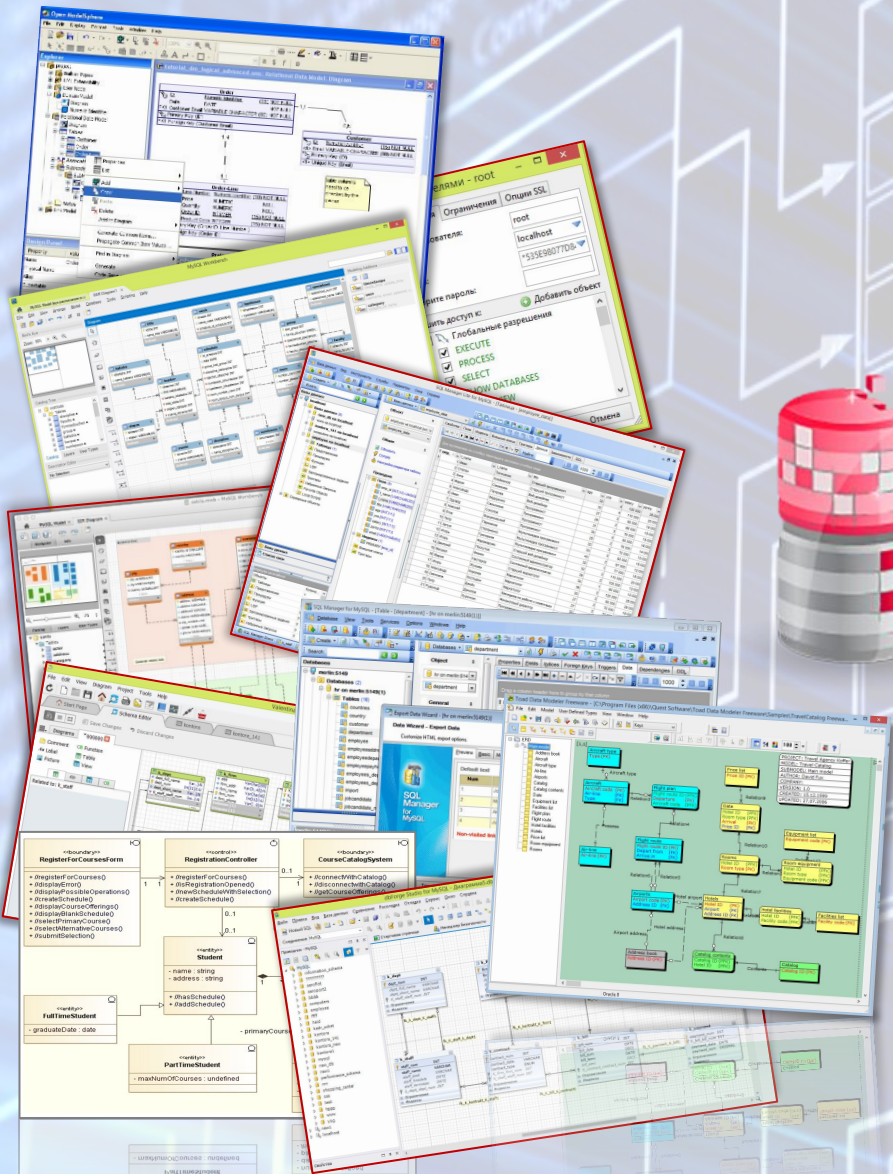


БАЗЫ ДАННЫХ



ORACLE

Microsoft®
SQL Server®

MYSQL®

PostgreSQL

Access

Гаверлов Александр Викторович
к.т.н., доцент



Тема 1. Введение в курс «Базы данных»

Лекция 1

Вопросы лекции:

1. История возникновения баз данных
2. Основные термины и определения
3. Классификация СУБД
4. Перспективы развития БД

Рекомендуемая литература

1. Веллинг Л., Томсон Л. MySQL. Учебное пособие: Пер. с англ. – М.: Издательский дом «Вильямс», 2005. – 304 с.: ил.
2. Гарсиа-Молима Г., Ульман Д. Д., Уидом Д. Системы баз данных. Полный курс.: Пер. с англ. – М. : Издательский дом «Вильямс», 2003. – 1088 с.: ил.
3. Гольцман Виктор. MySQL 5.0. Библиотека программиста. – СПб.: Питер, 2010.
4. Грофф Дж., Вайнберг П. SQL: Полное руководство: Пер. с англ. – 2-е изд., перераб. и доп. – К.: Издательская группа BHV, 2001. – 816 с., ил.
5. Гурвиц Г.А. Microsoft® Access 2010. Разработка приложений на реальном примере – СПб : БХВ-

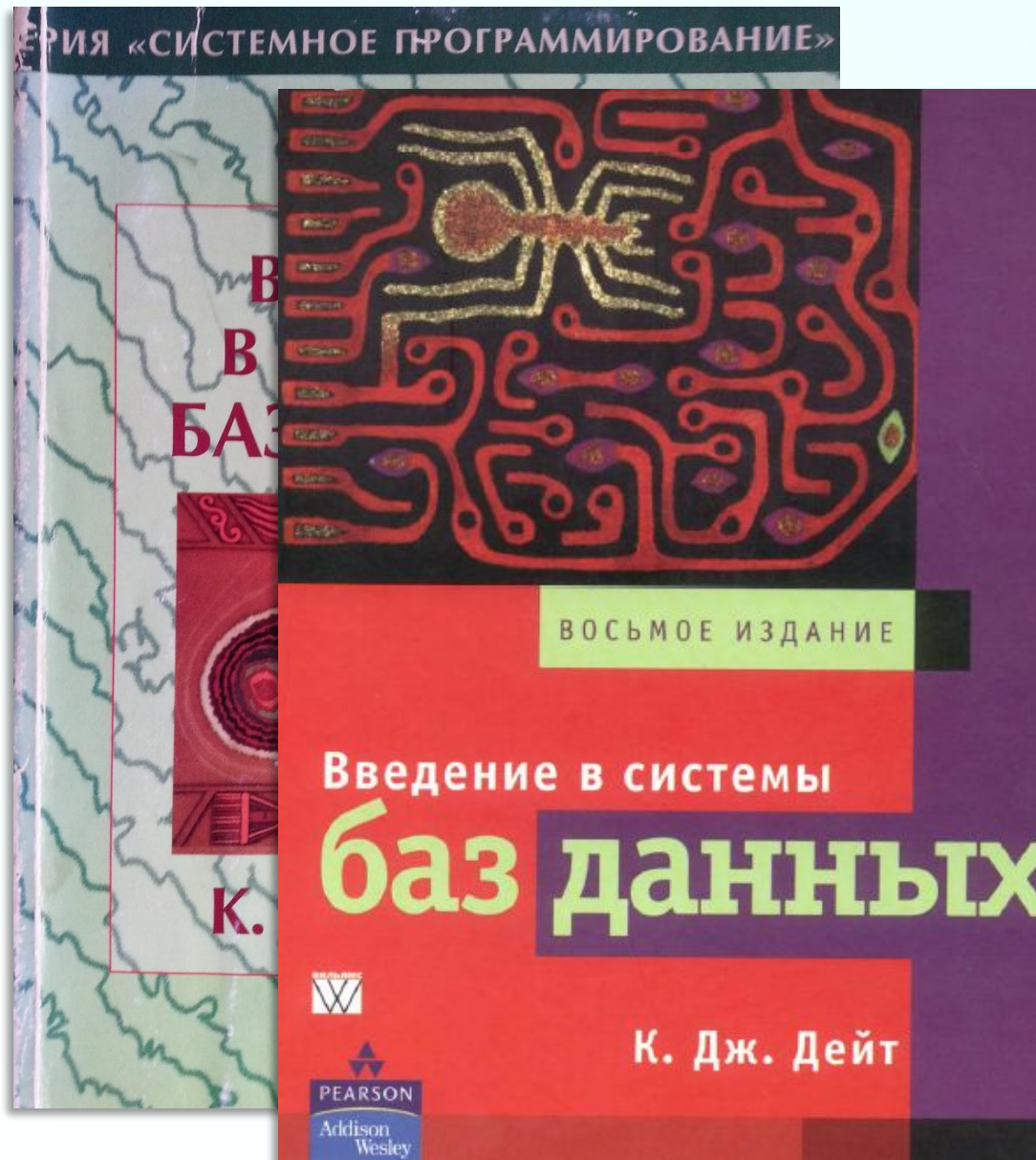
Рекомендуемая литература

6. Дейт, К. Дж. Введение в системы баз данных. Вильямс, 2002. - 1071 с.
7. Диго С.М. Базы данных: проектирование и использование. Учебник. М.: ФИС, 2005.
8. Карпова Т.С. Базы данных: модели, разработка, реализация. - СПб.: Питер, 2002. - 304 с.
9. Клайн К. SQL. Справочник. 2-е издание / Пер. с англ. - М.: КУДИЦ-ОБРАЗ, 2006 - 832 с.
10. Коннолли Т., Бегг К., Страчан А.. Базы данных. Проектирование, реализация и сопровождение. Теория и практика. - 3-е издание - М.: Издательский дом «Вильямс», 2003.
11. Кренке Д. Теория и практика построения баз данных. - 8-е изд. СПб.: Питер, 2003, 800 с.:

Рекомендуемая литература

12. Кузнецов М.В. MySQL 5 / М. В. Кузнецов, И. В. Симдянов. – СПб.: БХВ-Петербург, 2010. – 1024 с: ил.
13. Кузнецов С.Д. Базы данных: Модели и языки. Учебник. М.: ООО «Бином-Пресс», 2010 г.
14. Кузнецов С.Д. Основы баз данных: учебное пособие / С.Д. Кузнецов - 2-е изд., испр. – М.: Интернет-Университет Информационных Технологий; БИНОМ. Лаборатория знаний, 2007. – 484 с.: ил.
15. Райордан Р. Основы реляционных баз данных / Пер. с англ. – М.: Издательско-торговый дом «Русская Редакция», 2001. – 384 с.: ил.
16. Хомоненко А. Д., Цыганков В. М., Мальцев М. Г. Базы данных: Учебник для высших учебных заведений / Под ред. А. Д. Хомоненко - 6-е изд.

