

ЛЕКЦИЯ №2. КЛАССИФИКАЦИЯ БД. СУБД И ЕЕ КОМПОНЕНТЫ. ЛОГИЧЕСКОЕ И ФИЗИЧЕСКОЕ ОПИСАНИЕ ДАННЫХ. АРХИТЕКТУРА БАНКА ДАННЫХ.

2.1 Классификация БД.....	1
2.2 Свойства БД	2
2.3 Логическая и физическая независимость данных	3
2.4 Состав и структура СУБД.....	3
2.5 Уровни организации БД.....	5
2.5 Банки данных	6

2.1 Классификация БД

Базы данных можно классифицировать:

1. по модели представления данных
2. по организации хранения данных и обращения к ним
3. по типу хранимой в ней информации

2.1.1 По модели представления данных БД разделяются на:

иерархическая БД – база данных, в которой связь между элементами осуществляется по типу подчинения и схематично изображается в виде дерева. Иерархия начинается с корневого узла. Каждый узел имеет только одного «предка» и N «потомков».

(+) простота и однозначность представления, легкость адресации

(–) существенная зависимость от программно-аппаратных средств

Пример: дерево папок Windows, каталог товаров

сетевая БД, возможно существование любых взаимосвязей между объектами. Если изобразить эту модель графически, то получится набор узлов на плоскости, связанных линиями со стрелками.

(+) теоретически возможны сколь угодно сложные связи между объектами;

(–) сложность реализации, существенная зависимость от программно-аппаратных средств

Пример: служба WWW – документы, произвольно связанные ссылками.

реляционная БД, представление данных в виде системы взаимосвязанных таблиц. Каждый объект системы описывается в виде таблицы с набором свойств (атрибутов), а взаимосвязь между объектами – связями между таблицами.

(+) простота; относительная независимость от программных и аппаратных средств;

(–) существенная зависимость скорости обработки от объема БД

Использование: все существующие СУБД

объектно-ориентированная БД – БД в которой, данные представлены в виде абстрактных объектов, наделённых свойствами, в виде неструктурированных данных, и использующие методы взаимодействия с другими объектами окружающего мира.

По организации хранения данных и обращения к ним БД разделяются на:

1. локальные (персональные) - БД и СУБД находятся на одном компьютере и не может работать по сети;

2. сетевые (интегрированные) - БД и СУБД находится на сервере сети, а приложение на компьютере пользователя посылает запрос и получает результат;
3. распределенные базы данных – БД находятся на множестве компьютеров в сети, а СУБД осуществляет контроль над ними, а приложение на компьютере пользователя посылает запрос к СУБД и получает результат.

2.1.2 По типу хранимой информации БД разделяются на:

1. документальные,
2. фактографические,
3. лексикографические.

Среди документальных БД различают библиографические, реферативные и полнотекстовые.

К лексикографическим БД относятся различные словари (классификаторы, многоязычные словари, словари основ слов и т. п.).

2.2 Свойства БД

БД обладают следующими свойствами:

1. **Избыточность информации** - превышение количества информации, используемой для передачи или хранения сообщения, над его информационной энтропией (полезность).
2. **Логическая независимость данных** - общая логическая структура данных может быть изменена без изменения прикладных программ.
3. **Физическая независимость данных** - физическое расположение и организация данных могут изменяться, не вызывая при этом изменений ни общей логической структуры данных, ни прикладных программ.

Большая часть дисковых или ленточных библиотек, которые существовали до использования средств управления базами данных, содержала большое количество повторяющейся информации. При запоминании многих элементов данных допускалась избыточность, так как на носители информации для различных целей записывались одни и те же данные и, кроме того, хранились различные варианты модификаций одних и тех же данных. База данных представляет возможность в значительной степени избавиться от такой избыточности.

