

Базы данных и Системы управления базами данных

Лекция 1

О дисциплине: Литература

Основная

- 1. Петкович Д. Microsoft SQL Server 2008: рук-во для начинающих. - СПб. : БХВ-Петербург, 2012. - 730 с. : ил.. - ISBN 978-5-9775-0149-1.
- 2. Советов Б.Я., Цехановский В.В., Чертовский В.Д. Базы данных: теория и практика : учебник для студ. вузов. - 2-е изд.. - М. : Юрайт, 2012. - 463 с.. - (Бакалавр). - Библиогр.: с. 459-460 (49 назв.). - ISBN 978-5-9916-1479-5.

Дополнительная

- Хомоненко А.Д., Цыганков В.М., Мальцев М.Г. Базы данных и СУБД: Учебник для высших учебных заведений / под ред проф. А.Д. Хомоненко. – 6-е изд. - СПб. : КОРОНА-Век, 2010. - 736 с.
- Гузик В.Ф., Костюк А.И., Ляпунцова Е.В., Б.В.Катаев. Базы данных и СУБД: Учебно-методическое пособие. – Таганрог: Изд-во Технологического института ЮФУ, 2007. – 119.
- **Дейт К.Дж. Введение в системы баз данных. – 8-е изд. – М: Вильямс, 2005.- 1327с.**
- Кляйн К.Е. SQL [Текст] : справочник / пер. Е. Демьянова, А. Слинкина. - 3-е изд.. - СПб.-М. : Символ, 2010. - 651 с. : ил.. - ISBN 978-5-93286-165-3. - ISBN 978-596-51884-4.
- Култыгин О. П. Администрирование баз данных. СУБД MS SQL Server [Текст] : учеб. пособие. - М. : Московская финансово-промышленная академия, 2012. - 228 с.. - (Университетская серия). - Библиогр.: с. 220-221. - ISBN 978-5-4257-0026-1
- Астахова И. Ф., Мельников В. М., Толстобров А. П., Фертиков В. В. СУБД: язык SQL в примерах и задачах [Текст] : учеб. пособие для студ. вузов. - М. : Физматлит, 2009. - 165 с.. - (Информационные и компьютерные технологии). - ISBN 978-5-9221-0816-4
- Кузнецов С.Д. Основы баз данных [Текст] : учеб. пособие. - 2-е изд., испр.. - М. : Интернет-Университет Информационных Технологий : БИНОМ. Лаборатория знаний, 2010. - 484 с. : ил.. - (Основы информационных технологий). - Библиогр.: с. 480-484. - ISBN 978-5-94774-736-2.

Назначение и основные компоненты баз данных

Исторический экскурс

1 Этап: Автоматизация вычислительных задач (до 1955 года)

Предметная область: Инженерные и научные расчеты

Особенности работы:

- Простое представление данных – одиночные значения заданного типа;
- Сложный, многооперационный процесс вычислений

Способ представления данных: Значения переменных, вводимые пользователем

Модель данных: отсутствует

С примерами...



10 грамм



5 килограмм



Возможность обработки массивов данных

2 Этап: Программируемое оборудование обработки записей (1955- середина 1960х)

Предметная область: ...+ Статистическая обработка, учетные задачи и т.п.

Особенности работы:

- Простое представление данных – множества значений заданного типа;
- Упрощение вычислительных задач, появление прикладного программного обеспечения

Способ представления данных: файлы (перфокарты, магнитные ленты)

Модель данных: простые индексно-последовательные организации записей

Массовое производство и продажа...



Январь: 20 бутылок
по 300 рублей



Февраль: 15 бутылок
по 320 рублей



Март: 17 бутылок по
295 рублей



...



Сложности автоматизации



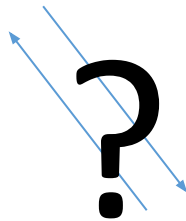
Файлы данных



Обмен
данными



Программное
обеспечение
VineTrader
(ВАШЕ! ЛИЧНОЕ!
😊)



Yet another
программное
обеспечение

Появление баз данных

3 Этап: Дореляционные базы данных (середина 1960х-1970 год)

Предметная область: + автоматизация обработки отраслевой информации

Особенности работы:

- Сложное представление обрабатываемых данных – структурированные множества разнотипных значений;
- Простые алгоритмы обработки данных, реализованные в программном обеспечении.

Способ представления данных: Базы данных

Модель данных: иерархическая, сетевая

Развитие продаж, появление склада...

