

Организация баз данных и знаний

Лекция 1



Введение в Базы Данных.
СУБД

Структура курсу

Розділи, теми, модулі	Види занять						Курсова робота	Залік	Іспит	Самостійна робота	
	Всього		лекції	практичні заняття	лабораторні роботи	контрольна робота				розрахунково-графічна робота	самостійна робота
	годин	кредитів									
1	2	3	4	5	6	7	8	9	10	11	12
Модуль 1. Організація баз даних.	24	1,5	16	-	8						48
Тестування за модулем 1.						2				-	
Модуль 2. Фізичні моделі баз даних.	24	2	16	-	8						50
Тестування за модулем 2.						2					
РАЗОМ ЗА СЕМЕСТР	64	3,5	32		16	4	-	-	+	-	98

Расписание занятий

среда, 9/4

GMT+02	
07:00	
08:00	
09:00	08:30 – 10:05 ОБДиЗ
10:00	
11:00	10:25 – 12:00 ОБДиЗ
12:00	
13:00	
14:00	

среда, 9/11

GMT+02	
07:00	
08:00	
09:00	08:30 – 10:05 ОБДиЗ
10:00	
11:00	
12:00	
13:00	

8-30	Основы баз данных и знаний	208	Мартыненко
10-25	Основы баз данных и знаний	208	Мартыненко
	Системное программирование	208	Васильев
12-35	Системное программирование	208	Васильев
14-30	Системное программирование	208	Васильев
	-		
16-25			
18-10			

Содержание курса:

Теоретическая часть

- Введение в базы данных.
- Понятие СУБД.
- Проектирование БД.
- Модели данных.
- Реляционная модель данных.
- Теория нормальных форм.
- Реляционная алгебра.
- Язык SQL.
- СУБД MySQL.

Содержание курса:

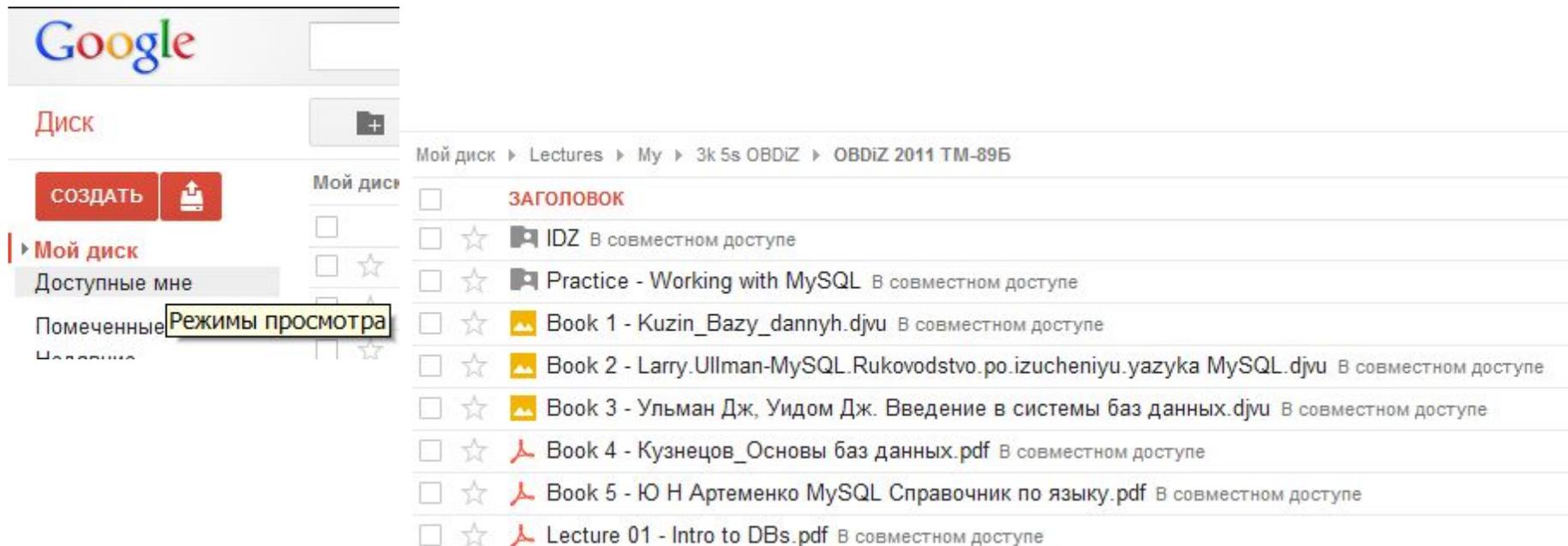
Практическая часть:

- Модуль 1 – Лабораторные занятия:
 - Проектирование баз данных.
 - Работа с логической структурой данных.
 - Построение реляционных баз данных.

- Модуль 2 – Лабораторные занятия:
 - Изучение основ работы с СУБД MySQL
 - Индивидуальное задание на проектирование базы данных.

Работа с курсом

- Лекции – презентации
- Практика – занятие в аудитории
 - Модуль 1 – доска и мел
 - Модуль 2 – компьютерный класс



The screenshot shows the Google Drive interface. At the top left is the Google logo. Below it is the 'Диск' (Drive) header. On the left side, there are buttons for 'СОЗДАТЬ' (Create) and 'Мой диск' (My Drive). The main area displays a folder path: 'Мой диск > Lectures > My > 3k 5s OBDiZ > OBDiZ 2011 TM-89B'. Below the path is a list of files and folders:

- ЗАГОЛОВОК
- ☆ IDZ В совместном доступе
- ☆ Practice - Working with MySQL В совместном доступе
- ☆ Book 1 - Kuzin_Bazy_dannyh.djvu В совместном доступе
- ☆ Book 2 - Larry.Ullman-MySQL.Rukovodstvo.po.izucheniyu.yazyka MySQL.djvu В совместном доступе
- ☆ Book 3 - Ульман Дж, Уидом Дж. Введение в системы баз данных.djvu В совместном доступе
- ☆ Book 4 - Кузнецов_Основы баз данных.pdf В совместном доступе
- ☆ Book 5 - Ю Н Артеменко MySQL Справочник по языку.pdf В совместном доступе
- ☆ Lecture 01 - Intro to DBs.pdf В совместном доступе

A box labeled 'Режимы просмотра' (View modes) is overlaid on the left side of the file list.

Контроль и оценка успеваемости

- Межсессионный теоретический контроль:
 - Модульная контрольная работа 1 - тест
 - Модульная контрольная работа 2 – тест

- Межсессионный практический контроль:
 - Модуль 1 – выполнение индивидуального задания на проектирование БД
 - Модуль 2 – создание и заполнение реляционной БД в СУБД MySQL

- Сессионный контроль:
 - Экзамен – в конце семестра

- Оценка «автоматом»
 - средний бал по всем пунктам.

Тест

Вариант 1

Модульная контрольная работа № 1 по курсу "Интернет-ориентированное программирование".

Имя и фамилия: _____ Группа: _____ Дата: _____

1. Какое из следующих определений верно для HTML?
HTML – расшифровывается как «HyperTransfer Markup Language»
HTML – расшифровывается как «HyperText Markup Language»
HTML – расшифровывается как «HyperText Made Language»
HTML – расшифровывается как «HyperTransfer Markup Language»
2. Какая из следующих функций НЕ возможна в браузере?
поддержка браузерных языков сценариев (JavaScript, VBScript) плагинов, и внешних программ
поддержка серверных языков сценариев (PHP, JSP, Perl, Java, ASP)
получение файлов web-страниц по http-протоколу
просмотр оглавления FTP-серверов и передача файлов по ftp-протоколу
3. Как называется сеть серверов, содержащих таблицы соответствия символьных имен IP-адресам?
Department for Name Servers
Distributed Name Service
Domain-Name Server
Domain Name System
4. Что такое DOM?
Document Object Model
Document Oriental Model
Document Object Mode
Differential Object Model
5. Как расшифровывается название организации W3C?
WarCraft 3 Campaign
World Wild Wet Chicks
World Wide Web Consortium
Warhammer 3 Commandos
6. Сколько октетов входит в IPv4-адрес?
20
3
4
8
7. Какой стране принадлежит домен верхнего уровня http://www.dfat.gov.au?
Armenia – Армения
Angola – Ангола
Australie – Австралия
Albania – Албания
8. Какой из следующих топологий НЕ существует:
линия (line);
шина (bus);
тетраэдр (tetrahedron);
дерево (tree).
9. Локальная вычислительная сеть называется:
LAN;
RUN;
BAN;
FTP.
10. Копирование файл с вашего компьютера на удаленный компьютер в сети называется:
unload;
load;
inload;
upload.

Вариант 1

11. Какая из перечисленных организаций НЕ относится к развитию Интернет?
Internet Society (ISOC)
Web Standards Group (W3G)
Internet Corporation for Assigned Names and Numbers (ICANN)
World Wide Fund for Nature (WWF)
12. Абонент (узел, хост, станция) вычислительной сети это:
устройство, подключенное к сети и активно участвующее в информационном обмене;
человек, имеющий договор с телекоммуникационной компанией;
сторона договора энергоснабжения (потребитель электроэнергии);
физическое лицо, зарегистрированное библиотекой как ее постоянный пользователь.
13. Сколько уровней в эталонной модели OSI:
5
125
7
3
14. Гипертекст это:
текст, в котором используется шрифт большого размера;
текст, набранный на компьютере;
структурированный текст, в котором могут осуществляться переходы по специальным меткам;
очень большой текст.
15. Информационное содержимое сервера, сайта называется:
контент;
NNTP;
ADSL;
Network.

Практические вопросы:

1. История создания и этапы развития сети Интернет.
2. Какие способы создания сайта вы знаете?
Перечислите преимущества и недостатки каждого из них.

