

Информатика

Биологический институт
Национальный исследовательский
Томский государственный
университет

Лекция

Базы данных

Дмитрий Владимирович Курбатский

**ст. преподаватель каф. ихтиологии и
гидробиологии, научный сотрудник ЛБМ БИ ТГУ**

- Зоологический музей (к. 123)
- Кафедра зоологии позвоночных (к. 126)
- Группа ВКонтакте - Курсы "Информатика" и "Информационные технологии:
http://vk.com/i_it_bi_tsu
- Персональный раздел: <http://zoo.tsu.ru/kdv>



Студенту на заметку

- Сайт <http://sql.ru> – про базы данных и работу с ними.

Блок 1

Общие понятия и классификация
БД

База данных (БД)

- организованная в соответствии с определёнными правилами и поддерживаемая в памяти компьютера совокупность данных, характеризующая актуальное состояние некоторой предметной области и используемая для удовлетворения информационных потребностей пользователей.
- совокупность данных, хранимых в соответствии со схемой данных, манипулирование которыми выполняют в соответствии с правилами средств моделирования данных.
- некоторый набор перманентных (постоянно хранимых) данных, используемых прикладными программными системами какого-либо предприятия.
- совместно используемый набор логически связанных данных (и описание этих данных), предназначенный для удовлетворения информационных потребностей организации.

и даже

- представленная в объективной форме совокупность самостоятельных материалов (статей, расчётов, нормативных актов, судебных решений и иных подобных материалов), систематизированных таким образом, чтобы эти материалы могли быть найдены и обработаны с помощью электронной вычислительной машины (ЭВМ) (Гражданский кодекс РФ, ст. 1260).

Вариант необходимых свойств

- БД хранится и обрабатывается в вычислительной системе.
- Данные в БД логически структурированы (систематизированы).
- БД включает схему, или метаданные, описывающие логическую структуру БД в формальном виде (в соответствии с некоторой метамоделью).

СУБД

- Система управления базами данных
- *database management system (DBMS)*
- – совокупность программных и лингвистических средств общего или специального назначения, обеспечивающих управление созданием и использованием баз данных.

Модель данных

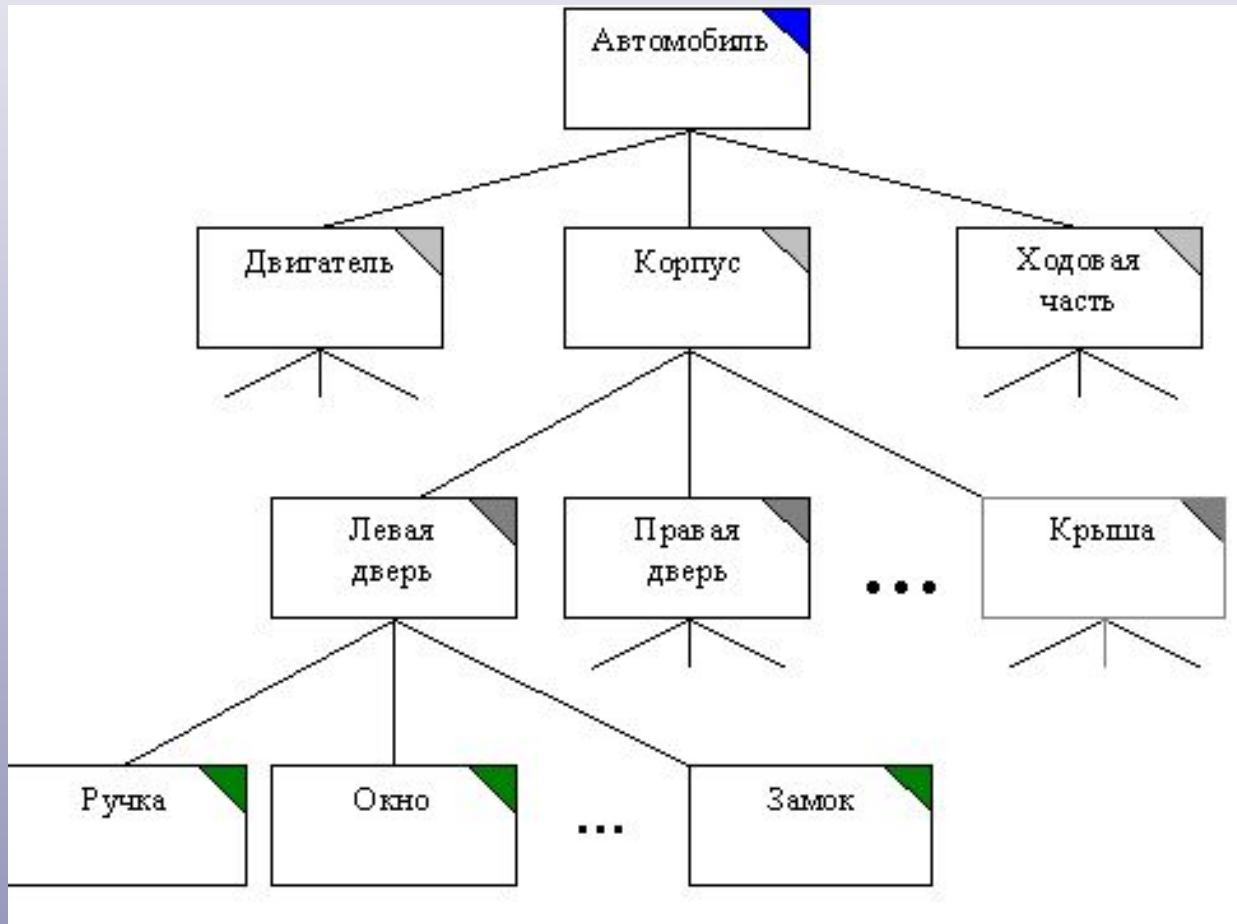
1. аспект структуры
 2. аспект манипуляции
 3. аспект целостности
- аспект защиты
 - аспект практической реализации и применения

Классификация БД

...по модели данных

- Иерархическая
- Сетевая
- Реляционная
- Объектная и объектно-ориентированная
- Объектно-реляционная
- Функциональная
- Документно-ориентированная
- NoSQL
- OBDA
- и др.

Иерархическая БД

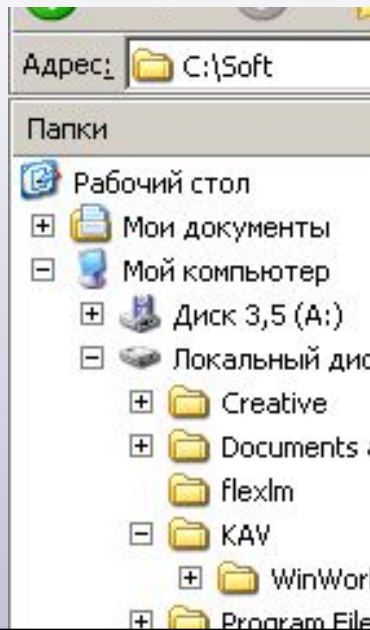


- Information Management System (IMS)
- Time-Shared Date Management System (TDMS)
- Mark IV Multi - Access Retrieval System
- System 2000
- InterSystems Caché
- Google App Engine Datastore API

в операционных системах:

- Серверы каталогов LDAP и Active Directory
- иерархические файловые системы
- Реестр Windows.

Иерархи

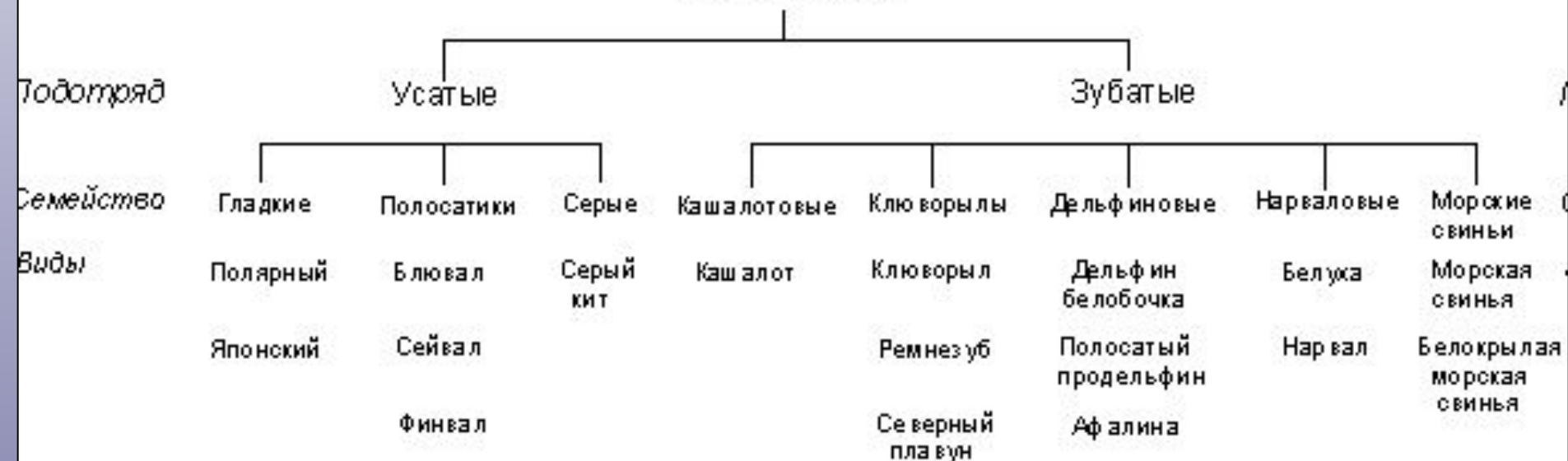


```

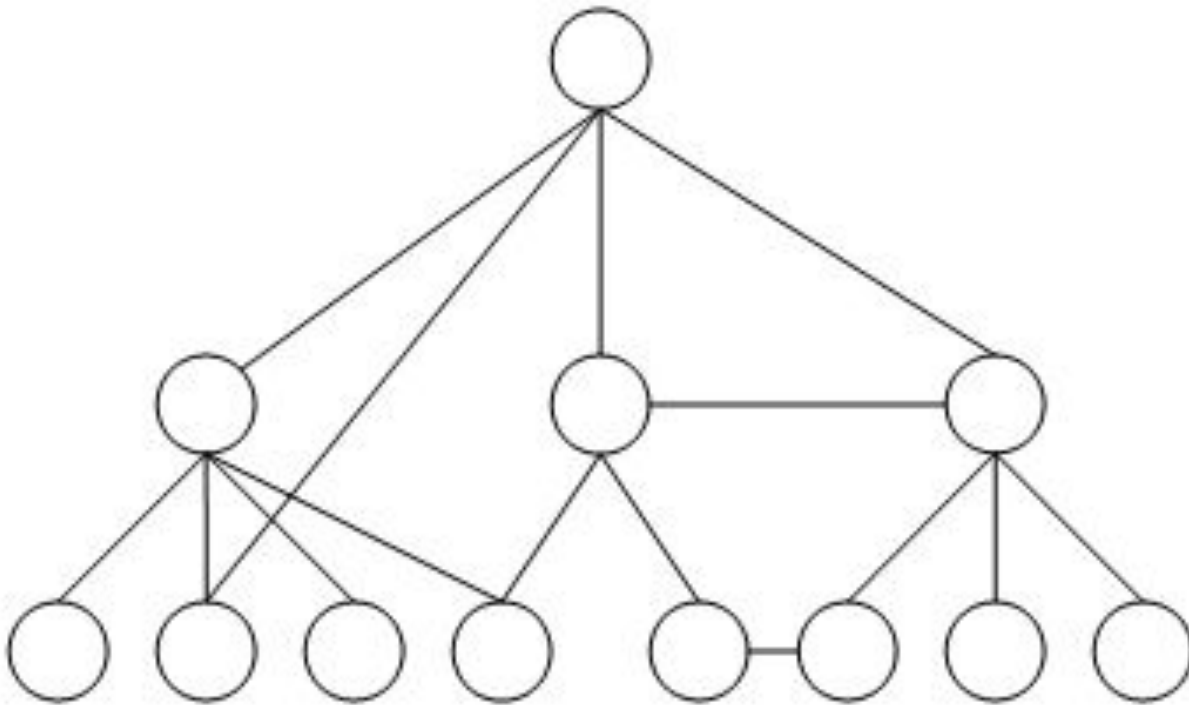
1  <!--DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "
2  <html>
3  <head>
4      <title>Сведения о посещении занятий по курсу "И
5      <meta http-equiv='Expires' content='Mon, 26 Jul
6      <meta http-equiv='Last-Modified' content='Wed,
7      <meta http-equiv='Cache-Control' content='no-sto
8      <meta http-equiv='Cache-Control' content='post-c
9      <meta http-equiv='Pragma' content='no-cache' />
10     <link rel='stylesheet' type='text/css' href='sty
11 </head>
12 <body style='font-size:large;background:#01e5a5;'>
13 <a name='top' />
14 <div style='font-size:20%;'>

