

Дилан: Я посвящаю эту работу Харпер, которой я так горжусь, и Ноэль, каждый день доставляющей радость нашей семье. Я также хотел бы посвятить эту книгу своим родителям, сестре и супруге, которые всегда оказывали мне всяческую поддержку.

Виктор: Эту работу я посвящаю своей супруге Марии за ее поддержку в процессе работы над этой книгой. Мне пришлось решать сложную задачу, выкраивая время то для того, то для другого. Без твоей поддержки у меня ничего бы не получилось. Я тебя люблю. Кроме того, хочу посвятить эту книгу (и выразить благодарность) моим детям, Эндрю и Майклу, за то, что они такие простодушные и прямолинейные. Когда люди спрашивают их, где работает папа, они отвечают: «Папа работает в Кафа».

Дейв: Я посвящаю эту книгу своей супруге Дебби и нашим детям Захарию, Эбигейл, Бенджамину, Сафе, Соломону, Ханне, Джоанне, Ребекке, Сюзанне, Ною, Самюэлю, Гидеону, Джошуа и Даниэлю. И наконец, все, что я делаю, я делаю во славу Творца и Спасителя нашего, Иисуса Христа.

Краткое оглавление

1	■	<i>Введение в Kafka</i>	26
2	■	<i>Знакомство с Kafka</i>	44
3	■	<i>Разработка проекта на основе Kafka</i>	75
4	■	<i>Производители: источники данных</i>	103
5	■	<i>Потребители: извлечение данных</i>	127
6	■	<i>Брокеры</i>	155
7	■	<i>Темы и разделы</i>	176
8	■	<i>Kafka как хранилище</i>	193
9	■	<i>Управление: инструменты и журналы</i>	211
10	■	<i>Защита Kafka</i>	236
11	■	<i>Реестр схем</i>	256
12	■	<i>Потоковая обработка с помощью Kafka Streams и ksqlDB</i>	270

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 *Реестр схем*

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

Предисловие от издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге, – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте www.dmkpress.com, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу dmkpress@gmail.com; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу http://dmkpress.com/authors/publish_book/ или напишите в издательство по адресу dmkpress@gmail.com.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг – возможно, ошибку в основном тексте или программном коде, – мы будем очень благодарны, если вы сообщите нам о ней. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Если вы найдете какие-либо ошибки в коде, пожалуйста, сообщите о них главному редактору по адресу dmkpress@gmail.com, и мы исправим это в следующих тиражах.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательство «ДМК Пресс» очень серьезно относится к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу dmkpress@gmail.com.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

Предисловие

Начиная с первого выпуска, вышедшего в 2011 году, технологии Apache Kafka® помогли создать новую категорию систем передачи данных, и теперь они являются основой бесчисленного множества современных приложений, управляемых событиями. В своей книге «Kafka в действии» Дилан Скотт (Dylan Scott), Виктор Гамов (Viktor Gamov) и Дейв Клейн (Dave Klein) делятся навыками проектирования и реализации приложений на основе событий, реализованных с использованием Apache Kafka. Авторы имеют богатый опыт работы с Kafka в реальном мире, что выделяет эту книгу среди других.

Давайте на минутку зададимся вопросом: «Зачем вообще нужна платформа Kafka?» Исторически сложилось так, что большинство приложений были основаны на системах хранения данных. Когда в мире происходили какие-то интересные события, они немедленно сохранялись в этих системах, но реакция на эти события происходила позже – либо когда пользователь явно запрашивал информацию, либо в ходе выполнения некоторых заданий пакетной обработки.

В системах передачи данных приложения строятся путем предварительного определения того, что они должны делать при появлении новых событий. Когда случаются новые события, приложения автоматически реагируют на них практически мгновенно. Такие приложения, управляемые событиями, привлекательны тем, что позволяют предприятиям гораздо быстрее извлекать новую информацию из своих данных. Однако переход к приложениям, управляемым событиями, требует изменения мышления, что не всегда легко. Эта книга предлагает исчерпывающее описание событийно-ориентированного мышления, а также реалистичные практические примеры, которые вы сможете опробовать.

«Kafka в действии» объясняет, как работает Kafka, и особое внимание уделяет созданию комплексных приложений, управляемых событиями, на основе Kafka. Здесь вы познакомитесь с компонентами, необходимыми для создания простого приложения Kafka, а также узнаете, как создавать сложные приложения с использованием таких библиотек, как Kafka Streams и `ksqlDB`. Также в этой книге рассказывается, как после создания приложения развернуть его в промышленном окружении, и освещаются такие ключевые темы, как мониторинг и безопасность.