Произведенны
е запасы

Погреб

Холодильник

Merlot

Chardonnay

Pinot noir

15

10

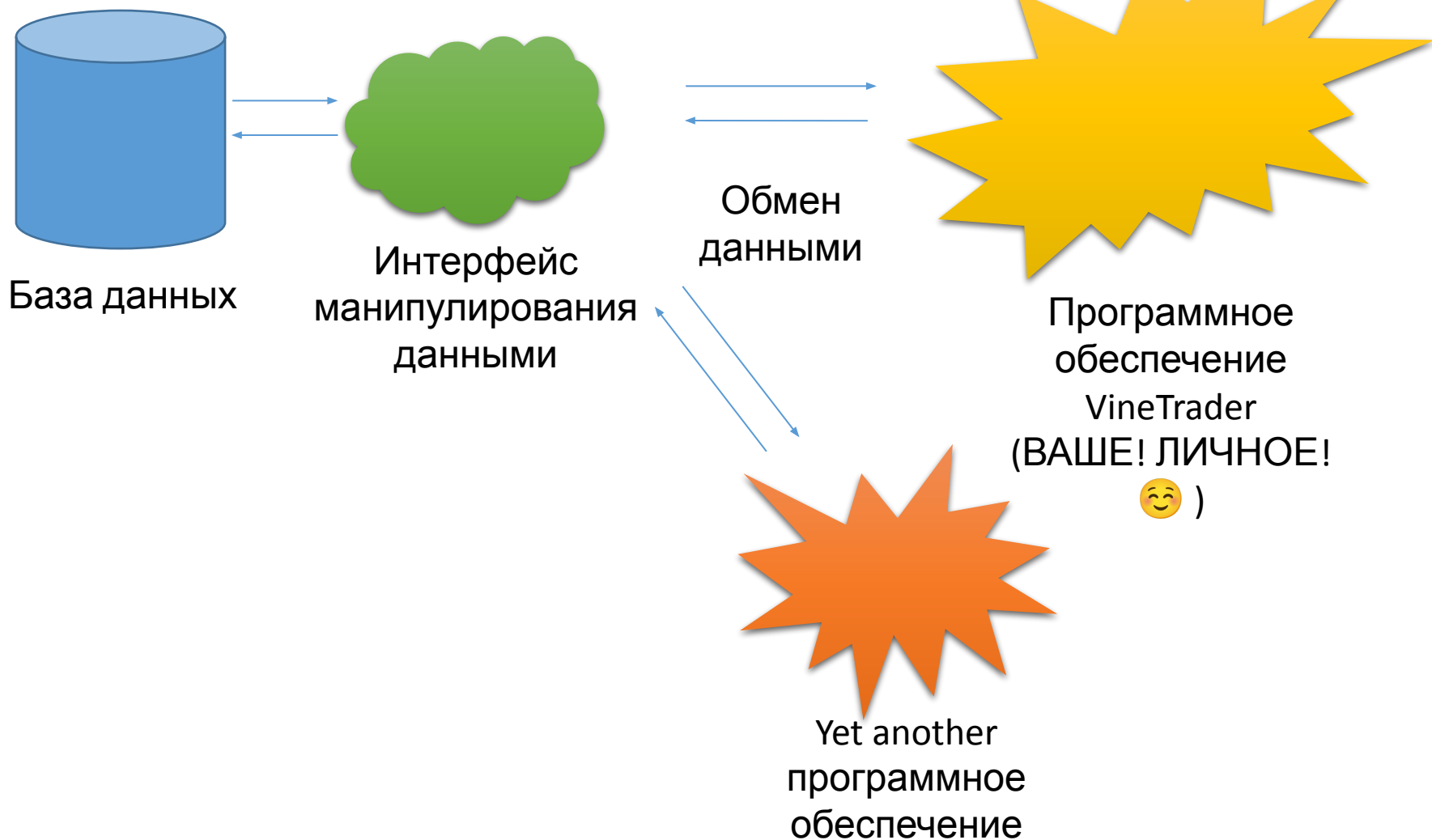
12

Списать со склада в
счет заказа №2 от
1.09.2019 :

10 бутылок Пино нуар

5 бутылок Мерло

Независимость программ и данных



Реляционная модель данных

Э. Кодда

4 Этап: Реляционные базы данных (1970 год – н.в.)

Предметная область: over9000 задач, связанных со сбором, обработкой, хранением и выдачей информации (т.н. информационные системы).

Особенности работы:

- Сложное представление обрабатываемых данных – табличное представление (отношения и кортежи);
- Простые алгоритмы обработки данных, реализованные в программном обеспечении Информационных систем.

Способ представления данных: Базы данных

Модель данных: Реляционная (таблицы и отношения)

Развитие деятельности...



- Учет продаж
- Планирование производства
- Складской учет
- Бухгалтерия
- Логистика
- ...

Определения...

Предметная область – часть реального мира, подлежащая изучению с целью организации управления и, в конечном итоге, автоматизации.

ДАННЫЕ – представление фактов, понятий или инструкций в форме, приемлемой для общения, интерпретации, или обработки человеком или с помощью автоматических средств

Модель данных – это абстрактное, самодостаточное, логическое определение объектов, операторов и прочих элементов, в совокупности составляющих абстрактную машину доступа к данным, с которой взаимодействует пользователь

База данных (БД) – это совокупность взаимосвязанных структурированных данных, относящихся к определенной предметной области и организованных так, чтобы обеспечить независимость данных от программ обработки.

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

[Автоматизированная] информационная система ([А]ИС) представляет собой совокупность данных, экономико-математических методов и моделей, технических, программных средств и специалистов, предназначенную для обработки информации и принятия управленческих решений

Примеры...

Предметная область	Данные	Модель данных	База данных	СУБД	[А]ИС
Образовательная деятельность ВУЗа	Посещаемость лекций по курсу	Иерархическая	Академическая успеваемость	MS SQL Server	АСУ Деканат
Статистические исследования	Температура в регионе за период	Сетевая	Метеорологические сводки	MySQL	АРМ Метеоролога
Торговля	Остаток n-го товара на складе	Реляционная	Складские запасы	Postgree	1С: Предприятие. Торговля и склад
Управление взаимоотношениями с клиентами (CRM)	Телефон клиента	Объектная	Клиенты компании	Redis	АРМ Менеджера по работе с клиентами
	

