

Тема №1 Концепция баз данных, понятие моделей данных

1. Основные подходы к обработке информации в автоматизированных информационных системах
2. Концепция баз данных
3. Системы управления базами данных
4. Модели данных
5. Общие принципы классификации СУБД
6. Обзор основных моделей данных

Основные подходы к обработке информации в автоматизированных информационных системах

Фундаментальным принципом в решении вопроса о соотношении программ и данных является *концепция независимости прикладных программ от данных*.

Суть этой концепции состоит не столько в отделении программ от данных, сколько в рассмотрении их как самостоятельных взаимодействующих объектов.

Одной из последних модификаций этого принципа является концепция независимости прикладных программ от данных вместе с процедурами их обработки - объектно-ориентированный подход в программировании.

Основные подходы к обработке информации в автоматизированных информационных системах

Развитие концепции независимости программ от данных привело к формированию в 1962 году концепции базы данных (БД) и созданию на ее основе метода баз данных для решения задач обработки информации.

До середины 60-х годов прошлого века основной концепцией построения программного обеспечения являлась концепция *файловой системы* и так называемый *позадачный метод*.

Основные подходы к обработке информации в автоматизированных информационных системах

Основные концепции обработки данных

Концепция файловой системы

- данные ИС представляются в виде совокупности файлов несущей Операционной системы;
- структура файла определяется разработчиком ИС (совмещение логической и физической структуры данных);
- любая программа, содержащая запрос к данным, должна уметь интерпретировать структуру необходимых для реализации запроса файлов (зависимость программ обработки от организации данных);
- каждый запрос требует в принципе своей программы обработки (разнообразие точек зрения алгоритмическую обработку данных).

Основные подходы к обработке информации в автоматизированных информационных системах

Концепция баз данных

- данные ИС размещаются в файлах несущей Операционной системы;
- физическая структура файлов фиксируется, а логическая структура данных представляется на ее основе (разделение и физической и логической структуры данных);
- обработка запроса к данным реализуется через специальный интерфейс манипулирования данными (независимость программ от организации данных);
- программы разрабатываются на основе единой точки зрения на процедуры обработки данных.

Основные подходы к обработке информации в автоматизированных информационных системах

Концепция объектно-ориентированных баз данных

- данные ИС размещаются в файлах несущей Операционной системы;
- информация ИС представляется в виде объектов (логическая структура и методы - принцип разделения логической и физической структуры остается в силе);
- обработка запроса к данным реализуется через интерфейс манипулирования объектами (независимость программ и объектов);
- программы разрабатываются на основе единой точки зрения на классы объектов и идеи повторного использования программного кода.

Основные подходы к обработке информации в автоматизированных информационных системах

Методы обработки данных

Позадачный метод

Суть: декомпозиция программы на подзадачи с своими блоками данных, своими алгоритмами.

Проблемы:

- избыточность данных;
- взаимосвязь между данными и программой;
- замкнутость системы;
- нарушается естественная для учреждения схема обработки данных.

Основные подходы к обработке информации в автоматизированных информационных системах

Метод баз данных

Суть: наличие отдельного описания логической структуры данных, единая точка зрения на процедуры обработки данных.

Проблемы:

- проблема представления данных;
- проблема обучения персонала;
- структурная реорганизация учреждения (реинжиниринг).

Метод объектно-ориентированных баз данных

Суть: информация (данные + программы) представляется объектами, которые обмениваются сообщениями (взаимосвязь процессов).

Проблемы:

- проблема определения объектов (классы);
- проблема обучения персонала;
- реорганизация учреждения.

Концепция баз данных

Идея повышения степени независимости обрабатывающих программ от способов хранения и содержания хранимых данных впервые была использована в концепции баз данных путем *разделения логического и физического уровней хранения данных* в 1964 году в исследованиях сотрудников фирмы IBM.

