

Научная библиотека

БНТУ



* 8 0 1 2 3 4 1 3 9 *

Ния Нархид, Гвен Шапира, Тодд Палино

Арасче Kafka

ПОТОВОЯ ОБРАБОТКА И АНАЛИЗ ДАННЫХ

НАВУКОВАЯ БІБЛІЯТЭКА

Беларускага нацыянальнага
тэхнічнага ўніверсітэта

Інв. № 1883836



Санкт-Петербург • Москва • Минск

2022

Краткое содержание

Предисловие	15
Введение.....	18
Глава 1. Знакомьтесь: Kafka.....	22
Глава 2. Установка Kafka.....	42
Глава 3. Производители Kafka: запись сообщений в Kafka	65
Глава 4. Потребители Kafka: чтение данных из Kafka.....	88
Глава 5. Внутреннее устройство Kafka	120
Глава 6. Надежная доставка данных.....	140
Глава 7. Создание конвейеров данных.....	161
Глава 8. Зеркальное копирование между кластерами.....	184
Глава 9. Администрирование Kafka.....	213
Глава 10. Мониторинг Kafka	243
Глава 11. Поточковая обработка.....	278
Приложение. Установка Kafka на других операционных системах.....	315

Мия Нархид, Гвен Шапира, Тодд Палино
Арасье Кафка. Потоковая обработка и анализ данных

Серия «Бестселлеры O'Reilly»

Перевод с английского И. Пальты

Оглавление

Предисловие	15
Введение.....	18
Для кого предназначена эта книга.....	19
Условные обозначения.....	19
Использование примеров кода	20
Благодарности	21
Глава 1. Знакомьтесь: Kafka	22
Обмен сообщениями по типу «публикация/подписка».....	22
С чего все начинается	23
Отдельные системы организации очередей.....	26
Открываем для себя систему Kafka	26
Сообщения и пакеты.....	26
Схемы.....	28
Темы и разделы.....	28
Производители и потребители.....	29
Брокеры и кластеры.....	31
Несколько кластеров	32
Почему Kafka?	34
Несколько производителей.....	34
Несколько потребителей.....	34
Сохранение информации на диске	34
Масштабируемость.....	35
Высокое быстродействие	35
Экосистема данных.....	35
Сценарии использования.....	36

История создания Kafka	38
Проблема LinkedIn	38
Рождение Kafka	40
Открытый исходный код.....	40
Название	41
Приступаем к работе с Kafka.....	41
Глава 2. Установка Kafka.....	42
Обо всем по порядку	42
Выбрать операционную систему	42
Установить Java.....	42
Установить ZooKeeper	43
Установка брокера Kafka	45
Конфигурация брокера.....	46
Основные настройки брокера.....	47
Настройки тем по умолчанию	49
Выбор аппаратного обеспечения	53
Пропускная способность дисков	54
Емкость диска	54
Память	55
Передача данных по сети.....	55
CPU.....	55
Кafka в облачной среде	56
Кластеры Kafka	56
Сколько должно быть брокеров?	57
Конфигурация брокеров.....	58
Тонкая настройка операционной системы.....	58
Промышленная эксплуатация	61
Параметры сборки мусора	61
Планировка ЦОД.....	62
Размещение приложений на ZooKeeper	63
Резюме.....	64
Глава 3. Производители Kafka: запись сообщений в Kafka	65
Обзор производителя.....	66
Создание производителя Kafka	68
Отправка сообщения в Kafka	70

Синхронная отправка сообщения	71
Асинхронная отправка сообщения	71
Настройка производителей.....	72
acks	73
buffer.memory	73
compression.type.....	74
retries	74
batch.size	75
linger.ms	75
client.id	75
max.in.flight.requests.per.connection	75
timeout.ms, request.timeout.ms и metadata.fetch.timeout.ms.....	76
max.block.ms	76
max.request.size	76
receive.buffer.bytes и send.buffer.bytes	76
Сериализаторы	77
Пользовательские сериализаторы	77
Сериализация с помощью Apache Avro	79
Использование записей Avro с Kafka	81
Разделы	84
Старые API производителей	86
Резюме	87
Глава 4. Потребители Kafka: чтение данных из Kafka.....	88
Принципы работы потребителей Kafka.....	88
Потребители и группы потребителей	88
Группы потребителей и ребалансировка разделов.....	92
Создание потребителя Kafka	94
Подписка на темы.....	94
Цикл опроса.....	95
Настройка потребителей	97
fetch.min.bytes.....	97
fetch.max.wait.ms.....	97
max.partition.fetch.bytes	98
session.timeout.ms	98
auto.offset.reset.....	99
enable.auto.commit.....	99

partition.assignment.strategy	99
client.id	100
max.poll.records.....	100
receive.buffer.bytes и send.buffer.bytes	100
Фиксация и смещения.....	101
Автоматическая фиксация.....	102
Фиксация текущего смещения.....	103
Асинхронная фиксация	104
Сочетание асинхронной и синхронной фиксации	105
Фиксация заданного смещения	106
Прослушивание на предмет перебалансировки	107
Получение записей с заданными смещениями	109
Выход из цикла.....	112
Десериализаторы.....	113
Пользовательские сериализаторы	114
Использование десериализации Avro в потребителе Kafka	116
Автономный потребитель: зачем и как использовать потребитель без группы	117
Старые API потребителей	118
Резюме.....	119
Глава 5. Внутреннее устройство Kafka	120
Членство в кластере.....	120
Контроллер.....	121
Репликация.....	122
Обработка запросов	124
Запросы от производителей.....	127
Запросы на извлечение	127
Другие запросы.....	129
Физическое хранилище	131
Распределение разделов	131
Управление файлами.....	133
Формат файлов.....	134
Индексы	136
Сжатие	136
Как происходит сжатие.....	137
Удаленные события	138
Когда выполняется сжатие тем	139
Резюме.....	139