```

Китообразные

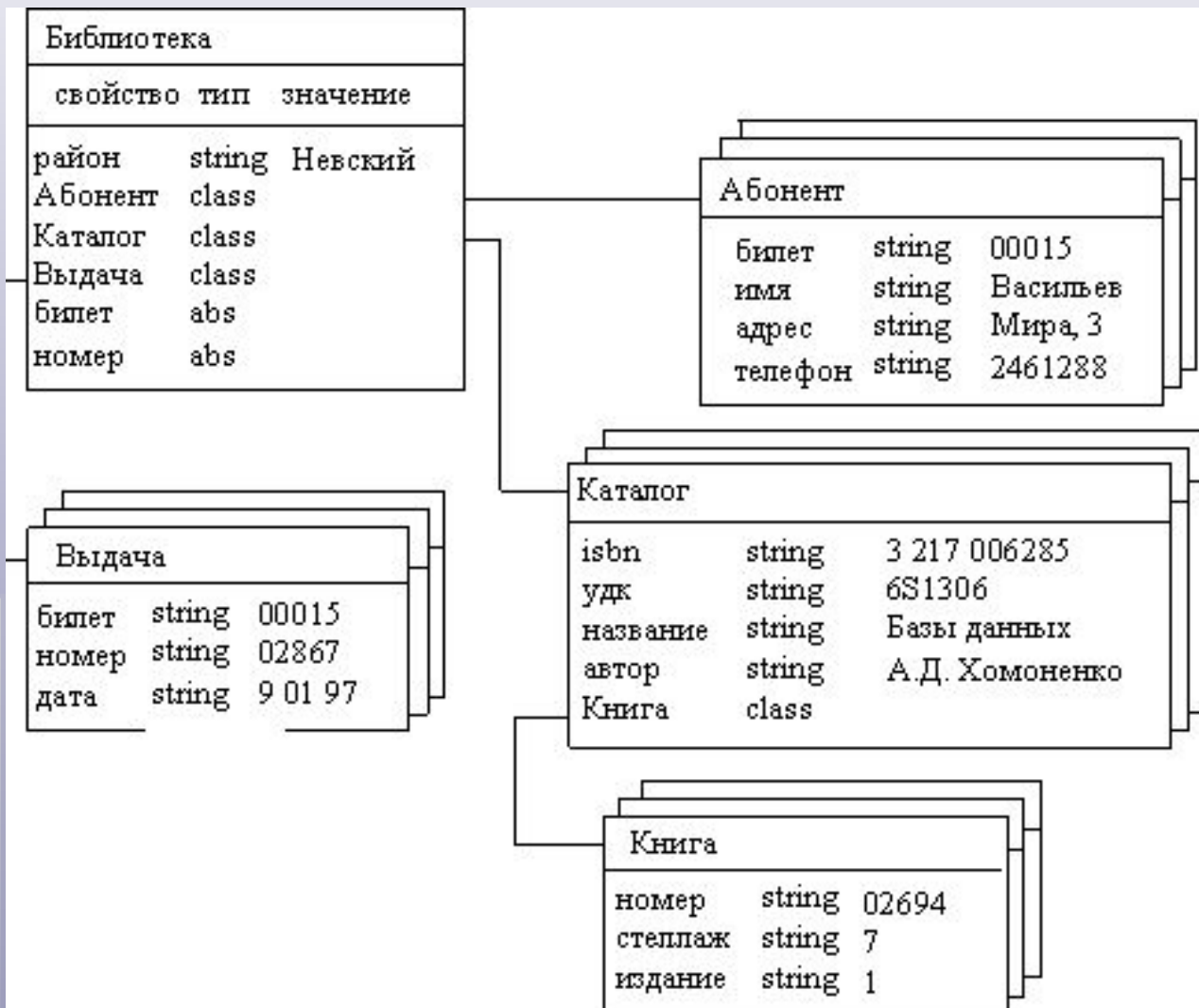


Сетевая модель данных



- СООБЗ
Cerebrum
- ИСУБД
CronosPRO
- dbVista
- Caché
- GT.M

Объектная и объектно-реляционные БД



- Jasmine
- Versant
- POET
- 1C

Документо-ориентированные СУБД

- MongoDB

Embedded Contact Information Document

User Information Document

```
array (  
  '_id' => new MongoId("5146bb52d8524270060001f2"),  
  'age' => new MongoInt32(31),  
  'city' => 'Dallas',  
  'contact_information' =>  
    array (  
      'work_phone' => '678-234-846',  
      'mobile_phone' => '9786756453',  
    ),  
  'email' => 'richard@abc.com',  
  'user_name' => 'Richard Peter',  
)  
  
array (  
  '_id' => new MongoId("5146bb52d8524270060001f3"),  
  'age' => new MongoInt32(25),  
  'city' => 'Los Angeles',  
  'contact_information' =>  
    array (  
      'home_phone' => '345-456-678',  
      'mobile_phone' => '9876543210',  
    ),  
  'email' => 'mark@abc.com',  
  'user_name' => 'Mark Hanks',  
)
```



БД, основанные на онтологиях

- OBDA

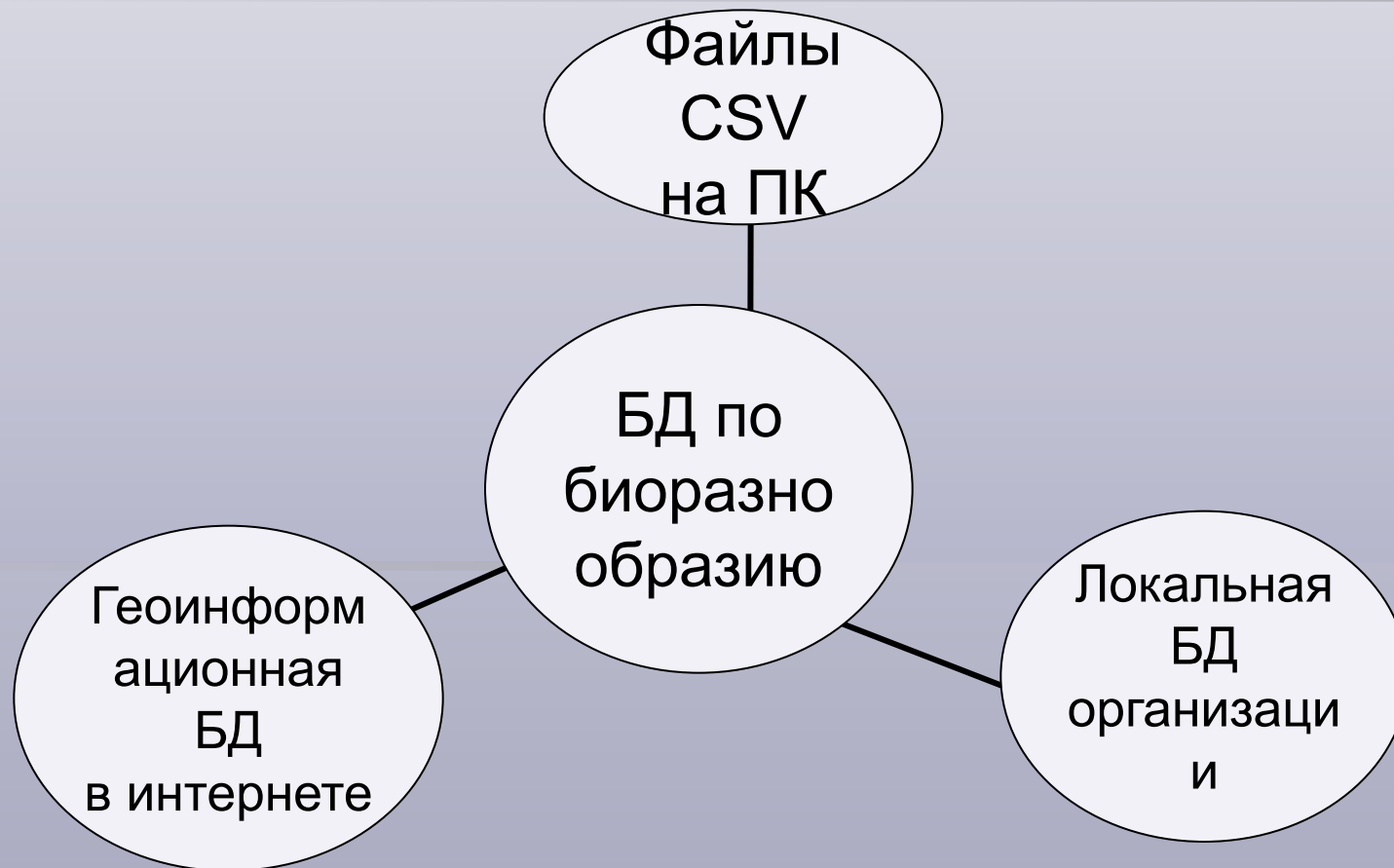
...по среде постоянного хранения

- Во вторичной памяти, или традиционная (*conventional database*)
 - средой постоянного хранения является периферийная энергонезависимая память
 - как правило жёсткий диск
 - в оперативной памяти лишь кеш и данные для текущей обработки
- В оперативной памяти (*in-memory database, memory-resident database, main memory database*)
 - все данные на стадии исполнения находятся в оперативной памяти.
- В третичной памяти (*tertiary database*)
 - средой постоянного хранения является отсоединяемое от сервера устройство массового хранения
 - как правило на основе магнитных лент (в прошлом) или оптических дисков
 - во вторичной памяти сервера хранится лишь каталог данных третичной памяти, файловый кеш и данные для текущей обработки; загрузка же самих данных требует специальной процедуры

...по степени распределённости

- Централизованная
 - или сосредоточенная (*centralized database*)
 - БД, полностью поддерживаемая на одном компьютере.
- Распределённая
 - (*distributed database*)
 - БД, составные части которой размещаются в различных узлах компьютерной сети в соответствии с каким-либо критерием.
 - Бывает:
 - Неоднородная (*heterogeneous distributed database*): фрагменты распределённой БД в разных узлах сети поддерживаются средствами более одной СУБД
 - Однородная (*homogeneous distributed database*): фрагменты распределённой БД в разных узлах сети поддерживаются средствами одной и той же СУБД.
 - Фрагментированная, или секционированная (*partitioned database*): методом распределения данных является фрагментирование (партиционирование, секционирование), вертикальное или горизонтальное.
 - Тиражированная (*replicated database*): методом распределения данных является тиражирование (репликация).

