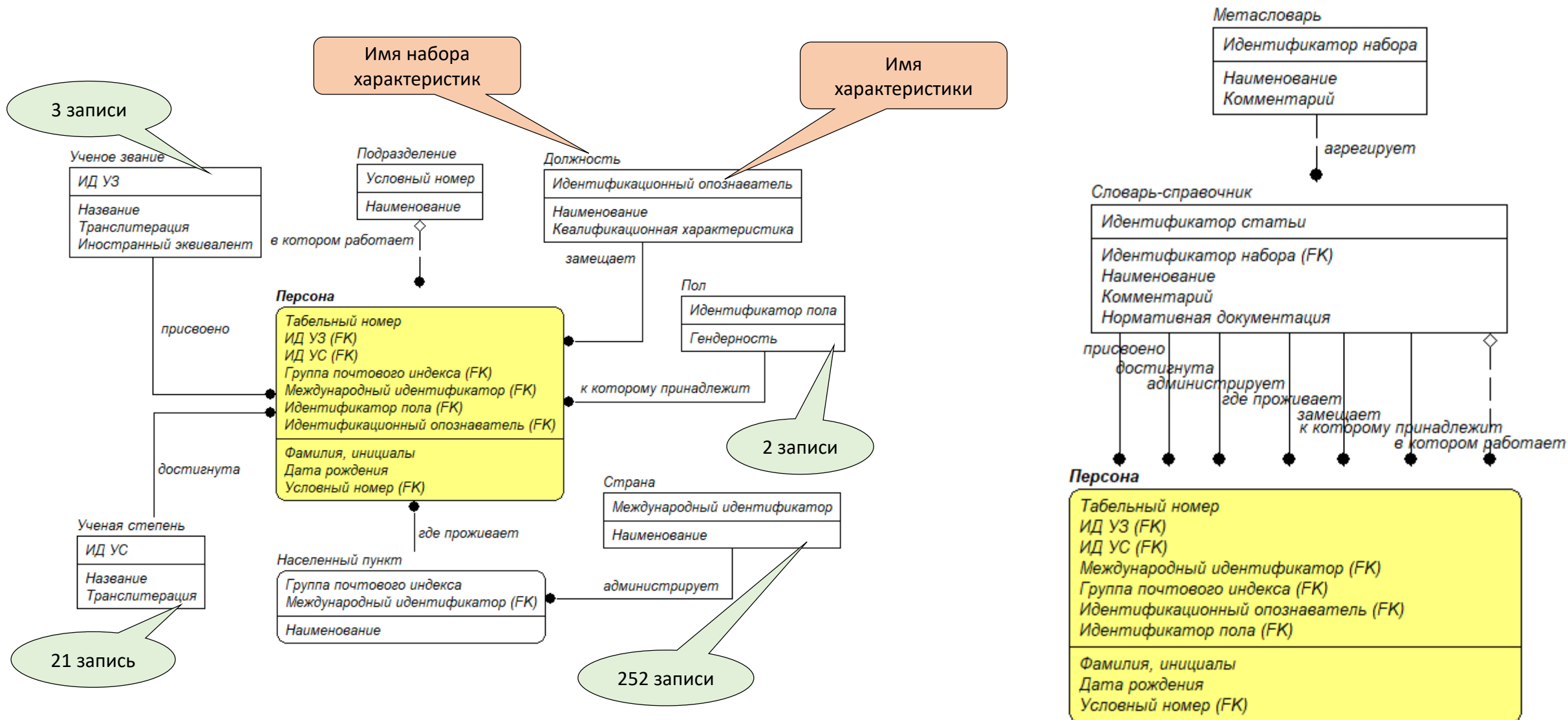


Характеристика

Идентификатор теракта	Наименование	Единица измерения	Значение
26.09.22 03:18	Число жертв	чел.	0
26.09.22 03:18	Вид ВУ	NULL	С4
26.09.22 03:18	Мощность ВУ	кг	1200
26.09.22 03:18	Место закладки 1	нитка	1
26.09.22 03:18	Место закладки 2	нитка	2
21.12.22 23:44	Число жертв	чел.	0
21.12.22 23:44	Горючее вещество	NULL	бензин
21.12.22 23:44	Число объектов	шт.	2
21.12.22 23:44	Площадь пожара	кв. м.	4
15.03.22 06:22	Число жертв	чел.	10
15.03.22 06:22	Отравляющее вещество	NULL	Стирол

