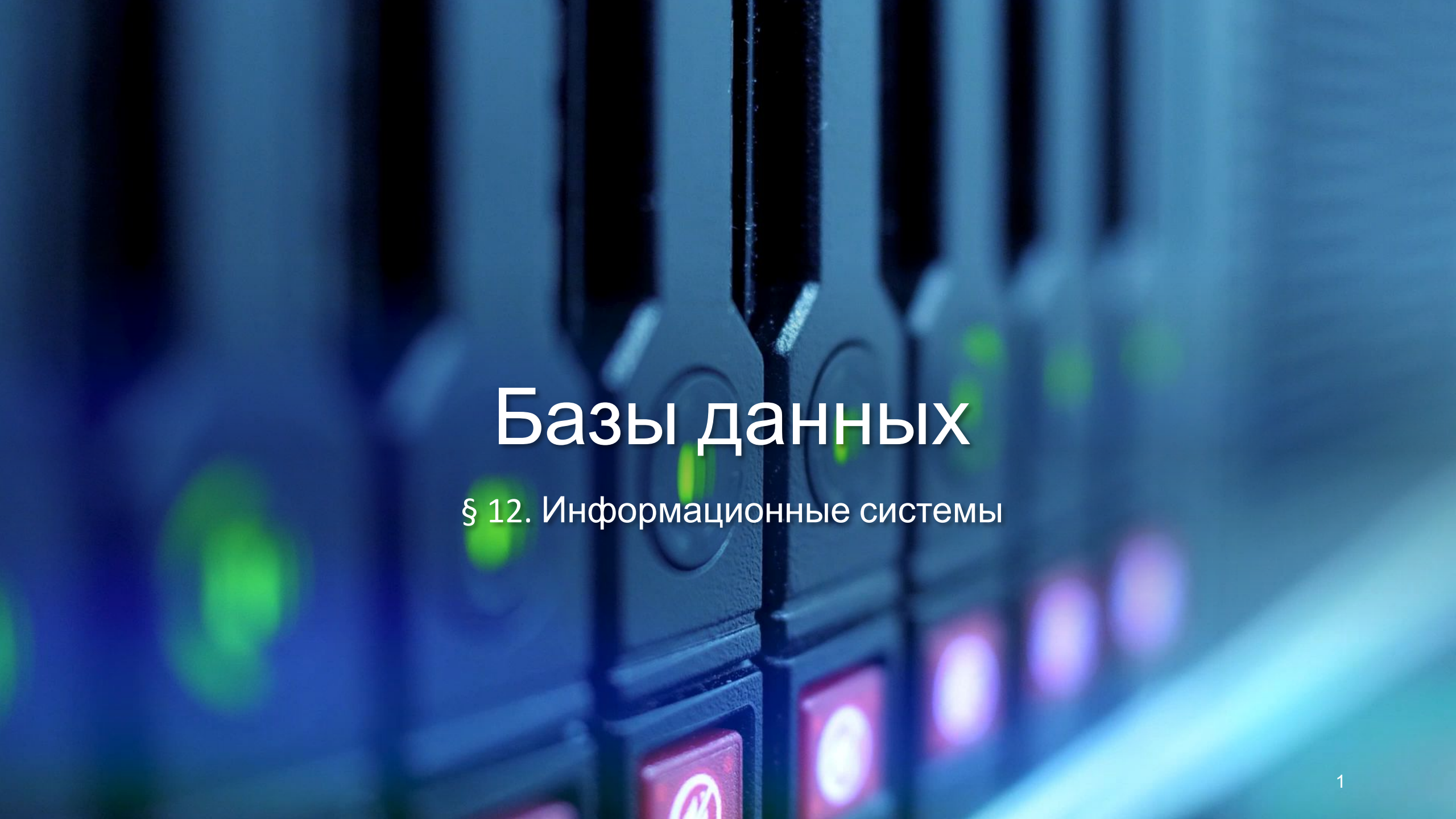


Базы данных

§ 12. Информационные системы

Что такое информационная система?

Информационная система (ИС) в широком смысле — это аппаратные и программные средства, предназначенные для того, чтобы своевременно обеспечить пользователей нужной информацией.

Задачи:

- хранение данных
- доступ к данным

© К.Ю. Поляков, Е.А. Ерёмин, 2013

БД и СУБД

База данных (БД) — это специальным образом организованная совокупность данных о некоторой предметной области, хранящаяся во внешней памяти компьютера.

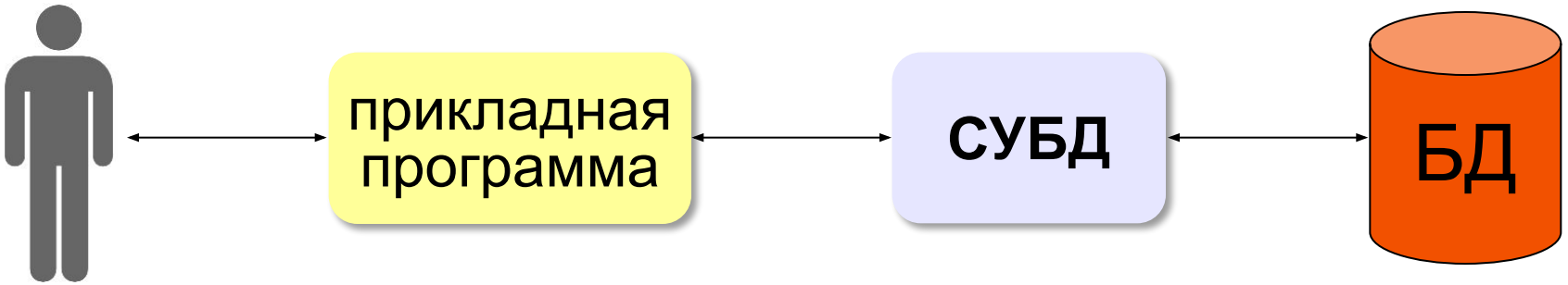
Система управления базой данных (СУБД) — это программные средства, которые позволяют выполнять все необходимые операции с базой данных.

