

УДК 004.42

ББК 32.973

С44

Дилан Скотт, Виктор Гамов, Дейв Клейн

С44 Kafka в действии / пер. с англ. А. Н. Киселева. – М.: ДМК Пресс, 2022. – 310 с.: ил.

ISBN 978-5-93700-118-4

Это практическое руководство показывает, как использовать распределенную потоковую платформу Apache Kafka для удовлетворения различных бизнес-требований. Рассказывается, как устроена Kafka и где она может пригодиться на практике; описываются характеристики проектов, в которых может пригодиться эта платформа. Рассматриваются основные ее компоненты – клиенты и кластер, представлены варианты улучшения работающего кластера.

Книга адресована разработчикам, желающим ознакомиться с идеей потоковой обработки данных. Для изучения примеров кода понадобятся базовые знания командной строки; желательно иметь навыки программирования на языке Java.

УДК 004.42

ББК 32.973

Original English language edition published by Manning Publications, USA.
Russian-language edition copyright (c) 2022 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 (анг.) 978-1-61729-523-2

ISBN (рус.) 978-5-93700-118-4

© 2022 by Manning Publications Co.

© Оформление, издание, перевод,
ДМК Пресс, 2022

Содержание

<i>Предисловие от издательства</i>	13
<i>Предисловие</i>	14
<i>Вступление</i>	15
<i>Благодарности</i>	16
<i>Об этой книге</i>	18
<i>Об авторах</i>	22
<i>Об иллюстрации на обложке</i>	23

ЧАСТЬ I. НАЧАЛО 25

1	<i>Введение в Kafka</i>	26
	1.1. Что такое Kafka?	27
	1.2. Использование Kafka.....	32
	1.2.1. Kafka – разработчикам	32
	1.2.2. Как преподнести Kafka вашему руководству	34
	1.3. Мифы о Kafka	35
	1.3.1. Kafka работает только с Hadoop®.....	35
	1.3.2. Kafka ничем не отличается от других брокеров сообщений	36
	1.4. Kafka в реальном мире	37
	1.4.1. Ранние примеры.....	37
	1.4.2. Более поздние примеры	39
	1.4.3. Когда Kafka может быть неприменима.....	40
	1.5. Онлайн-ресурсы	41
	Итоги	42
	Ссылки	42
2	<i>Знакомство с Kafka</i>	44
	2.1. Отправка и прием сообщения	45
	2.2. Что такое брокер?	46
	2.3. Экскурсия по Kafka	51
	2.3.1. Производители и потребители	51
	2.3.2. Темы	55

2.3.3. ZooKeeper	56
2.3.4. Высокоуровневая архитектура Kafka.....	58
2.3.5. Журнал коммитов	59
2.4. Различные пакеты исходного кода, и что они делают	60
2.4.1. Kafka Streams	60
2.4.2. Kafka Connect.....	62
2.4.3. Пакет AdminClient	62
2.4.4. ksqlDB.....	63
2.5. Клиенты Confluent	63
2.6. Поточковая обработка и терминология	67
2.6.1. Поточковая обработка	69
2.6.2. Что означает семантика «точно один раз»	69
Итоги	70
Ссылки	70

ЧАСТЬ II. ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ КАФКА.... 73

3 Разработка проекта на основе Kafka..... 75

3.1. Разработка проекта на основе Kafka.....	76
3.1.1. Использование существующей архитектуры данных	76
3.1.2. Первый шаг	76
3.1.3. Встроенные возможности	77
3.1.4. Данные для наших накладных	80
3.2. События датчиков.....	82
3.2.1. Имеющиеся проблемы	82
3.2.2. Почему Kafka – правильный выбор.....	85
3.2.3. Первые мысли об архитектуре	86
3.2.4. Требования к пользовательским данным.....	88
3.2.5. Общий план с учетом поставленных вопросов.....	88
3.2.6. Обзор и оценка плана.....	92
3.3. Формат представления данных	93
3.3.1. План для данных	93
3.3.2. Настройка зависимостей.....	95
Итоги	101
Ссылки	101

4 Производители: источники данных 103

4.1. Пример	104
4.1.1. Примечания в отношении производителя.....	107
4.2. Параметры производителя	108
4.2.1. Настройка списка брокеров.....	109

4.2.2. Быстрее или надежнее?	110
4.2.3. Отметки времени	113
4.3. Генерирование кода с учетом наших требований	115
4.3.1. Версии клиентов и брокеров	124
Итоги	125
Ссылки	125

5 Потребители: извлечение данных..... 127

5.1. Пример	128
5.1.1. Параметры потребителя.....	129
5.1.2. Наши координаты в потоке событий	133
5.2. Как взаимодействуют потребители	137
5.3. Трассировка	138
5.3.1. Координатор группы.....	139
5.3.2. Стратегия назначения разделов.....	141
5.4. Маркировка местонахождения.....	142
5.5. Чтение из сжатой темы.....	145
5.6. Реализация в коде наших заводских требований	145
5.6.1. Варианты чтения	146
5.6.2. Требования.....	148
Итоги	151
Ссылки	151

6 Брокеры 155

6.1. Знакомство с брокерами.....	155
6.2. Роль ZooKeeper.....	156
6.3. Конфигурационные параметры брокеров	158
6.3.1. Другие журналы Kafka: журналы приложений.....	160
6.3.2. Журнал сервера.....	160
6.3.3. Управление состоянием.....	160
6.4. Ведущие реплики разделов и их роль.....	162
6.4.1. Потеря данных	164
6.5. Взгляд внутрь Kafka.....	165
6.5.1. Обслуживание кластера	167
6.5.2. Добавление брокера	167
6.5.3. Обновление кластера	167
6.5.4. Обновление клиентов	168
6.5.5. Резервные копии.....	168
6.6. Примечание о системах с сохранением состояния.....	169
6.7. Упражнение	171
Итоги	172
Ссылки	173

7	<i>Темы и разделы.....</i>	<i>176</i>
7.1.	Темы	176
7.1.1.	Параметры создания темы	180
7.1.2.	Коэффициенты репликации.....	182
7.2.	Разделы	183
7.2.1.	Размещение раздела.....	183
7.2.2.	Просмотр журналов.....	184
7.3.	Тестирование с помощью EmbeddedKafkaCluster.....	186
7.3.1.	Использование Kafka Testcontainers	188
7.4.	Сжатые темы.....	188
	Итоги	191
	Ссылки	191
8	<i>Kafka как хранилище</i>	<i>193</i>
8.1.	Как долго можно хранить данные	194
8.2.	Перемещение данных	195
8.2.1.	Сохранение исходных событий	195
8.2.2.	Отказ от пакетного мышления	196
8.3.	Инструменты	196
8.3.1.	Apache Flume	197
8.3.2.	Red Hat® Debezium™.....	199
8.3.3.	Secor	200
8.3.4.	Пример сохранения данных	201
8.4.	Возврат данных в Kafka.....	201
8.4.1.	Многоуровневое хранилище.....	203
8.5.	Архитектуры с использованием Kafka.....	203
8.5.1.	Лямбда-архитектура.....	203
8.5.2.	Каппа-архитектура	205
8.6.	Окружения с несколькими кластерами	206
8.6.1.	Масштабирование путем добавления кластеров	206
8.7.	Варианты хранения в облаке и в контейнерах	207
8.7.1.	Кластеры Kubernetes	207
	Итоги	208
	Ссылки	208
9	<i>Управление: инструменты и журналы.....</i>	<i>211</i>
9.1.	Клиенты администрирования	212
9.1.1.	Решение задач администрирования в коде с помощью AdminClient	212
9.1.2.	kcat	214
9.1.3.	Confluent REST Proxy API	216

9.2. Запуск Kafka как службы systemd	217
9.3. Журналы.....	218
9.3.1. Журналы приложений Kafka.....	219
9.3.2. Журналы ZooKeeper	220
9.4. Брандмауэры.....	221
9.4.1. Публикуемые слушатели	221
9.5. Метрики.....	222
9.5.1. Консоль JMX	222
9.6. Способы трассировки	225
9.6.1. Логика на стороне производителя.....	226
9.6.2. Логика на стороне потребителя.....	228
9.6.3. Переопределение клиентов	230
9.7. Общие инструменты мониторинга.....	231
Итоги	232
Ссылки	232

ЧАСТЬ III. ДОПОЛНИТЕЛЬНЫЕ ВОЗМОЖНОСТИ ... 235

10 <i>Защита Kafka</i>	236
10.1. Основы безопасности	238
10.1.1. Шифрование с помощью SSL.....	239
10.1.2. Настройка соединений SSL между брокерами и клиентами	240
10.1.3. Настройка соединений SSL между брокерами.....	244
10.2. Kerberos и Simple Authentication and Security Layer (SASL).....	244
10.3. Авторизация в Kafka	245
10.3.1. Списки управления доступом.....	246
10.3.2. Управление доступом на основе ролей.....	247
10.4. ZooKeeper.....	248
10.4.1. Настройка Kerberos	248
10.5. Квоты	249
10.5.1. Ограничение пропускной способности сети	250
10.5.2. Ограничение частоты запросов	252
10.6. Данные в состоянии покоя.....	252
10.6.1. Управляемые варианты.....	253
Итоги	253
Ссылки	254

11 <i>Реестр схем</i>	256
------------------------------------	------------

11.1. Предлагаемая модель зрелости Kafka.....	257
11.1.1. Уровень 0	257

11.1.2. Уровень 1	258
11.1.3. Уровень 2	259
11.1.4. Уровень 3	259
11.2. Реестр схем	260
11.2.1. Установка Confluent Schema Registry	260
11.2.2. Конфигурация реестра	261
11.3. Компоненты реестра схем	262
11.3.1. REST API	262
11.3.2. Клиентская библиотека	263
11.4. Правила совместимости	265
11.4.1. Проверка изменений схемы	266
11.5. Альтернатива реестру схем	267
Итоги	268
Ссылки	268

12 *Потоковая обработка с помощью Kafka Streams и ksqlDB* 270

12.1. Kafka Streams	271
12.1.1. KStreams API DSL	273
12.1.2. KTable API	277
12.1.3. GlobalKTable API	278
12.1.4. Processor API	279
12.1.5. Настройка Kafka Streams	281
12.2. ksqlDB: база данных потоковой передачи событий	282
12.2.1. Запросы	284
12.2.2. Локальная разработка	284
12.2.3. Архитектура ksqlDB	286
12.3. Куда пойти дальше	287
12.3.1. Предложения по улучшению Kafka (KIP)	287
12.3.2. Проекты Kafka, которые вы можете исследовать ...	288
12.3.3. Каналы сообщества Slack	288
Итоги	288
Ссылки	289

Приложение А. Установка..... 290

Приложение В. Пример клиента..... 299

Предметный указатель 304