Золотой фонд компьютерной литературы



Дейт, К. Дж.

Введение в системы баз данных

8-е издание.: Пер. с англ. — М.: Издательский дом "Вильямс", 2005. — 1328 с.

http://www.ph4s.ru/book_pc_bazy.html

Рекомендуемые Интернет-ресурсы:

1. Базы данных. Вводный курс. Кузнецов С. Д. URL: http://citforum.ru/database/advanced_intro.
2. Интерактивный учебник по SQL. URL: <http://www.sql-tutorial.ru>.
3. Национальный Открытый Университет «ИНТУИТ». Васильев Ю. Работа в Microsoft Access. URL: <http://www.intuit.ru/studies/courses/1127/126/info>.
4. Национальный Открытый Университет «ИНТУИТ». Карпова Т. Базы данных: модели, разработка, реализация. URL: <http://www.intuit.ru/studies/courses/1001/297/info>.
5. Национальный Открытый Университет «ИНТУИТ». Каталог курсов: Базы данных. URL: http://www.intuit.ru/studies/courses?service=0&option_id=3&service_path=1.
6. Национальный Открытый Университет «ИНТУИТ». Кузнецов С. Введение в реляционные базы данных. URL: <http://www.intuit.ru/studies/courses/74/74/info>.

Рекомендуемые Интернет-ресурсы:

7. Национальный Открытый Университет «ИНТУИТ». Назаров А. Введение в СУБД MySQL. URL:
<http://www.intuit.ru/studies/courses/111/111/info>.
8. Национальный Открытый Университет «ИНТУИТ». Туманов В. Основы проектирования реляционных баз данных. URL:
<http://www.intuit.ru/studies/courses/1095/191/info>.
9. Основы современных баз данных. Кузнецов С. Д. URL:
<http://citforum.ru/database/osbd/contents.shtml>.
10. Официальный сайт базы данных MySQL. URL:
<http://www.mysql.com>.
11. Практическое владение языком SQL. Упражнения по SQL. URL: <http://sql-ex.ru>.
12. Профессиональный сайт по SQL. URL: <http://www.sql.ru>.
13. Самоучитель по языку SQL (SQL DML). URL:
<http://sql-ex.ru/help>.
14. Сервер Информационных Технологий (CIT Forum). Базы

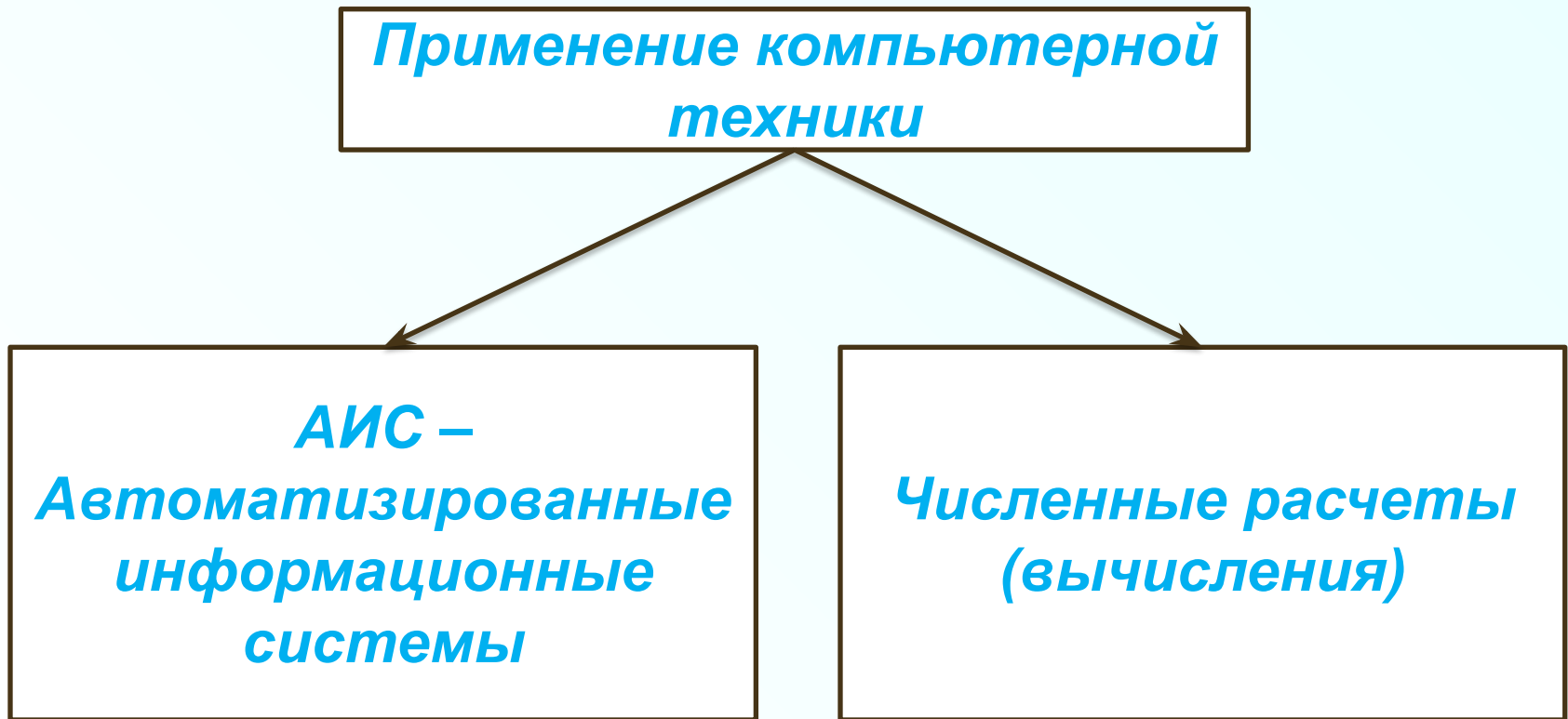
Рекомендуемые Интернет-ресурсы:

15. Справочное руководство по MySQL. URL:
<http://www.mysql.ru/docs/man>.
16. Справочник по MySQL. URL:
<http://www.spravkaweb.ru/mysql>.
17. Справочник по языкам SQL Server. URL:
[https://msdn.microsoft.com/ru-ru/library/dn198336\(v=sql.120\).aspx](https://msdn.microsoft.com/ru-ru/library/dn198336(v=sql.120).aspx).
18. Структуризированный язык запросов (SQL). URL:
<http://www.helloworld.ru/texts/comp/db/mysql/osnovisql>.
19. Учебник по MySQL. URL:
<http://www.helloworld.ru/texts/comp/db/mysql/mysql2/mysql.htm>.

История возникновения баз данных



Использование средств вычислительной техники в автоматических или автоматизированных информационных системах является одним из двух основных направлений ее применения



Этапы развития БД.

Этап 0. Файловые системы

Магнитные диски впервые были реализованы в 1956 году в исследовательской лаборатории корпорации IBM, расположенной в Сан-Хосе (Калифорния), где был выпущен серийный дисковый накопитель IBM 350 — первое устройство с подвижной головкой для чтения и записи.

Важным шагом в развитии баз данных явился переход к использованию централизованных систем управления файлами, или, используя общепринятый в данный момент термин, **файловым**

системам

Этапы развития БД.

Этап 0. Файловые системы


Магнитные диски впервые были реализованы в 1956 году в исследовательской лаборатории корпорации IBM, расположенной в Сан-Хосе (Калифорния), где был выпущен серийный дисковый накопитель IBM 350 — первое устройство с подвижной головкой для чтения и записи.



Важным шагом в развитии баз данных явился переход к использованию централизованных систем управления файлами, или, используя общепринятый в данный момент термин, **файловым системам**

Этапы развития БД.