Неоднородная БД



Секционирование БД

По горизонтали

№ п.п.	№ билета	Группа	Ф.И.О. студента	Место рождения	Место проживания	Дата рождения
1	5134 312321	181	Иванов А.Я.	Томск	общ. № 8, к. 23	12.31.1990
2	5134 457324	182	Валенков У.Ю.	Новосибирск	общ. № 8, к. 13	31.06.1989
3	5134 789312	181	Ушаков П.С.	Томск	общ. № 8, к. 24	11.11.1991
4	1451 463156	183	Сухоруков В.А.	Норильск	пр. Ленина, 36, кв. 123	12.00.2000
Сервер деканата ...						
98	5134 143536	181	Нехорошев О.Г.	Астана, Казахстан	общ. № 8, к. 23	15.05.1992

№ п.п.	№ билета	Группа	Ф.И.О. студента	Задание 1	Задание 2
1	5134 312321	181	Иванов А.Я.	+	+
2	5134 457324	182	Валенков У.Ю.	1	+
3	5134 789312	181	Ушаков П.С.	-	-
4	1451 463156	183	Сухоруков В.А.	+	1
Сервер преподавателя ...					
98	5134 143536	181	Нехорошев О.Г.	+ 19	+

Секционирование БД

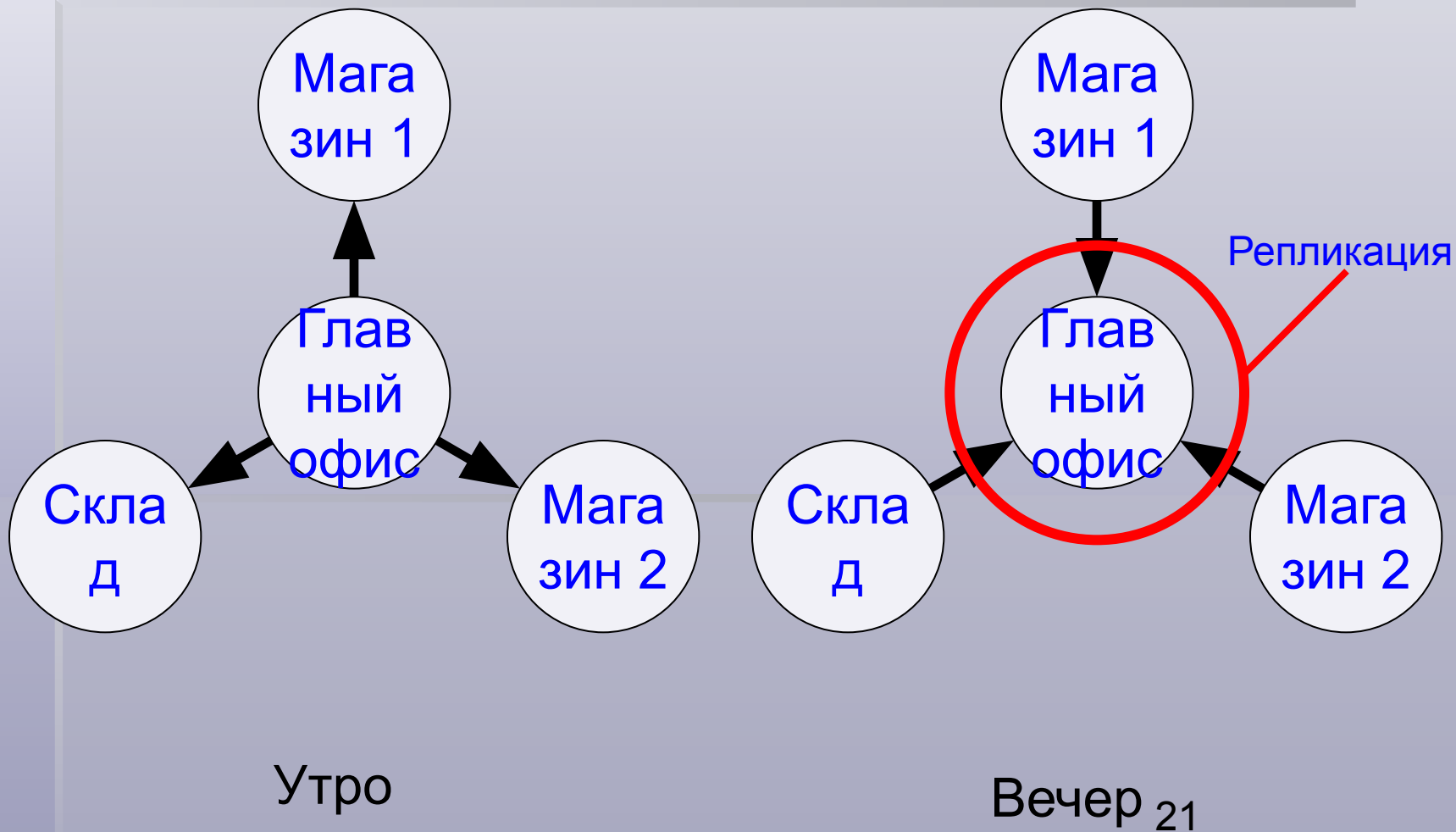
По вертикали

№ п. п.	№ билета	Группа	Ф.И.О. студента	Химия	Физика	Биология
1	5134 312321	181	Иванов А.Я.	+	+	+
2	5134 789312	181	Ушаков П.С.	-	-	-
Сервер куратора гр. 181						
29	5134 143536	181	Нехорошев О.Г.	+	+	+

№ п. п.	№ билета	Группа	Ф.И.О. студента	Химия	Физика	Биология
30	5134 457324	182	Валенков У.Ю.	1	+	+
Сервер куратора гр. 182						

№ п. п.	№ билета	Группа	Ф.И.О. студента	Химия	Физика	Биология
58	1451 463156	183	Сухоруков В.А.	+	1	1
Сервер куратора гр. 183						

Тиражирование



■■■ПО ДИНАМИКЕ ДАННЫХ

- Пространственная
 - spatial database
 - БД, в которой поддерживаются пространственные свойства сущностей предметной области.
 - в ГИС
- Временная, или темпоральная
 - temporal database
 - БД, в которой поддерживается какой-либо аспект времени, не считая времени, определяемого пользователем.
 - журналы и логи программ и ОС
 - блоги и форумы
- Пространственно-временная
 - spatial-temporal database
 - БД, в которой одновременно поддерживается одно или более измерений в аспектах как пространства, так и времени.
 - приёмники GPS
- Циклическая
 - round-robin database
 - БД, объём хранимых данных которой не меняется со временем, поскольку в процессе сохранения данных одни и те же записи используются циклически.
 - в видеорегистрации

...по размеру

Размер:	Способ доступа
■ маленькие ■ 10+...100+ значений ■ побольше ■ 1-2 таблицы ■ 100...100000 строк ■ ещё больше ■ сотни тыс. - миллионы строк ■ десятки - сотни таблиц ■ большие ■ размеры в терабайтах ■ сверхбольшие ■ размеры более петабайта	■ язык программирования ■ электронные таблицы ■ настольные и распределённые реляционные СУБД ■ быстродействующие распределённые СУБД ■ высокопроизводительные нереляционные СУБД

■ ■ ■ по содержанию

- Географическая
- Историческая
- Научная
- Мультимедийная
- и др.

В частности,

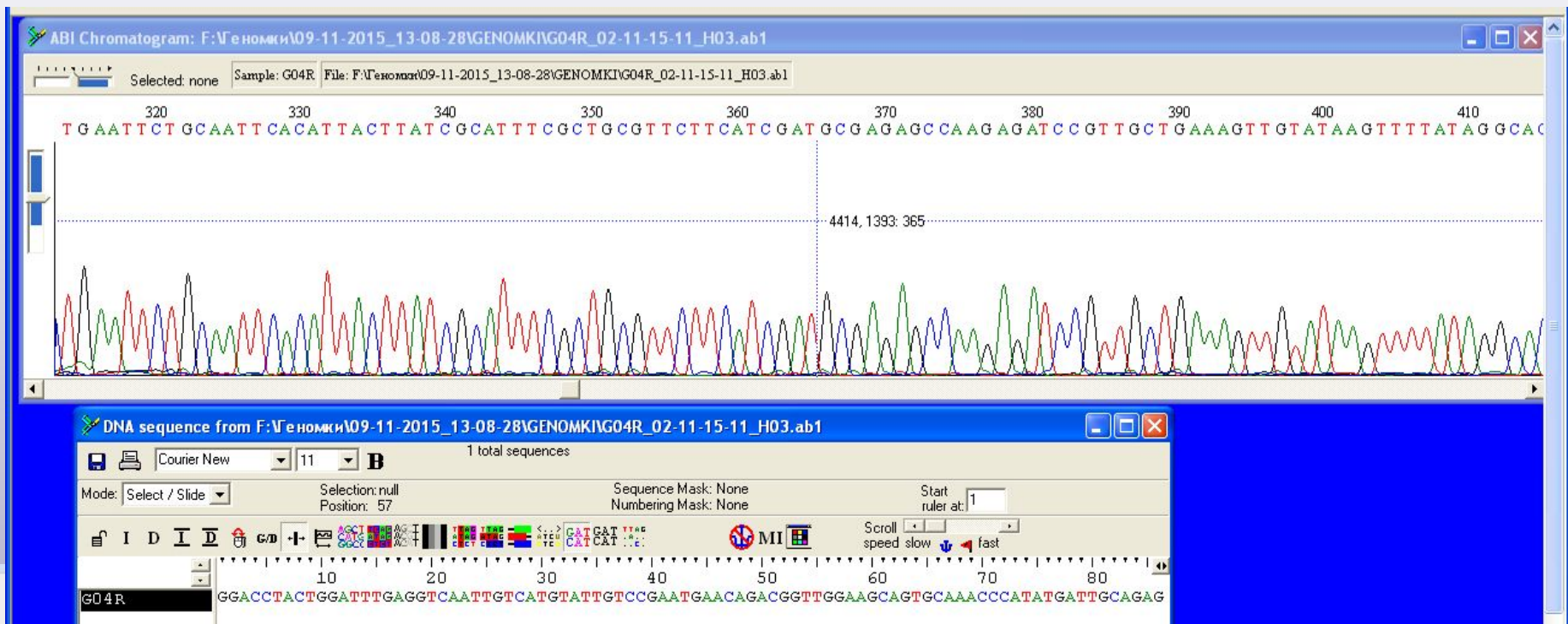
- каталоги музеев и коллекций
 - Зоомузей ТГУ
 - ИСиЭЖ СО РАН
- персональные БД

БД в биологии

- Global Biodiversity Information Facility
www.gbif.org
- www.biodat.eu Biomass Database
- [Обзор БД генома человека](#)
- Погода России и мира, с возможностью скачивания rp5.ru
- Геоинформационные
 - векторные
 - космоснимки
 - климат

GenBank

■ www.ncbi.nlm.nih.gov/Genbank/



Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

Alignments Download GenBank Graphics Distance tree of results						
	Description	Max score	Total score	Query cover	E value	Ident Accession
☐	Lentinus sajor-caju strain dd08058 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and i	1162	1162	98%	0.0	98% FJ810164.1
☐	Pleurotus sp. BAB-4072 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal tran	1158	1158	98%	0.0	98% KJ670292.1
☐	Pleurotus sapidus strain xiaxiu_990 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and	1157	1157	98%	0.0	98% KP877607.1
☐	Pleurotus sapidus strain tedafengwei 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, an	1157	1157	98%	0.0	98% KP877605.1
☐	Lentinus sajor-caju strain S2005 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and inte	1157	1157	98%	0.0	98% JQ868742.1
☐	Pleurotus sajor-caju strain S048 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and inte	1157	1157	98%	0.0	98% AY540328.1
☐	Pleurotus pulmonarius var. shenii 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA ge	1157	1157	98%	0.0	98% AY68200.1

Всемирная систематика

- www.catalogueoflife.org
- >1.6 млн видов
 - бактерии, грибы, животные, растения
 - ~84 %

- Выбрать
- Найти
- Введение

Homo sapiens cro-magnonensis Gregory, 1921 (синоним)

Homo sapiens cromagnonensis Gregory, 1921 (синоним)

Homo sapiens grimaldiensis Gregory, 1921 (синоним)

Homo scythicus Bory de St. Vincent, 1825 (синоним)

Homo sinicus Bory de St. Vincent, 1825 (синоним)

Homo spelaeus Lapouge, 1899 (синоним)

Homo troglodytes Linnaeus, 1758 (синоним)

Homo wadjakensis Dubois, 1921 (синоним)

Народные названия:

Народное название	Язык	Страна
Human	English	-

Классификация:

Царство	Animalia	CoL	LSID
Тип/Отдел	Chordata	CoL	LSID
Класс	Mammalia	ITIS Global	LSID
Порядок	Primates	ITIS Global	LSID
Надсемейство	Hominoidea	ITIS Global	LSID
Семейство	Hominidae	ITIS Global	LSID
Род	Homo	ITIS Global	LSID

Распространение:

Africa; Australia; Caribbean; Europe & Northern Asia (excluding China); Middle America; North America; Oceania; South America; Southern Asia

Окружающая среда:

-

Дополнительная информация:

-

Последняя таксономическая ревизия:

Gardner A.L, 26-Jun-2014

Источник:

ITIS Global, Apr 2016



Bibliographic citation:

Orrell T. (custodian) (2016). ITIS Global: The Integrated Taxonomic Information System (version Apr 2016). In: Species 2000 & ITIS Catalogue of Life, 31st October 2016 (Roskov Y., Abucay L., Orrell T., Nicolson D., Flann C., Bailly N., Kirk P., Bourgoin T., DeWalt R.E., Decock W., De Wever A., eds). Digital resource at www.catalogueoflife.org/col. Species 2000: Naturalis, Leiden, the Netherlands. ISSN 2405-8858.