База данных - унифицированная совокупность хранимых и воспроизводимых данных, используемых в рамках организации

Концепция баз данных

Несколько дополнительных определений базы данных:

- *База данных есть совокупность взаимосвязанных данных, совместно используемых несколькими приложениями и хранящимися с (минимальной) регулируемой избыточностью.*
- *База данных состоит из всех экземпляров записей, экземпляров наборов записей и областей, которые контролируются конкретной схемой.*

Концепция баз данных

Существенный момент для разработчика ИС - данные становятся определенным образом организованы, приобретают некую упорядоченность и внутреннюю структуру, а также имеется некоторый набор унифицированных операций обработки данных и декларативных средств представления данных.

К таким операциям следует отнести операции "Вставить" (Insert), "Добавить" (Add), "Удалить" (Delete) и ряд других.

К декларативным средствам представления данных следует отнести языки определения данных.

Концепция баз данных

Использование рассматриваемой концепции при создании ИС предполагает наличие *языка определения данных* и *языка манипулирования данными*, а также *правил построения интерфейсов программ (приложений) с БД и пользователем*.

Деление средств манипулирования данными и их представления является условным.

Язык определения данных служит для описания логической структуры (схемы) БД и способов хранения и доступа к данным.

Язык манипулирования данными предоставляет алгоритмические средства построения приложений для обработки сохраняемых в БД элементов данных.

Системы управления базами данных

СУБД являются сложными программными системами, работающими на различных операционных платформах.

Именно СУБД должна предоставить средства определения и манипулирования данными, сделав данные независимыми от прикладных программ, их использующих.

В последнее время набирает обороты концепция машин баз данных, которая предполагает *аппаратную реализацию некоторых процедур обработки данных.*

Системы управления базами данных

Основные функции СУБД

- обеспечить языковые средства описания и манипулирования данными;
- обеспечить поддержку логических моделей данных;
- обеспечить взаимодействие логической и физической структур данных;
- обеспечить защиту и целостность данных;
- обеспечить поддержку баз данных в актуальном состоянии.

Системой управления базами данных (Data-base Management System) называется совокупность программных средств, необходимых для использования базы данных и предоставляющих разработчикам и пользователям множество различных представлений данных.

Модели данных

Представление информации с помощью данных требует унифицированного подхода к понятию *данных как независимого объекта моделирования*.

Поэтому для разработчика ИС выбор соответствующей модели данных является одной из самых важных проблем. Выбор модели данных влечет за собой выбор средств анализа предметной области (ПО БД) как сферы реального мира, подлежащего изучению и обработке средствами ВТ.

Модели данных

- *Модель данных* ограничивает возможность выбора СУБД, так как обычно отдельно взятая СУБД поддерживает определенную модель данных.
- *Модель данных* определяет и методы создания дружественного интерфейса пользователя за счет средств СУБД (особенности конкретной реализации модели (замкнутость на свою среду) иногда весьма существенные, ибо коммерческие интересы фирм - разработчиков СУБД вступают в противоречие с требованиями рынка информационных услуг).

Модели данных

- *Модель данных* требует приведения представлений пользователя о данных и результатах их обработки к определенному уровню понимания, что может повлечь за собой необходимость обучения пользователя методам и средствам работы с данными.

Модели данных

Понятие *модели данных* является одним из фундаментальных понятий информатики, от которого во многом зависят механизмы реализации ИС как программно-аппаратного комплекса.

В самом общем случае модель данных - это логическое представление *данных* и совокупность операций над ними.

Модель данных (Data Model) есть логическая структура данных, которая представляет присущие этим данным свойства, не зависящие от аппаратного и программного обеспечения и не связанные с функционированием компьютера.

Модели данных

Рассматривают несколько аспектов моделирования в обработке данных:

информационное моделирование:

- концептуальное моделирование (моделирование семантики предметной области);
- логическое моделирование данных;

физическое моделирование:

- создание моделей доступа к данным;
- оптимизация физической организации данных в аппаратной среде.

Модели данных

Понятие об информационной модели данных

Принцип описания

Информационная модель данных определяется точкой зрения пользователя и в его терминах.

Средства описания

Сущности
Атрибуты
Связи

Результат моделирования

Исходные данные для определения внешней схемы базы данных ИС

Модели данных

Объектами информационной модели являются сущности реального мира из предметной области.

*Свойства объектов (сущностей) называют *атрибутами*.*

Сущности вступают в связи друг с другом через свои атрибуты.

