**Проектирование реляционных баз данных**

**Содержание**

1. Цель работы.
2. Последовательность выполнения задания.
3. Краткие теоретические сведения.
3.1. Этапы разработки базы данных
3.2. Концептуальное моделирование данных
	* анализ предметной области. Основные понятия концептуальной модели данных.
	* построение ER-диаграмм

3.3. Логическое моделирование данных

* + основные понятия реляционной модели данных
	+ способы построения логической модели данных.
1. Содержание отчёта
Приложение1. Варианты заданий к контрольной работе.
Список литературы

***1. Цель работы***

Целью выполнения контрольной работы по курсу “Банки и базы данных” является:

* изучение этапов проектирования реляционных баз данных;
* приобретение практических навыков в разработке и реализации информационных систем;
* приобретение навыков работы с реляционными базами данных.

***2. Задание****:*

По заданному в варианте описанию предметной области разработать и реализовать проект реляционной базы данных

**Последовательность выполнения задания**

1. Разработка концептуальной модели данных
	1. Анализ предметной области; выделение объектов, информация о которых должна храниться в базе данных, определение их атрибутов, связей между объектами и характеристик этих связей
	2. Построение ЕR – диаграмм
2. Разработка логической модели данных
	1. Создание таблиц, определение типов данных в каждом поле, ограничений на установка диапазонов допустимых значений, первичных ключей.
	2. Определение внешних ключей и связей между таблицами.
3. Разработка пользовательского интерфейса. Создание запросов и представлений, необходимых пользователю при работе с базой данных. Проектирование форм для просмотра, ввода, редактирования и поиска данных в базе данных. Разработка отчётов для вывода нужных данных на печать.
4. Реализация проекта базы данных средствами выбранной СУБД.

**Примечание**: пункты 3 и 4 задания выполняются на лабораторном практикуме. Для их выполнения используйте методические указания к лабораторному практикуму.

***3.Краткие теоретические сведения.***

**3.1. Этапы разработки базы данных. Многоуровневое представление данных.**

Целью разработки любой базы данных является хранение и использование информации о какой-либо предметной области.

При разработке базы данных обычно выделяется несколько уровней моделирования, при помощи которых происходит переход от предметной области к реализации базы данных средствами конкретной СУБД. Можно выделить следующие уровни:

* Сама предметная область
* Модель предметной области
* Концептуальная модель данных
* Логическая модель данных
* Физическая модель данных
* Собственно база данных и приложения

***Предметная область*** - это часть реального мира, данные о которой необходимо отразить в базе данных. Например, в качестве предметной области можно выбрать бухгалтерию какого-либо предприятия, отдел кадров, банк, магазин и т.д. Предметная область бесконечна и содержит как существенно важные понятия и данные, так и малозначащие или вообще не значащие данные. Так, если в качестве предметной области выбрать учет товаров на складе, то понятия "накладная" и "счет-фактура" являются существенно важнымипонятиями, а то, что сотрудница, принимающая накладные, имеет двоих детей - это для учета товаров неважно. Однако, с точки зрения отдела кадров данные о наличии детей являются существенно важными. Таким образом, важность данных зависит от выбора предметной области.

***Модель предметной области***. Модель предметной области - это наши знания о ней. Знания могут быть как в виде неформальных знаний в мозгу эксперта, так и выражены формально при помощи каких-либо средств. В качестве таких средств могут выступать текстовые описания предметной области, наборы должностных инструкций, правила ведения дел в компании и т.п. Опыт показывает, что текстовый способ представления модели предметной области неэффективен. Гораздо более информативными и полезными при разработке баз данных являются описания предметной области, выполненные при помощи специализированных графических схем. Модель предметной области описывает скорее процессы, происходящие в предметной области и данные, используемые этими процессами. От того, насколько правильно смоделирована предметная область, зависит успех дальнейшей разработки приложений.