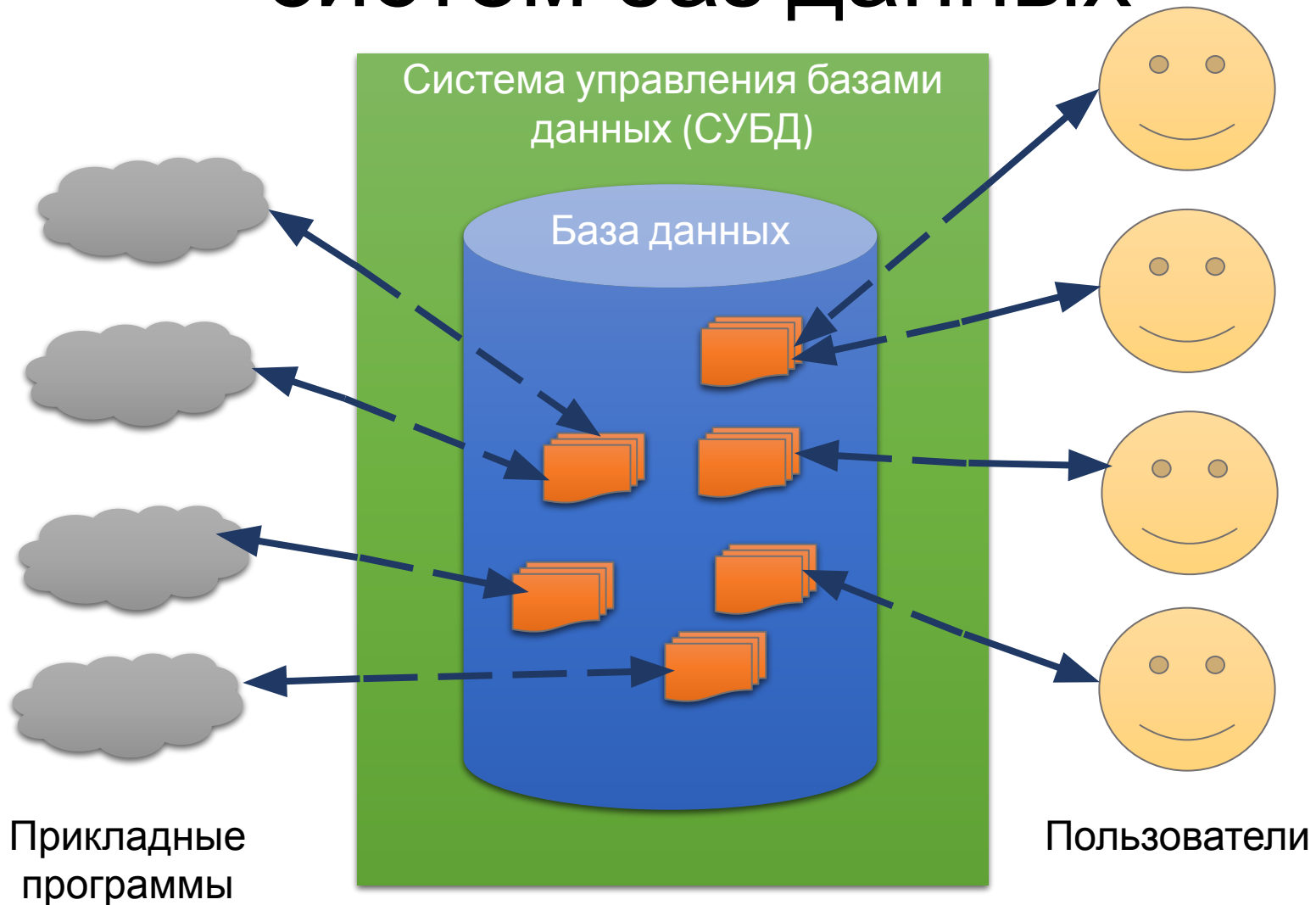
Особенности баз данных

- минимально необходимый объем хранимой информации о предметной области;
- атомарность базовых элементов (объектов);
- избыточность и целостность (согласованность) БД;
- полностью структурированная информация на основе модели данных;
- наличие метаданных;
- однозначный ответ на запрос;
- ориентированность на конкретную предметную область (сферу или объект применения).

Преимущества применения баз данных

- **Компактность.** Нет необходимости в создании и ведении, возможно, весьма объемистых бумажных картотек
- **Быстродействие.** Компьютер может выбирать и обновлять данные гораздо быстрее человека
- **Низкие трудозатраты.** Отпадает необходимость в утомительной работе над картотекой вручную
- **Актуальность.** В случае необходимости под рукой в любой момент имеется точная, свежая информация
- **Защита.** Данные могут быть лучше защищены от случайной потери и несанкционированного доступа
- **Возможность совместного доступа к данным**
- **Сокращение избыточности данных**
- **Устранение противоречивости данных** (до некоторой степени)
- **Возможность поддержки транзакций**

Основные компоненты систем баз данных



Основные компоненты систем баз данных: Данные

- Могут храниться в нескольких базах данных
- Бывают интегрируемые (объединяемые из нескольких источников) и разделяемые (предоставляемые по частям отдельным пользователям)
- Для каждого пользователя могут быть представлены различным образом

Основные компоненты систем баз данных: Аппаратное обеспечение

- тома внешней (вторичной) памяти, используемые для хранения информации, а также соответствующие устройства ввода—вывода, контроллеры устройств, каналы ввода—вывода и т.д.;
- аппаратный процессор (или процессоры) вместе с оперативной (первичной) памятью, предназначенные для поддержки работы программного обеспечения системы баз данных.

Основные компоненты систем баз данных: Программное обеспечение

- Система управления базами данных (диспетчер базы данных, сервер базы данных)
- Диспетчеры транзакций;
- Средства проектирования;
- Генераторы отчетов и т.д.

Основные компоненты систем баз данных:

Пользователи

- Первая группа— прикладные программисты
- Вторая группа— конечные пользователи
- Третья группа – администраторы баз данных

Уровни представления
баз данных.

Независимость от
данных

Трехуровневая система организации баз данных (стандарт ANSI)



позволяет обеспечить:

- логическую (между уровнями 1 и 2) независимость и
- физическую (между уровнями 2 и 3) независимость при работе с данными.

Трехуровневая система организации баз данных

Уровень внешних моделей:

- определяет точку зрения на БД отдельных приложений.
- Каждое приложение видит и обрабатывает только те данные, которые необходимы именно этому приложению.

