

УДК 004.65
ББК 32.973.26
Ч12

Чаллавала Ш., Лакхатария Дж., Мехта Ч., Патель К.
Ч12 MySQL 8 для больших данных / пер. с англ. А. В. Логунова. – М.: ДМК Пресс, 2018. – 226 с.: ил.

ISBN 978-5-97060-653-7

Среди организаций, работающих с крупными объемами данных на регулярной основе, реляционная система управления базами данных MySQL стала популярным решением по обработке структурированных больших данных. В книге вы познакомитесь с тем, как администраторы баз данных могут использовать MySQL для обработки миллиардов записей и извлечения данных с производительностью, сравнимой или превосходящей коммерческие решения для СУБД с более высокими затратами. Показано, как реализовывать успешную стратегию больших данных с помощью таких технологий, как Apache Hadoop, MapReduce и MySQL Applier. Также книга включает в себя практические примеры использования Apache Sqoop для обработки событий в режиме реального времени.

Издание будет полезно администраторам баз данных MySQL и специалистам по большим данным, которые хотят интегрировать MySQL и Hadoop с целью реализации высокопроизводительных решений.

УДК 004.65
ББК 32.973.26

Original English language edition published by Packt Publishing, Ltd. UK. Copyright © 2017 by Packt Publishing, Ltd. Russian-language edition copyright © 2018 by DMK Press. All rights reserved.

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.

ISBN 978-1-78839-718-6 (англ.)
ISBN 978-5-97060-653-7 (рус.)

Copyright © 2017 by Packt Publishing
© Оформление, издание, перевод, ДМК Пресс, 2018

Содержание

Об авторах	11
О рецензентах	14
Предисловие	17
Глава 1. Введение в большие данные и MySQL 8.....	21
Важность больших данных.....	22
Социальные медиа	22
Политика	23
Наука и исследование.....	24
Энергетика	24
Обнаружение мошенничества.....	24
Здравоохранение	25
Бизнес-картирование.....	25
Жизненный цикл больших данных.....	26
Объем	27
Разнообразие	27
Скорость	28
Правдивость	28
Фазы жизненного цикла больших данных	28
Структурированные базы данных.....	31
Основы MySQL	32
MySQL как реляционная система управления базами данных.....	32
Лицензирование	32
Надежность и масштабируемость	33
Совместимость платформ.....	33
Выпуски.....	33
Новые возможности в MySQL 8	33
Транзакционный словарь данных.....	34
Роли	35
Автоинкремент InnoDB.....	35
Поддержка невидимых индексов	36
Улучшение индексов, отсортированных по убыванию	36
SET PERSIST.....	36
Расширенная поддержка ГИС.....	36
Кодировка символов по умолчанию	37
Улучшение побитовых операций	37
InnoDB Memcached	37
NOWAIT и SKIP LOCKED	37
Преимущества использования MySQL.....	38
Безопасность.....	38

Масштабируемость	38
Реляционная система управления базами данных с открытым исходным кодом	39
Высокая производительность	39
Высокая доступность	39
Кросс-платформенность	39
Инсталляция MySQL 8	40
Получение MySQL 8	40
Инсталляция MySQL 8	40
Служебные команды MySQL	41
Эволюция MySQL для больших данных	42
Получение данных в MySQL	43
Организация данных в Hadoop	43
Аналитическая обработка данных	43
Результаты анализа	43
Резюме	44
Глава 2. Методы запроса данных в MySQL 8	45
Обзор SQL	45
Подсистемы (движки) хранения и их разновидности	46
InnoDB	48
MyISAM	49
Memory	50
Archive	50
Blackhole	51
CSV	51
Merge	51
Federated	52
NDB Cluster	53
Оператор SELECT в MySQL 8	54
Оператор WHERE	54
Предложение ORDER BY	56
Предложение LIMIT	57
Операции соединения SQL	57
UNION	59
Оптимизация запросов SELECT	60
Операторы INSERT, REPLACE и UPDATE в MySQL 8	62
INSERT	62
UPDATE	62
REPLACE	62
Транзакции в MySQL 8	63
Агрегирование данных в MySQL 8	63
Важность агрегатных функций	64
JSON	66
JSON_OBJECTAGG	67
JSON_ARRAYAGG	68
Резюме	70

Глава 3. Индексирование данных для высокопроизводительных запросов	71
Индексирование в MySQL	72
Индексные структуры	72
Создание или удаление индексов	75
Типы индексов СУБД MySQL 8	78
Определение первичного индекса	78
Уникальные ключи	80
Определение столбцового индекса	80
Полнотекстовая индексация	85
Пространственные индексы	89
Индексирование данных JSON	90
Генерируемые столбцы	90
Определение индексов на JSON	92
Резюме	94
Глава 4. Использование Memcached в MySQL 8	95
Обзор Memcached	95
Настройка плагина Memcached	97
Инсталляция	97
Верификация	98
Использование плагина Memcached	99
Наладчик производительности	99
Инструмент кеширования	99
Простота в использовании	99
Анализ хранящихся в Memcached данных	100
Конфигурирование репликации Memcached	101
API Memcached для различных технологий	103
Memcached с Java	103
Memcached с PHP	105
Memcached с Ruby	105
Memcached с Python	106
Резюме	106
Глава 5. Разделение больших объемов данных	107
Разделение данных в MySQL 8	108
Что такое разделение данных?	108
Типы разделения данных	109
Горизонтальное разделение в MySQL 8	109
Диапазонное разделение	110
Списковое разделение	112
Хеш-разделение	112
Столбцовое разделение	114
Разделение по ключу	116
Разбиение на подразделы	117
Вертикальное разделение	118

