

Słowo wstępne	17
---------------------	----

Przedmowa	19
-----------------	----

Część I. Fundamenty	23
----------------------------	-----------

1. Technologia Big Data	25
--------------------------------------	-----------

Czym jest technologia Big Data i jak może być pomocna?	26
--	----

Dojrzałość danych	29
-------------------	----

Etap 1: reakcyjny	30
-------------------	----

Etap 2: informacyjny	31
----------------------	----

Etap 3: predyktywny	31
---------------------	----

Etap 4: transformatywny	31
-------------------------	----

Samoobsługowa analityka biznesowa	32
-----------------------------------	----

Podsumowanie	33
--------------	----

2. Typy architektur danych	35
---	-----------

Ewolucja architektur danych	36
-----------------------------	----

Relacyjna hurtownia danych	38
----------------------------	----

Jezioro danych	40
----------------	----

Nowoczesna hurtownia danych	42
-----------------------------	----

Architektura Data Fabric	43
--------------------------	----

Architektura Data Lakehouse	43
-----------------------------	----

Siatka danych	44
---------------	----

Podsumowanie	45
--------------	----

3. Sesja projektowania architektury	46
--	-----------

Czym jest sesja projektowania architektury?	46
---	----

Dlaczego należy przeprowadzać sesję ADS?	46
--	----

Przed sesją ADS	47
Przygotowanie	47
Zaproszenie uczestników	50
Przeprowadzanie sesji ADS	51
Wprowadzenia	52
Ustalenia	52
Korzystanie z tablicy	57
Po zakończeniu sesji ADS	58
Wskazówki dotyczące prowadzenia sesji ADS	59
Podsumowanie	61

Część II. Typowe pojęcia związane z architekturami danych **63**

4. Relacyjna hurtownia danych	65
Czym jest relacyjna hurtownia danych?	65
Czym hurtownia danych nie jest?	67
Podejście odgórne	69
Dlaczego warto skorzystać z relacyjnej hurtowni danych?	71
Wady związane z korzystaniem z relacyjnej hurtowni danych	74
Zasilanie hurtowni danych	75
Częstotliwość wyodrębniania danych	75
Metody wyodrębniania	76
Metoda określania, jakie dane zmieniły się od ostatniej operacji wyodrębnienia	76
Informacje o kresie relacyjnej hurtowni danych okazały się mocno przesadzone	78
Podsumowanie	79
5. Jezioro danych	80
Czym jest jezioro danych?	80
Dlaczego warto używać jeziora danych?	81
Podejście oddolne	83
Najlepsze praktyki projektowania architektury jeziora danych	84
Wiele jezior danych	90
Zalety	90
Wady	93
Podsumowanie	94
6. Procesy i rozwiązania z zakresu magazynowania danych	95
Rozwiązania do przechowywania danych	96
Składnice danych	96
Magazyny danych operacyjnych	97
Centra danych	99

Procesy danych	101
Zarządzanie danymi głównymi	101
Wirtualizacja i federacja danych	102
Katalogi danych	107
Platformy danych	108
Podsumowanie	109
7. Metody projektowe	110
Porównanie systemów OLTP i OLAP	111
Dane operacyjne i analityczne	113
Przetwarzanie SMP i MPP	113
Architektura Lambda	114
Architektura Kappa	117
Trwałość poliglottyczna i poliglottyczne magazyny danych	118
Podsumowanie	120
8. Metody modelowania danych	121
Modelowanie relacyjne	121
Klucze	121
Diagramy relacji encji	122
Reguły i formy normalizacji	122
Śledzenie zmian	123
Modelowanie wymiarowe	124
Fakty, wymiary i klucze	125
Monitorowanie zmian	125
Denormalizacja	126
Wspólny model danych	128
Model Data Vault	129
Metodologie Kimballa i Inmona dotyczące hurtowni danych	131
Metodologia odgórna Inmona	132
Metodologia oddolna Kimballa	133
Wybór metodologii	134
Modele hybrydowe	135
Mity dotyczące metodologii	137
Podsumowanie	140
9. Metody pozyskiwania danych	141
Porównanie procesów ETL i ELT	141
Odwrócony proces ETL	143
Porównanie przetwarzania wsadowego oraz przetwarzania w czasie rzeczywistym	145
Zalety i wady przetwarzania wsadowego	146
Zalety i wady przetwarzania w czasie rzeczywistym	146

Nadzór nad danymi	147
Podsumowanie	147

Część III. Architektury danych 149

10. Nowoczesna hurtownia danych	151
Architektura nowoczesnej hurtowni danych	151
Zalety i wady architektury nowoczesnej hurtowni danych	156
Łączenie relacyjnej hurtowni danych z jeziorem danych	158
Jezioro danych	158
Relacyjna hurtownia danych	158
Kamienie milowe prowadzące do hurtowni MDW	159
Rozbudowa korporacyjnej hurtowni danych	159
Tymczasowe jezioro danych oraz korporacyjna hurtownia danych	161
Rozwiązanie kompleksowe	162
Studium przypadku: strategiczne przejście firmy Wilson & Gunkerk do nowoczesnej hurtowni danych	163
Wyzwanie	163
Rozwiązanie	163
Rezultat	164
Podsumowanie	164
11. Architektura Data Fabric	166
Architektura Data Fabric	167
Zasady dostępu do danych	167
Katalog metadanych	168
Zarządzanie danymi głównymi	169
Wirtualizacja danych	169
Przetwarzanie w czasie rzeczywistym	169
Interfejsy API	170
Usługi	170
Produkty	170
Dlaczego warto dokonać przejścia z hurtowni MDW na architekturę Data Fabric?	170
Potencjalne wady	171
Podsumowanie	171
12. Architektura Data Lakehouse	173
Opcje warstwy Delta Lake	174
Poprawa wydajności	176
Architektura Data Lakehouse	177
Co się stanie, gdy zrezygnujesz z relacyjnej hurtowni danych?	179