Я надеюсь, что вам понравится эта книга так же, как мне. Удачной передачи событий!

– Юн Рао (Jun Rao), соучредитель Confluent

Вступление

Один из вопросов, который часто задают нам, когда мы рассказываем о работе над технической книгой: почему был выбран именно формат печатной книги? Дилан, например, всегда предпочитал узнавать что-то новое, читая книги. Другой фактор – ностальгия, навеваемая воспоминаниями о первой технической книге по программированию, которую он прочитал, «Elements of Programming with Perl» Эндрю Л. Джонсона (Andrew L. Johnson), выпущенной издательством Manning в 2000 году. Эта книга особенно запомнилась ему, и он до сих пор вспоминает, насколько приятно было читать ее страницы. Мы надеемся доставить такое же удовольствие своим читателям, описывая Apache Kafka.

Предвкушение познания чего-то нового почувствовал каждый из нас, когда мы впервые начали работать с Kafka. На наш взгляд, Kafka существенно отличается от любых других брокеров сообщений или шин служб предприятия (Enterprise Service Bus, ESB), которые нам доводилось использовать раньше. Быстрота разработки производителей и потребителей сообщений, возможность повторной обработки данных и скорость, с которой независимые потребители перемещаются без удаления данных из других потребительских приложений, позволяя решать проблемы, встречавшиеся в прошлом, и впечатлили нас больше всего, когда мы стали изучать возможность применения Kafka.

Мы видим, что Kafka меняет стандарты для платформ данных; она способна помочь перенести пакетные рабочие процессы и рабочие процессы извлечения, преобразования и загрузки (Extract, Transform, Load, ETL) ближе во времени к потокам данных. Поскольку эта платформа отходит от архитектур обработки данных, использовавшихся раньше и известных многим корпоративным пользователям, мы хотели помочь пользователям, не знакомым с Kafka, научиться работать с производителями и потребителями Kafka, а также решать базовые задачи разработки и администрирования Kafka. Мы надеемся, что к концу этой книги вы почувствуете в себе готовность углубиться в исследование более сложных тем Kafka, таких как мониторинг кластеров, создание метрик и межсайтовая репликация данных.

Всегда помните, что эта книга запечатлела момент, как Kafka выглядит сегодня. Она почти наверняка будет меняться и, надеюсь, станет еще лучше к тому времени, когда вы будете читать эту работу. Мы верим, что эта книга направит вас на увлекательный путь изучения основ Apache Kafka.

Благодарности

Дилан. Прежде всего я хотел бы поблагодарить мою семью. Спасибо вам! Я никогда не устану благодарить за поддержку и любовь, которые вижу каждый день. Я люблю вас всех! Дэн и Дебби, я высоко ценю, что вы всегда были моими самыми преданными сторонниками и фанатами. Сара, Харпер и Ноэль, я не смогу на словах передать всю любовь и гордость, которую испытываю ко всем вам, и благодарность за поддержку, которую вы мне оказываете. Спасибо семье DG, что всегда были рядом со мной. Спасибо и вам, JC.

Также отдельное спасибо Виктору Гамову и Дейву Клейну, что были соавторами этой книги! В продвижении этого проекта мне также помогали мои друзья и коллеги по работе, которых я должен упомянуть: команда Team Serenity (Бекки Кэмпбелл (Becky Campbell), Адам Доман (Adam Doman), Джейсон Фер (Jason Fehr) и Дэн Рассел (Dan Russell)), Роберт Абейта (Robert Abeyta) и Джереми Кастл (Jeremy Castle). Спасибо также, Джабулани Симплизио Чибайя (Jabulani Simplisio Chibaya), не только за рецензию, но и за добрые слова.

Виктор. Я хотел бы сказать огромное спасибо моей супруге и поблагодарить ее за поддержку. Спасибо также членам команды по связям с разработчиками и сообщества Confluent: Але Мюррей (Ale Murray), Еве Байзек (Yeva Byzek), Робину Моффатт (Robin Moffatt) и Тиму Берглунд (Tim Berglund). Вы все так много делаете для сообщества Apache Kafka!

Дейв. Я хотел бы поблагодарить Дилана и Виктора за то, что позволили мне поучаствовать в этом захватывающем путешествии.

