

Оглавление

Предисловие	14
Введение	16
О книге	17
Для кого эта книга	17
Структура издания	17
О коде	18
Благодарности	19
От издательства	20
О научном редакторе русскоязычного издания	20
Об авторах	21
Иллюстрация на обложке	22

Часть I Начало работы

Глава 1. Введение в Apache Kafka	24
1.1. Что такое Apache Kafka и как он решает наши проблемы	25
1.2. Kafka в корпоративных экосистемах	26
1.3. Архитектура Kafka	29
1.4. Запуск и использование Kafka	31
1.5. Траектория обучения	32
Резюме	33
Глава 2. Первые шаги в работе с Kafka	34
2.1. Пример использования	34
2.2. Создание сообщений	35
2.3. Потребление сообщений	36
2.4. Параллельное потребление и генерация сообщений	37
2.5. Графические интерфейсы пользователя в Kafka	40
Резюме	42

Часть II

Основные концепции

Глава 3. Топики и сообщения в Kafka	44
3.1. Топики.....	44
3.1.1. Просмотр топиков.....	45
3.1.2. Создание, настройка и удаление топиков.....	48
3.2. Сообщения.....	51
3.2.1. Типы сообщений.....	52
3.2.2. Форматы данных.....	54
3.2.3. Структура сообщений.....	55
Резюме.....	57
Глава 4. Kafka как распределенный лог	59
4.1. Логи.....	59
4.1.1. Что такое лог.....	60
4.1.2. Основные свойства лога.....	61
4.1.3. Kafka как лог.....	62
4.2. Kafka как распределенная система.....	64
4.2.1. Партиционирование и ключи.....	65
4.2.2. Группы потребителей.....	70
4.2.3. Репликация.....	74
4.3. Компоненты Kafka.....	77
4.3.1. Координационный кластер.....	78
4.3.2. Брокер.....	79
4.3.3. Клиенты.....	79
4.4. Kafka в корпоративной среде.....	80
Резюме.....	82
Глава 5. Надежность	84
5.1. Подтверждения.....	85
5.1.1. Стратегии подтверждения в Kafka.....	86
5.1.2. Подтверждения и синхронизированные реплики.....	87
5.1.3. Гарантии доставки сообщений в Kafka.....	91
5.2. Транзакции.....	94
5.2.1. Транзакции в базе данных.....	95
5.2.2. Транзакции в Kafka.....	96
5.2.3. Транзакции и потребители.....	97
5.3. Репликация и принцип «лидер — фолловеры».....	99
Резюме.....	105

Глава 6. Производительность	107
6.1. Настройка топиков для повышения производительности	109
6.1.1. Масштабирование и балансировка нагрузки	109
6.1.2. Определение необходимого количества партиций	110
6.1.3. Изменение количества партиций	112
6.2. Производительность продюсера.....	115
6.2.1. Конфигурация продюсера.....	115
6.2.2. Тестирование производительности продюсера	118
6.3. Конфигурация и оптимизация брокеров.....	120
6.3.1. Оптимизация брокеров	121
6.3.2. Определение количества брокеров и их размера	122
6.4. Производительность потребителей.....	123
6.4.1. Конфигурация потребителей	123
6.4.2. Тестирование производительности потребителей	123
Резюме	125

Часть III Внутренние механизмы Kafka

Глава 7. Управление кластером	128
7.1. Управление кластером с помощью Apache Kafka Raft	129
7.2. Управление кластером с помощью ZooKeeper	131
7.3. Миграция с ZooKeeper на KRaft.....	133
7.4. Подключение к кластеру Kafka.....	134
Резюме	137
Глава 8. Создание и сохранение сообщений	139
8.1. Продюсер	140
8.1.1. Генерация сообщений	140
8.1.2. Процесс производства сообщений	142
8.1.3. Продюсер и подтверждения	142
8.2. Брокер.....	144
8.2.1. Прием и сохранение сообщений	144
8.2.2. Брокеры и подтверждения	145
8.3. Структуры данных и файлов	146
8.3.1. Метаданные, контрольные точки и топики.....	146
8.3.2. Каталог партиций.....	147
8.3.3. Данные логов и индексы.....	150
8.3.4. Сегменты	152
8.3.5. Удаленные топики	153
8.4. Репликация	154
8.4.1. Синхронизированные реплики	154

8.4.2. High Watermark	155
8.4.3. Влияние задержек при репликации.....	157
Резюме	158
Глава 9. Потребление сообщений	160
9.1. Получение сообщений	161
9.1.1. Запросы fetch	161
9.1.2. Получение данных из ближайшей реплики.....	162
9.2. Обработка fetch-запросов от потребителей брокером	162
9.3. Смещения и потребитель	164
9.3.1. Управление смещениями	164
9.3.2. Смещения в Kafka	165
9.4. Группы потребителей Kafka.....	168
9.4.1. Управление группами потребителей.....	168
9.4.2. Распределение партиций между потребителями.....	171
9.4.3. Статическое членство	174
Резюме	175
Глава 10. Очистка сообщений	177
10.1. Зачем удалять сообщения	177
10.2. Методы очистки в Kafka	178
10.3. Удаление логов по возрасту и размеру	179
10.3.1. Когда происходит очистка логов	180
10.3.2. Хранение смещений.....	182
10.4. Сжатие логов.....	182
10.4.1. Когда происходит очистка логов при использовании сжатия.....	184
10.4.2. Как работает очиститель логов	187
10.4.3. Tombstone	189
Резюме	189

Часть IV

Kafka в корпоративной среде

Глава 11. Интеграция внешних систем с помощью Kafka Connect	193
11.1. Что такое Kafka Connect	194
11.2. Кластер Kafka Connect: распределенный режим.....	196
11.2.1. Настройка кластера Kafka Connect	196
11.2.2. Создание коннектора	197
11.2.3. Тестирование коннектора	198
11.3. Масштабируемость и отказоустойчивость Kafka Connect	199
11.4. Конфигурация исполнителей	200

11.5. REST API Kafka Connect.....	203
11.5.1. Состояние кластера Kafka Connect.....	204
11.5.2. Создание, изменение и удаление коннекторов.....	206
11.6. Конфигурация коннектора.....	208
11.6.1. Общая конфигурация коннектора.....	208
11.6.2. Обработка ошибок в Kafka Connect.....	209
11.7. Преобразования отдельных сообщений.....	211
11.8. Пример Kafka Connect: коннектор-источник JDBC.....	213
11.8.1. Подготовка коннектора-источника JDBC.....	213
11.8.2. Настройка коннектора-источника JDBC.....	214
11.8.3. Источник-коннектор JDBC.....	215
11.9. Пример Kafka Connect: коннектор отслеживания изменений данных ...	217
11.9.1. Подготовка коннектора Debezium для PostgreSQL.....	217
11.9.2. Настройка коннектора Debezium для PostgreSQL.....	218
11.9.3. Тестирование коннектора Debezium для PostgreSQL.....	219
Резюме.....	221
Глава 12. Потокковая обработка.....	223
12.1. Обзор потоковой обработки.....	224
12.1.1. Библиотеки для потоковой обработки данных.....	225
12.1.2. Обработка данных.....	226
12.2. Потокковые обработчики.....	227
12.2.1. Типы обработчиков.....	228
12.2.2. Топологии обработчиков.....	231
12.3. Потокковая обработка с использованием SQL.....	232
12.4. Состояния потоков.....	234
12.4.1. Потоки и таблицы.....	235
12.4.2. Агрегации.....	237
12.4.3. Соединения потоков.....	238
12.4.4. Пример использования: уведомления.....	241
12.5. Потокковая обработка и время.....	244
12.5.1. Время относительно.....	244
12.5.2. Временные окна.....	245
12.5.3. Пример использования: обнаружение мошенничества.....	247
12.6. Масштабирование Kafka Streams.....	248
Резюме.....	251
Глава 13. Управление.....	254
13.1. Управление схемами.....	255
13.1.1. Зачем нужны схемы.....	256
13.1.2. Уровни совместимости.....	257

13.1.3. Реестры схем.....	261
13.1.4. Avro	263
13.2. Безопасность	264
13.2.1. Шифрование на транспортном уровне.....	265
13.2.2. Аутентификация	267
13.2.3. Авторизация.....	268
13.2.4. Шифрование при хранении.....	270
13.2.5. Сквозное шифрование	271
13.2.6. Безопасность в ZooKeeper.....	271
13.2.7. Защита незащищенного кластера Kafka	272
13.3. Квоты в Kafka: защита кластера от перегрузки.....	273
Резюме	277
Глава 14. Эталонная архитектура Kafka	279
14.1. Полезные компоненты и инструменты.....	280
14.1.1. kcat	281
14.1.2. Графические интерфейсы пользователя.....	282
14.1.3. Управление ресурсами Kafka.....	283
14.1.4. Cruise Control for Apache Kafka.....	283
14.2. Окружения развертывания	285
14.2.1. Kafka на собственном оборудовании компании.....	285
14.2.2. Kafka в виртуализированных окружениях	286
14.2.3. Kafka в Kubernetes: Strimzi	287
14.2.4. Запуск Kafka в публичном облаке	288
14.3. Требования к оборудованию	289
14.3.1. Брокеры	290
14.3.2. Координационный кластер.....	292
Резюме	292
Глава 15. Мониторинг и оповещения в Kafka.....	294
15.1. Метрики инфраструктуры	295
15.2. Метрики брокеров.....	296
15.2.1. Метрики сервера Kafka	296
15.2.2. Метрики лога Kafka.....	298
15.2.3. Метрики сети Kafka.....	298
15.2.4. Метрики контроллера Kafka	299
15.3. Метрики клиента.....	301
15.3.1. Общие метрики клиентов	301
15.3.2. Метрики продюсера	303
15.3.3. Метрики потребителей	304
15.3.4. Метрики Kafka Connect и Kafka Streams.....	307

12 Оглавление

15.4. Оповещения	309
15.4.1. Настройка оповещений на основе метрик	309
15.4.2. Решение проблем с помощью оповещений.....	310
15.5. Окружения развертывания Kafka и сложности их мониторинга.....	311
15.5.1. Kafka на собственном оборудовании компании.....	311
15.5.2. Kafka на виртуальных машинах.....	311
15.5.3. Kafka в публичном облаке.....	312
15.5.4. Kafka в Kubernetes	312
15.5.5. Kafka как управляемый сервис.....	313
15.5.6. Вопросы безопасности в различных средах.....	313
Резюме	313
Глава 16. Управление аварийными ситуациями.....	316
16.1. Источники и виды сбоев	317
16.1.1. Сбои в сети	317
16.1.2. Сбои вычислительных ресурсов.....	318
16.1.3. Сбои хранилища.....	320
16.1.4. Сбои в центре обработки данных.....	320
16.2. Резервное копирование в Kafka.....	323
16.3. Зеркалирование кластеров Kafka с помощью MirrorMaker	324
16.3.1. Активный/пассивный кластер	325
16.3.2. Активный/активный кластер	326
16.3.3. Топология «звезда»	327
Резюме	329
Глава 17. Сравнение Kafka с другими технологиями	331
17.1. Данные внутри и снаружи.....	332
17.2. Kafka и классические системы обмена сообщениями	333
17.2.1. Универсальность Kafka.....	334
17.2.2. Операционная сложность в классических системах обмена сообщениями.....	335
17.2.3. Управление классическими системами обмена сообщениями.....	336
17.3. REST и Kafka	338
17.3.1. Проблемы синхронного взаимодействия	338
17.3.2. Альтернативные стратегии взаимодействия	339
17.4. Реляционные базы данных и Kafka.....	340
17.4.1. Сильные и слабые стороны реляционных баз данных	340
17.4.2. Комбинация Kafka и реляционных баз данных в современных архитектурах.....	341
17.5. Kafka как ядро платформы потоковой обработки данных.....	343
Резюме	345

Глава 18. Корпоративные архитектуры	347
18.1. Kafka как основа Data Mesh.....	348
18.1.1. Сложности традиционного управления данными	348
18.1.2. Принципы Data Mesh.....	348
18.1.3. Data Mesh и традиционные подходы	349
18.1.4. Роль Kafka в Data Mesh	351
18.2. Освобождение данных из основных систем с помощью Kafka	353
18.3. Kafka для больших данных	355
18.4. Kafka для промышленного Интернета вещей.....	356
18.4.1. Применение Kafka в промышленном Интернете вещей	356
18.4.2. Хранение данных	357
18.4.3. Интеграция данных и управление доступом.....	358
18.4.4. Когда использовать несколько кластеров Kafka	359
18.5. С чем не стоит путать Kafka.....	360
18.5.1. Kafka — это не реляционная база данных.....	360
18.5.2. Kafka — это не синхронный интерфейс связи.....	361
18.5.3. Kafka — это не платформа для обмена файлами	362
18.5.4. Kafka для небольших приложений — под вопросом.....	363
18.5.5. Kafka не замена хорошей архитектуре.....	364
Резюме	365

Приложения

Приложение А. Настройка тестового окружения Kafka	368
А.1. Операционные системы.....	369
А.2. Скачивание Kafka	369
А.3. Конфигурация Kafka.....	370
А.4. Подготовка каталогов данных	370
А.5. Запуск Kafka	371
А.6. Остановка Kafka	372
Приложение Б. Настройка мониторинга.....	374
Б.1. Prometheus	374
Б.2. Экспортер для Prometheus.....	376
Б.3. Prometheus Alertmanager	379
Б.4. Grafana	380