БД + СУБД = и**нформационная система**

БД и СУБД

Задачи СУБД:

- поиск данных
- редактирование данных
- выполнение несложных расчетов
- обеспечение *целостности* (корректности, непротиворечивости) данных
- восстановление данных после сбоев



Классификация ИС

локальные

БД и СУБД на компьютере пользователя

- ➕ автономность
- ➖ нужно обновлять БД на каждом компьютере
- ➖ как учесть изменения, внесённые пользователями

удалённые

БД на удалённом компьютере (в сети)

- ➕ все работают с одной БД
- ➖ зависимость от сети

Как хранить данные?

- **Текстовый формат**:

Иванов;Иван;Петрович;1968;«Audi»;20 000 р.
Петров;Пётр;Иванович;1975;«Лада»;30 000 р.
Сидоров;Семён;Иванович;1975;«Нива»;15 000 р.

- **Собственный формат программы**: ? Что плохо?

⛔ нет ***переносимости*** (БД можно использовать только с этой программой)

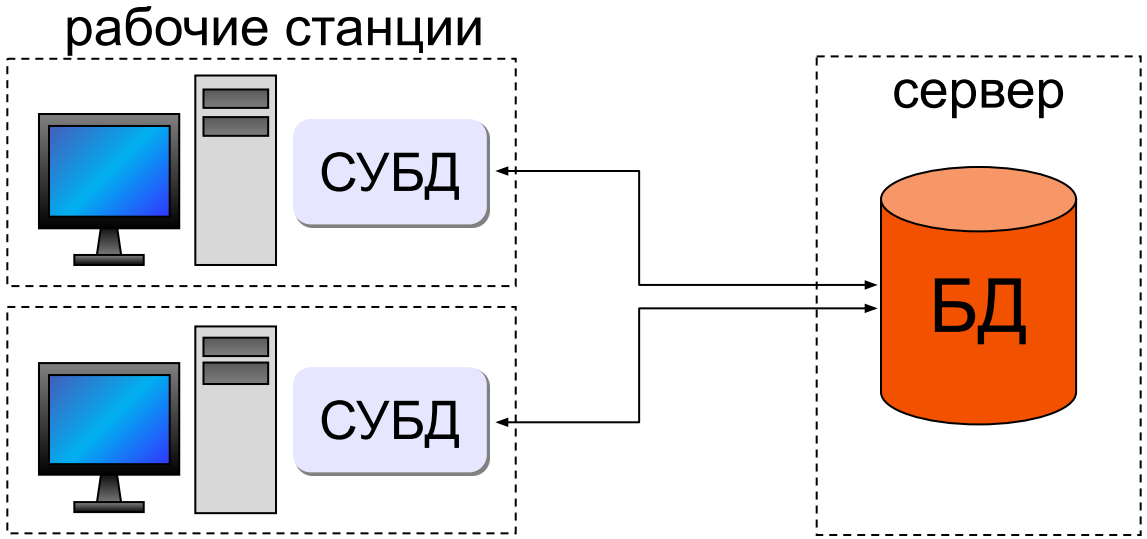
Переносимость – это возможность использовать БД в других информационных системах.

- **Универсальный формат**:

БД: «данные о данных»

данные

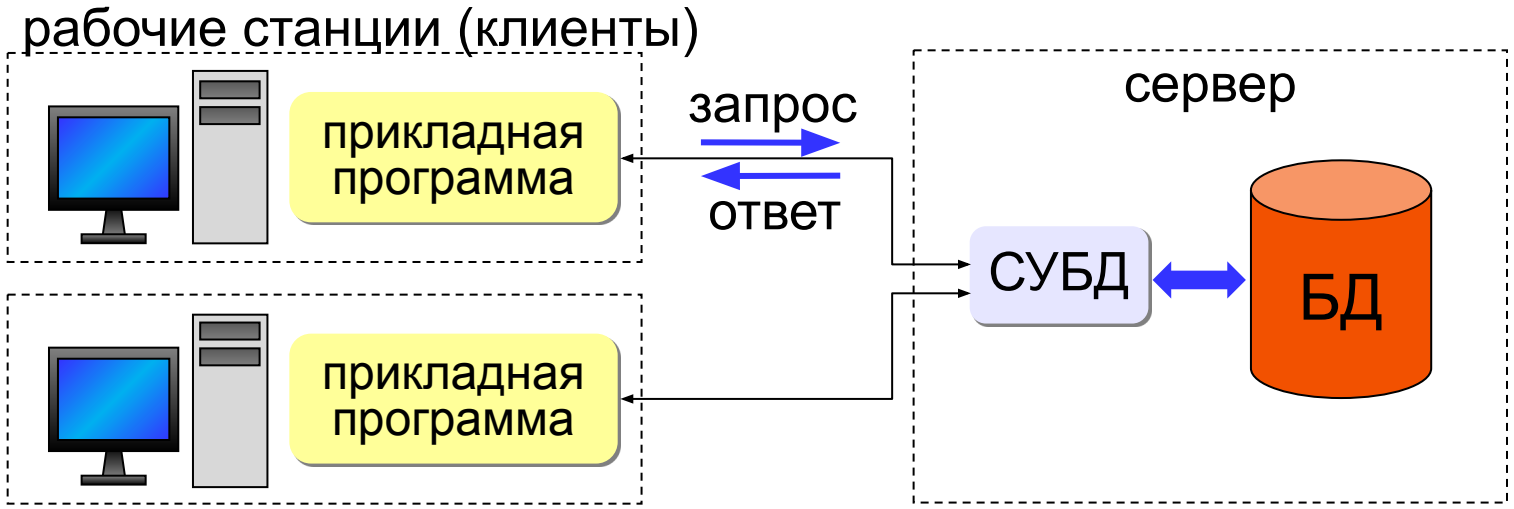
Файл-серверные СУБД



- рабочие станции должны быть мощными
- высокая нагрузка на сеть
- слабая защита данных
- ненадежность при большом количестве пользователей

Решение – перенести СУБД на сервер!

Клиент-серверные СУБД



Задачи клиента:

- отправить серверу запрос на языке SQL
- принять ответ сервера
- вывести результаты

Задачи сервера:

- при получении запроса поставить его в очередь
- выполнить запрос
- отправить ответ клиенту

SQL = *Structured Query Language* – язык структурных запросов для управления данными

Клиент-серверные СУБД

- обработка данных на сервере
- рабочие станции могут быть маломощными
- проще модернизация
- надежная защита данных (на сервере)
- снижается нагрузка на сеть
- надежная работа при большом количестве пользователей

- нужен мощный сервер
- высокая стоимость коммерческих СУБД

MS SQL Server

ORACLE

бесплатно!

