
Оглавление

Отзывы о книге.....	17
Предисловие ко второму изданию.....	19
Предисловие к первому изданию.....	21
Введение.....	24
Для кого предназначена эта книга.....	25
Условные обозначения.....	25
Использование примеров кода.....	26
Благодарности.....	27
От издательства.....	29
Глава 1. Знакомьтесь: Kafka.....	30
Обмен сообщениями по типу «публикация/подписка».....	30
С чего все начинается.....	31
Отдельные системы организации очередей.....	33
Открываем для себя систему Kafka.....	33
Сообщения и пакеты.....	36
Схемы.....	36
Топики и разделы.....	37
Производители и потребители.....	38
Брокеры и кластеры.....	39
Несколько кластеров.....	41
Почему Kafka?.....	43
Несколько производителей.....	43
Несколько потребителей.....	43
Сохранение информации на диске.....	43
Масштабируемость.....	44
Высокое быстродействие.....	44
Особенности платформы.....	44

Экосистема данных.....	45
Сценарии использования.....	46
История создания Kafka	48
Проблема LinkedIn	48
Рождение Kafka.....	49
Открытый исходный код	50
Коммерческое взаимодействие	51
Название	51
Приступаем к работе с Kafka.....	51
Глава 2. Установка Kafka	52
Настройка среды	52
Выбрать операционную систему.....	52
Установить Java.....	52
Установить ZooKeeper	53
Установка брокера Kafka.....	56
Настройка брокера.....	58
Основные параметры брокера	58
Настройки топиков по умолчанию.....	61
Выбор аппаратного обеспечения.....	67
Пропускная способность дисков	67
Емкость диска.....	68
Память.....	68
Передача данных по сети	69
CPU	69
Kafka в облачной среде	70
Microsoft Azure	70
Веб-сервисы Amazon Web Services.....	71
Настройка кластеров Kafka.....	71
Сколько должно быть брокеров	72
Конфигурация брокеров.....	74
Тонкая настройка операционной системы	74
Промышленная эксплуатация.....	78
Параметры сборки мусора.....	78
Планировка ЦОД.....	80
Размещение приложений на ZooKeeper	80
Резюме.....	82

Глава 3. Производители Kafka: запись сообщений в Kafka	83
Обзор производителя	84
Создание производителя Kafka	86
Отправка сообщения в Kafka	88
Синхронная отправка сообщения	89
Асинхронная отправка сообщения	90
Настройка производителей	91
client.id	91
acks	92
Время доставки сообщения	93
linger.ms	96
buffer.memory	96
compression.type	97
batch.size	97
max.in.flight.requests.per.connection	97
max.request.size	98
receive.buffer.bytes и send.buffer.bytes	98
enable.idempotence	99
Сериализаторы	99
Пользовательские сериализаторы	100
Сериализация с помощью Apache Avro	102
Использование записей Avro с Kafka	104
Разделы	107
Реализация пользовательской стратегии секционирования	108
Заголовки	110
Перехватчики	110
Квоты и регулирование запросов	112
Резюме	114
Глава 4. Потребители Kafka: чтение данных из Kafka	115
Принципы работы потребителей Kafka	115
Потребители и группы потребителей	115
Группы потребителей и перебалансировка разделов	118
Статические участники группы	122
Создание потребителя Kafka	123
Подписка на топики	123

Цикл опроса.....	124
Потокобезопасность.....	126
Настройка потребителей.....	127
fetch.min.bytes.....	127
fetch.max.wait.ms.....	127
fetch.max.bytes.....	128
max.poll.records.....	128
max.partition.fetch.bytes.....	128
session.timeout.ms и heartbeat.interval.ms.....	128
max.poll.interval.ms.....	129
default.api.timeout.ms.....	130
request.timeout.ms.....	130
auto.offset.reset.....	130
enable.auto.commit.....	131
partition.assignment.strategy.....	131
client.id.....	132
client.rack.....	132
group.instance.id.....	133
receive.buffer.bytes и send.buffer.bytes.....	133
offsets.retention.minutes.....	133
Фиксация и смещения.....	134
Автоматическая фиксация.....	135
Фиксация текущего смещения.....	136
Асинхронная фиксация.....	137
Сочетание асинхронной и синхронной фиксации.....	139
Фиксация заданного смещения.....	140
Прослушивание на предмет перебалансировки.....	141
Получение записей с заданными смещениями.....	144
Выход из цикла.....	145
Десериализаторы.....	147
Пользовательские сериализаторы.....	148
Использование десериализации Avro в потребителе Kafka.....	150
Автономный потребитель: зачем и как использовать потребитель без группы.....	151
Резюме.....	153