Трехуровневая система организации баз данных

Концептуальный уровень:

- база данных представлена в наиболее общем виде, который объединяет данные, используемые приложениями, работающими с БД.
- Отражает обобщенную модель предметной области (объектов реального мира), для которой создавалась база данных.
- Отражает только существенные с точки зрения обработки особенности объектов реального мира.

Трехуровневая система организации баз данных

Физический уровень:

- собственно данные, хранимые в файлах или в страничных структурах, расположенных на внешних носителях информации.

Независимость от данных

- **невосприимчивость приложений к изменениям в физическом представлении данных и в методах доступа к ним**
- может быть реализована на двух уровнях: физическом и логическом
- отражает зависимость уровней представления данных между собой, а также зависимость приложений и данных

Отсутствие независимости от данных

- Способ организации данных во вторичной памяти и способ доступа к ним диктуются требованиями приложений;
- Сведения об организации данных и способе доступа к ним встроены в саму логику и программный код приложения;
- Отражает зависимость уровней представления данных между собой, а также зависимость приложений и данных;
- Невозможно изменить физическое представление (способ физического размещения данных во вторичной памяти) или метод доступа (конкретный способ доступа к данным), не изменив самого приложения.

Инфологическое моделирование данных

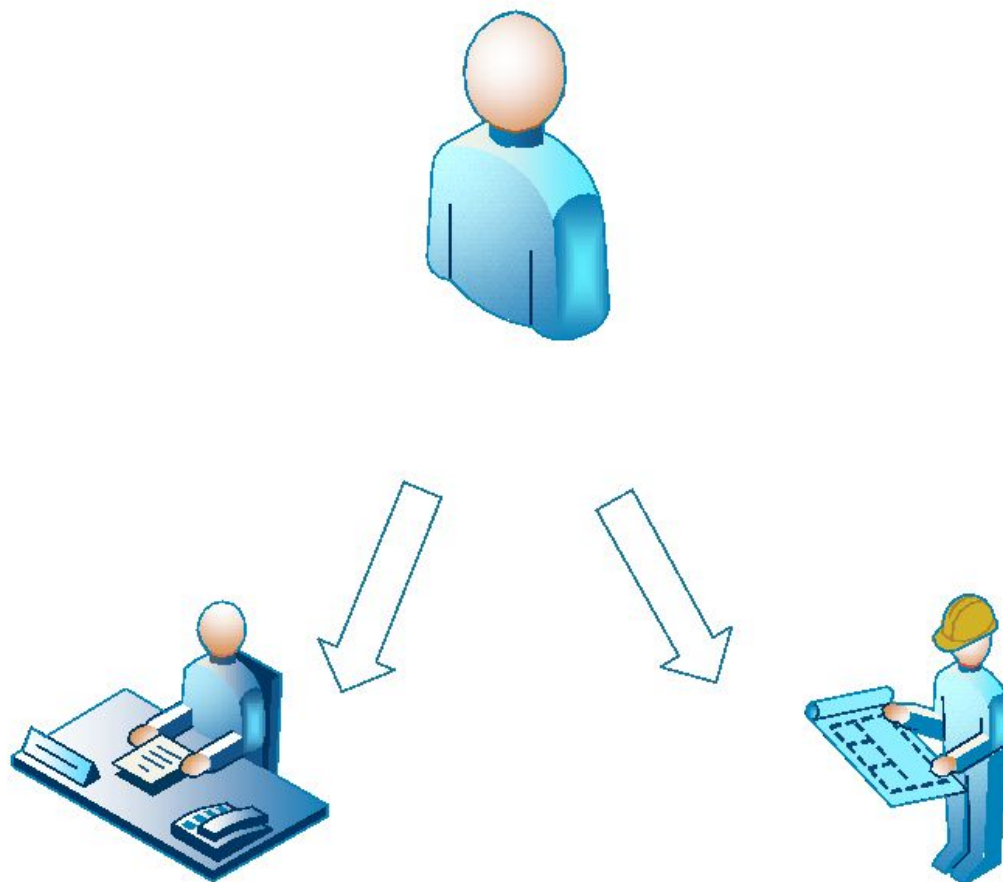
Инфологическое моделирование

- Информационная *модель данных* предназначена для представления семантики *предметной области* в терминах субъективных средств описания - сущностей, атрибутов, *идентификаторов сущностей, супертипов, подтипов* и т.д.

Содержание информационной модели

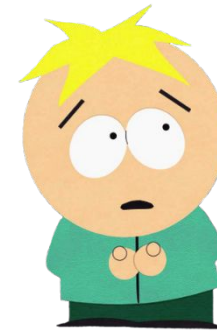
- диаграммы "*сущность-связь*" (Entity - *Relationship* Diagrams);
- определения сущностей;
- уникальные *идентификаторы сущностей* ;
- определения *атрибутов сущностей* ;
- отношения между сущностями;
- *супертипы* и подтипы.

Суццности



Сущность и экземпляр сущности

- Сущность описывает класс объектов предметной области
- Экземпляр сущности – один из объектов выделенного класса