Просмотр текущей успеваемости

		Модульная КР1 01.11.2012								
				Кол-во баллов			Итого	Оценка		
				Тест	Практика		Макс			
ФИО	Login	Попытка	Вариант	15	5	5	25	Универ	ESTC	
1 Аразгелдиев Мырат Базарбайович	s901	1	1	7	0	0	7	2	FX	
2 Беда Евгений Сергеевич	s902	1	1	6	0	0	6	2	FX	
3 Дмитренко Алексей Владиславович	s904	1	2	13	2	4	19	4	B	
4 Ерохин Алексей Олегович	s905	1	2	10	4	3	17	4	C	
5 Корнийчук Михаил	s919	1	н				0	н/а	н/а	
6 Краснов Артем Николаевич	s909	1	3	5	0	0	5	2	FX	
7 Кулинич Галина	s921	1	н				0	б	б	
8 Куценко Сергей Владимирович	s911	1	н				0	б	б	
9 Литвиненко Лилия Валериевна	s913	1	2	9	0	0	9	2	FX	
10 Макеев Даниил Геннадиевич	s914	1	1	2	0	0	2	2	FX	
11 Рева Диана Валериевна	s917	1	3	10	5	4	19	4	B	
12 Серебренников Сергей	s920	1	3	7,5	0	0	7,5	2	FX	
13 Татарина Дарина Игорьевна	s918	1	1	14	3	5	22	5	B	
Критерий оценки:										
от										
		0	10,9	2	FX					
		11	13,5	3	E					
		13,6	15,9	3	D					
		16	18,5	4	C					
		18,6	20,9	4	B					
		21	22,9	5	B					
		23	25	5	A					

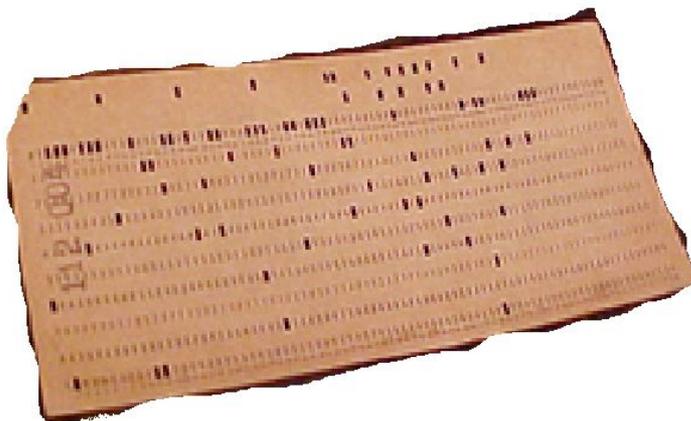
Итоговая оценка успеваемости

	Login	Модульная КР1 Попытка 2							Итоговая оценка за семестр		Оценка в зачетке	
					Кол-во баллов		Итого	Оценка				
		Попытка	Вариант	Тест	Практика	Макс	Универ	ESTC	Универ	ESTC		
ФИО				15	5	5	25					
1 Аразгелдиев Мырат Базарбайович	s901	2	3	15	3	3	21	5	B	5	B	V
2 Беда Евгений Сергеевич	s902									3	D	V
3 Дмитренко Алексей Владиславович	s904									5	A	V
4 Ерохин Алексей Олегович	s905									5	B	V
5 Краснов Артем Николаевич	s909									3	D	V
6 Кулинич Галина	s921									5	A	V
7 Куценко Сергей Владимирович	s911	1	3	15	4	5	24	5	A	5	A	V
8 Литвиненко Лилия Валериевна	s913									5	B	V
9 Макеев Даниил Геннадиевич	s914									3	D	V
10 Рева Диана Валериевна	s917									5	B	V
11 Серебренников Сергей	s920	2	2	10	0	5	15	3	D			
12 Татарина Дарина Игорьевна	s918									5	B	V
		Критерий оценки:										
		от										
		0	10,9	2	FX							
		11	13,5	3	E							
		13,6	15,9	3	D							
		16	18,5	4	C							
		18,6	20,9	4	B							
		21	22,9	5	B							
		23	25	5	A							

Определение Базы Данных

БД это...

«Набор связанных данных»!



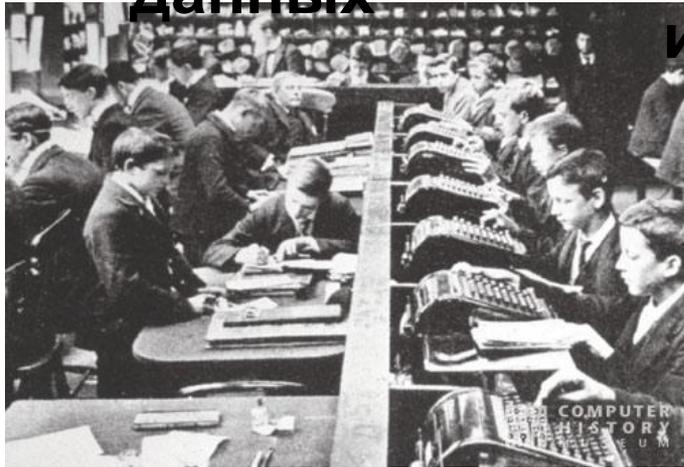
Предпосылки возникновения БД



**Ввод
данных**



**Хранение и извлечение
информации**



**Обработка
запросов**



**Сортиро
вка**

Картотека



- ▣ **Картотека** (др.-греч. χαρτης «лист папируса» и θήκη «место хранения») — упорядоченное собрание данных, как правило на карточках малого формата. Представляет собой каталог.

Картотека

- Каждая карта является информационной единицей и предоставляет сведения о каком-либо хранимом объекте, с целью облегчения поиска этого объекта по определённым признакам.
- Упорядочение осуществляется обязательно по логическим критериям, по алфавиту, дате и т. д.
- Примеры:
 - Картотека пациентов в поликлинике
 - Каталог библиотеки

Картотека

- Картотека, как правило, состоит из единственного собрания "однородных" логически связанных данных, является аналогом таблицы базы данных.
- Особенности:
 - хранение всех данных в одном месте: для поиска требуемой информации необходимо просмотреть всю картотеку;
 - хранение данных с использованием индексирования.

Предшественники баз данных.

□ Основные направления применения компьютерной техники:

1. Численные расчеты.

Исторически оно появилось раньше и способствовало развитию методов численного решения сложных математических задач, развитию языков программирования, ориентированных на решение вычислительных задач.

2. Хранение и обработка данных.

Целью любой информационной системы является хранение и обработка данных о каких-либо объектах реального мира.

Информация и данные

- Информация представляет собой сведения об окружающих человека предметах, явлениях и процессах и является объектом таких операций как восприятие, передача, преобразование, хранение и использование.
- Когда используется термин «данные», то речь идет об информации, представленной в формализованном виде, пригодной для автоматической обработки при возможном участии человека.

Файловые системы

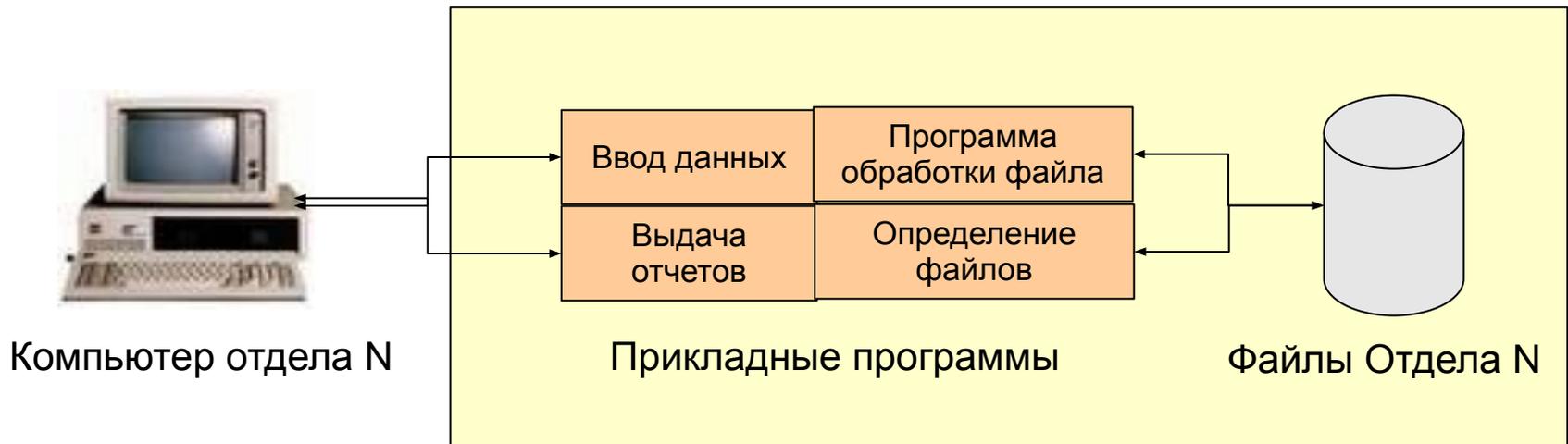
- **Файл** – это неструктурированный объем информации.
- **Файловая система** - набор программ, которые выполняют для пользователей некоторые операции, каждая программа определяет свои собственные данные и управляет ими.
- **Операции над файлами:**
 - удаление, создание файла или каталога, переименование, открытие, закрытие, чтение из файла, поиск, ...

