

УДК 004.6
ББК 32.973.26
Р33

Р33 Эрик Редмонд, Джим. Р. Уилсон

Семь баз данных за семь недель. Введение в современные базы данных и идеологию NoSQL. Под редакцией Жаклин Картер / Пер. с англ. Слинкин А. А. – М.: ДМК Пресс, 2018. – 384 с.: ил.

В книге описаны большинство из современных баз данных с открытым исходным кодом: Redis, Neo4J, CouchDB, MongoDB, HBase, PostgreSQL и Riak. Для каждой базы приведены примеры работы с реальными данными, демонстрирующие основные идеи и сильные стороны.

Эта книга прольет свет на сильные и слабые стороны каждой из семи баз данных и научит вас выбирать ту, которая лучше отвечает требованиям.

Издание предназначено для программистов разной квалификации, использующих базы данных в своей профессиональной деятельности.

ISBN 978-5-97060-615-5

УДК 004.6
ББК 32.973.26

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

Материал, изложенный в данной книге, многократно проверен. Но, поскольку вероятность технических ошибок все равно существует, издательство не может гарантировать абсолютную точность и правильность приводимых сведений. В связи с этим издательство не несет ответственности за возможные ошибки, связанные с использованием книги.

ISBN 978-1-93435-692-0 (англ.)
ISBN 978-5-97060-615-5 (рус.)

© Pragmatic Programmers, LLC.
© Оформление, перевод на русский язык
ДМК Пресс



СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ	13
БЛАГОДАРНОСТИ	15
ВСТУПЛЕНИЕ	17
Почему именно семь баз данных?	17
Что есть в этой книге?	17
Чего нет в этой книге?	18
Это не руководство по установке	18
Руководство администратора? Пожалуй, нет	19
Замечание для пользователей Windows	19
Примеры кода и соглашения	19
Сетевые ресурсы	20
Глава 1. Введение	21
1.1. Все начинается с вопроса	21
1.2. Жанры	23
Реляционные СУБД	24
Хранилища ключей и значений	25
Столбцовые базы данных	26
Документно-ориентированные базы данных	27
Графовые базы данных	28
Многостороннее хранение	28
1.3. Вперед и вверх	29
Глава 2. PostgreSQL	30
2.1. Произносится Post-greS-Q-L	30
2.2. День 1: отношения, операции CRUD и соединения	32
Введение в SQL	33

Быстрый поиск с применением индексов	41
День 1: итоги	43
День 1: домашнее задание	44
2.3. День 2: более сложные запросы, код и правила	45
Агрегатные функции	45
Оконные функции	48
Транзакции	49
Хранимые процедуры	50
Триггеры	52
Представление о мире	54
Правила	55
Создание сводных таблиц с помощью crosstab()	57
День 2: итоги	59
День 2: домашнее задание	59
2.4. День 3: полнотекстовый поиск и многомерные кубы	60
Нечеткий поиск	62
Полнотекстовый поиск	65
День 3: итоги	75
День 3: домашнее задание	75
2.5. Резюме	75
Сильные стороны PostgreSQL	76
Слабые стороны PostgreSQL	77
Перед расставанием	77
Глава 3. Riak	78
3.1. Riak дружит с веб	79
3.2. День 1: CRUD, ссылки и типы MIME	80
Лучше REST может быть только REST (или как заивать локоны)	82
Ссылки	85
Типы MIME в Riak	89
День 1: итоги	90
День 1: домашнее задание	90
3.3. День 2: mapreduce и кластеры серверов	91
Скрипт для загрузки данных	91
Введение в Mapreduce	92
Mapreduce в Riak	95
О согласованности и долговечности	101
День 2: итоги	109
День 2: домашнее задание	109

3.4. День 3: разрешение конфликтов и расширение Riak ..	110
Разрешение конфликтов с помощью векторных часов.....	110
Расширение Riak	117
День 3: итоги.....	121
День 3: домашнее задание	122
3.5. Резюме	122
Сильные стороны Riak.....	123
Слабые стороны Riak.....	123
Riak и теорема CAP	123
Перед расставанием	124

Глава 4. HBase..... 125

4.1. Введение в HBase	126
4.2. День 1: операции CRUD и администрирование таблиц	127
Конфигурирование HBase	128
Оболочка HBase	129
Создание таблицы	129
Вставка, обновление и выборка данных	131
Добавление данных из программы.....	136
День 1: итоги.....	137
День 1: домашнее задание	138
4.3. День 2: работа с «большими данными»	139
Импорт данных, выполнение скриптов	139
Потоковая загрузка XML.....	140
Загрузка википедии	141
Сжатие и фильтры Блума.....	143
Контакт? Есть контакт!.....	143
Знакомство с регионами и мониторингом места на диске.....	145
Опрос регионов	146
Сканирование одной таблицы для построения другой.....	149
Построение сканера.....	150
Запуск скрипта.....	152
Исследование результатов.....	153
День 2: итоги.....	154
День 2: домашнее задание	155
4.4. День 3: переходим в облако	156
Разработка «бережливого» приложения для HBase	156
Введение в Whirr	160
Подготовка к работе с EC2	160
Подготовка Whirr	161