Этап 0. Файловые системы



Файловая система — порядок, определяющий формат содержимого и способ физического хранения информации, которую принято группировать в виде файлов. Конкретная файловая система определяет размер имен файлов и (каталогов), максимальный возможный размер файла и раздела, набор атрибутов файла.



Первая развитая **файловая система** была разработана фирмой **IBM** для ее серии **System/360** в **1964** году

Недостатки применение файловых систем для хранения и обработки данных в информационных системах:

- ✓ **Избыточность данных.** Из-за дублирования данных в разных файлах память на внешних запоминающих устройствах используется неэкономно
- ✓ **Несогласованность данных.** Учитывая, что одна и та же информация может размещаться в разных файлах, технологически тяжело проследить за внесением изменений одновременно во все файлы.
- ✓ **Зависимость структур данных и прикладных программ.** Этот недостаток файловых систем приводит к значительному увеличению стоимости сопровождения программных средств. Иногда стоимость сопровождения программных средств может достигать близко 70 % стоимости их разработки.

Этапы развития БД.

Этап 1. Базы данных на больших ЭВМ 1960–1980 гг.



Получаемые в результате библиотеки, реализующие дополнительные индивидуальные средства управления данными, являлись существенной частью информационных систем и практически повторялись от одной системы к другой.



Стремление выделить и обобщить общую часть информационных систем, ответственную за управление сложно структурированными данными, вылилось в реализацию новых программных систем, названных впоследствии **Системами Управления Базами Данных (СУБД)**, а сами хранилища информации, которые работали под управлением данных систем, назывались **базами данных**.

Этапы развития БД.

Этап 1. Базы данных на больших ЭВМ 1960–1980 гг.

На сегодняшний день история развития СУБД насчитывает уже более 45 лет. В 1968 году компания IBM разработала первую промышленную СУБД:

IBM IMS (Information Management System)

Главным архитектором СУБД был **Верн Уоттс**. Начав работу в IBM в 1956 году, он непрерывно работал над IMS начиная от времени её первоначального проектирования вплоть до своей кончины 4 апреля 2009 года.

В задачу IMS входила обработка спецификации изделия для ракеты Сатурн-5 и кораблей Аполлон.

В качестве носителя информации использовалась **магнитная лента**, а в качестве структуры данных – **иерархическая модель**.

IBM System/360



Этапы развития БД.

Этап 2. Настольные (desktop) СУБД 1975 - 1995 гг.



Звание первого персонального компьютера принадлежит модели 5100 производства фирмы IBM, выпущенной в 1975 году. Он был более компактным, чем мэйнфреймы, имел встроенные монитор, клавиатуру и накопитель на магнитной ленте, и предназначался для решения научно-инженерных задач.



Первым же массовым **персональным компьютером** производства фирмы IBM, выпущенным в 1981 году, стал **IBM PC** модели **5150**, положивший начало семейству наиболее распространённых современных персональных компьютеров.

Этапы развития БД.

Этап 2. Настольные (desktop) СУБД 1975 - 1995 гг.



Спрос на развитые удобные программы обработки данных заставлял поставщиков программного обеспечения поставлять все новые системы, которые принято называть **настольными (desktop) СУБД.**



Наличие на рынке большого числа СУБД, выполняющих сходные функции, потребовало разработки методов экспорта, импорта и открытых форматов хранения данных. Так появились первые коммерческие **СУБД с реляционной моделью данных.**


Этапы развития БД.

Этап 2. Настольные (desktop) СУБД 1975 - 1995 гг.

- ✓ Основанные на реляционном подходе СУБД для персональных компьютеров принято считать **системами второго поколения.**
- ✓ В 80-х годах были созданы различные коммерческие реляционные СУБД - например, DB2 или SQL/DS корпорации IBM, Oracle и др. Большинство СУБД имели развитый и удобный пользовательский интерфейс, предлагающий интерактивный режим работы с БД, как в рамках описания БД, так и в рамках проектирования запросов.


Этапы развития БД.


Этап 2. Настольные (desktop) СУБД 1975 - 1995 гг.


 Главное **ограничение** при работе с настольными СУБД накладывалось **монопольным доступом**, поскольку первое время персональные компьютеры не были подключены к вычислительным сетям. Базы данных на них создавались для работы одного пользователя.

Этапы развития БД.


Этап 3. Распределенные базы данных с 1985 по наст. вр.

 **Третий этап** развития СУБД связывают с распространением локальных и глобальных компьютерных сетей.

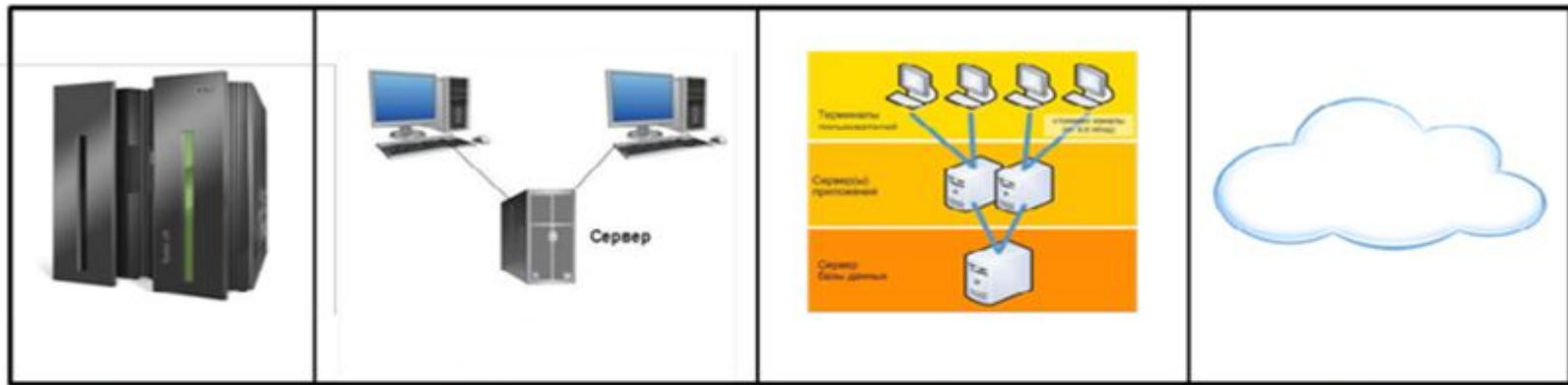
 На сегодняшний день третий этап можно считать незавершённым.

 К этому этапу можно отнести разработку ряда стандартов в рамках языков описания и манипулирования данными:

SQL-89, SQL-92, SQL-99, SQL:2003, SQL:2006, SQL:2008

 Представителями современных СУБД можно считать серверы баз данных **Oracle, MS SQL, Informix, DB2, MySQL** и другие.

Эволюция архитектуры информационных систем



1-й этап

2-й этап

3-й этап

4-й этап



1-й этап. Монолитная архитектура (mainframe), когда и база данных, и приложения работали на одном большом компьютере. Рабочие места пользователей – терминалы.

2-й этап. Архитектура «файл–сервер». Здесь уже был свой выделенный сервер баз данных, и пользователи работали на «толстых» клиентах, разгружая сервер БД.

3-й этап. Трехуровневая архитектура (клиент–сервер). где логика приложений вынесена на отдельный компьютер, называемый сервером приложений, а пользователи работали на «тонких» клиентах через web-браузеры. Большинство приложений сегодня выполнено именно в этой архитектуре. Она подразумевает развертывание всей IT-инфраструктуры на территории заказчика.

4-й этап. Облачные вычисления – следующий шаг в эволюции архитектуры построения информационных систем. Благодаря огромным преимуществам этого подхода очевидно, что многие информационные системы в ближайшее время будут перенесены в облако. Этот процесс уже начался.

РЕЗЮМЕ

История развития баз данных насчитывает более 50 лет.

Условно выделяют три этапа. При этом между ними нет жестких временных ограничений, этапы плавно переходили из одного в другой и существовали параллельно:

- Этап 1. Базы данных на больших ЭВМ 1960–1980 гг.
- Этап 2. Настольные (desktop) СУБД 1975–1995 гг.
- Этап 3. Распределенные базы данных 1985–... гг.

Основные термины и определения

Основные термины и определения

База данных (БД) представляет собой совокупность специальным образом организованных данных, хранимых в памяти вычислительной системы и отображающих состояние объектов и их взаимосвязей в рассматриваемой предметной области.

Основные термины и определения

В части четвертой **«Гражданского кодекса Российской Федерации»** дается

следующее определение базы данных:

«Базой данных является представленная в объективной форме совокупность самостоятельных материалов (статей, расчетов, 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)».