Файловые системы

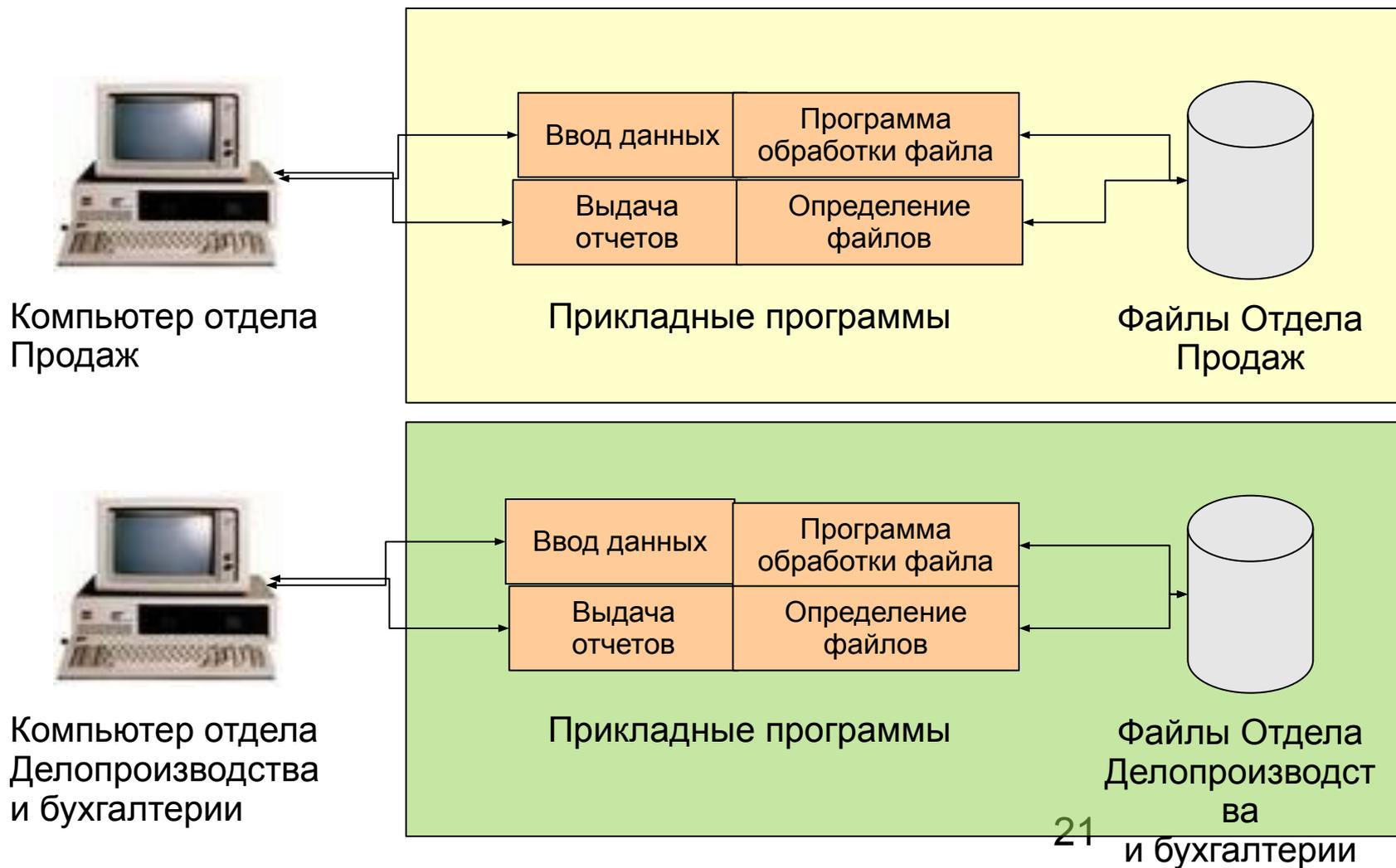
- Метод доступа к информации:
 - различные алгоритмы доступа, которые зависят от внутренней организации файла
- Например:
 - Файл с данными о заказчиках
 - Файл с данными о студентах
 - Метод доступа:
множество операций над множествами студентов или заказчиков.

Файловые системы

- Архитектура 70-х годов.
- Каждый отдел работает со своими данными с помощью своего специализированного приложения, которое в коде жестко определяет физическую структуру и методы хранения записей в файле.



Файловые системы



Недостатки (ограничения)

файловых систем :

- ▣ *разделение и изоляция данных:*
сложность доступа к данным в изолированном файле, возрастающая при увеличении количества файлов, требуемых для получения необходимой информации.
- ▣ *дублирование данных:*
неэкономичное расходование ресурсов, возможность нарушения целостности (появление противоречивых данных).

Недостатки (ограничения)

файловых систем :

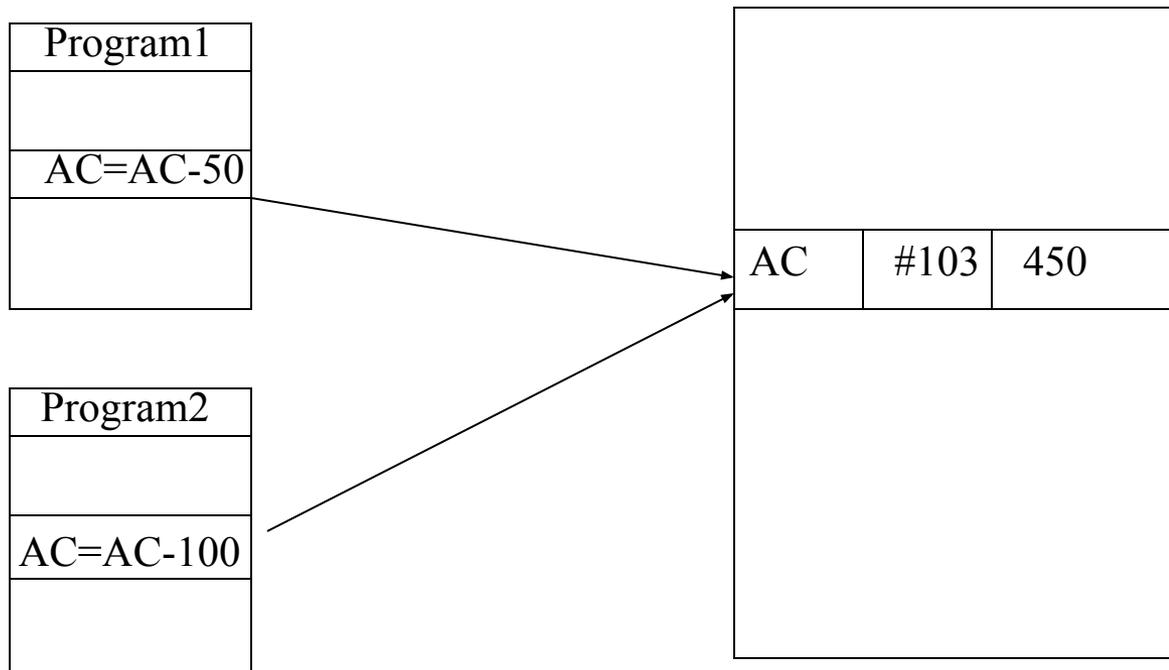
- ▣ *зависимость от данных (и программ):* сложность изменения структуры данных, новый вид данных – новая программа. Приложение - единственный способ доступа к данным.
- ▣ *несовместимость файлов:* вследствие различий языков программирования, использованных при создании приложения.
- ▣ *добавление новой информации* фиксированные запросы/быстрое увеличение количества приложений.

Недостатки файловых систем

- ▣ *целостность хранимой информации*
данные не защищены от неправильной интерпретации при выводе;
вся ли необходимая информация об объекте хранится?
- ▣ *совместный доступ к данным*
параллельное выполнение программ на том же самом файле
- ▣ *безопасность данных*
Трудности с организацией безопасности файлов

Параллельное выполнение программ

Какое конечное значение будет у переменной AC?



Проблемы безопасности

- Разрешить доступ к файлам только уполномоченному персоналу
- Возможность ограничить доступ к частям файлов
- Возможность управления использованием одних и тех же данных разными пользователями
- Защита от несанкционированного использования
- Защита от несанкционированного вывода информации

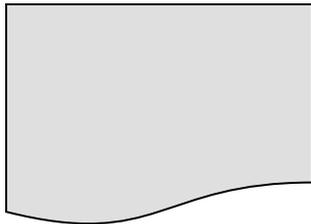
Пример использования файлов для хранения данных

- Допустим, что мы строим информационную систему, в которой нам необходимо хранить информацию о:
 - студентах
 - предметах
 - преподавателях
 - и о том, кто что преподает и кто чему учится

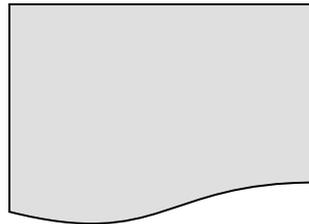
Можно ли это реализовать в файлах ?

Попробуем сохранить наши данные в
файлах:

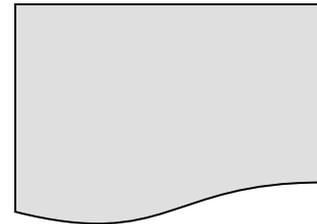
students.txt



courses.txt



professors.txt



Попробуем описать на языке
программирования задачи работы с
данными

Работа с данными

- Добавить преподавателя “Mary Johnson” в курс “CSE444”:

Write a C/Java program to do the following:

Read ‘students.txt’

Read ‘courses.txt’

Find&update the record “Mary Johnson”

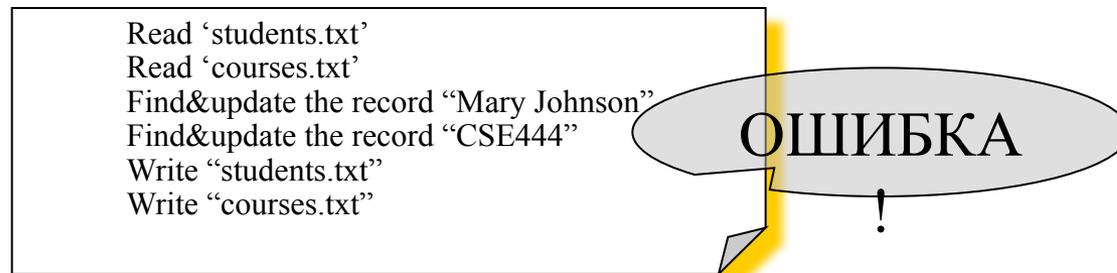
Find&update the record “CSE444”

Write “students.txt”

Write “courses.txt”

Проблемы...

