

Spis treści

<b>O autorze</b>	<b>15</b>
<b>Przedmowa</b>	<b>17</b>
<b>Wprowadzenie</b>	<b>19</b>
Kto powinien przeczytać tę książkę?	20
Cel tej książki	21
Jak czytać tę książkę	22
Jak zorganizowana jest ta książka	22
Część I: „Wprowadzenie”	22
Część II: „Bazy klucz-wartość”	22
Część III: „Bazy dokumentów”	23
Część IV: „Bazy rodziny kolumn”	23
Część V: „Bazy grafowe”	23
Część VI: „Wybór bazy danych dla Twojej aplikacji”	24
„Dodatki”	24
<b>Część I Wprowadzenie</b>	<b>25</b>
<b>Rozdział 1. Różne bazy do różnych zastosowań</b>	<b>27</b>
Projekt bazy relacyjnej	28
Aplikacja e-commerce	28
Wczesne systemy zarządzania bazami danych	29
Systemy oparte na plikach płaskich	29
Systemy z hierarchicznym modelem danych	33
Systemy z sieciowym modelem danych	35
Podsumowanie wczesnych systemów baz danych	37
Rewolucja baz relacyjnych	38
Relacyjne systemy zarządzania danymi	39
Przyczyny powstania baz NoSQL	45
Skalowalność	46
Koszt	47
Elastyczność	47
Dostępność	48
Podsumowanie	48
Studium przypadku	50
Pytania kontrolne	50
Odniesienia	51
Bibliografia	51

<b>Rozdział 2. Różnorodność baz NoSQL</b>	<b>53</b>
Zarządzanie danymi w bazach rozproszonych	54
Przechowywanie danych w sposób trwały	55
Utrzymanie spójności danych	56
Zapewnienie dostępności danych	57
Zrównoważenie czasów reakcji, spójności i trwałości	60
Spójność, dostępność i partycjonowanie: teoria CAP	62
ACID i BASE	64
ACID: atomowość, spójność, izolacja, trwałość	64
BASE: zasadnicza dostępność, miękki stan, ostateczna spójność	65
Rodzaje ostatecznej spójności	66
Cztery typy baz NoSQL	68
Bazy par klucz-wartość	68
Bazy dokumentów	73
Bazy rodziny kolumn	75
Bazy grafowe	77
Podsumowanie	79
Pytania kontrolne	80
Odniesienia	81
Bibliografia	81
 <b>Część II Bazy klucz-wartość</b>	 <b>83</b>
 <b>Rozdział 3. Wprowadzenie do baz klucz-wartość</b>	 <b>85</b>
Od tablic do baz klucz-wartość	86
Tablice: baza klucz-wartość z kółkami treningowymi	86
Tablice asocjacyjne: zdjęcie kółek treningowych	87
Pamięć podręczna: dodanie do roweru biegów	88
Bazy klucz-wartość w pamięci i na dysku: od rowerów do pojazdów zmotoryzowanych	91
Podstawowe funkcjonalności baz klucz-wartość	92
Prostota: komu w ogóle potrzebne są skomplikowane modele danych?	93
Szybkość: nie ma czegoś takiego jak „za szybko”	94
Skalowalność: nadążaj za wyścigiem	95
Klucze: więcej niż tylko nic nieznaczące identyfikatory	101
Jak konstruować klucze	101
Używanie kluczy do odnajdywania wartości	102
Wartości: przechowywanie prawie dowolnych danych	106
Wartości nie wymagają silnego typowania	106
Ograniczenia w wyszukiwaniu wartości	107
Podsumowanie	108
Pytania kontrolne	109
Odniesienia	109
Bibliografia	110

<b>Rozdział 4. Terminologia baz klucz-wartość</b>	<b>111</b>
Terminy związane z modelowaniem danych w bazach klucz-wartość	112
Klucz	114
Wartość	116
Przestrzeń nazw	117
Partycja	118
Klucz partycjonowania	120
Brak schematu	121
Terminy związane z architekturą baz klucz-wartość	122
Klaster	122
Pierścień	124
Replikacja	124
Terminy związane z implementacją baz klucz-wartość	126
Funkcje haszujące	126
Kolizja	127
Kompresja	128
Podsumowanie	129
Pytania kontrolne	129
Odniesienia	130
 <b>Rozdział 5. Projektowanie baz klucz-wartość</b>	 <b>131</b>
Projektowanie kluczy i partycjonowanie	132
Klucze powinny posiadać konwencję nazewnictwa	132
Dobrze zaprojektowane klucze wymagają mniej kodu	133
Praca z zakresami wartości	134
Klucze muszą brać pod uwagę ograniczenia	135
Wykorzystanie kluczy do partycjonowania	136
Projektowanie struktury wartości	137
Typy posiadające strukturę pomagają skrócić czas oczekiwania	138
Duże wartości mogą prowadzić do mało wydajnych operacji odczytu i zapisu	140
Ograniczenia baz klucz-wartość	142
Pobieranie wartości wyłącznie za pomocą klucza	143
Bazy klucz-wartość nie wspierają przeszukiwania zakresów	144
Brak standardowego języka zapytań podobnego do SQL dla baz relacyjnych	144
Wzorce projektowe dla baz klucz-wartość	145
Klucze z ograniczonym czasem życia	145
Emulowanie tabel	147
Agregacje	148
Agregacje atomowe	150
Indeksy	151
Podsumowanie	153
Studium przypadku: bazy klucz-wartość dla konfiguracji aplikacji mobilnej	153
Pytania kontrolne	156