Все вместе мы благодарим нашего редактора в Manning – Тони Арритола (Toni Arritola), чей опыт и наставничество помогли сделать эту книгу реальностью. Мы также выражаем благодарность Кристен Уоттерсон (Kristen Watterson), которая была нашим первым редактором до того, как на смену ей пришел Тони, нашим техническим редакторам Рафаэлю Вильеле (Raphael Villela), Ники Бакнер (Nickie Buckner), Фелипе Эстебану Вильдосо Кастильо (Felipe Esteban Vildoso Castillo), Маюру Патилу (Mayur Patil), Валентину Креттазу (Valentin Crettaz) и Уильяму Руденмальму (William Rudenmalm). Мы благодарим Чака Ларсона (Chuck Larson) за огромную помощь с графикой и Суманта Тамбе (Sumant Tambe) за техническую корректуру кода.

Сотрудники издательства Manning оказывали всемерную помощь по самым разным вопросам, от производства до продвижения, – это

очень продуктивная команда. Несмотря на все правки и обзоры, в текст книги и в исходный код могли просочиться опечатки и неточности (по крайней мере, мы никогда не видели книги без опечаток!), но эта команда, безусловно, помогла свести эти ошибки к минимуму.

Спасибо также Натану Марцу (Nathan Marz), Майклу Ноллу (Michael Noll), Джанакираму М. С. В. (Janakiram MSV), Биллу Беджеку (Bill Bejeck), Гуннару Морлингу (Gunnar Morling), Робину Моффатту (Robin Moffatt), Генри Каю (Henry Cai), Мартину Фаулеру (Martin Fowler), Александру Дину (Alexander Dean), Валентину Креттазу (Valentin Crettaz) и Ани Ли (Anyi Li). Вы очень помогли нам, рассказывая о своей работе и делаясь своими замечательными предложениями и отзывами.

Юн Рао, для нас большая честь, что вы нашли время написать предисловие к этой книге. Большое спасибо!

Мы очень признательны всему сообществу Apache Kafka (включая, конечно же, Джея Крепса (Jay Kreps), Неха Наркхеде (Neha Narkhede) и Юна Рао (Jun Rao)) и команде Confluent, продвигающей Kafka вперед и давшей разрешение на использование материалов, которые помогли наполнить эту книгу. По крайней мере, мы очень надеемся, что эта книга побудит разработчиков взглянуть на Kafka.

Наконец, большое спасибо всем рецензентам, это: Брайс Дарлинг (Bryce Darling), Кристофер Бейли (Christopher Bailey), Цицеро Зандона (Cicero Zandona), Конор Редмонд (Conor Redmond), Дэн Рассел (Dan Russell), Дэвид Криф (David Krief), Фелипе Эстебан Вилдосо Кастильо (Felipe Esteban Vildoso Castillo), Финн Ньюик (Finn Newick), Флорин-Габриэль Барбучану (Florin-Gabriel Barbuceanu), Грегор Райман (Gregor Rayman), Джейсон Фер (Jason Fehr), Хавьер Колладо Кабеза (Javier Collado Cabeza), Джон Мур (Jon Moore), Хорхе Эстебан Квилкате Отоя (Jorge Esteban Quilcate Otoya), Джошуа Хорвиц (Joshua Horwitz), Мадханмохан Савадамуту (Madhanmohan Savadamuthu), Мишель Мауро (Michele Mauro), Питер Перлепес (Peter Perlepes), Роман Левченко (Roman Levchenko), Санкет Найк (Sanket Naik), Шобха Айер (Shobha Iyer), Сумант Тамбе (Sumant Tambe), Витон Витанис (Viton Vitanis) и Уильям Руденмальм (William Rudenmalm) – ваши отзывы и предложения помогли сделать эту книгу лучше.

Мы могли кого-то упустить, и если это действительно так, то просим вас простить нас за нашу ошибку. Мы ценим вас.

Об этой книге

Мы писали «Kafka в действии» как практическое руководство по началу работы с Apache Kafka. В этой книге читатели встретят небольшие примеры, объясняющие некоторые параметры и настройки, которые можно использовать для изменения поведения Kafka в соответствии с конкретными вариантами применения. Ядро Kafka специально создавалось как настраиваемое в широких пределах и легко интегрирующееся с другими продуктами, такими как Kafka Streams и `ksqlDB`. Мы надеемся показать, как можно использовать платформу Kafka для удовлетворения различных бизнес-требований, чтобы вы освоились с ней к концу этой книги и знали, с чего начать решение ваших задач.

Кому адресована эта книга

«Kafka в действии» адресована разработчикам, желающим познакомиться с идеей потоковой обработки данных. От читателя не требуется обладать какими-либо знаниями о Kafka, но базовые знания командной строки и умение ею пользоваться не будут лишними. В Kafka есть несколько мощных инструментов командной строки, которые мы используем, и пользователь должен уметь по крайней мере вводить команды в командной строке.

