

Spis treści

Przedmowa do wydania II	15
Przedmowa do wydania I	17
Wstęp	21
1. Poznaj Kafkę	25
Publikuj-Subskrybuj	25
Jak to się zaczyna?	26
Indywidualne systemy kolejkowe	27
Pojawienie się Kafki	28
Komunikaty i partie	28
Schematy	28
Tematy i partycje	29
Producenci i konsumenci	30
Brokery i klastry	31
Wiele klastrów	32
Dlaczego Kafka?	33
Wiele producentów	33
Wiele konsumentów	34
Retencja na dysku	34
Skalowalność	34
Wysoka wydajność	34
Funkcjonalności platformy	35
Ekosystem danych	35
Przypadki użycia	36
Pochodzenie Kafki	37
Problem LinkedIna	37
Narodziny Kafki	38
Open source	39
Zaangażowanie komercyjne	39
Nazwa	40
Pierwsze kroki z Kafką	40

2. Instalowanie Kafki 41

Konfiguracja środowiska 41

Wybór systemu operacyjnego 41

Instalowanie Javy 41

Instalowanie ZooKeepera 42

Instalowanie brokera Kafki 45

Konfiguracja brokera 46

Ogólne parametry brokera 46

Domyślne ustawienia tematu 49

Wybór sprzętu 54

Przepustowość dysku 54

Pojemność dysku 55

Pamięć 55

Konfiguracja sieciowa 55

Procesor 56

Kafka w chmurze 56

Microsoft Azure 56

Amazon Web Services 57

Konfiguracja klastrów Kafki 57

Ile brokerów? 57

Konfiguracja brokera 59

Dostosowywanie systemu operacyjnego 59

Zagadnienia produkcyjne 62

Opcje mechanizmu odzyskiwania pamięci 62

Układ centrum danych 64

Współlokowanie aplikacji na ZooKeeperze 64

Podsumowanie 66

3. Producenty Kafki — zapisywanie komunikatów w Kafce 67

Omówienie producenta 68

Konstruowanie producenta Kafki 69

Wysyłanie komunikatów do Kafki 71

Synchroniczne wysyłanie komunikatów 72

Asynchroniczne wysyłanie komunikatów 73

- Konfiguracja producentów 74
 - client.id 74
 - acks 74
 - Czas dostarczania komunikatów 75
 - linger.ms 77
 - buffer.memory 78
 - compression.type 78
 - batch.size 78
 - max.in.flight.requests.per.connection 78
 - max.request.size 79
 - receive.buffer.bytes i send.buffer.bytes 79
 - enable.idempotence 80
- Serializatory 80
 - Serializatory niestandardowe 80
 - Serializacja przy użyciu Apache Avro 82
 - Używanie rekordów Avro z Kafką 84
- Partycje 86
 - Wdrażanie niestandardowej strategii partycjonowania 88
- Nagłówki 89
- Przechwytywacze 89
- Kwoty i dławienie przepływności 91
- Podsumowanie 92
- 4. Konsumenty Kafki — odczytywanie danych z Kafki 94
 - Koncepcje konsumenckie Kafki 94
 - Konsumenty i grupy konsumentów 94
 - Grupy konsumentów i równoważenie partycji 97
 - Statyczne członkostwo w grupie 100
 - Tworzenie konsumenta Kafki 101
 - Subskrybowanie tematów 101
 - Pętla odpytowania 102
 - Bezpieczeństwo wątków 103
 - Konfiguracja konsumentów 104
 - fetch.min.bytes 104

fetch.max.wait.ms 105

fetch.max.bytes 105

max.poll.records 105

max.partition.fetch.bytes 105

session.timeout.ms i heartbeat.interval.ms 106

max.poll.interval.ms 106

default.api.timeout.ms 107

request.timeout.ms 107

auto.offset.reset 107

enable.auto.commit 107

partition.assignment.strategy 107

client.id 109

client.rack 109

group.instance.id 109

receive.buffer.bytes i send.buffer.bytes 109

offsets.retention.minutes 109

Zatwierdzenia i przesunięcia 110

Zatwierdzanie automatyczne 111

Zatwierdzanie bieżących przesunięć 112

Zatwierdzanie asynchroniczne 113

łączenie zatwierdzeń synchronicznych i asynchronicznych 114

Zatwierdzanie określonego przesunięcia 115

Nasłuchiwanie równoważenia obciążenia 116

Konsumowanie rekordów z określonymi przesunięciami 118

Wychodzenie z pętli 119

Deserializatory 121

Niestandardowe deserializatory 122

Stosowanie deserializacji Avro z konsumentem Kafki 124

Samodzielny konsument — dlaczego i jak korzystać z konsumenta bez grupy? 124

Podsumowanie 125

5. Programowe zarządzanie Kafką 127

Przegląd interfejsu AdminClient 128

Asynchroniczny i ostatecznie spójny interfejs API 128

Opcje 128

Hierarchia płaska 128

Dodatkowe uwagi 129

Cykl życia AdminClient API — tworzenie, konfiguracja i zamykanie 129

client.dns.lookup 130

request.timeout.ms 131

Podstawowe zarządzanie tematami 131

Zarządzanie konfiguracją 135

Zarządzanie grupą konsumentów 136

Eksploracja grup konsumentów 136

Modyfikowanie grup konsumentów 138

Metadane klastra 139

Zaawansowane operacje administracyjne 140

Dodawanie partycji do tematu 140

Usuwanie rekordów z tematu 141

Wybór lidera 141

Ponowne przypisywanie replik 142

Testowanie 143

Podsumowanie 145

6. Wewnętrzne mechanizmy działania Kafki 146

Przynależność do klastra 146

Kontroler 147

KRaft — nowy kontroler Kafki oparty na algorytmie Raft 148

Replikacja 150

Przetwarzanie żądań 152

Żądania produkcji 154

Żądania pobierania 155

Inne żądania 157

Fizyczna pamięć masowa 159

Warstwowy system pamięci masowej 159

Przydzielanie partycji 161

Zarządzanie plikami 162

Format plików 163

Indeksy 165

Kompaktowanie 165

Jak działa kompaktowanie? 166

Usunięte zdarzenia 167

Kiedy tematy są kompaktowane? 168

Podsumowanie 169

7. Niezawodne dostarczanie danych 170

Gwarancje niezawodności 170

Replikacja 171

Konfiguracja brokera 173

Współczynnik replikacji 173

Wybór nieczystego lidera 175

Minimalna liczba replik zsynchronizowanych 176

Utrzymywanie synchronizacji replik 176

Utrwalanie na dysku 177

Korzystanie z producentów w niezawodnym systemie 177

Wysyłanie potwierdzeń 178

Konfigurowanie prób ponawiania przez producenta 179

Dodatkowa obsługa błędów 180

Korzystanie z konsumentów w niezawodnym systemie 180

Ważne właściwości konfiguracji konsumenta
w celu niezawodnego przetwarzania 181

Bezpośrednie zatwierdzanie przesunięć w konsumentach 182

Walidacja niezawodności systemu 184

Walidacja konfiguracji 184

Walidacja aplikacji 185

Monitorowanie niezawodności w środowisku produkcyjnym 186

Podsumowanie 187

8. Semantyka „dokładnie raz” 189

Producent idempotentny 190

Jak działa idempotentny producent? 190

Ograniczenia producenta idempotentnego 192

Jak korzystać z producenta idempotentnego Kafki? 193

Transakcje 193

Przypadki użycia transakcji 194

Jakie problemy rozwiązują transakcje? 194

W jaki sposób transakcje gwarantują semantykę „dokładnie raz”? 195

Jakich problemów nie rozwiązują transakcje? 198

Jak korzystać z transakcji? 200

Identyfikatory transakcyjne i odgradzanie 203

Jak działają transakcje? 204

Wydajność transakcji 206

Podsumowanie 206

9. Budowanie potoków danych 208

Zagadnienia związane z budowaniem potoków danych 209

Terminowość 209

Niezawodność 209

Wysoka i zmienna przepustowość 210

Format danych 211

Transformacje 211

Bezpieczeństwo 212

Postępowanie w razie awarii 213

Powiązania i zwinność 213

Przypadki użycia dla frameworku Kafka Connect
oraz dla producentów i konsumentów 214