Relacyjna warstwa udostępniająca	181
Podsumowanie	182
13. Fundamenty siatki danych	183
Zdecentralizowana architektura danych	184
Szum wokół siatki danych	185
Cztery zasady Dehghani dotyczące siatki danych	186
Pierwsza zasada: własność domeny	186
Druga zasada: dane jako produkt	187
Trzecia zasada: samoobsługowa infrastruktura danych jako platforma	189
Czwarta zasada: nadzór nad federacyjnymi zasobami obliczeniowymi	190
„Czysta” siatka danych	191
Domeny danych	192
Logiczna architektura siatki danych	193
Różne topologie	195
Porównanie siatki danych i architektury Data Fabric	197
Warianty zastosowania	197
Podsumowanie	199
14. Czy powinno się adaptować siatkę danych? Mity, obawy i przyszłość	200
Mity	200
Mit: użycie siatki danych to cudowny środek pozwalający	
szybko poradzić sobie z wszystkimi trudnościami towarzyszącymi danym	200
Mit: siatka danych zastąpi Twoje jezioro danych i hurtownię danych	201
Mit: wszystkie projekty z hurtownią danych nie udają się,	
a siatka danych rozwiąże ten problem	201
Mit: budowanie siatki danych oznacza decentralizację absolutnie wszystkiego	201
Mit: możesz użyć wirtualizacji danych, aby utworzyć siatkę danych	201
Obawy	202
Kwestie filozoficzne i koncepcyjne	203
Łączenie danych w środowisku zdecentralizowanym	204
Inne kwestie związane z decentralizacją	205
Złożoność	206
Duplikacja	206
Wykonalność	207
Ludzie	209
Bariery na poziomie domen	210
Ocena organizacyjna: czy powinno się adaptować siatkę danych?	211
Zalecenia dotyczące implementowania z powodzeniem siatki danych	212
Przyszłość siatki danych	214
Szersze spojrzenie: zrozumienie architektur danych i ich zastosowań	215
Podsumowanie	216

15. Ludzie i procesy	219
Organizacja zespołów: role i obowiązki	220
Role w przypadku nowoczesnej hurtowni danych oraz architektur Data Fabric lub Data Lakehouse	220
Role w przypadku siatki danych	222
Dlaczego projekty się nie udają: pułapki i zapobieganie im	225
Pułapka: pozwalanie szefostwu myśleć, że analityka biznesowa jest „łatwa”	225
Pułapka: używanie niewłaściwych technologii	225
Pułapka: zgromadzenie zbyt wielu wymagań biznesowych	225
Pułapka: zgromadzenie zbyt małej liczby wymagań biznesowych	226
Pułapka: prezentowanie raportów bez wcześniejszego sprawdzenia poprawności ich zawartości	226
Pułapka: zatrudnianie niedoświadczonej firmy konsultingowej	226
Pułapka: zatrudnianie firmy konsultingowej, która zleca prace projektowe pracownikom z innych krajów	226
Pułapka: przekazywanie konsultantom własności projektu	227
Pułapka: zlekceważenie konieczności transferu zasobów wiedzy z powrotem do organizacji	227
Pułapka: zmniejszanie budżetu w połowie trwania projektu	227
Pułapka: rozpoczynanie od daty końcowej i cofanie się z działaniami	228
Pułapka: określanie struktury hurtowni danych zgodnie z danymi źródłowymi, a nie wymogami firmy	228
Pułapka: zaprezentowanie użytkownikom rozwiązania z długim czasem odpowiedzi lub innymi problemami z wydajnością	228
Pułapka: przesadzenie z projektem architektury danych lub niedopracowanie go	229
Pułapka: kiepska komunikacja między działem informatycznym i domenami biznesowymi	229
Wskazówki dotyczące sukcesu	229
Nie oszczędzaj na inwestycjach	229
Angażuj użytkowników, prezentuj im wyniki i wzbudzaj ich entuzjazm	230
Zapewnij wartość w nowych raportach i panelach kontrolnych	231
Poproś użytkowników o zbudowanie prototypu	231
Znajdź orędownika/sponsora projektu	232
Stwórz plan projektu z celem 80% efektywności	232
Podsumowanie	232

16. Technologie	234
Wybór platformy	234
Rozwiązania <i>open source</i>	234
Rozwiązania lokalne	237
Rozwiązania dostawców usług w chmurze	238
Modele usług w chmurze	240
Główni dostawcy usług w chmurze	243
Rozwiązania z wieloma chmurami	243
Środowiska oprogramowania	246
Hadoop	246
Databricks	249
Snowflake	251
Podsumowanie	252