ТЕМА 4.

ОБЩИЕ ПРИНЦИПЫ ПОСТРОЕНИЯ МОДЕЛЕЙ ДАННЫХ В ГИС

- Гис использует разнообразные данные об объектах, характеристиках земной поверхности, информацию о формах и связях между объектами, различные описательные сведения.
- В существующих ГИС используются различные способы для организации реальности посредством моделей данных. Каждая модель более пригодна для определенных типов данных и областей применения, поэтому при необходимости решения большого числа задач следует использовать разные модели.
- В процессе функционирования ГИС все многообразие входных данных – информация об объектах, их характеристиках, о формах и связях между объектами, различные описательные сведения – преобразуется в единую общую модель (набор моделей), хранимую в базе данных.

- Интегрированная информационная основа базы данных не является просто суммой информационных моделей частей объекта, хотя включает данные о связях и дополнительную служебную информацию.
- Целостность, непротиворечивость и оптимальность этой общей модели ГИС обусловливается обоснованным выбором составляющих частей модели.
- Модели объектов ГИС, хранящиеся в базах данных, состоят из более простых частей, которые принято называть моделями данных. В свою очередь, модели данных в ГИС имеют сложную многоуровневую структуру, в которой нижние уровни состоят из элементарных (атомарных) моделей данных.

- Из элементарных моделей конструируются более сложные. Конструирование, или проектирование сложных моделей на основе более простых, зависит от выбора структуры сложной модели, от типа связей в сложной модели и от качественных характеристик элементарных моделей.
- Проблема организации базы данных в ГИС сводится к решению ряда задач, первой из которых является организация моделей объектов. Это определяет необходимость предварительного анализа свойств элементарных моделей данных, составляющих более сложные модели в БД, и выбора базовых теоретических моделей с учетом конкретной предметной области задач ГИС. Такой подход позволяет оптимизировать создание информационной основы и процессы обработки данных в БД.

4.1. Основные понятия моделей данных

- Реальная ГИС используется для решения заданного круга задач в конкретной области применения.
- ***Предметной областью*** называется подмножество (часть реального мира), на котором определяется набор данных и методы манипулирования с ними для решения конкретной задачи или исследований.
- Для построения модели объекта в виде составляющих частей и определение связей между этими частями применяют методы (процедуры) абстракции, которые тоже образуют целый ряд понятий:
- ***абстракция*** – процедура структуризации (типизации) данных. Различают два вида абстракции: обобщение и агрегация.

- Обобщение в свою очередь подразделяется на две категории: собственно обобщение и классификация.
- ***собственно обобщение*** – процедура соотнесения множества типов одному типу соотносится с понятием: «есть часть…»;
- ***классификация*** – процедура соотнесения множества знаков одному типу.
- ***Агрегация*** – процедура конструирования объекта из других базовых объектов; соотносится с понятием «есть некоторые…».
- ***Под агрегатными данными*** будем понимать набор данных для формирования объекта из его частей на основе процедур агрегации.
- Процедура, обратная агрегации, называется ***пошаговой детализацией***. Она применяется для разбиения агрегатной модели на составные части.

Рис. 4.1. Схема структурирования данных с применением процедур абстракции: а) прямые процедуры; б) обратные процедуры

