

4 Проектирование таблиц

Как важно быть нормальным ✨

...и тогда мама назвала меня своим маленьким помощником!

Да, это ненормально.



До настоящего момента мы не особо задумывались при создании таблиц. Работают — и ладно; в конце концов, с ними можно выполнять команды `SELECT`, `INSERT`, `DELETE` и `UPDATE`. Но при увеличении объема данных постепенно становится ясно, что следовало бы сделать при создании таблицы для упрощения условий `WHERE`: ее следовало бы сделать более нормальной.

Две таблицы

Джек и Марк создали таблицы для хранения информации о рекордах рыбной ловли. В таблице Марка имеются столбцы для бытового и научного названия рыбы, ее веса и места, где она была поймана. Столбца для имени человека, поймавшего рыбу, в этой таблице нет.

Таблица состоит из четырех столбцов. Сравните с таблицей fish_records на следующей странице.

fish_info

common	species	location	weight
большеротый окунь	M. salmoides	Монтгомери Лейк, GA	22 фт 4 унц
судак	S. vitreus	Олд Хикори Лейк, TN	25 фт 0 унц
лосось Кларка	O. Clarki	Пирамид Лейк, NV	41 фт 0 унц
желтый окунь	P. Flavescens	Бордентаун, NJ	4 фт 3 унц
синежаберник	L. Macrochirus	Кетона Лейк, AL	4 фт 12 унц
панцирник	L. Osseus	Тринити Ривер, TX	50 фт 5 унц
белый краппи	P. annularis	Дамба Энид, MS	5 фт 3 унц
красноперая щука	E. americanus	Дьюарт Лейк, IN	1 фт 0 унц
серебристый карась	C. auratus	Лейк Ходжес, CA	6 фт 10 унц
чавыча	O. Tshawytscha	Кенай Ривер, АК	97 фт 4 унц

Я ихтиолог. В своей таблице я буду проводить поиск только по бытовому или научному названию, чтобы узнать вес и место вылова рыбы.



Марк

В таблице Джека тоже хранятся бытовые и научные названия рыб, но в ней также имеются столбцы для имени и фамилии рыболова, а место вылова разбито на два столбца: название водоема хранится отдельно от штата.

Эта таблица тоже содержит информацию о рыболовных рекордах, но в ней почти вдвое больше столбцов.

fish_records

first_name	last_name	common	location	state	weight	date
Джордж	Перри	большеротый окунь	Монтгомери Лейк	GA	22 фт 4 унц	2/6/1932
Мабри	Харпер	судак	Олд Хикори Лейк	TN	25 фт 0 унц	2/8/1960
Джон	Скиммерхорн	лосось Кларка	Пирамид Лейк	NV	41 фт 0 унц	1/12/1925
С.С.	Эббот	желтый окунь	Бордентаун	NJ	4 фт 3 унц	1/5/1865
Т.С.	Хадсон	синежаберник	Кетона Лейк	AL	4 фт 12 унц	9/4/1950
Таунсенд	Миллер	панцирник	Тринити Ривер	TX	50 фт 5 унц	30/7/1954
Фред	Брайт	белый краппи	Дамба Энид	MS	5 фт 3 унц	31/7/1957
Майк	Берг	красноперая щука	Дьюарт Лейк	IN	1 фт 0 унц	9/6/1990
Флорентино	Абена	серебристый карась	Лейк Ходжес	CA	6 фт 10 унц	17/4/1996
Лес	Андерсон	чавыча	Кенай Ривер	AK	97 фт 4 унц	17/5/1985

Возьми в руку карандаш



Напишите запрос для каждой таблицы, возвращающий все записи для штата Нью-Джерси.

.....

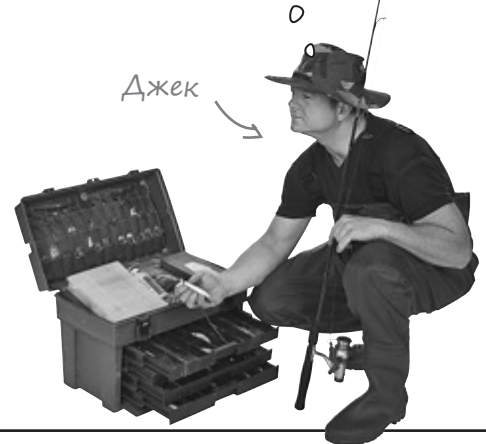
.....

.....

.....

А я пишу статьи для рыболовного журнала. И мне нужно знать имена рыбаков, даты и места рекордного вылова.

Джек





Возьми в руку карандаш Решение

Напишите для каждой таблицы запрос, возвращающий все записи для штата Нью-Джерси.

Для получения результатов из строки «город, штат» приходится использовать ключевое слово LIKE.

