

Лекция 7

Базы данных и банки данных в ГИС. Основные типы баз данных. Системы управления базами данных (СУБД). Функции СУБД.

Цель лекции: сформировать понятие о базах географических данных и их типах.

Краткое содержание лекции. В процессе проектирования БД обычно выделяют три основных уровня: концептуальный, логический и физический.

Концептуальный уровень не зависит от имеющихся аппаратных и программных средств. Для БД ГИС он связан с концептуальной Моделью географических данных и включает: описание и определение рассматриваемых объектов; установление способа представления географических объектов в базе данных; выбор базовых типов пространственных объектов — точки, линии, ареалы, ячейки Раstra; решение вопроса о способе представления размерности и взаимосвязей реального мира в БД (например, следует ли показать здание в виде ареала или точки). На концептуальном уровне определяется и содержание базы данных, в свою очередь определяемое сутью явления, характером его пространственного распространения и задачами, для которых создается БД. Здесь следует выделить задачи создания одной или серии карт, комплексного картографирования, создания синтетических карт для многоцелевого и многократного использования.

Логический уровень определяется имеющимися программными средствами и практически не зависит от технического обеспечения. Он включает разработку логической структуры элементов базы данных в соответствии с системой управления базами данных (СУБД), используемой в программном обеспечении. Наиболее распространенными логическими структурами — моделями БД и их СУБД — являются иерархическая, сетевая, реляционная. В иерархической модели записи данных образуют древовидную структуру, при этом каждая запись связана только с одной записью, находящейся на более высоком уровне. Доступ к любой записи осуществляется по строго определенным «веткам» и узлам такого дерева. Иерархические модели хорошо подходят для задач с явно выраженной иерархически соподчиненной структурой информации и запросов. Они обладают низким быстродействием, трудно модифицируемы, но эффективны с точки зрения организации машинной памяти.

В сетевых моделях каждая запись в каждом из узлов сети может быть связана с несколькими другими узлами; кроме данных записи содержат указатели, определяющие местоположение других записей, связанных с ними. Такие модели очень трудно редактировать, например удалять записи, так как вместе с данными нужно редактировать и указатели. Подобные модели хорошо работают в случае решения сетевых, коммуникационных задач.

В иерархической и сетевой моделях для поиска конкретной записи необходимо вначале определить путь доступа к записи, а затем просмотреть все записи, находящиеся на этом пути.

Реляционные СУБД завоевали самую широкую популярность. Они свободны от всех ограничений, связанных с организацией хранения данных и спецификой запоминающих устройств. Эти модели имеют табличную структуру (рис. 16, в): строки таблицы соответствуют одной записи сведений об объекте, а столбцы — поля — содержат однотипные характеристики всех объектов. Всевозможные способы индексации данных существенно сокращают время поиска и запроса к данным. К числу наиболее известных СУБД реляционного типа относятся dBASE, Clipper, Foxbase, Paradox, ORACLE (последняя особенно подходит для больших объемов данных).

Физический уровень связан с аппаратными и программными средствами. На этом уровне определяются объемы хранимой в БД информации и необходимые объемы памяти компьютера (оперативной и долговременной), рассматриваются вопросы о структурировании файлов на диске или других носителях информации для обеспечения

программного доступа к ним, представления данных в памяти компьютера (целые, действительные числа, байты или буквенно-цифровые характеристики).

Объектно-ориентированные модели применяют, если геометрия определенного объекта способна охватывать несколько слоев, атрибуты таких объектов могут наследоваться, для их обработки применяют специфические методы.

Для обработки данных, размещенных в таблицах необходимы дополнительные сведения о данных, их называют метаданными.

Метаданные - данные о данных: каталоги, справочники, реестры и иные формы описания наборов цифровых данных.

Как правило, ГИС создаются на основе уже существующих систем управления базами Данных (СУБД), приобретение или аренда СУБД составляет основную часть затрат на программное обеспечение системы. СУБД выполняет множество функций, которые в противном случае следовало бы программировать в ГИС. Различают два пути использования СУБД в ГИС:

1) выполнение ГИС-процедур полностью через СУБД, тогда Доступ ко всем данным осуществляется только через СУБД и все данные должны удовлетворять требованиям, заложенным при ее разработке;

2) некоторые данные (обычно таблицы атрибутов и их отношений) доступны через СУБД, поскольку они вполне соответствуют модели, а к некоторым данным (обычно пространственно локализованным) доступ прямой, так как они не удовлетворяют требованиям модели СУБД.