Глава 5. Программное управление Apache Kafka	154
Обзор AdminClient	155
Асинхронный и в конечном итоге согласованный API	155
Опции	156
Плоская иерархия	156
Дополнительные примечания	156
Жизненный цикл AdminClient: создание, настройка и закрытие	157
client.dns.lookup	158
request.timeout.ms	159
Управление основными топиками	160
Управление конфигурацией	164
Управление группами потребителей	165
Изучение групп потребителей	166
Модификация групп потребителей	168
Метаданные кластера	170
Расширенные операции администратора	170
Добавление разделов в топик	170
Удаление записей из топика	171
Выборы лидера	172
Перераспределение реплик	173
Тестирование	174
Резюме	177
Глава 6. Внутреннее устройство Kafka	178
Членство в кластере	178
Контроллер	179
KRaft: новый контроллер Kafka на основе Raft	181
Репликация	183
Обработка запросов	186
Запросы от производителей	189
Запросы на извлечение	189
Другие запросы	194
Физическое хранилище	195
Многоуровневое хранилище	196
Распределение разделов	198
Управление файлами	200
Формат файлов	200

Индексы	203
Сжатие	204
Как происходит сжатие	204
Удаленные события	206
Когда выполняется сжатие топиков.....	207
Резюме.....	208
Глава 7. Надежная доставка данных.....	209
Гарантии надежности	210
Репликация.....	211
Настройка брокера.....	212
Коэффициент репликации	213
«Нечистый» выбор ведущей реплики	214
Минимальное число согласованных реплик.....	216
Поддержание синхронизации реплик	217
Долговременное хранение на диске.....	217
Использование производителей в надежной системе.....	218
Отправка подтверждений	219
Настройка повторов отправки производителями.....	220
Дополнительная обработка ошибок	221
Использование потребителей в надежной системе	221
Свойства конфигурации потребителей, важные для надежной обработки.....	222
Фиксация смещений в потребителях явным образом	224
Проверка надежности системы.....	226
Проверка конфигурации	226
Проверка приложений	228
Мониторинг надежности при промышленной эксплуатации	228
Резюме.....	230
Глава 8. Семантика «точно один раз».....	231
Идемпотентный производитель.....	232
Как работает идемпотентный производитель	232
Ограничения идемпотентного производителя.....	235
Как использовать идемпотентный производитель Kafka	236
Транзакции	237
Сценарии использования транзакций	237
Какие проблемы решают транзакции	238

Как транзакции гарантируют «точно один раз»	239
Какие проблемы не решаются транзакциями.....	242
Как использовать транзакции	245
Идентификаторы транзакций и ограждения.....	248
Как работают транзакции	250
Производительность транзакций.....	252
Резюме.....	253
Глава 9. Создание конвейеров данных	254
Соображения по поводу создания конвейеров данных.....	255
Своевременность	255
Надежность	256
Высокая/переменная нагрузка	257
Форматы данных.....	257
Преобразования	258
Безопасность	259
Обработка сбоев.....	260
Связывание и гибкость	261
Когда использовать Kafka Connect, а когда — клиенты-производители и клиенты-потребители.....	262
Kafka Connect.....	263
Запуск Kafka Connect.....	263
Пример коннектора: файловый источник и файловый приемник.....	266
Пример коннектора: из MySQL в Elasticsearch	269
Преобразования одиночных сообщений	276
Взглянем на Kafka Connect поближе.....	278
Альтернативы Kafka Connect.....	282
Фреймворки ввода и обработки данных для других хранилищ	282
ETL-утилиты на основе GUI.....	283
Фреймворки потоковой обработки.....	283
Резюме.....	283
Глава 10. Зеркальное копирование между кластерами	285
Сценарии зеркального копирования данных между кластерами	286
Мультикластерные архитектуры.....	287
Реалии взаимодействия между различными ЦОД.....	287
Архитектура с топологией типа «звезда»	289
Архитектура типа «активный — активный»	291