Онлайн ресурс:

http://www.itis.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&search_value=180092

SEINet

Arizona - New Mexico Chapter



Home Specimen Search Images Flora Projects Agency Floras Dynamic Floras Additional Portals Resources Log In New Ac

Home > **Dynamic Map**

Pan, zoom and click on map to capture coordinates, then submit coordinates to build a species list. [More Details](#)

Build Checklist

Apiaceae

Point (Lat, Long): 34.81380,
-105.07324

Radius: 100

Kilometers

Карта Спутник



Taxon:

Display as:

Plant
longevity
☒ perennial

Inflorescence
type ⓘ
☐ umbel
☒ head/capitate/glomerules

flowering time
☐ April
☐ June

38.54817 -107.97363 within 100 km

Species Count: 2

Apiaceae
Aulospermum purpureum
Cymopterus purpurascens



Taxon:

Display as:

Plant
longevity
☐ annual or biennial
☐ perennial

Inflorescence
type ⓘ
☐ umbel
☐ head/capitate/glomerules

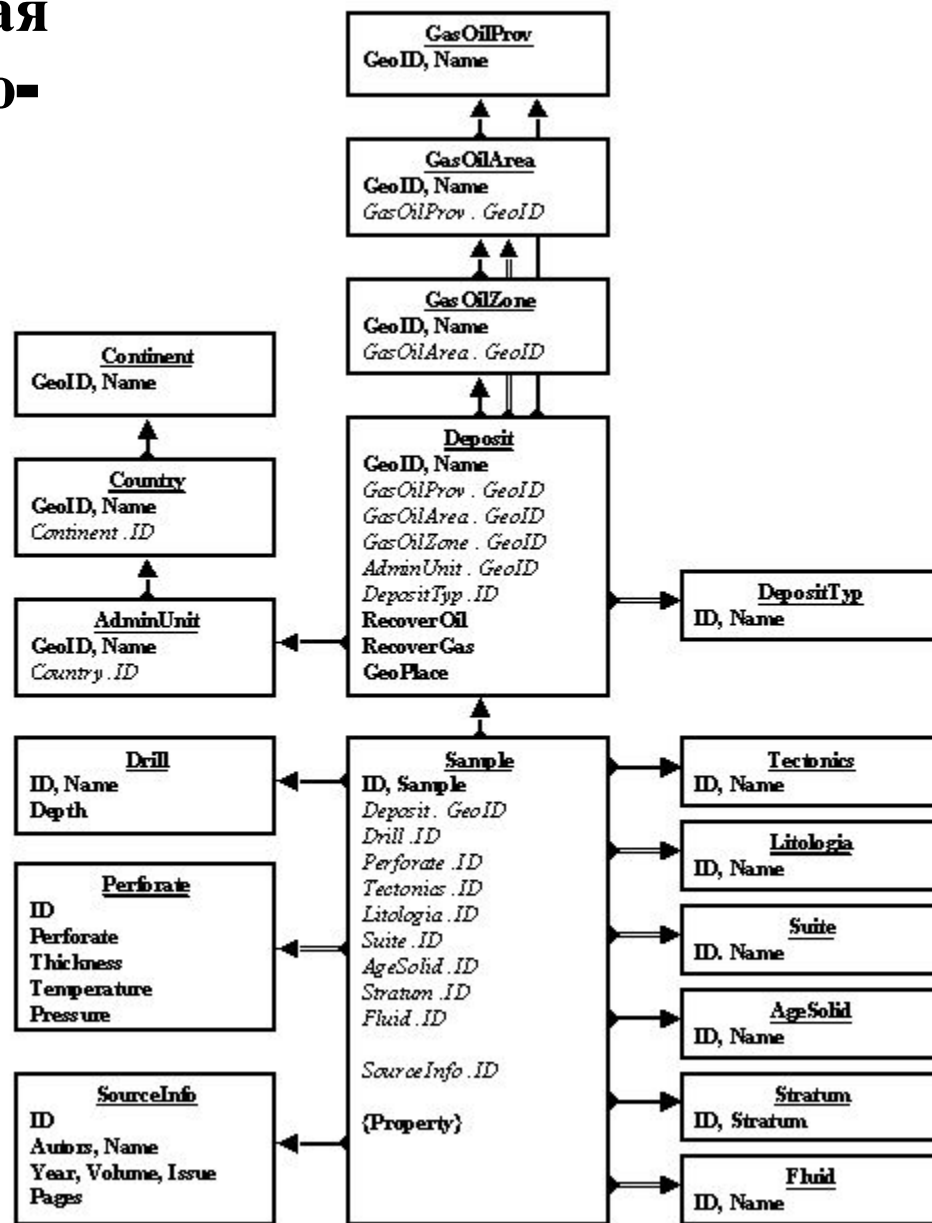
38.54817 -107.97363 within 100 km

Species Count: 43

Apiaceae
Aletes acaulis
Aletes eastwoodiae
Angelica ampla
Angelica grayi
Angelica pinnata
Aulospermum purpureum
Berula erecta
Cicuta bulbifera
Cicuta maculata
Conioselinum scopulorum
Cymopterus bulbosus
Cymopterus constancei
Cymopterus glomeratus
Cymopterus montanus
Cymopterus planosus
Cymopterus purpurascens
Heracleum maximum
Ligusticum porteri
Ligusticum tenuifolium
Lomatium bicolor
Lomatium concinnum

Экологическая и ресурсная информационная база юго-востока Западной Сибири

- enrit.ipc.tsc.ru
- Включает БД:
 - «Химия нефти и газа»
 - «Источники выбросов»
 - «Состояние водных бассейнов»
 - «Мониторинг лесных ресурсов»
 - «Экология почвы»
 - «Экологическая инспекция»
 - «Экологическая экспертиза»
 - «Кадастр экологических данных по лесным землям города Томска»
 - «Ландшафтно – экологическая карта Томской области»
- Другие базы данных



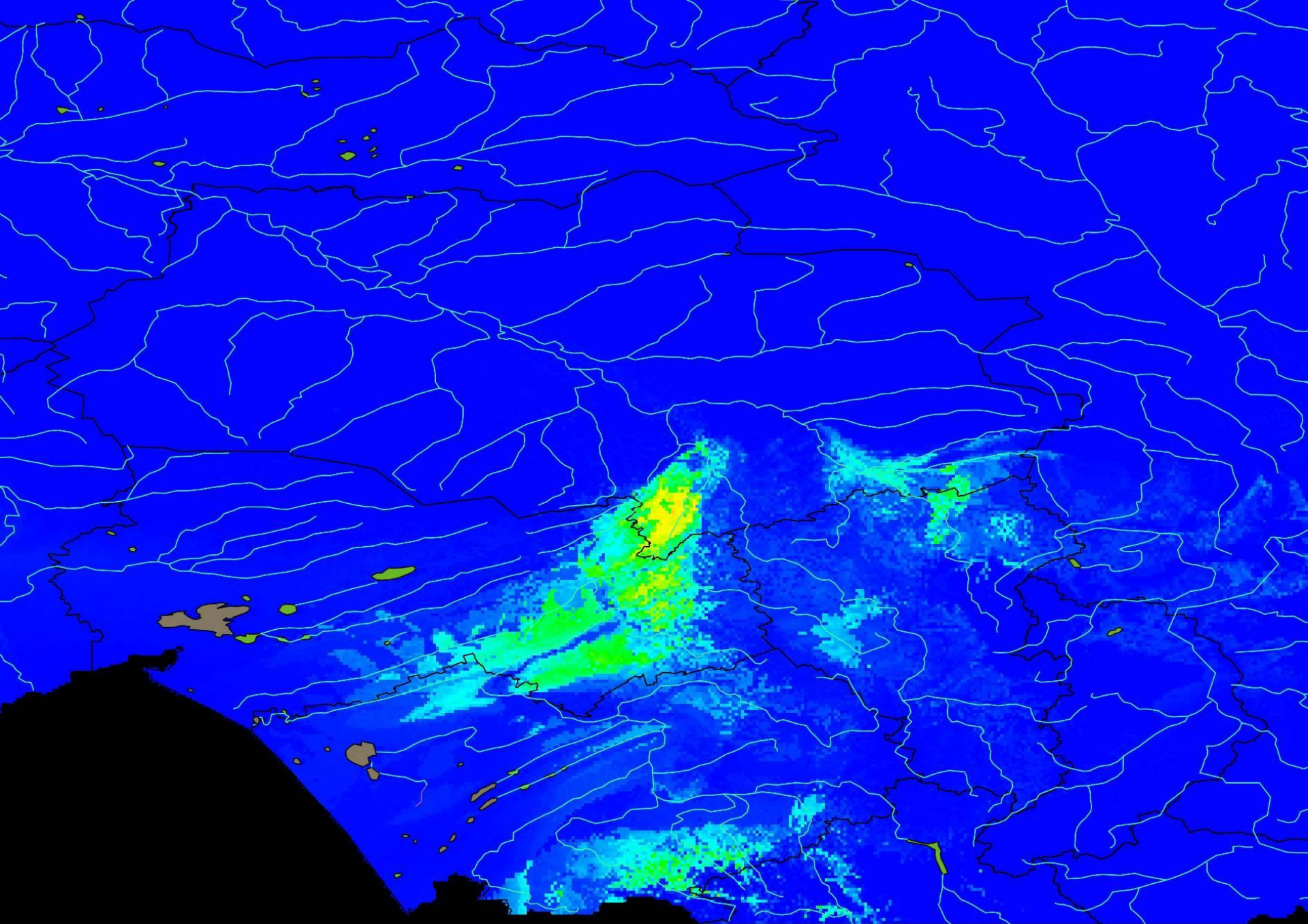
WorldClim - Global Climate Data

- worldclim.org
- Разрешение 30"..10'
- Периоды: 1950-2000; средний голоцен, последний ледниковый максимум; будущее (до 2100 г.)
- Min, max, avg T; осадки; а также
- Bioclim:
 - BIO1 = Annual Mean Temperature
 - BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
 - BIO3 = Isothermality (BIO2/BIO7) (* 100)
 - BIO4 = Temperature Seasonality (standard deviation *100)
 - BIO5 = Max Temperature of Warmest Month
 - BIO6 = Min Temperature of Coldest Month
 - BIO7 = Temperature Annual Range (BIO5-BIO6)
 - BIO8 = Mean Temperature of Wettest Quarter
 - BIO9 = Mean Temperature of Driest Quarter
 - BIO10 = Mean Temperature of Warmest Quarter
 - BIO11 = Mean Temperature of Coldest Quarter
 - BIO12 = Annual Precipitation
 - BIO13 = Precipitation of Wettest Month
 - BIO14 = Precipitation of Driest Month
 - BIO15 = Precipitation Seasonality (Coefficient of Variation)
 - BIO16 = Precipitation of Wettest Quarter
 - BIO17 = Precipitation of Driest Quarter
 - BIO18 = Precipitation of Warmest Quarter
 - BIO19 = Precipitation of Coldest Quarter