Мне почти никогда не приходится искать записи по штату. Я храню название штата в одном столбце с названием города.

```
SELECT * FROM fish_info
WHERE location LIKE '%NJ';
```



common	species	location	weight
желтый окунь	P. Flavescens	Бордентаун, NJ	4 фт 3 унц

Запрос напрямую обращается к столбцу state.

А мне часто приходится искать по штату, поэтому я выделил название штата в отдельный столбец при создании таблицы.

```
SELECT * FROM fish_records
WHERE state = 'NJ';
```



first_name	last_name	common	location	state	weight	date
C.C.	Эббот	желтый окунь	Бордентаун	NJ	4 фт 3 унц	1/5/1865

Часть Задаваемые Вопросы

В: Выходит, таблица Джека лучше, чем таблица Марка?

О: Нет. Это разные таблицы с разными целями. Марку редко приходится проводить поиск по штату, потому что его интересуют только названия (бытовое и научное) выловленных рыб и их вес.

С другой стороны, Джеку *потребуется* искать данные по штату в своих запросах. Именно поэтому он создал в своей таблице отдельный столбец, чтобы было удобнее указывать штат в запросах.

В: Следует ли избегать оператора LIKE в запросах? Что в нем плохого?

О: В операторе LIKE нет ничего плохого, но он усложняет структуру запроса и повышает риск получения посторонних результатов. Если столбцы содержат сложную информацию, LIKE не позволяет легко и однозначно определить критерий поиска.

В: Почему короткие запросы лучше длинных?

О: Чем проще запрос, тем лучше. С увеличением объема базы данных и добавлением новых таблиц запросы усложняются. Начинайте с самых простых запросов, позднее вы их оцените.

В: Значит, в моих столбцах всегда должны храниться как можно меньшие фрагменты данных?

О: Не обязательно. Как показывает пример с таблицами Марка и Джека, все зависит от *использования* данных. Для примера представьте таблицы со списком машин, предназначенные для автомеханика и продавца. Механику необходима подробная информация о каждой машине, а продавцу может быть достаточно фирмы-производителя, модели и номера.

В: Допустим, в записи хранится почтовый адрес. Почему бы не создать один столбец для хранения полного адреса и несколько других столбцов для хранения его составных частей?

О: Дублирование данных поначалу может показаться вполне разумной мерой, но подумайте, сколько лишнего пространства будет расходоваться на жестком диске, если база данных вырастет до значительных размеров. А еще при дублировании данных в команду UPDATE должно включаться дополнительное лишнее условие, и вы должны помнить о нем при каждом изменении данных.

Давайте более подробно разберемся в том, как спроектировать оптимальную структуру таблицы для ваших целей..

Структура таблицы зависит от того, как вы собираетесь использовать свои данные.



МОЗГОВОЙ ШТУРМ

SQL — язык, используемый реляционными базами данных. Как вы думаете, что означает термин «реляционный» в контексте баз данных SQL?

Логические связи как суть таблицы

SQL известен как язык Реляционных Систем Управления Базами Данных (РСУБД). Термин запоминать не обязательно, нас интересует только слово «РЕЛЯЦИОННЫХ*». Для нас оно означает, прежде всего, одно: чтобы правильно спроектировать таблицу, необходимо продумать, как столбцы связываются друг с другом для описания некоторого объекта.

Ваша задача — описать объект при помощи столбцов так, чтобы по возможности упростить получение необходимой информации. Конечно, выбор во многом зависит от ваших требований к таблице, но существуют некоторые общие меры, которые следует принять при выборе структуры таблицы.

1. Выберите один объект, который должна описывать таблица.

Какой основной объект описывает ваша таблица?

2. Составьте список того, что необходимо знать об этом объекте при работе с таблицей.

Как будет использоваться ваша таблица?

3. Используя список, разбейте необходимую информацию об объекте на фрагменты, которые могут использоваться для определения структуры таблицы.

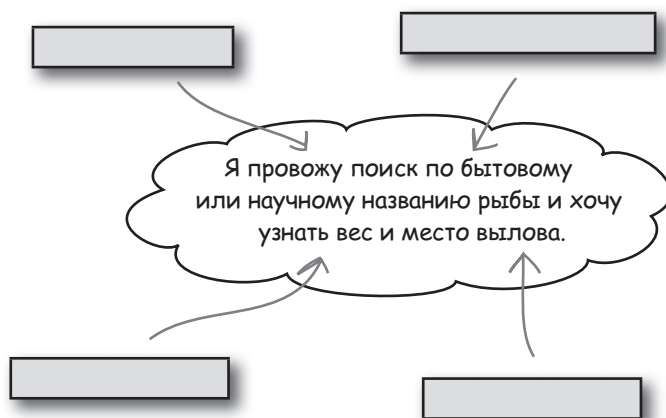
Как проще всего запросить данные из таблицы?

* Встречается мнение, что термин «РЕЛЯЦИОННЫЙ» относится к логическим связям между таблицами. Это неверно.

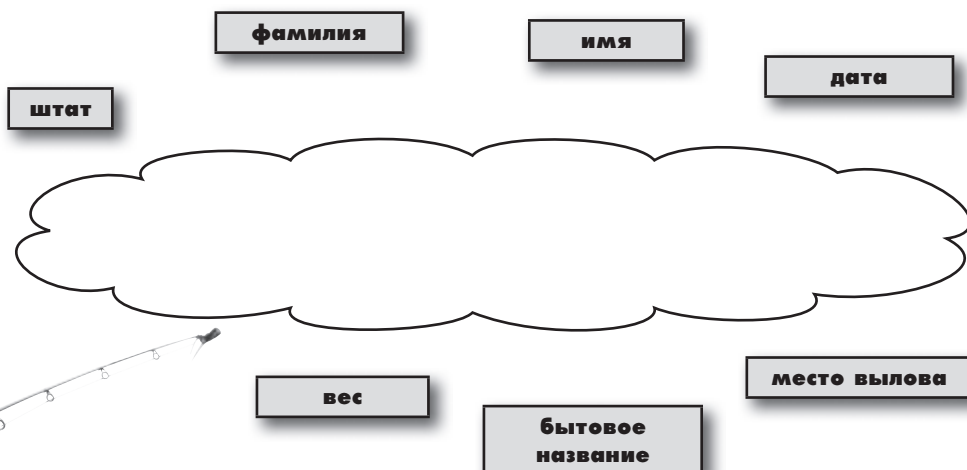


Упражнение

Сможете ли вы определить столбцы таблицы по тем словам, которыми ихтиолог Марк описывает выборку данных из таблицы? Запишите имена столбцов в прямоугольниках.



Теперь ваша очередь. Напишите аналогичную фразу для Джека, автора статей по рыбной ловле, который использует таблицу для получения подробной информации для своих статей. Затем проведите стрелки от каждого столбца к его упоминанию в описании.





Упражнение Решение

Сможете ли вы определить столбцы таблицы по тем словам, которыми ихтиолог Марк описывает выборку данных из таблицы? Запишите имена столбцов в прямоугольниках.



Теперь ваша очередь. Напишите аналогичную фразу для Джека, автора статей по рыбной ловле, который использует таблицу для получения подробной информации для своих статей. Затем проведите стрелки от каждого столбца к его упоминанию в описании.



Но почему мы остановились на таблице Джека? Ведь дату можно разбить на день, месяц и год? Да и место вылова можно уточнить до названия улицы и номера дома рыбака.



Да, можно, но такой уровень детализации данных просто не нужен.

По крайней мере не в этом конкретном случае. Если бы Джек писал статьи о том, куда лучше отправиться на выходных, чтобы поймать большую рыбу, *тогда* он, возможно, указал бы название улицы и номер дома, чтобы читатели могли поискать жилище где-нибудь поблизости.

Но Джека интересует только место вылова и штат, и он добавил только эти столбцы, чтобы не увеличивать объем базы данных без необходимости. На этой стадии он решил, что его данные достаточно детализированы — то есть являются *атомарными*.



Как вы думаете, что означает термин *атомарный* в контексте данных SQL?

Атомарные данные

Что такое «атом»? Маленький блок информации, который невозможно (или нежелательно) разделить на составные части меньшего размера. Это определение относится и к данным: АТОМАРНЫЕ данные были разделены на наименьшие компоненты, дальнейшее деление которых *невозможно или нежелательно*.

Доставка за 30 минут, или Пицца бесплатно

Для примера возьмем курьера, доставляющего пиццу клиентам. Чтобы выполнить свою работу, ему достаточно знать улицу и номер дома в одном столбце. Для него эти данные являются атомарными: курьеру никогда не приходится искать номер дома отдельно от названия улицы.

Более того, разбиение адреса доставки на название улицы и номер дома только усложнит его работу, а клиентам придется дольше дожидаться своих заказов.



Для курьера адрес доставки, объединяющий улицу и номер дома в одном столбце, достаточно атомарен.

```
File Edit Window Help SimplePizzaFactory
+-----+
| order_number | address |
+-----+
| 246          | 59 N. Ajax Rapids |
| 247          | 849 SQL Street    |
| 248          | 2348 E. PMP Plaza |
| 249          | 1978 HTML Heights |
| 250          | 24 S. Servlets Springs |
| 251          | 807 Infinite Circle |
| 252          | 32 Design Patterns Plaza |
| 253          | 9208 S. Java Ranch |
| 254          | 4653 W. EJB Estate |
| 255          | 8678 OOA&D Orchard |
+-----+
> SELECT address FROM pizza_deliveries WHERE order_num = 252;
+-----+
| address |
+-----+
| 32 Design Patterns Plaza |
+-----+
1 row in set (0.04 sec)
```