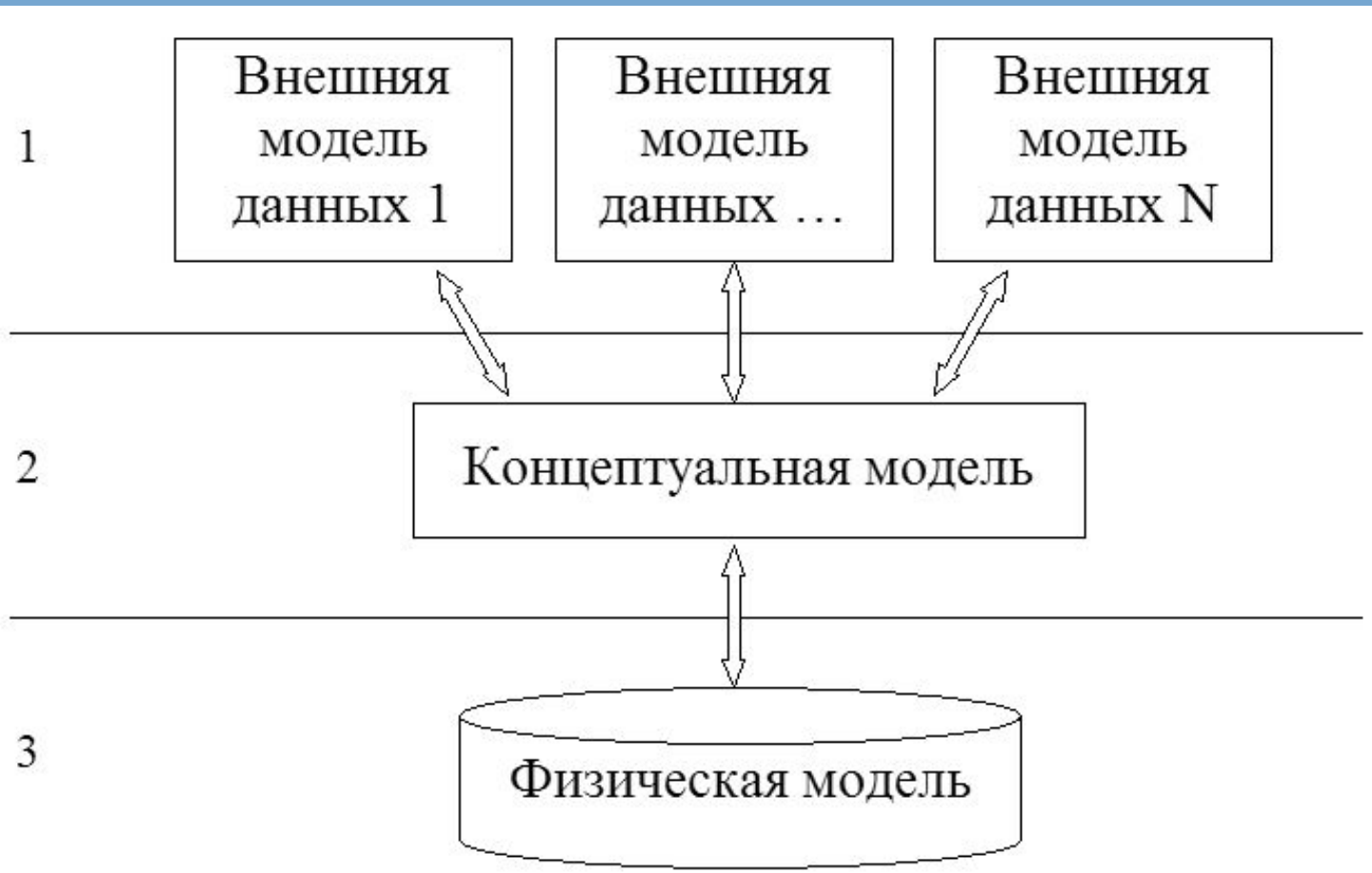
Модели данных.

