**Лекция 2. Классификация моделей данных**

Одними из основополагающих в концепции баз данных являются обобщенные категории  
«данные» и «модель данных».  
Понятие «*данные*» в концепции баз данных – это набор конкретных значений, параметров,  
характеризующих объект, условие, ситуацию или любые другие факторы. Примеры данных: ПетровНиколай Степанович, и т. д. Данные не обладают определенной структурой, данные становятсяинформацией тогда, когда пользователь задает им определенную структуру, то есть осознает ихсмысловое содержание. Поэтому центральным понятием в области баз данных является понятиемодели. Не существует однозначного определения этого термина, у разных авторов эта абстракцияопределяется с некоторыми различиями, но, тем не менее, можно выделить нечто общее в этихопределениях.  
*Модель данных —* это некоторая абстракция, которая, будучи приложима к конкретным  
данным, позволяет пользователям и разработчикам трактовать их уже как информацию, то есть  
сведения, содержащие не только данные, но и взаимосвязь между ними.  
На Рисунке 1 представлена классификация моделей данных.  
В соответствии с рассмотренной ранее трехуровневой архитектурой мы сталкиваемся с понятием модели данных по отношению к каждому уровню. И действительно, физическая модель данных оперирует категориями, касающимися организации внешней памяти и структур хранения, используемых в данной операционной среде. В настоящий момент в качестве физических моделей используются различные методы размещения данных, основанные на файловых структурах: это организация файлов прямого и последовательного доступа, индексных файлов и инвертированных файлов, файлов, использующих различные методы хеширования, взаимосвязанных файлов. Кроме того, современные СУБД широко используют страничную организацию данных. Физические модели данных, основанные на страничной организации, являются наиболее перспективными.

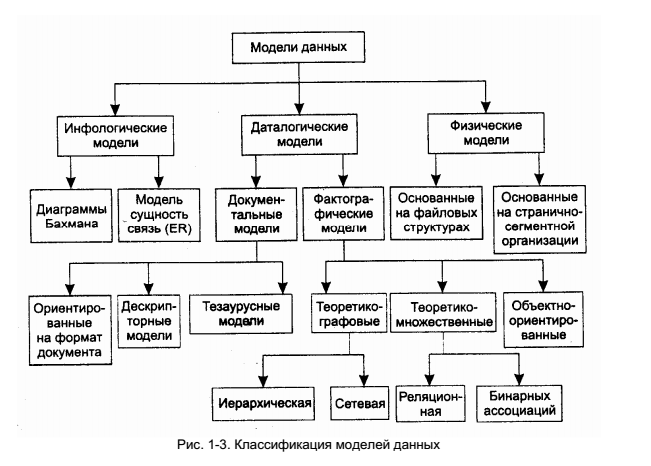


Рисунок 1. Классификация моделей данных.

Наибольший интерес вызывают модели данных, используемые на концептуальном уровне. Поотношению к ним внешние модели называются подсхемами и используют те же абстрактныекатегории, что и концептуальные модели данных.  
Модели концептуального уровня должны выражать информацию о предметной области в виде,независимом от используемой СУБД. Эти модели называются *инфологическими,* или *семантическими,*и отражают в естественной и удобной для разработчиков и других пользователей форме информационно-логический уровень абстрагирования, связанный с фиксацией и описанием объектовпредметной области, их свойств и их взаимосвязей.  
Инфологические модели данных используются на ранних стадиях проектирования для  
описания структур данных в процессе разработки приложения, а *даталогические*модели уже  
поддерживаются конкретной СУБД.  
*Фактографические модели данных* соответствуют представлению информации в виде  
определенных структур данных (дерево, сеть, таблица).  
*Документальные модели данных* соответствуют представлению о слабоструктурированной  
информации, ориентированной в основном на свободные форматы документов, текстов на  
естественном языке.  
Модели, основанные на языках разметки документов, связаны, прежде всего, со стандартным  
общим языком разметки — SGML (StandartGeneralisedMarkupLanguage), который был утвержден ISO в качестве стандарта еще в 80-х годах.