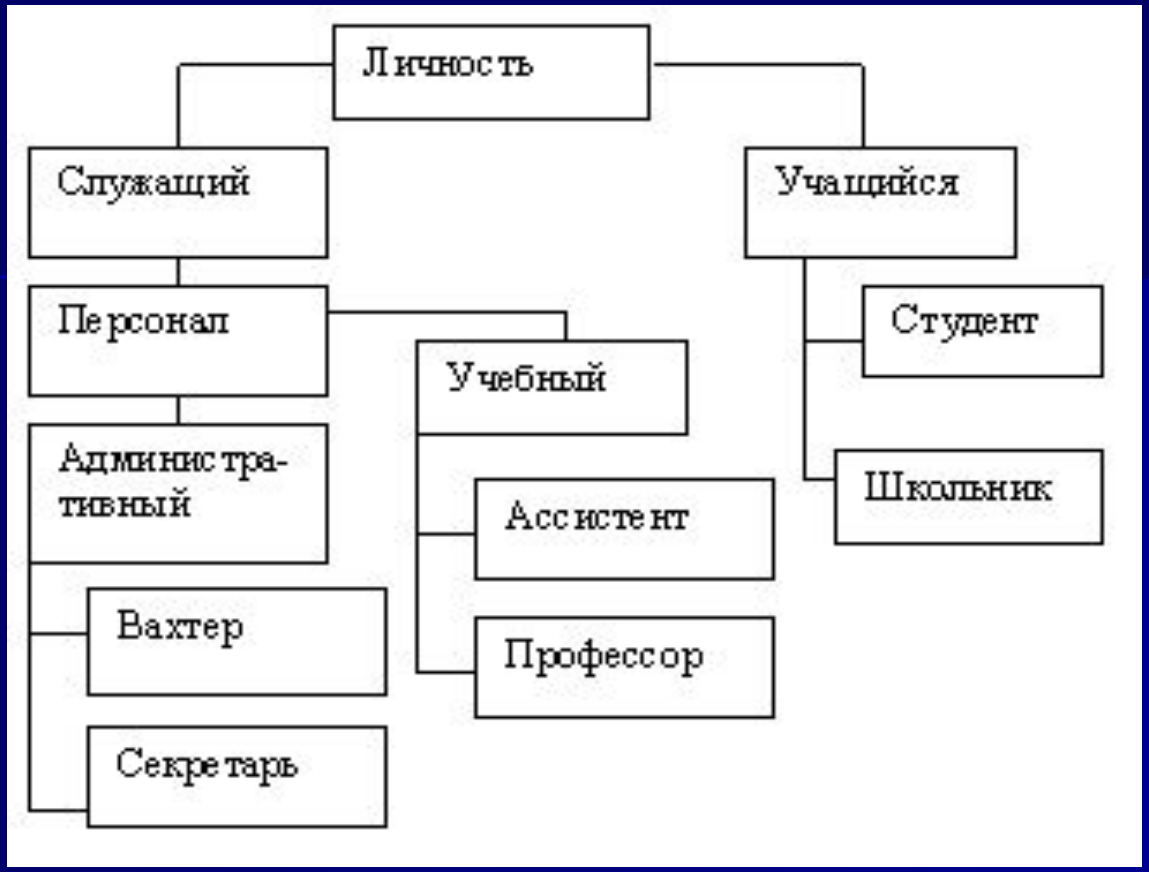

Рис. 4.2. Пример построения модели на основе процедур обобщения

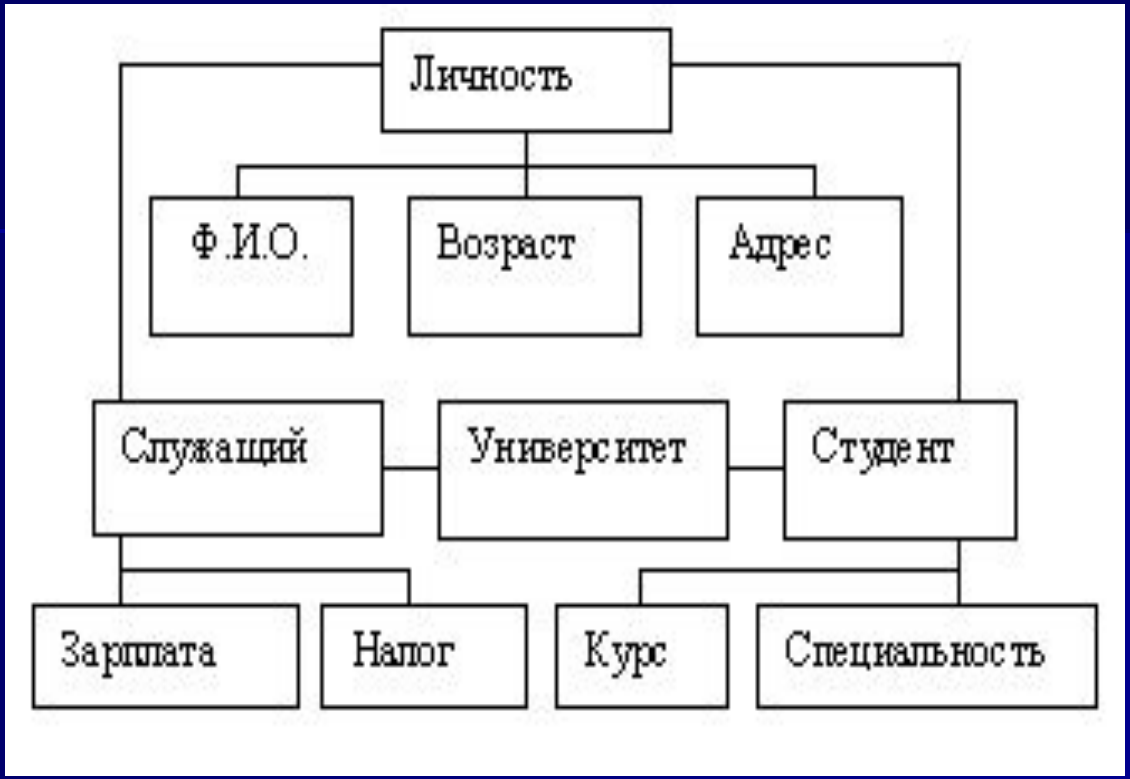

Рис. 4.3. Пример построения модели на основе процедур агрегации

4.2. Инфологическая модель

- Инфологическая модель является базовой моделью данных, используемой в ГИС. Она дает формализованное описание проблемной области независимо от структур данных.
- Одно из главных понятий инфологической модели – объект. Это понятие связано с событиями: возникновение, исчезновение и изменение. Объекты могут быть атомарными или составными.
- Атомарный объект – это объект определенного типа, дальнейшее разложение которого на более мелкие объекты внутри данного типа невозможно.

- Составные объекты включают в себя множества объектов, кортежи объектов.
- Целью инфологического моделирования является формализация объектов реального мира предметной области и методов обработки информации в соответствии с поставленными задачами обработки и требованиями представления данных.
- Инфологическая модель носит описательный характер и может включать в себя ряд компонентов (рис. 4.6.) .
- В настоящее время не существует общепринятых способов ее построения. Используют аналитические методы, методы графического описания и системный подход.

Рис. 4.6. Основные компоненты инфологической модели

4.3. Иерархическая модель

- Иерархическая модель относится к наиболее простым структурно определенным моделям. В этой модели данных связи между ее частями являются жесткими, а ее структурная диаграмма должна быть упорядоченным деревом. При этом, для описания различных уровней модели используют следующие понятия: корень, ствол, ветви, листья и лес.
- Обобщенная иерархическая модель представляет собой описание процесса или системы, состоящей из совокупности уровней, связанных одной дугой (рис. 4.7).