Базу данных иногда определяют как неизбыточную совокупность элементов данных; однако в действительности для уменьшения времени доступа к данным или упрощения способов адресации во многих базах данных избыточность в незначительной степени присутствует. Некоторые записи повторяются для того, чтобы обеспечить возможность восстановления данных при их случайной потере. Чтобы база данных была неизбыточной и удовлетворяла другим требованиям, приходится идти на компромисс. В этом случае говорят об управляемой, или минимальной, избыточности или о том, что хорошо разработанная база данных свободна от излишней избыточности.

Неуправляемая избыточность имеет несколько недостатков:

- хранение нескольких копий данных приводит к дополнительным затратам.
- (что особенно серьезно) приходится выполнять многократные операции обновления для нескольких избыточных копий. Избыточность поэтому обходится значительно дороже в тех случаях, когда при обработке файлов обновляется большое количество информации или, что еще хуже, часто вводятся новые элементы или уничтожаются старые.
- вследствие того, что различные копии данных могут соответствовать различным стадиям обновления, информация, выдаваемая системой, может быть противоречивой.

Если не использовать базы данных, то при обработке большого количества информация появится так много избыточных данных, что фактически станет невозможным сохранять их все на одном и том же уровне обновления. Очень часто пользователи обнаруживают явные противоречия в данных и поэтому испытывают недоверие к полученной от ЭВМ информации. Невозможность хранения избыточных данных на одинаковом уровне обновления является основным препятствием в обработке данных с помощью ЭВМ.

О независимости данных часто говорят как об одном из основных свойств базы данных. Под этим подразумевается независимость данных и использующих их прикладных программ друг от друга в том смысле, что изменение одних не приводит к изменению других. В частности, прикладной программист изолирован от влияния изменений данных и их организации, а также от изменения характеристик физических устройств, на которых они хранятся. В действительности же полностью независимыми данные бывают так же редко, как и полностью неизбыточными. Как мы увидим ниже, независимость данных определяется с различных точек зрения. Сведения, которыми должен располагать программист для доступа к данным, различны для различных баз данных. Тем не менее, независимость данных — это одна из основных причин использования систем управления базами данных.

2.3 Логическая и физическая независимость данных

По мере накопления опыта использования первых систем управления базами данных довольно скоро стало очевидным, что необходим дополнительный уровень независимости данных. Общая логическая структура данных, как правило, сложная, и по мере роста базы данных она неизбежно изменяется. Поэтому важно обеспечить возможность изменения общей логической структуры без изменения при этом использующих ее многочисленных прикладных программ. В некоторых системах изменение общей логической структуры данных составляет форму ее существования, т. е. эта структура находится в состоянии постоянного развития. Поэтому требуются два уровня независимости данных. Мы будем называть их логической и физической независимостью данных.

Логическая независимость данных означает, что общая логическая структура данных может быть изменена без изменения прикладных программ (изменение, конечно, не должно заключаться в удалении из базы данных таких элементов, которые используются прикладными программами).

Физическая независимость данных означает, что физическое расположение и организация данных могут изменяться, не вызывая при этом изменений ни общей логической структуры данных, ни прикладных программ.

Программное обеспечение баз данных будет фактически получать представление данных прикладного программиста из общей логической структуры, а затем будет отображать общую логическую структуру в физическое представление данных.

2.4 Состав и структура СУБД

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

1. Место хранения данных и метаданные
2. Менеджер памяти
3. Менеджер транзакций

ЛЕКЦИЯ №2. КЛАССИФИКАЦИЯ БД. СУБД И ЕЕ КОМПОНЕНТЫ. ЛОГИЧЕСКОЕ И ФИЗИЧЕСКОЕ ОПИСАНИЕ ДАННЫХ. АРХИТЕКТУРА БАНКА ДАННЫХ.

В нижней части схемы – место хранения данных. Принято, что компоненты схем, имеющие форму дисков, обозначают место хранения данных. В данном случае этот компонент содержит не только данные, но и метаданные – информацию о структуре данных. Например, если это реляционные СУБД, метаданные включают в себя имена отношений, имена атрибутов этих отношений и типы данных для этих атрибутов (например, число или строку длиной в 20 символов).

