

Spis treści

PRZEDMOWA	xvii
PODZIĘKOWANIA	xxi
WPROWADZENIE	xxiii
Czym jest SQL?	xxiv
Po co używać SQL?	xxiv
O książce	xxv
Korzystanie z przykładowych skryptów	xxvii
Korzystanie z PostgreSQL	xxviii
Instalacja PostgreSQL	xxix
Praca z pgAdmin	xxxii
Alternatywa dla pgAdmin	xxxv
Podsumowanie	xxxv
1	
TWORZENIE PIERWSZEJ BAZY DANYCH I TABELI	1
Tworzenie bazy danych	3
Wykonywanie SQL w pgAdmin	4
Połączenie z bazą danych analysis	5
Tworzenie tabeli	6
Polecenie CREATE TABLE	6
Tworzymy tabelę teachers	7
Wpisywanie wierszy do tabeli	8
Instrukcja INSERT	9
Przeglądanie danych	10
Gdy kod nie zadziała	10
Formatowanie SQL w celu zwiększenia czytelności	11
Podsumowanie	11
2	
EKSPLORACJA DANYCH ZA POMOCĄ INSTRUKCJI SELECT	13
Podstawowa składnia SELECT	14
Zapytania dla podzbioru kolumn	15
Użycie DISTINCT do znalezienia unikatowych wartości	16
Sortowanie danych przez ORDER BY	17
Filtrowanie wierszy za pomocą WHERE	20
Użycie LIKE i ILIKE z WHERE	22
Łączenie operatorów z użyciem AND i OR	23
Zastosowanie poznanych pojęć	24
Podsumowanie	25

3	
ZROZUMIEĆ TYPY DANYCH	27
Znaki	28
Liczby	30
Liczby całkowite	31
Autoinkrementacja liczb	32
Liczby niecałkowite	33
Wybór numerycznego typu danych	36
Daty i godziny	36
Użycie typu interval w obliczeniach	39
Typy dodatkowe	40
Konwersja wartości z jednego typu do innego z użyciem CAST	40
Skrócona notacja CAST	41
Podsumowanie	42
4	
IMPORT I EKSPORT DANYCH	43
Praca z rozdzielanymi plikami tekstowymi	44
Postępowanie z kolumnami zawierającymi separatory	45
Obsługa nagłówków	45
Użycie COPY do importu danych	46
Importowanie danych ze spisu ludności opisujących hrabstwa	48
Tworzenie tabeli us_counties_2010	49
Kolumny i typy danych w spisie ludności	50
Wykonanie importu danych ze spisu ludności za pomocą COPY	52
Import podzbioru kolumn za pomocą COPY	54
Wstawianie domyślnej wartości do kolumny podczas importu	56
Użycie COPY do eksportu danych	57
Eksport wszystkich danych	57
Eksport wybranych kolumn	58
Eksport wyników zapytania	58
Import i eksport przez pgAdmin	59
Podsumowanie	60
5	
PODSTAWY MATEMATYKI I STATYSTYKI W SQL	61
Operatory arytmetyczne	62
Matematyka i typy danych	62
Dodawanie, odejmowanie i mnożenie	63
Dzielenie i modulo	64
Potęgowanie, pierwiastkowanie i silnie	65
Kolejność działań	65
Matematyka na tabelach z danymi ze spisu ludności	66
Dodawanie i odejmowanie kolumn	67
Znajdowanie wartości procentowych z całości	69
Sledzenie zmian w wartości procentowej	70
Funkcje agregujące do liczenia średnich i sum	71
Wyznaczanie mediany	72

Znajdowanie mediany z funkcjami percentylowymi	73
Mediana i percentyle z danymi ze spisu ludności	74
Znajdowanie innych kwantyli za pomocą funkcji percentylowych	75
Tworzenie funkcji median()	76
Znajdowanie najczęściej występujących wartości	78
Podsumowanie	79