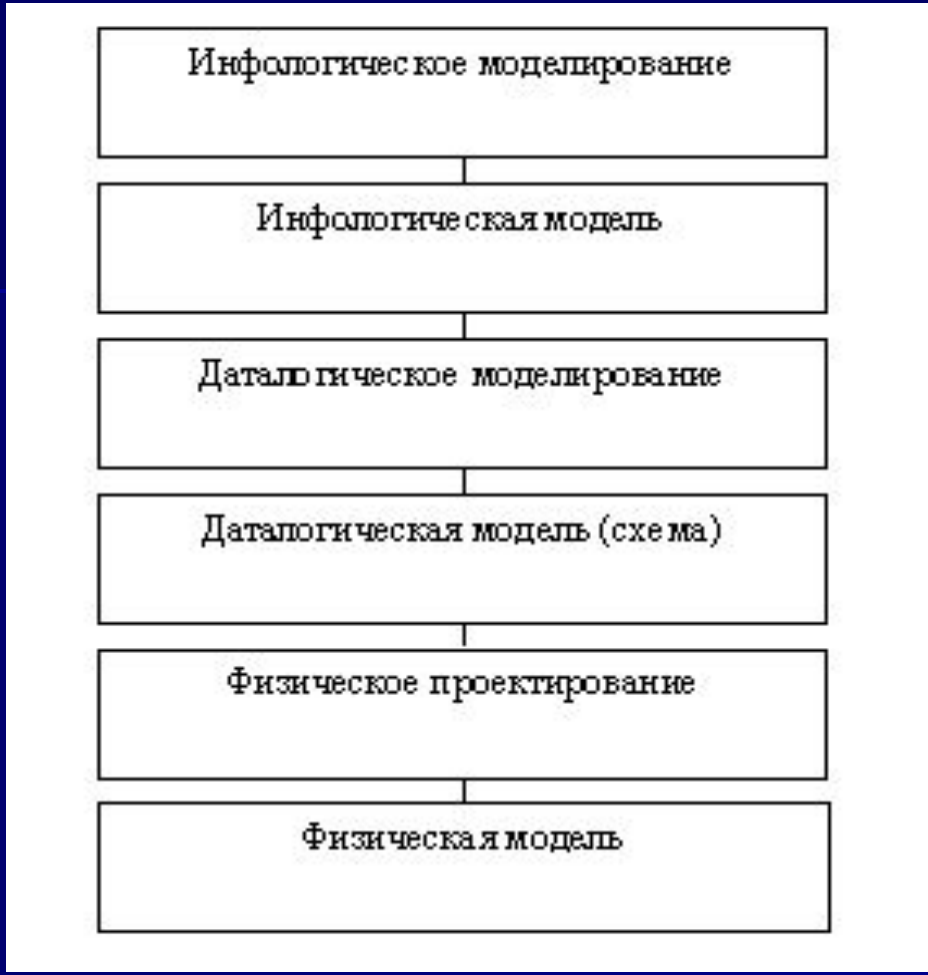

Рис. 4.7. Иерархическая модель проектирования ГИС

- В иерархических моделях данных существует два внутренних ограничения: 1) все типы связей должны быть фукнциональными; 2) структура связей должна быть древовидной. Следствием этих ограничений является необходимость соответствующей структуризации данных. В силу функциональности связей запись может иметь не более одной исходной записи любого типа, т.е. связь должна иметь жесткий вид – один ко многим.
- Недостатком иерархической модели является снижение времени доступа при большом числе уровней, поэтому в ГИС не используются модели при большом числе уровней (более 10). Однако, иерархические модели довольно устойчиво применяются для составления различного рода классификаторов.

4.4. Квадротомическое дерево

- Квадротомическое дерево относится к иерархической структуре данных и используется для накопления и хранения географической информации.
- Технология построения квадротомического дерева основана на рекурсивном разделении квадрата на квадранты и подквадранты до тех пор, пока все подквадранты не станут однородными по отношению к значению изображения (цвета) или пока не будет достигнут предопределенный заранее наименьший уровень разрешения.
- Преимущество такой структуры состоит в том, что регулярное разделение обеспечивает накопление, восстановление и обработку данных простым и эффективным способом.

4.5. Реляционная модель

- В современных информационных системах и базах данных наиболее широко представлены реляционные модели.
- Термин «реляционный» указывает прежде всего на то, что такая модель хранения данных построена на взаимоотношении составляющих ее частей, которые удобно хранить в виде таблицы. Таблица состоит из строк и столбцов и имеет имя, уникальное внутри базы данных. Информация в реляционной модели данных представлена в виде совокупности взаимосвязанных таблиц, которые принято называть отношениями и реляциями.