Odniesienia	156
<b>Część III Bazy dokumentów</b>	<b>157</b>
<b>Rozdział 6. Wprowadzenie do baz dokumentów</b>	<b>159</b>
Czym jest dokument?	160
Dokumenty nie są wcale takie proste	160
Dokumenty a pary klucz-wartość	163
Zarządzanie wieloma dokumentami w kolekcji	164
Unikaj jawnego definiowania schematu	172
Podstawowe operacje na bazach dokumentów	173
Wstawianie dokumentów do kolekcji	175
Usuwanie dokumentów z kolekcji	176
Aktualizacja dokumentów w kolekcji	177
Pobieranie dokumentów z kolekcji	178
Podsumowanie	180
Pytania kontrolne	180
Odniesienia	181
<b>Rozdział 7. Terminologia baz dokumentów</b>	<b>183</b>
Terminy dotyczące dokumentów i kolekcji	184
Dokument	184
Kolekcja	186
Dokument osadzony	187
Brak schematu	188
Schemat polimorficzny	190
Typy partycji	191
Partycjonowanie pionowe	192
Partycjonowanie poziome, czyli sharding	194
Modelowanie danych i przetwarzanie zapytań	198
Normalizacja	198
Denormalizacja	200
Procesor zapytań	200
Podsumowanie	201
Pytania kontrolne	201
Odniesienia	201
<b>Rozdział 8. Projektowanie baz dokumentów</b>	<b>203</b>
Normalizacja, denormalizacja i poszukiwanie równowagi	204
Relacja jeden-do-wielu	206
Relacja wiele-do-wielu	206
Potrzeba złączeń	206
Wykonywanie złączeń: podnoszenie ciężarów dla baz relacyjnych	208
Co zrobiłby projektant bazy dokumentów?	210
Planowanie z uwzględnieniem dokumentów zmiennych	215
Unikanie przenoszenia dużych dokumentów	218

Strefa Złotowłosej w indeksach	218
Aplikacje zorientowane na odczyt	218
Aplikacje zorientowane na zapis	219
Modelowanie powszechnych relacji	221
Relacja jeden-do-wielu w bazach dokumentów	221
Relacja wiele-do-wielu w bazach dokumentów	222
Modelowanie hierarchii w bazach dokumentów	223
Podsumowanie	225
Studium przypadku: manifesty użytkowników	226
Osadzać czy nie?	227
Wybór indeksów	228
Osobne kolekcje dla typów?	228
Pytania kontrolne	229
Odniesienia	229

## **Część IV Bazy rodziny kolumn** **231**

### **Rozdział 9. Wprowadzenie do baz rodziny kolumn** **233**

Na początku było Google BigTable	234
Wykorzystanie dynamicznej kontroli nad kolumnami	236
Indeksowanie po rekordzie, nazwie kolumny i stemplu czasowym	236
Kontrolowanie lokalizacji danych	237
Odczyt i zapis wierszy atomowych	237
Utrzymywanie posortowanych wierszy	238
Podobieństwa i różnice między bazami rodziny kolumn a bazami klucz-wartość i bazami dokumentów	240
Cechy baz rodziny kolumn	240
Podobieństwa i różnice między bazami rodziny kolumn i bazami dokumentów	241
Bazy rodziny kolumn kontra bazy relacyjne	242
Architektura baz rodziny kolumn	245
Architektura HBase: różnorodność węzłów	245
Architektura Cassandra: peer-to-peer	247
Rozgłaszanie: protokół plotki	248
Termodynamika i bazy rozproszone: po co nam antyentropia	250
Przechowaj to dla mnie: przekazanie ze wskazaniem	251
Kiedy korzystać z baz rodziny kolumn	252
Podsumowanie	254
Pytania kontrolne	254
Odniesienia	255

### **Rozdział 10. Terminologia baz rodziny kolumn** **257**

Podstawowe komponenty baz rodziny kolumn	258
Przestrzeń kluczy	258
Klucz wiersza	258
Kolumna	259