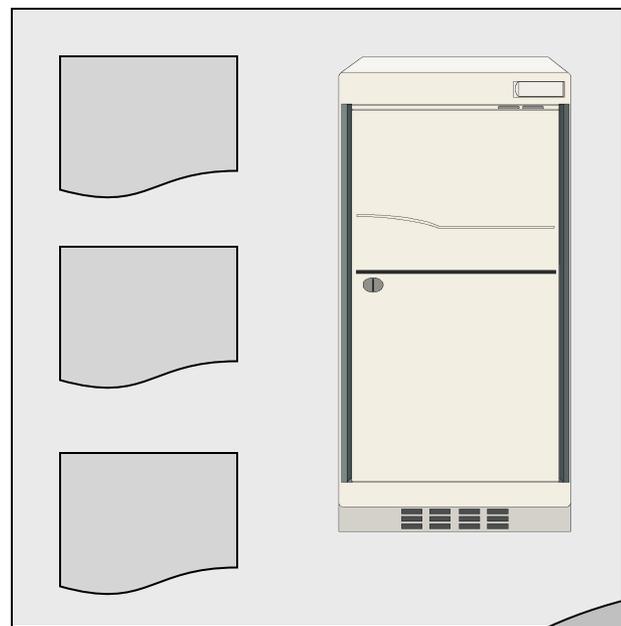
□ Системная ошибка:



- В чем дело?
- Большой объем данных (более 50GB)
 - Проблема с открытием такого файла.
- Одновременный доступ большим числом пользователей
 - Файл students.txt – заблокирован для записи

Необходимо трансформировать архитектуру

Система “client-server” - СУБД



подключение



Файлы
с

данными

Database server
СУБД (еще одна
программа)

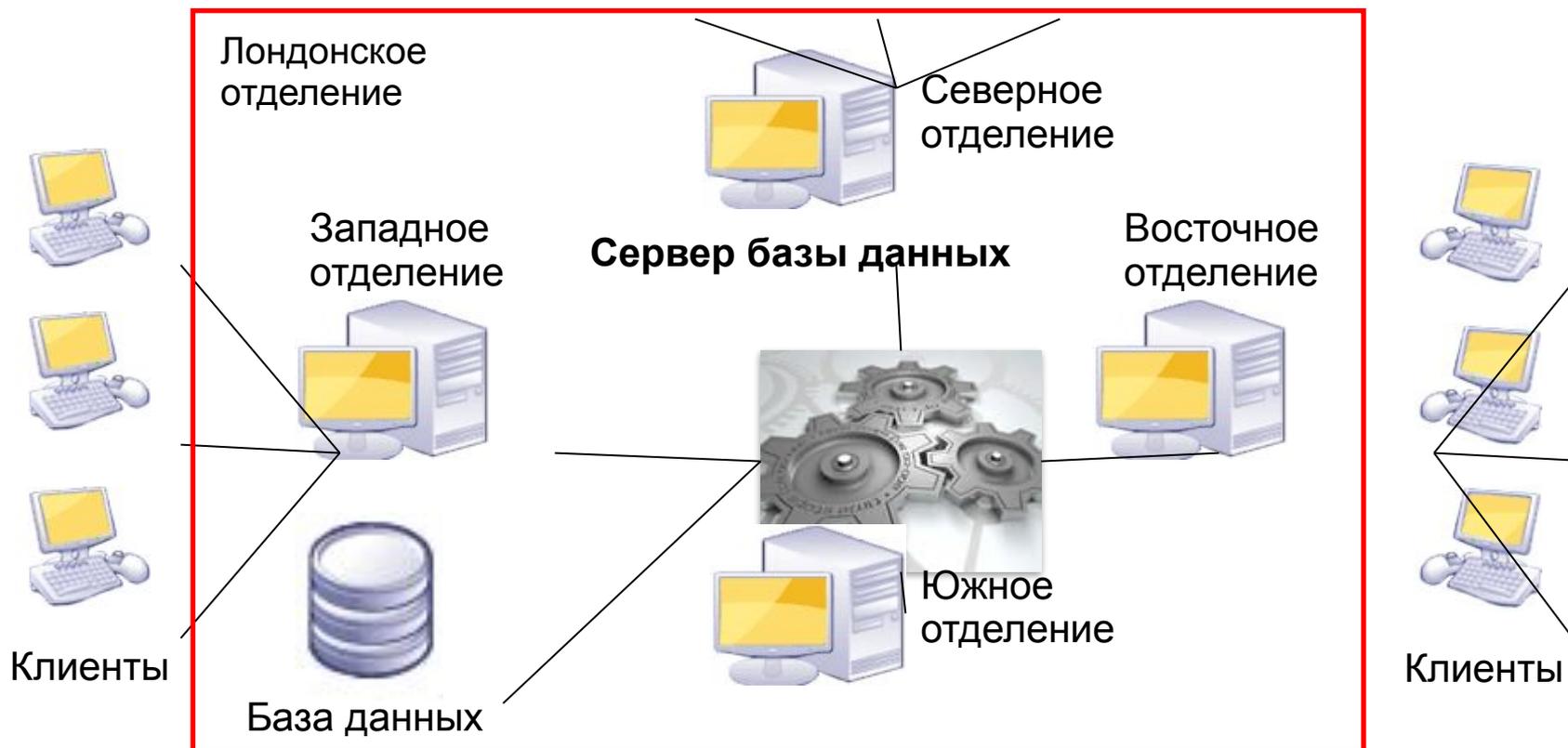
Приложения

я

Новая архитектура



Клиенты



Клиенты

БД – База Данных

- Совместно используемый набор логически связанных данных (и их описание! - Метаданные), предназначенных для удовлетворения информационных потребностей организации
- **База данных** представляет собой большой, структурированный набор данных
- **База данных** содержит модель чего-нибудь из реального мира

Что такое База Данных?

- “Набор информации, хранящейся в компьютере”

Oxford English Dictionary

- “Один или несколько больших структурированных множеств постоянных данных, обычно связанных с программным обеспечением для обновления и запросов данных ”

Free On-Line Dictionary of Computing

- “Совокупность данных, упорядоченная для удобства и быстроты поиска и извлечения ”

Dictionary.com

Понятие «База Данных»

- **База данных** – это совокупность связанных данных, организованных по определенным правилам, предусматривающим общие принципы описания, хранения и манипулирования, независимая от прикладных программ.

- **Отличительные признаки БД:**
 - База данных хранится и обрабатывается в вычислительной системе
 - Данные в базе данных логически структурированы (систематизированы)
 - База данных включает метаданные, описывающие логическую структуру БД

Другие варианты определения «База Данных»

1. совместно используемый набор логически связанных данных (и их описание!), предназначенных для удовлетворения информационных потребностей организации.
2. структурированный организованный набор данных, описывающих характеристики каких-либо физических или виртуальных систем.
3. совместно используемый набор логически связанных данных (и описание этих данных), предназначенный для удовлетворения информационных потребностей организации.

Какие бывают Базы Данных

- Библиотечные каталоги
- Медицинская документация
- Банковские счета
- Складской учет
- Учет персонала
- Каталоги продукции
- Телефонные справочники
- Расписание поездов
- Заказы авиабилетов
- Данные кредитных карт
- Данные о студентах
- История клиентов
- Цены на фондовом рынке
- Форумы
- и так далее ...

Классификации БД

- По технологии хранения:
 - БД во вторичной памяти (традиционные)
 - БД в оперативной памяти (in-memory databases)
 - БД в третичной памяти (tertiary databases)
- По содержанию:
 - Географические
 - Исторические
 - Научные
 - Мультимедийные
 - ...
- По степени распределенности:
 - Централизованные (сосредоточенные)
 - Распределенные

СУБД

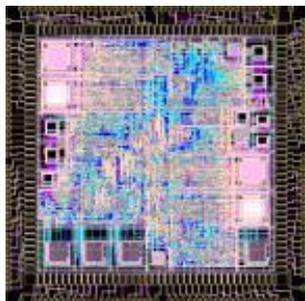
Система управления базами данных

- ▣ **СУБД** — программное обеспечение, с помощью которого пользователи могут определять, создавать и поддерживать актуальной базу данных, а также осуществлять к ней контролируемый доступ
 - или
- ▣ **СУБД** — специализированная программа (чаще комплекс программ), предназначенная для организации и ведения базы данных.

СУБД

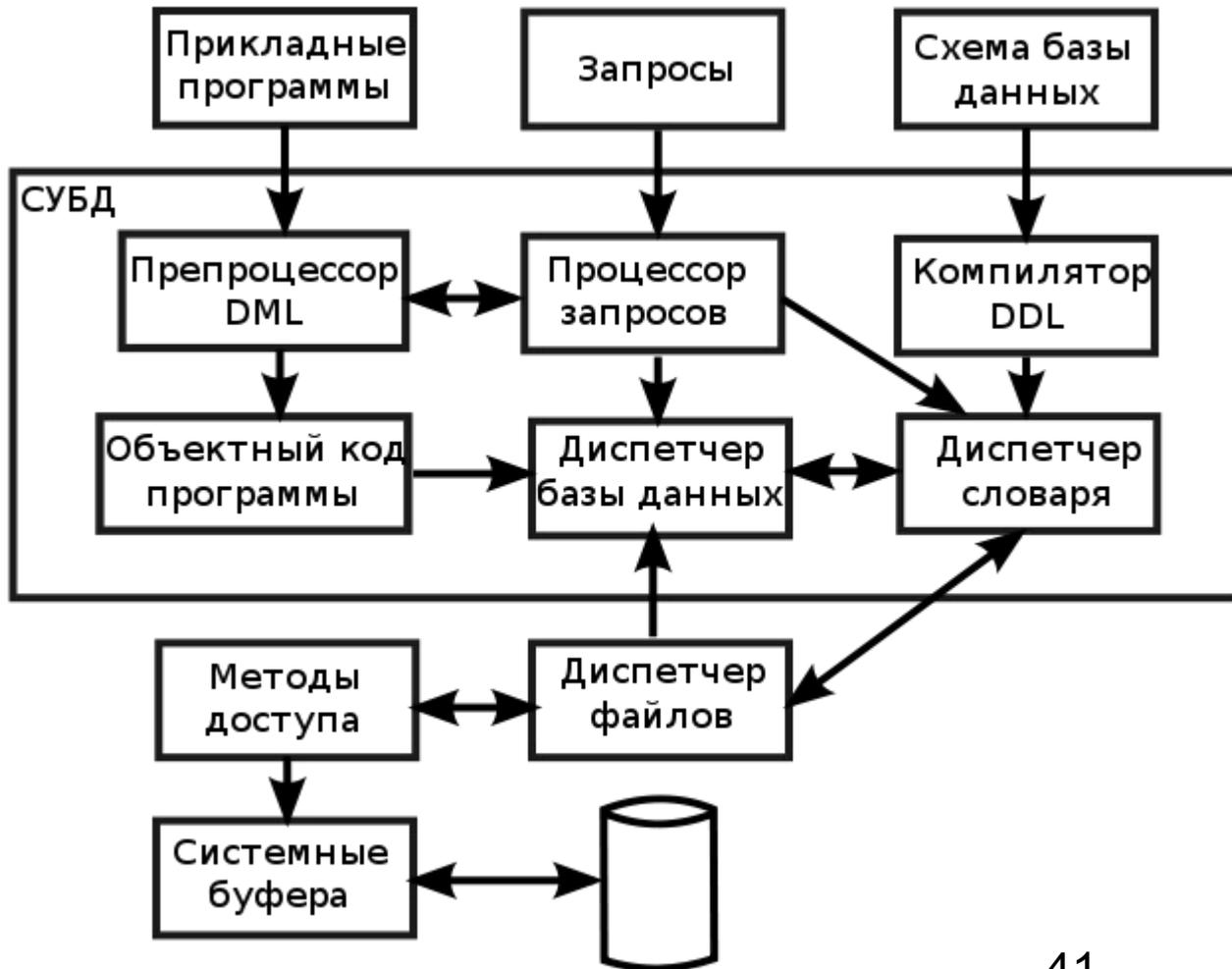
Приложения

СУБД



Абстракция данных
позволяет
данному
интерфейсу
быть четко
определенным,
и, соответственно,
разделять
приложения
и системы
управления
данными.

Основные программные компоненты среды СУБД :



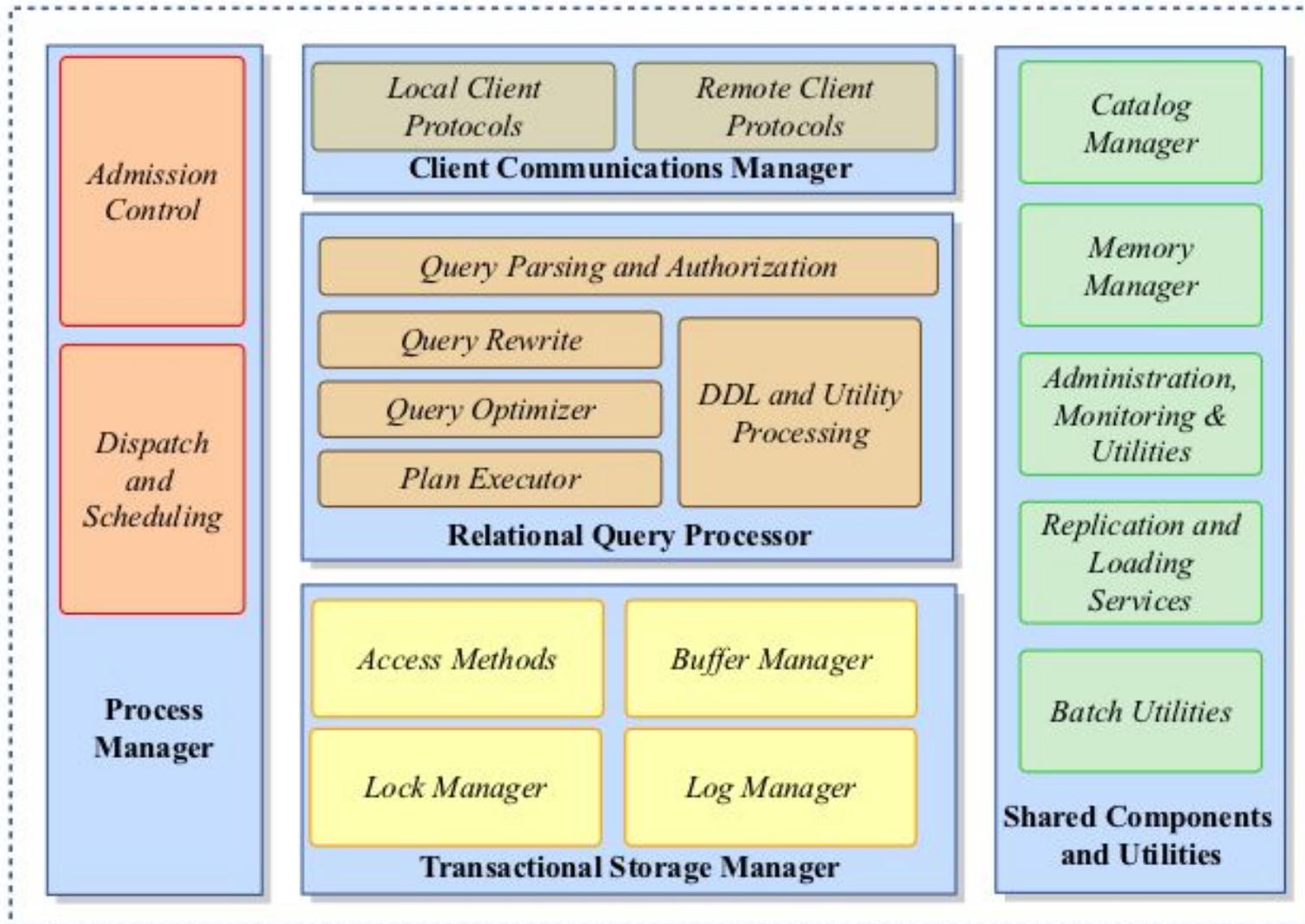
Основные программные компоненты среды СУБД :

- **процессор запросов:**
преобразует запросы в последовательность низкоуровневых инструкций для контроллера базы данных.
- **контроллер базы данных:**
взаимодействует с запущенными пользователями прикладными программами и запросами (принимает запросы; проверяет внешние и концептуальные схемы для определения концептуальных записей, удовлетворяющих требованиям запроса; затем вызывает контроллер файлов для выполнения поступившего запроса).
- **контроллер файлов:**
манипулирует предназначенными для хранения данных файлами и отвечает за распределение доступного дискового пространства; создает и поддерживает список структур и индексов, определенных во внутренней схеме (в случае использования хешированных файлов, вызывает функцию хеширования для генерации адресов и записей); не управляет физическим вводом и выводом, лишь передает команды соответствующим методам доступа, которые считывают данные в системные буферы или записывают их оттуда на диск.

Основные программные компоненты среды СУБД :

- **препроцессор языка DML:**
преобразует внедренные в прикладные программы DML-операторы в вызовы стандартных функций базового языка (для генерации соответствующего кода препроцессор языка DML должен взаимодействовать с процессором запросов).
- **компилятор языка DDL:**
преобразует DDL-команды в набор таблиц, содержащих метаданные, после чего таблицы сохраняются в системном каталоге, а управляющая информация - в заголовках файлов с данными.
- **контроллер словаря:**
управляет доступом к системному каталогу и обеспечивает работу с ним (системный каталог доступен большинству компонентов СУБД).

Основные программные компоненты среды СУБД :



Изменение архитектуры

- Ранняя архитектура:
 - Приложение →
 - Файлы

- С использованием СУБД:
 - Приложение →
 - Система управления БД →
 - База данных

Почему вам это нужно?

- Базы данных — неизбежность
- Базы данных — везде
- Базы данных — огромный рынок (~20 млрд USD в год)
- Базы данных — основа для бизнес-аналитики и анализа данных
- Базы данных — это интересно!

БД – широкая область для компьютерных наук

- Языки программирования и разработка программного обеспечения
- Структуры данных и алгоритмы
- Логика, дискретная математика, теория вычислений
- Задачи систем:
 - параллелизм, операционные системы, организация файлов, сети, распределенные системы ...

Три аспекта изучения СУБД

1. Построение модели и проектирование базы данных.
 - ▣ Позволяет исследовать различные проблемы до реализации.
2. Программирование – описание различных запросов к базе данных и операций над данными.
 - ▣ SQL – Structured Query Language
язык структурированных запросов
3. Непосредственная реализация СУБД.

Уровни пользователей БД

1. Программист системы баз данных (DBMS programmer)
 - Непосредственно создает программное обеспечение системы базы данных
2. Конструктор баз данных (Database designer)
 - Проектирует логическую схему базы данных
3. Разработчик приложений (Database application developer)
 - Создание программного обеспечения, которое позволяет пользователям взаимодействовать с системой баз данных
4. Администратор баз данных (Database administrator)
 - Управляет системой баз данных
5. Конечные пользователи (End users)
 - Использование системы баз данных для достижения какой-либо цели

Словарь данных - Метаданные

- Словарь или каталог хранит информацию о самой базе данных
- Это данные о данных или 'метаданные'
- Почти каждый аспект СУБД использует словарь
- Словарь содержит
 - Описание объектов базы данных (таблицы, пользователи, правила, виды, индексы,...)
 - Информацию о доступе к данным (кто в какой момент что использует)
 - Схемы и отображения

Системы Баз Данных

- Система Баз Данных состоит из
 - **Данных (Базы Данных)**
 - Процедур по работе с данными
 - **Программного обеспечения**
 - Аппаратных средств
 - Пользователей

- (мы будем подробно рассматривать в курсе выделенное жирным)

Данные

- Наиболее важный компонент СУБД для конечных пользователей. В системном каталоге содержатся:
 - имена, типы и размеры элементов данных;
 - имена связей;
 - ограничения целостности данных;
 - имена зарегистрированных пользователей, которым предоставлены некоторые права доступа к данным;
 - используемые индексы и структуры хранения.

Процедуры по работе с данными

- К процедурам относятся инструкции и правила, которые должны учитываться при проектировании и использовании базы данных.
- Описание процедур — инструкции о правилах выполнения:
 - регистрация в СУБД;
 - использование отдельного инструмента СУБД или приложения;
 - запуск и остановка СУБД;
 - создание резервных копий СУБД;
 - обработка сбоев аппаратного и программного обеспечения, включая процедуры идентификации вышедшего из строя компонента, исправления отказавшего компонента, а также восстановления базы данных после устранения неисправности;
 - изменение структуры таблицы, реорганизация базы данных, размещенной на нескольких дисках, способы улучшения производительности и методы архивирования данных на вторичных устройствах хранения.