ГИС добавляет географический аспект к уже существующим методам поиска и запроса. Сложность и разнообразие представления данных в ГИС, различимость в представлении позиционной и атрибутивной составляющей информации, необходимость ее обработки в контексте пространственной близости предъявляют своеобразные и повышенные требования к СУБД по сравнению с традиционной формой их использования.

Функции СУБД. Каждую СУБД принято характеризовать способностью выполнять следующие основные функции [К. Дейт, 1980 Дж. Ульман, 1983]:

- о управление данными во внешней памяти;
- о управление буферами оперативной памяти;
- о операции над БД;
- обеспечение надежности хранения данных в БД;
- о поддержка языка управления БД.

Управление данными во внешней памяти. Эта функция обеспечивает организацию структуры внешней памяти как для хранения данных, входящих в БД, так и для служебных целей, например, для ускорения доступа к данным. В некоторых СУБД используются возможности файловых систем, в других работа производится на уровне функционирования устройств внешней памяти. В любом случае пользователи СУБД не обязаны знать, какая структура используется или как организованы файлы. Обычно в СУБД создается собственная система наименования объектов БД.

Управление буферами оперативной памяти. СУБД обычно работают с БД значительного размера, существенно большего доступного объема оперативной памяти. Для того чтобы СУБД не зависела от скорости работы устройств внешней памяти, используется организация собственных наборов буферов оперативной памяти с определенными правилами замены и обновления буферов.

Операции над БД. Последовательность операций над БД, рассматриваемых СУБД как единое целое, называется транзакцией. При выполнении транзакции СУБД либо фиксирует во внешней памяти изменения в БД, произведенные этой транзакцией, либо не производит никаких изменений. Понятие транзакции важно для сохранения логической целостности БД, особенно в многопользовательских СУБД. Каждая транзакция начинается при целостном состоянии БД и оставляет это состояние целостным после своего завершения. При эффективном управлении параллельно выполняющимися транзакциями со стороны

СУБД каждый из пользователей может ощущать себя единственным ее пользователем. Управление транзакциями в многопользовательской СУБД осуществляется с помощью специальных операций, которые объединяют транзакции одного пользователя в серии (сериализация транзакций), при этом суммарный эффект смеси транзакций эквивалентен эффекту их последовательного выполнения. Существует несколько базовых алгоритмов сериализации транзакций, среди которых наиболее распространены алгоритмы, основанные на синхронизационных захватах объектов БД.

Следует рассматривать системы управления базами данных в отношении ГИС не столько в качестве модуля управления базой пространственных данных (создание, наполнение, поддержание БД в корректном состоянии), сколько в качестве специализированных средств пространственного анализа геоинформации. Рассматриваемая с данной точки зрения СУБД является компонентой, позволяющей обеспечить основное назначение геоинформационных систем – анализ и обобщение информации с пространственно-временной привязкой в целях принятия решения на уровне управления геопространством. В данном случае под пространственным анализом следует понимать получение таких данных в результате запроса пользователя, которые в явном виде не содержатся в базе данных.

Для работы с базами данных Используются специальные языки, называемые языками баз данных. Первоначально в СУБД поддерживалось несколько специализированных по функциям языков. В современных СУБД обычно поддерживается единый интегрированный язык, содержащий все необходимые средства для работы с БД, начиная от ее создания, и обеспечивающий базовый пользовательский интерфейс с базами данных. Стандартным языком наиболее распространенных в настоящее время реляционных СУБД является язык SQL (Structured Query Language). Язык SQL позволяет определять схему реляционной БД и манипулировать данными. При этом именованное объектов БД (таблиц и их столбцов) поддерживается на языковом уровне в том смысле, что компилятор языка SQL производит преобразование имен объектов в их внутренние идентификаторы на основании специально поддерживаемых служебных таблиц-каталогов. Внутренняя часть СУБД (ядро) вообще не работает с именами таблиц и их столбцов.