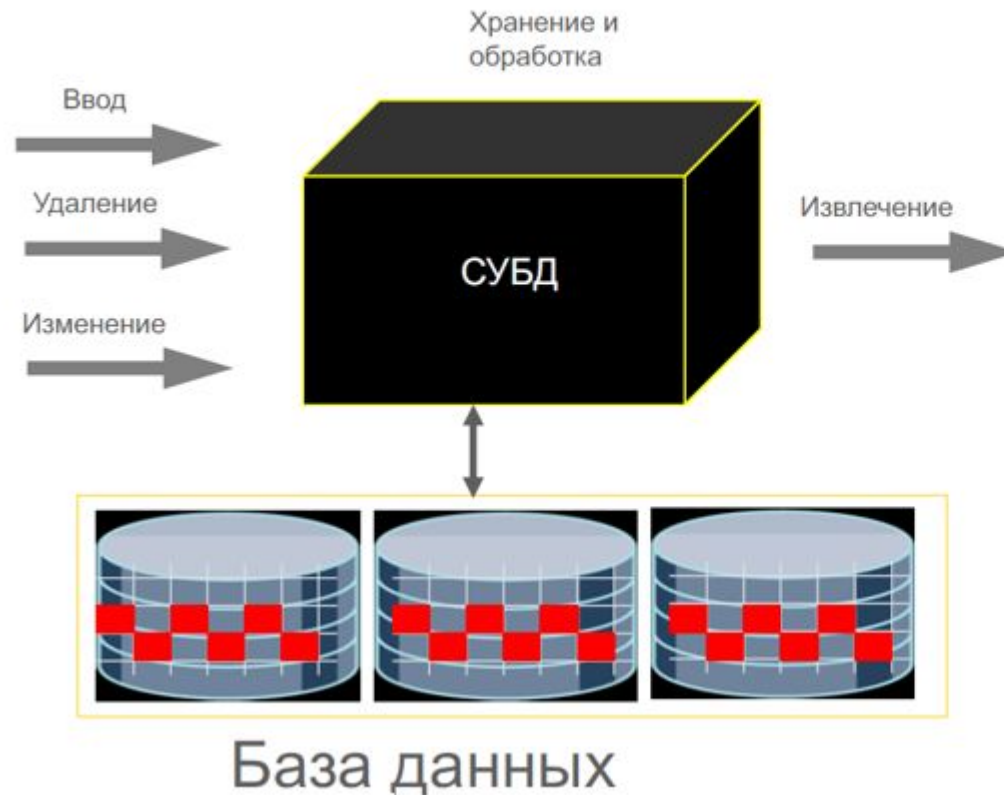
Основные термины и определения

Данные (data – факт) – это совокупность сведений, зафиксированных на определенном носителе в форме, пригодной для постоянного хранения, передачи и обработки.

Преобразование и обработка данных позволяет получить **информацию**.

Система управления базами данных (СУБД) – это комплекс языковых и программных средств, предназначенный для создания, ведения и совместного использования БД многими пользователями.

Система управления базой данных (СУБД)



- Черный ящик
- СУБД и БД
- Единый язык описания
- SQL
- Разные типы данных
 - Числа
 - Строчки
 - Дата и время
 - Видео
 - Аудио
 - Фото и картинки
 - Документы
 - Пространственная инф
 - XML

Система управления базами данных (СУБД) – это комплекс языковых и программных средств, предназначенный для создания, ведения и совместного использования БД многими пользователями.

Основные функции СУБД:

- ✓ **Создание БД**
- ✓ **Создание пользователей и указание привилегий**
- ✓ **Обеспечение работы пользователей с БД с учетом привилегий**
- ✓ **Поддержание целостности данных**
- ✓ **Поддержание механизма транзакций**
- ✓ **Журналирование**
- ✓ **Управление оперативной памятью**

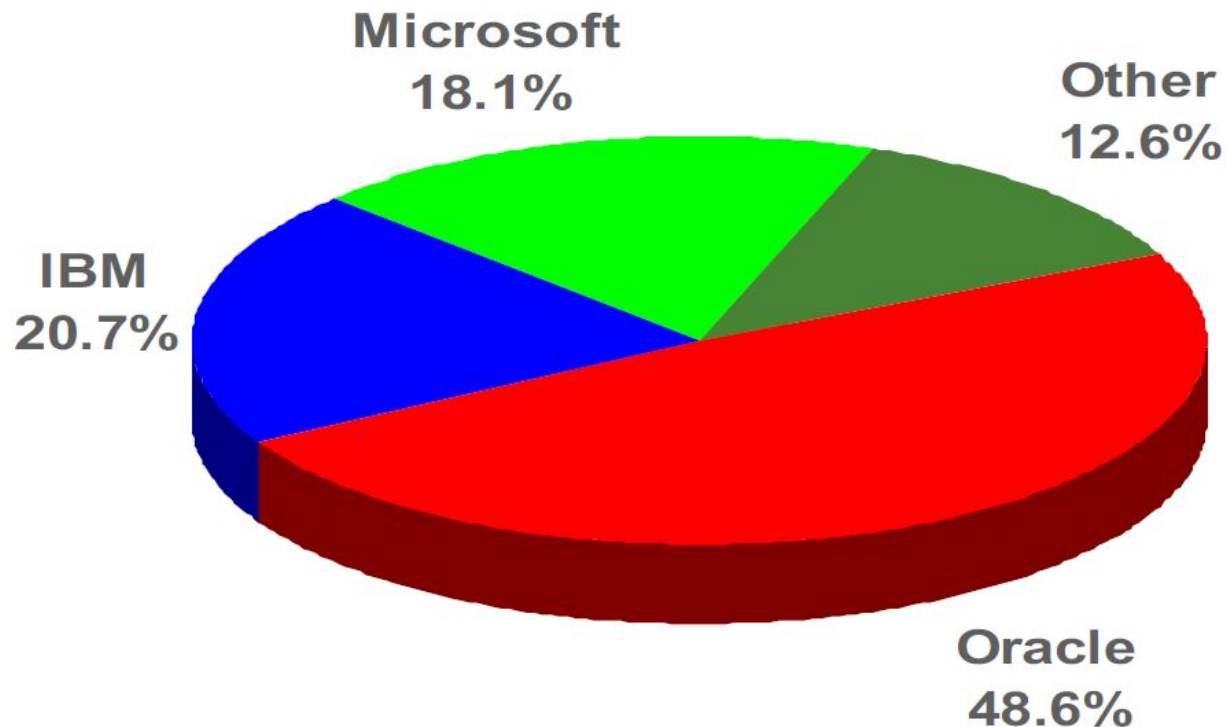
СУБД на рынке, прошлое и настоящее

15 лет назад - большая четверка коммерческих СУБД

✓ - Oracle, Informix, Sybase, Ingress

Сегодня в мире - большая тройка

✓ Oracle, Microsoft SQL Server, IBM DB2



Основные термины и определения

Структурирование – это введение соглашений о способах представления данных.

Создавая базу данных, пользователь стремится упорядочить информацию по различным признакам и быстро извлекать выборку с произвольным сочетанием признаков. Сделать это возможно, только если данные структурированы.

Неструктурированными называют данные, записанные, например, в текстовом файле.

Основные термины и определения

✱ По степени структурированности выделяют следующие формы представления данных:

- ☑ неструктурированные
- ☑ структурированные
- ☑ слабоструктурированные

✱ К **неструктурированным** относятся данные, произвольные по форме, включающие тексты и графику, мультимедиа (видео, речь, аудио). Эта форма представления данных широко используется, например, в Интернете, а сами данные предоставляются пользователю в виде отклика поисковыми системами.

Основные термины и определения



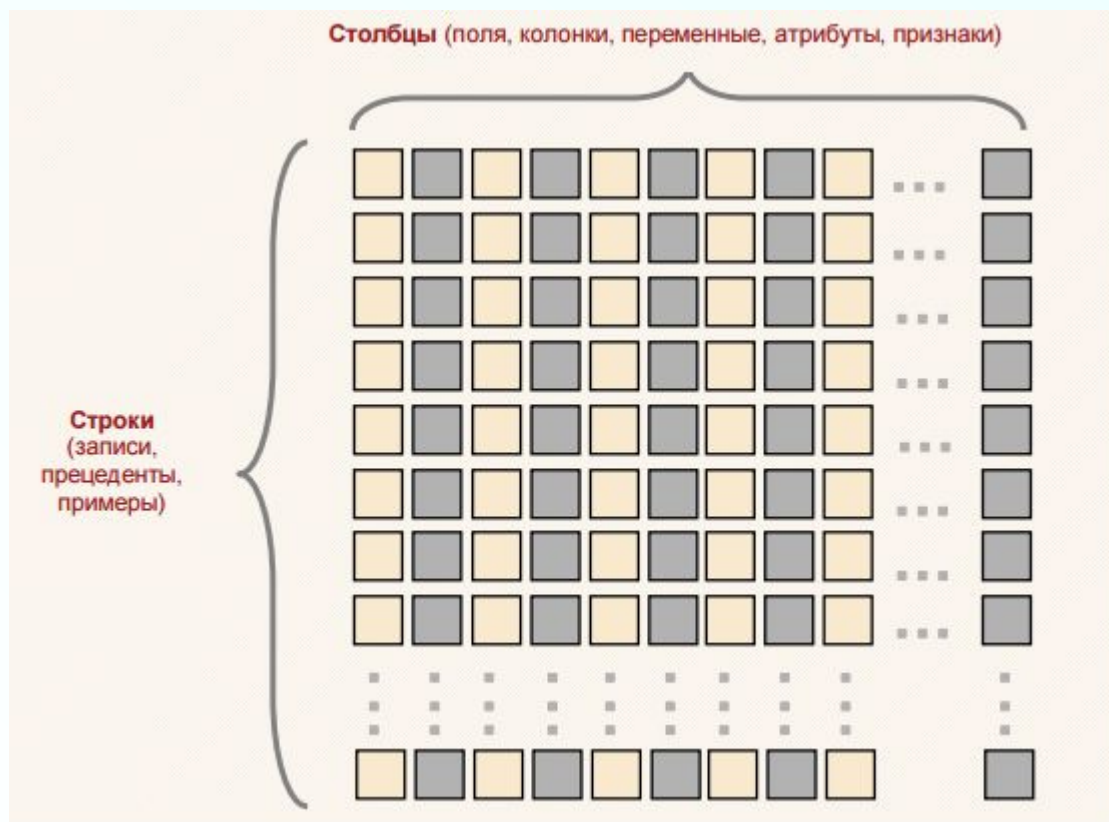
Структурированные данные отражают отдельные факты предметной области. Структурированными называются данные, определенным образом упорядоченные и организованные с целью обеспечения возможности применения к ним некоторых действий (например, визуального или машинного анализа). Это основная форма представления сведений в **базах данных**.



