

Модель процесса хранения данных в информационных системах

Хранение и накопление являются одними из основных операций, осуществляемых над информацией и главным средством обеспечения ее доступности в течение некоторого промежутка времени.

Информационная система – это хранилище информации, снабженное процедурами ввода, поиска и размещения и выдачи информации. Наличие таких процедур – главная особенность информационных систем, отличающих их от простых скоплений информационных материалов.

База данных – организованная совокупность данных, предназначенная для длительного хранения во внешней памяти компьютера, постоянного обновления и использования.

Банк данных – система, представляющая определенные услуги по хранению и поиску данных определенной группе пользователей по определенной тематике.

Система управления базами данных (СУБД) – программное обеспечение, предназначенное для работы с базами данных. Операции для большинства распространенных СУБД делятся на 4 основные группы:

- Определение Данных (DataDefinition).
- Хранение Данных (DataMaintenance).
- Выборка Данных (DataRetrieval).
- Управление Данными (DataControl).

Каждая БД и СУБД строится на основе некоторой явной или неявной модели данных. Все СУБД, построенные на одной и той же модели данных, относят к одному типу. Например, основой реляционных СУБД является реляционная модель данных, сетевых СУБД — сетевая модель данных, иерархических СУБД — иерархическая модель данных и т.д.

Существуют два основных направления реализации СУБД: программное и аппаратное.

Программная реализация выполняет следующие функции:

- описание данных на концептуальном и логическом уровнях;
- загрузку данных;
- хранение данных;
- поиск и ответ на запрос (транзакцию);
- внесение изменений;
- обеспечение безопасности и целостности.

Обеспечивает пользователя следующими языковыми средствами:

- языком описания данных (ЯОД);
- языком манипулирования данными (ЯМД);
- прикладным (встроенным) языком данных (ПЯД, ВЯД).

Аппаратная реализация предусматривает использование так называемых машин баз данных (МБД). Их появление вызвано возросшими

объемами информации и требованиями к скорости доступа. Термин «компьютер БД» – автономный процессор баз данных или процессор, поддерживающий СУБД. Основные направления МБД:

- параллельная обработка;
- распределенная логика;
- ассоциативные ЗУ;
- конвейерные ЗУ;
- фильтры данных и др.

Хранилище данных (англ.*Data Warehouse*) — очень большая предметно-ориентированная информационная корпоративная база данных, специально разработанная и предназначенная для подготовки отчётов, анализа бизнес-процессов с целью поддержки принятия решений в организации. Строится на базе клиент-серверной архитектуры, реляционной СУБД и утилит поддержки принятия решений. Данные, поступающие в хранилище данных, становятся доступны только для чтения. Данные из промышленной OLTP-системы копируются в хранилище данных таким образом, чтобы построение отчётов и OLAP-анализ не использовал ресурсы промышленной системы и не нарушал её стабильность.

Существуют два архитектурных направления - **нормализованные** хранилища данных и **размерностные** хранилища.

Основные принципы организации хранилищ данных следующие.

- Предметная ориентация.
- Средства интеграции.
- Постоянство данных.
- Хронология данных.

Еще одним важным направлением развития баз данных являются репозитории. **Репозиторий** – место, где хранятся и поддерживаются какие-либо данные. Чаще всего данные в репозитории хранятся в виде файлов, доступных для дальнейшего распространения по сети (не пользовательские, а системные данные). Примером репозитория может служить репозиторий свободного программного обеспечения Сизиф команды ALT Linux Team.

Основные функции репозитариев:

- парадигма включения/выключения и некоторые формальные процедуры для объектов;
- поддержка множественных версий объектов и процедуры управления конфигурациями для объектов;
- оповещение инструментальных и рабочих систем об интересующих их событиях;
- управление контекстом и разные способы обзора объектов репозитария; определение потоков работ.

1. Трехуровневое представление для описания предметной области

По отношению к пользователям применяют трехуровневое представление для описания предметной области: концептуальное, логическое и внутреннее (физическое).

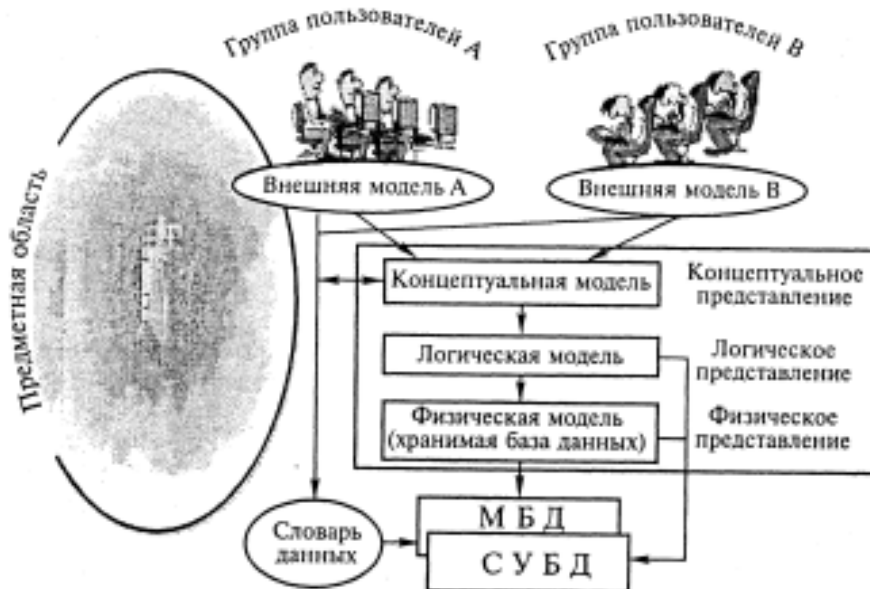


Рис. 1. Описание предметной области

Концептуальный уровень. Этот уровень характеризуется системным описанием предметной области на основе информационных потребностей пользователей ИС – инфологической моделью. Для представления инфологических моделей используются ER-диаграммы типа «сущность – связь» (модель Чена).

Логический уровень. На основе инфологической модели предметной области создается модель данных. Используются разновидности моделей данных: иерархические, сетевые и реляционные (табличные).

1). *Иерархическая модель* данных графически может быть представлена как перевернутое дерево, состоящее из объектов различных уровней. На рис.2. узлы и ветви образуют иерархическую древовидную структуру. Узел является совокупностью атрибутов, описывающих объект. Наивысший в иерархии узел называется корневым (это главный тип объекта). Корневой узел находится на первом уровне. Зависимые узлы (подчиненные типы объектов) находятся на втором, третьем и т. д. уровнях. Между главным и подчиненными объектами устанавливается взаимосвязь «один ко многим».

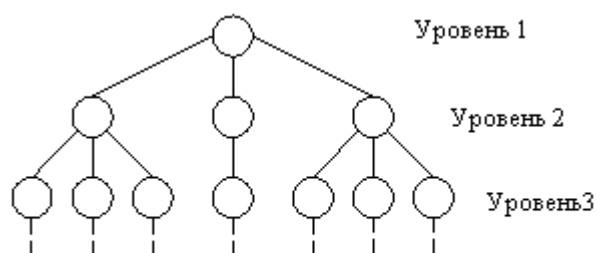


Рис. 2. Структура иерархической модели данных

2) *Сетевая модель*— логическая модель данных, являющаяся расширением иерархического подхода и использует для описания модель графов.

Разница между иерархической моделью данных и сетевой состоит в том, что в иерархических структурах запись-потомок должна иметь в точности одного предка, а в сетевой структуре данных у потомка может иметься любое число предков.

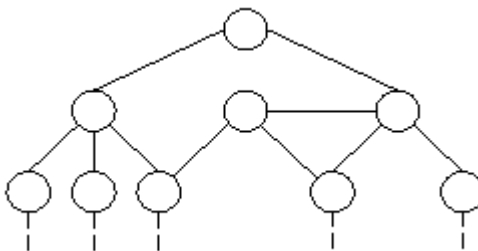


Рис.3. Структура сетевой модели данных

3) *Реляционная модель*

Основные идеи реляционной модели данных были предложены Эдгаром Коддом в 1969 г. Для обозначения родовой структуры Кодд стал использовать термин отношение (*relation*), а для обозначения элементов отношения — термин кортеж. Соответственно, модель данных получила название реляционной модели.

Реляционная модель данных — логическая модель данных, прикладная теория, описывающая структурный аспект, аспект целостности и аспект обработки данных в реляционных базах данных.

Реляционная база данных — база данных, основанная на реляционной модели данных. Для работы с реляционными БД применяют реляционные СУБД.

- Структурный аспект (составляющая) — данные в базе данных представляют собой набор отношений.

- Аспект целостности — отношения (таблицы) отвечают определенным условиям целостности. РМД поддерживает декларативные ограничения целостности уровня домена (типа данных), уровня отношения и уровня базы данных.

- Аспект (составляющая) обработки (манипулирования) — РМД поддерживает операторы манипулирования отношениями (реляционная алгебра, реляционное исчисление).

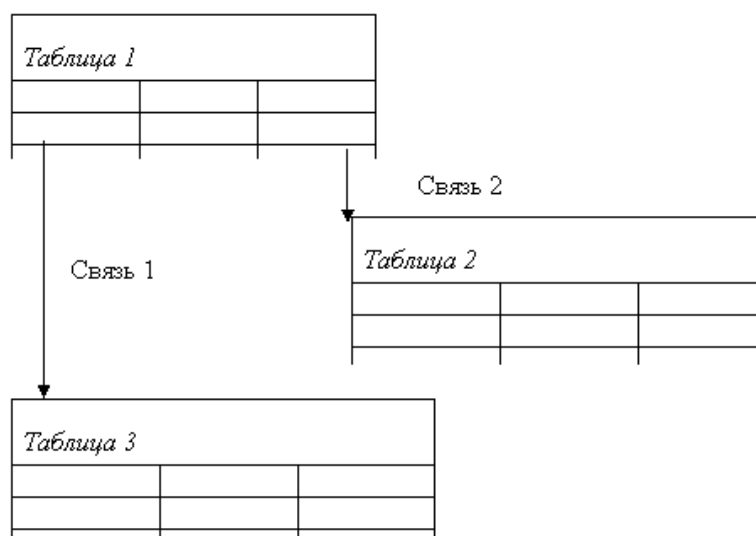


Рис.4. Структура реляционной модели данных

Отношения удобно представлять в виде таблиц.

Физический (внутренний) уровень связан со способом фактического хранения данных в физической памяти ЭВМ. Основными компонентами физического уровня являются хранимые записи, объединяемые в блоки; указатели, необходимые для поиска данных; данные переполнения; промежутки между блоками; служебная информация.

2. Классификация БД

Используется классификация баз данных по разным признакам. Первый признак классификации баз данных – **по содержанию хранимой информации**. **Фактографические БД** содержат данные, представляемые в краткой форме, в строго фиксированных форматах. Такие БД являются аналогами бумажных картотек, например библиотечного каталога или каталога видеотеки. Другой тип баз данных – **документальные БД**. Здесь аналогом являются архивы документов, например архив судебных дел, архив исторических документов и пр.

Классификация **по способу хранения данных** делит БД на **централизованные БД** и **распределенные БД**. Вся информация в централизованной БД хранится на одном компьютере. Это может быть автономный ПК или сервер сети, к которому имеют доступ пользователи-клиенты. Распределенные БД используются в локальных и глобальных компьютерных сетях. В таком случае разные части базы хранятся на разных компьютерах.

Третий признак классификации баз данных – **по структуре модели данных**. Известны три разновидности структур данных: **иерархическая**, **сетевая** и **табличная**. Соответственно по признаку структуры базы данных делятся на **иерархические БД**, **сетевые БД** и **реляционные (табличные) БД** (РБД).

3. Проектирование БД

Отдельные БД могут объединять все данные, необходимые для решения одной или нескольких прикладных задач, или данные, относящиеся к какой-либо предметной области (например, финансам, студентам, преподавателям, кулинарии и т.п.). Первые обычно называют **прикладными БД**, а вторые – **предметными БД** (соотносящимся с предметами организации, а не с ее информационными приложениями). (Первые можно сравнить с базами материально-технического снабжения или отдыха, а вторые – с овощными и обувными базами).