Kafka Connect 215

Uruchamianie frameworku Kafka Connect 216

Przykład konektora — źródło plików i ujście plików 218

Przykład konektora — z MySQL-a do Elasticsearcha 220

Pojedyncze transformacje komunikatów 226

Jak działa framework Kafka Connect? 228

Alternatywy dla frameworku Kafka Connect 231

Frameworki pobierania dla innych magazynów danych 231

Narzędzia ETL oparte na GUI 231

Frameworki przetwarzania strumieniowego 232

Podsumowanie 232

10. Mirroring danych między klastrami 234

Przypadki użycia dla mirroringu między klastrami 234

Architektury wieloklastrowe 236

Realia komunikacji między centrami danych 236

Architektura piasty i szprych 237

Architektura aktywny-aktywny 239

Architektura aktywny-pasywny 241

Klastry rozciągnięte 246

MirrorMaker platformy Apache Kafki 247

Konfigurowanie MirrorMakera 249

Topologia replikacji wieloklastrowej 251

Zabezpieczanie MirrorMakera 252

Wdrażanie MirrorMakera w środowisku produkcyjnym 253

Regulowanie MirrorMakera 256

Inne rozwiązania do mirroringu między klastrami 258

uReplicator Ubera 259

Brooklin LinkedIna 260

Rozwiązania do mirroringu między centrami danych firmy Confluent 260

Podsumowanie 262

11. Zabezpieczanie Kafki 263

Zamykanie dostępu do Kafki 263

Protokoły bezpieczeństwa 265

Uwierzytelnianie 267

SSL 267

SASL 272

Ponowne uwierzytelnianie 283

Aktualizowanie zabezpieczeń bez przestojów 284

Szyfrowanie 285

Szyfrowanie kompleksowe 286

Autoryzacja 287

AcIAuthorizer 288

Dostosowywanie autoryzacji 291

Kwestie bezpieczeństwa 293

Audyt 294

Zabezpieczanie ZooKeepera 294

SASL 295

SSL 295

Autoryzacja 296

Zabezpieczanie platformy 296

Zabezpieczanie hasłem 297

Podsumowanie 298

12. Administrowanie Kafką 300

Operacje na tematach 300

Tworzenie nowego tematu 301

Wyświetlanie listy wszystkich tematów w klastrze 302

Generowanie opisu tematu 302

Dodawanie partycji 304

Redukcja liczby partycji 305

Usuwanie tematu 305

Grupy konsumentów 306

Wyświetlanie listy i generowanie opisu grup 306

Usuwanie grup 307

Zarządzanie przesunięciami 307

Dynamiczne zmiany konfiguracji 309

Nadpisywanie wartości domyślnych konfiguracji tematu 309

Nadpisywanie wartości domyślnych konfiguracji klienta i użytkownika 311

Nadpisywanie wartości domyślnych konfiguracji brokera 312

Generowanie opisów nadpisanych konfiguracji 312

Usuwanie nadpisanych konfiguracji 313

Produkowanie i konsumowanie 313

Producent konsolowy 313

Konsument konsolowy 315

Zarządzanie partycjami 318

Wybór repliki preferowanej 318

Zmienianie replik partycji 320

Zrzucanie segmentów dziennika 324

Weryfikacja replik 326

Pozostałe narzędzia 326

Niebezpieczne operacje 327

Przenoszenie kontrolera klastra 327

Rozwiązywanie problemów z usuwaniem tematów 328

Ręczne usuwanie tematów 328

Podsumowanie 329

13. Monitorowanie Kafki

Podstawy monitorowania 330

Gdzie są wskaźniki? 330

Jakich wskaźników potrzebujesz? 332

Kontrolowanie kondycji aplikacji 333

SLO 333

Definicje dotyczące poziomu usług 334

Jakie wskaźniki nadają się na SLI? 335

Stosowanie SLO w ostrzeganiu 335

Wskaźniki brokerów Kafki 336

Diagnozowanie problemów z klastrami 337

Sztuka posługiwania się wskaźnikiem URP 338

Wskaźniki brokera 343

Wskaźniki tematów i partycji 351

Monitorowanie JVM 353

Monitorowanie systemu operacyjnego 354

Rejestrowanie 356

Monitorowanie klienta 357

Wskaźniki producenta 357

Wskaźniki konsumenta 359

Kwoty 362

Monitorowanie opóźnień 363

Monitorowanie kompleksowe 364

Podsumowanie 364

14. Przetwarzanie strumieniowe 366

Czym jest przetwarzanie strumieniowe? 367

Koncepcje przetwarzania strumieniowego 370

Topologia 370

Czas 370

Stan 372

Dualność strumieniowo-tablicowa 373

Okna czasowe 374

Gwarancje przetwarzania 376

Wzorce projektowe przetwarzania strumieniowego 376

Przetwarzanie pojedynczego zdarzenia 376

Przetwarzanie ze stanem lokalnym 377

Przetwarzanie wielofazowe (repartycjonowanie) 379

Przetwarzanie z wyszukiwaniem zewnętrznym

- łączenie strumieniowo-tablicowe 380

Łączenie tablicowo-tablicowe 381

Łączenie strumieniowe 382

Zdarzenia poza kolejnością 383

Ponowne przetwarzanie 384

Zapytania interaktywne 384

Kafka Streams na przykładach 385

Licznik słów 385

Statystyki giełdowe 388

Wzbogacanie strumienia kliknięć 390

Strumień Kafki — przegląd architektury 392

Budowanie topologii 392

Optymalizacja topologii 393

Testowanie topologii 393

Skalowanie topologii 394

Obsługa awarii 397

Przypadki użycia dla przetwarzania strumieniowego 398

Jak wybrać framework przetwarzania strumieniowego 399

Podsumowanie 401

A Instalowanie Kafki w innych systemach operacyjnych 403

Instalowanie w systemie Windows 403

Korzystanie z windowsowego podsystemu dla Linuksa 403

Korzystanie z natywnej Javy 404

Instalowanie w systemie macOS 406

Korzystanie z Homebrew 406

Instalacja ręczna 407

B Dodatkowe narzędzia Kafki 409

Kompleksowe platformy 409

Wdrażanie klastra i zarządzanie nim 410

Monitorowanie i eksploracja danych 411

Biblioteki klienckie 412

Przetwarzanie strumieniowe 412