6

ŁĄCZENIE TABEL W RELACYJNEJ BAZIE DANYCH	81
Łączenie tabel z użyciem JOIN	82
Łączenie tabel z użyciem kolumn kluczy	82
Zapytania do wielu tabel z użyciem JOIN	86
Typy JOIN	87
JOIN	89
LEFT JOIN i RIGHT JOIN	89
FULL OUTER JOIN	91
CROSS JOIN	91
Użycie NULL do znalezienia wierszy z brakującymi wartościami	93
Trzy typy relacji pomiędzy tabelami	94
Relacja jeden-do-jednego	94
Relacja jeden-do-wielu	94
Relacja wiele-do-wielu	94
Wybór określonych kolumn w złączeniu	95
Upraszczanie składni JOIN z aliasami tabel	96
Złączanie wielu tabel	97
Obliczenia matematyczne na kolumnach łączonych tabel	99
Podsumowanie	101

7

DOBRE PRAKTYKI W PROJEKTOWANIU TABEL	103
Nazywanie tabel, kolumn i innych identyfikatorów	104
Używanie cudzysłowów z identyfikatorami	105
Pułapki z identyfikatorami	106
Wytyczne do nazewnictwa identyfikatorów	106
Sprawdzanie wartości kolumn za pomocą ograniczeń	107
Klucze główne: naturalny kontra sztuczny	108
Klucze obce	114
Automatyczne usuwanie rekordów z użyciem CASCADE	115
Ograniczenie CHECK	116
Ograniczenie UNIQUE	118
Ograniczenie NOT NULL	119
Usuwanie ograniczeń lub dodawanie ich później	119
Przyspieszanie zapytań za pomocą indeksów	120
B-drzewo: domyślny indeks PostgreSQL	121
Uwagi dotyczące korzystania z indeksów	124
Podsumowanie	125

8

POZYSKIWANIE INFORMACJI PRZEZ GRUPOWANIE

I SUMOWANIE

127

Tworzenie tabel ankiet bibliotecznych	128
Tworzenie tabeli dla danych z 2014 roku	128
Tworzenie tabeli dla danych z 2009 roku	130
Eksploracja danych o bibliotekach z użyciem funkcji agregujących	131
Zliczanie wierszy i wartości za pomocą count()	131
Wyszukiwanie wartości maksymalnych i minimalnych przy użyciu max() i min()	134
Agregowanie danych z użyciem GROUP BY	135
Podsumowanie	144

9

WERYFIKOWANIE I MODYFIKOWANIE DANYCH

147

Importowanie danych o producentach mięsa, drobiu i jaj	148
Wywiad z zestawem danych	149
Sprawdzanie brakujących wartości	151
Sprawdzanie niespójności w wartościach danych	153
Sprawdzanie zniekształconych wartości za pomocą length()	154
Modyfikacja tabel, kolumn i danych	155
Modyfikowanie tabel za pomocą ALTER TABLE	157
Modyfikowanie wartości za pomocą UPDATE	158
Tworzenie kopii zapasowych tabel	159
Przywracanie brakujących wartości danych w kolumnach	160
Aktualizowanie wartości w celu zachowania spójności	163
Naprawianie kodów pocztowych przy użyciu konkatenacji	164
Aktualizacja danych pomiędzy tabelami	166
Usuwanie niepotrzebnych danych	168
Usuwanie wierszy z tabeli	169
Usuwanie kolumny z tabeli	169
Usuwanie tabeli z bazy danych	170
Użycie bloków transakcji do zapisywania lub cofania zmian	170
Poprawianie wydajności podczas aktualizowania dużych tabel	173
Podsumowanie	174