Firebird (*www.firebirdsql.org*)

PostgreSQL (*www.postgresql.org*)

MySQL (*www.mysql.com*)

Транзакции

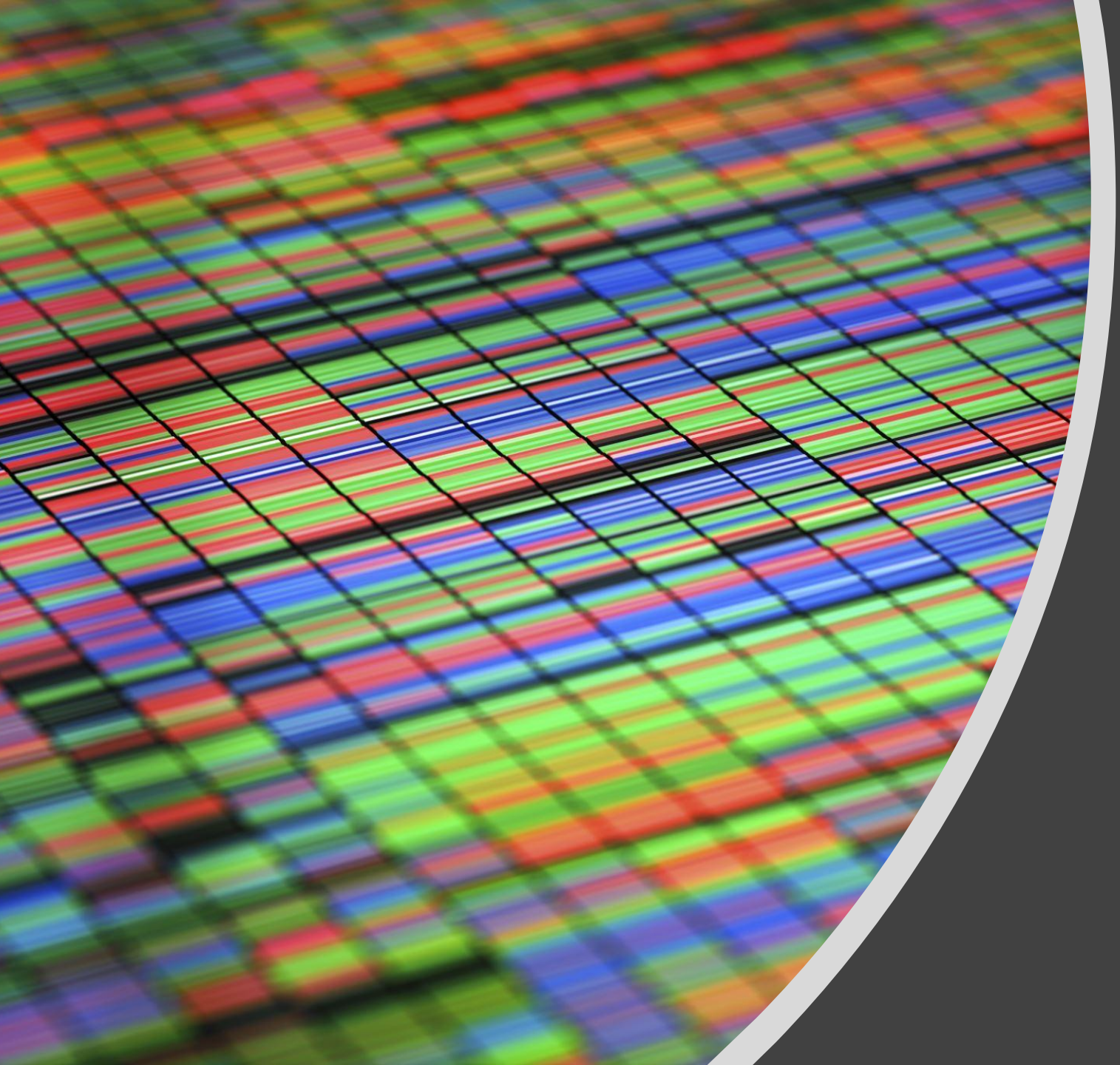
Задача: перевести 100 000 рублей со счета 12345 на счет 54321.

- прочитать сумму на счету 12345
- уменьшить ее на 100 000 рублей
- прочитать сумму на счету 54321
- увеличить ее на 100 000 рублей

сбой питания!

Транзакция – это группа операций, которая представляет собой одно законченное действие. Транзакция должна быть выполнена целиком или не выполнена вообще.

- журналирование
- создание новых записей (помечаются как действительные после завершения транзакции)



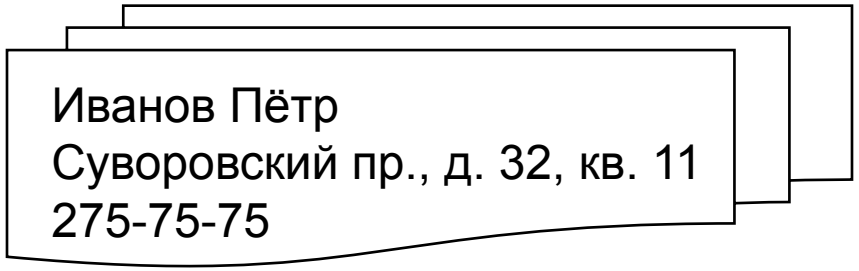
Базы данных

§ 13. Таблицы

Таблицы

Примеры:

- записная книжка
- каталог в библиотеке



поля

записи

Фамилия	Имя	Адрес	Телефон
Иванов	Петр	Суворовский пр., д. 32, кв. 11	275-75-75
Петров	Василий	Кутузовский пр., д. 12, кв. 20	276-76-76
Васильев	Иван	Нахимовский пр., д. 23, кв. 33	277-77-77

Типы полей:

- целые числа
- вещественные числа
- денежные суммы
- логические значения
- текстовые данные
- время, дата
- произвольные двоичные данные (рисунки, звук, видео)

Ключ

Ключ – это поле или комбинация полей, однозначно определяющие запись.

Могут ли эти данные быть ключом?

- ~~фамилия~~
- ~~имя~~
- номер паспорта
- ~~номер дома~~
- регистрационный номер автомобиля
- ~~город проживания~~
- адрес электронной почты
- дата выполнения работы
- марка стиральной машины ?