Введение

* Информация об определенной предметной области представлена в базе данных моделями нескольких уровней. По числу уровней в архитектуре различают одноуровневые, двухуровневые, трехуровневые системы. На различных уровнях архитектуры СУБД поддерживается разный уровень абстракции данных.

* В настоящее время наиболее распространенной является предложенная американским комитетом по стандартизации ANSI (American National Standards Institute) трехуровневая система организации БД. При проектировании баз данных выделяют три уровня: концептуальный, внутренний (физический) и внешний.

Уровни представления данных



Уровень внешних моделей

* 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 ему. Например, система распределения работ использует сведения о квалификации сотрудника, но ее не интересуют сведения об окладе, домашнем адресе и телефоне, и наоборот, именно эти сведения используются в подсистеме отдела кадров.

Концептуальный уровень

* здесь БД представлена в наиболее общем виде, который объединяет данные, используемые всеми приложениями, работающими с данной базой данных. Фактически, концептуальный уровень отражает обобщенную модель предметной области, для которой создавалась база данных.

Концептуальный уровень (2)

* Как любая модель, она отражает только существенные, с точки зрения обработки, особенности объектов предметной области. Этот уровень не зависит от особенностей используемой СУБД.
* Часто на этом уровне строится 2 модели БД – концептуальная и логическая. Соответственно выполняется концептуальное и логическое проектирование БД.

Физический уровень

* собственно данные, расположенные в файлах или в страничных структурах, расположенных на внешних носителях информации. Физическое представление БД относится к внутреннему уровню. Он описывает способы организации данных на внешних носителях информации и предназначен для достижения оптимальной производительности и эффективности использования ресурсов вычислительной системы. Описание физической структуры БД называется схемой хранения, а соответствующий этап проектирования БД – физическим проектированием.

* В этом семестре вы будете выполнять логическое и концептуальное проектирование БД – т.е. строить концептуальный уровень представления БД.
* В следующем семестре будете выполнять физ моделирование БД – внутренний уровень представления БД.

Данные

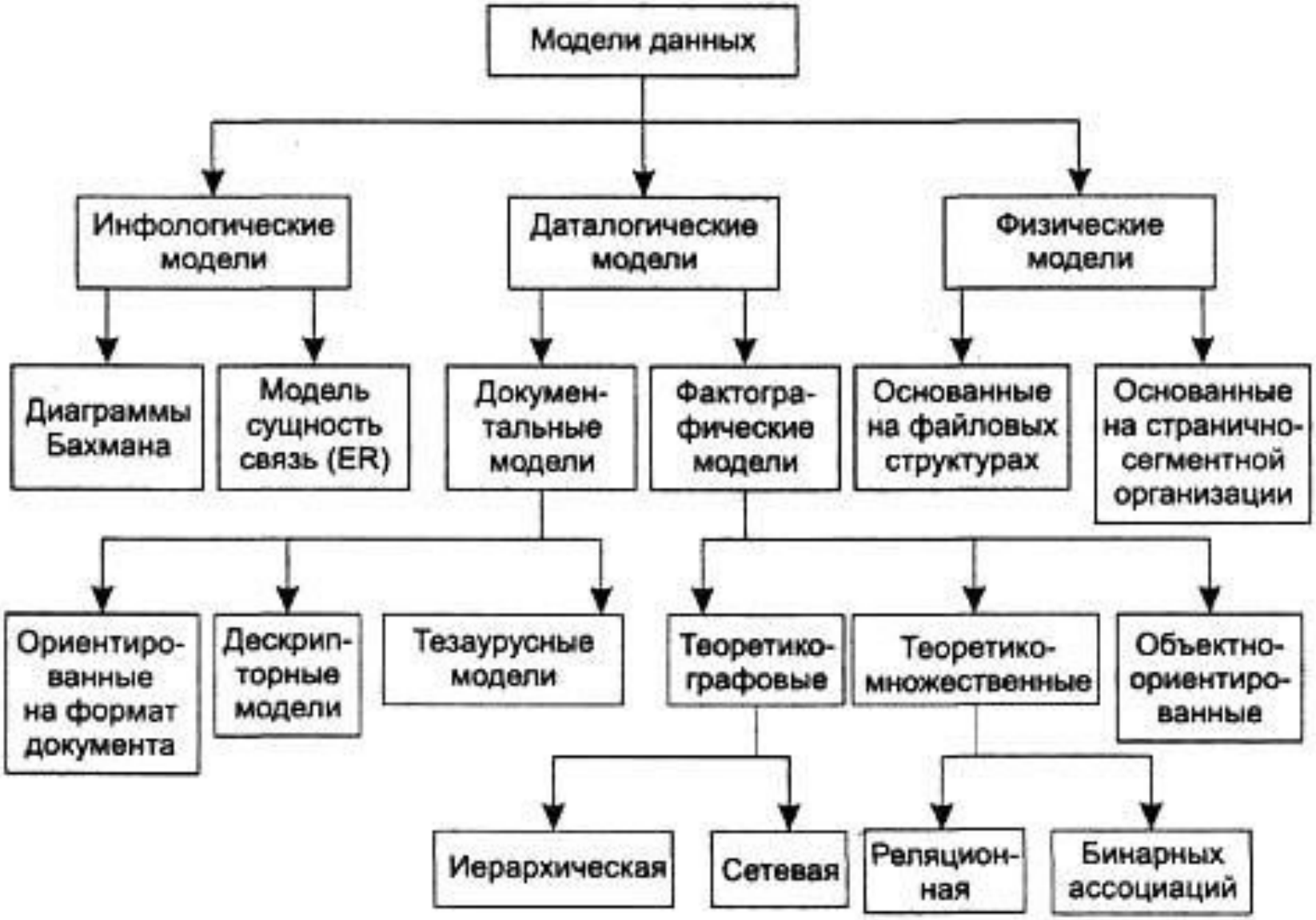
* Понятие «данные» в концепции баз данных — это набор конкретных значений, параметров, характеризующих объект, условие, ситуацию или любые другие факторы.
* Примеры данных: Петров Николай Степанович, 1998 год, $30, ул. Ленина и т. д.