Архитектура типа «активный — резервный»	293
Эластичные кластеры	301
Утилита MirrorMaker (Apache Kafka)	302
Настройка MirrorMaker	304
Топология мультикластерной репликации	307
Обеспечение безопасности MirrorMaker	308
Развертывание MirrorMaker для промышленной эксплуатации	309
Тонкая настройка MirrorMaker	314
Другие программные решения для зеркального копирования между кластерами	317
uReplicator компании Uber	317
LinkedIn Brooklin	318
Решения Confluent для зеркального копирования между ЦОД	319
Резюме	322
Глава 11. Обеспечение безопасности Kafka	323
Блокировка Kafka	324
Протоколы безопасности	326
Аутентификация	328
SSL	329
SASL	334
Повторная аутентификация	347
Обновления системы безопасности без простоя	349
Шифрование	350
Сквозное шифрование	351
Авторизация	353
AclAuthorizer	354
Настройка авторизации	358
Вопросы безопасности	360
Аудит	361
Обеспечение безопасности ZooKeeper	362
SASL	362
SSL	363
Авторизация	364
Обеспечение безопасности платформы	364
Защита паролей	365
Резюме	367

Глава 12. Администрирование Kafka	369
Операции с топиками.....	369
Создание нового топика	370
Вывод списка всех топиков в кластере	371
Подробное описание топиков.....	372
Добавление разделов.....	373
Уменьшение количества разделов.....	374
Удаление топика.....	375
Группы потребителей.....	376
Вывод списка и описание групп	376
Удаление группы.....	377
Управление смещениями	378
Динамические изменения конфигурации	379
Переопределение значений настроек топиков по умолчанию	380
Переопределение настроек клиентов и пользователей по умолчанию ...	382
Переопределение настроек конфигурации брокера по умолчанию	383
Описание переопределений настроек	384
Удаление переопределений настроек	385
Производство и потребление	385
Консольный производитель.....	385
Консольный потребитель	388
Управление разделами.....	391
Выбор предпочтительной ведущей реплики	391
Изменение реплик раздела	393
Сброс на диск сегментов журнала.....	398
Проверка реплик.....	400
Другие утилиты	401
Небезопасные операции	402
Перенос контроллера кластера	402
Отмена удаления топиков	403
Удаление топиков вручную.....	403
Резюме.....	404
Глава 13. Мониторинг Kafka	405
Основы показателей	405
Как получить доступ к показателям.....	405
Какие показатели нам нужны.....	407
Контроль состояния приложения.....	409

Цели на уровне обслуживания	410
Определения уровня сервиса	410
Какие показатели являются хорошими индикаторами уровня обслуживания	411
Использование целей уровня обслуживания для оповещений	412
Показатели брокеров Kafka.....	414
Диагностика проблем с кластером	414
Искусство недореплицированных разделов	416
Показатели брокеров	422
Показатели топиков и разделов	432
Мониторинг JVM	435
Мониторинг ОС	436
Журналирование	438
Мониторинг клиентов.....	439
Показатели производителя	439
Показатели потребителей	442
Квоты	446
Мониторинг отставания.....	447
Сквозной мониторинг	448
Резюме.....	449
Глава 14. Поточковая обработка	450
Что такое потоковая обработка.....	452
Основные понятия потоковой обработки.....	455
Топология.....	455
Время.....	456
Состояние.....	458
Таблично-поточковый дуализм	459
Временные окна.....	461
Гарантии обработки	462
Паттерны проектирования потоковой обработки.....	462
Обработка событий по отдельности	463
Обработка с использованием локального состояния	463
Многоэтапная обработка/повторное разделение на разделы.....	465
Обработка с применением внешнего справочника: соединение потока данных с таблицей.....	467
Соединение таблицы с таблицей	468
Соединение потоков.....	470

Внеочередные события	471
Повторная обработка	472
Интерактивные запросы	473
Кafka Streams в примерах	474
Подсчет количества слов	474
Сводные показатели фондовой биржи	477
Обогащение потока событий перехода по ссылкам	480
Кafka Streams: обзор архитектуры	483
Построение топологии	483
Оптимизация топологии	484
Тестирование топологии	484
Масштабирование топологии	485
Как пережить отказ	489
Сценарии использования потоковой обработки	490
Как выбрать фреймворк потоковой обработки	492
Резюме	494
Приложение А. Установка Kafka в других операционных системах	495
Установка в Windows	495
Использование Windows Subsystem для Linux	495
Использование Java естественным образом	496
Установка в macOS	498
Использование Homebrew	499
Установка вручную	500
Приложение Б. Дополнительные инструменты Kafka	501
Комплексные платформы	501
Развертывание и управление кластером	503
Мониторинг и исследование данных	505
Клиентские библиотеки	506
Потоковая обработка	507
Об авторах	509
Иллюстрация на обложке	511