***Концептуальная модель данных***. На следующем, более низком уровне находится концептуальная модель данных предметной области. Концептуальная модель описывает понятия предметной области, их взаимосвязь, а также ограничения на данные, налагаемые предметной областью. Примеры понятий - "сотрудник", "отдел", "проект", "зарплата". Примеры взаимосвязей между понятиями - "сотрудник числится ровно в одном отделе", "сотрудник может выполнять несколько проектов", "над одним проектом может работать несколько сотрудников". Примеры ограничений - "возраст сотрудника не менее 16 и не более 60 лет".

Концептуальная модель данных является начальным прототипом будущей базы данных. Эта модель строится в терминах информационных единиц, но *без привязки к конкретной СУБД*. Основным средством разработки концептуальной модели данных в настоящий момент являются различные варианты ***ER-диаграмм***(***Entity-Relationship***, ***диаграммы сущность-связь***).

***Логическая модель данных –***это данные, представленные на языке описания данных конкретной СУБД.
Логическая модель данных включает в себя следующие составляющие:

* структура данных;
* ограничения, накладываемые на данные
* операции, производимые над данными.

Одну и ту же ER-модель можно преобразовать как в реляционную модель данных, так и в модель данных для иерархических и сетевых СУБД, или в постреляционную (объектно- ориентированную модель данных.

Наиболее распространённая модель данных, используемая большинством СУБД – ***реляционная модель.*** Поэтому можно считать, что логическая модель данных для нас формулируется в терминах реляционной модели данных.

***Физическая модель данных***. На еще более низком уровне находится физическая модель данных. Физическая модель данных описывает хранение данных средствами конкретной СУБД. Ограничения, имеющиеся в логической модели данных, реализуются различными средствами СУБД. При этом решения, принятые на уровне логического моделирования определяют некоторые границы, в пределах которых можно развивать физическую модель данных. Качество физической модели во многом зависит от выбора СУБД.

***Собственно база данных и приложения***. И, наконец, как результат предыдущих этапов появляется собственно сама база данных. База данных реализована на конкретной программно-аппаратной основе, и выбор этой основы позволяет существенно повысить скорость работы. Например, можно выбирать различные типы компьютеров, менять количество процессоров, объем оперативной памяти, дисковые подсистемы и т.п. Очень большое значение имеет также настройка СУБД в пределах выбранной программно-аппаратной платформы.

Многоуровневая архитектура (концептуальный, логический и физический уровни) позволяет обеспечить *независимость хранимых данных* от использующих их программ. При необходимости можно переписать хранимые данные на другие носители информации и (или) реорганизовать их физическую структуру, изменив лишь физическую модель данных. Можно подключить к системе любое число новых пользователей (новых приложений), дополнив, если надо, логическую модель. Указанные изменения физической и логической моделей не будут замечены существующими пользователями системы (окажутся "прозрачными" для них), так же как не будут замечены и новые пользователи. Следовательно, независимость данных обеспечивает возможность развития системы баз данных без разрушения существующих приложений.

Решения, принятые на каждом этапе моделирования и разработки базы данных, будут сказываться на дальнейших этапах. Поэтому особую роль играет принятие правильных решений *на ранних этапах проектирования*.

**3.2. Концептуальное моделирование данных**

. Одна из наиболее распространённых концептуальных моделей данных - модель "Сущность-Связь" (часто ее называют кратко ER-моделью). На использовании разновидностей ER-модели основано большинство современных подходов к проектированию реляционных баз данных. Основными понятиями ER-модели являются сущность, связь и атрибут.

***Сущность*** – это класс однотипных объектов, информация о котором должна сохраняться и быть доступна. Каждая сущность должна иметь имя, выраженное существительным в единственном числе. Примерами сущностей могут быть такие классы объектов как "Поставщик", "Сотрудник", "Накладная". В ER-диаграммах сущность изображается в виде прямоугольника, содержащего имя сущности (рисунок 1).



Рисунок 1 – Изображение сущности в ER-диаграмме

***Экземпляр сущности*** – это конкретный представитель данной сущности. Например, представителем сущности "Сотрудник" может быть "Сотрудник Иванов". Экземпляры сущностей должны быть *различимы*, т.е. сущности должны иметь некоторые свойства, уникальные для каждого экземпляра этой сущности.

