

Тема №1

Концепция моделей данных

1. Модели данных
2. Обзор основных моделей данных

Данные

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.

Особенности моделей 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).

Особенности моделей 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.

Модель данных

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.

Моделирование

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.

Информационная модель 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

Компоненты моделей 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.

Схемы абстрагирования 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.*

Концептуальная схема 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.*

Концепция трех схем

Концепция трех схем

Внешняя схема

Определяется точкой зрения пользователя (заказчика).

Концептуальная схема

Нейтральная точка зрения

Внутренняя схема

Определяется возможностями компьютерной технологии.

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.

Структурированность 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*.

Структурированность 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.

Типы логических моделей 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.*

Типы логических моделей данных

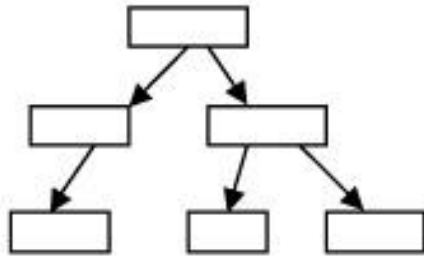
Модель данных	Характер связи между объектами	Формальное представление
Сетевая	Полужесткие связи	Произвольный граф
Иерархическая	Жесткие связи	Древовидная структура
Реляционная	Изменчивые связи	Плоский файл

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*.

Типы моделей данных

Основные типы моделей данных

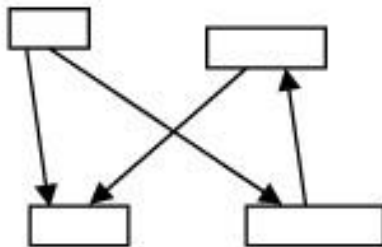


Иерархическая

Взаимосвязи между данными жестко фиксированы.

Изменение связи ведет к реорганизации структуры.

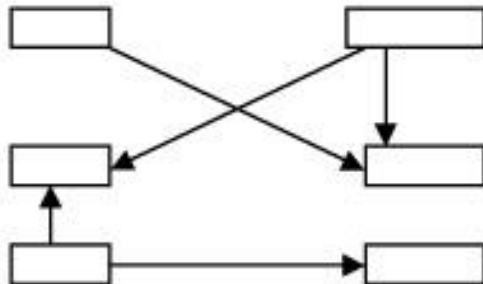
Число связей ограничено.



Сетевая

Характер связей более разнообразен.

Трудно вводить изменения.



Реляционная

Таблицы независимы.

Связи полностью изменчивы.

Простота расширения.

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.

Иерархические модели 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.

Иерархические модели 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.

Иерархические модели 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.

Сетевые модели 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.

Сетевые модели 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.