Организация того или иного вида хранения данных (структурированных или неструктурированных) связана с обеспечением доступа к ним. Под **доступом** понимается возможность выделения элемента данных (или множества элементов) среди других элементов по каким-либо признакам с целью выполнения некоторых действий над элементом.

Основные термины и определения

❄️ Одной из самых распространенных моделей хранения структурированных данных является таблица. В ней все данные упорядочиваются в двумерную структуру, состоящую из столбцов и строк.



❄️ В ячейках такой таблицы содержатся элементы данных: символы, числа, логические значения.

Основные термины и определения

- ❄️ **Неструктурированные** данные непригодны для обработки напрямую методами анализа данных, поэтому такие данные подвергаются специальным приемам структуризации, причем сам характер данных в процессе структуризации может существенно измениться. Например, в анализе текстов (Text Mining) при структурировании из исходного текста может быть сформирована таблица с частотами встречаемости слов, и уже такой набор данных будет обрабатываться методами, пригодными для структурированных данных.

Пример структуризации данных

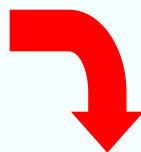
Личное дело № 16493, Сергеев Петр Михайлович, дата рождения 1 января 1976 г.;

Л/д № 16593, Петрова Анна Владимировна, дата рожд. 15 марта 1975 г.;

№ личн. дела 16693, д.р. 14.04.76, Анохин Андрей Борисович.

Структурированные
данные

Неструктурированные
данные



№ личного дела	Фамилия	Имя	Отчество	Дата рождения
16493	Сергеев	Петр	Михайлович	01.01.76
16593	Петрова	Анна	Владимировна	15.03.75
16693	Анохин	Андрей	Борисович	14.04.76

Основные термины и определения

- ** Слабоструктурированные данные — это данные, для которых определены некоторые правила и форматы, но в самом общем виде. Например, строка с адресом, строка в прайс-листе, ФИО и т.п.
- ** В отличие от неструктурированных, такие данные с меньшими усилиями преобразуются к структурированной форме, однако без процедуры преобразования они тоже непригодны для анализа.

Пример структуризации строки с адресом



Основные термины и определения

Целостность БД — соответствие имеющейся в базе данных информации её внутренней логике, структуре и всем явно заданным правилам. Каждое правило, налагающее некоторое ограничение на возможное состояние базы данных, называется **ограничением целостности** (integrity constraint).

Транзакция – законченная совокупность действий над БД, которая переводит БД из одного целостного состояния в другое целостное состояние. Совокупность простых операций над БД, объединенных в единое целое, и выполняемых по принципу «все или ни одной». Т.е. в случае возникновения ошибок при выполнении какой-либо операции, входящей в транзакцию, БД возвращается в состояние до выполнения транзакции.

Привилегия пользователя – права пользователя на выполнение операций с данными (запись, корректировка, чтение, удаление), а также выполнение других действий над БД.

Классификация СУБД и БД

Классификация СУБД и БД

По сфере возможного применения:

✓ **универсальные**

Пример: MS Access, PostgreSQL.

✓ **специализированные (проблемно-ориентированные) СУБД и БД**

Примеры ИС, в которых необходимо использовать специализированные СУБД:

- отдельные АСУТП, где нужна СУБД реального времени, обладающая полной функциональностью универсальной СУБД;
- биометрические системы;
- системы военного назначения;
- государственные информационные системы и т.д.

Пример: В АС военного назначения используются СУБД ЛИНТЕР и Линтер-ВС.

Классификация СУБД

По «мощности» СУБД делятся на:

- ✓ **«Настольные»** – невысокие требования к техническим средствам, ориентация на конечного пользователя («дружелюбность» интерфейса, простота создания БД и обработки информации), низкая стоимость.
Пример: MS Access, Visual FoxPro.
- ✓ **Корпоративные** – обеспечивают работу в распределенной среде, высокую производительность, имеют развитые средства администрирования и более широкие возможности поддержания целостности. Системы сложны, дороги, требуют значительных вычислительных мощностей.
Примеры: Oracle, DB2, Sybase, MS SQL Server, Progress

Корпоративные СУБД, как правило, реализуют архитектуру клиент-сервер.

Помимо хранения централизованной базы данных центральная машина – **сервер базы данных**, должна обеспечивать выполнение основного объема обработки данных. Запрос на данные, выдаваемый клиентом (рабочей станцией), порождает поиск и извлечение данных на сервере. Извлеченные данные транспортируются по сети от сервера к клиенту. Спецификой архитектуры клиент-сервер является использование языка запросов SQL.



Классификация СУБД

По характеру использования СУБД делятся на:

- ✓ **Персональные.** Обеспечивают возможность создания персональных БД и недорогих приложений, работающих с ними. Персональные СУБД или разработанные с их помощью приложения зачастую могут выступать в роли клиентской части многопользовательской СУБД
Пример: MS Access, Visual FoxPro, Paradox.
- ✓ **Многопользовательские.** Включают в себя **сервер БД** и **клиентскую часть** и, как правило, могут работать в неоднородной вычислительной среде (с разными типами ЭВМ и операционными системами).
Примеры: Oracle, MySQL.

Классификация СУБД

По степени доступности БД выделяют:

✓ Общедоступные БД.

Примеры: Банк документов на сайте Президента Российской Федерации (<http://kremlin.ru/>), Информационно-правовая система «Законодательство России» (<http://pravo.gov.ru/>).

✓ БД с ограниченным доступом пользователей.

В качестве примера можно привести БД, используемые в системе органов внутренних дел (криминалистические учеты, розыскные учеты, оперативно-справочные учеты, автоматизированные банки данных дактилоскопической информации (АДИС "Папилон")

Классификация СУБД

По способу доступа к БД выделяют:

✓ Файл-серверные СУБД.

Примеры: Microsoft Access, Paradox, dBase, FoxPro, Visual FoxPro.

✓ Клиент-серверные СУБД.

Примеры: Oracle, Firebird, Interbase, IBM DB2, Informix, MS SQL Server, Sybase Adaptive Server Enterprise, PostgreSQL, MySQL, Caché, ЛИНТЕР

Классификация СУБД

В **файл-серверных СУБД** файлы данных располагаются централизованно на файл-сервере. СУБД располагается на каждом клиентском компьютере (рабочей станции). Доступ СУБД к данным осуществляется через локальную сеть. Синхронизация чтений и обновлений осуществляется посредством файловых блокировок.

Преимуществом этой архитектуры является низкая нагрузка на процессор файлового сервера.

Недостатки: потенциально высокая загрузка локальной сети; затруднённость или невозможность централизованного управления; затруднённость или невозможность обеспечения таких важных характеристик, как высокая надёжность, высокая доступность и высокая безопасность. Применяются чаще всего в локальных приложениях, которые используют функции управления БД; в системах с низкой интенсивностью обработки данных и низкими пиковыми нагрузками на БД.

На данный момент файл-серверная технология считается устаревшей, а её использование в крупных ИС — недостатком.

Классификация СУБД

Клиент-серверная СУБД располагается на сервере вместе с БД и осуществляет доступ к БД непосредственно, в монопольном режиме. Все клиентские запросы на обработку данных обрабатываются клиент-серверной СУБД централизованно.

Недостаток клиент-серверных СУБД состоит в повышенных требованиях к серверу.

Достоинства: потенциально более низкая загрузка локальной сети; удобство централизованного управления; удобство обеспечения таких важных характеристик, как высокая надёжность, высокая доступность и высокая безопасность.

