

Севастопольский государственный университет
Кафедра «Информационные системы»

Управление данными

курс лекций

лектор:
ст. преподаватель кафедры ИС Абрамович А.Ю.



Лекция 1

Концепции систем БД.
Классификация баз данных по типам.
Основные свойства и критерии
оценки баз данных.

Функции системы управления базами
данных (СУБД) и классификация
СУБД

ОСНОВНЫЕ КОНЦЕПЦИИ СИСТЕМ БАЗ ДАННЫХ

Основные черты концепции БД:

- **данные отделяются от прикладной программы (ПП)**, появляется специальная программная надстройка для управления данными, называемая **системой управления базами данных (СУБД)**; СУБД управляет данными и служит посредником между ними и ПП; ПП упрощаются, освобождаются от функций структуризации, хранения и поиска данных;
- появляются **стандартизированные данные о фактографических данных – метаданные**, управляемые СУБД; метаданные описывают информационные параметры и взаимосвязи фактографических данных о ПО; СУБД совместно с метаданными представляет собой **стандартизированное инструментальное средство** для моделирования ПО различной природы;
- происходит **централизация (интеграция) данных**, их многоаспектное использование для различных приложений, что сокращает избыточность данных, позволяет обеспечить более высокий уровень достоверности данных и оптимизировать различные процедуры ведения и использования БД.

ОСНОВНЫЕ КОНЦЕПЦИИ СИСТЕМ БАЗ ДАННЫХ

Использование концепции баз данных позволяет:

- **повысить надежность, целостность и сохранность данных;**
- **сохранить затраты интеллектуального труда;**
- **обеспечить простоту и легкость использования данных;**
- **обеспечить независимость прикладных программ от данных (изменений их описаний и способов хранения);**
- **обеспечить достоверность данных;**
- **обеспечить требуемую скорость доступа к данным;**
- **стандартизовать данные в пределах одной предметной области;**
- **автоматизировать реорганизацию данных;**
- **обеспечить защиту от искажения и уничтожения данных;**
- **сократить дублирование информации за счет структурирования данных;**
- **обеспечить обработку незапланированных запросов к хранимой информации;**
- **создать предпосылки для создания распределенной обработки данных.**

ОСНОВНЫЕ СВОЙСТВА БАЗ ДАННЫХ

Простота
обновления
данных

Какой бы **высокой** ни была **скорость**, это бессмысленно, если **нужно сделать много сложных операций**, чтобы получить, обновить или добавить данные в базу.

Адекватность
отображения
данных

Адекватное отображение данных - представление данных в форме, удобной для человека.

Стандартизация
построения и
эксплуатации БД

Стандартизация обеспечивает преемственность поколений СУБД, упрощает взаимодействие БД одного поколения СУБД с одинаковыми и различными моделями данных.

Совместное
использование
данных

Совместное использование данных - это предоставление общего доступа к БД.

Независимость
данных

Независимость данных - возможность изменения логической и физической структуры БД без изменения представлений пользователей.

Высокое
быстродействие

Время отклика - промежуток времени от момента запроса к БД до фактического получения данных.

Безопасность
данных

Безопасность данных включает их целостность и защиту.

КРИТЕРИИ ОЦЕНКИ КАЧЕСТВА БАЗ ДАННЫХ

При анализе качества баз данных целесообразно рассматривать два компонента: **систему программ управления данными и совокупность данных, упорядоченных по некоторым правилам.**

Различия требований к характеристикам качества привели к созданию широкого спектра локальных, специализированных и распределенных СУБД. **В зависимости от области применения, приоритет при оценке качества может отдаваться различным конструктивным характеристикам: надежности и защищенности применения, удобству использования малоквалифицированными пользователями, эффективности использования ресурсов.**

В системах баз данных **доминирующее значение приобретают сами данные, их хранение и обработка.**

При разработке базы данных в техническом задании и спецификации на нее должен **формализоваться набор функциональных требований к качеству базы данных, адекватный ее назначению и области применения, а также требованиям заказчика и потенциальных пользователей.** Так же как для программных систем, **характеристики качества информации можно разделить на функциональные и конструктивные.** Их номенклатура, содержание и субхарактеристики базируются на ISO 9126.

