

Севастопольский государственный университет
Кафедра «Информационные системы»

Управление данными

курс лекций

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

СОСТАВ КУРСА

- 😊 **18 ЛЕКЦИЙ**
- 😊 **5 ЛАБОРАТОРНЫХ РАБОТ**
- 😊 **РАСЧЕТНО-ГРАФИЧЕСКАЯ РАБОТА**

ЗАЧЕТ С ОЦЕНКОЙ





Лекция 1

Эволюция методов хранения данных.

Понятие информации, данных, знаний, предметной области, базы и банка данных

ИСТОРИЯ РАЗВИТИЯ БАЗ ДАННЫХ

Цель существования систем хранения информации – обеспечить выдачу достоверной информации в определенное время, определенному лицу, в определенном месте, за определенную плату.

В истории вычислительной техники можно проследить развитие **двух основных областей ее использования:**

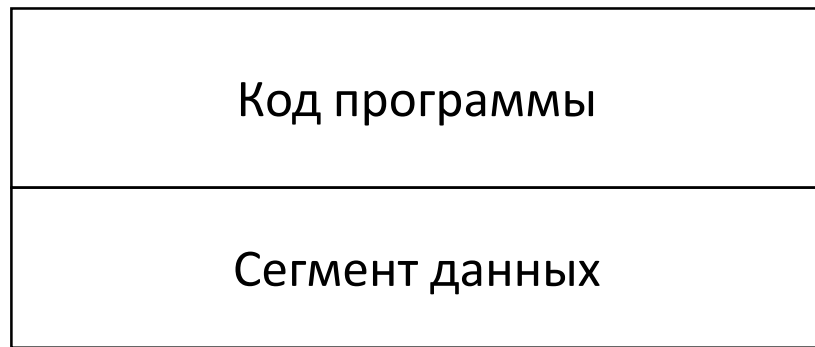
Применение вычислительной техники для выполнения численных расчетов

Использование средств вычислительной техники в автоматизированных ИС

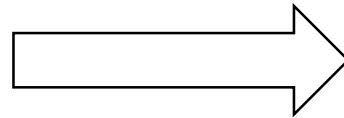
Вторая область использования вычислительной техники **возникла позже первой**, т.к. **возможности компьютеров** по хранению информации были **очень ограниченными**. В первых компьютерах использовались два вида устройств внешней памяти — **магнитные ленты и барабаны**. Емкость магнитных лент была достаточно велика, но они обеспечивали только **последовательный доступ к данным**. Магнитные барабаны давали возможность **произвольного доступа к данным**, но имели ограниченный объем хранимой информации.

В ИС совокупность взаимосвязанных информационных объектов фактически отражает сложную модель объектов реальной предметной области, что требует быстрой реакции системы на запросы пользователей.

ИСТОРИЯ РАЗВИТИЯ БАЗ ДАННЫХ



Расположение данных сразу за кодом программы (в сегменте данных)



Хранение данных в отдельных файлах

Важный шаг в развитии ИС — переход к использованию централизованных **систем управления файлами (СУФ)**. С точки зрения прикладной программы, **файл** — это **именованная область внешней памяти, в которую можно записывать и из которой можно считывать данные**.

От СУФ зависят: правила именования файлов, способ доступа к данным, хранящимся в файле, и структура этих данных (и, возможно, от типа файла). **СУФ берет на себя** распределение внешней памяти, отображение имен файлов в соответствующие адреса во внешней памяти и обеспечение доступа к данным

ИСТОРИЯ РАЗВИТИЯ БАЗ ДАННЫХ

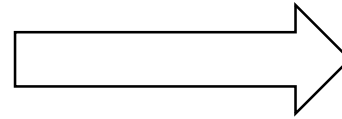
Проблемы использования файлов для хранения данных:

Зависимость программ от данных	<p>При изменении структуры файла требовалось изменять структуру программы, а это требовало новой компиляции, то есть процесса перевода программы в исполняемые машинные коды.</p> <p>Для ИС характерно наличие большого числа пользователей, обращающихся к информации в одних и тех же файлах. Изменение структуры файла, необходимое для одной программы, требовало исправления перекомпиляции и дополнительной отладки всех остальных программ, работающих с этим же файлом.</p>
Децентрализованное администрирование режимов доступа к файлу	<p>СУФ должны обеспечивать авторизацию доступа к файлам.</p> <p>В общем виде подход состоит в том, что по отношению к каждому зарегистрированному пользователю данной вычислительной системы для каждого существующего файла указываются действия, которые разрешены или запрещены данному пользователю. Администрирование режимов доступа к файлу выполняет создатель-владелец.</p>
Низкая эффективность параллельной работы с одним файлом	<p>Если все пользователи собираются только читать файл, ничего страшного не произойдет. Но если хотя бы один из них будет изменять файл, для корректной работы этих пользователей нужна синхронизация их действий по отношению к файлу</p>

