

Оглавление

Предисловие.....	15
Введение	17
Благодарности.....	18
Об этой книге	19
Кому стоит прочитать эту книгу.....	19
Структура издания.....	19
О коде	21
Другие онлайн-ресурсы	21
Об авторе.....	22

ЧАСТЬ I. ЗНАКОМСТВО С KAFKA STREAMS

Глава 1. Добро пожаловать в Kafka Streams	24
1.1. Движение больших данных, и как оно повлияло на программирование	24
1.1.1. Возникновение больших данных	25
1.1.2. Важнейшие понятия парадигмы MapReduce	26
1.1.3. Одной пакетной обработки недостаточно	30
1.2. Знакомство с потоковой обработкой	31

1.3. Обработка транзакции покупки товара	32
1.3.1. Рассматриваем вариант с потоковой обработкой	32
1.3.2. Представление требований в виде графа.....	33
1.4. Транзакция покупки с другой точки зрения.....	34
1.4.1. Узел-источник.....	34
1.4.2. Узел маскирования номеров платежных карт.....	35
1.4.3. Узел паттернов	35
1.4.4. Узел поощрений.....	35
1.4.5. Узел хранения	37
1.5. Kafka Streams как граф узлов обработки	37
1.6. Использование Kafka Streams для потока данных транзакций покупок	39
1.6.1. Задаем источник.....	40
1.6.2. Первый узел-обработчик: маскирование номеров платежных карт	40
1.6.3. Второй узел-обработчик: паттерны покупок	41
1.6.4. Третий узел-обработчик: поощрение покупателей.....	43
1.6.5. Четвертый узел-обработчик: запись данных о покупках	44
Резюме	45
Глава 2. Kafka в двух словах.....	46
2.1. Проблема данных.....	46
2.2. Использование Kafka для обработки данных	47
2.2.1. Первоначальная платформа данных компании ZMart.....	47
2.2.2. Концентратор информации о транзакциях продаж на основе Kafka.....	48
2.3. Архитектура Kafka	50
2.3.1. Kafka — это брокер сообщений	50
2.3.2. Kafka — это журнал	51
2.3.3. Функционирование журналов в Kafka.....	52

2.3.4. Kafka и секции	53
2.3.5. Секции группируют данные по ключу	54
2.3.6. Написание пользовательского класса секционирования.....	55
2.3.7. Настройка пользовательского секционирования.....	56
2.3.8. Выбор правильного числа секций	57
2.3.9. Распределенный журнал.....	57
2.3.10. ZooKeeper: ведущие/ведомые брокеры и репликация.....	58
2.3.11. Apache ZooKeeper.....	59
2.3.12. Выборы контроллера	59
2.3.13. Репликация.....	60
2.3.14. Обязанности контроллера.....	61
2.3.15. Управление журналами.....	62
2.3.16. Удаление журналов.....	62
2.3.17. Сжатие журналов	64
2.4. Отправка сообщений с помощью генераторов.....	65
2.4.1. Свойства генераторов.....	68
2.4.2. Указание секции или метки даты/времени	69
2.4.3. Указание секции	69
2.4.4. Метки даты/времени в Kafka	70
2.5. Чтение сообщений с помощью потребителей	70
2.5.1. Управление смещениями	71
2.5.2. Автоматическая фиксация смещений	73
2.5.3. Фиксация смещения вручную.....	73
2.5.4. Создание потребителя	73
2.5.5. Потребители и секции	74
2.5.6. Перебалансировка	74
2.5.7. Более точное назначение топиков/секций потребителям	75
2.5.8. Пример потребителя.....	75

2.6. Установка и запуск Kafka.....	76
2.6.1. Локальные настройки Kafka	77
2.6.2. Запуск Kafka	77
2.6.3. Отправляем наше первое сообщение.....	79
Резюме	81

ЧАСТЬ II. РАЗРАБОТКА С ПОМОЩЬЮ KAFKA STREAMS

Глава 3. Разработка приложений Kafka Streams	84
3.1. API потоковых узлов-обработчиков	84
3.2. Программа Hello World для Kafka Streams	85
3.2.1. Создание топологии для Yelling.....	86
3.2.2. Настройка Kafka Streams.....	90
3.2.3. Создание объектов Serde	91
3.3. Работа с данными покупателей	93
3.3.1. Конструирование топологии	94
3.3.2. Создание пользовательского объекта Serde	101
3.4. Интерактивная разработка.....	103
3.5. Дальнейшие шаги	105
3.5.1. Новые требования	105
3.5.2. Сохранение записей вне Kafka	112
Резюме	114
Глава 4. Потоки данных и состояние	115
4.1. Обработка событий	116
4.2. Операции с сохранением состояния в Kafka Streams	117
4.2.1. Узел-обработчик transformValues	118
4.2.2. Поощрения покупателей с сохранением состояния	119
4.2.3. Инициализация преобразователя значений	121