Идентификация сущностей

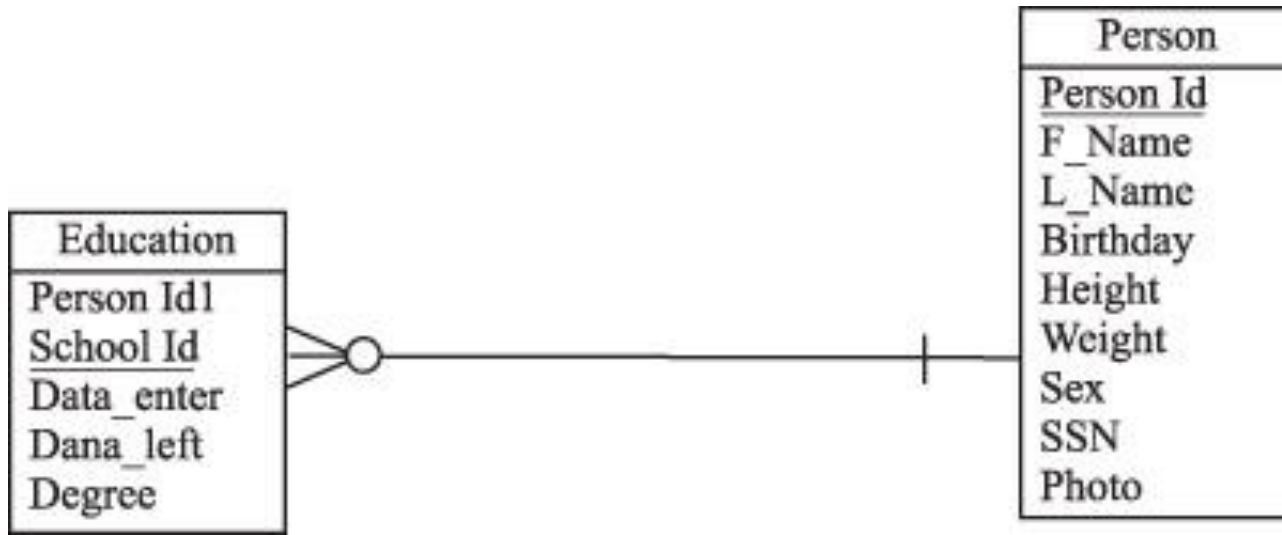
- Искусственный или естественный атрибут (набор атрибутов), значение которого уникально для каждого экземпляра
- **Правильно:** Номер зачетной книжки
- **Неправильно:** ФИО

Визуализация информационной модели на ER-диаграммах

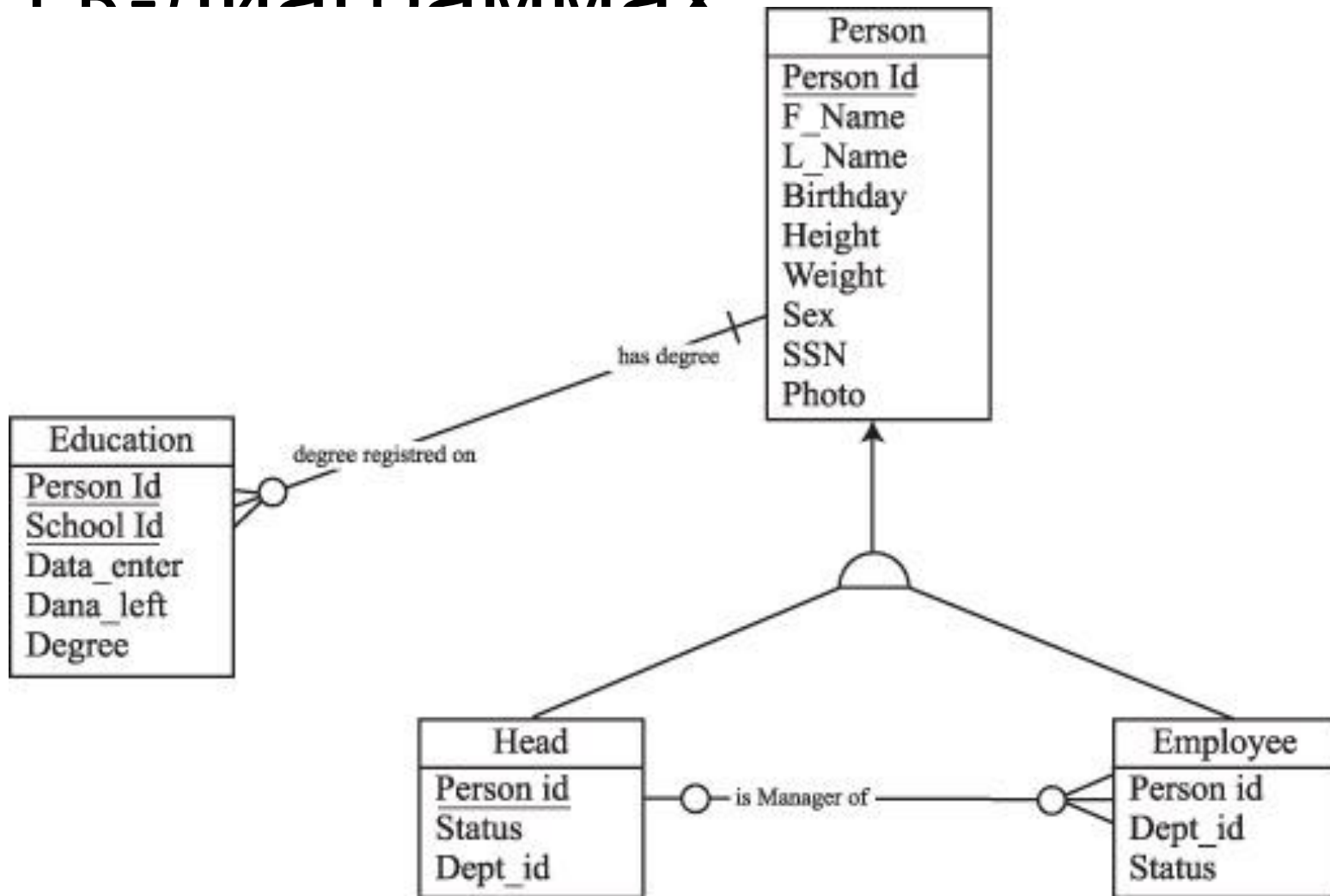
Person

Person
<u>Person Id</u>
F_Name
L_Name
Birthday
Height
Weight
Sex
SSN
Photo

PERSON	
<u>PERSON ID</u>	<u>INTEGER</u>
F_NAME	CHAR(30)
L_NAME	CHAR(20)
BIRTHDAY	DATE
HEIGHT	HUMBER(5,2)
WEIGHT	HUMBER(4,1)
SEX	HUMBER(1)
SSN	CHAR(30)
PHOTO	LONG RAW



Визуализация информационной модели на ER-диаграммах



Проектирование баз данных

Особенности проектирования баз данных

- Является частью процесса проектирования информационных систем
- Определяет вероятность удовлетворения информационной системой функциональных требований
- Необходимо учитывать не только требования, но и ограничения
- ❖ *Выдача отчетов по продажам по регионам, автоматический расчет скидки*
- ❖ *Максимально допустимое время расчета, используемые технологии*

Как оценивают базы данных?