***Атрибут сущности*** – это именованная независимая характеристика, являющаяся некоторым свойством сущности. Наименование атрибута должно быть выражено существительным в единственном числе (возможно, с характеризующими прилагательными). Примерами атрибутов сущности "Сотрудник" могут быть такие атрибуты как "Табельный номер", "Фамилия", "Имя", "Отчество", "Должность", "Зарплата" и т.п. . Атрибут изображается в виде эллипса, содержащего имя атрибута (рисунок 2).



Рисунок 2 – Изображение атрибутов сущности на ER-диаграмме

***Связь*** – это ассоциация, поясняющая отношения между сущностями. Например, связи между сущностями могут выражаться следующими фразами – "СОТРУДНИК может *иметь*несколько ДЕТЕЙ", "каждый СОТРУДНИК обязан *числиться* ровно в одном ОТДЕЛЕ", и т.д. Графически связь изображается линией, соединяющей две сущности. Каждая связь имеет два конца и одно или два наименования, например "иметь", "принадлежать" и т.п. Каждое из наименований относится к своему концу связи. Иногда наименования не пишутся ввиду их очевидности.

Каждая связь может иметь один из следующих ***типов связи по значности***:

Связь ***один к одному*** означает, что одному экземпляру первой сущности соответствует один экземпляр второй сущности (рисунок 3).

Связь ***один ко многим***(1:m) означает, что одному экземпляру первой сущности со- ответствует несколько экземпляров второй сущности, но не наоборот (рисунок 4).

Связь ***многие ко многим*** (m:m) означает, что одному экземпляру первой сущности соответствует несколько экземпляров второй сущности и наоборот (рисунок 5).

Каждая связь может иметь один из следующих ***типов связи по членству***:

* ***Обязательная*** связь означает, что обе сущности зависят от наличия связи. Т.е, экземпляр одной сущности *обязан быть связан не менее чем с одним* экземпляром другой сущности, и наоборот;
* ***Необязательная*** связь означает, что ни одна из сущностей не зависит от наличия связи. Т.е, экземпляр одной сущности *может быть связан* с одним или несколькими экземплярами другой сущности, *а может быть и не связан* ни с одним экземпляром.
* ***Возможная***связь означает, что одна из сущностей зависит от наличия связи. Т.е. один конец связи необязательный, а другой- обязательный, например: "каждый сотрудник *может* иметь несколько детей", и "каждый ребенок *обязан* принадлежать ровно одному сотруднику" (рис.6)

При разработке ER-моделей необходимо получить следующую информацию о предметной области:

1. Список сущностей предметной области.
2. Список атрибутов сущностей.
3. Описание взаимосвязей между сущностями.

***Пример***: Необходимо разработать информационную систему по заказу оптовой торговой фирмы, которая должна выполнять следующие действия:

* Хранить информацию о покупателях.
* Печатать накладные на отпущенные товары.
* Следить за наличием товаров на складе.

Анализируя эту информацию, выделим сущности:

* *Покупатель;*
* *Накладная;*
* *Товар;*
* *Склад*(если фирма имеет несколько складов).

Каждый покупатель является юридическим лицом и имеет наименование, адрес, банковские реквизиты. Каждый товар имеет наименование, цену, а также характеризуется единицами измерения. Каждая накладная имеет уникальный номер, дату выписки, список товаров с количествами и ценами, а также общую сумму накладной. Накладная выписывается с определенного склада и на определенного покупателя. Каждый склад имеет свое наименование. Исходя из этих данных, выпишем атрибуты сущностей:

* *Покупатель: Наименование покупателя, Адрес*,*Банковские реквизиты*;
* *Товар: Наименование товара, Цена товара*, *Единица измерения*;
* *Склад: Наименование склада;*
* *Накладная: Номер накладной, Дата выписки, Склад, Покупатель;*

