

Inhalt

Vorwort	19
Einführung	21

Teil I: Grundlagen

1 Big Data	27
Was ist Big Data, und wie kann Big Data Ihnen helfen?	28
Data Maturity	32
Stufe 1: Reaktiv	33
Stufe 2: Informativ	34
Stufe 3: Prädiktiv	34
Stufe 4: Transformativ	35
Self-Service Business Intelligence	35
Zusammenfassung	36
2 Arten von Datenarchitekturen	39
Entwicklung von Datenarchitekturen	40
Relationales Data Warehouse	43
Data Lake	45
Modern Data Warehouses	48
Data Fabric	48
Data Lakehouse	49
Data Mesh	50
Zusammenfassung	51

3	Die Architektur-Design-Sitzung	53
	Was ist eine ADS?	53
	Warum eine ADS abhalten?	54
	Vor der ADS	55
	Vorbereiten	55
	Teilnehmerinnen und Teilnehmer einladen	58
	Die ADS leiten	60
	Einführungen	60
	Erkundung	61
	Whiteboarding	67
	Nach der ADS	68
	Tipps für die Durchführung einer ADS	70
	Zusammenfassung	72

Teil II: Allgemeine Datenarchitekturkonzepte

4	Das relationale Data Warehouse	75
	Was ist ein relationales Data Warehouse?	75
	Was ein Data Warehouse nicht ist	78
	Der Top-down-Ansatz	80
	Warum ein relationales Data Warehouse verwenden?	82
	Nachteile bei Verwendung eines relationalen Data Warehouse	86
	Ein Data Warehouse füllen	88
	Wie oft sollen die Daten extrahiert werden?	88
	Extraktionsmethoden	88
	Wie man feststellt, welche Daten sich seit der letzten Extraktion geändert haben	89
	Der Tod des relationalen Data Warehouse wurde übertrieben dargestellt	91
	Zusammenfassung	92
5	Data Lake	93
	Was ist ein Data Lake?	94
	Warum einen Data Lake verwenden?	94

Bottom-up-Ansatz	97
Best Practices für das Design von Data Lakes	98
Mehrere Data Lakes	105
Vorteile	106
Nachteile	109
Zusammenfassung	110
6 Lösungen und Prozesse zur Datenspeicherung	111
Datenspeicherlösungen	112
Data Marts	112
Operational Data Stores	113
Data Hubs	116
Datenprozesse	118
Stammdatenverwaltung	119
Datenvirtualisierung und Datenföderierung	120
Datenkataloge	126
Datenmarktplätze	127
Zusammenfassung	129
7 Ansätze für das Design	131
OLTP vs. OLAP	132
Operative und analytische Daten	135
Symmetrisches Multiprocessing und massive Parallelverarbeitung	135
Lambda-Architektur	137
Kappa-Architektur	140
Polyglotte Persistenz und polyglotte Datenspeicher	142
Zusammenfassung	143
8 Ansätze zur Datenmodellierung	145
Relationale Modellierung	145
Schlüssel	146
Entity-Relationship-Diagramme	146
Normalisierungsregeln und -formen	147
Änderungen verfolgen	149

Dimensionale Modellierung	149
Fakten, Dimensionen und Schlüssel	149
Änderungen verfolgen	150
Denormalisierung	152
Common Data Model	153
Data Vault	154
Die Methodiken von Kimball und Inmon für das Data Warehousing	156
Die Top-down-Methodik von Inmon	157
Die Bottom-up-Methodik von Kimball	159
Eine Methodik auswählen	160
Hybride Modelle	161
Mythen über die Methodiken	164
Zusammenfassung	167
9 Ansätze für die Datenaufnahme	169
ETL vs. ELT	169
Reverse ETL	172
Stapel- vs. Echtzeitverarbeitung	173
Stapelverarbeitung – Vor- und Nachteile	175
Echtzeitverarbeitung – Vor- und Nachteile	175
Data Governance	176
Zusammenfassung	177

Teil III: Datenarchitekturen

10 Das Modern Data Warehouse	181
Die MDW-Architektur	181
Die MDW-Architektur – Vor- und Nachteile	187
RDW und Data Lake kombinieren	189
Data Lake	189
Relationales Data Warehouse	189
Schritt für Schritt zum MDW	190
EDW-Erweiterung	191