Эти три компонента информационной модели представляют субъективные средства описания модели, которые после определенной формализации дают внешнюю схему данных БД ИС.

Модели данных

В рамках информационного моделирования существует несколько точек зрения (схем) на абстрагирование *данных*.

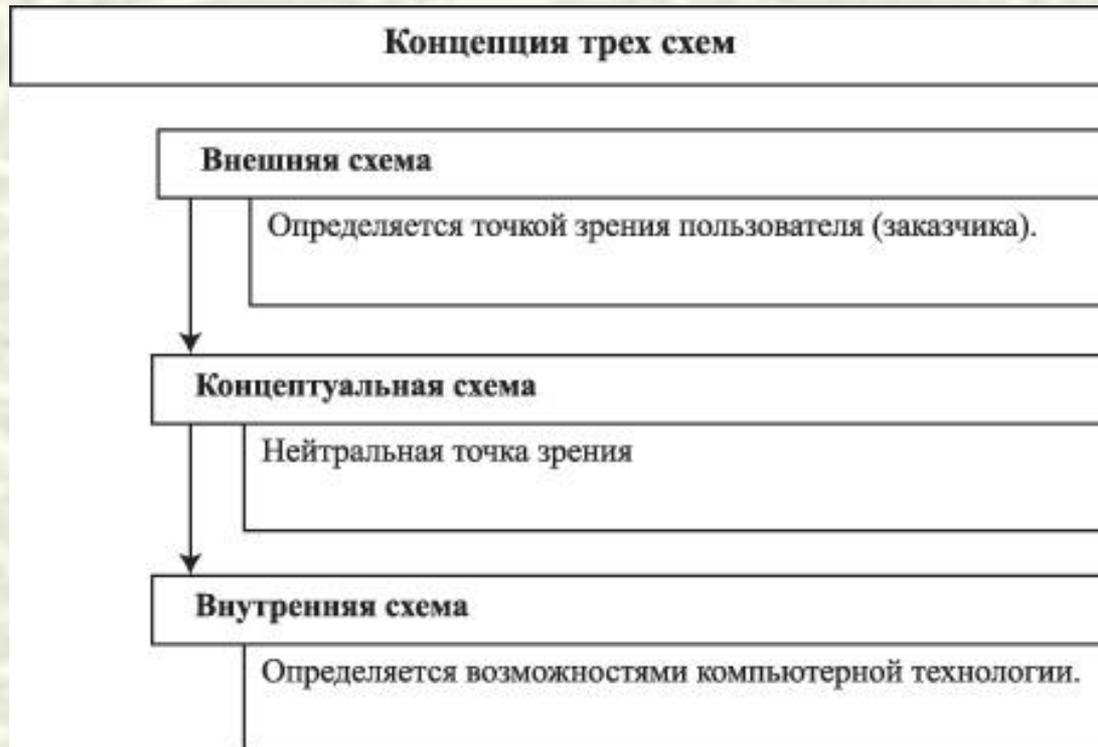
1. С точки зрения пользователя (называемой *внешней схемой*), определение данных представляется в контексте языка предметной области. *Структура данных и содержание* меняется в зависимости от сферы деятельности и особенностей конкретного пользователя.
2. С точки зрения компьютера (называемой *внутренней схемой*), данные определяются в терминах файловых структур для хранения и поиска. *Структура данных в этом случае* зависит от конкретной компьютерной технологии и от требований эффективности обработки данных.

Модели данных

Исследовательская группа по СУБД ANSI/X3/SPARC пришла к выводу, что для создания идеальной среды управления данными необходимо определение их с *третьей, промежуточной точки зрения* (концепция трех схем ANSI/X3/SPARC).

Эта точка зрения (называемая *концептуальной схемой*) сводится к *единообразному определению данных в рамках предметной области, не ориентированному на какое-либо конкретное использование их и не зависящему от того, как данные физически обрабатываются на компьютере.*

Модели данных



Основные свойства концептуальной схемы:

- Должна быть согласованной с инфраструктурой предметной области и верной во всех сферах ее применения.
- При ее расширении новые данные должны определяться без изменения ранее определенных данных.
- Должна адаптироваться как к точкам зрения пользователей, так и к многообразию структур хранения и поиска данных.