Поскольку список товаров в накладной не может быть атрибутом, выделим его в отдельную сущность:
*Список товаров в накладной: Наименование товара, Цена товара в накладной, Количество товара в накладной.*Сумма накладной равна сумме стоимостей всех товаров, входящих в накладную. Эта характеристика не является независимой, и не может считаться атрибутом

Определим связи между сущностями. Покупатели могут покупать много товаров, и товары могут продаваться многим покупателям. Если фирма имеет несколько складов, то каждый товар может храниться на нескольких складах. Покупатели покупают товары, получая при этом накладные, в которые внесены данные о количестве и цене купленного товара. Каждый покупатель может получить несколько накладных. Каждая накладная *обязана* выписываться на одного покупателя. Каждая накладная *обязана*содержать не менее одного товара (не бывает пустых накладных). Каждый товар может быть продан нескольким покупателям через несколько накладных. Кроме того, каждая накладная должна быть выписана с определенного склада, и с любого склада может быть выписано много накладных. Один из вариантов ER-диаграммы приведён на рисунке 6.

**3.2.Логическое моделирование данных**

Наиболее распространённой логической моделью на сегодняшний день является **реляционная модель**.Основы реляционной модели данных были впервые изложены в статье Е.Кодда [43] в 1970 г. Наиболее распространенная трактовка реляционной модели данных принадлежит К.Дейту. []. Согласно Дейту, реляционная модель состоит из трех частей:

* Структуры данных.(структурная часть)
* Ограничений, накладываемых на данные (целостная часть).
* Набора допустимых операций над данными (манипуляционная часть).

**3.2.2. Структура реляцинных данных**

Единственной структурой данных, используемой в реляционной модели, является ***отношение*** (relation). Отношение представляет собой связь между элементами нескольких множеств атомарных однотипных значений, именуемых ***доменами***. Говорят, что значения принадлежат к одному и тому же домену, если имеет смысл их сравннение. Например, фамилию сотрудника можно сравнить с фамилией другого человека, но нет смысла сравнивать её с наименованием товара. ***Атрибутом отношения*** называют набор значений, принадлежащих к одному и тому же домену. Например: атрибут “Фамилия сотрудника” взят из домена фамилий.

**Сотрудник**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Фамилия сотрудника | Имя сотрудника | Отчество сотрудника | Должность | Оклад |
| ИвановПетров | ИванПетр | ИвановичПетрович | СлесарьЭлектрик | 10002000 |

Рисунок 7 – Пример реализации отношения “Сотрудник”

1. Отношения реализуются в виде двумерных таблиц (рисунок 7), обладающих следующими свойствами:
2. Каждая таблица состоит из однотипных строк и имеет уникальное имя.
3. В каждой позиции таблицы на пересечении строки и столбца может содержаться только одно значение.
4. Строки таблицы обязательно отличаются друг от друга хотя бы одним значением.
5. Столбцам таблицы однозначно присваиваются имена, и в каждом из них размещаются однородные значения данных (например: даты, фамилии, наименования, целые числа или денежные суммы).
6. При выполнении операций с таблицей ее строки и столбцы можно обрабатывать в любом порядке, независимо от содержания.

***Реляционная база данных*** представляет собой совокупность взаимосвязанных таблиц, содержащий всю информацию, которую необходимо хранить и обрабатывать ***Схемой базы данных*** называют список, содержащий имена таблиц, имена атрибутов таблиц, ключевые атрибуты и внешние ключи.

**3.2.3. Ограничения целостности**

Во второй части реляционной модели данных определяются два ограничения, которые должны выполняться в *любой*реляционной базе данных. Это:

* Целостность сущностей.
* Целостность внешних ключей.

*Определение 1*. Пусть дано отношение . Подмножество атрибутов отношения будем называть ***потенциальным ключом***, если обладает следующими свойствами:

1. ***Свойством уникальности***- в отношении не может быть двух различных кортежей, с одинаковым значением .
2. ***Свойством минимальности***- никакое подмножество в не обладает свойством уникальности.