ФУНКЦИОНАЛЬНАЯ ПРИГОДНОСТЬ ИНФОРМАЦИИ БАЗЫ ДАННЫХ

Мерой качества функциональной пригодности может быть **степень покрытия целей, назначения и функций баз данных доступной пользователям информацией**. Как и для программных систем, для баз данных целесообразно использовать группу субхарактеристик, **определяющих функциональные, структурные и эксплуатационные требования**.

На содержательном уровне функциональную пригодность многих баз данных отражают:

- **полнота накопленных описаний объектов** — относительное число объектов или документов, имеющих в базе данных, к общему числу объектов по данной тематике или по отношению к числу объектов в аналогичных базах данных;
- **идентичность** — относительное число описаний объектов, не содержащих дефекты и ошибки, к общему числу документов об объектах в базах данных;
- **актуальность** — относительное число устаревших данных об объектах в базах данных к общему числу накопленных и обрабатываемых данных.

Требования к информации баз данных также должны содержать особенности обеспечения ее **надежности, эффективности использования ресурсов компьютера, практичности, применимости, сопровождаемости и мобильности**. Содержание и атрибуты этих характеристик несколько отличаются от тех, которые применяются для программ.

КОРРЕКТНОСТЬ ИЛИ ДОСТОВЕРНОСТЬ ДАННЫХ

Корректность или достоверность данных — это степень соответствия данных об объектах в базах данных реальным объектам в данный момент времени, определяющаяся изменениями самих объектов, некорректностями записей об их состоянии или некорректностями расчетов их характеристик. Сюда же можно отнести и **некоторые объемно-временные характеристики сохраняемых и обрабатываемых данных:**

- **объем базы данных** — относительное число записей описаний объектов или документов, доступных для хранения и обработки, по сравнению с полным числом реальных объектов во внешней среде;
- **оперативность** — степень соответствия динамики изменения данных состояниям реальных объектов;
- **глубина ретроспективы** — интервал времени от даты выпуска и/или записи в базу данных самого раннего документа до настоящего времени;
- **динамичность** — относительное число изменяемых описаний объектов к общему числу записей в базе данных за некоторый интервал времени, определяемый периодичностью издания версий базы.

Защищенность информации реализуется средствами СУБД в сочетании с поддерживающими их средствами защиты данных. Цели, назначение и функции защиты тесно связаны с особенностями функциональной пригодности каждой базы данных.

Надежность информации баз данных может основываться на применении понятий и методов теории надежности, которая позволяет получить ряд четких, хорошо измеряемых интегральных показателей. **Надежная база данных, прежде всего, должна обеспечивать низкую вероятность потери работоспособности. Быстрое реагирование на потерю или искажение данных и восстановление их достоверности и работоспособности за время меньшее, чем порог между сбоем и отказом, обеспечивают высокую надежность.**

Используемость ресурсов (или ресурсная экономичность) в стандартах отражается занятостью ресурсов центрального процессора, оперативной, внешней и виртуальной памяти, каналов ввода-вывода, терминалов и каналов связи. В зависимости от конкретных задач и особенностей базы данных при выборе атрибутов качества может доминировать либо величина абсолютной занятости ресурсов различных видов, либо относительная величина использования ресурсов каждого вида при нормальном функционировании базы данных.

Практичность (применимость) — трудно формализуемое понятие, но значительно определяющее функциональную пригодность и полезность применения базы данных для определенных пользователей. В эту группу показателей входят субхарактеристики, **с различных сторон отражающие функциональную понятность, удобство освоения, системную эффективность и простоту использования данных.** Некоторые субхарактеристики можно оценивать экономическими показателями — затратами труда и времени специалистов на реализацию определенных функций взаимодействия с данными.