Классификация СУБД

По форме представляемой информации выделяют фактографические, документальные, мультимедийные БД.

- ✓ фактографические БД, в которых хранится информация об интересующих пользователя объектах предметной области в виде «фактов» (facts – «данные» (англ.)). Например, данные о сотрудниках, данные о поставщиках и поставках продукции и т.п. При этом в качестве факта рассматривается неделимый по смыслу информационный элемент, отражающий значение какого-либо свойства объекта. Примеры: Oracle, PostgreSQL, MySQL, Microsoft Access.

Классификация СУБД

По форме представляемой информации выделяют:

- ✓ **документальные БД.** Единицей хранения является какой-либо документ (например, текст закона или статьи), и пользователю в ответ на его запрос выдается либо ссылка на документ, либо сам документ, в котором он может найти интересующую его информацию. К документальным БД относятся, например, базы данных научного цитирования: РИНЦ, Web of Science, Scopus и т.д. Большое распространение получили документальные справочные правовые системы КонсультантПлюс, ГАРАНТ, Кодекс.
- ✓ **мультимедийные БД,** содержащие мультимедийную информацию: картографические, видео-, аудио-, графические и др. Пример: банк данных дактилоскопической информации (АДИС «Папилон»).

Классификация СУБД

По типологии хранения – **локальные и распределенные**.

По функциональному назначению (характеру решаемых задач и, соответственно, характеру использования данных) – **операционные и справочно-информационные**. К последним относятся ретроспективные БД (электронные каталоги библиотек), которые используются для информационной поддержки основной деятельности и не предполагают внесения изменений в уже существующие записи, например, по результатам этой деятельности. Операционные БД предназначены для управления различными технологическими процессами.

Классификация СУБД и БД

По способу распространения:

- ✓ **Commercial Software** — коммерческие (с ограниченными лицензией возможностями на использование), разрабатываемое для получения прибыли.
Примеры: Oracle, Microsoft Access.
- ✓ **Freeware** — свободные, распространяемые без ограничений на использование, модификацию и распространение.
Примеры: MySQL.
- ✓ **Shareware** — условно-бесплатные.

Классификация СУБД

По характеру преобладающей обработки информации:

- ✓ **OLTP (On-Line Transaction Processing)** – системы оперативной обработки информации
- ✓ **OLAP (On-Line Analytical Processing)** – системы для сложной аналитической обработки информации

Характеристика	OLTP	OLAP
Преобладающие операции	Ввод данных, поиск	Анализ данных
Характер запросов	Преобладают простые транзакции и запросы	Сложные запросы
Характеристика хранимых данных	Оперативные, детализированные	Агрегированные, охватывающие большой период времени

Классификация СУБД

По используемой Модели данных

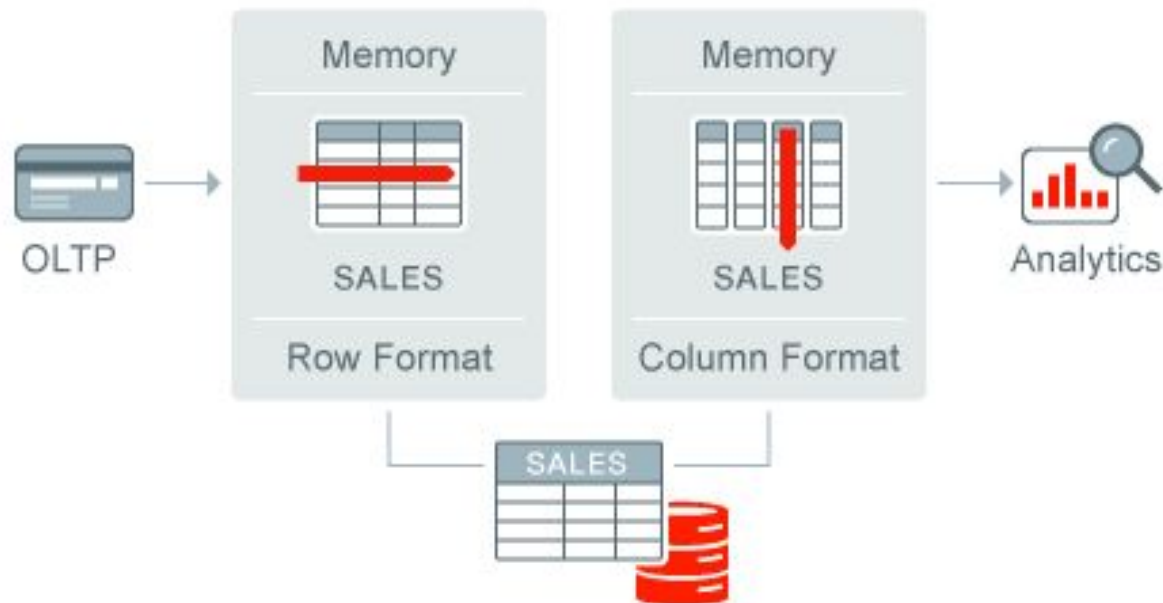
- ✓ Иерархические
- ✓ Сетевые
- ✓ Реляционные
- ✓ Объектно-ориентированные
- ✓ Объектно-реляционные
- ✓ NoSQL:
 - “Ключ-значение”
 - Документные
 - Поколоночные (столбцовые)



Перспективы развития баз данных

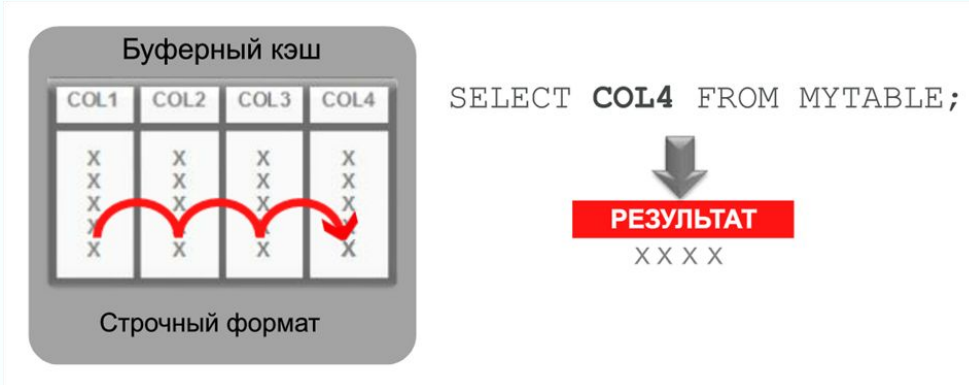
Технология In-Memory

- Технология In-Memory появилась в версии Oracle Database 12.1.0.2. Суть ее заключается в том, что рядом привычным буферным кэшем, который хранит строки таблиц и блоки индексов, находится новый кэш, точнее новая разделяемая область для данных в оперативной памяти, в которой данные из таблиц хранятся в колоночном формате.
- При использовании технологии In-Memory аналитика работает в сотни раз быстрее, потому что колоночное представление для нее более эффективно.

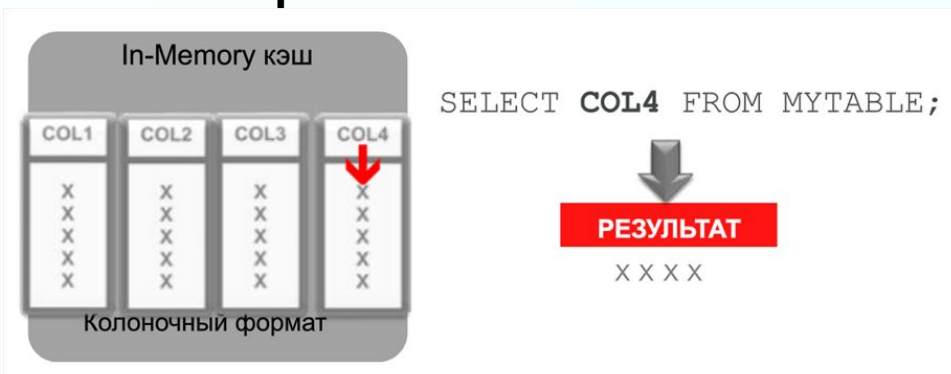


Технология In-Memory

- В обычном буферном кэше информация хранится по строкам. Вот пример — из четырехколоночной таблицы нужно извлечь колонку №4. Для этого придется полностью просканировать всю эту табличку в оперативной памяти:



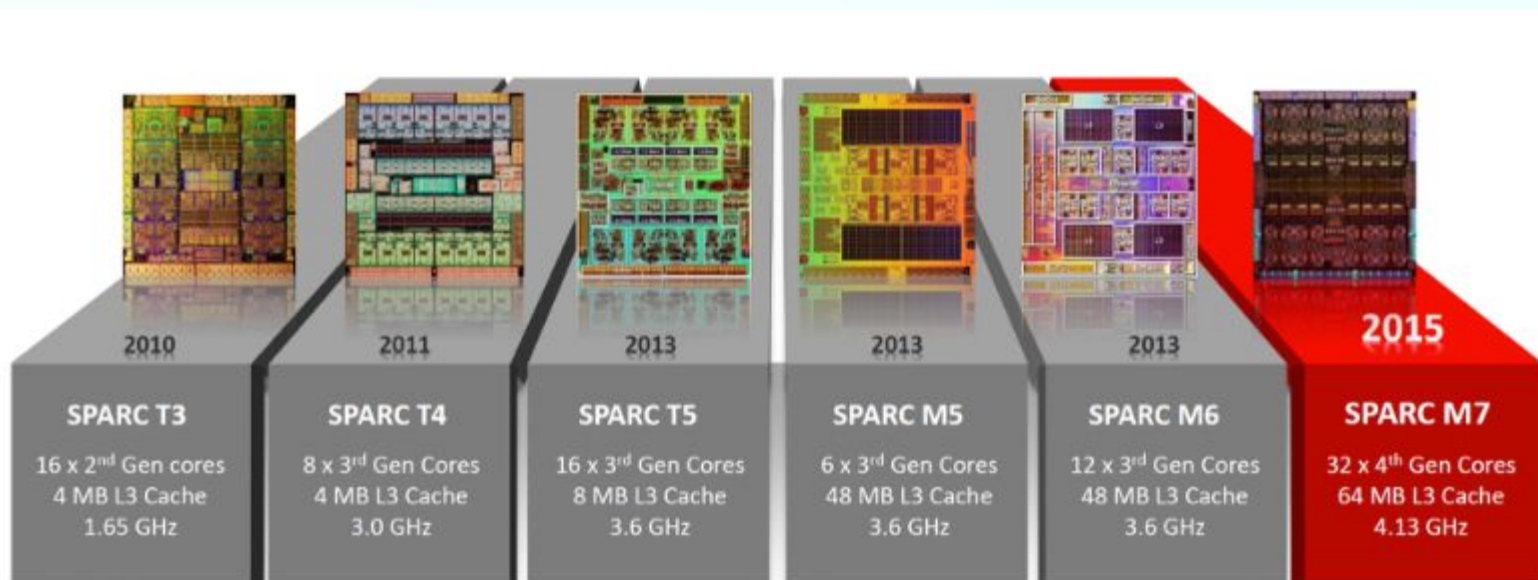
- Если та же таблица хранится в колоночном формате, то вся четвертая колонка нашей таблички находится в одном экстенсте, т.е. в одном блоке памяти. Мы можем сразу выделить ее, тут же прочитать и вернуть приложению. Уменьшаются затраты на сканирование, на пересылку этих данных процессору, снижается загрузка процессора. Все работает значительно быстрее.



Такие операции сканирования очень характерны для ERP-приложений, для хранилищ данных в аналитических системах.

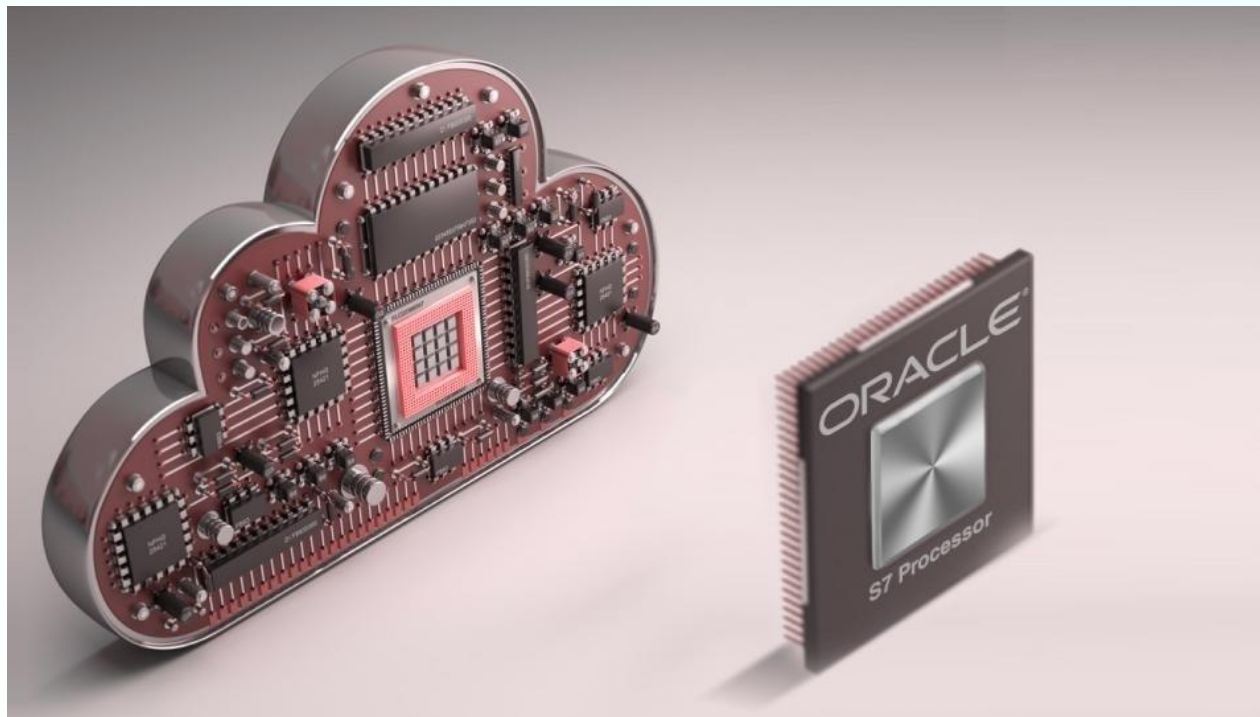
Технология SPARC

☀️ Технология SPARC принадлежит Oracle уже пять лет. За это время корпорация Oracle выпустила микропроцессоры SPARC T3, SPARC T4, SPARC T5, SPARC M5 и SPARC M6, каждый из которых был важным шагом на пути эволюции технологий — причем системы SPARC T3 и T4 разрабатывались еще компанией **Sun Microsystems**, и последующие процессоры многое унаследовали от них.



Процессор Oracle SPARC M7

✦ **SPARC M7** - первый процессор, который полностью, начиная с идеологии и базового дизайна, разрабатывался Oracle и для Oracle. Основной целью проекта разработки было обеспечить максимальную эффективность работы ПО Oracle — и в результате был создан первый в индустрии 32-ядерный процессор с беспрецедентными нагрузочной способностью, производительностью ядра, возможностями быстрого шифрования и аппаратной декомпрессии.



Процессор Oracle SPARC M7

Если сравнить параметры микропроцессора SPARC M7 с параметрами самого совершенного выпущенного ранее процессора SPARC T5, обнаружится, что многие параметры увеличились в два раза, а некоторые — в четыре. У процессора SPARC M7 32 ядра общего назначения, т. е. вдвое больше, чем у процессора T5. Также у процессора SPARC M7 вдвое больше вычислительных потоков и вчетверо больше кэша на каждое ядро, а новая архитектура существенно повысила производительность каждого ядра. Новые контроллеры памяти позволили увеличить пропускную способность памяти и скорость доступа к памяти, а пропускная способность ввода-вывода выросла в четыре раза. Выросла и тактовая частота процессора. В целом процессор SPARC M7 работает примерно в три раза быстрее, чем процессор SPARC T5.

	SPARC T5	SPARC M7	Повышение
Число ядер	16	32	2^x
Число потоков	128	256	2^x
Объем кэш-памяти на ядро	0,5 МБ	2 МБ	4^x
Пропускная способность памяти	79 ГБ/с	160 ГБ/с	2^x
Скорость доступа к памяти	163 нс	131 нс	20%
Пропускная способность ввода-вывода	32 ГБ/с	145 ГБ/с	4^x
Частота	3,60 ГГц	4,13 ГГц	15%

Процессор Oracle SPARC T7

Исторический революционный шаг, сделанный Oracle новым процессором, — это реализация программных функций непосредственно на кристалле. Это безопасность на кристалле, т. е. ускорение шифрования и аппаратная защита памяти, и, что еще важнее, — SQL на кристалле, т. е. первая в мире аппаратная реализация ускорения обработки SQL-запросов и декомпрессии для Oracle Database In-Memory.