*Тезаурусные* модели основаны на принципе организации словарей, содержат определенные  
языковые конструкции и принципы их взаимодействия в заданной грамматике. Эти модели  
эффективно используются в системах-переводчиках, особенно многоязыковых переводчиках.  
Принцип хранения информации в этих системах и подчиняется тезаурусным моделям,  
*Дескрипторные* модели — самые простые из документальных моделей, они широко  
использовались на ранних стадиях использования документальных баз данных. В этих моделяхкаждому документу соответствовал дескриптор — описатель. Этот дескриптор имел жесткуюструктуру и описывал документ в соответствии с теми характеристиками, которые требуются дляработы с документами в разрабатываемой документальной БД. Например, для БД, содержащейописание патентов, дескриптор содержал название области, к которой относился патент, номерпатента, дату выдачи патента и еще ряд ключевых параметров, которые заполнялись для каждогопатента. Обработка информации в таких базах данных велась исключительно по дескрипторам, тоесть по тем параметрам, которые характеризовали патент, а не по самому тексту патента.

**Жизненный цикл БД**

Под ***жизненным циклом*** базы данных понимаются этапы развития БД, начиная от анализа  
предметной области, и заканчивая эксплуатацией БД.  
Этапы жизненного цикла базы данных изображены на Рис. 0-4. Они аналогичны, в основном,  
развитию любой программной системы, однако в них есть определенная специфика, касающаясятолько баз данных.  
Процесс проектирования БД представляет собой последовательность переходов от неформального словесного описания информационной структуры предметной области к формализованному описанию объектов предметной области в терминах некоторой модели. В общем случае можно выделить следующие этапы проектирования:  
1. Системный анализ и словесное описание информационных объектов предметной области.  
2. Проектирование инфологической модели предметной области – частично формализованное  
описание объектов предметной области в терминах некоторой семантической модели,  
например, в терминах ЕR-модели.  
3. Даталогическое или логическое проектирование БД, то есть описание БД в терминах  
принятой даталогической модели данных.  
4. Физическое проектирование БД, то есть выбор эффективного размещения БД на внешних  
носителях для обеспечения наиболее эффективной работы приложения.



Рисунок 2.Этапы жизненного цикла ИС

Если мы учтем, что между вторым и третьим этапами необходимо принять решение, с  
использованием какой стандартной СУБД будет реализовываться наш проект, то условно процесс проектирования БД можно представить последовательностью выполнения пяти соответствующих этапов (Рисунке 3). Рассмотрим более подробно этапы проектирования БД.



Рисунок 3. Этапы проектирования БД

**Системный анализ предметной области.**

С точки зрения проектирования БД в рамках системного анализа, необходимо осуществить первый этап, то есть провести подробное словесное описание объектов предметной области и реальных связей, которые присутствуют между описываемыми объектами. Желательно, чтобы данное описание позволяло корректно определить все взаимосвязи между объектами предметной области.  
В общем случае существуют два подхода к выбору состава и структуры предметной области:  
***Функциональный подход*** – он реализует принцип движения «от задач» и применяется тогда, когда заранее известны функции некоторой группы лиц и комплексов задач, для обслуживания информационных потребностей которых создается рассматриваемая БД. В этом случае мы можемчетко выделить минимальный необходимый набор объектов предметной области, которые должныбыть описаны.  
***Предметный подход*** – когда информационные потребности будущих пользователей БД  
жестко не фиксируются. Они могут быть многоаспектными и весьма динамичными. Мы не можемточно выделить минимальный набор объектов предметной области, которые необходимо описывать.В описание предметной области в этом случае включаются такие объекты и взаимосвязи, которыенаиболее характерны и наиболее существенны для нее. БД, конструируемая при этом, называетсяпредметной, то есть она может быть использована при решении множества разнообразных, заранеене определенных задач. Конструирование предметной БД в некотором смысле кажется гораздо болеезаманчивым, однако трудность всеобщего охвата предметной области с невозможностьюконкретизации потребностей пользователей может привести к избыточно сложной схеме БД, котораядля конкретных задач будет неэффективной.  
Чаще всего на практике рекомендуется использовать некоторый компромиссный вариант,  
который, с одной стороны, ориентирован на конкретные задачи или функциональные потребностипользователей, а с другой стороны, учитывает возможность наращивания новых приложений.  
Системный анализ должен заканчиваться подробным описанием информации об объектах  
предметной области, которая требуется для решения конкретных задач и которая должна храниться вБД, формулировкой конкретных задач, которые будут решаться с использованием данной БД скратким описанием алгоритмов их решения, описанием выходных документов, которые должныгенерироваться в системе, описанием входных документов, которые служат основанием длязаполнения данными БД.