Понятность зависит от качества документации и субъективных впечатлений потенциальных пользователей. Ее можно описать качественно четкостью функциональной концепции, широтой демонстрационных возможностей, полнотой, комплектностью и наглядностью представления в эксплуатационной документации возможных функций и особенностей реализации данных.

Простота использования — возможность удобно и комфортно эксплуатировать базу данных и манипулировать данными. Она соответствует управляемости, устойчивости к дефектам данных и согласованности с ожиданиями и навыками пользователей.

Изучаемость может определяться трудоемкостью и длительностью подготовки пользователя. Качество изучаемости зависит от внутренних свойств и сложности структуры информации базы данных. Изучаемость может также характеризоваться объемом эксплуатационной документации или объемом и качеством электронных учебников.

Мобильность баз данных, как и программ, можно характеризовать в основном длительностью и трудоемкостью их инсталляции, адаптации при переносе на иные аппаратные и операционные платформы. Часто возможность переноса при первичном формировании и наполнении базы данных не предусматривается и проявляется после длительной эксплуатации. Сложность, трудоемкость и длительность переноса в этом случае значительно возрастают и требуют тщательного планирования и организации работ, приближающихся к созданию новой базы данных.

КЛАССИФИКАЦИЯ БАЗ ДАННЫХ

СУБД может использоваться для создания разных типов БД. **Каждая БД хранит определенную коллекцию данных и используется для определенной цели.** По мере развития технологий, для классификации БД использовались разные методы.



КЛАССИФИКАЦИЯ БАЗ ДАННЫХ

Классификация по количеству пользователей:

Однопользовательская БД – поддерживает только одного пользователя одновременно. Если пользователь А использует БД, пользователи В и С должны ждать, пока пользователь А не завершит работу.

Многопользовательская БД – поддерживает несколько пользователей одновременно. Когда многопользовательская БД поддерживает относительно небольшое количество пользователей (обычно менее 50) или определенный отдел в организации, она называется БД ***рабочей группы***. Когда БД используется всей организацией и поддерживает множество пользователей (более 50, обычно несколько сотен) во многих отделах, она называется ***корпоративной БД***.

Классификация по расположению:

Централизованная БД – поддерживает данные, расположенные на одном сервере.

Распределенная БД – поддерживает данные, распределенные по нескольким различным сервером.

Как централизованные, так и распределенные БД требуют **четко определенной инфраструктуры** (аппаратное обеспечение, операционные системы, сетевые технологии и т.д.) для реализации и эксплуатации БД.

Облачная БД – это БД, которая создается и поддерживается с использованием облачных служб данных.

КЛАССИФИКАЦИЯ БАЗ ДАННЫХ

Классификация по типу хранимых данных:

Универсальные БД – содержат широкий спектр данных, используемых в разных областях, например, БД переписи населения, которая содержит общие демографические данные.

Специальные БД – содержат данные, ориентированные на конкретные предметные области. Данные в БД этого типа используются главным образом для академических или исследовательских целей. Примером специальных БД можно показать БД географических информационных систем (ГИС), в которых хранятся геопространственные данные.

Классификация по времени отклика:

Оперативная БД – предназначена для поддержки повседневных операций компании, также известная как БД оперативной обработки транзакций (online transaction processing – OLTP), транзакционная БД или производственная БД.

Аналитическая БД – ориентирована на хранение исторических данных и бизнеспоказателей, используемых для принятия решений. Аналитические БД позволяют пользователю выполнять расширенный анализ бизнес-данных с использованием сложных инструментов. **Аналитические БД** включают два основных компонента: хранилище данных и инструменты аналитической обработки.

КЛАССИФИКАЦИЯ БАЗ ДАННЫХ

Классификация по степени структурирования данных:

Неструктурированные данные – это данные, которые существуют в исходном (необработанном) состоянии, то есть в том формате, в котором они были собраны. Следовательно, неструктурированные данные существуют в формате, который не поддается обработке, дающей информацию.

Структурированные данные – это результат форматирования неструктурированных данных для облегчения хранения, использования и генерации информации.

