

Korporacyjne jezioro danych : wykorzystaj potencjał big data w swojej organizacji / Alex Gorelik. – Gliwice, © 2020

Spis treści

Wstęp	9
1. Wprowadzenie do jezior danych	13
Dojrzewanie jeziora danych	15
Kałuże danych	17
Stawy danych	17
Udane tworzenie jeziora danych	18
Właściwa platforma	18
Właściwe dane	19
Właściwy interfejs	20
Bagno danych	22
Wskazówki dotyczące sukcesu w budowaniu jezior danych	23
Tworzenie jeziora danych	24
Organizowanie jeziora danych	24
Konfiguracja jeziora danych pod kątem samoobsługi	26
Architektury jeziora danych	30
Jeziora danych w chmurze publicznej	31
Logiczne jeziora danych	31
Podsumowanie	34
2. Perspektywa historyczna	37
Dysk do danych samoobsługowych — narodziny baz danych	37
Imperatyw analityczny — narodziny hurtowni danych	40
Ekosystem hurtowni danych	41
Przechowywanie i kwerendowanie danych	42
Ładowanie danych — narzędzia do integracji danych	47
Organizowanie danych i zarządzanie nimi	50
Konsumowanie danych	55
Podsumowanie	56
3. Wprowadzenie do big data i nauki o danych	57
Hadoop przewodzi historycznemu przejściu na big data	57
System plików Hadoop	58
Współdziałanie przetwarzania i przechowywania w zadaniu MapReduce	59
Schemat odczytu	60
Projekty Hadoop	61
Nauka o danych	62
Uczenie maszynowe	66

Zdolność wyjaśnienia	67
Zarządzanie zmianami	68
Podsumowanie	69
4. Budowanie jeziora danych	71
Co to jest Hadoop i dlaczego z niego korzystamy?	71
Zapobieganie rozprzestrzenianiu się kałuż danych	74
Wykorzystanie big data	74
Nauka o danych jako główny czynnik	75
Strategia 1. — przeniesienie istniejącej funkcjonalności	77
Strategia 2. — jeziora danych dla nowych projektów	79
Strategia 3. — ustanowienie centralnego punktu zarządzania	79
Który sposób jest odpowiedni dla Ciebie?	80
Podsumowanie	82
5. Od stawów danych, czyli hurtowni danych big data, do jezior danych	83
Podstawowe funkcje hurtowni danych	84
Modelowanie wymiarowe dla analityki	85
Integrowanie danych z różnych źródeł	86
Zachowywanie historii za pomocą powoli zmieniających się wymiarów	86
Ograniczenia hurtowni danych jako repozytorium historycznego	86
Przejście do stawu danych	87
Utrzymywanie historii w stawie danych	87
Wdrażanie powoli zmieniających się wymiarów w stawie danych	88
Rozrastanie się stawów danych w jeziora danych — ładowanie danych, które nie znajdują się w hurtowni danych	90
Surowe dane	91
Dane zewnętrzne	91
Internet rzeczy (IoT) i inne dane strumieniowe	94
Architektura Lambda	94
Transformacje danych	97
Systemy docelowe	99
Hurtownie danych	100
Operacyjne magazyny danych	100
Aplikacje czasu rzeczywistego i produkty oparte na danych	100
Podsumowanie	101
6. Optymalizacja pod kątem samoobsługi	103
Początki samoobsługi	103
Analitycy biznesowi	105
Znajdowanie i zrozumienie danych — dokumentowanie przedsiębiorstwa	106
Budowanie zaufania	109
Dostarczanie	115

Przygotowanie danych do analizy	116
Przygotowywanie danych w jeziorze danych	117
Umiejszczenie przygotowywania danych w Hadoop	118
Powszechne przypadki użycia dla przygotowywania danych	119
Analiza i wizualizacja	120
Podsumowanie	123
7. Architektura jeziora danych	125
Organizacja jeziora danych	125
Strefa ładowania lub surowa	126
Strefa złota	127
Strefa robocza	129
Strefa wrażliwa	129
Wiele jezior danych	131
Zalety utrzymywania osobnych jezior danych	131
Zalety scalania jezior danych	131
Jeziora danych w chmurze	132
Wirtualne jeziora danych	135
Federacja danych	135
Wirtualizacja big data	136
Eliminacja redundancji	137
Podsumowanie	139
8. Katalogowanie jeziora danych	141
Organizowanie danych	141
Metadane techniczne	142
Metadane biznesowe	146
Znakowanie	148
Automatyczne katalogowanie	149
Logiczne zarządzanie danymi	150
Zarządzanie wrażliwymi danymi i kontrola dostępu	150
Jakość danych	152
Powiązanie różnych danych	154
Ustanawianie pochodzenia	155
Dostarczanie danych	156
Narzędzia służące do budowania katalogu	157
Porównanie narzędzi	158
Ocean danych	159
Podsumowanie	159
9. Zarządzanie dostępem do danych	161
Autoryzacja lub kontrola dostępu	162
Zasady dostępu do danych oparte na znacznikach	163
Anonimizacja wrażliwych danych	166
Suwerenność danych i zgodność z przepisami	169

Samoobsługowe zarządzanie dostępem	171
Dostarczanie danych	174
Podsumowanie	180
10. Perspektywy dla różnych branż	181
Big data w usługach finansowych	182
Konsumenci, cyfryzacja i dane zmieniają znane nam finanse	182
Ratowanie banku	183
Nowe możliwości oferowane przez nowe dane	186
Kluczowe procesy korzystania z jeziora danych	188
Wartość dodana przez jeziora danych w usługach finansowych	190
Jeziora danych w branży ubezpieczeniowej	191
Inteligentne miasta	193
Big data w medycynie	194
Skorowidz	196

oprac. BPK