- Функциональность и адаптируемость
- Производительность обработки транзакций
- Пропускная способность
- Время реакции
- Безопасность

Что такое проектирование баз данных?

Проектирование базы данных - это поиск способов удовлетворения *функциональных требований* средствами имеющейся компьютерной технологии с учетом заданных ограничений

Основной задачей проектировщика *базы данных* является обоснованный выбор такой ее структуры, которая обеспечит согласованное взаимодействие всех ее компонентов согласно заданным функциональным требованиям в рамках заданных ограничений

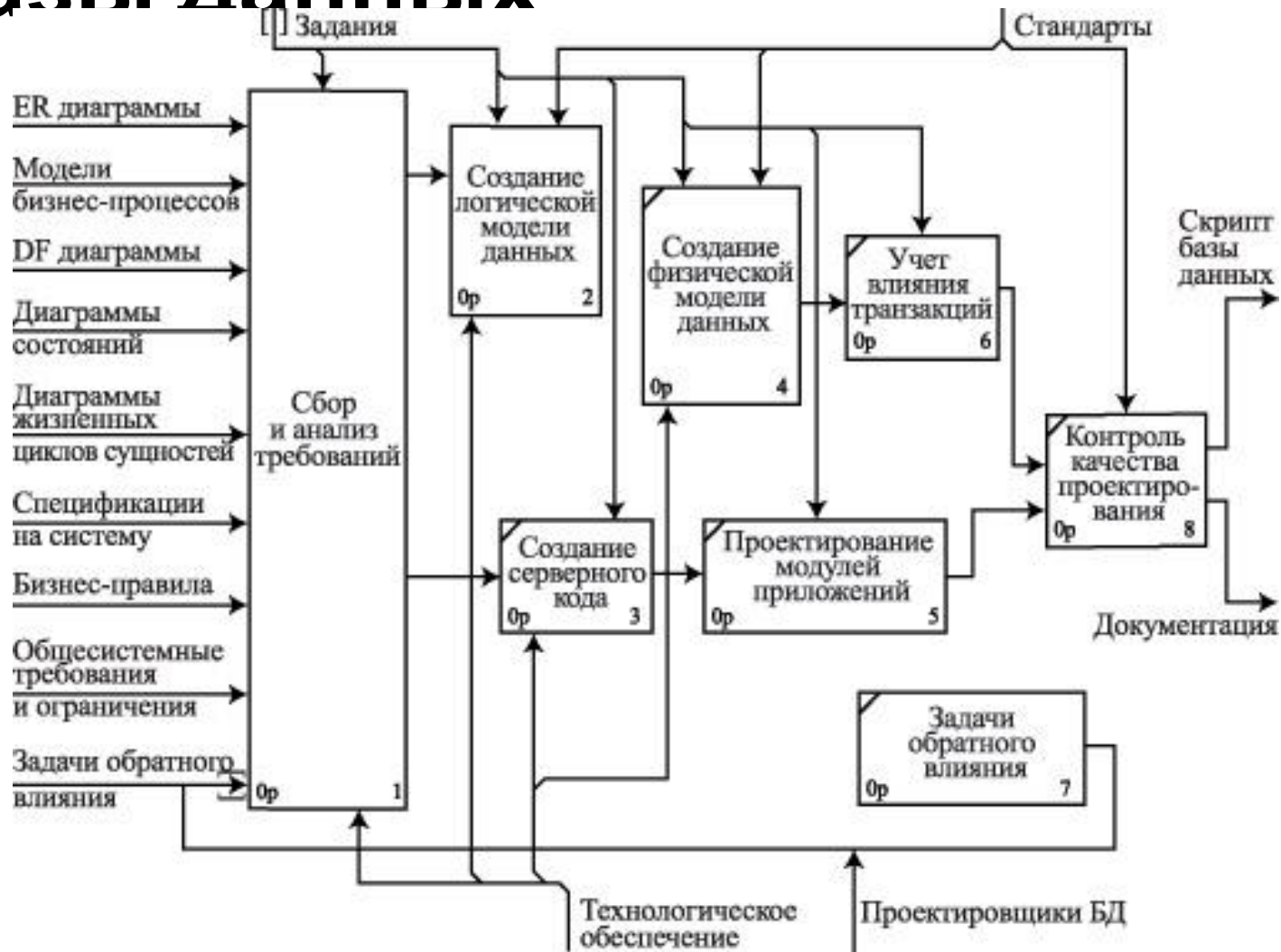
Основные задачи при проектировании баз данных

- Проектирование объектов базы данных (таблицы, представления, индексы, триггеры, хранимые процедуры, функции, пакеты) для представления данных предметной области в базе данных.
- Проектирование интерфейса взаимодействия с базой данных (формы, отчеты и т.д.), т.е. проектирование приложений, которые будут сопровождать данные в базе данных и реализовывать вопросно-ответные отношения на этих данных.
- Проектирование баз данных под конкретную вычислительную среду или информационную технологию (архитектура "клиент-сервер", параллельные архитектуры, распределенная вычислительная среда).
- Проектирование баз данных под назначение системы (интеллектуальный анализ данных, OLAP, OLTP и т.д.).