Полуструктурированные данные – данные, обработанные в определенной степени. Например, если посмотреть на типичную веб-страницу, данные будут представлены в заранее подготовленном формате для передачи некоторой информации.

Данные типы БД сосредоточены на хранении и управлении высокоструктурированными данными. Но организации также используют полуструктурированные и неструктурированные данные. **Неструктурированные и полуструктурированные потребности хранения и управления данными решаются с помощью нового поколения баз данных, известных как базы данных XML. Расширенный язык разметки (XML) – это специальный язык, используемый для представления и обработки элементов данных в текстовом формате. БД XML поддерживает хранение и управление полуструктурированными XML-данными.**

СИСТЕМА УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ

Система управления базами данных (СУБД) – это набор программ, которые управляют структурой БД и контролируют доступ к данным, хранящимся в БД.

СУБД служит посредником между пользователем и БД. Сама структура БД хранится в виде набора файлов, и единственный способ получить доступ к данным в этих файлах – через СУБД. Она представляет конечному пользователю (или прикладной программе) единое интегрированное представление данных в БД. СУБД получает запросы приложений и переводит их в сложные операции, необходимые для выполнения. Она скрывает большую часть внутренней сложности БД от прикладных программ и пользователей.

Наличие СУБД между приложениями пользователя и БД дает некоторые важные преимущества. Во-первых, **СУБД позволяет совместно использовать данные в БД нескольким приложениям или пользователям.** Во-вторых, **СУБД объединяет представления данных разных пользователей в единый универсальный репозиторий данных.**

ПРЕИМУЩЕСТВА ИСПОЛЬЗОВАНИЯ СУБД

Улучшенный обмен данными. СУБД помогает создать среду, в которой пользователи получают доступ к управляемым данным.

Улучшенная безопасность данных. Чем больше пользователей получают доступ к данным, тем выше риск нарушения безопасности. СУБД обеспечивает соблюдение политики конфиденциальности и безопасности данных. Улучшенная интеграция данных.

Минимизация несогласованности данных. Несогласованность данных существует, когда разные версии одних и тех же данных появляются в разных местах. Вероятность несогласованности данных значительно снижается в правильно спроектированной БД.

Улучшенный доступ к данным. СУБД позволяет быстро получать ответы на специальные запросы.

Улучшенное принятие решений. Улучшенный доступ к управляемым данным позволяют генерировать качественную информацию для принятия правильных решений.

Повышение производительности пользователя. Доступность данных в сочетании с инструментами, которые преобразуют данные в полезную информацию, дает возможность пользователям принимать быстрые, обоснованные решения.

ОСНОВНЫЕ ФУНКЦИИ СУБД

Создание баз данных, изменение, удаление и объединение их по определённым признакам.

Хранение данных, в том числе больших массивов, в структурированном виде и нужном формате

Защита данных от взлома и нежелательных изменений при помощи распределённого доступа: когда разным группам пользователей доступны разный объём и сегменты данных

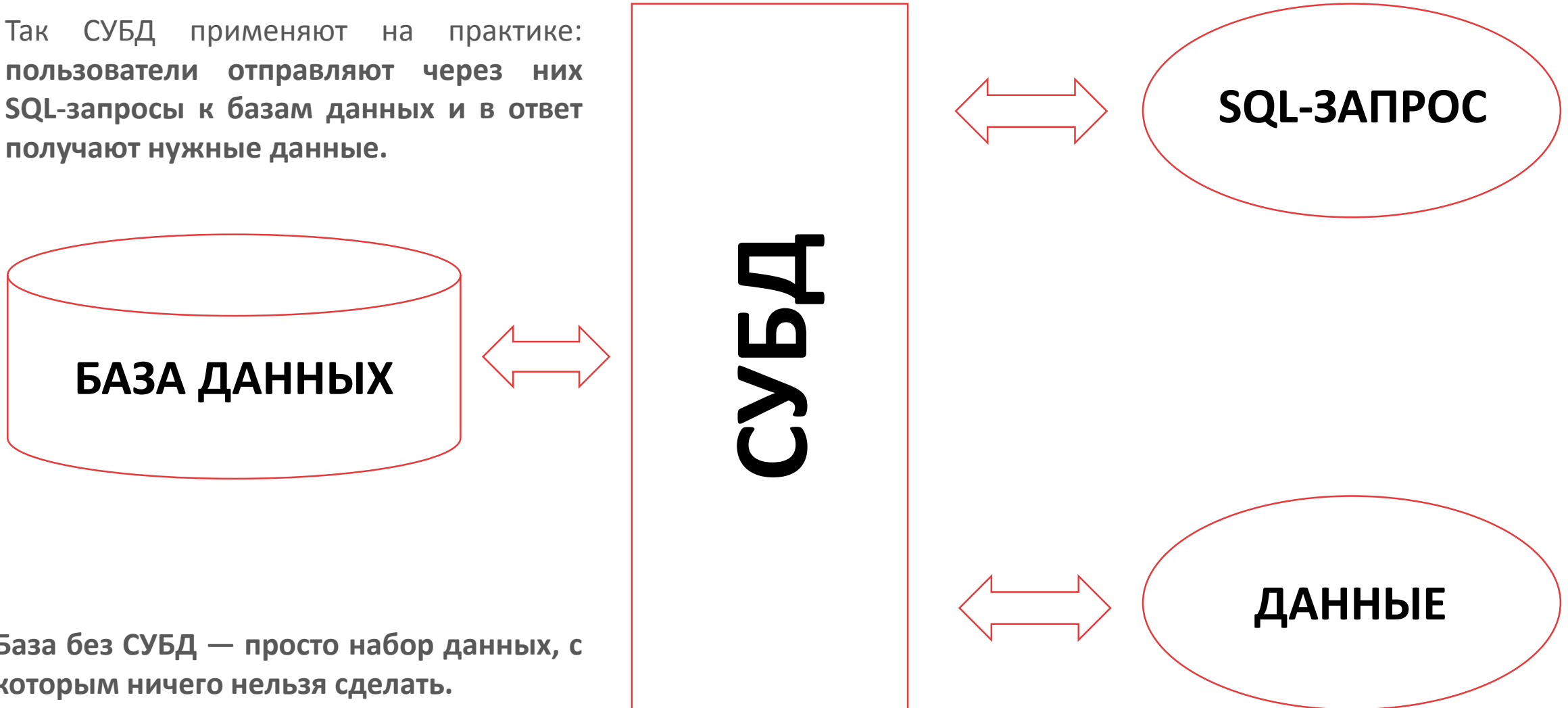
Выгрузка и сортировка данных по заданным фильтрам при помощи SQL-запросов

Поддержка целостности баз данных, резервное копирование и восстановление после сбоев

Основная функция СУБД – это предоставление пользователю БД возможности работы с ней, не вникая в детали на уровне аппаратного обеспечения. То есть все запросы пользователя к БД, добавление и удаление данных, выборки, обновление данных – все это обеспечивает СУБД.