? Может ли быть несколько ключей в таблице?

Ключ

Первичный ключ – это ключ, выбранный в качестве основного.

Простой ключ – состоит из одного поля.

Составной ключ – состоит из нескольких полей.

Дата	Время	Температура	Влажность	Скорость ветра
21.07.2012	12:00	25	75	4
21.07.2012	15:00	23	70	3
…	…	…	…	…

Какой ключ?

Составной ключ ← **Дата** + **Время**

Свойства ключа:

- уникальность
- несократимость (Дата + Время + Влажность)

Суррогатный ключ

	Фамилия	Имя	Адрес	Телефон
	Иванов	Петр	Суворовский пр., д. 32, кв. 11	275-75-75
	Петров	Василий	Кутузовский пр., д. 12, кв. 20	276-76-76
3	Васильев	И	Нахимовский пр., д. 23, кв. 33	277-77-77

? Какой ключ?

Суррогатный ключ – это дополнительное поле, которое служит первичным ключом.

- операции с числами выполняются быстрее
- уникальность поддерживается СУБД (поле-счётчик)

Поиск в БД

Линейный поиск – это перебор всех записей до тех пор, пока не будет найдена нужная.

Код	Фамилия
1	Сидоров
2	Ветров
…	
1024	Померанцев

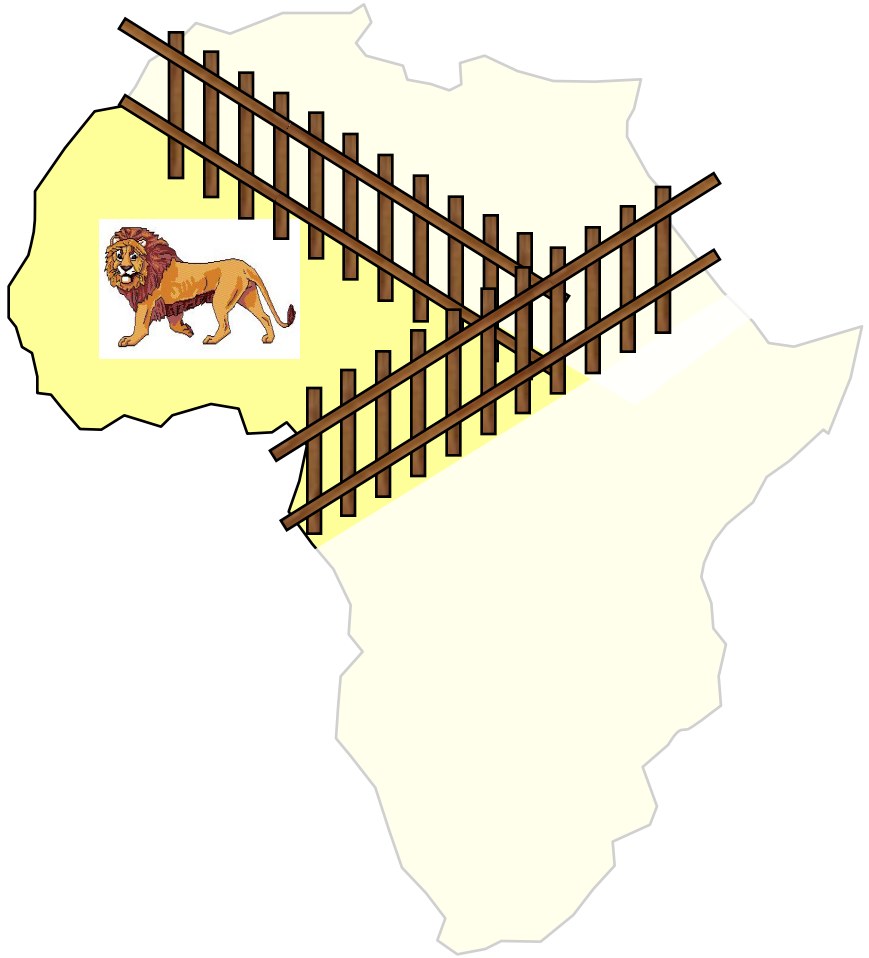
Иванов?

1024 сравнения!