Модели данных

Наличие в СУБД определенной, допустимой структуры данных приводит к понятию *баз структурированных данных*, то есть данные в таких БД должны быть представлены как *совокупность взаимосвязанных элементов*.

Если допустить возможность порождения новых типов и динамический процесс установления связей (во время появления объекта в БД), то мы приходим к понятию *баз неструктурированных данных*.

Допустимы и промежуточные варианты, которые носят название *БД с частично детерминированной схемой*.

Модели данных

Такое деление БД с точки зрения степени структурированности сохраняемых данных оказывается существенным моментом при выборе несущей СУБД для реализации ИС, поскольку конкретная СУБД обычно поддерживает определенную *модель данных*.

С другой стороны, следует иметь в виду, что для каждого из приведенных типов БД используются соответствующие модели данных, т.е. существует некоторое множество моделей данных.

Модели данных

В настоящее время для баз структурированных данных различают *три основных типа логических моделей данных* в зависимости от характера поддерживаемых ими связей между элементами данных:

- сетевую,
- иерархическую,
- реляционную.

Классифицирующими признаками в этих моделях являются: *степень жесткости (фиксации) связи, математическое представление структуры модели и допустимые типы данных.*

Модели данных

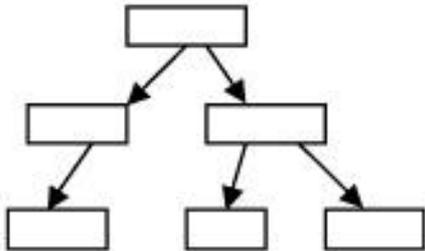
Модель данных	Характер связи между объектами	Формальное представление
Сетевая	Полужесткие связи	Произвольный граф
Иерархическая	Жесткие связи	Древовидная структура
Реляционная	Изменчивые связи	Плоский файл

При сопоставлении моделей следует помнить, что все они *теоретически эквивалентны*.

Эквивалентность моделей состоит в том, что *они могут быть сведены одна к другой путем формальных преобразований*.

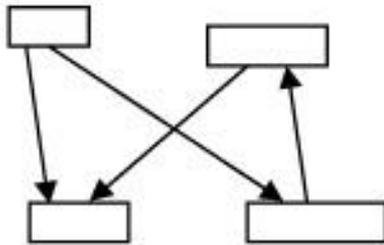
Модели данных

Основные типы моделей данных



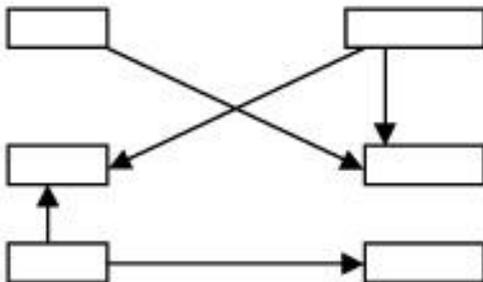
Иерархическая

Взаимосвязи между данными жестко фиксированы.
Изменение связи ведет к реорганизации структуры.
Число связей ограничено.



Сетевая

Характер связей более разнообразен.
Трудно вводить изменения.



Реляционная

Таблицы независимы.
Связи полностью изменчивы.
Простота расширения.

Общие принципы классификации СУБД

СУБД классифицируются по типу модели данных, которую они поддерживают. Следовательно, различают СУБД *сетевые*, *иерархические* и *реляционные*.

С точки зрения практики обработки данных СУБД в самом общем виде БД подразделяют на:

- *фактографические*, которые хранят совокупность фактов интегрированных, возможно, из различных документов;
- *документальные*, которые ориентированы на хранение документов;
- *документально-фактографические*, которые обладают чертами и тех и других.

Общие принципы классификации СУБД

Примеры:

СУБД CDS/ISIS в первую очередь ориентирована на поддержку работы с документом, который состоит из определенного числа рубрик, проиндексированных по тезаурусу ключевых слов.

СУБД ADABAS хорошо подходит для организации фактографических БД.

СУБД ORACLE - для БД смешанного типа.

