

Spis treści

O autorze	11
O recenzencie merytorycznym	13
Podziękowania	15

Część I	Bazy danych następnej generacji	17
----------------	--	-----------

Rozdział 1.	Trzy rewolucje związane z bazami danych	19
	Wczesne systemy baz danych	19
	Pierwsza rewolucja związana z bazami danych	21
	Druga rewolucja związana z bazami danych	23
	Teoria modelu relacyjnego	23
	Modele transakcji	25
	Pierwsze relacyjne bazy danych	25
	Wojny baz danych!	26
	Model przetwarzania klient-serwer	26
	Programowanie obiektowe i system OODBMS	27
	Okres stabilizacji relacyjnych baz danych	29
	Trzecia rewolucja związana z bazami danych	29
	Google i Hadoop	29
	Reszta witryn internetowych	30
	Chmura obliczeniowa	30
	Bazy danych dokumentów	31
	NewSQL	32
	Eksplzja nierelacyjnych baz danych	32
	Podsumowanie: jeden rozmiar nie pasuje wszystkim	33
	Źródła	34

Rozdział 2. Google, Big Data i Hadoop	35
Rewolucja związana z koncepcją Big Data	35
Chmura, urządzenia przenośne, serwisy społecznościowe i Big Data	36
Google: pionier koncepcji Big Data	37
Sprzęt używany przez firmę Google	38
Stos oprogramowania firmy Google	38
Dodatkowe informacje o modelu MapReduce	40
Hadoop: stos open source firmy Google	42
Początki technologii Hadoop	42
Siła technologii Hadoop	43
Architektura technologii Hadoop	43
HBase	46
Hive	46
Pig	49
Ekosystem Hadoop	49
Podsumowanie	50
Źródła	51
Rozdział 3. Sharding, Amazon i narodziny systemu NoSQL	53
Skalowanie standardu Web 2.0	53
Historia triumfu standardu Web 2.0	54
Rozwiązanie open source	55
Sharding	55
Fatalny koniec spowodowany przez tysiąc segmentów	57
Twierdzenie CAP	58
Spójność ostateczna	58
System Dynamo firmy Amazon	59
Mieszanie spójne	62
Spójność umożliwiająca dostosowanie	62
System Dynamo i rodzina magazynów klucz-wartość	63
Podsumowanie	65
Źródła	65
Rozdział 4. Bazy danych dokumentów	67
Format XML i bazy danych XML	68
Narzędzia i standardy związane z formatem XML	68
Bazy danych XML	69
Obsługa danych XML w systemach relacyjnych	69
Bazy danych dokumentów JSON	70
JSON i AJAX	71
Bazy danych dokumentów JSON	71
Modele danych w bazach danych dokumentów	73
Pierwsze bazy danych dokumentów JSON	74
MemBase i Couchbase	75
MongoDB	75
Format JSON, wszędzie format JSON	76
Podsumowanie	77

Rozdział 5. Tabele nie są przyjazne: grafowe bazy danych	79
Czym jest graf?	79
Wzorce systemu RDBMS związane z grafami	81
RDF i SPARQL	82
Grafy właściwości i Neo4j	83
Gremlin	84
Wewnętrzne szczegóły baz danych grafów	86
Silniki obliczeniowe grafów	87
Podsumowanie	88
Rozdział 6. Kolumnowe bazy danych	89
Schematy hurtowni danych	89
Alternatywa kolumnowa	91
Kompresja kolumnowa	91
Konsekwencje zapisu kolumnowego	93
Sybase IQ, C-Store i Vertica	94
Architektury kolumnowych baz danych	94
Projekcje	96
Technologia kolumnowa w innych bazach danych	97
Podsumowanie	98
Źródła	98
Rozdział 7. Koniec dysku? Pamięciowe bazy danych i bazy oparte na dyskach SSD	99
Koniec dysku?	99
Dysk SSD	100
Ekonomia dysku	101
Bazy danych oparte na dyskach SSD	102
Pamięciowe bazy danych	103
TimesTen	104
Redis	105
SAP HANA	107
VoltDB	109
„Pamięciowa” baza danych Oracle 12c	111
Stos Berkeley Analytics Data Stack i Spark	112
Architektura środowiska Spark	113
Podsumowanie	115
Źródła	115
Część II Ze wszystkimi szczegółami	117
Rozdział 8. Wzorce rozproszonych baz danych	119
Rozproszone relacyjne bazy danych	120
Replikacja	120
Współużytkowany dysk i brak współużytkowania	120
Nierelacyjne rozproszone bazy danych	124