Rodziny kolumn	260
Struktury i procesy: implementacja baz rodziny kolumn	261
Wewnętrzne struktury i parametry konfiguracyjne baz rodziny kolumn	261
Starzy znajomi: klastry i partycje	262
Rzut oka pod maskę: inne komponenty baz rodziny kolumn	264
Procesy i protokoły	268
Replikacja	268
Antyentropia	268
Protokół plotki	269
Przekazanie ze wskazaniem	270
Podsumowanie	271
Pytania kontrolne	271
Odniesienia	272

## **Rozdział 11. Projektowanie baz rodziny kolumn** **273**

Wskazówki dotyczące projektowania tabel	275
Denormalizuj, zamiast łączyć	276
Wykorzystuj kolumny bez wartości	276
Używaj zarówno nazwy kolumny, jak i wartości kolumn	
do przechowywania danych	277
Modeluj encje za pomocą pojedynczego wiersza	278
Unikaj punktów zapalnych w kluczach wierszy	279
Utrzymuj odpowiednią liczbę wersji wartości kolumn	280
Unikaj rozbudowanych struktur danych w wartościach kolumn	281
Wskazówki dotyczące indeksowania	282
Kiedy korzystać z indeksów pomocniczych zarządzanych przez system	
bazy rodziny kolumn	282
Kiedy tworzyć indeksy pomocnicze i zarządzać nimi za pomocą tabeli	286
Narzędzia do pracy z bazami Big Data	288
Ekstrakcja, transformacja i ładowanie danych Big Data	289
Analizowanie danych Big Data	290
Narzędzia do monitorowania Big Data	293
Podsumowanie	294
Studium przypadku: analiza danych klienta	294
Zrozumienie potrzeb użytkownika	295
Pytania kontrolne	296
Odniesienia	297

## **Część V Bazy grafowe** **299**

### **Rozdział 12. Wprowadzenie do baz grafowych** **301**

Czym jest graf?	301
Modelowanie grafów i sieci	302
Modelowanie lokalizacji geograficznych	303
Modelowanie chorób zakaźnych	303
Modelowanie encji abstrakcyjnych i konkretnych	305

Modelowanie mediów społecznościowych	307
Zalety baz grafowych	308
Szybsze wykonywanie zapytań dzięki unikaniu złączeń	308
Upraszczenie modelowania	310
Wiele relacji pomiędzy encjami	310
Podsumowanie	311
Pytania kontrolne	311
Odniesienia	312

### **Rozdział 13. Terminologia baz grafowych** **313**

Elementy grafów	314
Wierzchołek	314
Krawędź	315
Ścieżka	317
Pętla	317
Operacje na grafach	318
Unia grafów	318
Przecięcie grafów	319
Przeszukiwanie grafu	320
Właściwości grafów i krawędzi	320
Izomorfizm	321
Rząd i rozmiar	321
Stopień	322
Bliskość	322
Pośrednictwo	322
Typy grafów	323
Grafy skierowane i nieskierowane	324
Sieć przepływowa	324
Grafy dwudzielne	325
Multigraf	325
Graf ważony	326
Podsumowanie	327
Pytania kontrolne	327
Odniesienia	327

### **Rozdział 14. Projektowanie baz grafowych** **329**

Początki projektowania grafów	329
Projektowanie bazy grafowej sieci społecznościowej	331
Projektowanie sterowane przez zapytania (znowu)	334
Odpytywanie grafu	336
Cypher: zapytania deklaratywne	336
Gremlin: zapytania przez trawersowanie grafu	337
Wskazówki i sztuczki przydatne w projektowaniu baz grafowych	341
Użyj indeksów do poprawienia czasu pobierania danych	342
Używaj krawędzi odpowiedniego rodzaju	342
Podczas przeszukiwania grafu uważaj na cykle	343

Weź pod uwagę skalowalność swojej bazy grafowej	344
Podsumowanie	345
Studium przypadku: optymalizacja tras transportowych	345
Zrozumieć potrzeby użytkownika	345
Projektowanie rozwiązania polegającego na analizie grafu	346
Pytania kontrolne	347
Odniesienia	347
<b>Część VI Wybór bazy danych dla Twojej aplikacji</b>	<b>349</b>
<b>Rozdział 15. Wytyczne do wyboru bazy danych</b>	<b>351</b>
Wybór bazy danych NoSQL	352
Przypadki użycia i kryteria wyboru baz klucz-wartość	353
Przypadki użycia i kryteria wyboru baz dokumentów	354
Przypadki użycia i kryteria wyboru baz rodziny kolumn	354
Przypadki użycia i kryteria wyboru baz grafowych	356
Używanie baz NoSQL i baz relacyjnych razem	357
Podsumowanie	358
Pytania kontrolne	358
Odniesienia	359
<b>Dodatki</b>	<b>361</b>
<b>Dodatek A Odpowiedzi do pytań kontrolnych</b>	<b>363</b>
<b>Dodatek B Lista baz NoSQL</b>	<b>389</b>
<b>Dodatek C Słowniczek</b>	<b>393</b>
<b>Skorowidz</b>	<b>401</b>