- Основными понятиями реляционной модели данных являются: тип данных, домен, атрибут, кортеж, ключ.
- ***Доменом*** называется совокупность значений, не повторяющихся в одном столбце таблицы.
- Столбцы отношения называют ***атрибутами***, им присваиваются имена, по которым к ним затем производится обращение.
- Список имен атрибутов отношения с указанием имен доменов называется ***схемой отношения. Степень отношения*** – это число его атрибутов.
- Каждая строка в таблице – это ***запись или кортеж***. Кортеж представляет собой множество пар {имя атрибута, значение}, которое содержит одно вхождение каждого имени атрибута, принадлежащего схеме отношений. «Значение» является допустимым значением домена данного атрибута.
- ***Ключ отношения*** – это подмножество атрибутов, имеющее следующие свойства:

- 1) уникальную идентификацию;
- 2) неизбыточность;
- 3) ни один из атрибутов ключа нельзя удалить, не нарушив его уникальность.
- В зависимости от количества атрибутов, входящих в ключ, различают простые и сложные (или составные) ключи.
- На рисунке 4.8 представлен пример реляционной модели. Таблица «Успеваемость» содержит сведения об успеваемости студентов по предмету «Высшая математика».
- Каждый столбец таблицы имеет уникальное имя, которое записывается в верхней части таблицы. Любая таблица имеет один или несколько столбцов, значения в которых однозначно идентифицируют каждую ее строку. Такой столбец (или комбинация столбцов) называются первичным ключом.

№ студ. билета	Ф.И.О.	Оценка
2398021	Иванов И.И.	4
2398022	Петров А.П.	3
2398023	Рыбин И.И.	4
2398024	Семин П.П.	5
2398025	Теркина Ю.И.	5

Рис. 4.8. Реляционная модель «Успеваемость»

- В таблице «Успеваемость» первичным ключом является столбец «№ студ. билета».
- Взаимосвязь таблиц – важнейший элемент реляционной модели данных. Она поддерживается внешними ключами. Внешний ключ – это атрибут (или множество атрибутов) одного отношения, являющийся ключом другого (или того же самого) отношения.
- В рассматриваемом примере атрибут «№ студ. билета» иллюстрирует понятие внешнего ключа.
- Внешние ключи используются для установления логических связей между отношениями. Связь между двумя таблицами устанавливается путем присваивания значений внешнего ключа одной таблицы значениям ключа другой.

- Помимо таблиц в ГИС могут храниться и другие объекты, такие как экранные формы, отчеты, представления, прикладные программы, работающие с информацией, размещенной в реляционной модели.
- Данные информационные системы должны быть однозначными и непротиворечивыми, т.е. для них устанавливаются некоторые ограничительные условия, обеспечивающие логическую основу для поддержания корректных значений данных в базе. Ограничения целостности позволяют свести к минимуму ошибки, возникающие при обновлении и обработке данных.
- Важнейшими ограничениями целостности данных являются: категорийная целостность и ссылочная целостность.

4.6. Модель «сущность-связь»

- Модель данных «сущность-связь» (часто называемая также ER-моделью) дает представление о предметной области в виде объектов, называемых сущностями, между которыми фиксируются связи.
- Основными понятиями ER-модели являются сущность, связь и атрибут.
- На рис. 4.9 приведена схема проектирования геоинформационной системы, построенная на основе модели «сущность-связь». ER-модели могут рассматриваться как обобщение и развитие иерархических и сетевых моделей, что допускает спецификацию ограничений целостности и непосредственное представление связей типа «один к одному» (1:1), «один ко многим» (1:M), «многие к одному» (M:1), «многие ко многим» (M:N).