- данные не надо предварительно готовить

- низкая скорость поиска

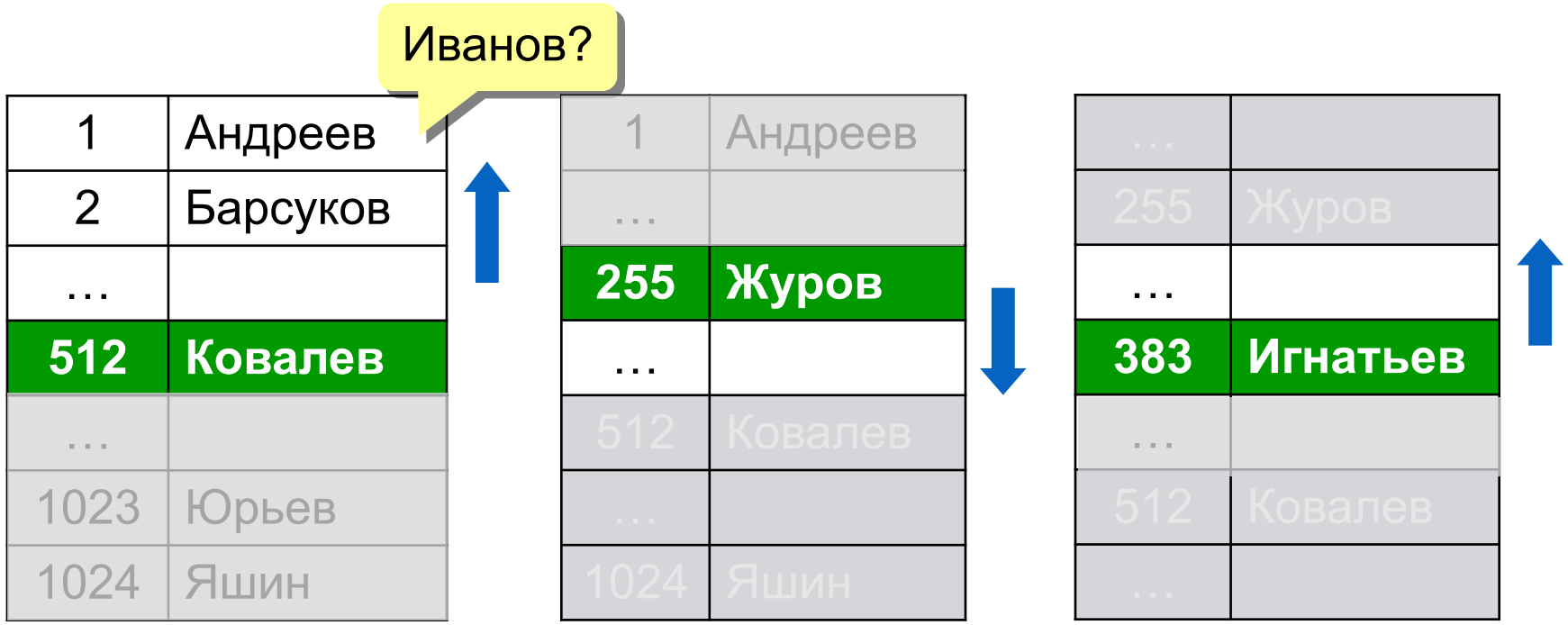
Двоичный поиск



1. Разделить область поиска на две равные части.
2. Определить, в какой половине находится нужный объект.
3. Перейти к шагу 1 для этой половины.
4. Повторять шаги 1-3 пока объект не будет «пойман».

Двоичный поиск в БД

требует предварительной сортировки.



? Сколько сравнений?

11 сравнений!

- быстрый поиск
- записи надо отсортировать по нужному полю
- можно использовать только для одного поля

Индексный поиск в БД

Индекс – это вспомогательная таблица, которая служит для ускорения поиска в основной таблице.

Номер	Фамилия	Имя	Адрес	Телефон
1	Иванов	Петр	Суворовский пр., д. 32, кв. 11	275-75-75
2	Петров	Василий	Кутузовский пр., д. 12, кв. 20	276-76-76
3	Васильев	Иван	Нахимовский пр., д. 23, кв. 33	277-77-77

отсортированы!

Фамилия	Номера записей
Васильев	3
Иванов	1
Петров	2

- двоичный поиск по всем столбцам, для которых построены индексы

- индексы занимают место на диске
- при изменении таблицы надо перестраивать все индексы (в СУБД – автоматически)

Целостность БД

Целостность базы данных означает, что она содержит полную и непротиворечивую информацию и удовлетворяет всем заданным ограничениям.

Физическая целостность – сохранность данных в случае отказа оборудования:

- транзакции
- резервное копирование
- RAID-массивы жёстких дисков (дублирование)

Логическая целостность – непротиворечивость:

- типы полей
- обязательные поля
- уникальные индексы
- ограничения на значения
- шаблоны ввода (###) ###-##-##
- условия на несколько полей (увольнение позже приёма)

Задача

Постройте индексы по всем полям:

Номер	Дата	Заказ	Товар	Количество, т
1	12.09.13	12	Ананасы	12
2	12.09.13	13	Апельсины	12
3	13.09.13	14	Ананасы	15
4	13.09.13	14	Бананы	13
5	13.09.13	15	Апельсины	11

© К.Ю. Поляков, Е.А. Ерёмин, 2013

Базы данных

§ 14. Многотабличные базы данных

Однотабличная БД

Альбомы

Код	Название	Группа	Год	Число композиций
1	Реки и мосты	Машина времени	1987	16
2	В круге света	Машина времени	1988	11
3	Группа крови	Кино	1988	11
4	Последний герой	Кино	1989	10

Что плохо?

- дублирование данных
- при изменении каких-то данных, возможно, придется менять несколько записей
- нет защиты от ошибок ввода (опечаток)

Многотабличная БД

Группы

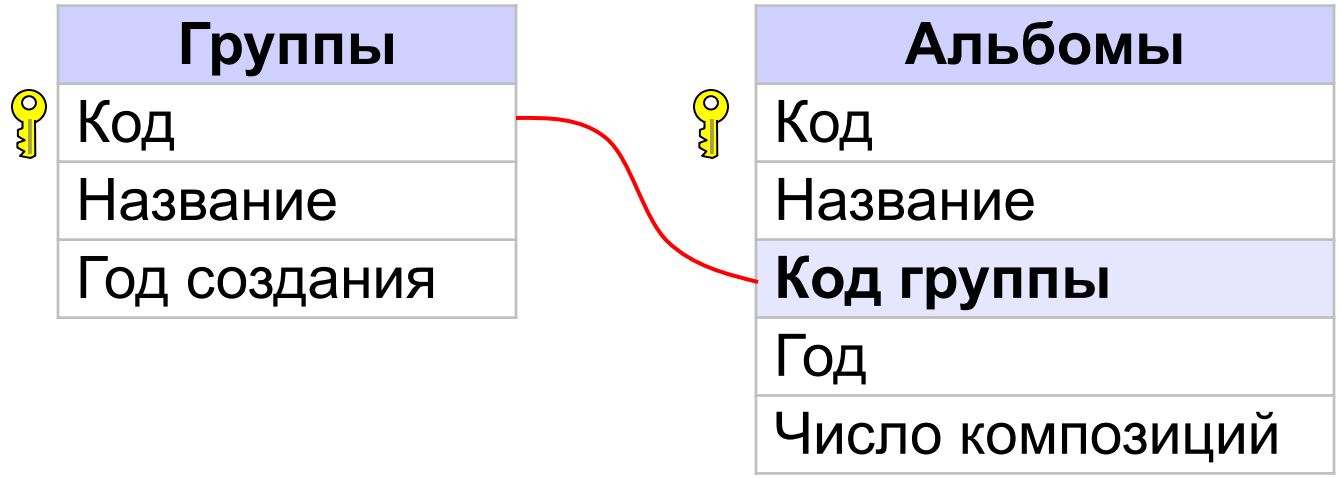
Код	Название	Год создания
1	Машина времени	1969
2	Кино	1981

Альбомы

Код	Название	Код группы	Год	Число композиций
1	Реки и мосты	1	1987	16
2	В круге света	1	1988	11
3	Группа крови	2	1988	11
4	Последний герой	2	1989	10

? Что улучшилось?