4.2.4. Отображение объекта Purchase в объект RewardAccumulator на основе состояния	121
4.2.5. Обновление узла-обработчика поощрений.....	126
4.3. Использование хранилищ состояния для поиска и ранее просмотренные данные.....	128
4.3.1. Локальность данных	128
4.3.2. Восстановление после сбоя и отказоустойчивость	130
4.3.3. Использование хранилищ состояния в Kafka Streams	131
4.3.4. Другие поставщики хранилищ пар «ключ/значение»	132
4.3.5. Отказоустойчивость StateStore.....	132
4.3.6. Настройки топиков журналов изменений	132
4.4. Получение дополнительной информации путем соединения потоков данных.....	134
4.4.1. Подготовка данных	136
4.4.2. Генерация ключей с идентификаторами покупателей для соединения	137
4.4.3. Конструирование соединения	139
4.4.4. Другие варианты соединений	144
4.5. Метки даты/времени в Kafka Streams	146
4.5.1. Готовые реализации интерфейса TimestampExtractor	149
4.5.2. Класс WallclockTimestampExtractor.....	150
4.5.3. Пользовательская реализация интерфейса TimestampExtractor.....	150
4.5.4. Указываем, какой TimestampExtractor использовать.....	151
Резюме	152
Глава 5. API KTable	153
5.1. Взаимосвязь между потоками данных и таблицами	154
5.1.1. Поток записей.....	154
5.1.2. Обновления записей (журнал изменений).....	156

5.1.3. Поток событий по сравнению с потоком обновлений	158
5.2. Обновления записей и настройки KTable	160
5.2.1. Задание размера буфера кэша	161
5.2.2. Задание интервала фиксации	162
5.3. Агрегирование и оконные операции	163
5.3.1. Агрегирование объема продаж акций по отраслям промышленности	164
5.3.2. Оконные операции	169
5.3.3. Соединение объектов KStream и KTable	176
5.3.4. Объекты GlobalKTable	179
5.3.5. Доступное для запросов состояние	181
Резюме	182
Глава 6. API узлов-обработчиков	184
6.1. Компромисс между повышением уровня абстракции и расширением возможностей контроля	184
6.2. Создание топологии с использованием источников, узлов-обработчиков и стоков	185
6.2.1. Добавление узла-источника	186
6.2.2. Добавление узла-обработчика	187
6.2.3. Добавление узла-стока	190
6.3. Углубляемся в API узлов-обработчиков на примере узла биржевой аналитики	191
6.3.1. Узел-обработчик показателей акций	193
6.3.2. Метод process()	196
6.3.3. Выполнение пунктуатора	198
6.4. Узел совместной группировки	199
6.4.1. Создание узла совместной группировки	201
6.5. Интеграция API узлов-обработчиков и API Kafka Streams	211
Резюме	212

ЧАСТЬ III. АДМИНИСТРИРОВАНИЕ KAFKA STREAMS

Глава 7. Мониторинг и производительность	214
7.1. Основы мониторинга Kafka	214
7.1.1. Оценка производительности потребителей и генераторов	215
7.1.2. Проверка отставания потребителя	217
7.1.3. Перехват информации о поведении генераторов и потребителей	218
7.2. Метрики приложения	222
7.2.1. Настройки метрик	224
7.2.2. Как получить доступ к собранным метрикам	225
7.2.3. Использование JMX	225
7.2.4. Просмотр метрик	230
7.3. Дополнительные методики отладки Kafka Streams	231
7.3.1. Просмотр структуры приложения	231
7.3.2. Получение уведомлений о различных состояниях приложения	233
7.3.3. Использование интерфейса StateListener	234
7.3.4. Прослушиватель восстановления состояния	236
7.3.5. Обработчик неперехваченных исключений	239
Резюме	240
Глава 8. Тестирование приложения Kafka Streams	241
8.1. Тестирование топологии	242
8.1.1. Создание теста	245
8.1.2. Тестирование хранилища состояния в топологии	247
8.1.3. Тестирование узлов-обработчиков и преобразователей	248
8.2. Комплексное тестирование	251
8.2.1. Создание комплексного теста	252
Резюме	257

ЧАСТЬ IV. ПЕРЕДОВЫЕ ВОЗМОЖНОСТИ KAFKA STREAMS

Глава 9. Создание продвинутых приложений с помощью Kafka Streams	260
9.1. Интеграция Kafka с другими источниками данных	261
9.1.1. Интеграция данных с помощью Kafka Connect.....	262
9.1.2. Настройка Kafka Connect.....	263
9.1.3. Преобразование данных	265
9.2. Выбрасываем базу данных за борт	269
9.2.1. Как работают интерактивные запросы	272
9.2.2. Распределение хранилищ состояния	273
9.2.3. Настройка и обнаружение распределенного хранилища состояния	274
9.2.4. Написание кода для интерактивных запросов	276
9.2.5. Внутри сервера запросов	278
9.3. KSQL	282
9.3.1. Потоки и таблицы KSQL	283
9.3.2. Архитектура KSQL.....	283
9.3.3. Установка и запуск KSQL.....	285
9.3.4. Создание потока данных KSQL	286
9.3.5. Написание KSQL-запроса.....	288
9.3.6. Создание таблицы KSQL.....	288
9.3.7. Настройка KSQL	289
Резюме	290

ПРИЛОЖЕНИЯ

Приложение А. Дополнительная информация о настройках.....	292
А.1. Ограничение количества перебалансировок при запуске приложения	292
А.2. Устойчивость к отказам брокеров	293

A.3. Обработка ошибок десериализации	293
A.4. Масштабирование приложения	294
A.5. Конфигурация RocksDB	295
A.6. Заблаговременное создание топиков повторного секционирования	295
A.7. Настройка внутренних топиков	296
A.8. Перезапуск приложения Kafka Streams.....	297
A.9. Очистка локального состояния.....	298
Приложение Б. Строго однократная доставка	299