10

FUNKCJE STATYSTYCZNE W SQL

177

Tworzenie tabeli z danymi statystycznymi ze spisu ludności	178
Wyznaczanie korelacji za pomocą corr(Y, X)	180
Sprawdzenie dodatkowych korelacji	182
Przewidywanie wartości za pomocą analizy regresji	184
Określanie wpływu zmiennej niezależnej za pomocą r-kwadrat	186
Tworzenie rankingów w SQL	187
Rankingi z użyciem rank() oraz dense_rank()	188
Ranking w podgrupach z użyciem PARTITION BY	189
Obliczanie współczynników dla porównań	191
Podsumowanie	193

11

PRACA Z DATAMI I GODZINAMI

195

Typy danych i funkcje dla dat i godzin	196
Manipulowanie datą i godziną	196
Wyodrębnianie komponentów z wartości timestamp	197
Tworzenie wartości datetime ze składników timestamp	199
Pobieranie aktualnej daty i godziny	200
Praca ze strefami czasowymi	201
Znajdowanie ustawień Twojej strefy czasowej	202
Ustawianie strefy czasowej	203
Obliczenia z datami i godzinami	205
Znajdowanie wzorców w danych o nowojorskich taksówkach	206
Znajdowanie wzorców w danych Amtrak	212
Podsumowanie	216

12

ZAAWANSOWANE TECHNIKI ZAPYTAŃ

219

Użycie podzapytań	220
Filtrowanie w klauzuli WHERE za pomocą podzapytań	221
Tworzenie tabel pochodnych za pomocą podzapytań	223
Złączanie tabel pochodnych	224
Generowanie kolumn z podzapytań	226
Wyrażenia z podzapytaniami	227
Wspólne wyrażenia tablicowe	229
Tabele krzyżowe	232
Instalowanie funkcji crosstab()	233
Tablicowanie wyników ankiety	233
Tablicowanie odczytów temperatury w mieście	235
Przeklasyfikowanie wartości za pomocą CASE	238
Użycie CASE we wspólnych wyrażeniach tablicowych	239
Podsumowanie	241

13

EKSPLORACJA TEKSTÓW W CELU WYSZUKIWANIA ISTOTNYCH DANYCH

243

Formatowanie za pomocą funkcji tekstowych	244
Formatowanie wielkości znaków	244
Informacje o łańcuchach znaków	245
Usuwanie znaków	245
Pobieranie i zmiana znaków	246
Dopasowywanie wzorców tekstowych za pomocą wyrażeń regularnych	246
Notacja wyrażeń regularnych	247
Przekształcanie tekstu w dane za pomocą funkcji z wyrażeniami regularnymi	249
Używanie wyrażeń regularnych z WHERE	263
Dodatkowe funkcje dotyczące wyrażeń regularnych	264
Wyszukiwanie pełnotekstowe w PostgreSQL	266
Typy danych wyszukiwania tekstowego	266
Tworzenie tabeli do wyszukiwania pełnotekstowego	268

Przeszukiwanie tekstów przemówień	269
Klasyfikowanie dopasowań zapytań według trafności	273
Podsumowanie	275

14

ANALIZA DANYCH PRZESTRZENNYCH ZA POMOCĄ POSTGIS 277

Instalowanie PostGIS i tworzenie przestrzennej bazy danych	278
Budowanie bloków danych przestrzennych	279
Geometrie dwuwymiarowe	280
Format Well-Known Text	281
Uwaga dotycząca układów współrzędnych	282
Spatial Referencing System Identifier	282
Typy danych PostGIS	283
Tworzenie obiektów przestrzennych za pomocą funkcji PostGIS	284
Tworzenie typu geometry z użyciem Well-Known Text	284
Tworzenie typu geography z Well-Known Text	285
Funkcje dla Point	286
Funkcje dla LineString	286
Funkcje dla Polygon	287
Analizowanie danych rynku rolniczego	288
Tworzenie i wypełnianie kolumny geography	289
Dodanie indeksu GiST	289
Wyszukiwanie obiektów geograficznych w obrębie podanej odległości	290
Wyznaczanie odległości pomiędzy obiektami geograficznymi	292
Praca z plikami shape ze spisu ludności	294
Zawartość pliku shape	294
Ładowanie plików shape przez GUI	295
Eksplorowanie pliku shape ze spisu ludności 2010	297
Wykonywanie połączeń przestrzennych	300
Odkrywanie danych o drogach i szlakach wodnych	300
Złączanie tabel dróg i szlaków wodnych ze spisu ludności	301
Znajdowanie lokalizacji, w której przecinają się obiekty	303
Podsumowanie	304