Пользователи

- Обслуживание информационных потребностей пользователей — цель проектирования, создания и поддерживания базы данных.
- По способу использования пользователями системы:
 - **Рядовые пользователи**, обычно не знающие о наличии СУБД (обращающиеся к базе данных с помощью специальных приложений, позволяющих в максимальной степени упростить выполняемые ими операции).
 - **Опытные пользователи**, знакомые со структурой базы данных и возможностями СУБД (для выполнения требуемых операций могут использовать язык запросов высокого уровня, например SQL, и даже способны создавать собственные прикладные программы).

Языки по работе с данными, существующие в СУБД

- **Data definition language (DDL) –
Язык определения данных**
 - описательный язык, который позволяет администратору баз данных или пользователю описать и именовать сущности и атрибуты, необходимые для работы некоторого приложения, а также связи, имеющиеся между различными сущностями, кроме того, указать ограничения целостности и защиты.
 - определяет структуру базы данных.
 - Результат работы DDL-операторов — набор таблиц, хранимых в системном каталоге.

Языки по работе с данными, существующие в СУБД

- Data manipulation language (DML) –
Язык манипулирования данными
 - содержит набор операторов для поддержки основных операций манипулирования содержащимися в базе данными
 - позволяет выполнять операции с данными (добавление, удаление, извлечение, модификация)
 - программист, указывает на то, какие данные ему необходимы и как их можно получить - определяет все операции доступа к данным, которые должны быть выполнены для получения требуемой информации.

Языки по работе с данными, существующие в СУБД

- **Data Control Language (DCL) –
Язык контроля над данными**
 - язык баз данных для осуществления административных функций, присваивающих или отменяющих право (привилегию) использовать базу данных, таблицы в базе данных, а также выполнять те или иные операторы.
 - GRANT — применяется для присвоения привилегии;
 - REVOKE — применяется для отмены привилегии.

Функции компонентов баз данных



Преимущества СУБД

1 Контроль за избыточностью данных

- Традиционные файловые системы хранят одинаковые данные в нескольких файлах. Базы данных максимально ограничивают степень избыточности данных за счет интеграции файлов.
- Избыточность допускается при дублировании ключевых элементов для моделирования связей
- СУБД допускают дублирование данных для повышения производительности системы.

2 Непротиворечивость данных

- Устранение избыточности данных или контроль над ней позволяет уменьшить риск возникновения противоречивых состояний. Изменение одного значения элемента данных требует выполнить только одну операцию обновления, причем новое значение станет доступным сразу всем пользователям базы данных.
- Если элемент данных, с ведома системы, хранится в базе данных в нескольких экземплярах, то система сможет следить за тем, чтобы копии не противоречили друг другу.

Преимущества СУБД

3 Больше полезной информации при том же объеме хранимых данных

- Благодаря интеграции рабочих данных организации, на основе тех же данных можно получать дополнительную информацию.

4 Совместное использование данных

- Большое количество пользователей может работать с большим объемом данных.
- Возможность создания новых приложений на основе уже существующей в БД информации, позволяющей добавлять в БД только новые данные; а также возможность исключать из функционала программы функции, реализованные в СУБД

5 Поддержка целостности данных

- Целостность базы данных означает корректность и непротиворечивость хранимых в ней данных, и, обычно, описывается с помощью ограничений, которые не должны нарушаться в базе данных.
- Ограничения можно применять к элементам данных внутри одной записи или к связям между записями.
- Интеграция данных позволяет АБД задавать требования по поддержке целостности данных, а СУБД применять их.

Преимущества СУБД

6 Повышенная безопасность

- Без привлечения соответствующих мер безопасности интегрированные данные становятся более уязвимыми для несанкционированного доступа со стороны пользователей, чем данные в файловой системе.
- Интеграция позволяет АБД определить требуемую систему безопасности базы данных, а СУБД привести ее в действие.
- Система обеспечения безопасности может быть выражена в форме имен и паролей для идентификации пользователей, которые зарегистрированы в этой базе данных.
- Доступ к данным со стороны зарегистрированного пользователя может быть ограничен только некоторыми операциями.

7 Применение стандартов

- Интеграция позволяет АБД определять и применять необходимые стандарты для регламентирования формата данных при обмене ими между системами, соглашениях об именах, формах представления документации, процедурах обновления и правилах доступа.

Преимущества СУБД

8 Повышение эффективности с увеличением масштабов системы

9 Возможность нахождения компромисса для противоречивых требований

- Потребности одних пользователей/отделов могут противоречить потребностям других пользователей.
- АБД, может принимать решения о проектировании и способе использования базы данных, при которых имеющиеся ресурсы всей организации в целом будут использоваться наилучшим образом.
- Эти решения обеспечивают оптимальную производительность для самых важных приложений, причем чаще всего за счет менее критичных.

10 Повышение доступности данных и их готовности к работе

- Данные, которые пересекают границы отделов, в результате интеграции становятся непосредственно доступными конечным пользователям.
- Во многих СУБД предусмотрены языки запросов или инструменты для создания отчетов, которые позволяют пользователям вводить не предусмотренные заранее запросы и почти немедленно получать требуемую информацию на своих терминалах, не прибегая к помощи программиста, который для извлечения этой информации из базы данных должен был бы создать специальное программное обеспечение.

Преимущества СУБД

11 Улучшение показателей производительности

- На базовом уровне СУБД обеспечивает все низкоуровневые процедуры работы с файлами, которую обычно выполняют приложения. Наличие этих процедур позволяет программисту сконцентрироваться на разработке более специальных, необходимых пользователям функций, не заботясь о подробностях их воплощения на более низком уровне.
- Во многих СУБД предусмотрена также среда разработки четвертого поколения с инструментами, упрощающими создание приложений баз данных. Результатом является повышение производительности работы программистов и сокращение времени разработки новых приложений (с соответствующей экономией средств).

12 Упрощение сопровождения системы за счет независимости от данных

- В файловых системах описания данных и логика доступа к данным встроены в каждое приложение, поэтому программы становятся зависимыми от данных.
- В СУБД описания данных отделены от приложений, а потому приложения защищены от изменений в описаниях данных. Эта особенность называется независимостью от данных.
- Наличие независимости программ от данных значительно упрощает обслуживание и сопровождение приложений, работающих с базой данных.

Преимущества СУБД

13 Улучшенное управление параллельной работой

- В некоторых файловых системах при одновременном доступе к одному и тому же файлу двух пользователей может возникнуть конфликт двух запросов, результатом которого будет потеря информации или утрата ее целостности.
- Во многих СУБД предусмотрена возможность параллельного доступа к базе данных и гарантируется отсутствие подобных проблем.

14 Развитые службы резервного копирования и восстановления

- Ответственность за обеспечение защиты данных от сбоев аппаратного и программного обеспечения в файловых системах возлагается на пользователя.
- В современных СУБД предусмотрены средства снижения вероятности потерь информации при возникновении различных сбоев.

Недостатки СУБД

1 Сложность

- Обеспечение необходимой функциональности СУБД влечет за собой усложнение проектирования и усложнение программного обеспечения СУБД, а также необходимость понимания принципов работы СУБД всеми сотрудниками, работающими с ней.

2 Размер

- Сложность и широта функциональных возможностей СУБД, превращает ее в чрезвычайно сложный программный продукт, требующий много места на диске и нуждающийся в большом объеме оперативной памяти для эффективной работы.

3 Стоимость СУБД

- В зависимости от имеющейся вычислительной среды и требуемых функциональных возможностей стоимость СУБД может изменяться в очень широких пределах. Кроме того, сопровождение системы составляет некоторый процент от ее общей стоимости.

Недостатки СУБД

4 Дополнительные затраты на аппаратное обеспечение

- Для удовлетворения требований, предъявляемых СУБД и базой данных, может понадобиться приобрести дополнительные устройства хранения информации или более мощный компьютер.

5 Затраты на преобразование

- Стоимость подготовки персонала для работы с новой системой, оплату услуг специалистов, оказывающих помощь в преобразовании и запуске новой системы.

Недостатки СУБД

6 Производительность

- Обычно файловая система создается для некоторых специализированных приложений, потому ее производительность может быть весьма высока. А СУБД предназначены для решения более общих задач и обслуживания сразу нескольких приложений, что сказывается на производительности.

7 Более серьезные последствия при выходе системы из строя

- Централизация ресурсов повышает уязвимость системы. Выход из строя одного из ее компонентов может привести к полному прекращению всей работы организации.

Функции СУБД

1 Хранение, извлечение и обновление данных

- СУБД должна предоставлять пользователям возможность сохранять, извлекать и обновлять данные в базе данных (самая фундаментальная функция СУБД).
- Способ реализации этой функции в СУБД должен позволять скрывать от конечного пользователя внутренние детали физической реализации системы (например, файловую организацию или используемые структуры хранения).

2 Каталог, доступный конечным пользователям

- СУБД должна иметь доступный конечным пользователям каталог, в котором хранится описание элементов данных (системный каталог).

3 Поддержка транзакций

- СУБД должна иметь механизм, который гарантирует выполнение либо всех операций обновления данной транзакции, либо ни одной из них

Функции СУБД

4 Службы управления параллельной работой

- СУБД должна иметь механизм, который гарантирует корректное обновление баз данных при параллельном выполнении операций обновления многими пользователями.

5 Службы восстановления

- СУБД должна предоставлять средства восстановления базы данных на случай какого-либо ее повреждения или разрушения.

6 Службы контроля доступа к данным

- СУБД должна иметь механизм, гарантирующий возможность доступа к базе данных только санкционированных пользователей.
- Термин безопасность относится к защите базы данных от преднамеренного или случайного несанкционированного доступа. Предполагается, что СУБД обеспечивает механизмы подобной защиты данных.

Функции СУБД

7 Поддержка обмена данными

- СУБД должна обладать способностью к интеграции с коммуникационным программным обеспечением.
- Любая СУБД должна обладать способностью интеграции с разнообразными существующими диспетчерами обмена данными (DEM — Data Exchange Manager).

8 Службы поддержки целостности данных

- СУБД должна обладать инструментами контроля за тем, чтобы данные и их изменения соответствовали заданным правилам (еще один тип защиты базы данных). Целостность базы данных означает корректность и непротиворечивость хранимых данных.

Функции СУБД

9 Службы поддержки независимости от данных

- СУБД должна обладать инструментами поддержки независимости программ от фактической структуры базы данных.

10 Вспомогательные службы

- СУБД должна предоставлять некоторый набор различных вспомогательных служб, предназначенных для оказания помощи АБД в эффективном администрировании базы данных.
- Утилиты, работающие на внешнем уровне, могут быть созданы самим АБД, но утилиты, работающие на внутреннем уровне системы должны быть предоставлены разработчиком СУБД.

Некоторые современные СУБД

1. MySQL:

- www.mysql.org
- Open source, quite powerful

2. PostgreSQL:

- www.postgresql.org
- Open source, powerful

3. Microsoft Access:

- Simple system, lots of nice GUI wrappers

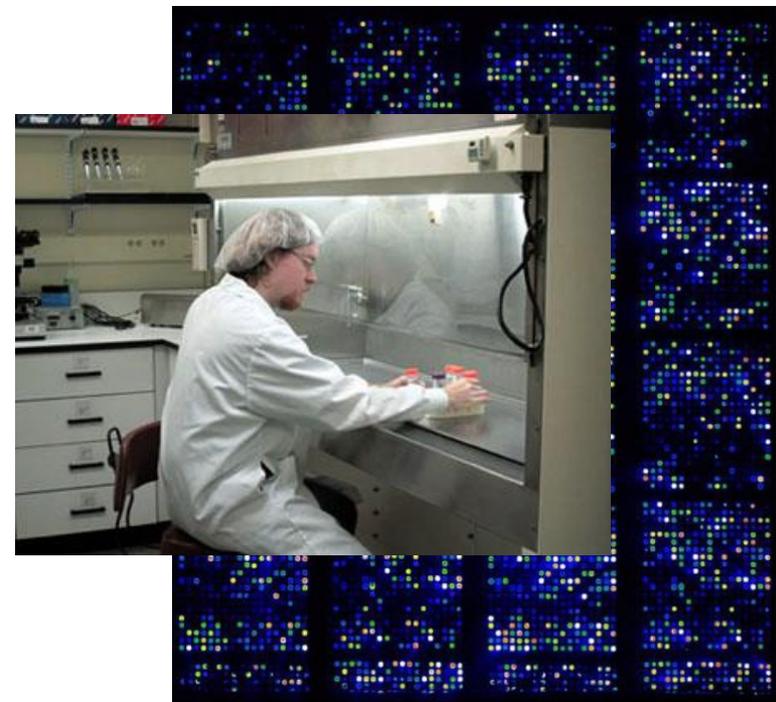
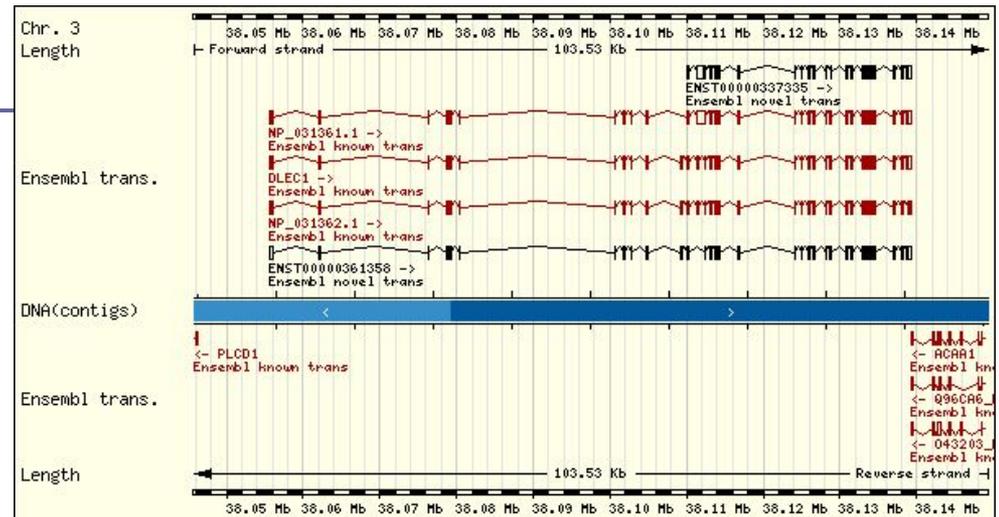
4. Commercial systems:

- Oracle (www.oracle.com)
- SQL Server (www.microsoft.com/sql)
- DB2 (www.ibm.com/db2)

Области применения БД:

Биология

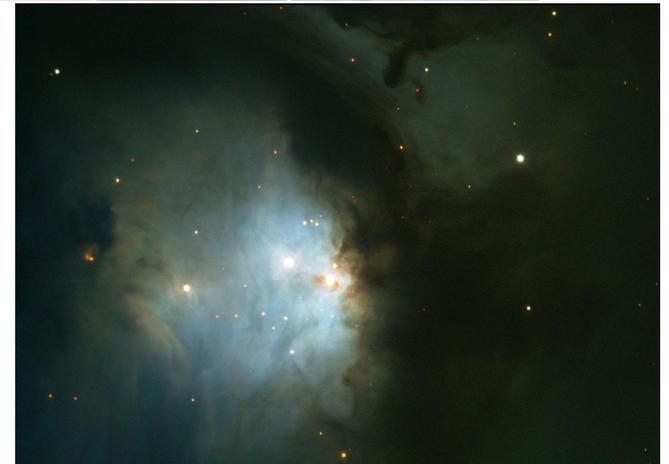
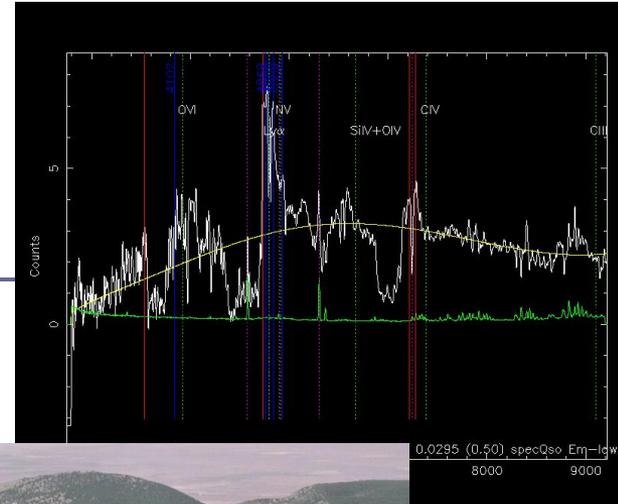
- Старый способ:
 - «Мокрая» химия
- Новый способ:
 - Микроматрицы ДНК
 - Поиск в банках данных GenBank, Ensembl, GDB, SwissProt, Entrez using BLAST, FASTA, GCG, EMBOSS



Области применения БД: Астрономия

- Старый способ:
 - Запись на время работы телескопа

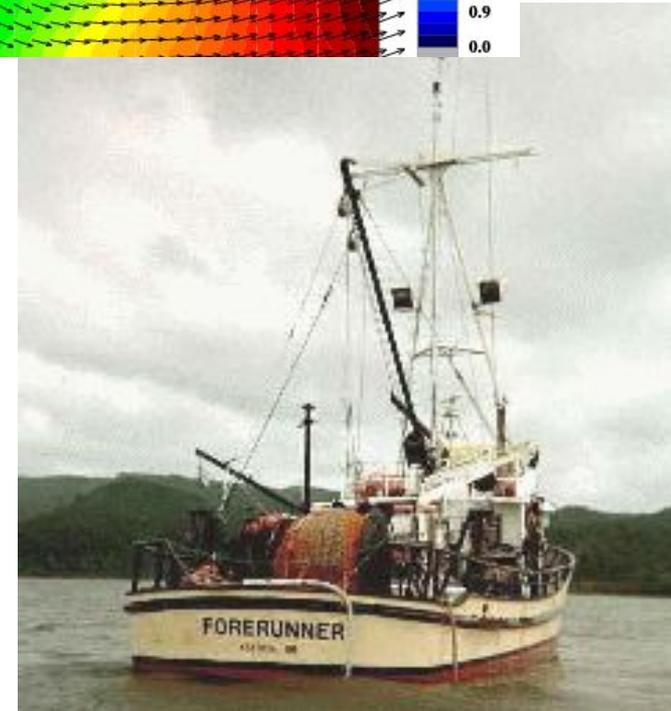
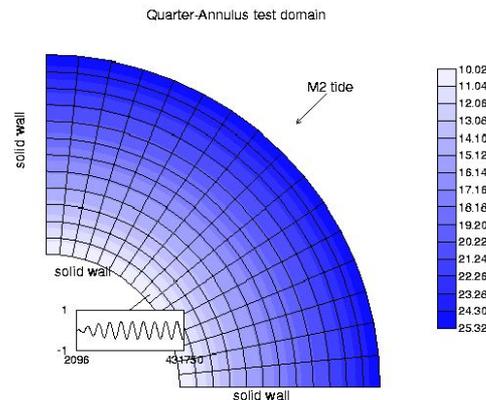
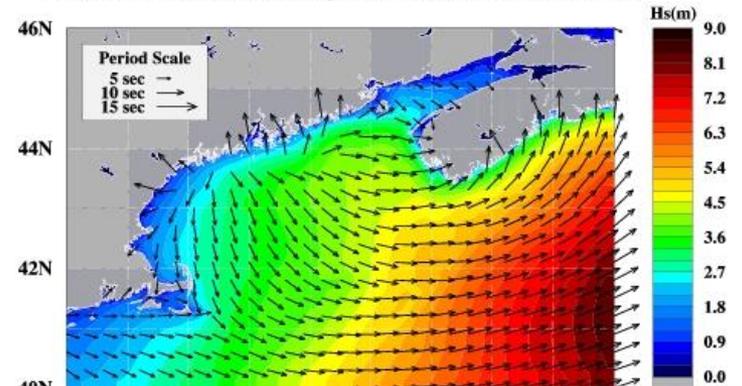
- Новый способ:
 - Sloan Digital Sky Survey
 - Systematically mapping $\frac{1}{4}$ of the entire sky
 - 12 TB to date,
 - 15 TB final in 2007
 - Совместное распознавание звезд



Области применения БД: Океанография

- Старый способ:
 - Field work
 - Simplified Calculations
- Новый способ:
 - Finite Element Analysis
 - In situ sensors
 - CODAR

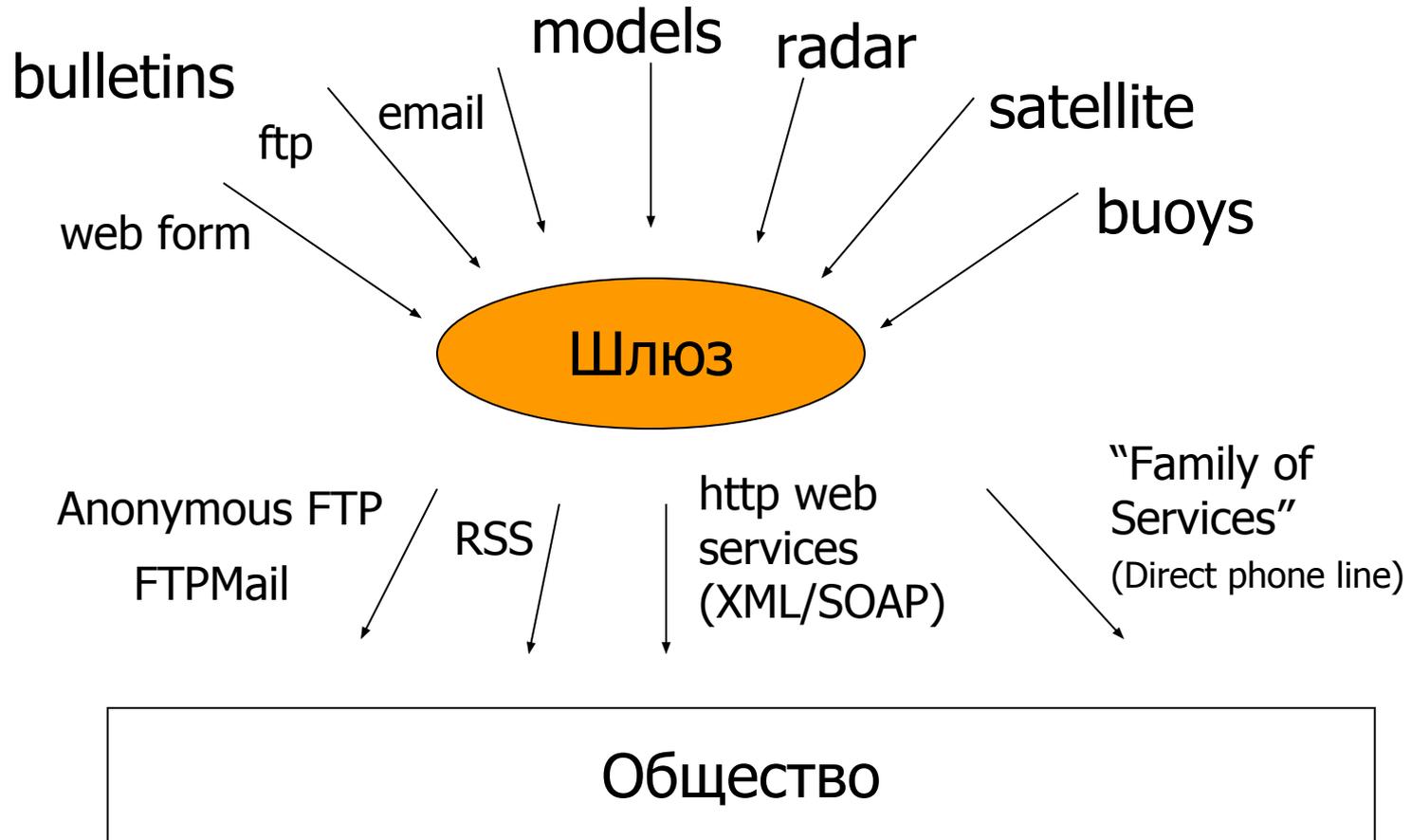
Fig. Wave Heights (contours), Wave Directions (dir. of vectors), Wave Periods (length of vectors) At 2006021312



Национальная служба погоды США

- Сбор данных
 - Радары погоды
 - Спутники
 - Прогнозы
 - Сводки
- Распространение данных
 - По радио: aviation, marine, military channels
 - FTP, HTTP, email, RSS: public
- Часть глобальной сети ООН

НСП: Шлюз данных



Национальная служба погоды

□ **Компьютерные модели**

- Содержит более 10 расчетных моделей погоды от регионального до мирового масштаба
- Хранит факсы/изображения
- Хранят текстовую информацию, полученную из расчетных моделей
- Специальные данные в оригинальных форматах

□ **Дополнительная текстовая информация**

- Предупреждения, прогнозы, советы, обсуждения
- Более ~100 различных типов текстов

Национальная служба погоды

□ **Данные наблюдений** -

- Хранятся как минимум 24 часа
- Собираются с самолетов, буев, кораблей, метеозондов
- Специальные форматы, преобразовываемые в XML

□ **Данные с радаров** -

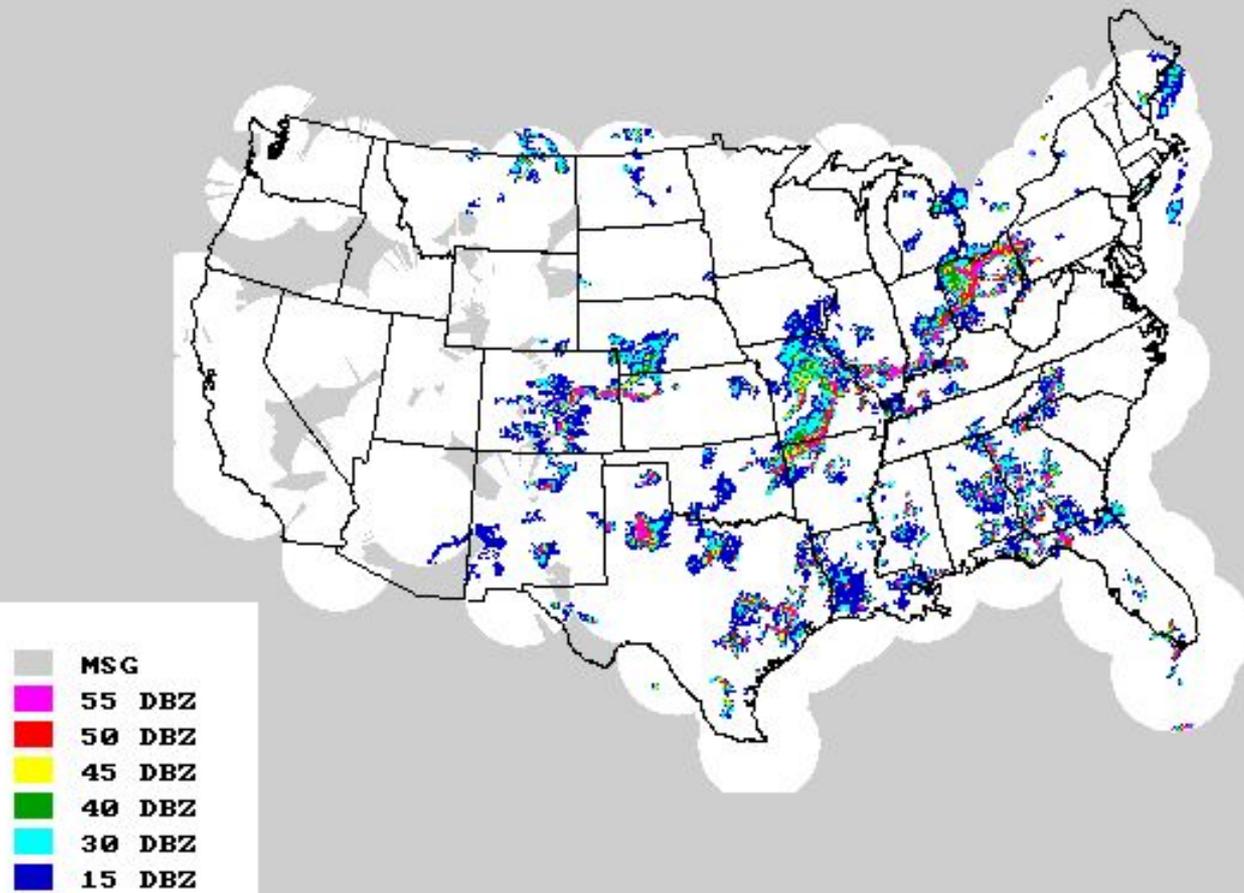
- Multicast by connecting a router directly to NWS as well as FTP
- SL.us008001/DF.of/DC.radar/DS.p19r1/SI.kfws/sn.0114

□ **Данные со спутников** -

- Cloud Water Vapor, Cloud Liquid Water, Rain Rate, Sea Ice Concentration, Sea Ice Age, Sea Ice Edge, Soil Moisture, Surface Wind, Water Vapor over oceans, Surface Temperature, Snow Water Content, Cloud Amount, and EDR Surface Type

Национальная служба погоды

RADAR REFLECTIVITY FROM RADAR CODED MESSAGES
METEOROLOGICAL DEVELOPMENT LABORATORY
NATIONAL WEATHER SERVICE
AP/CLUTTER EDITING APPLIED
JUN 22, 2006 20:19 UTC



Рекомендуемые источники информации по лекции

- http://ru.wikipedia.org/wiki/База_данных
- http://ru.wikipedia.org/wiki/Система_упр_авления_базами_данных
- <http://ru.wikipedia.org/wiki/Картотека>
- http://citforum.ru/database/osbd/contents_shtml

Рекомендуемые источники информации по курсу

- <https://www.coursera.org/course/db>
- <http://www.intuit.ru/department/database/workwdb/>
- <http://www.mstu.edu.ru/study/materials/zelenkov/toc.html>
- http://www.intuit.ru/studies/courses?page=1&service=0&option_id=3

Рекомендуемые источники информации по курсу

- http://wiki.auditory.ru/Категория:Базы_данных
- Кузин А. В., Левонисова С. В.
Базы данных. – Академия, 2008
- Ульман Л.
MySQL Руководство по изучению языка. – 2004
- Ульман Дж, Уидом Дж.
Введение в системы баз данных
- К. Дж. Дейт
Введение в системы баз данных
- С.Д. Кузнецов
Основы современных баз данных
- ...