Оглавление

От издательства **11**

О научном редакторе русского издания **12**

Предисловие **13**

Какие задачи мы попытаемся решить 14

О чем мы будем говорить 14

Для кого эта книга 17

Почему именно SQL 18

Осторожно! Впереди мнение! 19

Условные обозначения, используемые в книге 19

Другие обозначения, принятые в книге 20

«Модель» данных 23

Использование исходного кода примеров 27

Благодарности 27

ЧАСТЬ I
ОБЗОР

Глава 1. Обзор оператора SELECT **30**

Простые запросы SELECT 30

Обобщенные табличные выражения 34

Случаи использования оператора CASE 38

Операторы соединения 43

И еще немного о SELECT 53

Глава 2. Сочетание функций.....55

Агрегатные функции.....56

Функции преобразования.....59

Криптографические функции: HASHBYTES.....63

Функции даты и времени.....63

Логические функции: IIF.....66

Строковые функции.....67

Системные функции.....81

И еще немного о функциях.....83

ЧАСТЬ II
СЛОЖНОСТИ ПРИ РАБОТЕ С ДАННЫМИ

Глава 3. Имена, имена, имена.....86

Что в имени?.....86

Фамилии.....90

Имена.....98

Второе имя.....99

Прозвище (ник).....100

Название компании.....101

Полные имена.....102

Организации — физические лица.....103

И еще немного об именах.....104

Глава 4. Адреса, адреса, адреса.....106

Из чего состоит адрес?.....106

Улица и дом.....108

Город.....115

Округ.....116

Штат или аббревиатура штата 116
Почтовый индекс, или ZIP 117
Страна 118
И еще немного об адресах 119

Глава 5. Даты, даты, даты 121
Время относительно 121
И еще немного о датах 126

Глава 6. Электронная почта 130
Какие адреса электронной почты действительны? 130
И еще немного об электронной почте 132

Глава 7. Телефонные номера 134
Что такое телефонный номер? 134
Немного о налоговых идентификаторах 140
И еще немного о телефонных номерах (и налоговых идентификаторах) 141

Глава 8 . Недопустимые символы 143
Представления данных 143
Невидимые пробелы 147
Сортировка 149
Очистка входных данных 153
И еще немного о недопустимых символах 156

Глава 9. Ортогональные данные 158
Общая проблема, общее решение, новая общая проблема 159
Намылить, смыть, повторить 169
И еще немного об ортогональных данных 172

ЧАСТЬ III
СБОРКА

Глава 10. Большой куш 176

Что нам нужно? 176

Настройка оценок 192

Удаление дубликатов 199

И еще немного об оценке 209

Глава 11. Качество данных, или GIGO 211

Как сообщить о проблеме с качеством 211

Невозможные данные 212

ETL — ваш путь к успеху 216

И еще немного о качестве данных 221

Глава 12. Все целиком 223

Подход 223

Какая оценка? 225

Первый проход: наивное сопоставление 229

Второй проход: нормализация отношений 233

Третий проход: оценка! 238

А что с настройкой? 246

И напоследок о практических вопросах 246

Глава 13. Код — это тоже данные! 248

Работа с данными XML 248

Работа с данными в формате JSON 250

Извлечение данных из HTML 251

Код, генерирующий код 253

Импакт-анализ: второй пример 255

И напоследок о коде как о данных 272

И напоследок обо всем сказанном 274

Приложение. «Модель» данных 276

Таблица клиентов 276

Представление NormalizedCustomer 276

Таблица потенциальных совпадений PotentialMatches 276

Представление CustomerCountByState 280

Таблица почтовых сокращений PostalAbbreviations 280

Глоссарий 281

Об авторе 283

Иллюстрация на обложке 284

Технологический стек 285