КОМПОНЕНТЫ СУБД

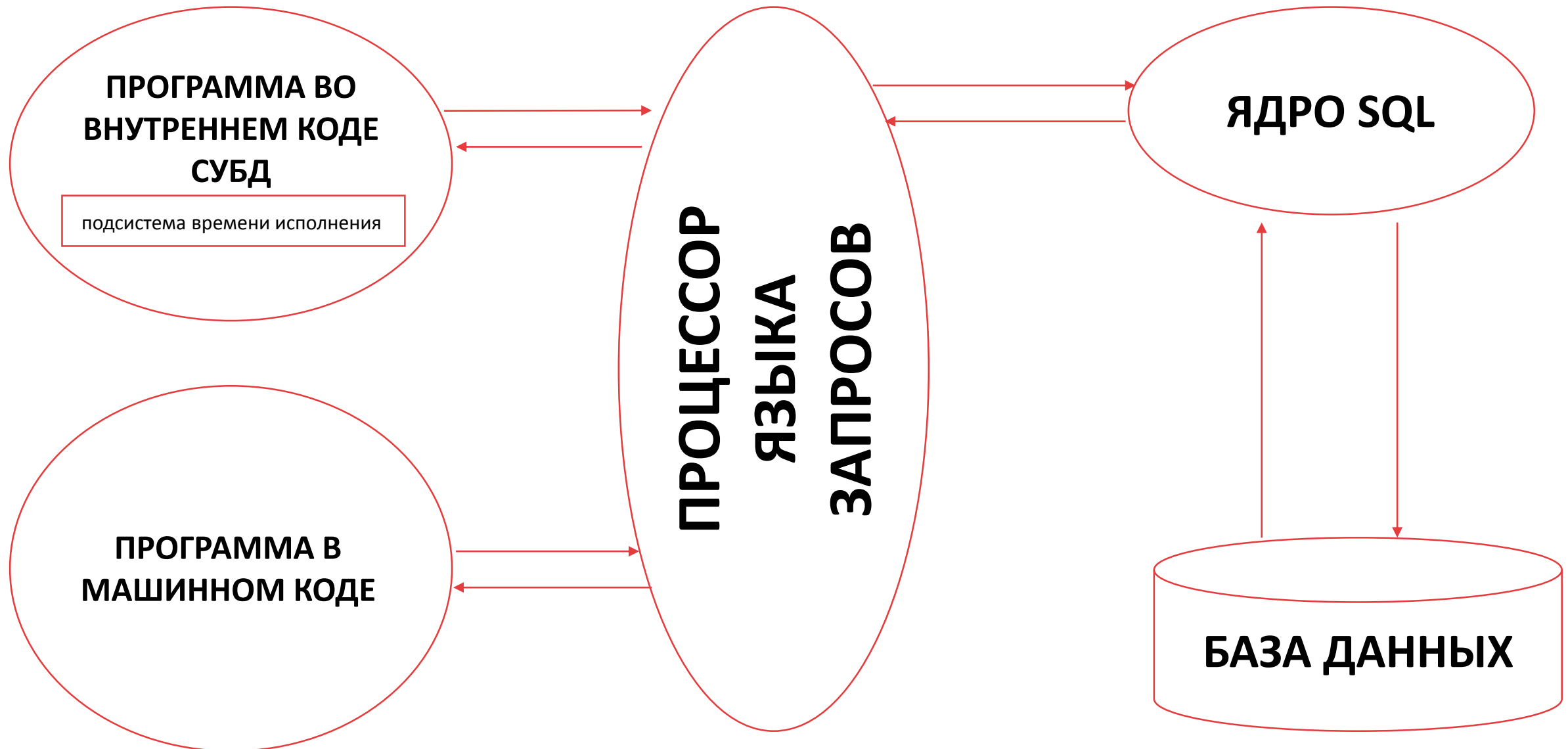
Так СУБД применяют на практике: пользователи отправляют через них SQL-запросы к базам данных и в ответ получают нужные данные.



База без СУБД — просто набор данных, с которым ничего нельзя сделать.

КОМПОНЕНТЫ СУБД

Структура СУБД: когда кто-то отправляет запрос к базе данных, он проходит через специальное ПО, процессор языка запросов и ядро, а потом тот же путь проходят результаты в виде данных



КОМПОНЕНТЫ СУБД

Главные элементы, которые есть в каждой СУБД, и их функции:

Ядро. Это основа всей системы, которая отвечает за хранение и обработку баз данных. В ядре фиксируются все изменения: добавление, удаление или исправление целых баз и отдельных ячеек.

Процессор, или компилятор. Обрабатывает запросы к базам данных на внутренних языках и SQL, преобразуя их в нужные команды и передавая результаты.

Программные средства, или утилиты. С их помощью пользователи вводят запросы, а администраторы баз данных настраивают доступ и другие параметры.

Базы данных. То, где хранятся данные, организованные особым образом, иногда — в зашифрованном виде.