Также могут пригодиться некоторые навыки программирования на языке Java и способность распознавать идеи программирования на любом языке. Эти навыки помогут понять представленные примеры кода, которые реализованы в основном в стиле Java 11 (а также Java 8). Кроме того, хотя это и необязательно, будет полезно общее понимание архитектуры распределенных приложений. Чем больше пользователь знает о репликациях и сбоях, тем проще ему будет понять, например, как Kafka использует реплики.

Организация книги

Эта книга состоит из трех частей, разбитых на 12 глав. Часть I представляет ментальную модель Kafka и рассказывает, где может пригодиться Kafka в реальном мире:

- глава 1 содержит общее введение в Kafka, опровергает некоторые мифы и описывает примеры использования в реальных условиях;
- глава 2 исследует архитектуру Kafka в общих чертах и вводит важную терминологию.

Часть II переходит к основным компонентам Kafka – клиентам и самому кластеру:

- глава 3 рассматривает характеристики проектов, в которых с успехом можно было бы использовать Kafka, и описывает некоторые подходы к разработке новых проектов. Здесь также обсуждается необходимость схем, на которые следует обратить особое внимание на этапе создания проекта на основе Kafka, но не позже;
- глава 4 иллюстрирует некоторые детали создания клиента-производителя и параметры управления передачей ваших данных в кластер Kafka;
- глава 5 меняет фокус главы 4 и иллюстрирует приемы получения данных из Kafka с помощью клиента-потребителя. Здесь будет представлена идея смещений и повторной обработки данных, обусловленная возможностью хранения сообщений;
- глава 6 рассматривает роль брокеров в кластере и как они взаимодействуют с вашими клиентами. Здесь мы исследуем различные компоненты, такие как контроллер и реплика;
- глава 7 обсуждает понятия тем и разделов, включая возможность компактификации тем и особенности хранения разделов;
- глава 8 описывает инструменты и архитектуры для обработки данных, которые может потребоваться сохранить или обработать повторно. Необходимость хранения данных в течение нескольких месяцев или лет может привести к тому, что вам придется подумать о вариантах хранения за пределами кластера;
- глава 9 завершает часть II обзором журналов, метрик и административных функций, помогающих поддерживать работоспособность кластера.

В части III мы перейдем от знакомства с основными компонентами частей Kafka к изучению вариантов улучшения работающего кластера:

- глава 10 представляет варианты усиления защиты кластера Kafka с помощью SSL, списков управления доступом ACL и квот;
- глава 11 посвящена реестру схем Schema Registry и особенностям его использования для работы с данными и сохранения совместимости с предыдущими и будущими версиями наборов данных. Считается, что эта возможность предназначена для приложений корпоративного уровня, однако она с успехом может использоваться для обслуживания любых данных, изменяющихся с течением времени;

- глава 12, последняя, посвящена знакомству с Kafka Streams и ksqlDB. Эти продукты относятся к более высоким уровням абстракции и основаны на ядре, которому посвящена вся вторая часть книги. Kafka Streams и ksqlDB – достаточно обширные темы, поэтому в нашем введении мы представим ровно столько информации, сколько действительно необходимо, чтобы приступить к самостоятельному изучению этих продуктов.

О примерах программного кода

Эта книга содержит множество примеров исходного кода как в пронумерованных листингах, так и в обычном тексте. В обоих случаях исходный код оформлен моноширинным шрифтом, чтобы визуально отделить его от обычного текста. Во многих случаях исходный код был дополнительно отформатирован; мы добавили разрывы строк и изменили отступы, чтобы уместить примеры по ширине книжной страницы. В некоторых случаях даже этого оказалось недостаточно, и в каких-то листингах вы можете встретить символы ➡, обозначающие продолжение строк. Многие листинги сопровождаются дополнительными комментариями, поясняющими важные понятия.

Наконец, следует отметить, что многие примеры кода не предназначены для выполнения в форме самостоятельных программ, – это выдержки, иллюстрирующие наиболее важные стороны обсуждаемого. Все примеры из книги и сопровождающий их исходный код в полной форме вы найдете на GitHub по адресу <https://github.com/Kafka-In-Action-Book/Kafka-In-Action-Source-Code> и на сайте издателя www.manning.com/books/kafka-in-action. Также выполняемые фрагменты кода можно получить в онлайн-версии этой книги по адресу <https://livebook.manning.com/book/kafka-in-action>.

