

2.1. Общая нотация модели «Сущность - Связь»

В основе любой информационной системы – лежит база данных. Разберем, из каких этапов строится создание базы данных или её разработка.

Процесс создания базы данных

- Определение требований.
- Выбор СУБД и разработка структуры БД с учетом особенностей СУБД.
- Реализация.
- Тестирование, разработка документации, сопровождение.

1. Вначале мы должны определиться с предметной областью, информацию о которой мы будем хранить, и специфицировать те требования, которые мы к ней предъявляем. Это могут быть объекты реального мира, их связи, отношения друг с другом; также мы должны предусмотреть возможные размеры, которые достигнет база данных; какое количество пользователей будет с нею обращаться.

2. В зависимости от этих требований подбирается некая реальная СУБД, в контексте которой будут описаны основные структуры и объекты, необходимые для представления данных. Также СУБД должна позволять нам описывать правила целостности и некую бизнес-логику.

3. Когда СУБД выбрана, структура базы данных описана, можно переходить непосредственно к процессу реализации базы данных. На этом этапе будут созданы объекты базы данных, и они заполняются непосредственными данными.

4. Далее потребуется труд тестировщиков, которые проверят, что все данные описаны корректно, правила бизнес-логики реализованы. Также должна быть разработана документация, которая позволит пользователям «общаться» с базой данных, а администраторам её поддерживать.

Рассмотрим первый этап – один из важнейших этапов – каким образом определяются требования к базе данных. Назовем это – процессом моделирования.

Определение требований – моделирование

Процесс моделирования начинается с изучения понятий и описаний. Мы должны понять какие данные мы хотим хранить и, самое важное, на какие вопросы пользователей к этим данным мы должны находить ответы в базе данных.

- Изучения понятий и описаний.
- Описание информационных объектов или понятий предметной области и связей между ними.
- Описание ограничений целостности.

При моделировании мы должны описать объекты и понятия выбранной предметной области и проследить связи между ними.

Также должны быть описаны ограничения целостности, то есть требования к допустимым значениям данных и связей между ними.

Рассмотрим предметную область. Например, нам нужно создать информационную систему, которая будет называться «Экзаменационная сессия». В данной системе нам необходимо хранить информацию о том, как студенты распределены по группам, необходимо хранить информацию о студентах, скорее всего информация о студенте будет содержать номер его зачетки, фамилию, имя, отчество, и мы догадываемся, что в номере зачетки не бывает букв, в фамилии не бывает цифр, сам студент должен состоять только в одной группе. Когда мы назначаем экзамен, то экзамен должен быть по конкретному предмету, его должен принимать определенный преподаватель, результатом экзамена является экзаменационная оценка, которая бывает в допустимом диапазоне.

Таким образом мы привели пример описания предметной области, и неких ограничений целостности и правил бизнес-логики.

Чтобы подчеркнуть важность стадии моделирования можно привести так называемый 16-ый закон систематики, который называют законом Мерфи.

Закон Мерфи **16-ый закон систематики**

Сложная система, спроектированная насухо, никогда не работает, и исправить ее, чтобы заставить работать – невозможно.

Это говорит о том, что стадия проектирования данных чрезвычайно важна.

На этапе моделирования данных мы должны составить семантическую модель.

Семантические модели

Семантическая модель (или концептуальная модель, инфологическая модель) – модель предметной области, предназначенная для представления семантики предметной области на самом высоком уровне абстракции. То есть мы должны

абстрагироваться, при составлении семантической модели, от конкретной СУБД, в рамках которой будет реализована база данных, а тем более от особенности физического хранения.

Процесс моделирования данных нужно начинать с анализа предметной области из выявления требований к ней отдельных пользователей. Если мы рассматриваем базу данных «Экзаменационная сессия», то скорей всего у разных пользователей будут разные задачи, связанные с этой базой данных: студентам нужно знать, когда и в какую аудиторию они должны прийти на экзамен; преподаватель должен иметь возможность проставить оценку за экзамен; а администратор учебного заведения должен составлять расписание экзаменационной сессии, указывать предметы и назначать преподавателей для проведения экзамена.

Представления отдельных пользователей обобщаются в концептуальную схему базы данных. Она может быть описана на естественном языке, при помощи математических формул, таблиц, графиков и других средств, которые позволяют людям, которые будут разрабатывать проект базы данных, понять структуру предметной области. Эту модель еще называют инфологической моделью данных.

После разработки инфологической модели можно приступать к даталогическому моделированию. На этом уровне мы описываем структуры хранения в контексте выбранной СУБД, а с физическим уровнем хранения данных имеют дело уже инструменты выбранной СУБД.

Трехуровневая архитектура

Трехуровневая архитектура (инфологический, даталогический и физический уровни) позволяет нам обеспечить **независимость хранимых данных** от использующих их программ.

Администратор базы данных при необходимости может переписать данные на другое устройство или разбить их на части, если база данных достигла очень больших размеров, может подключать к ней различных пользователей, но указанные изменения физической модели данных никоим образом не скажутся на логической структуре данных.

Так же если нам потребуется каким-то образом видоизменить логическую структуру данных, то это не потребует изменения физических устройств хранения данных.