Глава 6. Надежная доставка данных	140
Гарантии надежности.....	141
Репликация.....	142
Настройка брокера.....	143
Коэффициент репликации	143
«Нечистый» выбор ведущей реплики	145
Минимальное число согласованных реплик.....	146
Использование производителей в надежной системе	147
Отправка подтверждений	148
Настройка повторов отправки производителями	149
Дополнительная обработка ошибок.....	150
Использование потребителей в надежной системе	151
Свойства конфигурации потребителей, важные для надежной обработки	152
Фиксация смещений в потребителях явным образом	153
Проверка надежности системы.....	156
Проверка конфигурации	157
Проверка приложений	158
Мониторинг надежности при промышленной эксплуатации.....	158
Резюме.....	160
Глава 7. Создание конвейеров данных	161
Соображения по поводу создания конвейеров данных	162
Своевременность.....	162
Надежность	163
Высокая/переменная нагрузка	164
Форматы данных.....	164
Преобразования.....	165
Безопасность.....	166
Обработка сбоев	166
Связывание и быстрота адаптации.....	167
Когда использовать Kafka Connect, а когда клиенты-производители и клиенты-потребители.....	168
Kafka Connect.....	168
Запуск Connect	169
Пример коннектора: файловый источник и файловый приемник.....	171
Пример коннектора: из MySQL в Elasticsearch.....	172
Взглянем на Connect поближе	178

Альтернативы Kafka Connect	181
Фреймворки ввода и обработки данных для других хранилищ.....	182
ETL-утилиты на основе GUI	182
Фреймворки потоковой обработки	182
Резюме.....	183
Глава 8. Зеркальное копирование между кластерами.....	184
Сценарии зеркального копирования данных между кластерами.....	185
Мультикластерные архитектуры	186
Реалии взаимодействия между различными ЦОД.....	186
Архитектура с топологией типа «звезда»	187
Архитектура типа «активный – активный».....	189
Архитектура типа «активный – резервный»	192
Потери данных и несогласованности при внеплановом восстановлении после сбоя.....	193
Начальное смещение для приложений после аварийного переключения.....	194
После аварийного переключения.....	198
Несколько слов об обнаружении кластеров.....	198
Эластичные кластеры.....	199
Утилита MirrorMaker (Apache Kafka)	200
Настройка MirrorMaker.....	201
Развертывание MirrorMaker для промышленной эксплуатации	202
Тонкая настройка MirrorMaker	206
Другие программные решения для зеркального копирования между кластерами	209
uReplicator компании Uber.....	209
Replicator компании Confluent.....	210
Резюме.....	211
Глава 9. Администрирование Kafka.....	213
Операции с темами	213
Создание новой темы.....	214
Добавление разделов	215
Удаление темы	216
Вывод списка всех тем кластера	216
Подробное описание тем	217

Группы потребителей.....	218
Вывод списка и описание групп	218
Удаление группы.....	220
Управление смещениями.....	220
Динамические изменения конфигурации	222
Переопределение значений настроек тем по умолчанию.....	222
Переопределение настроек клиентов по умолчанию	224
Описание переопределений настроек.....	225
Удаление переопределений настроек	225
Управление разделами	225
Выбор предпочтительной ведущей реплики.....	226
Смена реплик раздела	227
Изменение коэффициента репликации	230
Сброс на диск сегментов журнала.....	231
Проверка реплик	233
Потребление и генерация	234
Консольный потребитель.....	234
Консольный производитель.....	237
Списки управления доступом клиентов	239
Небезопасные операции.....	239
Перенос контроллера кластера	240
Отмена перемещения раздела	240
Отмена удаления тем.....	241
Удаление тем вручную.....	241
Резюме.....	242
Глава 10. Мониторинг Kafka	243
Основы показателей.....	243
Как получить доступ к показателям.....	243
Внешние и внутренние показатели	244
Контроль состояния приложения	245
Охват показателей	245
Показатели брокеров Kafka	246
Недореплицированные разделы.....	246
Показатели брокеров	252
Показатели тем и разделов	261
Мониторинг JVM.....	263

Мониторинг ОС	265
Журналирование.....	266
Мониторинг клиентов	267
Показатели производителя	267
Показатели потребителей	271
Квоты.....	274
Мониторинг отставания.....	275
Сквозной мониторинг.....	276
Резюме.....	277
Глава 11. Поточковая обработка.....	278
Что такое потоковая обработка	279
Основные понятия потоковой обработки.....	282
Время.....	282
Состояние	283
Таблично-поточковый дуализм	284
Временные окна.....	286
Паттерны проектирования потоковой обработки	287
Обработка событий по отдельности.....	288
Обработка с использованием локального состояния.....	288
Многоэтапная обработка/повторное разделение на разделы	290
Обработка с применением внешнего справочника: соединение потока данных с таблицей.....	292
Соединение потоков	294
Внеочередные события	295
Повторная обработка.....	296
Kafka Streams в примерах	297
Подсчет количества слов.....	298
Сводные показатели фондовой биржи.....	301
Обогащение потока событий перехода по ссылкам.....	303
Kafka Streams: обзор архитектуры.....	305
Построение топологии	306
Масштабирование топологии.....	306
Как пережить отказ	310
Сценарии использования потоковой обработки.....	310
Как выбрать фреймворк потоковой обработки	312
Резюме.....	314

Приложение. Установка Kafka на других операционных системах.....315

Установка на Windows 315

 Использование Windows Subsystem для Linux..... 315

 Использование Java естественным образом..... 316

Установка на MacOS..... 318

 Использование Homebrew 319

 Установка вручную 319