

УДК 004.6
ББК 32.972
Р58

- Р58 **Робинсон Ян, Вебер Джим, Эфрем Эмиль**
Графовые базы данных: новые возможности для работы со связанными данными / пер. с англ. Р. Н. Рагимова; науч. ред. А. Н. Кисилев. – 2-е изд. – М.: ДМК Пресс, 2016. – 256 с.: ил.

ISBN 978-5-97060-201-0

Из книги вы узнаете, как проектировать и реализовывать приложения, основанные на графовых базах данных, приносящих мощь графов в широкий круг прикладных областей. Если вам необходимо уменьшить время выполнения запросов пользователей или создать базу данных, способную приспособливаться под быстро развивающийся бизнес, эта книга продемонстрирует вам практическое применение графовых моделей.

Второе издание книги содержит обновленные примеры кода и схемы, соответствующие актуальному синтаксису графовой базы данных Neo4j, а также информацию о новом функционале Neo4j.

Издание предназначено для программистов, желающих изучить работу графовых баз данных и научиться максимально использовать их мощь в своей работе.

УДК 004.6
ББК 32.972

Authorized Russian translation of the English edition of Graph Databases, 2nd Edition. This translation is published and sold by permission of O'Reilly Media, Inc., which owns or controls all rights to publish and sell the same.

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 978-1-491-93089-2 (анг.) © 2015 Neo Technology, Inc.

ISBN 978-5-97060-201-0 (рус.) © Оформление, перевод, ДМК Пресс, 2016

Содержание

Пролог	10
Предисловие	13
Глава 1. Введение	17
Что такое граф?	17
Обзор областей применения графов	19
Графовые базы данных	20
Механизмы вычисления графов	22
Преимущества графовых баз данных	23
Производительность	24
Гибкость	24
Оперативность	25
Итоги	25
Глава 2. Варианты хранения взаимосвязанных данных	26
Недостатки NOSQL-баз данных при работе со взаимосвязями	30
Взаимосвязи в графовых базах данных	35
Итоги	41
Глава 3. Моделирование данных графами	42
Модели и задачи	42
Графовая модель со свойствами и метками	43
Графовые запросы: введение в Cypher	44
Философия языка Cypher	45
MATCH	47
RETURN	48
Другие фразы языка Cypher	48
Сравнение реляционного и графового моделирования	49
Реляционная модель системы управления	51
Графовое моделирование системы управления	56
Тестирование модели	58
Кросс-модели нескольких прикладных областей	60
Создание графа творчества Шекспира	64

Введение в запросы	66
Определение шаблонов для поиска	68
Ограничение совпадений	70
Обработка результатов	71
Цепочки в запросах	72
Распространенные просчеты при моделировании	73
Проблемы анализа источников электронных писем	73
Первый блин комом?	73
Со второго раза все получится	76
Эволюция прикладной области	79
Идентификация узлов и взаимосвязей	85
Как избежать антишаблонов	85
Итоги	86

Глава 4. Разработка приложений графовых баз данных...88

Моделирование данных	88
Описание модели с учетом потребностей приложения	89
Узлы представляют сущности, взаимосвязи формируют структуру	90
Подробные имена или свойства взаимосвязей	91
Моделирование фактов в виде узлов	92
Работа	92
Исполнение ролей	92
Электронная переписка	94
Рецензии	94
Представление комплексных типовых значений в виде узлов	94
Время	95
Хронологическое дерево	95
Связанные списки	97
Управление версиями	98
Итеративная и поэтапная разработка	98
Архитектура приложений	100
Встроенная поддержка или сервер	100
Встроенная поддержка Neo4j	100
Серверный режим	102
Серверные расширения	103
Кластеризация	106
Репликация	106
Запись буфера с помощью очередей	107

Глобальные кластеры.....	107
Балансировка нагрузки.....	107
Отделение трафика чтения от трафика записи.....	107
Распределительный кэш.....	109
Чтение собственных записей.....	110
Тестирование.....	111
Разработка модели данных, основанная на тестировании.....	111
Пример: модель данных социальной сети и ее тестирование.....	112
Тестирование серверных расширений.....	116
Тестирование производительности.....	118
Тесты производительности запросов.....	119
Тесты производительности приложений.....	120
Тестирование с помощью репрезентативных данных.....	121
Планирование производственных мощностей.....	123
Критерии оптимизации.....	124
Производительность.....	124
Калькуляция затрат на увеличение производительности графовой базы данных.....	125
Варианты оптимизации производительности.....	125
Избыточность.....	127
Нагрузка.....	127
Импорт и массовая загрузка данных.....	128
Первоначальный импорт.....	128
Пакетный импорт.....	130
Итоги.....	133

Глава 5. Графы в реальном мире 134

Почему выбирают графовые базы данных.....	134
Типичные примеры использования.....	136
Социальные сети.....	136
Рекомендации.....	137
Геоинформационные системы.....	138
Управление справочными данными.....	139
Сети и управление центром обработки данных.....	139
Авторизация и контроль доступа (для коммуникаций).....	141
Реальные примеры.....	142
Социальные рекомендации (Профессиональная социальная сеть).....	142

8 ❖ Содержание

Модель данных Talent.net	144
Выявление социальных взаимосвязей	145
Поиск коллег с определенными интересами.....	149
Добавление взаимосвязей WORKED_WITH.....	152
Авторизация и контроль доступа	155
Модель данных компании TeleGraph.....	156
Поиск доступных администратору ресурсов	159
Определение доступности ресурса администратору	160
Поиск администраторов по учетной записи.....	163
Геоинформационные системы и логистика	165
Модель данных Global Post.....	166
Расчет маршрута.....	169
Поиск кратчайшего маршрута с помощью Cypher.....	172
Реализация расчета маршрутов с помощью фреймворка Traversal.....	175
Итоги	180

Глава 6. Внутреннее устройство графовых баз данных 182

Нативная обработка графов	182
Нативное хранилище графов	186
Программные интерфейсы	192
Программный интерфейс ядра.....	193
Базовый интерфейс	193
Фреймворк Traversal.....	194
Нефункциональные характеристики.....	196
Транзакции	197
Восстанавливаемость.....	199
Доступность	200
Масштабирование.....	202
Мощность.....	203
Задержки.....	203
Производительность.....	204
Итоги	206

Глава 7. Интеллектуальный анализ с помощью теории графов 207

Поиск в глубину и ширину	207
Поиск маршрутов с помощью алгоритма Дейкстры	209
Алгоритм A*	217

Теория графов и прогнозное моделирование.....	218
Триадические замыкания	219
Структурный баланс	221
Локальные перемиčky	226
Итоги	228
Приложение А. Обзор NOSQL-баз данных	229
Движение NOSQL	229
ACID или BASE.....	231
Секторы NOSQL	233
Хранилища документов.....	233
Хранилища пар ключ-значение	236
Семейства столбцов	239
Запросы или обработка в агрегированных хранилищах.....	242
Графовые базы данных.....	243
Графы со свойствами	244
Гиперграфы	245
Триплеты	246
Предметный указатель	249
Об авторах	254
Заключение.....	255