Sharding i replikacja w bazie danych MongoDB	125
Sharding	125
Mechanizmy shardingu	125
Równoważenie klastra	128
Replikacja	128
Potwierdzenie zapisu i preferowanie odczytu	128
HBase	130
Tabele, regiony i serwery regionów	130
Buforowanie i lokalność danych	131
Uporządkowanie oparte na kluczach wierszy	132
Serwer regionów — podziały, równoważenie i awarie	133
Repliki regionów	134
Cassandra	134
Protokół „plotek”	134
Mieszanie spójne	135
„Wtyczki”	137
Podsumowanie	140
Rozdział 9. Modele spójności	141
Typy spójności	141
Modele ACID i MVCC	142
Globalne numery sekwencyjne transakcji	144
Potwierdzanie dwuetapowe	144
Inne poziomy spójności	144
Spójność w bazie danych MongoDB	145
Blokowanie w bazie danych MongoDB	145
Zestawy replik i spójność ostateczna	146
Spójność w bazie danych HBase	146
Repliki regionów o ewentualnej spójności	146
Spójność w bazie danych Cassandra	147
Współczynnik replikacji	148
Spójność zapisu	148
Spójność odczytu	149
Interakcja między poziomami spójności	150
Mechanizm hinted handoff i naprawianie odczytu	150
Znaczniki czasu i szczegółowość	151
Zegary wektorowe	152
Uprozczone transakcje	153
Podsumowanie	157
Rozdział 10. Modele danych i magazynowanie	159
Modele danych	159
Przegląd relacyjnego modelu danych	160
Magazyny klucz-wartość	160
Modele danych w bazach danych BigTable i HBase	164
Cassandra	166
Modele danych JSON	168

Magazynowanie	169
Typowy model magazynu relacyjnego	170
Drzewa LSM	172
Dodatkowe indeksowanie	175
Podsumowanie	178

Rozdział 11. Języki i interfejsy programowania 181

Język SQL	182
Interfejsy API baz danych NoSQL	183
Riak	183
HBase	185
MongoDB	187
Język CQL (Cassandra Query Language)	189
MapReduce	191
Pig	193
Grafy DAG	194
Cascading	195
Spark	195
Powrót języka SQL	196
Hive	197
Impala	198
Spark SQL	199
Couchbase N1QL	199
Apache Drill	201
Inne narzędzia SQL baz danych NoSQL	203
Podsumowanie	204
Źródła	204

Rozdział 12. Bazy danych przyszłości 205

Powrót do rewolucji	205
Kontrewolucje	206
Czy zatoczono pełne koło?	207
Problem z wyborem	208
Czy można mieć wszystko?	208
Modele spójności	209
Schemat	210
Języki baz danych	211
Przechowywanie	213
Wizja zbieżnej bazy danych	214
A tymczasem powróćmy do centrali firmy Oracle	215
Obsługa formatu JSON przez firmę Oracle	216
Uzyskiwanie dostępu do danych JSON za pośrednictwem interfejsu Oracle REST	218
Dostęp do tabel bazy danych Oracle przy użyciu interfejsu REST	218
Obsługa grafów w bazie danych Oracle	219
Sharding firmy Oracle	221
Oracle jako hybrydowa baza danych	222

Inne zbieżne bazy danych	223
Przełomowe technologie baz danych	224
Technologie magazynowania	224
Łańcuch bloków	225
Komputer kwantowy	226
Podsumowanie	227
Źródła	228
Dodatek A Przegląd baz danych	229
Aerospike	230
Cassandra	230
Couchbase	231
DynamoDB	232
HBase	232
MarkLogic	233
MongoDB	234
Neo4j	234
NuoDB	235
Oracle RDBMS	236
Redis	236
Riak	237
SAP HANA	238
TimesTen	238
Vertica	239
VoltDB	239
Skorowidz	241