Живое обсуждение книги

Приобретая книгу «Kafka в действии», вы получаете бесплатный доступ к частному веб-форуму, организованному издательством Manning Publications, где можно оставлять комментарии о книге, задавать технические вопросы, а также получать помощь от автора и других пользователей. Чтобы иметь доступ к форуму и зарегистрироваться на нем, откройте в веб-браузере страницу <https://livebook.manning.com/#!/book/kafka-in-action/discussion>. Узнать больше о форумах Manning и познакомиться с правилами поведения можно по адресу <https://livebook.manning.com/#!/discussion>.

Издательство Manning обязуется предоставить своим читателям место встречи, где может состояться содержательный диалог между отдельными читателями и между читателями и автором. Но со стороны авторов отсутствуют какие-либо обязательства уделять фо-

руму какое-то определенное внимание – их присутствие на форуме остается добровольным (и неоплачиваемым). Мы предлагаем задавать авторам стимулирующие вопросы, чтобы их интерес не угасал! Форум и архив с предыдущими обсуждениями остается доступным на сайте издательства, пока книга продолжает издаваться.

Другие онлайн-ресурсы

Все изменения, происходящие в Kafka с течением времени, неизменно отражаются перечисленными ниже ресурсами. В большинстве случаев на этих сайтах можно найти документацию с описанием прошлых версий:

- документация по Apache Kafka – <http://kafka.apache.org/documentation.html>;
- документация Confluent – <https://docs.confluent.io/current>;
- портал разработчиков Confluent – <https://developer.confluent.io>.

Об авторах

Дилан Скотт (Dylan Scott) – разработчик программного обеспечения с более чем десятилетним опытом программирования на Java и Perl. После знакомства с системой обмена сообщениями Kafka как средством передачи больших объемов данных Дилан начал погружаться в мир Kafka и технологий потоковой обработки. Он имеет значительный опыт использования таких технологий и очередей, как Mule, RabbitMQ, MQSeries и Kafka.

Дилан обладает различными сертификатами, свидетельствующими об опыте работы в отрасли: PMP, ITIL, CSM, Sun Java SE 1.6, Oracle Web EE 6, Neo4j и Jenkins Engineer.

Виктор Гамов (Viktor Gamov) – пропагандист передовых практик разработки в Confluent, компании, разрабатывающей платформу потоковой передачи событий на основе Apache Kafka. За свою долгую карьеру Виктор накопил богатый опыт в сфере разработки архитектур корпоративных приложений с использованием технологий с открытым исходным кодом. Ему нравится помогать архитекторам и разработчикам проектировать и создавать масштабируемые и высокодоступные распределенные системы с малым временем реакции.

Виктор является профессиональным докладчиком на конференциях по темам распределенных систем, потоковой передачи данных, JVM и DevOps, а также регулярно посещает мероприятия, такие как JavaOne, Devoxx, OSCON, QCon и др. Является соавтором книги «Enterprise Web Development» (O'Reilly Media, Inc.).

Следуйте за Виктором в Твиттере [@gamussa](#), где он пишет о спортивной жизни, вкусной и здоровой пище, открытом исходном коде и, конечно же, о Kafka!

Дейв Клейн (Dave Klein) имеет 28-летний опыт работы разработчиком, архитектором, руководителем проекта, автором, преподавателем, организатором конференций и семейного учителя, пока недавно не получил работу своей мечты пропагандиста передовых практик разработки в Confluent. Дейв восхищается удивительным миром потоковой передачи событий Apache Kafka и всеми силами старается помочь другим исследовать его.

Об иллюстрации на обложке

На обложке «Kafka в действии» изображена иллюстрация, подписанная как «Femme du Madagascar» (Мадагаскарская женщина), из четырехтомного сборника изображений национальной одежды Сильвена Марешаля (Sylvain Maréchal), изданного во Франции в XIX веке. Каждая иллюстрация тщательно прорисована и раскрашена вручную. Богатое разнообразие коллекции Марешаля напоминает нам о том, насколько далекими друг от друга были культурные традиции городов и регионов мира всего 200 лет назад. Изолированные друг от друга люди говорили на разных диалектах и языках. Встретив человека на улице, по его одежде было легко определить, где он живет, чем зарабатывает на жизнь или какое положение в обществе занимает.

С тех пор стиль одежды сильно изменился и исчезло разнообразие, характеризующее различные области и страны. В настоящее время трудно отличить по одежде даже жителей разных континентов, не говоря уже об обитателях разных городов, регионов или стран. Мы заменили культурное разнообразие более разносторонней личной жизнью и, безусловно, не менее интересной интеллектуальной жизнью.

Мы в издательстве Manning славим изобретательность, предприимчивость и радость компьютерного бизнеса обложками книг, изображающими богатство региональных различий двухвековой давности, оживших благодаря иллюстрациям Марешаля.

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru