

Глава 6 Теория баз данных

- 6.1. Общие понятия
- 6.2. Модели данных
- 6.3. Реляционные базы данных
- 6.4. Постреляционные модели и базы данных
- 6.5. Проектирование баз данных

Современные информационные системы, основанные на концепции интеграции данных, характеризуются огромными объемами хранимых данных, сложной организацией, необходимостью удовлетворять разнообразные требования многочисленных пользователей.

Данная глава призвана сформировать представление о базах данных (БД), в ней рассказывается о возможностях систем управления базами данных (СУБД) и их использовании. Основные функциональные возможности и технологические операции при работе с СУБД рассматриваются без привязки к конкретному типу программного продукта. Знания, полученные при изучении данной главы, являются базовыми при практическом знакомстве с любым новым видом СУБД.

6.1. Общие понятия

Целью любой информационной системы является обработка данных об объектах реального мира. В широком смысле база данных — это совокупность сведений о конкретных объектах реального мира в какой-либо предметной области. Под *предметной областью* принято понимать часть реального мира, например, предприятие, вуз и т. д., подлежащую изучению с целью организации управления и автоматизации.

Рассмотрим несколько определений термина «база данных» (database).

База данных (БД) — это именованная совокупность данных, отражающая состояние объектов и их отношений в предметной области.

База данных — это набор постоянно хранимых данных, используемых прикладными программными системами предприятия.

База данных — это реализация схемы и модели данных на физическом уровне.

База данных — это объект управления в банке данных. База данных описывает состояние объектов предметной области на определенный момент времени совокупностью предложений на некотором формализованном языке.

База данных — это совокупность хранящихся вместе данных при наличии такой минимальной избыточности, которая допускает их использование оптимальным образом несколькими приложениями.

База данных — это совокупность материалов (статей, расчетов, нормативных актов, судебных решений и иных подобных материалов), систематизированных таким образом, чтобы эти материалы могли быть найдены и обработаны с помощью компьютера.

Все эти определения не являются противоречивыми или взаимоисключающими. Скорее, они представляют разные точки зрения авторов на одно и то же понятие. Сложность определения заключается в том, что компьютерные базы данных за свою не очень длинную историю прошли несколько этапов развития, от файловых систем, хранящих в себе «сырые» (неупорядоченные) данные, до постреляционных СУБД, содержимым которых являются данные, обладающие поведением (объекты). Остановимся на еще одном определении.

База данных — это информационная модель предметной области в виде совокупности данных, хранимых в памяти компьютера и связанных между собой по правилам, которые определяют их общие принципы описания, хранения и манипулирования.

Под *информационной моделью* понимают информацию об объекте, отобранную и структурированную в соответствии с заданной целью.

Модель данных — это описание методов представления и обработки данных в СУБД, в том числе методов определения типов и логических структур в базе данных, методов манипулирования данными и методов определения и поддержки целостности базы данных.

Исторически первые базы данных создавались на основе файловых систем, и вся ответственность за работу с ними возлагалась на прикладное программное обеспечение, использовавшее эти базы. Файловые базы данных сегодня практически не применяются. В современной технологии баз данных предполагается, что создание базы данных, ее поддержка и обеспечение доступа пользователей к ней осуществляются централизованно с помощью специального программного инструментария — системы управления базами данных.

Система управления базами данных (СУБД) — это комплекс языковых и программных средств, предназначенный для создания, ведения и применения базы данных пользователями.

Кроме базы данных и программного обеспечения, обеспечивающего основную функциональность СУБД, в состав современных серверов баз данных входят всевозможные средства разработки и механизмы взаимодействия с пользователем на высоком уровне (генераторы отчетов, конструкторы таблиц, построители запросов и форм). Эти средства разработки, сами являясь *приложениями* пользователя, позволяют создавать *приложения*, функционирующие как часть СУБД (например, формы и отчеты MS Access или веб-публикации в Oracle и MS SQL Server).

На рис. 6.1 представлена схема, в которой определены основные термины, используемые при обсуждении СУБД.



Рис. 6.1. Терминология СУБД

6.1.1. Компоненты среды функционирования СУБД

СУБД представляет собой комплекс программных средств, в работе которого принимает участие множество людей, как обслуживающих эти программы, так и использующих результат их работы. На рис. 6.2 представлены основные компоненты СУБД.

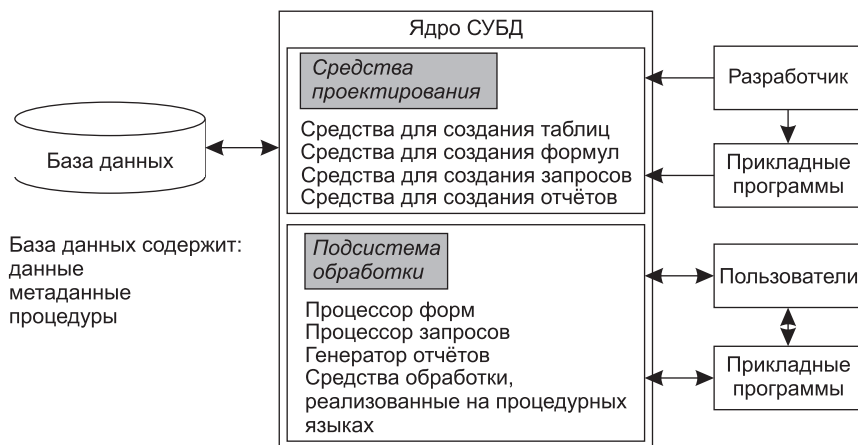


Рис. 6.2. Компоненты среды функционирования СУБД

Программное обеспечение

К программному обеспечению относятся все компьютерные программы, используемые в работе системы управления базами данных. Для выполнения всех функций СУБД требуется программное обеспечение трех видов: системное программное обеспечение, программное обеспечение СУБД, а также прикладные программы и утилиты. Поскольку программное обеспечение СУБД функционально располагается между системным и приложениями пользователя, его относят к разряду промежуточного (middleware) программного обеспечения.

Системное программное обеспечение управляет всеми компонентами оборудования и обеспечивает доступ к нему всех остальных приложений, работающих на компьютере. Примеры системного программного обеспечения: Windows, Linux, UNIX, MVS, MacOS, OpenSolaris и др.

Подсистема обработки СУБД управляет базой данных, реализуя функции СУБД.

Средства проектирования СУБД предназначены для получения доступа к данным и манипулирования ими в среде СУБД. Прикладные программы (приложения пользователя) в большинстве случаев служат для представления данных, хранящихся в БД, в виде отчетов и таблиц.

Люди

Сюда относятся все пользователи системы управления базой данных. Если взять за основу функциональные обязанности, то в системе управления базами данных можно выделить шесть основных групп пользователей: системные администраторы, администраторы баз данных, системные аналитики, проектировщики баз данных, программисты и конечные пользователи.

Системные администраторы несут ответственность и обеспечивают надежное функционирование системного программного обеспечения.

Администраторы баз данных (Data Base Administrator, DBA) управляют работой СУБД, обеспечивают функционирование СУБД, создают учетные записи пользователей СУБД, назначают права, ограничивают доступ, выполняют различные процедуры, связанные с обеспечением безопасности и надежности хранения данных.

Системные аналитики выполняют работу по сбору, систематизации и уточнению требований к структуре данных, приложениям и отчетам.

Проектировщики базы данных (системные архитекторы) проектируют структуру БД.

Программисты разрабатывают прикладное программное обеспечение. Они проектируют и создают формы ввода и отображения данных, отчеты и процедуры, с помощью которых конечные пользователи получают доступ к данным и возможность манипулирования ими.

Конечные пользователи применяют прикладные программы с целью выполнения ежедневных операций, например, в компании — это продавцы, заведующие складами, работники бухгалтерии, руководители и управляющие. Конечные пользователи высшего руководящего звена применяют информацию, полученную из базы данных, для решения тактических и стратегических задач предприятия.

База данных

База данных включает в себя данные, метаданные и процедуры.

Данные. Под терминами «данные», «информация» или «сведения» в данном контексте понимается весь фактический материал, хранящийся в базе данных. Данные являются необработанным сырьем, которое подлежит соответствующему структурированию. Принятие решения о том, какую информацию поместить в БД, каким образом ее упорядочить и структурировать, является важнейшей частью работы системных архитекторов (проектировщиков) базы данных.

Метаданные составляют содержимое системного каталога базы данных и представляют собой сведения об именах и структуре таблиц, именах и правах пользователей, наименовании и типах ограничений, о процедурах, функциях и других объектах базы данных.

Процедуры являются важным компонентом системы. Они устанавливают стандарты ведения коммерческой, технологической и производственно-технической деятельности в рамках предприятия и в отношениях с клиентами. Процедуры также используются для организации наблюдения и аудита как за вводимой в БД информацией, так и за информацией, порождаемой на основе извлекаемых данных.

6.1.2. Классификация СУБД

Классификация по типу принятой модели данных

Классификацию баз данных по модели данных иллюстрирует рис. 6.3.

Иерархические базы данных основаны на иерархической модели данных, в которой связь между объектами базы данных образует перевернутое дерево. При такой модели каждый нижележащий элемент иерархии соединен только с одним расположенным выше элементом

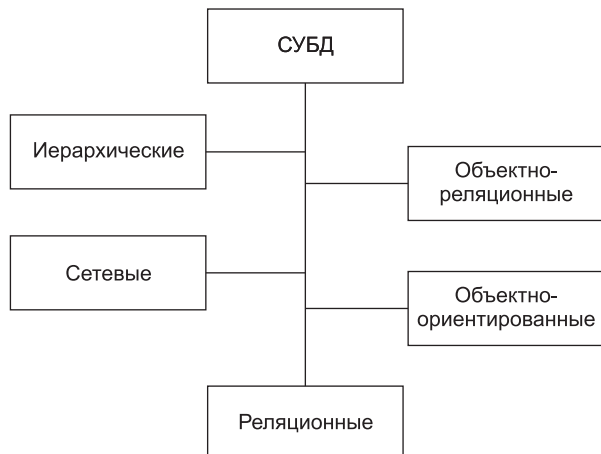


Рис. 6.3. Классификация баз данных по модели данных

Сетевые базы данных основаны на сетевой модели данных, в которой связи между объектами данных могут быть установлены в произвольном порядке.

Реляционные базы данных основаны на реляционной модели данных, в которой каждая единица данных в базе данных однозначно определяется именем таблицы (называемой отношением), идентификатором записи (кортежа) и именем поля.

Объектно-реляционные базы данных содержат объектно-ориентированные механизмы построения структур данных (как минимум, механизмы наследования и поддержки методов) в виде расширений языка и программных надстроек над ядром СУБД.

Объектно-ориентированные базы данных определяют как новое поколение баз данных, основанное на сочетании трех принципов: реляционной модели, стандартов на описание объектов и принципов объектно-ориентированного программирования.

Классификация по архитектуре

Классификацию баз данных по архитектуре иллюстрирует рис. 6.4.



Рис. 6.4. Классификация баз данных по архитектуре

В *локальных* базах данных все данные и объекты СУБД находятся на одном компьютере.

В *распределенных* базах данных различные части данных (группы таблиц, таблицы и даже фрагменты таблиц) и объекты СУБД могут находиться на разных компьютерах.

Пример. В качестве примера можно привести сложное производство (или сеть супермаркетов), разные части которого находятся в разных городах. Каждое предприятие накапливает «свои» данные. Необходимо, чтобы каждое из предприятий имело доступ к одним и тем же данным, как своим, так и данным других предприятий. Решением данной проблемы может быть создание одной локальной базы данных на одном компьютере с механизмом удаленного доступа. Однако это решение нерационально, поскольку быстрый доступ к данным будут получать клиентские компьютеры только того предприятия, на котором находится СУБД. Другим решением данной проблемы может быть создание на каждом предприятии своей копии СУБД. В этом случае возникает затруднение с синхронизацией данных между копиями (особенно в масштабах нашей страны, где в Хабаровске может быть разгар рабочего дня, а в Москве — глубокая ночь). Распределенная СУБД в этом случае обеспечивает механизм хранения данных в разных базах данных таким образом, что при обращении совокупность разных баз данных выглядит как одна база. Тогда часто используемые данные («свои» данные) находятся в той части базы данных, которая расположена на предприятии. А при необходимости обратиться к «чужим» данным, СУБД делает запрос к удаленной СУБД и получает данные оттуда. Совокупность разных баз данных на разных компьютерах с точки зрения клиента выглядит как одна база данных.

Классификация по способу доступа к БД

Классификацию баз данных по способу доступа иллюстрирует рис. 6.5.

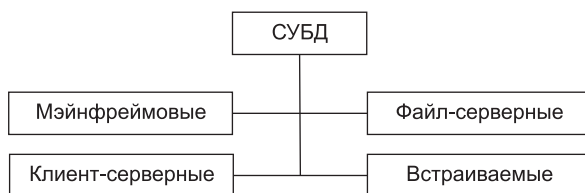


Рис. 6.5. Классификация баз данных по способу доступа

В *мэйнфреймовых* базах данных пользовательское рабочее место представляет собой текстовый или графический терминал, а вся информация обрабатывается на том же компьютере, где находится СУБД.

В *файл-серверных* СУБД файлы данных располагаются централизованно на файл-сервере, а ядро СУБД находится на каждом клиентском компьютере. Доступ к данным осуществляется через локальную сеть. Синхронизация чтений и обновлений осуществляется посредством файловых блокировок. Преимуществом этой архитектуры является низкая нагрузка на ЦП сервера, а недостатком — высокая нагрузка локальной сети.

Клиент-серверные СУБД состоят из клиентской части (которая входит в состав прикладной программы) и сервера. Клиент-серверные СУБД, в отличие от файл-серверных, обеспечивают разграничение доступа между пользователями и мало загружают сеть и клиентские машины. Сервер является внешней по отношению к клиенту программой, и при необходимости его можно заменить другим. Недостаток клиент-серверных СУБД состоит в самом факте существования сервера (что

плохо для локальных программ — в них удобнее встраиваемые СУБД) и больших вычислительных ресурсах, потребляемых сервером.

Встраиваемая СУБД представляет собой программную библиотеку, которая позволяет унифицированным образом хранить большие объемы данных на локальной машине. Доступ к данным может происходить посредством запросов на языке SQL либо путем вызова функций библиотеки из приложения пользователя. Встраиваемые СУБД быстрее обычных клиент-серверных и не требуют развертывания сервера.

Классификация по скорости обработки информации

Классификацию баз данных по скорости обработки информации иллюстрирует рис. 6.6.

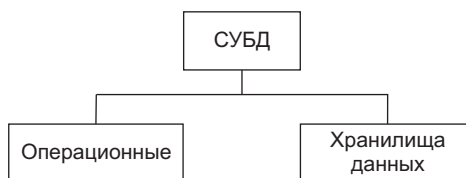


Рис. 6.6. Классификация баз данных по скорости обработки информации

Операционные (operational), или *рабочие* (production), базы данных обладают высокими скоростями реакции на запрос, извлечения и представления информации.

Хранилища данных и *многомерные хранилища данных* (data warehouse, OLAP) — это базы данных с очень большим объемом информации, подготовка представления которой занимает значительный объем времени.

6.1.3. Функции СУБД

Абстракция данных, управление словарем данных. Функционирование СУБД предусматривает, что определения элементов данных и их отношений (метаданные) хранятся в *словаре данных* (data dictionary). В свою очередь любые программы получают доступ к данным посредством СУБД. Для поиска необходимых структур данных и их отношений СУБД использует словарь данных, помогая избежать кодирования таких сложных взаимосвязей в каждой программе. Вдобавок любые изменения, которые делаются в структуре базы данных, автоматически регистрируются в словаре данных, что также освобождает программиста от необходимости модифицировать программы доступа к изменившимся структурам данных. СУБД обеспечивает абстракцию данных, тем самым устраняя в системе структурную зависимость и зависимость по данным.

Управление хранением данных. СУБД создает сложные структуры, необходимые для хранения данных, освобождая программистов от определения и программирования физических свойств данных. Современные СУБД обеспечивают хранение

не только данных, но и связанных с данными экранных форм, схем отчетов, правил проверки данных, кода процедур, систем обработки мультимедиа, форматов изображений, и т. п.

Преобразование и представление данных. СУБД берет на себя задачу структурирования вводимых данных, преобразуя их в форму, удобную для хранения. Поэтому СУБД и в данном случае избавляет человека от рутинной работы по преобразованию логического формата данных в физический формат. Обеспечивая независимость данных, СУБД преобразует логические запросы в команды, определяющие их физическое местоположение и извлечение. Таким образом, СУБД обеспечивает *программную независимость и абстракцию данных*.

Управление безопасностью. СУБД создает систему безопасности, которая обеспечивает защиту пользователя и конфиденциальность данных внутри БД. Правила безопасности устанавливают, какие пользователи могут получить доступ к базе данных, к каким элементам данных пользователь может получить доступ, какие операции с данными (чтение, добавление, удаление или изменение) может выполнять пользователь.

Управление многопользовательским доступом. СУБД создает сложные структуры, обеспечивающие доступ к данным нескольких пользователей одновременно. Для того чтобы обеспечить целостность и непротиворечивость данных, в СУБД применяются сложные алгоритмы, гарантирующие, что несколько пользователей могут получить одновременный доступ к базе данных без риска нарушить ее целостность.

Управление резервным копированием и восстановлением. В СУБД имеются процедуры резервного копирования и восстановления данных, обеспечивающие их безопасность и целостность. Современные СУБД содержат специальные утилиты, с помощью которых администраторы базы данных могут выполнять регулярные и экстренные процедуры резервного копирования и восстановления данных. Восстановление данных производится после повреждения БД, например, в случае появления сбойного сектора на жестком диске или после аварийного отключения питания. Такая возможность необходима для обеспечения целостности данных.

Управление целостностью данных. В СУБД предусмотрены правила, обеспечивающие целостность данных, что позволяет минимизировать избыточность данных и гарантировать их непротиворечивость. Для обеспечения целостности данных используются их связи, которые хранятся в словаре данных.

Поддержка языка доступа к данным и интерфейсов прикладного программирования. СУБД обеспечивает доступ к данным при помощи языка запросов. *Язык запросов* — это непроцедурный язык, то есть он предоставляет пользователю возможность определить, что необходимо выполнить, не указывая, как это сделать. В состав языка запросов СУБД входят два основных компонента: *язык определения данных* (Data Definition Language, *DDL*) и *язык манипулирования данными* (Data Manipulation Language, *DML*). *DDL* определяет структуры, в которых размещаются данные, а *DML* позволяет конечным пользователям извлекать данные из