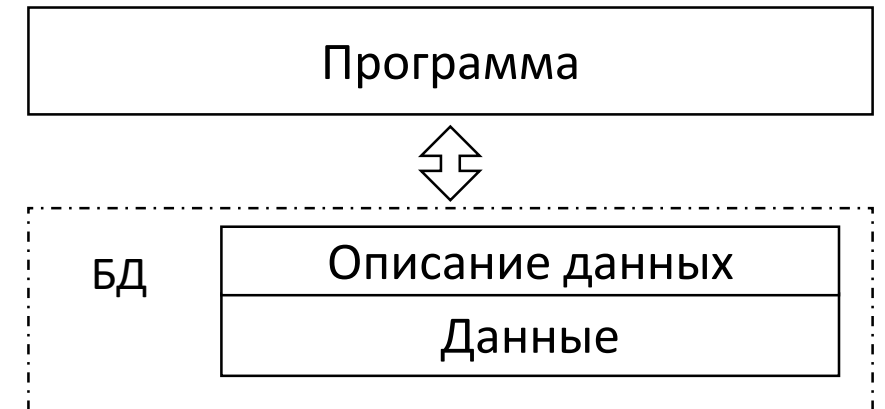
ИСТОРИЯ РАЗВИТИЯ БАЗ ДАННЫХ



Хранение данных в отдельных файлах



Подход был реализован в рамках новых программных систем, названных впоследствии **Системами Управления Базами Данных (СУБД)**, а сами хранилища информации, которые работали под управлением данных систем, назывались **базами данных (БД)**.



Перенос описания данных в массив данных

I этап — БД на больших ЭВМ начало 60-х - начало 70-х гг. Особенности этапа:

- все СУБД базируются на **мощных мультипрограммных операционных системах**, поэтому в основном **поддерживается работа с централизованной базой данных** в режиме **распределенного доступа**;
- используются **языки низкого уровня манипулирования данными**, ориентированные на **навигационные методы доступа к данным**;
- значительная роль отводится **администрированию данных**;
- идет обоснование и формализация **реляционной модели данных**;
- проводятся теоретические работы по **оптимизации запросов и управлению распределенным доступом** к централизованной БД, введено понятие **транзакции**.

ИСТОРИЯ РАЗВИТИЯ БАЗ ДАННЫХ

II этап — Эпоха персональных компьютеров 80-е годы. Особенности этапа:

- все СУБД были рассчитаны на **создание БД в основном с монопольным доступом** (в редких случаях предполагалась последовательная работа нескольких пользователей);
- большинство СУБД имели **развитый и удобный пользовательский интерфейс**. В большинстве существовал интерактивный режим работы с БД как в рамках описания БД, так и в рамках проектирования запросов;
- во всех настольных СУБД поддерживался **только внешний уровень представления реляционной модели**, то есть только внешний табличный вид структур данных;
- при наличии высокоуровневых языков манипулирования данными типа реляционной алгебры и SQL в настольных СУБД **поддерживались низкоуровневые языки манипулирования данными на уровне отдельных строк таблиц**;
- в настольных СУБД **отсутствовали средства поддержки ссылочной и структурной целостности базы данных**;
- наличие монопольного режима работы фактически привело к **вырождению функций администрирования БД** и в связи с этим — к **отсутствию инструментальных средств администрирования БД**;
- **сравнительно скромные требования к аппаратному обеспечению** со стороны настольных СУБД.

ИСТОРИЯ РАЗВИТИЯ БАЗ ДАННЫХ

III этап — Распределенные базы данных 90-е годы. Особенности этапа:

- практически все современные СУБД обеспечивают **поддержку полной реляционной модели** (структурная целостность, языковая целостность, ссылочная целостность);
- большинство современных СУБД **рассчитаны на многоплатформенную архитектуру**, то есть они могут работать на компьютерах с разной архитектурой и под разными операционными системами, при этом для пользователей доступ к данным, управляемым СУБД на разных платформах, практически неразличим;
- появились **развитые средства администрирования БД** на основе общей концепции средств защиты данных (для поддержки многопользовательской работы с БД и возможности децентрализованного хранения данных);
- существуют **средства подключения клиентских приложений**, разработанных с использованием настольных СУБД и **средства экспорта данных** из форматов настольных СУБД;
- разработан **ряд стандартов в рамках языков описания и манипулирования данными и технологий по обмену данными между различными СУБД**, к которым можно отнести и протокол ODBC (Open DataBase Connectivity), предложенный фирмой Microsoft;
- начаты работы, связанные с концепцией **объектно-ориентированных БД — ООБД**.

ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

Информационная система представляет собой систему, реализующую автоматизированный сбор, обработку и манипулирование данными и включающую технические средства обработки данных, программное обеспечение и обслуживающий персонал.

Банк данных – это система специальным образом организованных данных – баз данных, а также технических, программных, языковых и организационно-методических средств, предназначенных для обеспечения централизованного накопления и коллективного многоцелевого использования данных.

Основные компоненты банка данных:

- вычислительная система;
- база данных;
- система управления базой данных, СУБД;
- набор прикладных программ.

Основные функции банка данных:

- хранение данных и их защита;
- изменение хранимых данных;
- поиск и отбор данных по запросам;
- обработка данных и вывод результатов.

База данных (БД) является ядром банка данных и представляет совокупность взаимосвязанных и вместе хранящихся данных из определенной предметной области, организованных специальным образом и хранимых во внешней памяти (файлах базы данных).

ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

В литературе понятие «*база данных (БД)*» *трактуются по-разному*, так как «*число определений БД сравнимо с числом существующих систем управления базами данных (СУБД)*».

База данных (БД) – совокупность данных, организованных по определённым правилам, которые предусматривают общие принципы описания, хранения и манипулирования данными, независимо от прикладных программ.

База данных (БД) – поименованная совокупность взаимосвязанных данных, находящихся под управлением СУБД.

База данных (БД) – именованная совокупность данных, отражающая состояние объектов и их отношений в рассматриваемой предметной области.

В компьютерных базах данных **может содержаться любая информация**: от простого текста (например, фамилия, имя и адрес) до сложной структуры, включая рисунки, звуки и изображения. Хранение данных в заранее известном формате позволяет извлекать данные в желаемом формате благодаря использованию разных методов обработки. **Функционирование базы данных обеспечивает администратор базы данных.**

Администратор базы данных — лицо, отвечающее за выработку требований к базе данных, её проектирование, реализацию, эффективное использование и сопровождение, включая управление учётными записями пользователей БД и защиту от несанкционированного доступа.

ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

Не менее **важной функцией** администратора БД является поддержка *целостности базы данных*.

Целостность БД – свойство БД, означающее, что база данных содержит полную и непротиворечивую информацию, необходимую и достаточную для корректного функционирования приложений.

Данные из одной БД взаимосвязаны и предназначены для одного или нескольких типов приложений и хранятся так, чтобы быть независимыми от использующих их программ.

Приложения – программы, с помощью которых пользователи работают с БД.

Общее управление БД осуществляется специально предназначенной для этого **системой управления БД**, состоящей из языковых, алгоритмических и программных средств.

СУБД – совокупность языковых и программных средств, предназначенных для создания, ведения и использования информации, хранящейся в БД.

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.

ПРЕИМУЩЕСТВА И НЕДОСТАТКИ ТЕХНОЛОГИИ БАЗ ДАННЫХ

Преимущества технологии баз данных:

- контроль за **избыточностью данных** и автоматическая поддержка их целостности;
- **повышение безопасности данных** за счет наличия соответствующих эффективных средств в составе СУБД;
- **применение стандартных форматов данных**, что облегчает взаимодействие с другими системами;
- более экономичное развитие системы за счет объединения технических и финансовых средств, которые ранее распределялись между разными отделами и использовались недостаточно эффективно;
- **повышение доступности данных для конечных пользователей** за счет возможности самостоятельного ввода любых запросов с помощью языка SQL;
- более **простое сопровождение приложений**, работающих с БД, поскольку реализован принцип независимости программ и данных;
- более **эффективное управление параллельным доступом к БД**, поскольку во многих современных СУБД для исключения конфликтов при одновременном обращении к данным предусмотрена блокировка на уровне более мелких структурных элементов (например, не весь файл, а только отдельная строка);
- **наличие развитых средств СУБД** для резервного копирования и восстановления данных в случае нарушения целостности (в файловых системах аналогичные задачи часто возлагаются на самого пользователя).