Worldclim

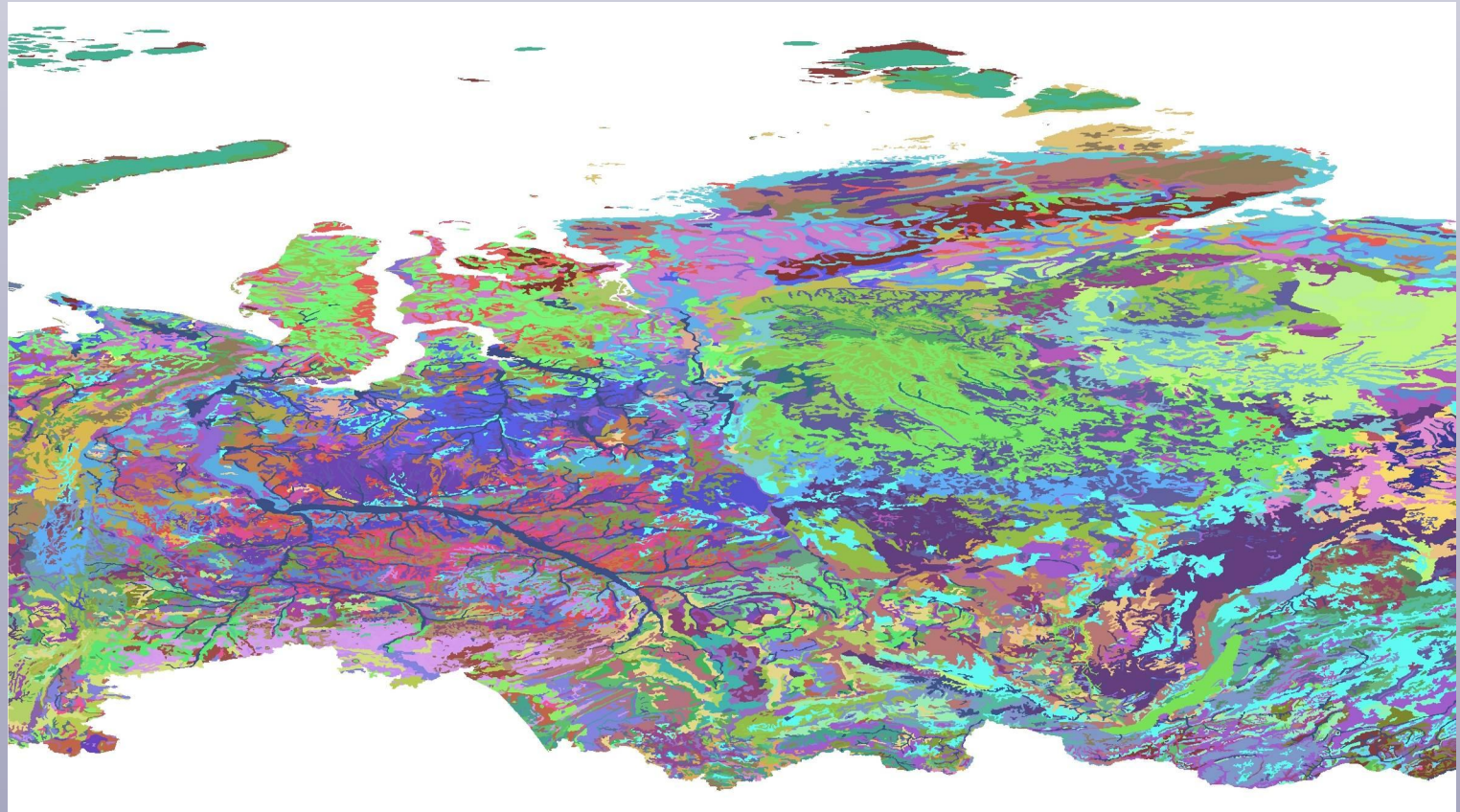
- GeoTIFF
- Grid





Единый государственный реестр почвенных ресурсов России

- egrpr.esoil.ru
- 1 : 2 500 000
- векторные данные (!)



Единый государственный реестр почвенных ресурсов России

- Поле Описание данных
- POLYGON_ID Уникальный номер полигона почвенной карты
- SOIL0_ID Индекс основной почвы (почвенный комплекс, непочвенное образование)
- SOIL0_TXT Наименование выдела легенды основной почвы (почвенный комплекс, непочвенное образование)
- AREA0 Доля основной почвы в полигоне карты, %
- SOIL0_TRL Наименование выдела легенды основной почвы - транслит
- SOIL0_EN Наименование выдела легенды основной почвы - перевод на английский
- WRB_ID Индекс основной почвы по WRB
- WRB_TXT Наименование основной почвы по WRB
- FAO_ID Индекс основной почвы по FAO
- FAO_TXT Наименование основной почвы по FAO
- SYMBOL Символьный индекс почвы
- ZONE Групповая принадлежность выдела легенды
- SOIL1_ID Индекс 1 сопутствующей почвы (почвенный комплекс, непочвенное образование)
- SOIL1_TXT Наименование выдела легенды 1 сопутствующей почвы (почвенный комплекс, непочвенное образование)
- AREA1 Доля 1 сопутствующей почвы в полигоне карты, %
- SOIL2_ID Индекс 2 сопутствующей почвы (почвенный комплекс, непочвенное образование)
- SOIL2_TXT Наименование выдела легенды 1 сопутствующей почвы (почвенный комплекс, непочвенное образование)
- AREA2 Доля 3 сопутствующей почвы в полигоне карты, %
- SOIL3_ID Индекс 3 сопутствующей почвы (почвенный комплекс, непочвенное образование)
- SOIL3_TXT Наименование выдела легенды 1 сопутствующей почвы (почвенный комплекс, непочвенное образование)
- AREA3 Доля 3 сопутствующей почвы в полигоне карты, %
- PARENT1_ID Индекс 1 почвообразующей породы
- PARENT1_TXT Наименование 1 почвообразующей породы
- PARENT2_ID Индекс 2 почвообразующей породы
- PARENT2_TXT Наименование 2 почвообразующей породы
- AREA_M2 Площадь, м.кв.

СУБД

- Система управления базами данных
- *database management system (DBMS)*
- – совокупность программных и лингвистических средств общего или специального назначения, обеспечивающих управление созданием и использованием баз данных.

Функции СУБД

- управление данными во внешней памяти (на дисках);
- управление данными в оперативной памяти с использованием дискового кэша;
- журнализация изменений, резервное копирование и восстановление базы данных после сбоев;
- **поддержка языков БД** (язык определения данных, язык манипулирования данными).

Компоненты СУБД

- ядро
 - отвечает за управление данными во внешней и оперативной памяти, за журнализацию
- процессор языка базы данных
 - обеспечивает оптимизацию запросов на извлечение и изменение данных и создание, как правило, машинно-независимого исполняемого внутреннего кода
- подсистема поддержки времени исполнения (runtime)
 - интерпретирует программы манипуляции данными, создающие пользовательский интерфейс с СУБД
- сервисные программы (внешние утилиты)
 - обеспечивают ряд дополнительных возможностей по обслуживанию информационной системы

Классификации СУБД

- По модели данных
 - Иерархические
 - Сетевые
 - Реляционные
 - Объектно-реляционные
 - Смешанные
- По степени распределённости
 - Локальные СУБД
 - Распределённые СУБД
- По способу доступа к БД
 - Файл-серверные
 - Клиент-серверные
 - Встраиваемые

Связанные понятия

- Информационная система
- База знаний
- Хранилище данных
- Язык Пролог

Блок 2

Реляционные БД

Компоненты БД, СУБД

Некоторые СУБД

Вспомогательная функция =ВПР()...

№ п. п.	Серия и № паспорта	Ф.И.О. преподавателя	Должность
1	ИЗ 905 12321	Москвитин С.С.	ст.преподаватель
2	ПЯ 134 45724	Кулижский С.П.	профессор
3	УЮ 134 78932	Тынзыбаев Т.М.	доцент
4	БЫ 151 463156	Москвитин С.С.	профессор
...			
24	CZ 112343536	Курбацкий М.И.	доцент

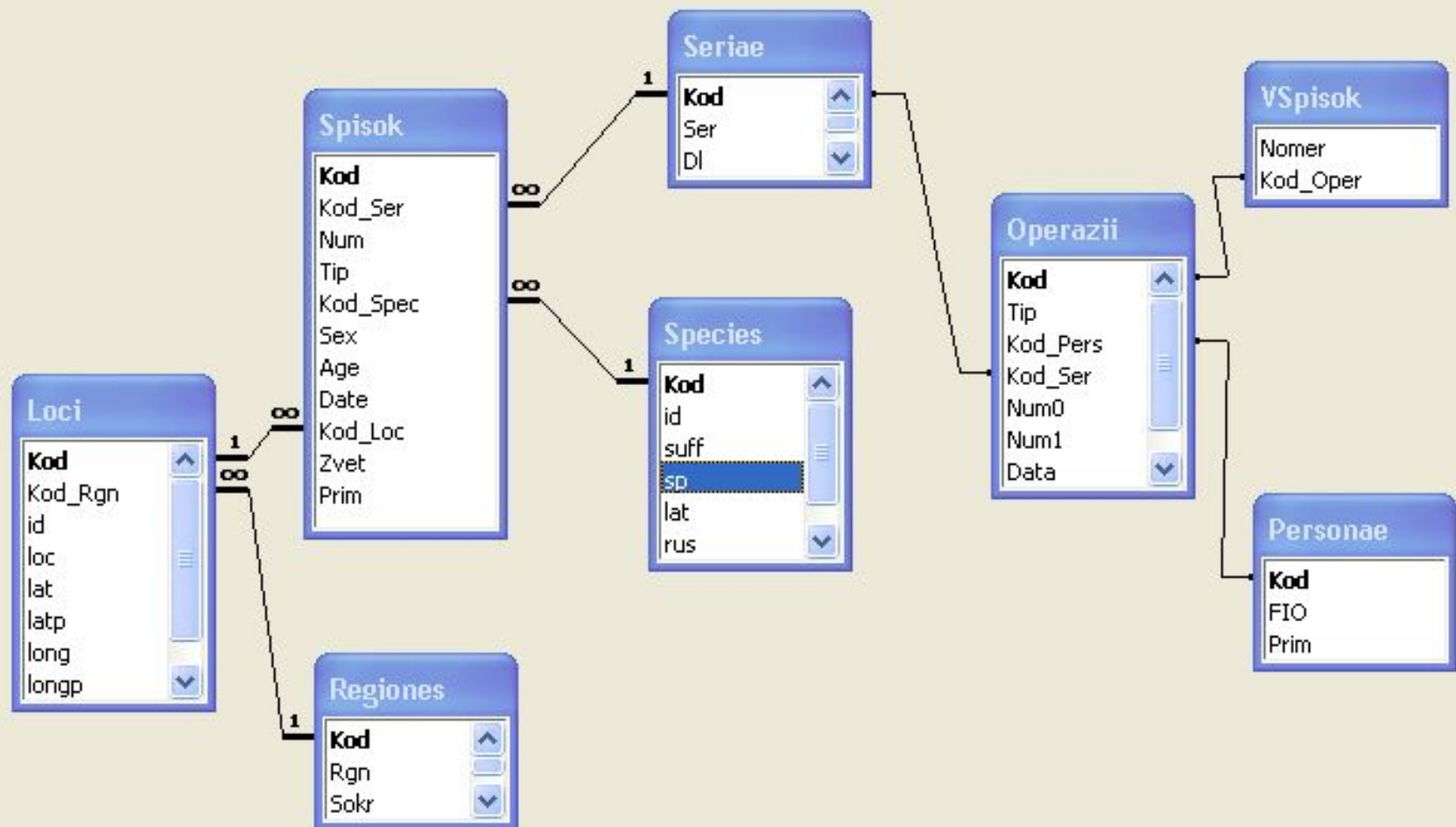
№ п.п.	№ билета	Ф.И.О. студента	Ф.И.О. руководителя
1	5134 312321	Иванов А.Я.	Кулижский С.П.
2	5134 457324	Валенков У.Ю.	Москвитин С.С.
3	5134 789312	Ушаков П.С.	Москвитин С.С.
4	1451 463156	Сухоруков В.А.	Тынзыбаев Т.М.
...			
98	5134 143536	Нехорошев О.Г.	Кулижский С.П.

Вспомогательную функцию =ВПР()...