Многотабличная БД



Внешний ключ – это неключевое поле таблицы, связанное с первичным ключом другой таблицы.

- убрано дублирование
- изменения нужно делать в одном месте
- некоторая защита от опечаток (выбор из списка)

- усложнение структуры (> 40-50 таблиц – много!)
- при поиске нужно «собирать» данные разных таблиц

Ссылочная целостность

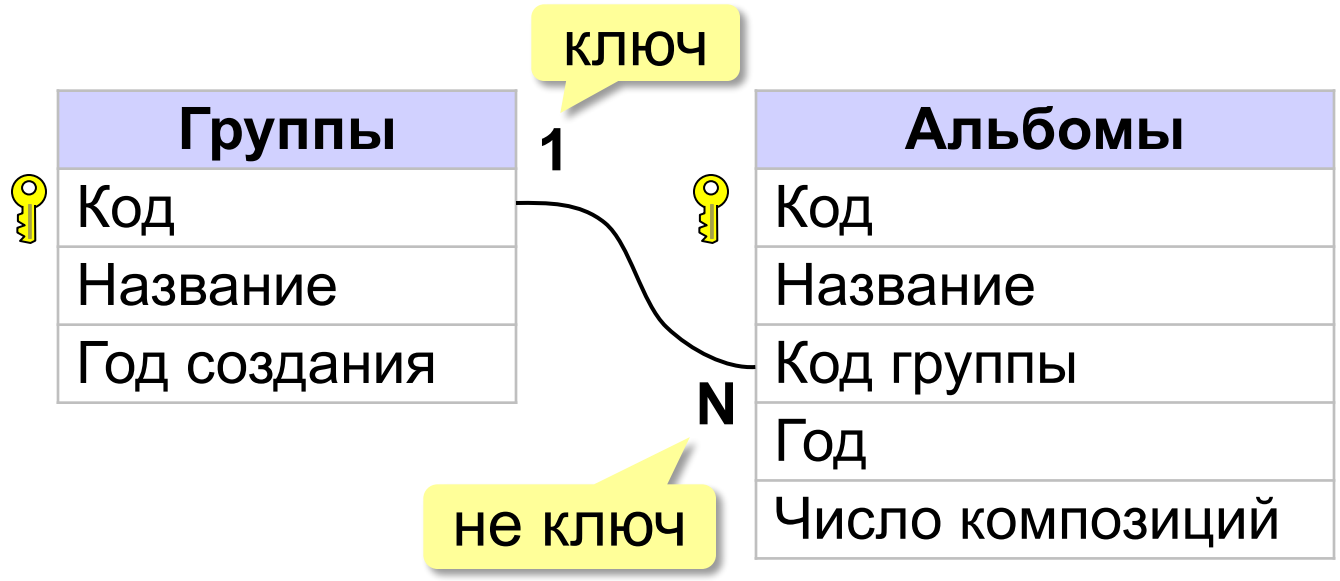
Удаление группы: что делать с альбомами?

СУБД:

- запретить удаление записи
- выполнить каскадное удаление (удалить все связанные с ней записи в других таблицах)
- разрешить внести изменения

нарушится ссылочная целостность!

Типы связей между таблицами



Связь 1:N – с одной записью в первой таблице могут быть связаны сколько угодно записей во второй таблице.

Типы связей между таблицами

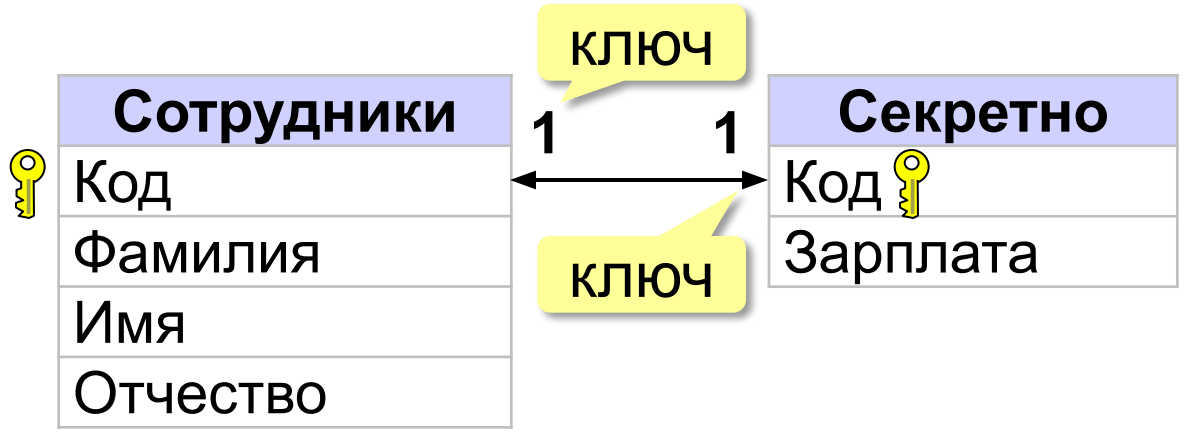
Связь 1:1 – с одной записью в первой таблице связана ровно одна запись во второй таблице.

Сотрудники

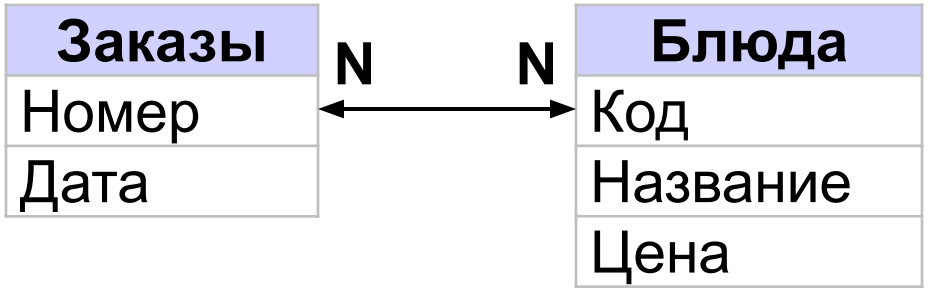
Код	Фамилия	Имя	Отчество
1	Иванов	Петр	Сидорович
2	Петров	Сидор	Иванович
3	Сидоров	Иван	Петрович

Секретно

Код	Зарплата
1	20 000 р.
2	30 000 р.
3	40 000 р.