Часто СУБД поддерживает индексы данных. Индекс – это структура данных, помогающая быстро найти элементы данных при наличии части их значения. Например, отношение, хранящее номера счетов и балансы, может иметь индекс счета-номера, что позволяет быстро найти баланс при наличии номера счета. Индексы – часть хранимых данных, а описание, указывающее какие атрибуты имеют индексы, часть метаданных.

Рассмотрим *менеджер памяти*, задача которого получать требуемую информацию из хранилища данных и изменять в нем информацию по требованию выше уровней системы. Следующий компонент называется процессором запроса или менеджером запросов. Его задача – найти лучший способ выполнения требуемой операции и дать соответствующие команды менеджеру памяти. Этот элемент не только обрабатывает запросы, но и запрашивает изменения данных или метаданных.

Компонент менеджер транзакций отвечает за целостность системы. Он должен обеспечить одновременную обработку множества запросов и защиту данных на случай выхода системы из строя. Он взаимодействует с менеджером запросов, т.к. должен знать, на какие данные воздействуют текущие запросы и может отложить определенные запросы или операции. Он взаимодействует с менеджером памяти, т.к. схемы защиты данных обычно включают в себя хранение файла регистрации изменений данных. В верхней части рис. 5.1 находятся три типа обращений СУБД:

1. Запросы – вопросы по поводу данных, которые генерируются двумя способами:
 - а) с помощью общего интерфейса запросов, например, реляционные СУБД позволяют печатать запросы SQL, передаваемые процессору запросов и получать на них ответы.
 - б) с помощью интерфейсов прикладных программ. Типовая СУБД позволяет писать прикладные программы, которые через вызовы СУБД запрашивают БД.
2. Модификации – это операции по изменению данных.
3. Модификации схемы – это команды, которые обычно даются персоналом, администраторами БД, имеющими право изменять схемы БД или создавать новую БД.

В простых системах БД менеджером памяти может быть просто система файлов базовой ОС. Менеджер памяти состоит из двух компонентов:

- менеджера буфера. Он управляет основной памятью.
- менеджера файлов. Он контролирует расположение файлов на диске и получает блок или блоки, содержащие файлы по запросу менеджера буфера.

Задача менеджера запросов – превращать запрос или действие из БД, которые могут быть выражены на очень высоком уровне (например, в виде запроса SQL), в последовательность запросов на хранимые данные типа отдельных кортежей отношения или частей индекса на отношении. Иногда самой трудной частью обработки запроса является его организация – выбор запроса или последовательности запросов к системе памяти, отвечающей на запрос.

СУБД должна гарантировать выполнение определенных операций. Важно, чтобы результат выполнения операций никогда не был утрачен, даже в случае серьезной поломки системы. Типовая СУБД позволяет пользователю сгруппировать несколько

запросов и/или изменений в одной транзакции. **Транзакция** – это группа операций, которые необходимо выполнить последовательно как единое целое.

Часто БД допускает параллельную поддержку множества транзакций («что-то» происходит на нескольких банкоматах одновременно). Гарантировать правильное выполнение всех таких транзакций – задача компонента СУБД, называемого менеджером транзакций. Основные требования, предъявляемые к выполнению транзакций (их называют ACID – свойства):

- атомарность. Требуется, чтобы были выполнены все транзакции или не была выполнена ни одна из них. Например, изъятие денег из банкомата и внесение соответствующего дебета в счет клиента должны быть единственной атомарной транзакцией. Отдельное выполнение одной из этих операций не допускается.
- непротиворечивость. Например, условие непротиворечивости для БД авиационных линий состоит в том, что ни одно из мест в самолете не бронируется для двух клиентов.
- изоляция. При параллельном выполнении двух или более транзакций их результаты должны быть изолированы друг от друга, т.е. одновременное выполнение двух транзакций не должно привести к результату, которого не было бы, если бы они выполнялись последовательно. Например, когда агенты продают билеты на один и тот же рейс и остается единственное свободное место, запрос одного из них должен быть удовлетворен, а запрос другого отвергнут. Недопустимо, чтобы по причине параллельных операций одно место было продано дважды.
- долговременность. Если транзакция завершена, ее результат не должен быть утрачен в результате сбоя системы.