Классификация БД не завершена: попытки ввести новые типы БД продолжаются (активные, дедуктивные, нечеткие реляционные, графические БД и т.д.).

Общие принципы классификации СУБД

Во многих случаях для разработчиков ИС бывает важно деление СУБД (и БД) по характеру обработки: на *централизованные* и *распределенные*.

При использовании распределенной обработки следует обратить внимание на характер обработки *транзакций*, которые оказывают существенное влияние на производительность системы.

Под *транзакцией* в самом общем случае понимают *единицу работы, требуемой пользователем от БД, независимо от характера обработки*.

Общие принципы классификации СУБД

Чаще всего в результате обработки *транзакции* реализуется запрос пользователя либо на выборку данных из БД, либо на обновление БД, либо на выполнение каких-то иных действий над БД.

Предполагается, что выполнение запроса сопровождается выполнением комплекса внутрисистемных действий СУБД, направленных на поддержание целостности данных, разграничение доступа и т.п.

Общие принципы классификации СУБД

Существуют различные концептуальные подходы к обработке *транзакций* при распределенной обработке. Принципиальным здесь является не только вопрос как, но и где локализуется обработка *транзакции*:

- на файлах компьютера конечного пользователя или
- на выделенном в сети компьютере.

От выбора той или иной концепции будет зависеть *время отклика системы на запрос пользователя*.

Параметр "время отклика системы на запрос пользователя" очень часто выступает в качестве определяющего или желательного параметра разрабатываемой системы.

Обзор основных моделей данных

Тремя основными типами моделей структурированных данных являются *иерархическая, сетевая и реляционная*.

Иерархическая модель данных организует данные в виде древовидной структуры и является реализацией логических связей между данными типа родовидовых отношений или отношений "часть-целое".

Примером простого иерархического представления может служить административная структура организации.

Обзор основных моделей данных

Деревом в информатике называют совокупность корневого элемента и множества подчиненных ему элементов, в которой отношения между элементами носят подчиненный вертикальный характер. Горизонтальные связи в такой системе отношений не допускаются.

В иерархической модели имеется корневой узел или корень дерева.

Остальные узлы называются порожденными и связаны между собой следующим образом: каждый узел имеет исходный, находящийся на вышестоящем уровне.

Узлы, не имеющие порожденных, называются листьями.

Обзор основных моделей данных

Иерархия должна удовлетворять следующим условиям:

- Иерархия имеет исходный узел (корень), из которого строится дерево. Каждое дерево имеет только один корень.
- Узел имеет непустое множество атрибутов, которые описывают объект, моделируемый в данном узле.
- Порожденные узлы могут добавляться в дерево как в вертикальном, так и в горизонтальном направлении.
- Доступ к порожденным узлам возможен только через исходный узел, поэтому существует только один путь доступа к каждому узлу.
- Возможно существование нескольких экземпляров каждого узла каждого уровня. При этом каждый экземпляр исходного узла начинает логическую запись.

Обзор основных моделей данных

К основным недостаткам *иерархической модели* можно отнести:

- сложность отображения связи "многие-к-многим"
- усложнение операции включения новых объектов и удаления устаревших объектов непосредственно в базе данных (в особенности обновление и удаление связей);
- неоднозначность представления данных о предметной области.

Обзор основных моделей данных

Сетевая структура может быть описана в терминах исходных и порождаемых узлов, а также представлена таким образом, чтобы порожденные узлы располагались ниже исходных.

При рассмотрении некоторых сетевых структур можно говорить об уровнях.

Сетевая модель данных является моделью объектов-связей, где допускаются только бинарные связи типа "многие-к-одному", что позволяет использовать для представления данных простую модель ориентированных графов.

В некоторых определениях *сетевой модели* допускаются связи типа "многие-ко-многим", но требование бинарности связи остается в силе.

Обзор основных моделей данных

К недостаткам сетевого подхода в БД относят сложность самой модели данных и сложность освоения средств манипулирования данными в ней.

При анализе ПО БД и программировании особенно тщательно приходится отслеживать цепочки связанных групповыми отношениями данных при операциях вставки, обновления и удаления.