Настройка кластера	162
Запуск кластера	163
Подключение к кластеру.....	163
Уничтожение кластера	164
День 3: итоги.....	164
День 3: домашнее задание	165
4.5. Резюме	166
Сильные стороны HBase.....	166
Слабые стороны HBase	167
HBase и теорема CAP	167
Перед расставанием	168
Глава 5. MongoDB	169
5.1. Монстр.....	169
5.2. День 1: операции CRUD и вложенность	171
Поработаем с командной строкой	171
JavaScript	173
Чтение: продолжаем изучать Mongo.....	175
Копнем глубже	177
Обновление	181
Ссылки	183
Удаление	184
Функциональные критерии.....	185
День 1: итоги.....	186
День 1: домашнее задание	186
5.3. День 2: индексирование, группировка, mapreduce	187
Индексирование: когда быстродействия не хватает	187
Агрегированные запросы	191
Команды на стороне сервера	194
Mapreduce (и Finalize).....	197
День 2: итоги.....	201
День 2: домашнее задание	201
5.4. День 3: наборы реплик, сегментирование, пространственные данные и GridFS	201
Наборы реплик.....	202
Сегментирование.....	206
Пространственные запросы	208
GridFS	210
День 3: итоги.....	211
День 3: домашнее задание	211
5.5. Резюме	212

Сильные стороны Mongo	212
Слабые стороны Mongo	212
Перед расставанием	213
Глава 6. CouchDB	214
6.1. Располагайтесь на кушетке	214
Сравнение CouchDB с MongoDB	215
6.2. День 1: операции CRUD, Futon и снова cURL	215
Знакомство с Futon	216
Выполнение операций CRUD с помощью REST-интерфейса и cURL	219
Чтение документа с помощью GET	220
Создание документа с помощью POST	221
Обновление документа с помощью PUT	222
Удаление документа с помощью DELETE	223
День 1: итоги	223
День 1: домашнее задание	223
6.3. День 2: создание и опрос представлений	224
Доступ к документам через представления	224
Создание первого представления	226
Сохранение представления в виде проектного документа	229
Поиск исполнителей по имени	229
Поиск альбомов по названию	230
Опрос представлений исполнителей и альбомов	231
Импорт данных в CouchDB с помощью программы на Ruby	233
День 2: итоги	238
День 2: домашнее задание	238
6.4. День 3: более сложные представления, Changes API и репликация данных	239
Создание более сложных представлений с помощью редукторов	239
Отслеживание изменений в CouchDB	243
Непрерывное отслеживание изменений	249
Фильтрация изменений	250
Репликация данных в CouchDB	252
День 3: итоги	256
День 3: домашнее задание	256
6.5. Резюме	257
Сильные стороны CouchDB	257
Слабые стороны CouchDB	258
Перед расставанием	258

Глава 7. Neo4J.....	259
7.1. Neo4J дружит с доской.....	259
7.2. День 1: графы, Groovy и операции CRUD	261
Веб-интерфейс Neo4j.....	262
Neo4j и Gremlin	264
Конвейеры	267
Конвейер и вершина	269
Бессхемная социальная сеть.....	270
Дорога меряется шагами	271
Обновляем, удаляем, стираем	278
День 1: итоги.....	279
День 1: домашнее задание.....	279
7.3. День 2: REST, индексы и алгоритмы.....	279
REST-интерфейс.....	279
Интересные алгоритмы.....	286
День 2: итоги.....	292
День 2: домашнее задание.....	292
7.4. День 3: распределенность и высокая доступность	293
Транзакции	293
Высокая доступность	294
HA-кластер.....	295
Резервное копирование	301
День 3: итоги.....	302
День 3: домашнее задание.....	302
7.5. Резюме	302
Сильные стороны Neo4j.....	303
Слабые стороны Neo4j	303
Neo4j и теорема CAP	304
Перед расставанием	304
Глава 8. Redis.....	305
8.1. Хранилище сервера структур данных	305
8.2. День 1: операции CRUD и типы данных.....	306
Приступая к работе	307
Транзакции	309
Составные типы данных	309
Блокирующие списки	313
Диапазоны	316
Пространства имен	319
И это еще не всё.....	320

День 1: итоги.....	321
День 1: домашнее задание.....	321
8.3. День 2: более сложные применения, распределенные вычисления.....	321
Простой интерфейс	322
Информация о сервере	325
Настройка Redis	325
Репликация главный-подчиненный.....	330
Загрузка данных.....	330
Кластер Redis.....	333
Фильтры Блума	334
SETBIT и GETBIT	337
День 2: итоги.....	338
День 2: домашнее задание.....	338
8.4. День 3: комбинирование с другими базами данных....	339
Служба многостороннего хранения.....	339
Заполнение данными	341
Фаза 1: трансформация данных	342
Фаза 2: вставка в каноническую систему.....	344
Хранилище связей	347
Веб-служба	349
Развитие веб-службы.....	351
День 3: итоги.....	352
День 3: домашнее задание.....	353
8.5. Резюме	353
Сильные стороны Redis.....	353
Слабые стороны Redis.....	354
Перед расставанием	354
Глава 9. Подводя итоги	356
9.1. Снова о жанрах	356
Реляционные базы данных	356
Хранилища ключей и значений	357
Столбцовые базы данных	358
Документные базы данных	359
Графовые базы данных	360
9.2. Как сделать выбор?.....	361
9.3. В каком направлении двигаться дальше?	362
ПРИЛОЖЕНИЕ 1. Сравнительный обзор баз данных	363

ПРИЛОЖЕНИЕ 2. Теорема CAP	367
A2.1. Согласованность в конечном счете.....	368
A2.2. CAP на практике	369
A2.3. Компромиссный выбор задержки.....	370
СПИСОК ЛИТЕРАТУРЫ	371
ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ	372