2.5 Уровни организации БД

Американским комитетом по стандартизации ANSI (American National Standards Institute) предложена трехуровневая система организации БД.

1. Уровень внешних моделей — самый верхний уровень, где каждая модель имеет свое "видение" данных. Этот уровень определяет точку зрения на БД отдельных приложений. Каждое приложение видит и обрабатывает только те данные, которые необходимы именно этому приложению. Например, система распределения работ использует сведения о квалификации сотрудника, но ее не интересуют сведения об окладе, домашнем адресе и телефоне сотрудника, и наоборот, именно эти сведения используются в подсистеме отдела кадров.
2. Концептуальный уровень — центральное управляющее звено, здесь база данных представлена в наиболее общем виде, который объединяет данные, используемые всеми приложениями, работающими с данной базой данных. Фактически 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.
3. Физический уровень — собственно данные, расположенные в файлах или в страничных структурах, расположенных на внешних носителях информации.

Эта 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. Это именно то, чего не хватало при использовании файловых систем.

2.5 Банки данных

Банк данных (БНД) является современной формой организации хранения и доступа к информации. Существует множество определений банка данных. В «Общепромышленных руководящих материалах по созданию банков данных» (М.: ГКНТ, 1982) дано следующее определение: «**Банк данных** — это система специальным образом организованных данных (баз данных), программных, технических, языковых, организационно-методических средств, предназначенных для обеспечения централизованного накопления и коллективного многоцелевого использования данных».

В этом определении обозначены и основные отличительные особенности банков данных. Прежде всего это то, что базы данных (БД) создаются обычно не для решения какой-либо одной задачи для одного пользователя, а для многоцелевого использования. БД отражают определенную часть реального мира. Эта информация должна по возможности фиксироваться в базе данных однократно, и все пользователи, которым эта информация нужна, должны иметь возможность работать с ней.

Другой отличительной особенностью банков данных является наличие специальных языковых и программных средств, облегчающих для пользователей выполнение всех операций, связанных с организацией хранения данных, их корректировки и доступа к ним. Такая совокупность языковых и программных средств называется системой управления базой данных (СУБД).

Нельзя сказать, что термин «банк данных» является общепризнанным. В некоторой англоязычной литературе в последнее время используется термин «система баз данных» (database system), который по своему содержанию близок введенному понятию банка данных (система баз данных включает базу данных, систему управления базами данных, соответствующее оборудование и персонал).

2.5.1 Требования к банкам данных.

Особенности «банковской» организации данных позволяют сформулировать основные требования, предъявляемые к БНД:

- адекватность отображения предметной области (полнота, целостность и непротиворечивость данных, актуальность информации, т. е. ее соответствие состоянию объекта на данный момент времени);
- возможность взаимодействия пользователей разных категорий и в разных режимах, обеспечение высокой эффективности доступа для разных приложений;
- дружелюбность интерфейсов и малое время на освоение системы, особенно для конечных пользователей;
- обеспечение секретности и конфиденциальности для некоторой части данных; определение групп пользователей и их полномочий;
- обеспечение взаимной независимости программ и данных;
- обеспечение надежности функционирования БНД; защита данных от случайного и преднамеренного разрушения; возможность быстрого и полного восстановления данных в случае их разрушения; технологичность обработки данных, приемлемые характеристики функционирования БНД (стоимость обработки, время реакции системы на запросы, требуемые машинные ресурсы и др.).

2.5.2 Компоненты банка данных

Информационная компонента. Ядром БнД является база данных. База данных — это поименованная совокупность взаимосвязанных данных, находящихся под управлением СУБД.