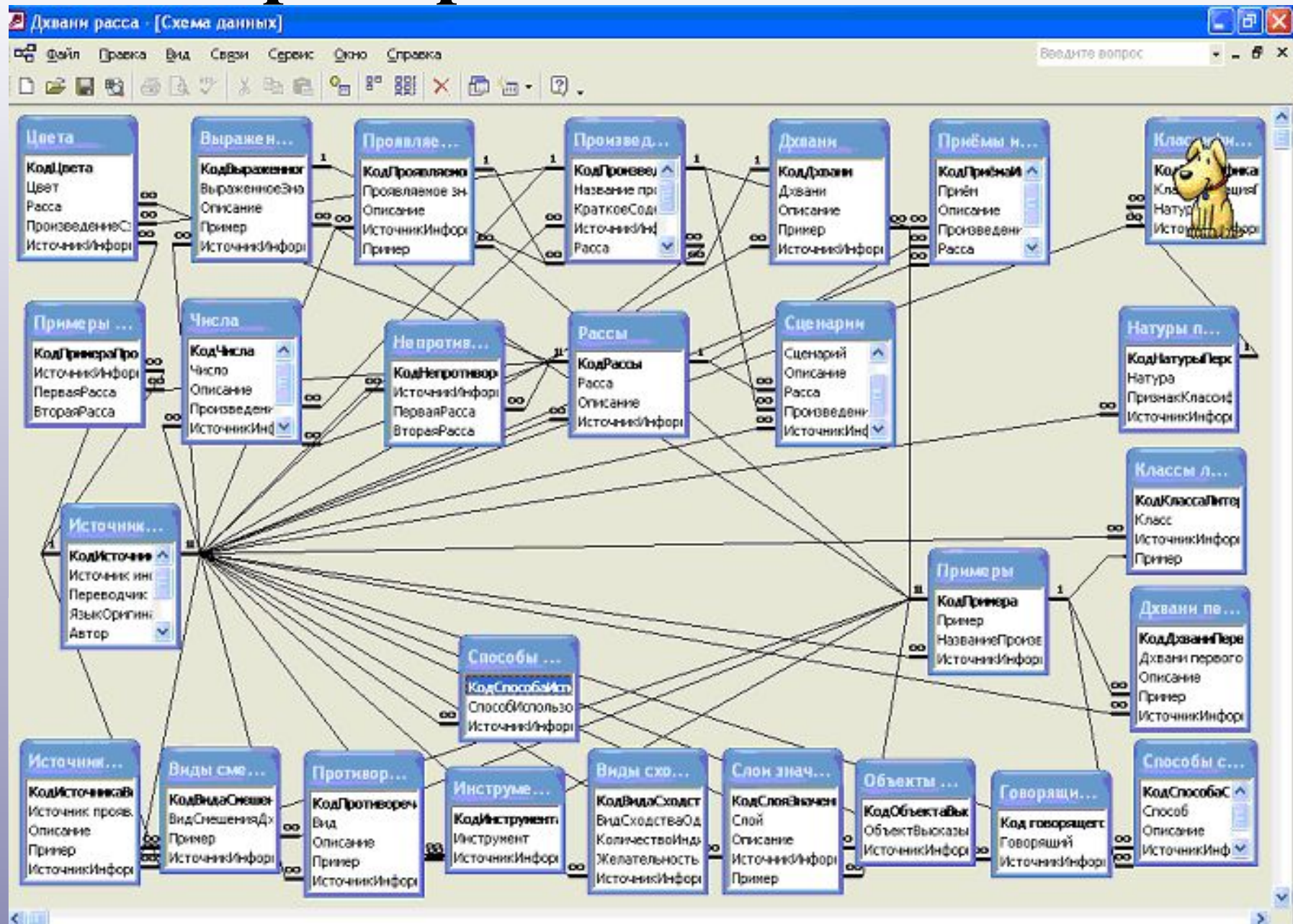
Код

№ п.п.	№ билета	Ф.И.О. студента	Ф.И.О. руководителя
1	5134 312321	Иванов А.Я.	ПЯ 134 45724
2	5134 457324	Валенков У.Ю.	ИЗ 905 12321
3	5134 789312	Ушаков П.С.	ЫЫ 151 463156
4	1451 463156	Сухоруков В.А.	УЮ 134 78932
...			
98	5134 143536	Нехорошев О.Г.	ПЯ 134 45724

Реляционные БД



Ещё пример



Аспекты реляционных БД

1. структура данных (объекты данных)
2. целостность данных
3. обработка данных (операторы)

Свойства реляционных БД

1. В таблице нет двух одинаковых строк.
2. Таблица имеет столбцы, соответствующие атрибутам отношения.
3. Каждый атрибут в отношении имеет уникальное имя.
4. Порядок строк в таблице произвольный.

12 правил Кодда

правило 0: Основное правило (Foundation Rule): Реляционная СУБД должна быть способна полностью управлять базой данных, используя связи между данными:

- Чтобы быть реляционной системой управления базами данных (СУБД), система должна использовать исключительно свои реляционные возможности для управления базой данных.

правило 1: Явное представление данных (The Information Rule):

- Информация должна быть представлена в виде данных, хранящихся в ячейках. Данные, хранящиеся в ячейках, должны быть атомарны. Порядок строк в реляционной таблице не должен влиять на смысл данных.

правило 2: Гарантированный доступ к данным (Guaranteed Access Rule):

- Доступ к данным должен быть свободен от двусмысленности. К каждому элементу данных должен быть гарантирован доступ с помощью комбинации имени таблицы, первичного ключа строки и имени столбца.

правило 3: Полная обработка неизвестных значений (Systematic Treatment of Null Values):

- Неизвестные значения NULL, отличные от любого известного значения, должны поддерживаться для всех типов данных при выполнении любых операций. Например, для числовых данных неизвестные значения не должны рассматриваться как нули, а для символьных данных — как пустые строки.

12 правил Кодда

правило 4: Доступ к словарю данных в терминах реляционной модели (Active On-Line Catalog Based on the Relational Model):

- Словарь данных должен сохраняться в форме реляционных таблиц, и СУБД должна поддерживать доступ к нему при помощи стандартных языковых средств, тех же самых, которые используются для работы с реляционными таблицами, содержащими пользовательские данные.

правило 5: Полнота подмножества языка (Comprehensive Data Sublanguage Rule):

- Система управления реляционными базами данных должна поддерживать хотя бы один реляционный язык, который
 - (а) имеет линейный синтаксис,
 - (б) может использоваться как интерактивно, так и в прикладных программах,
 - (в) поддерживает операции определения данных, определения представлений, манипулирования данными (интерактивные и программные), ограничители целостности, управления доступом и операции управления транзакциями (begin, commit и rollback).

правило 6: Возможность модификации представлений (View Updating Rule):

- Каждое представление должно поддерживать все операции манипулирования данными, которые поддерживают реляционные таблицы: операции выборки, вставки, модификации и удаления данных.

правило 7: Наличие высокоуровневых операций управления данными (High-Level Insert, Update, and Delete):

- Операции вставки, модификации и удаления данных должны поддерживаться не только по отношению к одной строке реляционной таблицы, но по отношению к любому множеству строк.

12 правил Кодда

правило 8: Физическая независимость данных (Physical Data Independence):

- Приложения не должны зависеть от используемых способов хранения данных на носителях, от аппаратного обеспечения компьютеров, на которых находится реляционная база данных.

правило 9: Логическая независимость данных (Logical Data Independence):

- Представление данных в приложении не должно зависеть от структуры реляционных таблиц. Если в процессе нормализации одна реляционная таблица разделяется на две, представление должно обеспечить объединение этих данных, чтобы изменение структуры реляционных таблиц не сказывалось на работе приложений.

правило 10: Независимость контроля целостности (Integrity Independence):

- Вся информация, необходимая для поддержания целостности, должна находиться в словаре данных. Язык для работы с данными должен выполнять проверку входных данных и автоматически поддерживать целостность данных.

правило 11: Дистрибутивная независимость (Distribution Independence):

- База данных может быть распределённой, может находиться на нескольких компьютерах, и это не должно оказывать влияние на приложения. Перенос базы данных на другой компьютер не должен оказывать влияния на приложения.

правило 12: Согласование языковых уровней (The Nonsubversion Rule):

- Если используется низкоуровневый язык доступа к данным, он не должен игнорировать правила безопасности и правила целостности, которые поддерживаются языком более высокого уровня.

Пример таблицы данных

Species : таблица							
	Kod	Код	Доп.	Sp.	Латынь	Русское	
+	67	66		☐	Branta canadensis	Канадская казарка	
+	68	67		☐	Branta leucopsis	Белошекая казарка	
+	69	68		☐	Branta bernicla	Черная казарка	
+	70	69		☐	Branta nigricans	Американская казарка	
+	71	70		☐	Rufibrenta ruficollis	Краснозобая казарка	
▶	72	71		☐	Anser anser	Серый гусь	
+	73	72		☐	Anser albifrons	Белолобый гусь	
+	74	73		☐	Anser erythropus	Пискулька	
+	75	74		☐	Anser fabalis	Гуменник	
+	76	75		☐	Chen caerulescens	Белый гусь	
+	77	75 a		☐	Chen rossii	Гусь Росса	
+	78	76		☐	Philacte canagica	Белошей	
+	79	77		☐	Eulabeia indica	Горный гусь	
+	80	78		☐	Cygnopsis cygnoides	Сухонос	
+	81	79		☐	Cygnus olor	Лебедь-шипун	
+	82	80		☐	Cygnus cygnus	Лебедь-кликун	
+	83	81		☐	Cygnus bewickii	Малый лебедь	
+	84	82		☐	Cygnus columbianus	Американский лебедь	
+	85	83		☐	Tadorna ferruginea	Огарь	
+	86	84		☐	Tadorna tadorna	Пеганка	
+	87	85		☐	Tadorna cristata	Хохлатая пеганка	
+	88	86		☐	Anas platyrhynchos	Кряква	
+	89	87		☐	Anas poecilorhyncha	Черная кряква	
+	90	88		☐	Anas crecca	Чирок-свистунок	
+	91	88 a		☐	Anas carolinensis	Зеленокрылый чирок	
+	92	89		☐	Anas formosa	Клоктун	
+	93	90		☐	Anas falcata	Касатка	
+	94	91		☐	Anas strepera	Серая утка	
+	95	92		☐	Anas penelope	Связь	

Связанные термины и понятия

- Отношение
- Заголовок
- Атрибут
- Кортёж
- Домен
- Атом
- Декомпозиция
- Нормализация
- Ключ
- Индекс
- Реляционная алгебра

Термины

отношение

~~таблица~~

атрибуты

~~столбцы~~

заголовки

кортежи

~~строки~~

Фамилия	Имя	Группа	Пол
Иванов	Алексей	123	м
Путин	Е.И.В.	123	м
Колосова	Ольга	123	ж
Конькова	Елена	122	ж
Тыквин	Джордж	123	н
Рабинович	Прохор	120	м

домен

Нормализация

Вид	Конт. лица
Поставщик	Иванов И.И., зам. дир., тел (3254)76-15-95
	Петров П.П., нач. отд. сбыта, тел (3254)76-15-35

АТОМЫ

Декомпозиция

Вид	Должность	Ф.И.О.	Код города	Тел.
Поставщик	зам. дир.	Иванов И.И.	3254	76-15-95
Поставщик	нач. отд. сбыта	Петров П.П.	3254	76-15-35

Нормализация

- Первая нормальная форма (1NF)
- Вторая нормальная форма (2NF)
- Третья нормальная форма (3NF)
- Нормальная форма Бойса — Кодда (BCNF)
- Четвёртая нормальная форма (4NF)
- Пятая нормальная форма (5NF)
- Доменно-ключевая нормальная форма (DKNF)
- Шестая нормальная форма (6NF)

Первая нормальная форма

Наим.	Город	Адрес	Эл. почта	Вид	Конт. лица
Поршнево й з-д	Владими р	Ул. 2-я Кольцевая, 17	info@plunger.ru	Поставщ ик	Иванов И.И., зам. дир., тел (3254)76-15-95 Петров П.П., нач. отд. сбыта, тел (3254)76-15-35
ООО Вымпел	Курск	Ул. Гоголя, 25	pennon@mail.ru	Клиент	Сидоров С.С., директор, тел. (7634)66-65-38
ИЧП Альфа	Владими р	Ул. Пушкинская, 37, оф. 565	alpha323@list.ru	Клиент	Васильев В.В., директор, тел (3254)74-57-45

Исходные данные

Первая нормальная форма

Наим.	Город	Адрес	Эл. почта	Вид
Поршневой з-д	Владимир	Ул. 2-я Кольцевая, 17	info@plunger.ru	Поставщик
Поршневой з-д	Владимир	Ул. 2-я Кольцевая, 17	info@plunger.ru	Поставщик
ООО Вымпел	Курск	Ул. Гоголя, 25	pennon@mail.ru	Клиент
ИЧП Альфа	Владимир	Ул. Пушкинская, 37, оф. 565	alpha323@list.ru	Клиент

Первая нормальная форма

Должность	Ф.И.О.	Код города	Тел.
зам. дир.	Иванов И.И.	3254	76-15-95
нач. отд. сбыта	Петров П.П.	3254	76-15-35
директор	Сидоров С.С.	7634	66-65-38
директор	Васильев В.В.	3254	74-57-45



Студенту на заметку

- Первые 3 формы нормализации в примерах

Отношения **relations**

- один к одному *one-to-one*
- один ко многим *one-to-many*
- многие ко многим *many-to-many*

Один к одному

Первичный ключ

№ п.п.	№ билета	Ф.И.О. студента	Ф.И.О. руководителя
1	5134 312321	Иванов А.Я.	Кулижский С.П.
2	5134 457324	Валенков У.Ю.	Москвитин С.С.
3	5134 789312	Ушаков П.С.	Москвитин С.С.
4	1451 463156	Сухоруков В.А.	Тынзыбаев Т.М.
...			
98	5134 143536	Нехорошев О.Г.	Кулижский С.П.

