

Spis treści

O autorze (15)

Przedmowa (17)

Wprowadzenie (19)

- o Kto powinien przeczytać tę książkę? (20)
- o Cel tej książki (21)
- o Jak czytać tę książkę (22)
- o 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)

CZĘŚĆ I. WPROWADZENIE (25)

Rozdział 1. Różne bazy do różnych zastosowań (27)

- o Projekt bazy relacyjnej (28)
 - Aplikacja e-commerce (28)
- o 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)
- o Rewolucja baz relacyjnych (38)
 - Relacyjne systemy zarządzania danymi (39)
- o Przyczyny powstania baz NoSQL (45)
 - Skalowalność (46)
 - Koszt (47)
 - Elastyczność (47)
 - Dostępność (48)
- o Podsumowanie (48)
- o Studium przypadku (50)
- o Pytania kontrolne (50)
- o Odniesienia (51)
- o Bibliografia (51)

Rozdział 2. Różnorodność baz NoSQL (53)

- o 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)
- o ACID i BASE (64)
 - ACID: atomowość, spójność, izolacja, trwałość (64)
 - BASE: zasadnicza dostępność, miękki stan, ostateczna spójność (65)
 - Rodzaje ostatecznej spójności (66)
- o Cztery typy baz NoSQL (68)
 - Bazy par klucz-wartość (68)
 - Bazy dokumentów (73)
 - Bazy rodziny kolumn (75)
 - Bazy grafowe (77)
- o Podsumowanie (79)
- o Pytania kontrolne (80)
- o Odniesienia (81)
- o Bibliografia (81)

CZĘŚĆ II. BAZY KLUCZ-WARTOŚĆ (83)

Rozdział 3. Wprowadzenie do baz klucz-wartość (85)

- o 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)
- o 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)
- o Klucze: więcej niż tylko nic nieznaczające identyfikatory (101)
 - Jak konstruować klucze (101)
 - Używanie kluczy do odnajdywania wartości (102)
- o Wartości: przechowywanie prawie dowolnych danych (106)
 - Wartości nie wymagają silnego typowania (106)
 - Ograniczenia w wyszukiwaniu wartości (107)
- o Podsumowanie (108)
- o Pytania kontrolne (109)
- o Odniesienia (109)
- o Bibliografia (110)

Rozdział 4. Terminologia baz klucz-wartość (111)

- o 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)
- o Terminy związane z architekturą baz klucz-wartość (122)
 - Klaster (122)
 - Pierścień (124)
 - Replikacja (124)
- o Terminy związane z implementacją baz klucz-wartość (126)
 - Funkcje haszujące (126)
 - Kolizja (127)
 - Kompresja (128)
- o Podsumowanie (129)
- o Pytania kontrolne (129)
- o Odniesienia (130)

Rozdział 5. Projektowanie baz klucz-wartość (131)

- o 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)
- o Projektowanie struktury wartości (137)
 - Typy posiadające strukturę pomagają skrócić czas oczekiwania (138)
 - Duże wartości mogą prowadzić do mało wydajnych operacji odczytu i zapisu (140)
- o 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)
- o Wzorce projektowe dla baz klucz-wartość (145)
 - Klucze z ograniczonym czasem życia (145)
 - Emulowanie tabel (147)
 - Agregacje (148)
 - Agregacje atomowe (150)
 - Indeksy (151)
- o Podsumowanie (153)
- o Studium przypadku: bazy klucz-wartość dla konfiguracji aplikacji mobilnej (153)
- o Pytania kontrolne (156)
- o Odniesienia (156)

CZĘŚĆ III. BAZY DOKUMENTÓW (157)

Rozdział 6. Wprowadzenie do baz dokumentów (159)

- o Czym jest dokument? (160)
 - Dokumenty nie są wcale takie proste (160)
 - Dokumenty a pary klucz-wartość (163)
 - Zarządzanie wieloma dokumentami w kolekcji (164)

- o Unikaj jawnego definiowania schematu (172)
- o 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)
- o Podsumowanie (180)
- o Pytania kontrolne (180)
- o Odniesienia (181)

Rozdział 7. Terminologia baz dokumentów (183)

- o Terminy dotyczące dokumentów i kolekcji (184)
 - Dokument (184)
 - Kolekcja (186)
 - Dokument osadzony (187)
 - Brak schematu (188)
 - Schemat polimorficzny (190)
- o Typy partycji (191)
 - Partycjonowanie pionowe (192)
 - Partycjonowanie poziome, czyli sharding (194)
- o Modelowanie danych i przetwarzanie zapytań (198)
 - Normalizacja (198)
 - Denormalizacja (200)
 - Procesor zapytań (200)
- o Podsumowanie (201)
- o Pytania kontrolne (201)
- o Odniesienia (201)

Rozdział 8. Projektowanie baz dokumentów (203)