15

OSZCZĘDZANIE CZASU DZIĘKI WIDOKOM, FUNKCJOM I WYZWALACZOM 305

Użycie widoków do uproszczenia zapytań	306
Tworzenie i odpytywanie widoków	307
Wstawianie, aktualizowanie i usuwanie danych przy użyciu widoku	310
Programowanie własnych funkcji	314
Tworzenie funkcji percent_change()	315
Użycie funkcji percent_change()	316
Aktualizowanie danych za pomocą funkcji	318
Używanie języka Python w funkcji	321
Automatyzowanie działań w bazie danych za pomocą wyzwalaczy	322
Rejestrowanie zmian ocen w tabeli	323
Automatyczna klasyfikacja temperatury	327
Podsumowanie	330

16

UŻYWANIE POSTGRESQL Z WIERSZA POLECEŃ 331

Konfigurowanie wiersza poleceń dla psql	332
Konfiguracja psql w Windows	332
Konfiguracja psql w macOS	336
Ustawienia psql w systemie Linux	339
Praca z psql	340
Uruchamianie psql i łączenie z bazą danych	340
Uzyskiwanie pomocy	341
Zmiana użytkownika i połączenia z bazą danych	341
Uruchamianie zapytań SQL w psql	342
Nawigacja i formatowanie wyników	344
Metapolecenia do uzyskiwania informacji o bazie danych	348
Importowanie, eksportowanie i używanie plików	349
Dodatkowe narzędzia wiersza poleceń do przyspieszania pracy	352
Dodawanie bazy danych za pomocą createdb	353
Ładowanie plików shape z użyciem shp2psql	353
Podsumowanie	354

17

UTRZYMANIE NASZEJ BAZY DANYCH 357

Odzyskiwanie nieużywanego miejsca za pomocą VACUUM	358
Siedzenie wielkości tabeli	358
Monitorowanie procesu autovacuum	361
Ręczne wykonywanie VACUUM	362
Zmniejszanie rozmiaru tabeli za pomocą VACUUM FULL	363
Zmiana ustawień serwera	363
Lokalizacja i edycja postgresql.conf	364
Ponowne załadowanie ustawień za pomocą pg_ctl	366
Kopia zapasowa bazy danych i jej przywracanie	367
Wykorzystanie pg_dump do tworzenia kopii zapasowej bazy danych lub tabeli	367
Przywracanie bazy danych z kopii zapasowej za pomocą pg_restore	368
Dodatkowe opcje kopii zapasowej i przywracania	368
Podsumowanie	369

18

IDENTYFIKACJA I OPOWIEŚĆ KRYJĄCA SIĘ ZA NASZYM DANYMI 371

Zaczynamy od pytania	372
Dokumentacja naszych działań	372
Zbieranie danych	373
Brak danych? Utwórz własną bazę danych	373
Dostęp do miejsca pochodzenia danych	374
Wywiad na temat danych za pomocą zapytań	375
Skonsultuj się z właścicielem danych	375
Identyfikacja kluczowych wskaźników i trendów czasowych	376
Zapytaj dlaczego	378

Przekaż swoje ustalenia	378
Podsumowanie	379

ANEKS

DODATKOWE ZASOBY POSTGRESQL 381

Środowiska programistyczne PostgreSQL	381
---------------------------------------	-----

Narzędzia, dodatki i rozszerzenia PostgreSQL	382
--	-----

Aktualności PostgreSQL	383
------------------------	-----

Dokumentacja	384
--------------	-----

INDEKS 385

oprac. BPK