Внешний ключ

№ билета	Специальность	Рейтинг
5134 312321	Почвоведение	90
5134 457324	Зоология	85
5134 789312	Энтомология	30
1451 463156	Аскаридология	78
...		
5134 143536	Экология	91

Один студент –
одно значение
рейтинга

Один ко многим

Внешний ключ

Первичный ключ

~~Паспорт~~

№ п.п.	№ билета	Ф.И.О. студента	Ф.И.О. руководителя
1	5134 312321	Иванов А.Я.	ПЯ 134 45724
2	5134 457324	Валенков У.Ю.	ИЗ 905 12321
3	5134 789312	Ушаков П.С.	БЫБ 151 463156
4	1451 463156	Сухоруков В.А.	УЮ 134 78932
...			
98	5134 143536	Нехорошев О.Г.	ПЯ 134 45724

Один руководитель –
много студентов

Первичный ключ

Серия и № паспорта	Ф.И.О. преподавателя	Должность
ИЗ 905 12321	Москвитин С.С.	ст.преподаватель
ПЯ 134 45724	Кулижский С.П.	профессор
УЮ 134 78	Имбаев Т.М.	доцент
БЫБ 151 463156	Москвитин С.С.	профессор
...		
СЗ 112343536	Курбацкий М.И.	доцент

Несколько
связанных записей

Нет связанных записей

Многие ко многим

№ п.п.	№ билета	Ф.И.О. студента	Ф.И.О. преподавателя
1	5134 312321	Иванов А.Я.	Кулижский С.П.
2	5134 312321	Иванов А.Я.	Курбацкий М.И.
2	5134 457324	Валенков У.Ю.	Москвитин С.С.
2	5134 457324	Валенков У.Ю.	Кулижский С.П.
2	5134 457324	Валенков У.Ю.	Москвитин С.С.
3	5134 789312	Ушаков П.С.	Тынзыбаев Т.М.
4	1451 463156	Сухоруков В.А.	Кулижский С.П.
4	1451 463156	Сухоруков В.А.	Москвитин С.С.
4	1451 463156	Сухоруков В.А.	Москвитин С.С.
4	1451 463156	Сухоруков В.А.	Курбацкий М.И.

**Много студентов
сдают много
экзаменов**

Исходные данные

Должность	Предмет	Оценка
профессор	Почвоведение	3
доцент	Экология	4
профессор	Энтомология	3
профессор	Почвоведение	5
ст.преподаватель	Зоология	3
доцент	Аскаридология	2
профессор	Почвоведение	3
ст.преподаватель	Зоология	5
профессор	Энтомология	4
доцент	Экология	4

63

Многие ко многим

Первичный ключ

№ п.п.	№ билета	Ф.И.О. студент
1	5134 312321	Иванов А.Я.
2	5134 457324	Валенков У.Ю.
3	5134 789312	Ушаков П.С.
4	1451 463156	Сухоруков В.А.
...		
98	5134 143536	Нехорешев О.Г.

Внешний ключ

Внешний ключ

№	№ билета	Серия и №	Оце
1	5134 312321	ПЯ 134 45724	3
2	5134 312321	CZ 112343536	4
3	5134 457324	БЫ 151 463156	3
4	5134 457324	ПЯ 134 45724	5
5	5134 457324	ИЗ 905 12321	3
6	5134 789312	УЮ 134 78932	2
7	1451 463156	ПЯ 134 45724	3
8	1451 463156	ИЗ 905 12321	5
9	1451 463156	БЫ 151 463156	4
10	1451 463156	CZ 112343536	4

Таблица для связи

Первичный ключ

№ п.п.	Серия и № паспорта	Ф.И.О. преподавателя	Предмет	Должность
1	ИЗ 905 12321	Москвитин С.С,	Зоология	ст.преподаватель
2	ПЯ 134 45724	Кулижский С.П.	Почвоведение	профессор
3	УЮ 134 78932	Тынзыбаев Т.М.	Аскаридология	доцент
4	БЫ 151 463156	Москвитин С.С.	Энтомология	профессор
...				
24	CZ 112343536	Курбацкий М.И.	Экология	доцент

Ключи

Код	№ билета	Группа	Ф.И.О. студента	Код преподавателя
1231	5134 312321	181	Иванов А.Я.	131
2311	5134 457324	182	Валенков У.Ю.	2

- По типу
 - первичный
 - вторичные
 - внешний
- По строению
 - простой
 - составной
- По происхождению
 - естественный
 - суррогатный

01402A
группа + подгруппа

Суррогатный ключ

Первичный ключ

Внешний ключ

№ п. п.	№ билета Вторичный ключ	Ф.И.О. студента	Ф.И.О. руководителя
1	5134 312321	Иванов А.Я	2
2	5134 457324	Валенков В.Ю.	1
3	5134 789312	Ушаков П.С.	4
4	1451 463156	Сухоруков В.А.	3
...			
28	5134 143536	Нехорошев О.Г.	2

Суррогатные
ключи

Первичный ключ

№ п.п.	Серия и № паспорта	Ф.И.О. преподавателя	Предмет	Должность
1	ИЗ 905 12321	Москвитин С.С.,	Зоология	ст.преподаватель
2	ПЯ 134 45724	Кулижский С.П.	Почвоведение	профессор
3	УЮ 134 78932	Тынзыбаев Т.М.	Аскаридология	доцент
4	БЫ 151 463156	Москвитин С.С.	Энтомология	профессор
...				
24	CZ 112343536	Курбацкий М.И.	Экология	доцент

Вторичный
ключ

Индексирование

П

- Пекáри
- Питекантропы
- Питоны
- Пумы

Р

- Росомахи
- Рыбы
- Рыси

Индекс

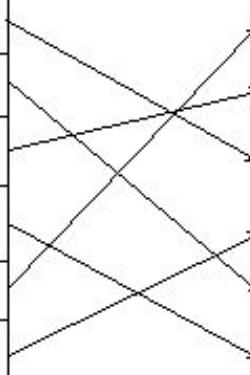
IDX_NAME

3	Алексеев
5	Иванов
2	Казанский
6	Родионов
1	Сергеева
4	Федоров

Таблица

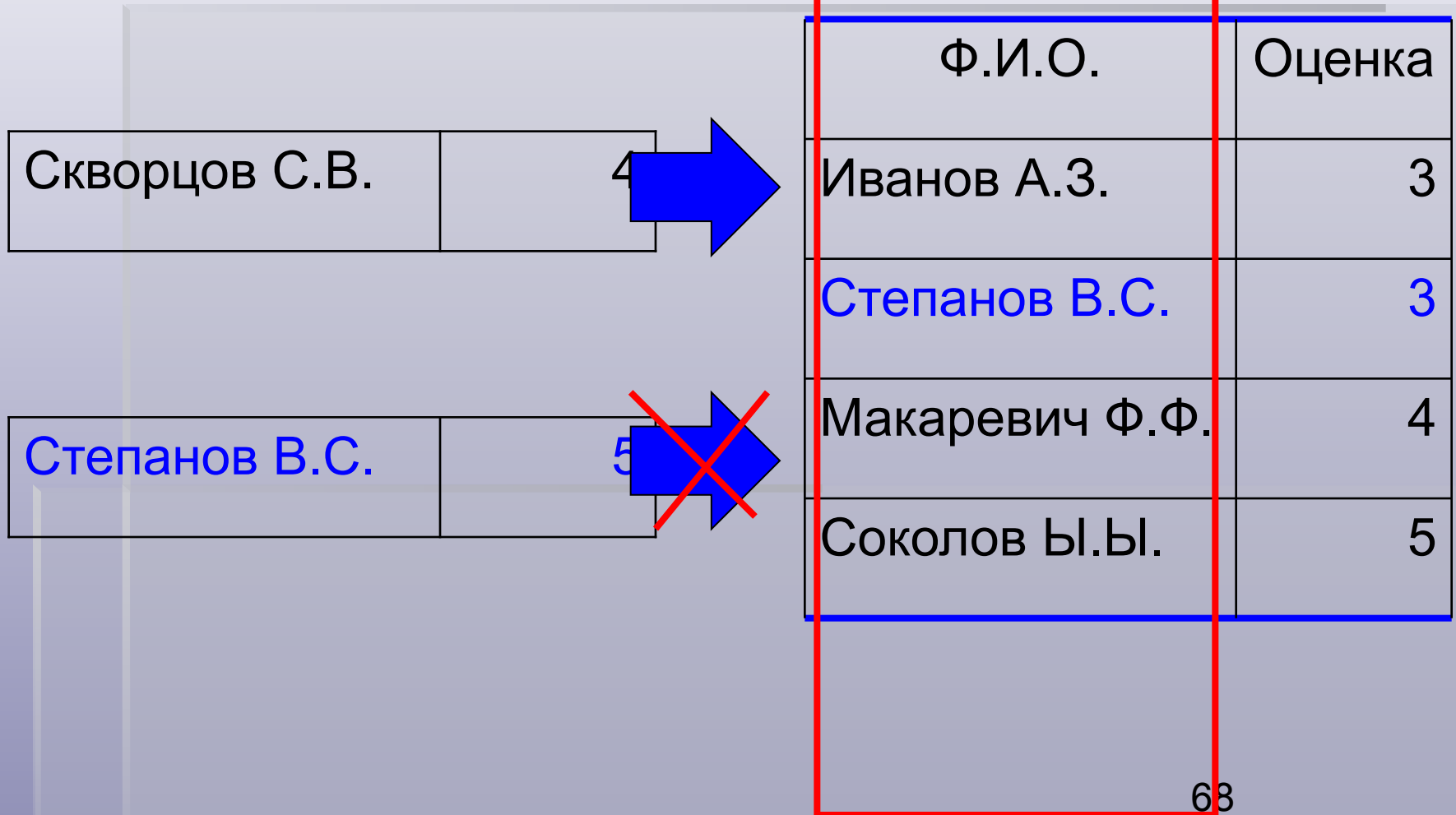
CUSTOMERS

1	Родионов
2	Казанский
3	Алексеев
4	Федоров
5	Иванов
6	Сергеева



- Карточка: пер. №№ : KRegN (407)
- Карточка: русское : KRTaxon (149)
- Карточка: хранение : KFM (10)
- Коллекторы : Coll (128)
- Наборщики : KSSz (4)
- Определители : Det (17)
- По возрасту : Age (6)
- По полу : Sex (3)
- Пол/возр. доп. : SA_Dop (1)
- Препараторы : Praep (22)
- Примечания сверщика : Prim (16)
- Прочее : Etc (14)
- Регистрационные №№ : Reg_N (407)
- Сохранность : Sohran (4)
- Таксоны: Виды : Taxon (124)
- Таксоны: Виды (рус.) : Taxon_R (140)
- Таксоны: Классы : Classis (5)
- Таксоны: Отряды : Ordo (29)
- Таксоны: Перечни охраняемых видов : OSpisok...
- Таксоны: Семейства : Familia (44)
- Таксоны: Типы : Typus (2)
- Типы коллекций : Koll_Tip (1)
- Типы хранения : Hran_Tip (4)
- Удалённые записи : KDel (1)
 - False (163)
- Формы хранения : Forma (11)
 - влажный препарат (22)
 - мумия (3)
 - сверить (47)
 - скелет (13)
 - сухой препарат, имаго (57)
 - тушка (113)
 - часть скелета (1)
 - череп (29)
 - чучело (125)
 - шкура без черепа (7)

Уникальный индекс



Блок 3

Компоненты СУБД

БД и программирование

Язык SQL

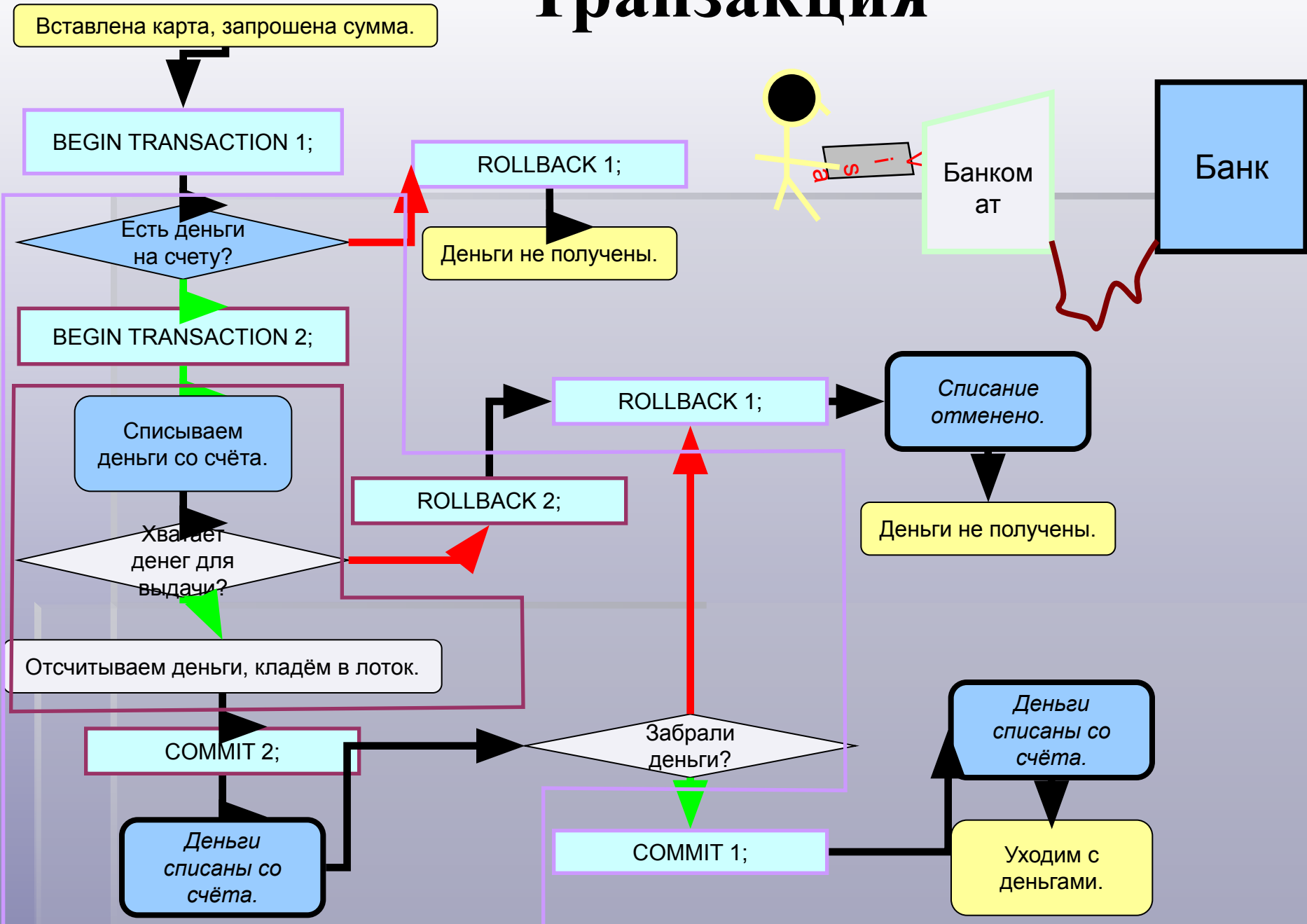
Компоненты СУБД

- Отношение (таблица)
- Представление
- Снимок
- Хранимая процедура
- Триггер
- Курсор
- Индекс
- Запрос

Дополнительные механизмы СУБД

- Транзакция
- Журнализация
- Секционирование
- Репликация

Транзакция



Журналирование

The screenshot displays a software interface with two main panels. The left panel, titled 'Designer', shows a 'Session List' with a tree structure. The right panel shows a data table with columns 'post_id' and 'cat_name'.

Session List:

- Date
- Agent
 - Local (No Agent) - 6
 - 87 - Refresh_Oracle_Tables - Jun 9, 2014 7:28:29 AM
 - Variables
 - 1 - Refresh_Oracle_Tables - Jun 9, 2014 7:28:29 AM
 - 1 - Procedure - Refresh_Oracle_Tables - Load Hive from

Data Table:

	post_id	cat_name
1	25	Uncategorized
2	30	Uncategorized
3	34	Uncategorized
4	42	Uncategorized
5	43	Uncategorized
6	51	Uncategorized
7	54	Uncategorized
8	57	Uncategorized
9	78	Uncategorized
10	83	Uncategorized
11	87	Oracle Database

Некоторые СУБД

- Старые
 - Clipper
 - FoxPro
 - IBM DB2
 - dBase
 - Paradox
- От Micro\$oft
 - Visual FoxPro
 - Microsoft Access
 - MS SQL Server
- Встраиваемые
 - BerkeleyDB
 - Firebird Embedded
 - OpenEdge
 - Microsoft SQL Server Compact
 - SQLite
- Бесплатные
 - MySQL
 - PostgreSQL
 - [OpenOffice Base](#)
- Прочие
 - Caché
 - Firebird
 - Informix
 - Interbase
 - Oracle
 - Sybase Adaptive Server Enterprise
 - ЛИНТЕР

Типы данных БД

(на примере **PostgreSQL**)

- Численные типы
 - Целые
 - С фиксированной точкой
 - С плавающей точкой
 - Денежный тип (отличается специальным форматом вывода, а в остальном аналогичен числам с фиксированной точкой с двумя знаками после запятой)
- Символьные типы произвольной длины
- Двоичные типы (включая BLOB)
- Булев тип
- Типы «дата/время»
- Перечисление
- Геометрические примитивы
- Сетевые типы
 - IP и IPv6-адреса
 - CIDR-формат
 - MAC-адрес
- UUID-идентификатор
- XML-данные
- Массивы
- Идентификаторы объектов БД
- Псевдотипы

Ещё о данных

- Тип Variant
- Пустые и нулевые значения:
 - NULL
 - 0
 - пустая строка ""
- Операторы
 - =, ==, ===
 - Like
 - BETWEEN

Язык запросов

- – искусственный язык, на котором делаются запросы к базам данных и другим информационным системам, особенно к информационно-поисковым системам.
- Примеры:
 - SQL
 - Language Integrated Query — расширение для некоторых языков программирования в .NET Framework
 - XQuery
 - XPath
 - поисковые запросы

Связанные понятия

- Регулярные выражения
- Релевантность

SQL

- Structured Query Language
- язык структурированных запросов



Студенту на заметку

- 3 варианта динамики данных
 - создание
 - удаление
 - модификация
- а также отбор и выдача результатов
- и обмен *между*

Примеры запросов **SQL**

- `SELECT [Имя] FROM [Студенты] WHERE [НомерГр] IN (1, 3, 5, 10);`
- `SELECT * FROM [Студенты] WHERE [Имя] LIKE 'Натал_я';`
- `INSERT INTO [Результаты] ([ФИО], [Оценка] ) SELECT [ФИО], [В_Оценка] FROM [Зачёты];`
- и др.

Состав **SQL**

- операторы
- инструкции
- вычисляемые функции
- таблицы
- поля
- связи и объединения

Возможности SQL

- Изменение структуры данных
 - таблицы
 - поля
 - связи, индексы и пр.
- Изменение данных
 - создание (добавление)
 - модификация
 - удаление
- Поиск и выборка данных
 - расчёты
 - транспонирование
- Управление пользователями и доступом к БД
- Управление работой СУБД
 - транзакции

Операторы SQL

- операторы определения данных (Data Definition Language, DDL)
 - CREATE создает объект БД
 - ALTER изменяет объект
 - DROP удаляет объект
- операторы манипуляции данными (Data Manipulation Language, DML)
 - SELECT считывает данные, удовлетворяющие заданным условиям
 - INSERT добавляет новые данные
 - UPDATE изменяет существующие данные
 - DELETE удаляет данные
- операторы определения доступа к данным (Data Control Language, DCL)
 - GRANT предоставляет пользователю (группе) разрешения на определенные операции с объектом
 - REVOKE отзывает ранее выданные разрешения
 - DENY задает запрет, имеющий приоритет над разрешением
- операторы управления транзакциями (Transaction Control Language, TCL)
 - COMMIT применяет транзакцию.
 - ROLLBACK откатывает все изменения, сделанные в контексте текущей транзакции.
 - SAVEPOINT делит транзакцию на более мелкие участки.

Структура запроса **SQL** на выборку данных

SELECT [ALL | DISTINCT] возвращаемый_столбец, ... | *

FROM спецификатор_таблицы, ...

WHERE условие_поиска

GROUP BY имя_столбца, ...

HAVING условие_поиска

ORDER BY спецификатор_сортировки, ...

Объединения таблиц

- JOIN

Объединение «по горизонтали»

- INNER

- OUTER (Left, Right)

- FULL

- UNION

Объединение «по вертикали»

Преимущества и недостатки **SQL**

■ Преимущества

- Независимость от конкретной СУБД
- Наличие стандартов
- Декларативность

■ Недостатки

- Несоответствие реляционной модели данных
 - Повторяющиеся строки
 - Неопределённые значения (nulls)
 - Явное указание порядка колонок слева направо
 - Колонки без имени и дублирующиеся имена колонок
 - Отсутствие поддержки свойства «=»
 - Использование указателей
 - Высокая избыточность
- Сложность
- Отступления от стандартов
- Сложность работы с иерархическими структурами

SQL-инъекции

Выборка данных из таблицы по условию:

- `SELECT * FROM [Товары] WHERE [Товары]![Название] = "$nazvanie";`
- Инъекция SQL:
 - ~~`$nazvanie="лыжи"`~~
 - `$nazvanie="\"; DELETE * FROM [Пароли] WHERE \"\" = \"\"`
- Результат – очистка паролей:
 - `SELECT * FROM [Товары] WHERE [Товары]![Название] = ""; DELETE * FROM [Пароли] WHERE "" = "";`

Взаимодействие БД и ОС

- ODBC
 - Open Database Connectivity
- DAO
 - Data Access Objects
- ADO
 - ActiveX Data Objects
- ADOdb
- BDE
 - Borland Database Engine



Студенту на заметку

■ API

- интерфейс программирования приложений
- *или* интерфейс прикладного программирования
- *application programming interface*
- набор готовых классов, процедур, функций, структур и констант, предоставляемых приложением (библиотекой, сервисом) для использования во внешних программных продуктах

Объекты **ADO**

- Connection (представляет подключение к удалённому источнику данных)
- Recordset (представляет набор строк, полученный от источника данных)
- Command (используется для выполнения команд и SQL-запросов с параметрами)
- Record (может представлять одну запись объекта Recordset или же иерархическую структуру, состоящую из текстовых данных)
- Stream (используется для чтения и записи потоковых данных, например, документов XML или двоичных объектов)
- Errors (представляет ошибки)
- Fields (представляет столбцы таблицы базы данных)
- Parameters (представляет набор параметров SQL-инструкции)
- Properties (представляет набор свойств объекта)

Bcë.