ПРЕИМУЩЕСТВА И НЕДОСТАТКИ ТЕХНОЛОГИИ БАЗ ДАННЫХ

Недостатки технологии баз данных:

- **сложность программных средств хорошей СУБД**, что обусловлено более широкими функциональными возможностями. Следовательно, возрастают требования к квалификации персонала, т.к. непонимание принципов работы системы может привести к печальным результатам;
- для работы **мощной СУБД** требуется соответствующая компьютерная техника;
- мощная многопользовательская СУБД может оказаться **очень дорогостоящей**. Кроме того, следует учитывать ежегодные затраты на ее сопровождение высококвалифицированными специалистами;
- **более серьезные последствия при выходе системы из строя**, т.к. любая централизация ресурсов делает систему более уязвимой (существует вероятность полного прекращения работы всей организации).

ЖИЗНЕННЫЙ ЦИКЛ БД

Как и любой программный продукт, база данных обладает собственным жизненным циклом (ЖЦБД). Главной составляющей в жизненном цикле БД является создание единой базы данных и программ, необходимых для ее работы.

ЖЦБД ВКЛЮЧАЕТ В СЕБЯ СЛЕДУЮЩИЕ ОСНОВНЫЕ ЭТАПЫ



ПРИНЦИП ЦЕНТРАЛИЗОВАННОГО УПРАВЛЕНИЯ ДАННЫМИ

Основой централизованного управления данными является база данных как набор структурированных и логически взаимосвязанных данных, которые удовлетворяют информационные потребности большого количества пользователей. **Вместо разрозненных файлов с избыточными данными, в которых могут присутствовать противоречия, здесь все данные собраны вместе, причем каждый элемент данных имеется в единственном экземпляре.**

Централизованное управление данными принципиально упрощает контроль и обеспечение безопасности и целостности данных, возможность их восстановления при возникновении аварийных ситуаций с аппаратными и программными средствами, обеспечение одновременной совместной работы с данными большого числа пользователей.



Каким образом унификация структур представления данных в единой базе данных соотносится с необходимостью многоаспектного использования этих данных разными пользовательскими задачами, которые, могут предъявлять свои специфические требования к организации данных?

Единственным конструктивным решением в этом случае является **выполнение еще одного важного принципа**, а именно: **обеспечения взаимной независимости представления данных, хранимых в общей базе данных, и работающих с этой базой прикладных программ, решающих частные пользовательские задачи.**

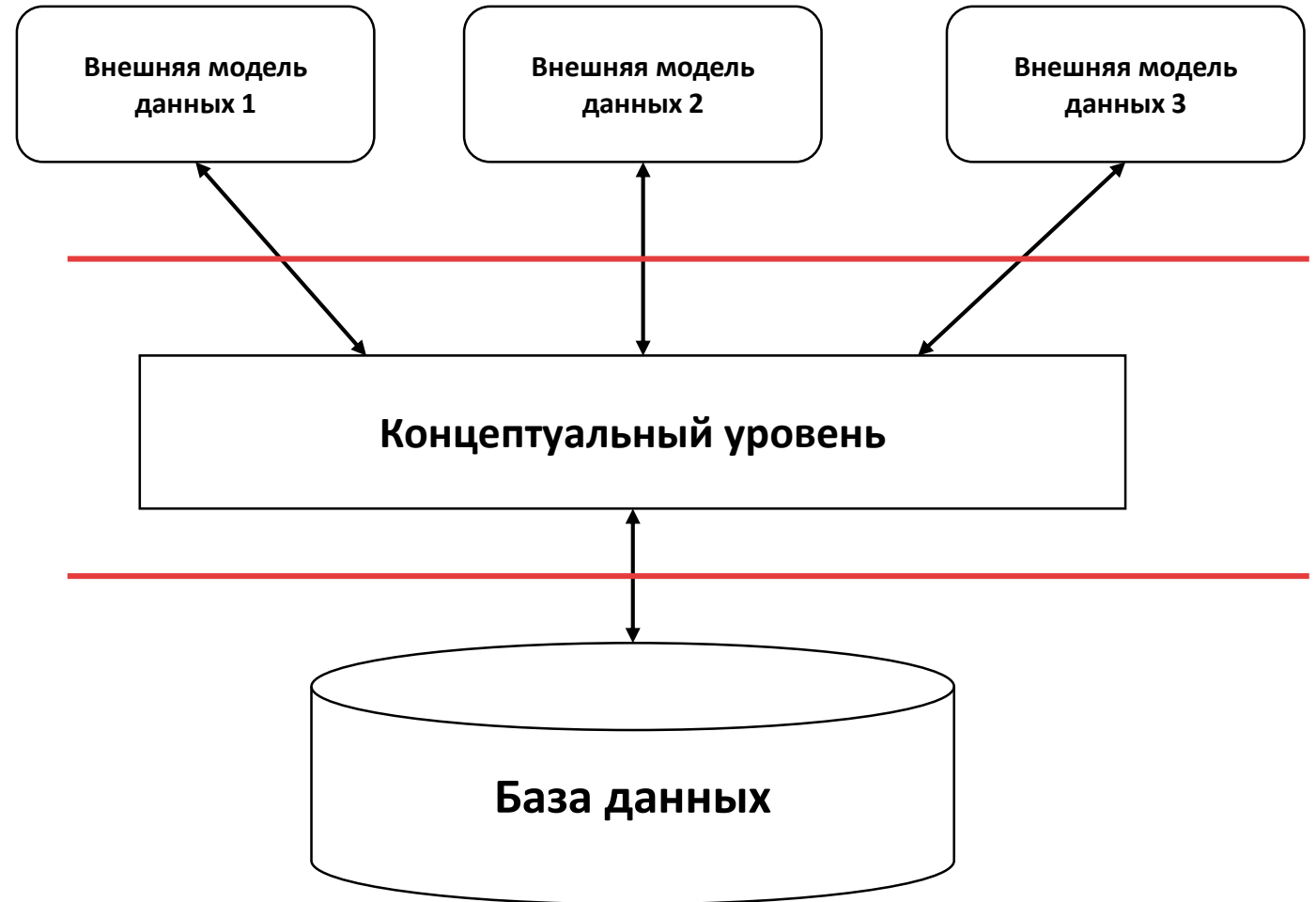
АРХИТЕКТУРА БАЗЫ ДАННЫХ

Трехуровневая система организации БД

Уровень внешних моделей — является самым верхним уровнем или уровнем пользователя. Это совокупность внешних представлений данных, которые обрабатывают приложения и какими их видит пользователь на экране.

Концептуальный уровень — центральное управляющее звено, здесь база данных представлена в наиболее общем виде, объединяет данные. Фактически отражает обобщенную модель предметной области (объектов реального мира), для которой создавалась БД.

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



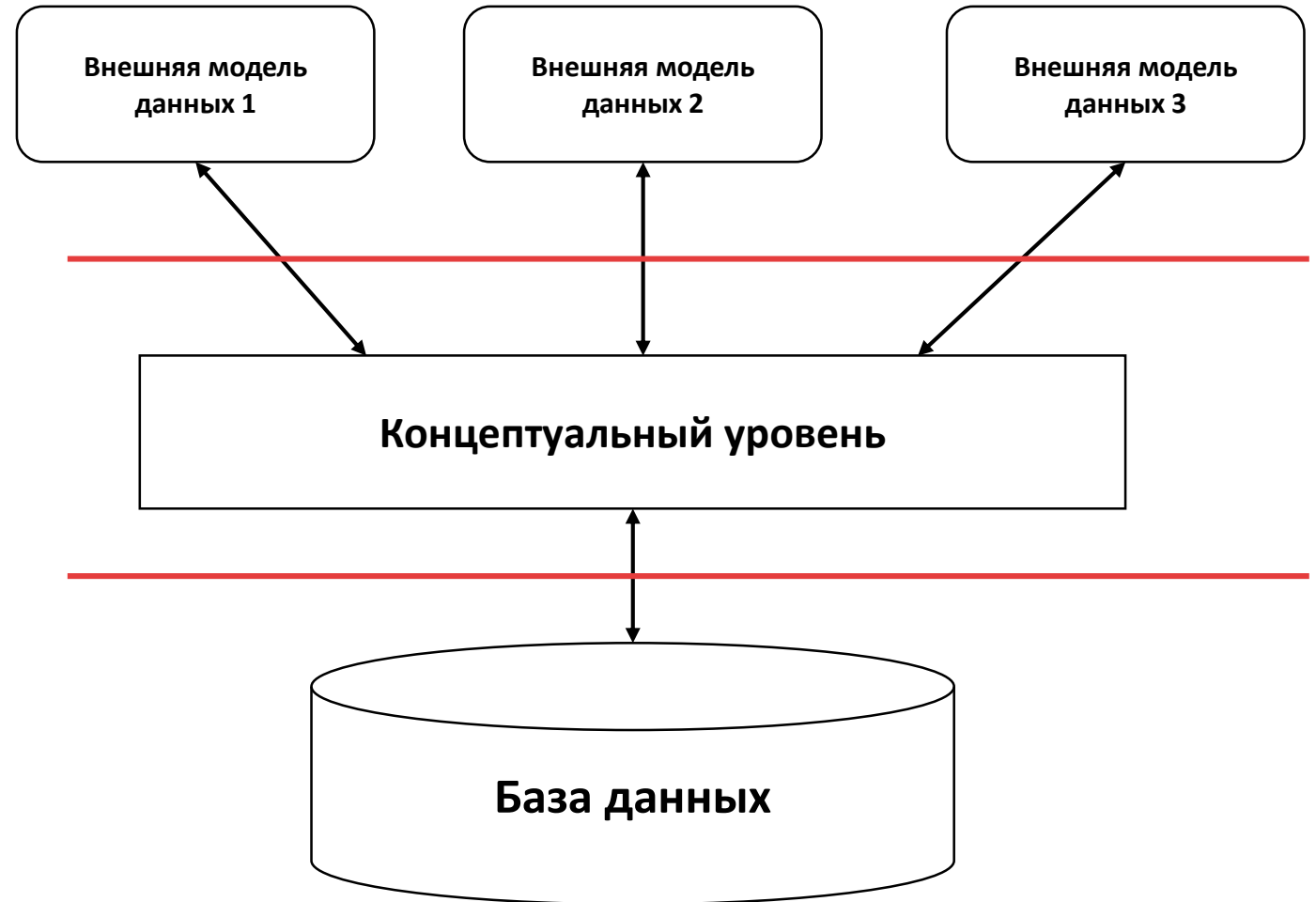
АРХИТЕКТУРА БАЗЫ ДАННЫХ

Трехуровневая система организации БД

Данная архитектура позволяет **обеспечить логическую (между уровнями 1 и 2) и физическую (между уровнями 2 и 3) независимость** при работе с данными.

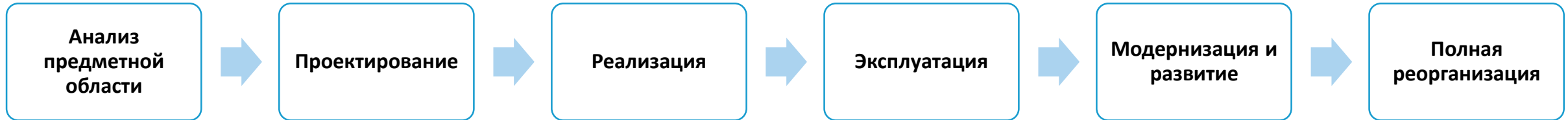
Логическая независимость предполагает возможность изменения одного приложения без корректировки других приложений, работающих с этой же базой данных.

Физическая независимость предполагает возможность переноса хранимой информации с одних носителей на другие при сохранении работоспособности всех приложений, работающих с данной базой данных.



ПОЛЬЗОВАТЕЛИ БАЗ ДАННЫХ

Как любой программно-организационно-технический комплекс, БД существует во времени и в пространстве, имеющая **определенные стадии развития**:



На каждом этапе своего существования с БД **связаны разные категории пользователей**.

Конечные пользователи. Это основная категория пользователей, в интересах которых и создается база данных. Это могут быть случайные пользователи, обращающиеся к БД время от времени за получением некоторой информации, а могут быть регулярные пользователи. Главный принцип – от конечных пользователей не должно требоваться каких-либо специальных знаний в области вычислительной техники и языковых средств.

Администраторы базы данных. Это группа пользователей, которая на начальной стадии разработки базы данных отвечает за его оптимальную организацию с точки зрения одновременной работы множества конечных пользователей, на стадии эксплуатации отвечает за корректность работы данной базы в многопользовательском режиме. На стадии развития и реорганизации эта группа пользователей отвечает за возможность корректной реорганизации базы без изменения или прекращения его текущей эксплуатации.

Разработчики и администраторы приложений. Это группа пользователей, которая функционирует во время проектирования, создания и реорганизации базы данных. Администраторы приложений координируют работу разработчиков при разработке конкретного приложения или группы приложений, объединенных в функциональную подсистему. Разработчики конкретных приложений работают с той частью информации из базы данных, которая требуется для конкретного приложения