Любое отношение имеет по крайней мере один потенциальный ключ. Действительно, если никакой атрибут или группа атрибутов не являются потенциальным ключом, то, в силу уникальности кортежей, все атрибуты вместе образуют потенциальный ключ. Потенциальный ключ, состоящий из одного атрибута, называется ***простым***, а состоящий из нескольких атрибутов - ***составным***. Отношение может иметь несколько потенциальных ключей. Традиционно, один из потенциальных ключей объявляется ***первичным***, а остальные -***альтернативными***. Т.к. потенциальные ключи служат идентификаторами объектов предметной области (т.е. предназначены для *различения*объектов), то значения этих идентификаторов не могут содержать неизвестные значения. Действительно, если бы идентификаторы могли содержать null-значения, то мы не могли бы дать ответ "да" или "нет" на вопрос, совпадают или нет два идентификатора.
Это определяет следующее ***правило целостности сущностей***:
*Правило целостности сущностей*. Атрибуты, входящие в состав некоторого потенциального ключа не могут принимать неопределённых (null)-значений.
Различные объекты предметной области, информация о которых хранится в базе данных, всегда взаимосвязаны друг с другом. Такие взаимосвязи отражаются в реляционных базах данных при помощи ***внешних ключей***, связывающих несколько отношений.
Т.к. внешние ключи фактически служат ссылками на строки в другом (или в том же самом) отношении, то эти ссылки не должны указывать на несуществующие объекты. Это определяет следующее ***правило целостности внешних ключей***:
*Правило целостности внешних ключей*. Внешние ключи не должны быть несогласованными, т.е. для каждого значения внешнего ключа таблицы должно существовать соответствующее значение первичного ключа в связанной с ней таблице.

**3.2.4. Способы построения логической модели данных**

При разработке логической модели данных возникают вопросы: хорошо ли спроектированы отношения? Правильно ли они отражают модель предметной области, а следовательно и саму предметную область?
Рассмотрим некоторые из таких критериев, которые являются безусловно важными с точки зрения получения качественной базы данных:

* Адекватность базы данных предметной области
* Легкость разработки и сопровождения базы данных
* Скорость выполнения операций обновления данных (вставка, обновление, удаление кортежей)
* Скорость выполнения операций выборки данных

При разработке схемы базы данных необходимо выполнить следующие условия:

1. Информация в таблицах не должна повторятся (не должно быть избыточности). Избыточность приводит к проблемам при поиске и обработке данных. Поэтому желательно, чтобы каждый факт хранился только в одном месте.
2. Поля таблиц по возможности не должны принимать неопределённых (пустых) значений. Неопределённые значения могут привести к ошибкам при выполнении вычислений над теми полями таблиц, в которых они встречаются.

Очевидно, что эти условия не будут выполняться, если разместить в одной таблице разнородную информацию о нескольких объектах сразу. На рисунке 8 приведён пример универсального отношения “Сотрудник”.

**Сотрудник**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Код сотрудника | Фамилия сотрудника | Возрастдетей сотрудников | НазваниеОтдела | Должность | Оклад |
| 11234 | ИвановИвановПетровСидоровВасильев | 101238 | ОГЭОГЭОГЭОГМОГМ | СлесарьСлесарьСлесарьЭлектрикЭлектрик | 10001000100030003000 |

Рисунок 8– Пример универсального отношения, содержащего аномалии

Поскольку в одном отделе может числиться несколько сотрудников, наименование отдела приходится повторить несколько раз. Кроме того, повторяются фамилии сотрудников, имеющих более одного ребёнка. Не все сотрудники имеют детей, поэтому атрибут “Возраст детей сотрудников” может принимать неопределённые значения. Избыточность и неопределённость данных приводят к проблемам при их вводе, обновлении и удалении. Например, при переводе сотрудника Иванова на другую должность необходимо изменить значения атрибутов “Должность” и “Оклад” дважды. При удалении данных о последнем сотруднике отдела ОГЭ будет удалена и информация о том, что такой отдел существует. Ошибки в структуре данных, приводящие к проблемам при их обработке, называют **аномалиями**. Решением проблемы будет разделение таблицы “Сотрудник” на несколько таблиц. При этом атрибуты, содержащие повторяющиеся или неопределёные значения, выносятся в отдельные таблицы, связанные с исходной таблицей при помощи внешних ключей (рисунок 10).