Существует множество определений базы данных. Некоторые из них имеют право на существование. Другие устарели и не соответствуют современным представлениям о БнД. Так, в ранних определениях базы данных указывалось на их избыточность, отсутствие дублирования данных в них. На самом деле это не так. В базах данных может наблюдаться избыточность информации. Она может быть вызвана спецификой используемой модели данных, не позволяющей полностью устранить дублирование, или технологическими причинами (обеспечение большей надежности, сокращение времени реакции системы и др.). Но это должна быть управляемая избыточность, причины и цели возникновения которой известны администратору базы данных и управляются как им, так и СУБД.

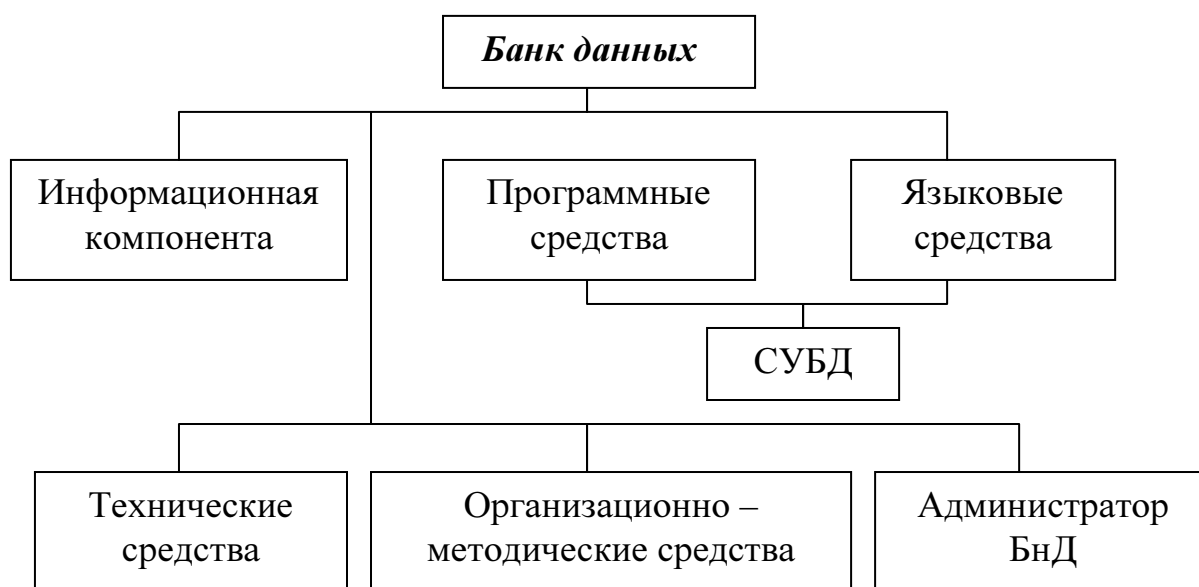


Рисунок 1 Компоненты банка данных

База данных — это объективная форма представления и организации совокупности данных (например, статей, расчетов), систематизированных таким образом, чтобы эти данные могли быть найдены и обработаны с помощью ЭВМ.

Описания баз данных относятся к метainформации, т. е. информации об информации. Описание баз данных часто называют *схемой*. Кроме того, в БнД могут присутствовать описания отдельных частей баз данных с точки зрения отдельных пользователей — *подсхемы*.

Кроме описания баз данных в состав метainформации, хранимой в БнД, может включаться информация о предметной области, необходимая для проектирования системы, о пользователях БнД, о проектных решениях и некоторая другая информация.

Централизованное хранилище метainформации называется *словарем данных*. В литературе используются также термины словарь-справочник, энциклопедия, репозиторий. В некоторых источниках выявляются различия между ними, в других они используются как синонимы. Для данного уровня рассмотрения для нас эти различия несущественны.