Предметные БД позволяют обеспечить поддержку любых текущих и будущих приложений, поскольку набор их элементов данных включает в себя наборы элементов данных прикладных БД. Вследствие этого предметные БД создают основу для обработки неформализованных, изменяющихся и неизвестных запросов и приложений (приложений, для которых невозможно заранее определить требования к данным). Такая **гибкость** и **адаптируемость** позволяет создавать **на основе предметных БД достаточно стабильные информационные системы**, т.е. системы, в которых большинство изменений можно осуществить без вынужденного переписывания старых приложений.

Основывая же проектирование БД на текущих и предвидимых приложениях, можно существенно ускорить создание **высокоэффективной** информационной системы, т.е. системы, структура которой учитывает наиболее часто встречающиеся пути доступа к данным. Поэтому прикладное проектирование до сих пор привлекает некоторых разработчиков. Однако по мере роста числа приложений таких информационных систем быстроувеличивается число прикладных БД, резко возрастает уровень дублирования данных и повышается стоимость их ведения.

Таким образом, каждый из рассмотренных подходов к проектированию воздействует на результаты проектирования в разных направлениях. Желание достичь и гибкости, и эффективности привело к формированию методологии проектирования, использующей как предметный, так и прикладной подходы. В общем случае предметный подход используется для построения первоначальной информационной структуры, а прикладной – для ее совершенствования с целью повышения эффективности обработки данных.

Другим важным аспектом проектирования БД является проблема избыточности информации.

Нормализация данных – одно из самых важных понятий и концепций реляционной системы. Нормализованная система сводит к минимуму количество избыточных данных, при этом сохраняя их целостность. Нормализованной можно назвать базу данных, в которой все таблицы следуют правилам нормальных форм. В большинстве случаев, вполне хватает следования первым трем нормальным формам:

4. Последовательность этапов проектирования БД

Проектирование БД можно объединить в четыре этапа. На этапе *формулирования и анализа требований* устанавливаются цели организации, определяются требования к БД. Эти требования документируются в форме, доступной конечному пользователю и проектировщику БД. Обычно при этом используется методика интервьюирования персонала различных уровней управления.

Этап *концептуального проектирования* заключается в описании и синтезе информационных требований пользователей в первоначальный проект БД. Результатом этого этапа является высокоуровневое представление информационных требований пользователей на основе различных подходов.

В процессе *логического проектирования* высокоуровневое представление данных преобразуется в структуре используемой СУБД. Полученная логическая структура БД может быть оценена количественно с помощью различных характеристик (число обращений к логическим записям, объем данных в каждом приложении, общий объем данных и т.д.).

На этапе *физического проектирования* решаются вопросы, связанные с производительностью системы, определяются структуры хранения данных и методы доступа.

Взаимодействие между этапами проектирования и словарной системой необходимо рассматривать отдельно. Процедуры проектирования могут использоваться независимо в случае отсутствия словарной системы.

Инструментальные средства проектирования и оценочные критерии используются на всех этапах разработки. Любой метод проектирования (аналитический, эвристический, процедурный), реализованный в виде программы, становится инструментальным средством проектирования.

Оценочные критерии принято делить на количественные и качественные. **Количественные критерии:** время, необходимое для ответа на запрос, стоимость модификации, стоимость памяти, время на создание, стоимость на реорганизацию. **Качественные критерии:** гибкость, адаптивность, доступность для новых пользователей, совместимость с другими системами, возможность конвертирования в другую вычислительную среду, возможность восстановления, возможность распределения и расширения.

Контрольные вопросы

1. Укажите отличия базы данных, хранилища данных, витрины данных, репозитария.
2. Какие модели используются для описания предметной области?
3. Какие модели используются на концептуальном уровне?
4. Какие модели используются на логическом уровне?
5. Какие модели используются на физическом уровне?
6. Дайте краткую характеристику основных типов баз данных.
7. Что такое СУБД и каковы ее стандарты?
8. Сформулируйте подходы к проектированию баз данных?