Типы связей между таблицами



? Может ли быть несколько одинаковых блюд в заказе?

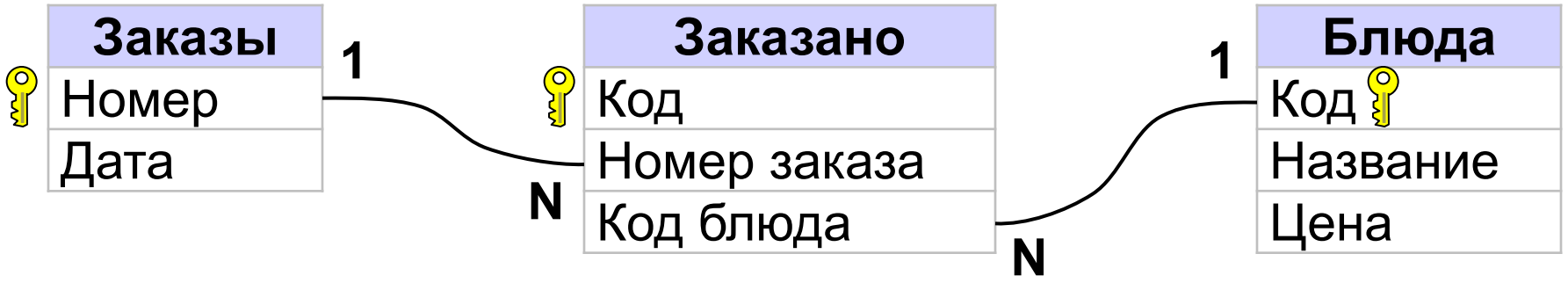
? Может ли быть одно блюдо в нескольких заказах?

Связь N:N – с одной записью в первой таблице могут быть связаны сколько угодно записей во второй таблице, **и наоборот**.

в СУБД не поддерживаются

Типы связей между таблицами

Связь N:N



Пример:

Заказы

Номер	Дата
1	11.12.12
2	12.12.12

Заказано

Код	Номер заказа	Код блюда
1	1	1
2	1	3
3	1	4
4	2	1
5	2	2
6	2	2
7	2	5

Блюда

Код	Название	Цена
1	борщ	80 р.
2	бифштекс	110 р.
3	гуляш	70 р.
4	чай	10 р.
5	кофе	50 р.

? Состав заказов?

Задачи

Сколько заказчиков располагаются в Перми?

Заказчики

Код	Название	Код города
1	ООО «Альфа»	3
2	ЗАО «Бета»	2
3	ООО «Гамма»	3
4	ОАО «Дельта»	2
5	ООО «Каппа»	1

Города

Код	Название
1	Москва
2	Санкт-Петербург
3	Пермь
4	Воронеж
5	Липецк

Задачи

Заказчики

Код	Название	Код города
1	ООО «Альфа»	3
2	ЗАО «Бета»	2
3	ООО «Гамма»	3
4	ОАО «Дельта»	2
5	ООО «Каппа»	1

Города

Код	Название
1	Москва
2	Санкт-Петербург
3	Пермь
4	Воронеж
5	Липецк

Заказы

Накладная	Код заказчика	Артикул	Кол-во упаковок
1011	3	7576	10
1012	5	7576	20
1013	4	3889	25
1014	1	7825	30
1015	3	7576	10

Товары

Артикул	Название	Цена за упаковку
7576	Бумага	150 руб.
2325	Карандаши	200 руб.
3889	Фломастеры	350 руб.
2987	Дневники	400 руб.
7825	Пеналы	250 руб.

Определите:

1) какие товары отправлены в каждый из городов;
2) сколько бумаги отправлено в каждый из городов;
3) общую стоимость товаров, отправленных в каждый из городов.

Задачи

Персоны

Код	ФИО	Пол
71	Иванов Т.М.	М
85	Пановко И.Т.	М
13	Черненко И.А.	Ж
42	Пановко А.И.	Ж
23	Иванова А.И.	Ж
96	Пановко Н.Н.	Ж
82	Черненко А.Н.	М
95	Фукс Т.Н.	Ж
10	Фукс Н.А.	М
…	…	…

Дети

Код родителя	Код ребенка
23	71
13	23
85	23
82	13
95	13
85	42
82	10
95	10
…	…

Определить фамилию и инициалы
а) бабушки А.И.Ивановой
б) родного брата И.А. Черненко
в) прадеда Т.М. Иванова
г) внука И.Т. Пановко

Задачи

Персоны

Код	ФИО	Пол
86	Сизых И.Т.	М
83	Сизых А.И.	М
50	Малых А.Т.	Ж
79	Сидоров Т.М.	М
23	Сидоров А.Т.	М
13	Малых И.И.	Ж
98	Симоняк Т.Н.	Ж
11	Симоняк Н.И.	М
…	…	…

Дети

Код родителя	Код ребенка
98	83
86	13
79	50
86	83
13	50
79	23
13	23
98	13
86	11
…	…

Определить фамилию и инициалы

а) племянника Н.И. Симоняка

б) всех родных братьев и сестёр Н.И. Симоняка

в) бабушки А.Т. Малых

г) дедушки А.Т. Сидорова

Задачи

Персоны

Код	ФИО	Пол
11	Косарева Л.П.	Ж
12	Левитин И.А.	М
24	Шумахер А.Ф.	Ж
45	Бланш А.А.	М
56	Васильева М.А.	Ж
83	Левитин Б.И.	М
94	Левитина В.И.	Ж
115	Кузнецов А.П.	М
140	Левитина Р.Б.	Ж
162	Левитин Л.Б.	М
171	Гайдарова З.Н.	Ж
186	Мурина С.А.	Ж
201	Кузнецов П.А.	М
…	…	…

Дети

Код родителя	Код ребенка
11	83
11	94
12	83
12	94
24	115
56	140
56	162
83	140
83	162
94	186
94	201
115	186
115	201
…	…

Определить фамилию и инициалы
а) всех внуков и внучек И.А. Левитина;
б) родной сестры П.А. Кузнецова;
в) родного брата С.А. Муриной;
г) бабушки Р.Б. Левитиной.

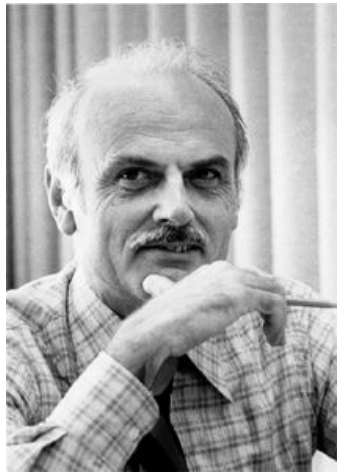
Базы данных

§ 15. Реляционная модель данных

Математическое описание БД

Требуется:

- описание БД, независимое от способа хранения данных
- методы управления этими данными



Эдгар Кодд

Реляционная модель данных (Э. Кодд, 1970):

- все данные – свойства некоторых объектов
- объекты делятся на классы (*сущности*)
- данные об объекте – набор свойств (*атрибутов*):

(*Название: «Кино», Лидер: «В. Цой»*)

кортеж

- порядок свойств не определён
- отношение (*relation*) – множество кортежей, описывающих объекты одного класса
- в отношении нет одинаковых кортежей
- порядок кортежей в отношении не определен

Реляционные БД

Реляционная БД – это БД, которая основана на реляционной модели, то есть представляет собой набор отношений.

Удобно хранить в виде таблицы:

Реляционные БД

На языке таблиц:

- каждая таблица описывает один класс объектов
- порядок расположения полей в таблице не имеет значения
- все значения одного поля относятся к одному и тому же типу данных
- в таблице нет двух одинаковых записей
- порядок записей в таблице не определён

Реляционная БД – это база данных, которая представлена в виде набора таблиц и управляется с помощью языка SQL.

SQL: обращение к данным по названиям таблиц (отношений) и полей (атрибутов).

Основные операции:

- создание и удаление таблиц
- добавление, изменение и удаление записей
- выборка данных

Нормализация

Рейс	От	До	Самолет	Дата
ZX 001	Москва	Берлин	Boeing 737	11.12.2013
ZX 002	Москва	Санкт-Петербург	Airbus A321	12.12. 2013
ZX 003	Санкт-Петербург	Берлин	Boeing 737	13.12. 2013

Что плохо?

Есть дублирование!

Нормализация – это изменение структуры базы данных, которое устраняет избыточность и предотвращает возможные нарушения целостности.

Нормализация на примерах

1. **Любое поле должно быть неделимым**:

Сотрудник	Телефоны
Иванов Петр Сидорович	123-45-67, (901) 111-22-33
Петров Сидор Иванович	345-67-89, (902) 222-33-44

? Что плохо?

Фамилия	Имя	Отчество	Телефон-Дом	Телефон-Моб
Иванов	Петр	Сидорович	123-45-67	(901) 111-22-33
Петров	Сидор	Иванович	345-67-89	(902) 222-33-44

Нормализация на примерах

2. Любое неключевое поле должно зависеть от ключа:

Номер	Автомобиль	Владелец	Телефон
А123АА47	«Лада-Калина»	Иванов	155-77-23
Т234ТТ78	«Ока»	Петров	277-34-67
В345ВВ98	«Мерседес»	Васильев	322-98-44
А345СС47	«Ауди»	Иванов	155-77-23

? Что плохо?

Автомобили **N** **1** **Владельцы**

Номер	Автомобиль	Владелец
А123АА47	«Лада-Калина»	1
Б234ББ78	«Ока»	2
В345ВВ98	«Мерседес»	3
А345СС47	«Ауди»	1

Код	Фамилия	Телефон
1	Иванов	155-77-23
2	Петров	277-34-67
3	Васильев	322-98-44

Нормализация на примерах

3. **Не должно быть одинаковых по смыслу полей**:

Дата	~~Бананы~~	~~Апельсины~~	~~Яблоки~~
21.05.2013	120	78	101
22.05.2013	153	99	65
23.05.2013	87	55	123

! П ? Что плохо? … новый товар!

Продажи **N** — **1** **Товары**

Дата	Товар	Продано
21.05.2013	1	120
21.05.2013	2	78
21.05.2013	3	101
22.05.2013	1	153
…	…	…

Код	Название
1	Бананы
2	Апельсины
3	Яблоки

Нормализация на примерах

4. **Не нужно хранить то, что может быть вычислено**:

Дата	Доходы	Расходы	Прибыль
03.2013	155	128	27
02.2013	178	105	73
01.2013	194	159	35

Что плохо? ...вычислить!

Дата	Доходы	Расходы
03.2013	155	128
02.2013	178	105
01.2013	194	159

прибыль вычисляется через запрос!

Нормализация: «за» и «против»

- устранено дублирование
- все изменения делаются в одном месте
- некоторая защита от опечаток (выбор из списка)
- СУБД поддерживает ссылочную целостность

- усложнение структуры (> 40-50 таблиц – много!)
- замедляется поиск (нужно «собирать» данные из разных таблиц)

Задача

Выполните нормализацию:

Год	Животные	Район	Количество
2009	белки	Нижняя Балка	12
2009	бурундуки	Верхняя Балка	5
2010	еноты	Нижняя Балка	7
2010	еноты	Овраг	3
2010	белки	Верхняя Балка	10

© К.Ю. Поляков, Е.А. Ерёмин, 2013

Задача

Выполните нормализацию:

Год спуска на воду	Название	Проект	Экипаж
1980	Удалой	1155	220 чел.
1985	Адмирал Трибуц	1155	220 чел.
1987	Североморск	1155	220 чел.
1982	Москва	1164	510 чел.
1983	Варяг	1164	510 чел.

Задача

Выполните нормализацию:

Год	Изготовитель	Город	Модель	Скорость	Цена
2007	ВАЗ	Тольятти	1119	165 км/ч	120000 р.
1995	ВАЗ	Тольятти	11113	130 км/ч	50000 р.
1992	КАМАЗ	Набережные Челны	5320	90 км/ч	200000 р.
2006	КАМАЗ	Набережные Челны	55102	90 км/ч	450000 р.
2007	БелАЗ	Жодино	75600	64 км/ч	1200000 р.

Задача

Выполните нормализацию:

Страна	Фамилия	Инструмент	Автор произведения	Место
Россия	Иванов	фортепьяно	Рахманинов	1
Россия	Петров	флейта	Лист	2
Германия	Шмидт	скрипка	Моцарт	3
США	Смит	скрипка	Рахманинов	4
США	Браун	гобой	Моцарт	5