Особенности баз данных с точки зрения проектирования

- имеет свою внутреннюю архитектуру
- имеет свое собственное лингвистическое содержание
- действует в рамках некоторой внешней среды
- имеет свои средства взаимодействия с внешней средой
- функционирует на конкретной программно-аппаратной платформе
- поддерживается в рамках определенных организационно-технологических мероприятий

Типовая бизнес-модель процесса проектирования базы данных



Основные этапы процесса проектирования

- сбор и анализ входных данных;
- создание логической модели базы данных ;
- создание физической модели базы данных: внутренняя схема;
- создание физической модели базы данных: учет влияния транзакций ;
- создание серверного кода ;
- проектирование модулей приложений базы данных;
- контроль качества проектирования базы данных ;
- задачи обратного влияния.

Входные данные

- *информационная модель предметной области* базы данных: диаграммы "сущность-связь" (ER -диаграммы)
- *функциональная модель предметной области* базы данных: бизнес-модель процессов, диаграммы потока данных (DF -диаграммы), диаграммы состояний, - диаграммы жизненных циклов сущностей, спецификации на системы (требования), бизнес-правила
- общесистемные требования и ограничения;
- задачи обратного влияния

Выходные данные

- физическая модель базы данных, которая может быть преобразована в скрипт для *создания базы данных*
- физическая база данных
- спецификация модулей приложений базы данных
- план тестирования базы данных

1 Этап: Концептуальное проектирование

- *Концептуальный уровень* — центральное управляющее звено, здесь база данных представлена в наиболее общем виде, который объединяет данные, используемые всеми приложениями, работающими с данной базой данных.
- *Концептуальный уровень* отражает обобщенную модель предметной области (объектов реального мира), для которой создавалась база данных. Как любая модель, концептуальная модель отражает только существенные, с точки зрения обработки, особенности объектов реального мира

2 Этап: Разработка внешних моделей

- Уровень внешних моделей — самый верхний уровень, где каждая модель имеет свое "видение" данных.
- Определяет точку зрения на БД отдельных приложений. Каждое приложение видит и обрабатывает только те данные, которые необходимы именно этому приложению.

3 Этап: Физическая реализация

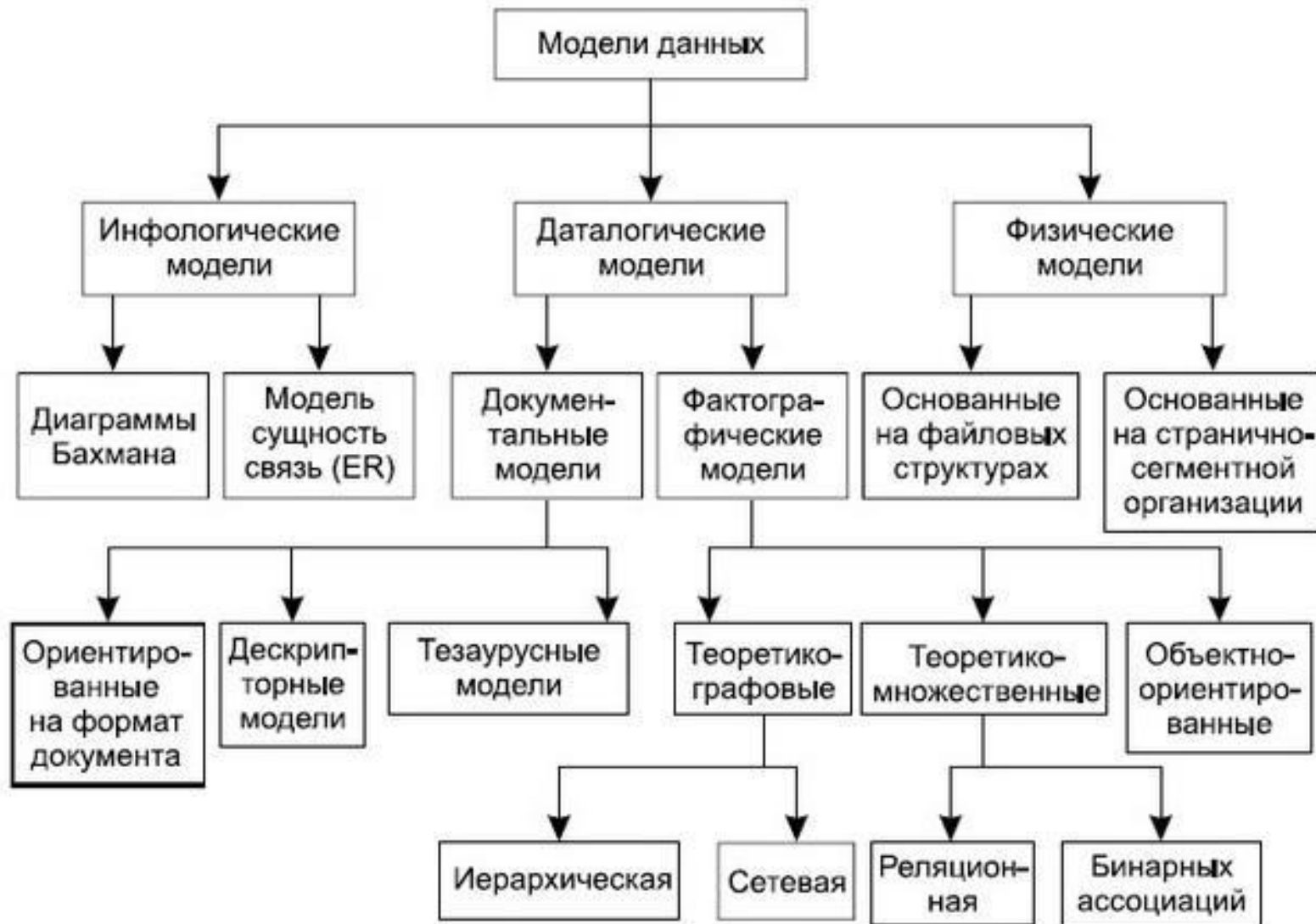
- Физический уровень — собственно данные, расположенные в файлах или в страничных структурах, расположенных на внешних носителях информации

Модели данных

Что такое модель данных?