Информация

* Данные не обладают определенной структурой, данные становятся информацией тогда, когда пользователь задает им определенную структуру, то есть осознает их смысловое содержание.
* Центральным понятием в области баз данных является понятие модели. Не существует однозначного определения этого термина, у разных авторов эта абстракция определяется с некоторыми различиями, но тем не менее можно выделить нечто общее в этих определениях.

Модель

* 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.
* Рассмотрим классификацию моделей данных.


Модели данных
Инфологические модели
Даталогические модели
Физические модели
Диаграммы Бахмана
Модель сущность связь (ER)
Докумен-тальные модели
Фактогра-фические модели
Основанные на файловых структурах
Основанные на странично-сегментной организации
Ориентиро-ванные на формат документа
Дескрип-торные модели
Тезаурусные модели
Теоретико-графовые
Теоретико-множественные
Объектно-ориентиро-ванные
Иерархическая
Сетевая
Реляцион-ная
Бинарных ассоциаций

* Сами модели мы рассмотрим на следующей паре – завтра. Сейчас насчет литературы.
* А сколько вас вообще человек Как вы по другим предметам ходите?
* Я поговорю с Феликсом Владимировичем. Но так можно вас объединить – чтобы вы ходили по четвергам. А они, если что, по средам.

* По концептуальному и логическому проектированию такого хорошего источника литературы.
* Вторую лекцию по моделям скину завтра, так как её еще буду доделывать.
* Вопросы есть какие-нибудь?
* Тема у нас одна и та же. Курсовые по ИТ и по УД строятся по одной предметной области. Актуальность, описание предметной области надо будет делать вначале – это общие разделы в обоих курсовых.

* А проектирование – сроки определим, когда пройдем. Потому что сейчас сначала пойдут темы, которые являются скорее теоретическими. Так как я уже писала, что построение модели БД пойдет во второй половине семестра – после того, как вы пройдете построение моделей по ИТ. То, что вы будете изучать сейчас по ИТ – является отправной точкой для построения моделей по УД. Поэтому и защиту мы принимаем вместе и эти модели очень взаимосвязаны.