Безопасность на кристалле

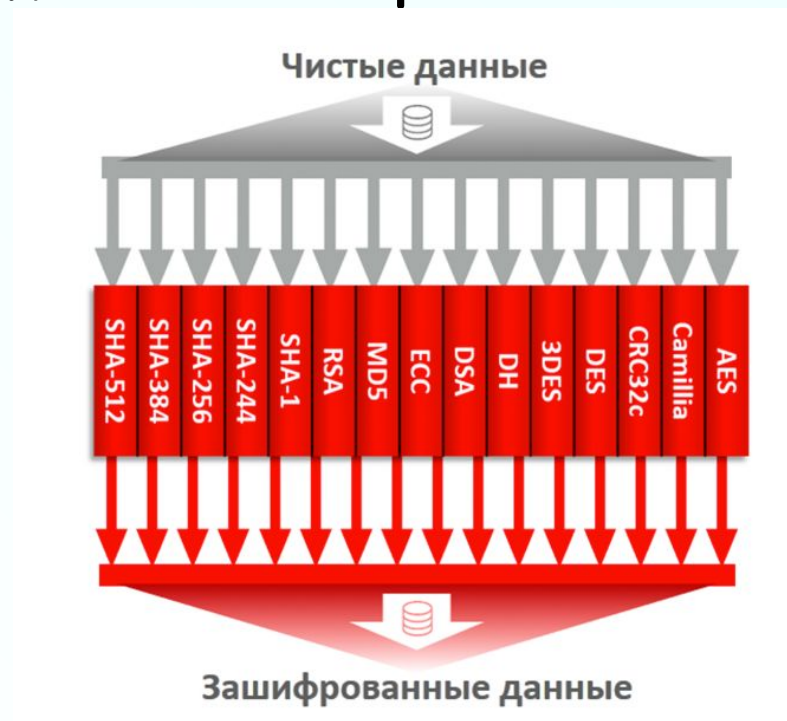
☀ Количество похищенных строк данных в мире за 2014 год, согласно отчету CSO Online Market Pulse, составило сотни миллионов, а понесенные бизнесом потери измеряются миллионами и миллионами долларов. Но, к сожалению, в системе корпоративной безопасности защита баз данных фактически является сейчас самым слабым местом. Более 50 % компаний считают, что самые важные и самые ценные данные хранятся у них в базах данных, но при этом большинство компаний инвестируют в первую очередь в защиту сети, а на защиту баз данных предпочитают тратить как можно меньше средств. В результате 76 % всех успешных атак на корпоративные данные не были остановлены именно средствами сетевой безопасности.

☀ Существует три основных вида угроз безопасности:

- ☑ Уязвимости базы данных как физического носителя
- ☑ Уязвимости операционной системы, приводящие к проникновению и получению несанкционированного доступа к данным
- ☑ Ошибки доступа к памяти

Безопасность на кристалле

- Процессор SPARC M7 обладает уникальной функциональностью, позволяющей обеспечить прозрачное шифрование данных с использованием 15 наиболее известных алгоритмов шифрования: в каждое ядро процессора встроен специализированный математический блок обработки инструкций шифрования, который обеспечивает скорость шифрования, практически равную скорости работы основного ядра и скорости работы с памятью. Поскольку все алгоритмы обрабатываются непосредственно в процессоре, падение производительности при этом составляет менее 3 %.



Безопасность на кристалле

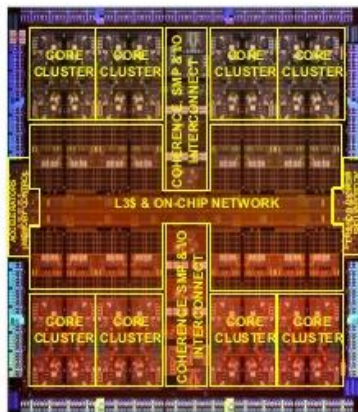
☀ Системы на базе SPARC M7 предлагают также аппаратную поддержку безопасной миграции доменов. В процессе миграции виртуальная машина с критичными данными передается через сеть, и образ виртуальной машины шифруется для передачи. При этом данные защищены во время передачи сложным алгоритмом шифрования. Таким образом обеспечивается защита передаваемых данных с минимальным влиянием на производительность мигрирующей виртуальной машины во время переноса.



Безопасность на кристалле

Большинство вирусов для систем RISC/UNIX пытаются напрямую адресовать память за рамками отведенных им буферов, и используют для этого либо механизмы переполнения стека, либо механизмы переполнения буфера. Система SPARC M7 впервые в истории имеет аппаратную защиту памяти и позволяет предотвращать несанкционированный доступ к памяти на уровне аппаратных процессорных ресурсов. Эта функция предотвращает доступ вредоносных программ и к памяти приложений, и к каким-либо функциям операционной системы, при этом она не влияет на производительность и ее невозможно обойти.

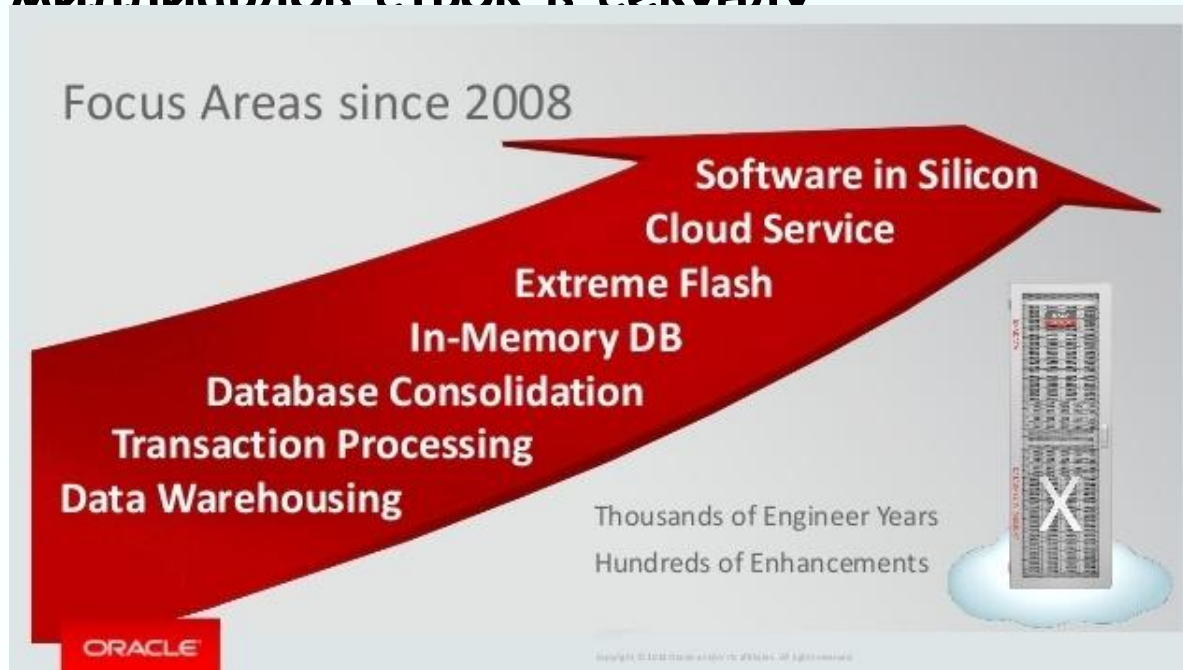
SPARC M7 Processor



- 32 SPARC Cores
 - Fourth Generation CMT Core (S4)
 - Dynamically Threaded, 1 to 8 Threads Per Core
- New Cache Organizations
 - Shared Level 2 Data and Instruction Caches
 - 64MB Shared & Partitioned Level 3 Cache
- DDR4 DRAM
 - Up to 2TB Physical Memory per Processor
 - 2X-3X Memory Bandwidth over Prior Generations
- Application Acceleration
 - Real-time Application Data Integrity
 - Data Base Query Offload Engines
 - In-Memory Columnar Decompression at Full Bandwidth
- SMP Scalability from 1 to 32 Processors
- Technology: 20nm, 13ML
- Oracle SW Core Factor 0.5

SQL на кристалле

- ★ SQL in Silicon — обработка запросов к базе данных, реализованная непосредственно на процессоре.
- ★ В процессоре SPARC M7 имеются специализированные ускорители SQL-инструкций, которые работают независимо, в синхронном и асинхронном режиме. И если с переходом на In-Memory скорость обработки инструкций составила миллионы строк в секунду, то с использованием специализированных ускорителей в процессоре M7 она достигла миллиардов строк в секунду



Серверы Oracle SPARC T7 и M7

☀ Максимальный результат, достигнутый на внутренних тестах Oracle, составил 170 млрд строк в секунду на процессорах SPARC M7 с использованием механизма In-Memory и встроенных сопроцессоров. Встроенные сопроцессоры не только повышают скорость обработки SQL-запросов, но и освобождают процессорные ядра общего назначения для работы других приложений — OLTP-запросов и пр.



Серверы Oracle SPARC T7 и M7

✳ В результате аналитика на SPARC M7 работает более чем в восемь раз быстрее, чем на системной архитектуре x86 платформы. OLTP работает примерно в три раза быстрее. Это значит, что там, где раньше требовалось пять двухпроцессорных серверов для обработки OLTP и аналитики, теперь можно обойтись одним однопроцессорным сервером на базе SPARC M7, который будет одновременно обрабатывать и OLTP-, и аналитические запросы.

✳ Один из крупных заказчиков Oracle, который занимается онлайн-торговлей, при тестировании сервера SPARC T7-4 на базе процессоров SPARC M7 с Oracle Database 12.1.0.2 и опцией In-Memory получил повышение скорости обработки запросов в 83 раза.



Спасибо за внимание!