- *Модель данных* - это некоторая *абстракция*, которая, будучи приложенной к конкретным данным, позволяет пользователям и разработчикам трактовать их уже как информацию, то есть сведения, содержащие не только данные, но и взаимосвязь между ними

Классификация моделей данных



Инфологические модели данных

- Не зависят от используемой СУБД
- Отражают в естественной и удобной форме информационно-логический уровень представления данных предметной области
- Используются на ранних стадиях проектирования

Даталогические модели

- Является основным средством логического проектирования баз данных
- Зависит от выбранного типа СУБД
- Отражает логические взаимосвязи между элементами данных
- Не зависит от физической организации и содержания данных

Физические модели данных

- Оперируют категориями, касающимися организации внешней памяти и структур хранения, используемых в операционной среде
- в качестве физических моделей используются различные методы размещения данных, основанные на файловых структурах: *организация файлов прямого и последовательного доступа, индексных файлов и инвертированных файлов, файлов, использующих различные методы хэширования, взаимосвязанных файлов*
- Физические модели данных, основанные на страничной организации, являются наиболее перспективными

Обзор современных систем управления базами данных

Функции системы управления базой данных (СУБД)

- описание данных на основе какой-либо модели данных (создание пустой структуры данных) , обеспечение избыточности
- манипулирование данными (поиск/отбор, вставка, удаление, замена)
- эффективное хранение данных, обеспечение физической независимости данных
- управление данными в оперативной памяти (буферизация)
- обеспечение многопользовательской работы, разграничение пользователей
- обеспечение целостности, безопасности и восстановления, поддержка транзакций
- предоставление консольных и программных интерфейсов пользователям, в т.ч. к системному каталогу
- утилиты: начальная загрузка, обмен данными с другими ИС, мониторинг и оптимизация БД, средства проектирования

Разделение рынка СУБД*

- **Реляционные СУБД: 93%**

(Oracle Database, IBM DB2 и Microsoft SQL Server занимают до 90% рынка, остальное у MySQL, PostgreSQL и т.д.)

- **СУБД нереляционного типа: 7% (активно развиваются в области социальных сетей, BigData)**

Лидеры рынка РСУБД

Oracle database

- Фактический стандарт для государственных информационных систем в России;
- Лучшие решения по переносимости, кроссплатформенности;
- Наивысшие стандарты надежности;

MS SQL Server

- Второй после Oracle продукт на рынке СУБД;
- Развитые средства анализа данных;
- Поддержка построения корпоративных хранилищ;
- Глубокая интеграция с другими решениями от Microsoft.

IBM DB2

- наиболее высокопроизводительная и мощная реляционная СУБД в мире;
- лучшие по сравнению с конкурентами механизмы масштабируемости;
- единственная реляционная СУБД общего назначения, имеющая реализации на аппаратно-программном уровне;
- собственный диалект языка SQL, направленный на оптимизацию запросов.

Лидеры рынка РСУБД

MySQL

- Одна из наиболее распространенных бесплатных СУБД;
- Принадлежит Oracle;
- Решение для малых и средних приложений;
- Включена в состав большинства «готовых» серверных пакетов WAMP, LAMP, XAMPP, Denwer и т.д.
- API к большинству известных языков программирования;
- Наивысшие стандарты надежности;

PostgreSQL

- Одна из наиболее функциональных бесплатных СУБД;
- Является объектно-реляционной СУБД;
- Поддерживает работу с процедурными, скрипотовыми, классическими и статистическими языками;
- Реализует некоторые принципы объектно-ориентированной концепции;
- Обладает высокой надежностью.

Линтер

- Единственная коммерчески-успешная российская СУБД;
- Поддерживает различные платформы, в том числе операционные системы реального времени;
- Поддерживает геометрические форматы и типы данных (для ГИС).

NoSQL СУБД

- Обозначает ряд подходов, направленных на реализацию хранилищ баз данных, имеющих существенные отличия от моделей, используемых в традиционных реляционных СУБД с доступом к данным средствами языка SQL;
- Направлена на решение проблем масштабируемости и согласованности данных;
- Один из первопроходцев современного NoSQL – компания Google;

Характерными чертами NoSQL-решений являются:

- Применение различных типов хранилищ.
- Возможность разработки базы данных без задания схемы.
- Использование многопроцессорности.
- Линейная масштабируемость (добавление процессоров увеличивает производительность).
- Инновационность: «не только SQL» открывает много возможностей для хранения и обработки данных
- Сокращение времени разработки
- Скорость: даже при небольшом количестве данных конечные пользователи могут оценить снижение времени отклика системы с сотен миллисекунд до

Лидеры рынка NoSQL СУБД

Ключ-Значение (Key-Value)

- Redis - БД в оперативной памяти с дополнительной отказоустойчивостью
- Riak - Распределенное, репликационное хранилище
- Memcached / MemcacheDB - распределённая БД в оперативной памяти

Распределённые СУБД

- Cassandra - структура данных основана на BigTable и DynamoDB
- HBase - хранилище для Apache Hadoop основанное на принципах BigTable

Документо-ориентированные СУБД

- Couchbase - основанное на JSON, совместимое с Memcached хранилище
- CouchDB - передовое документо-ориентированное хранилище
- MongoDB - очень популярное и функциональное хранилище

Графовые базы данных

- OrientDB - очень быстрое документо-ориентированное хранилище гибрида типа граф написанное на Java. Включает в себя разные режимы работы
- Neo4J - безсхемное, очень мощное и популярное хранилище написанное на Java