Инфологические модели

* используются на ранних стадиях проектирования баз данных для формального описания предметной области. Они содержат информацию о классах объектов, их свойствах и взаимосвязях, описания структур данных без привязки к какой-либо конкретной СУБД. Инфологические (или семантические) модели отражают в естественной и удобной для разработчиков и других пользователей форме информацию о предметной области в процессе разработки структуры будущей базы данных.

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: организация файлов прямого и последовательного доступа, индексных файлов и инвертированных списков.

Физическая модель

* Кроме того, современные СУБД широко используют страничную организацию данных. В этом случае база данных представлена минимальным количеством файлов, а задачи поиска, чтения и записи данных выполняет сама СУБД, а не операционная система. Физические модели данных, основанные на страничной организации, сейчас чаще всего используются.

Даталогические модели

* являются моделями концептуального и внешнего уровня.
* Модели концептуального и внешнего уровня строятся по одним и тем же правилам, но отображают разные данные. Модели внешнего уровня являются локальными концептуальными моделями, т.е. предназначены для одной задачи.

Документальные модели

* данных соответствуют представлению о слабоструктурированной информации, ориентированной в основном на свободные форматы документов, текстов на естественном языке.
* **Модели, ориентированные на формат документов**, связаны прежде всего со стандартным общим языком разметки — SGML (Standart Generalised Markup Language), который был утвержден ISO в качестве стандарта еще в 80-х годах. Этот язык предназначен для создания других языков разметки, он определяет допустимый набор тегов (ссылок), их атрибуты и внутреннюю структуру документа.
* По этим правилам образованы HTML и XML.

Дескрипторные модели

* самые простые из документальных моделей, они широко использовались на ранних стадиях использования документальных баз данных. В этих моделях каждому документу соответствовал дескриптор — описатель. Этот дескриптор имеет жесткую структуру и описывает документ в соответствии с теми характеристиками, которые требуются для работы с документами в разрабатываемой БД. Например, для БД, содержащей описание патентов, дескриптор содержит название области, к которой относился патент, номер, дату выдачи и еще ряд ключевых параметров, которые заполнялись для каждого патента. Обработка информации в таких базах данных ведется исключительно по дескрипторам, то есть по тем параметрам, которые характеризуют патент, а не по самому тексту патента.

Тезаурусные модели

* основаны на принципе организации словарей. Они содержат определенные языковые конструкции и принципы их взаимодействия в заданной грамматике. Эти модели эффективно используются в системах-переводчиках, особенно многоязыковых. Принцип хранения информации в этих системах и подчиняется тезаурусным моделям.

Теоретико-графовые модели

* отражают совокупность объектов реального мира в виде графа взаимосвязанных информационных объектов. Математической основой таких моделей является теория графов.

Теоретико-множественные модели

* Были созданы позже теоретико-графовых моделей. В них данные представляются в виде взаимосвязанных множеств.
* Математической основой таких моделей является теория множеств.

Развитие

* Эти модели (сетевая и иерархическая) использовались до создания реляционной модели данных Эдгаром Коддом в 70-е годы XX века.
* Реляционная модель была проще для понимания и использования, чем графовые модели.
* Это была первая модель данных, для работы с которой был создан полностью математически обоснованный аппарат – реляционная алгебра.

Объектно-ориентированные модели

* Но несмотря на простоту, реляционные модели не отражали реальный мир, таким какой он есть.
* Поэтому для решения более сложных задач, этой модели стало уже недостаточно. И была создана объектно-ориентированная модель данных, основывающаяся на принципах объектно-ориентированного программирования.

Настоящее время

* Для большинства приложений в настоящее время используются реляционные модели данных.
* Но современные СУБД включают возможности объектно-ориентированного программирования, т.е. создания объектно-реляционных БД.

* Поэтому мы будем изучать реляционные модели данных.

Реляционные СУБД (РСУБД)

* Microsoft SQL Server,
* Microsoft Access,
* MySQL.

Объектно-реляционные СУБД (ОРСУБД)

* Oracle Database,
* Informix,
* DB2,
* PostgreSQL,
* FirstSQL/J