Сотрудник                                                                Дети

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Код сотрудника | Фамилия | Код должности | Код отдела |   | Код сотрудника | Возраст детей |
| 1 | Иванов | 1 | 1 |   | 1 | 10 |
| 2 | Петров | 1 | 1 |   | 1 | 12 |
| 3 | Сидоров | 2 | 2 |   | 2 | 3 |
| 4 | Васильев | 2 | 2 |   | 4 | 8 |

Отдел                                                             Должность

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Код отдела | Название отдела |   | Код должности | Должность | Оклад |
| 1 | ОГЭ |   | 1 | Слесарь | 1000 |
| 2 | ОГМ |   | 2 | Электрик | 3000 |

Рисунок 9 – Примеры нормализованных таблиц

Процесс разделения таблиц с целью улучшения их свойств называется декомпозицией, или ***нормализацией,*** а полученные в результате таблицы – ***нормализованными***.
Заметим, что после завершения нормализации большинство таблиц содержит информацию только об одном объекте или явлении (сущности). Поэтому наиболее простым способом создания системы нормализованных таблиц является получение её из диаграммы “Сущность – связь”. Перехода от ER-диаграммы к таблицам состоит из следующих шагов:

1. Преобразование сущностей.

1. Каждая простая сущность становится таблицей.
2. Каждый атрибут сущности становится атрибутом (столбцом) таблицы.
3. Уникальный идентификатор сущности становится ключом таблицы.
4. Если в ER-диаграмме присутствовали подтипы сущности, они выносятся в отдельные таблицы.

2. Преобразование связей

1. Сущности, связанные обязательной связью типа 1:1,можно объединить в одну таблицу.
2. Связи типа 1:1 (возможные) и связи типа 1:m реализуются путем переноса ключевых атрибутов таблиц, соответствующих сущностям, стоящим со стороны “один” или с обязательного конца связи, в таблицы, соответствующие сущностям, стоящим со стороны “много” или с необязательного конца связи, в качестве внешних ключей (рисунок 10).

Сотрудник                                                             Дети

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Код сотрудника | Фамилия сотрудника |   | Код сотрудника | Имя ребёнка | Возраст ребёнка |
|   |   |   |   |   |   |

Рисунок 10 – Реализация связи типа 1:m

1. Связи типа m:m и необязательные связи реализуются при помощи промежуточной таблицы, содержащей ключевые атрибуты связываемых таблиц в качестве внешних ключей (рисунок 11).

Поставщик                                         Товар

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Код поставщика | Наименованиеорганизации | № счета |   | Код товара | Наименованиетовара | ЕдиницыИзмерения |
|   |   |   |   |   |   |   |

Поставки

|  |  |  |  |
| --- | --- | --- | --- |
| Код поставщика | Код товара | Дата поставки | Оптовая цена |
|   |   |   |   |

Рисунок 11 – Реализация связи типа m:m

***4. Содержание отчёта***.

В пояснительную записку к контрольной работе необходимо включить:

1. Краткое описание предметной области
2. Изображение сущностей и их атрибутов
3. ЕR-диаграмму
4. Таблицы базы данных с указанием первичных и внешних ключей

**Вариант 11.**Телеателье
База данных должна содержать сведения о следующих объектах:

1. Сотрудники - фамилия, имя, отчество, должность, оклад, адрес, дата рождения.
2. Мастера - фамилия, имя, отчество, разряд, специализация, район обслуживания.
3. Абонементное обслуживание - фио клиента, адрес, телевизор, дата регистрации, номер квитанции об уплате абонементной платы за текущий период.
4. Заявки - фио клиента, адрес, телевизор, неисправность, дата регистрации, дата выполнения, мастер, номер квитанции.
5. Склад - наименование детали, остаток на складе, распределение по мастерам.