Модель «сущность – связь»

Для семантического моделирования наиболее известной моделью является модель «сущность – связь» (Entity-Relationship model, или ER-модель), разработанная Питером Ченом в 1976 году. Эта модель обычно представляется в

графической форме с использованием оригинальной нотации Чена, которая называется «диаграммой сущность – связь».

Основные преимущества ER-моделей:

- наглядность;
- возможность проектировать базы данных с большим количеством объектов и атрибутов;
- на этой модели основаны множество систем автоматизированного проектирования баз данных (например, ERWin).

В дальнейшем многими авторами были разработаны подобные модели, но главное достоинство таких моделей состоит в наглядности, потому что рисунок всегда понятнее текстового описания.

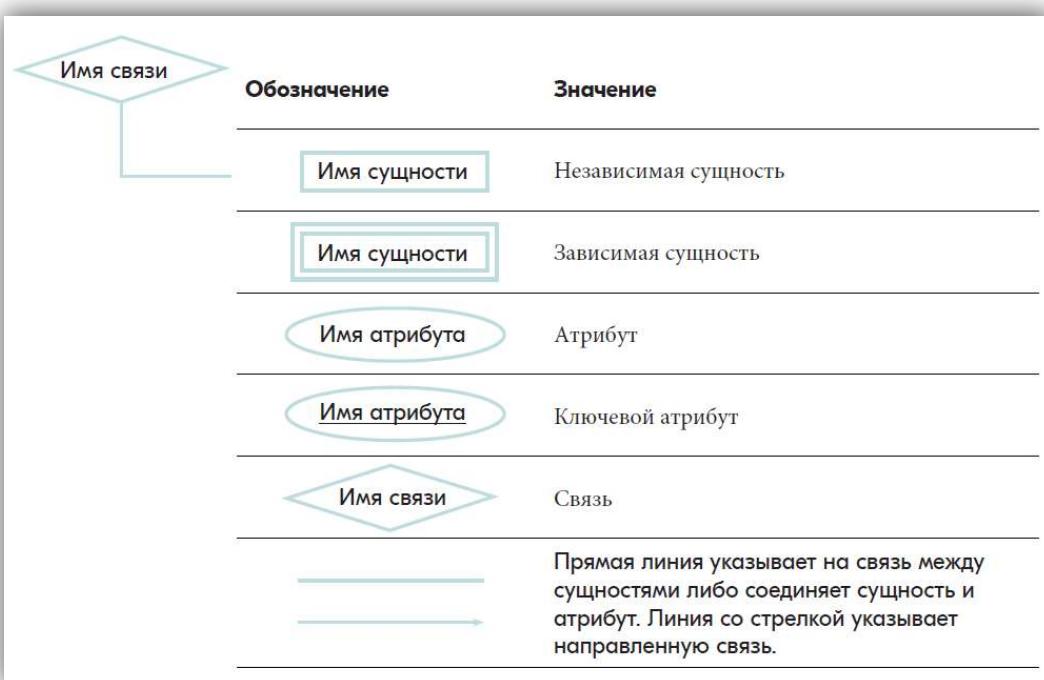
Диаграмму «сущность – связь» можно считать инструментом концептуального уровня, и важная ее особенность заключается в том, что она не содержит операций, поэтому не может быть использована непосредственно, но диаграммы «сущность – связь» являются развитой системой зависимостей и ограничений целостности.

Назначение диаграмм «сущность – связь»

Эти диаграммы являются способом проектирования данных, позволяют нам идентифицировать понятия предметной области и связей между ними, и дают нам легкий удобный графический интерфейс, который описывает логическую структуру базы данных.

Основные элементы диаграммы «сущность – связь»:

- сущности (прямоугольники) – это любые объекты или понятия, информацию о которых мы собираемся хранить;
- атрибуты (ovalы) – свойства сущностей;
- связи (ромбы) – некие отношения сущностей друг с другом.



На рисунке выше можно увидеть примеры основных понятий, которые могут быть использованы в диаграмме «сущность – связь».

Пример

Рассмотрим пример описанной нами экзаменационной сессии. В нашей базе данных должна храниться информация о студентах, сданных ими экзаменах по определенным предметам. Студент характеризуется номером зачетки, фамилией, именем, отчеством, датой рождения; студент может иметь несколько телефонов, все студенты распределены по учебным группам, для группы назначается экзамен по определенному предмету, назначается экзаменатор, указывается дата экзамена и аудитория, за каждый сданный экзамен студенту ставится оценка.

Сущностью является любой объект или понятие, информацию о которых мы храним. Мы видим, что сущностями у нас будут студенты, преподаватели, сданные предметы.

Свойства сущностей или их атрибуты: фамилия, имя, отчество, дата рождения, номер зачетки.

Ограничения, которые могут быть наложены на наши данные: номер зачетки – положительное число; фамилия, имя, отчество – не бывает цифр; оценка – бывает от двух до пяти; студент – состоит в одной учебной группе; в один день можно сдать один экзамен.

Использование диаграммы «сущность – связь»

Любой проектировщик или аналитик базы данных при помощи составления диаграммы может получить лучшее понимание предметной области, также диаграммы могут служить инструментом для документации, диаграмма «сущность – связь» передает нам логическую структуру базы данных и поможет использовать ее конечным пользователям.