- o 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)
- o Planowanie z uwzględnieniem dokumentów zmiennych (215)
 - Unikanie przenoszenia dużych dokumentów (218)
- o Strefa Złotowłosej w indeksach (218)
 - Aplikacje zorientowane na odczyt (218)
 - Aplikacje zorientowane na zapis (219)
- o 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)
- o Podsumowanie (225)
- o Studium przypadku: manifesty użytkowników (226)
 - Osadzać czy nie? (227)

- Wybór indeksów (228)
- Osobne kolekcje dla typów? (228)
- o Pytania kontrolne (229)
- o Odniesienia (229)

CZĘŚĆ IV. BAZY RODZINY COLUMN (231)

Rozdział 9. Wprowadzenie do baz rodziny column (233)

- o 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)
- o Podobieństwa i różnice między bazami rodziny column a bazami klucz-wartość i bazami dokumentów (240)
 - Cechy baz rodziny column (240)
 - Podobieństwa i różnice między bazami rodziny column i bazami dokumentów (241)
 - Bazy rodziny column kontra bazy relacyjne (242)
- o Architektura baz rodziny column (245)
 - Architektura HBase: różnorodność węzłów (245)
 - Architektura Cassandra: peer-to-peer (247)
 - Rozgłaszanie: protokół plotki (248)
 - Termodynamika i bazy rozproszone: po co nam entropia (250)
 - Przechowaj to dla mnie: przekazanie ze wskazaniem (251)
- o Kiedy korzystać z baz rodziny column (252)
- o Podsumowanie (254)
- o Pytania kontrolne (254)
- o Odniesienia (255)

Rozdział 10. Terminologia baz rodziny column (257)

- o Podstawowe komponenty baz rodziny column (258)
 - Przestrzeń kluczy (258)
 - Klucz wiersza (258)
 - Kolumna (259)
 - Rodziny column (260)
- o Struktury i procesy: implementacja baz rodziny column (261)
 - Wewnętrzne struktury i parametry konfiguracyjne baz rodziny column (261)
 - Starzy znajomi: klastry i partycje (262)
 - Rzut oka pod maskę: inne komponenty baz rodziny column (264)
- o Procesy i protokoły (268)
 - Replikacja (268)
 - Entropia (268)
 - Protokół plotki (269)
 - Przekazanie ze wskazaniem (270)
- o Podsumowanie (271)

- o Pytania kontrolne (271)
- o Odniesienia (272)

Rozdział 11. Projektowanie baz rodziny kolumn (273)

- o 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)
- o 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)
- o 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)
- o Podsumowanie (294)
- o Studium przypadku: analiza danych klienta (294)
 - Zrozumienie potrzeb użytkownika (295)
- o Pytania kontrolne (296)
- o Odniesienia (297)

CZĘŚĆ V. BAZY GRAFOWE (299)

Rozdział 12. Wprowadzenie do baz grafowych (301)

- o Czym jest graf? (301)
- o Modelowanie grafów i sieci (302)
 - Modelowanie lokalizacji geograficznych (303)
 - Modelowanie chorób zakaźnych (303)
 - Modelowanie encji abstrakcyjnych i konkretnych (305)
 - Modelowanie mediów społecznościowych (307)
- o Zalety baz grafowych (308)
 - Szybsze wykonywanie zapytań dzięki unikaniu złączeń (308)
 - Upraszczanie modelowania (310)
 - Wiele relacji pomiędzy encjami (310)
- o Podsumowanie (311)
- o Pytania kontrolne (311)
- o Odniesienia (312)

Rozdział 13. Terminologia baz grafowych (313)

- o Elementy grafów (314)

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

Rozdział 14. Projektowanie baz grafowych (329)

- o Początki projektowania grafów (329)
 - Projektowanie bazy grafowej sieci społecznościowej (331)
 - Projektowanie sterowane przez zapytania (znowu) (334)
- o Odpytywanie grafu (336)
 - Cypher: zapytania deklaratywne (336)
 - Gremlin: zapytania przez trawersowanie grafu (337)
- o 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)
- o Podsumowanie (345)
- o Studium przypadku: optymalizacja tras transportowych (345)
 - Zrozumieć potrzeby użytkownika (345)
 - Projektowanie rozwiązania polegającego na analizie grafu (346)
- o Pytania kontrolne (347)
- o Odniesienia (347)

CZĘŚĆ VI. WYBÓR BAZY DANYCH DLA TWOJEJ APLIKACJI (349)

Rozdział 15. Wytyczne do wyboru bazy danych (351)

- o 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)
- o Używanie baz NoSQL i baz relacyjnych razem (357)
- o Podsumowanie (358)
- o Pytania kontrolne (358)
- o Odniesienia (359)

DODATKI (361)

Dodatek A. Odpowiedzi do pytań kontrolnych (363)

Dodatek B. Lista baz NoSQL (389)

Dodatek C. Słowniczek (393)

Skorowidz (401)