Программные средства банка данных. Программные средства БнД представляют собой сложный комплекс, обеспечивающий взаимодействие всех частей информационной системы при ее функционировании (рис. 2).

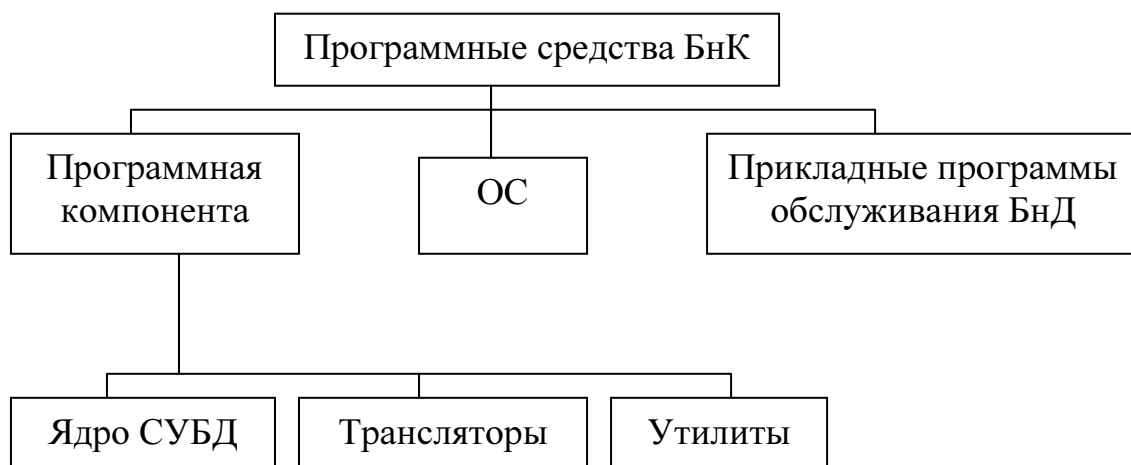


Рисунок 2 Программные средства БнД

Основу программных средств БнД представляет СУБД. В ней можно выделить ядро СУБД, обеспечивающее организацию ввода, обработки и хранения данных, а также другие компоненты, обеспечивающие настройку системы, средства тестирования, утилиты, обеспечивающие выполнение вспомогательных функций, таких, как восстановление баз данных, сбор статистики о функционировании БнД и др. Важной компонентой СУБД являются трансляторы или компиляторы для используемых ею языковых средств.

Подавляющее большинство СУБД работает в среде универсальных операционных систем и взаимодействует с ОС при обработке обращений к БнД. Поэтому можно считать, что ОС также входит в состав БнД.

Для обработки запросов к БД пишутся соответствующие программы, которые представляют прикладное программное обеспечение БнД.

Языковые средства БнД. Языки четвертого поколения создавались по принципу «люди стоят дороже, чем машины». При их проектировании используются следующие принципы.

1. Принцип минимума работы: язык должен обеспечить минимум усилий, чтобы «заставить» машину работать.
2. Принцип минимума мастерства: работа должна быть так проста, как только это возможно; она не должна быть уделом избранных и быть понятной лишь посвященным.
3. Принцип естественности языка, упразднения «инородного» синтаксиса и мнемоники. Язык не должен требовать от пользователей значительных усилий в изучении синтаксиса или содержать много мнемонических или иных обозначений, которые быстро забываются.
4. Принцип минимума времени. Язык должен позволять без существенной задержки реализовывать возникающие потребности в доступе к информации и ее обработке.
5. Принцип минимума ошибок. Технология должна быть спроектирована таким образом, чтобы минимизировать ошибки человека, а уж если они возникли, то по возможности «выловить» их автоматически.

6. Принцип минимума поддержки. Механизм языков четвертого поколения должен позволить легко вносить изменения в имеющиеся приложения.
7. Принцип максимума результата. Языки четвертого поколения предоставляют пользователям мощный инструмент для решения разнообразных задач.

Технические средства БД. В качестве технических средств для банков данных чаще всего используются универсальные ЭВМ, периферийные средства для ввода информации в базу данных и отображения выводимой информации. Если банк данных реализуется в сети, то необходимы соответствующие технические средства для обеспечения ее работы.

До недавнего времени БД реализовывались в основном на больших ЭВМ. В связи со значительным и постоянным улучшением характеристик персональных ЭВМ в настоящее время широко распространено создание банков данных и на машинах этого класса.

Организационно-методические средства. Организационно-методические средства банка данных представляют собой различные инструкции, методические и регламентирующие материалы, предназначенные для пользователей разных категорий, взаимодействующих с банком данных.

Администраторы банка данных. Функционирование БД невозможно без участия специалистов, обеспечивающих создание, функционирование и развитие БД. Такая группа специалистов называется администратором банка данных (АБД). Эта группа специалистов считается составной частью банка данных.

2.5.3 Функции администратор банка данных

Администраторы банка данных выполняют большой круг разнообразных функций. Далее в учебнике мы будем подробно рассматривать некоторые из них. Сейчас же просто перечислим основные из этих функций.

1. Анализ предметной области: описание предметной области, выявление ограничений целостности, определение статуса информации, определение потребностей пользователей, определение статуса пользователей, определение соответствия «данные — пользователь», определение объемно-временных характеристик обработки данных.
2. Проектирование структуры базы данных: определение состава и структуры файлов базы данных, связей между ними, выбор методов упорядочения данных и методов доступа к информации, описание структуры БД на ЯОД.
3. Задание ограничений целостности при описании структуры базы данных и процедур обработки БД: задание ограничений целостности, присущих предметной области, определение ограничений целостности, вызванных структурой базы данных, разработка процедур обеспечения целостности БД при вводе и корректировке данных, обеспечение ограничений целостности при параллельной работе пользователей в многопользовательском режиме.
4. Первоначальная загрузка и ведение базы данных: разработка технологии первоначальной загрузки и ведения (изменения, добавления, удаления записей) БД, проектирование форм ввода, создание программных модулей, подготовка исходных данных, ввод и контроль ввода.
5. Защита данных.
 - а) Обеспечение парольного входа в систему: регистрация пользователей, назначение и изменение паролей.
 - б) Обеспечение защиты конкретных данных: определение прав доступа групп пользователей и отдельных пользователей, определение допустимых операций

над данными для отдельных пользователей, выбор/создание программно-технологических средств защиты данных.

- в) Тестирование средств защиты данных.
 - г) Фиксация попыток несанкционированного доступа к информации.
 - д) Исследование возникающих случаев нарушения защиты данных и проведение мероприятий по их предотвращению.
6. Обеспечение восстановления БД: разработка программно-технологических средств восстановления БД, организация ведения системных журналов.
 7. Анализ обращений пользователей к БД: сбор статистики обращений пользователей к БД, ее хранение и анализ (кто из пользователей, к какой информации, как часто обращался, какие, выполнял операции, время выполнения запросов, анализ причин безуспешных (в том числе и аварийных) обращений к БД).
 8. Анализ эффективности функционирования БнД и развитие системы: анализ показателей функционирования системы (время обработки, объем памяти, стоимостные показатели), реорганизация и реструктуризация баз данных, изменение состава баз данных, развитие программных и технических средств.
 9. Работа с пользователями: сбор информации об изменениях в предметной области, об оценке пользователями работы БнД, определение регламента работы пользователей с БнД, обучение пользователей, консультирование пользователей.
 10. Подготовка и поддержание системных программных средств: сбор и анализ информации о СУБД и ППП, приобретение программных средств, их установка, проверка работоспособности, поддержание системных библиотек, развитие программных средств.
 11. Организационно-методическая работа: выбор или создание методики проектирования БД, определение целей и направлений развития системы, планирование этапов развития БнД, разработка и выпуск организационно-методических материалов.