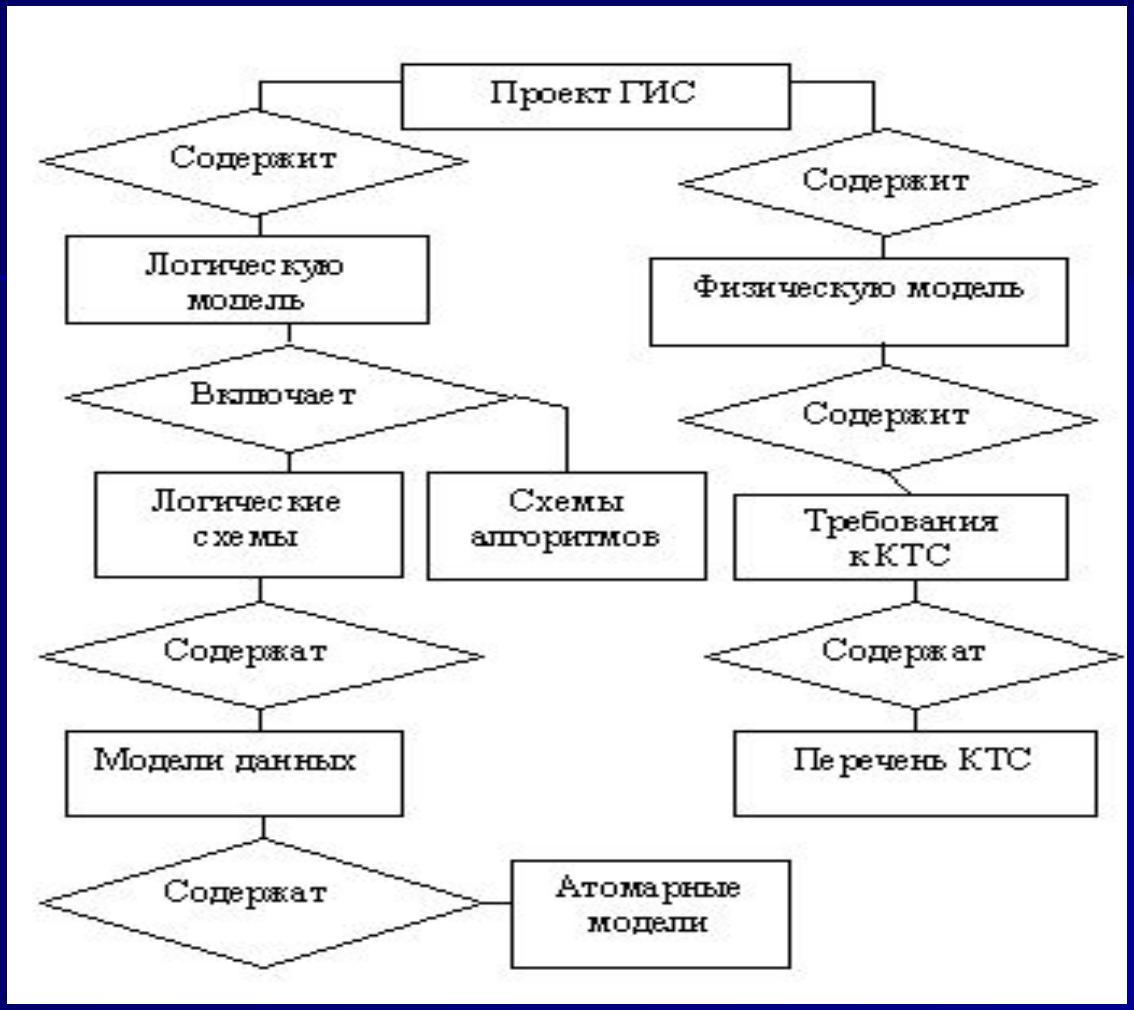

Рис. 4.9. Модель «сущность-связь» проекта ГИС:

КТС – комплекс технических средств

- При построении модели «сущность-связь» следует учитывать разновидность объектов. Объекты могут быть простыми или сложными. Сложные объекты в свою очередь подразделяются на составные, обобщенные и агрегированные.
- ER-модели, в связи с наглядностью представления концептуальных схем баз данных, получили широкое распространение в CASE-средствах, предназначенных для автоматизированного проектирования реляционных баз данных.

4.7. Прочие модели

- ***Сетевые модели*** дают представление о проблемной области в виде объектов, связанных бинарным отношением «многие ко многим», при этом каждый объект может иметь несколько «подчиненных» и несколько «старших» объектов. Эти модели используют табличные и графовые представления.
- ***Бинарные модели*** дают представление о проблемной области объекта в виде бинарных отношений, характеризуемых триадой: объект, атрибут, значение. Эти модели имеют графовое представление в виде B-дерева. Вершина графа соответствует классификационному обобщению данных в типы, называемые категориями. Дуги соответствуют бинарному отношению категорий.
- ***Семантические сети*** как модели данных созданы для изучения проблем искусственного интеллекта. Базовые структуры в этих моделях могут быть представлены графом, множество вершин и дуг которого, как для бинарной, так и сетевой модели образует сеть.