Разделение данных на многочисленные таблицы	119
Подрезание разделов в MySQL	122
Подрезание со списковым разделением	125
Подрезание с разделением по ключу	125
Выполнение запросов на разделенных данных	126
Запрос DELETE с параметром PARTITION	128
Запрос UPDATE с параметром PARTITION	129
Запрос INSERT с параметром PARTITION	129
Резюме	130

Глава 6. Репликация для построения высокодоступных

решений	131
Высокая доступность	131
Репликация в MySQL	132
Кластер MySQL	132
Облачная служба Oracle MySQL	133
MySQL с кластером Solaris	133
Репликация с помощью MySQL	134
Преимущества репликации в MySQL 8	134
Методы репликации в MySQL 8	135
Конфигурация репликации	135
Групповая репликация	148
Предварительные условия для групповой репликации	149
Конфигурирование групповой репликации	149
Конфигурирование пользователя репликации и активация плагина групповой репликации	151
Запуск групповой репликации	152
Резюме	153

Глава 7. Практические рекомендации по работе с MySQL 8

Сравнительные испытания и конфигурации MySQL	155
Использование ресурсов	155
Увеличение длительности нагрузочных тестов	155
Репликация параметров производственной среды	156
Сопоставимость пропускной способности и задержки	156
Sysbench может сделать больше	156
Мир виртуализации	156
Параллелизм	156
Фоновая нагрузка	157
Суть вашего запроса	157
Сравнительные испытания	157
Рекомендации в отношении вопросов MySQL	159
Типы данных	159
Not null	160
Индексация	160
Извлекайте все данные	161

Приложение сделает работу.....	161
Существование данных.....	161
Ограничивайте себя	161
Анализируйте медленные запросы.....	161
Стоимость запроса	161
Рекомендации в отношении конфигурации Memcached	162
Распределение ресурсов	162
Архитектура операционной системы	162
Конфигурации по умолчанию	162
Максимальный размер объекта.....	163
Ограничение очереди незавершенных заданий.....	163
Поддержка больших страниц.....	163
Конфиденциальные данные	163
Ограничение открытости.....	163
Отказоустойчивость	164
Пространства имен.....	164
Механизм кеширования	164
Общая статистика Memcached	164
Рекомендации в отношении репликации.....	166
Пропускная способность в групповой репликации	166
Определение размеров инфраструктуры	166
Постоянная пропускная способность	166
Неподходящая нагрузка	166
Масштабируемость операции записи	167
Резюме.....	167

Глава 8. Прикладной программный интерфейс NoSQL

для интеграции с решениями для больших данных	169
Обзор NoSQL	169
Быстрое изменение с течением времени	170
Масштабирование	170
Меньше управленческой деятельности	170
Лучшее для больших данных.....	171
NoSQL против SQL	171
Реализация API NoSQL	171
NoSQL со слоем API Memcached	172
NDB API Cluster	178
Резюме.....	189

Глава 9. Практический пример: часть I. Apache Sqoop

для обмена данными между MySQL и платформой Hadoop.....	190
Практический пример анализа журналов операций.....	191
Использование MySQL 8 и Hadoop для анализа журналов операций.....	191
Обзор Apache Sqoop.....	192
Интеграция Apache Sqoop с MySQL и Hadoop.....	194
Hadoop	194

Настройка Hadoop в Linux.....	196
Инсталляция Apache Sqoop	198
Конфигурирование коннектора MySQL	199
Импортирование неструктурированных данных в Hadoop HDFS из MySQL	199
Импорт Sqoop для извлечения данных из MySQL 8	199
Инкрементный импорт с использованием Sqoop	202
Загрузка структурированных данных в MySQL с помощью Apache Sqoop.....	202
Экспорт Sqoop для хранения структурированных данных в MySQL 8.....	202
Сохраненные задания Sqoop	204
Резюме	205

Глава 10. Практический пример: часть II. Обработка событий

в режиме реального времени с помощью MySQL Applier..... 206

Обзор практического примера	206
MySQL Applier	208
Дамп и импорт SQL	208
Sqoop.....	209
Репликатор Tungsten.....	209
Apache Kafka.....	209
Talend.....	210
Dell Shareplex	210
Сравнение инструментов.....	210
Обзор MySQL Applier	210
Инсталляция MySQL Applier	212
Интеграция в режиме реального времени с MySQL Applier.....	214
Организация и анализ данных в Hadoop	216
Резюме.....	218

Предметный указатель 219