# Проектирование нерегулярных структур

## Разработка компактного словаря-справочника данных



# Проектирование нерегулярных структур

## Организация хранения иерархий

**Задача:** Организовать учет организационно-штатной структуры общеобразовательной школы



### Решение 1: Матрица связности

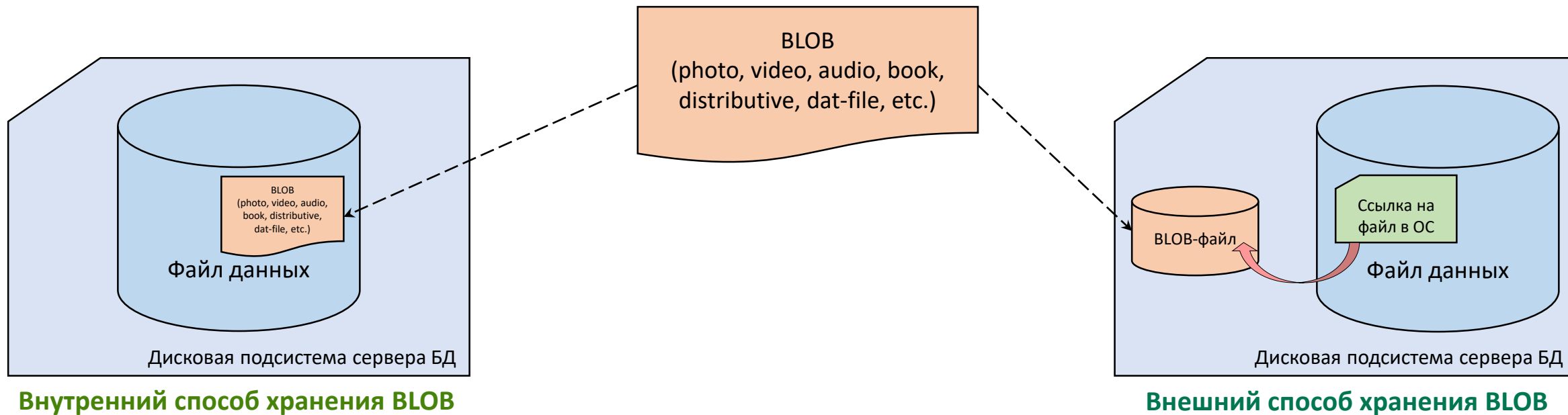
	Директор	Зав. мл.кл.	1 зав. отд.пр.	4 класс	...	Иванов М.И.	Петров В.С.	...	Боширов М.А.
Директор	1	1	1	1		0	0		0
Зав.мл.кл.	1	1	0	0		0	0		0
1 зав. отд.пр.	1	0	1	1		0	0		0
4 класс	0	1	0	1		1	1		1
...									
Иванов М.И.	0	0	0	1		1	0		0
Петров В.С.	0	0	0	1		0	1		0
...									
Боширов М.А.	0	0	0	1		0	0		1

### Решение 2: Список инцидентов

Узел А	Узел Б	Сателлит
root	Директор	1
Директор	Зав.мл.кл.	1
Директор	Зав. отд.пр.	2
...		
Зав.мл.кл.	4 класс	3
4 класс	Иванов М.И.	NULL
4 класс	Петров В.С.	NULL
...		
4 класс	Боширов М.А.	NULL

# Проектирование нерегулярных структур

## Обоснованный выбор способа хранения больших двоичных последовательностей



Высокая переносимость БД за счет жесткой привязки BLOB к файлу данных.

Полный контроль целостности BLOB со стороны СУБД.

Увеличение объема, занимаемого БД в памяти сервера (внешней и внутренней).

Отсутствие возможности хотя бы косвенного индексирования данных BLOB.

Накладные расходы при необходимости модификации BLOB внешним ПО.

Существенное нарушение «выровненности» строк.

Сложная переносимость БД из-за отсутствия привязки BLOB к файлу данных.

Отсутствие контроля целостности BLOB со стороны СУБД.

Отсутствие увеличения объема, занимаемого БД в памяти сервера.

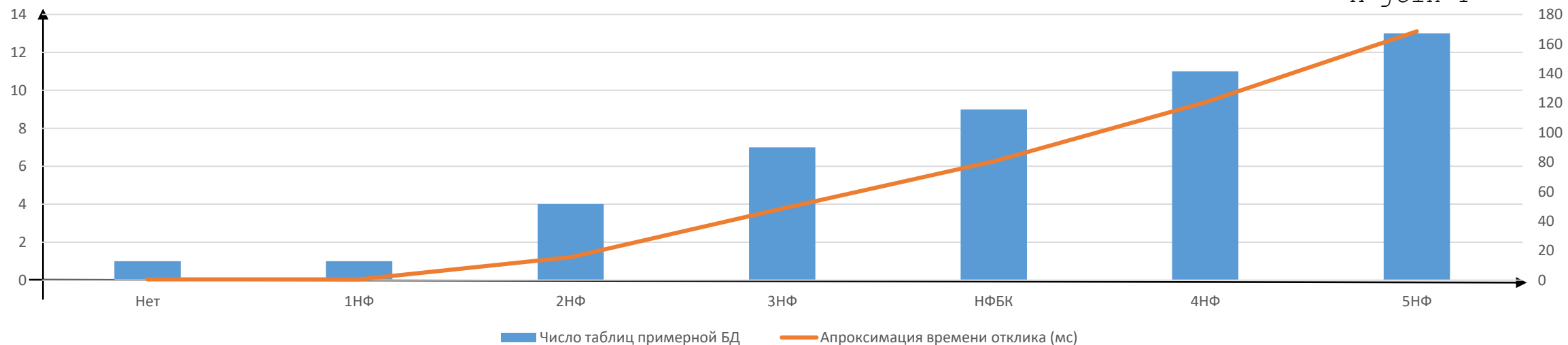
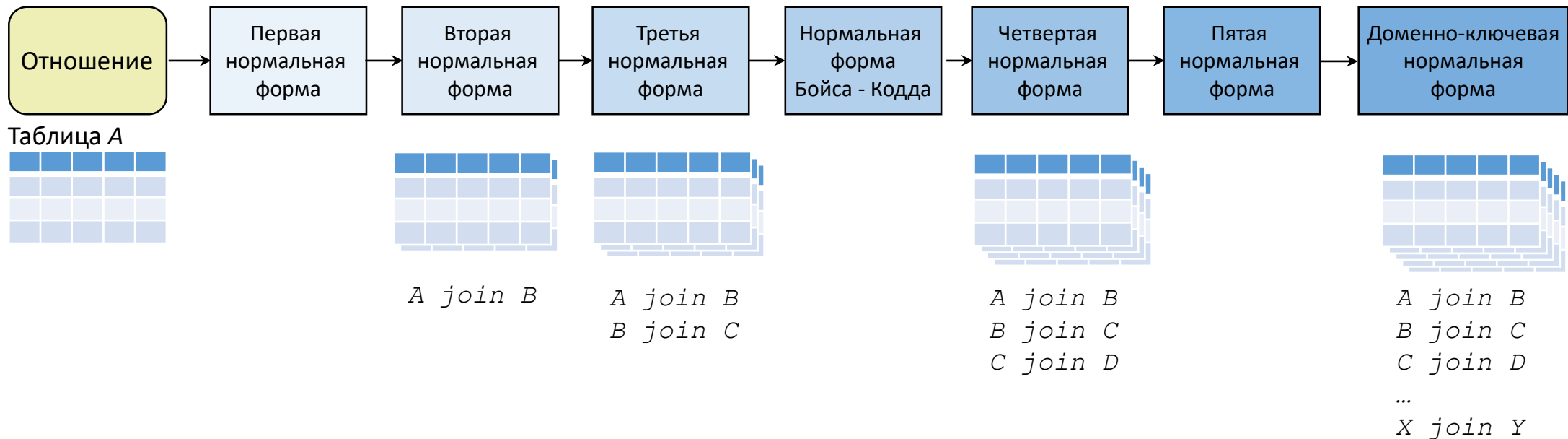
Возможно косвенное индексирование данных BLOB (по текстовой ссылке – адресу).

Прямая модификация BLOB внешним ПО (не зависимо от СУБД).

Не влияет на «выровненность» строк.

# Проектирование нерегулярных структур

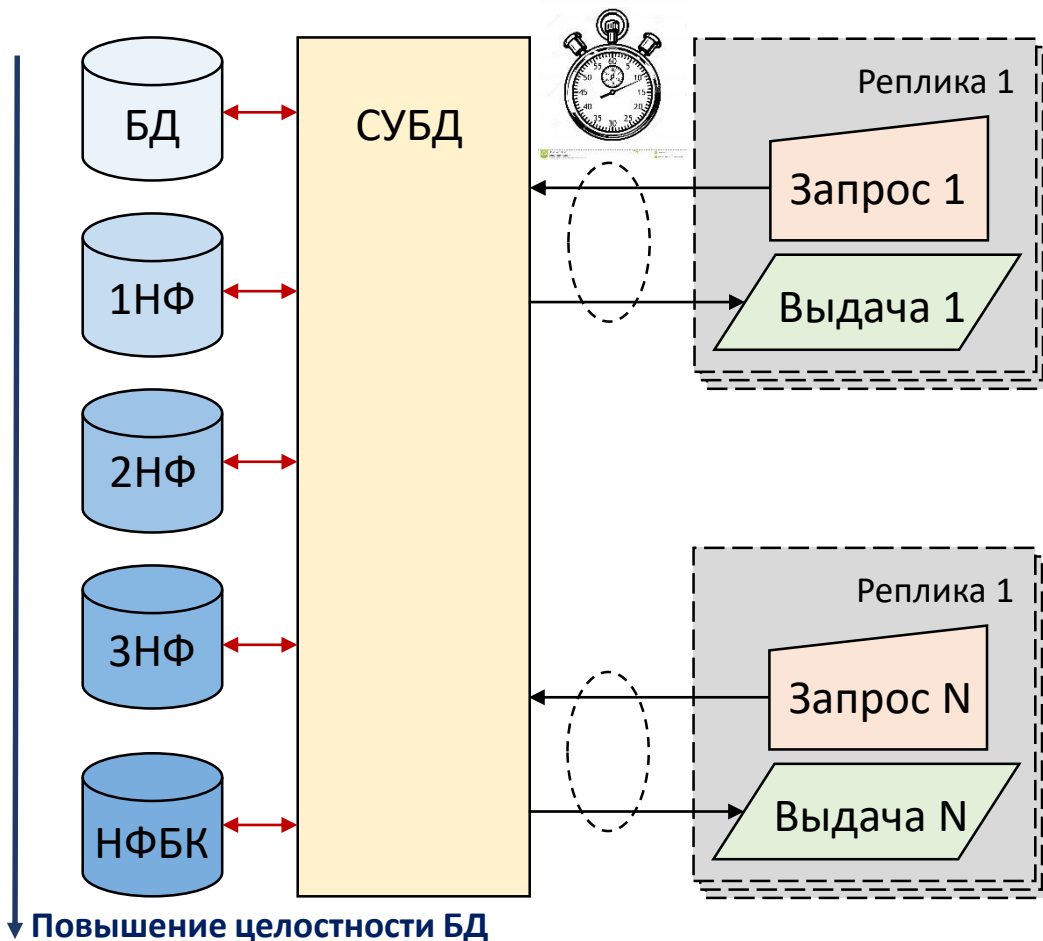
## «Парадокс нормализации»



# Проектирование нерегулярных структур

## «Парадокс нормализации»

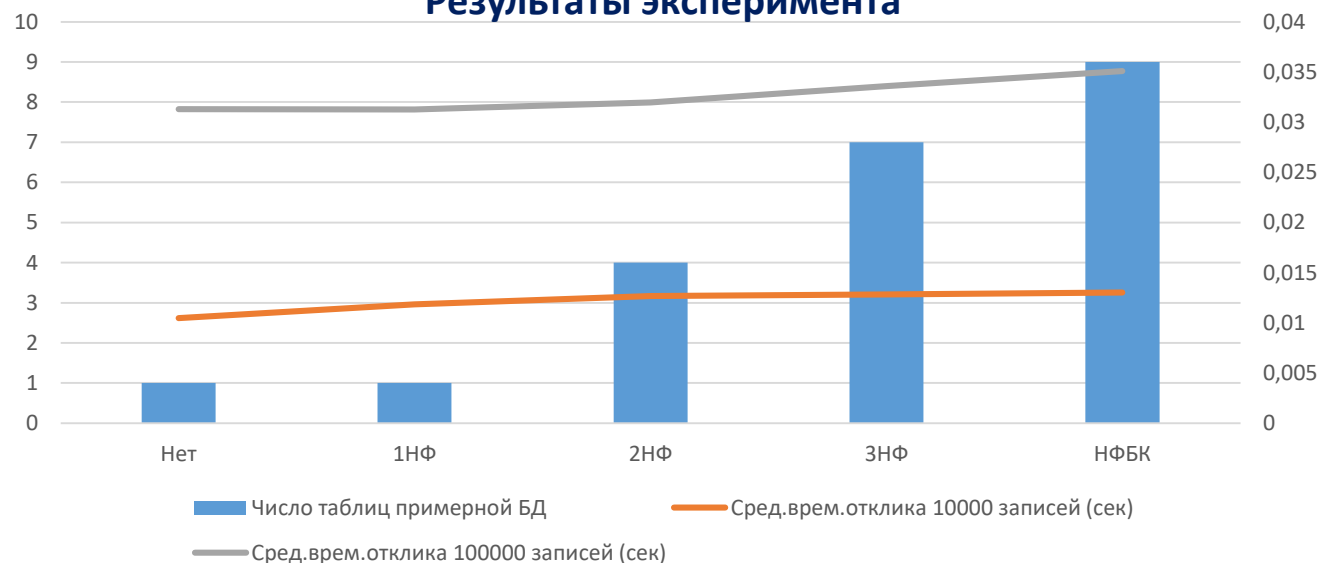
### Схема экспериментального стенда



**Использовано:** 5 вариантов БД разной степени нормализованности  
10 тестовых запросов с разной глубиной рекурсии (числом JOINов)  
10 реплик над каждой БД с каждым тестовым запросом

**Мощность эксперимента:**  $P_{\beta} = 0.012$

### Результаты эксперимента



**Коэффициент корреляции:**

**K = 0.36**

**K = 0.44**

# Языки информационного моделирования и проектирования баз данных

Entity-Relationship Diagram – ERD

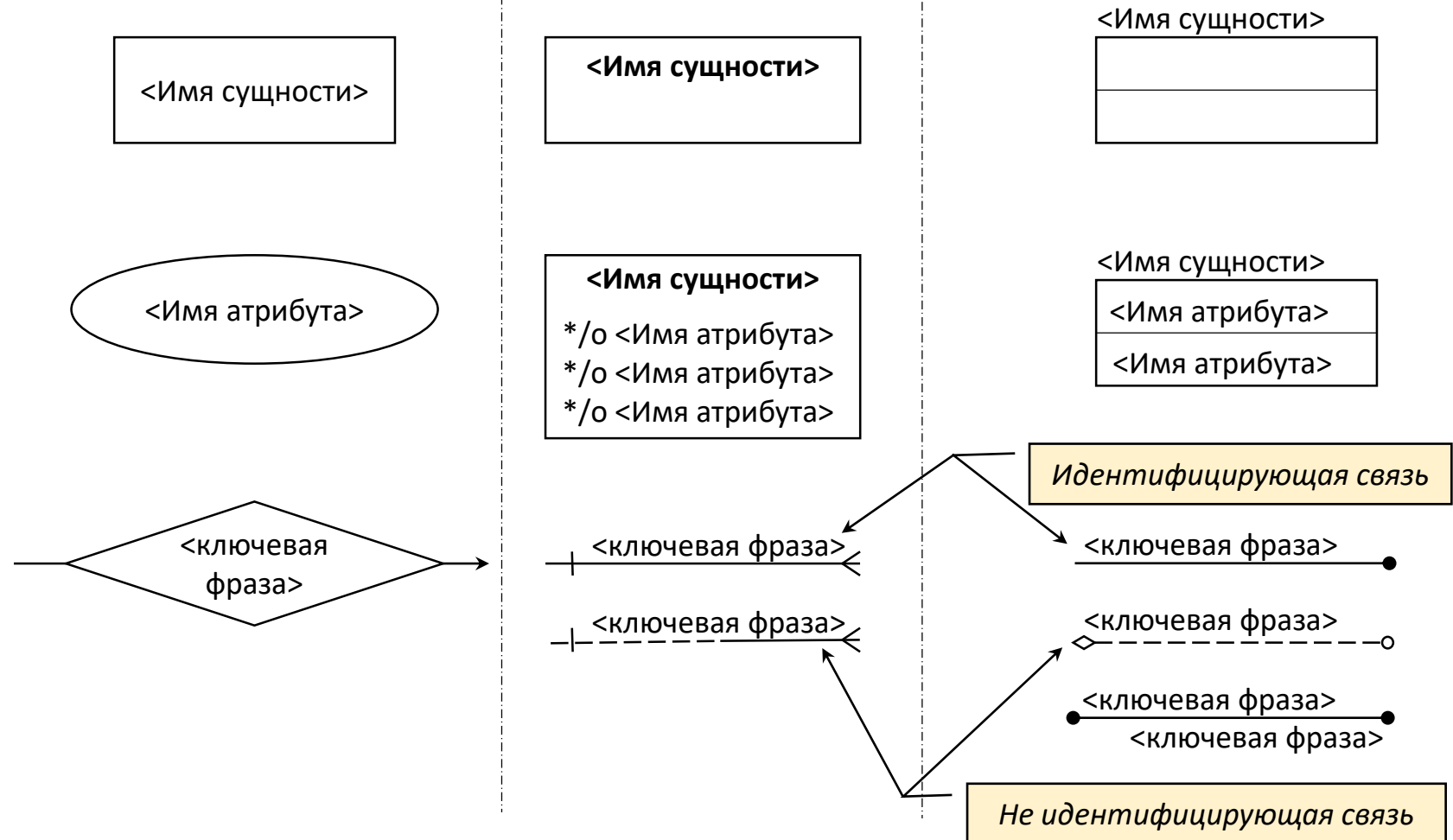
Нотация Мартина (Баркепа)

Integrated DEFINITION 1 eXchange – IDEF1x

**Сущность** – абстракция объекта учета в совокупности его экземпляров, отличающихся значениями одинакового набора значимых характеристик

**Атрибут** – абстракция значимого свойства объекта учета, принимающая значения из области допустимых – домена

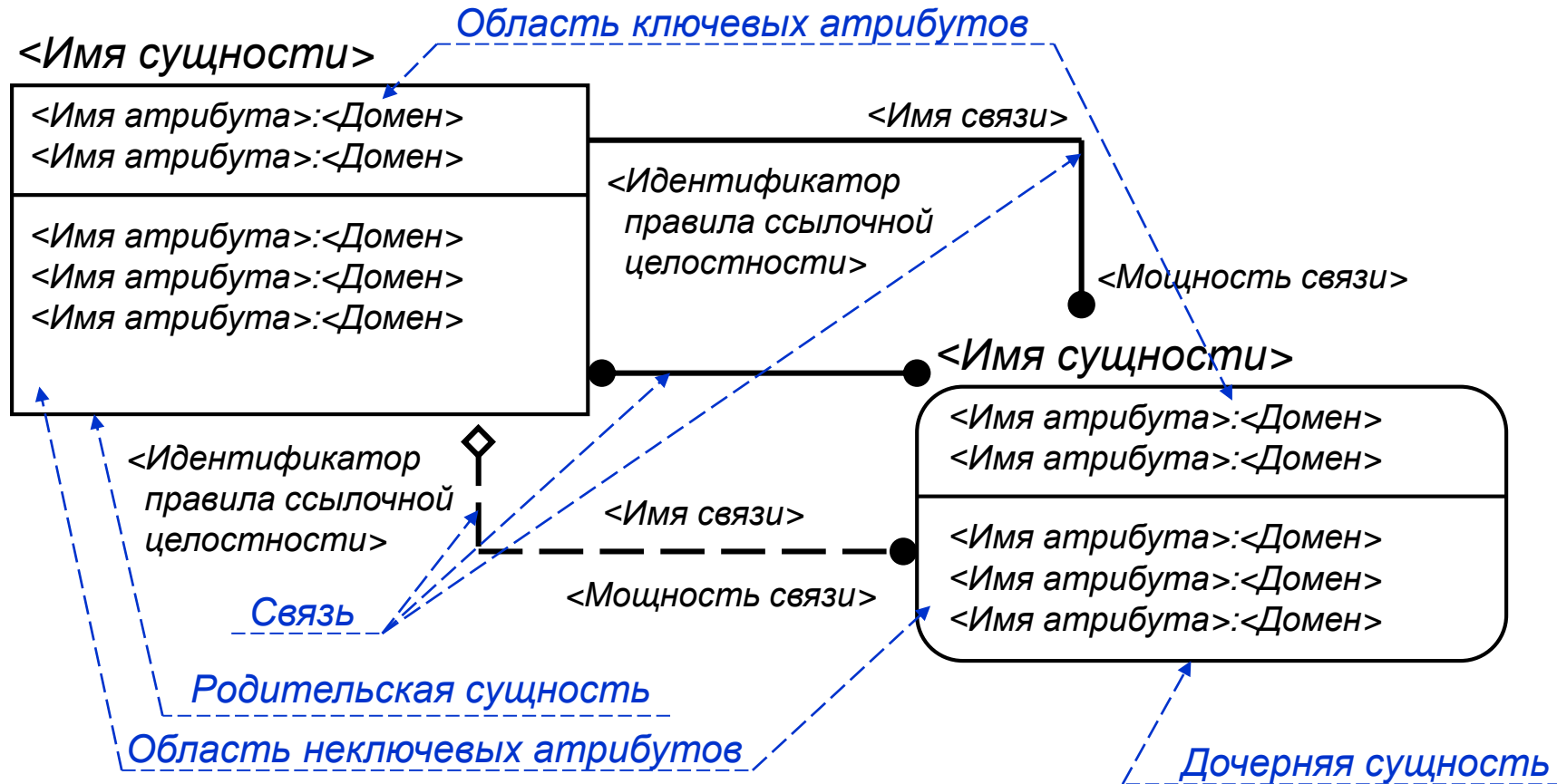
**Связь** – абстракция функционального взаимодействия между атрибутами разных или одной и той же сущности





# Языки информационного моделирования и проектирования баз данных

## Integrated DEFinition 1 eXchange – IDEF1x



# 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. **Дейт, К. Дж.** Введение в системы баз данных: пер. с англ. – 7-е изд. / К. Дж. Дейт. – Москва: Издательский дом "Вильямс", 2001. – 1072 с.
5. **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.
6. **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.
7. **Бойко, В. В.,** Проектирование баз данных информационных систем / В. В. Бойко, В. М. Савинков. – Москва: Финансы и статистика, 1989.
8. **Смирнов С. Н. Задворьев И.С.** Работаю с ORACLE.: Учебное пособие/2-е изд., испр. и доп. – М: Гелиос АРВ, 2002 г. – 496 с.
9. **Фейерштейн, С., Прибыл, Б.,** Oracle PL/SQL. Для профессионалов. 6-е изд. – Санкт-Петербург: Питер, 2015. – 1024 с.
10. **ISO/TR 16044:** 2004 – Graphic technology – Database architecture model and control parameter coding for process control and workflow (Database AMPAC).
11. **Маклаков, С. В.,** BPwin и ERwin. CASE-средства разработки информационных систем / С. В. Маклаков – Москва : ДИАЛОГ-МИФИ, 1999. – 256 с.
12. **Задворьев, И. С.,** Язык PL/SQL. Учебно-методическое пособие. – Москва: Онто-Принт, 2017. – 178 с.
13. **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://ORACLEPLSQL.RU](https://oracleplusql.ru).

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