Temporärer Data Lake plus EDW	192
All-in-one	193
Fallstudie: Die strategische Umstellung bei Wilson & Gunkerk auf ein MDW	194
Herausforderung	195
Lösung	195
Ergebnis	195
Zusammenfassung	196
11 Data Fabric	199
Die Data-Fabric-Architektur	200
Datenzugriffsrichtlinien	201
Metadatenkatalog	202
Stammdatenverwaltung	203
Datenvirtualisierung	203
Echtzeitverarbeitung	203
APIs	204
Dienste	204
Produkte	204
Weshalb von einem MDW auf eine Data-Fabric-Architektur umsteigen?	204
Potenzielle Nachteile	205
Zusammenfassung	206
12 Data Lakehouse	207
Delta-Lake-Features	208
Performanceverbesserungen	211
Die Data-Lakehouse-Architektur	212
Was, wenn man das RDW überspringt?	214
Relationale Serving-Schicht	217
Zusammenfassung	217
13 Data-Mesh-Grundlagen	219
Eine dezentralisierte Architektur	220
Der Hype um Data Mesh	221

Dehghanis vier Prinzipien des Data Mesh	223
Prinzip #1: Domain Ownership	223
Prinzip #2: Data-as-a-Product	224
Prinzip #3: Self-Serve-Infrastructure-as-a-Platform	226
Prinzip #4: Federated Computational Governance	228
Das »reine« Data Mesh	229
Datendomains	231
Logische Data-Mesh-Architektur	232
Verschiedene Topologien	234
Data Mesh vs. Data Fabric	236
Anwendungsfälle	237
Zusammenfassung	239
14 Data Mesh einführen? – Mythen, Bedenken und die Zukunft	241
Mythen	241
Mythos: Data Mesh ist eine Silberkugel, mit der sich alle Datenprobleme schnell lösen lassen	242
Mythos: Ein Data Mesh ersetzt Ihren Data Lake und Ihr Data Warehouse	242
Mythos: Data-Warehouse-Projekte scheitern alle – ein Data Mesh löst dieses Problem	242
Mythos: Ein Data Mesh aufzubauen bedeutet, absolut alles zu dezentralisieren	243
Mythos: Mit Datenvirtualisierungen lässt sich ein Data Mesh erstellen	243
Bedenken	244
Philosophische und konzeptionelle Fragen	245
Daten in einer dezentralisierten Umgebung kombinieren	246
Andere Probleme der Dezentralisierung	247
Komplexität	249
Duplizierung	249
Machbarkeit	250
Mitarbeiterinnen und Mitarbeiter	253
Hürden auf Domainebene	254
Organisatorische Bewertung: Sollten Sie ein Data Mesh einführen?	256

Empfehlungen für die Implementierung eines erfolgreichen Data Mesh	258
Die Zukunft von Data Mesh	259
Blick über den Tellerrand: Datenarchitekturen und ihre Anwendungen	260
Zusammenfassung	262

Teil IV: Menschen, Prozesse und Technologien

15 Menschen und Prozesse	265
Teamorganisation: Rollen und Verantwortlichkeiten	266
Rollen für MDW, Data Fabric oder Data Lakehouse	266
Rollen für Data Mesh	268
Warum Projekte scheitern: Fallstricke und Prävention	272
Fallstrick: Führungskräfte denken, dass BI »einfach« ist	272
Fallstrick: Die falschen Technologien verwenden	272
Fallstrick: Zu viele Geschäftsanforderungen sammeln	273
Fallstrick: Zu wenige Geschäftsanforderungen sammeln	273
Fallstrick: Berichte präsentieren, ohne ihren Inhalt zuvor zu validieren	274
Fallstrick: Unerfahrene Berater beauftragen	274
Fallstrick: Eine Beratungsfirma beauftragen, die die Entwicklung an Offshore-Arbeiter outsourct	274
Fallstrick: Projektbesitz an Berater abgeben	275
Fallstrick: Den notwendigen Wissenstransfer zurück in die Organisation vernachlässigen	275
Fallstrick: Das Budget auf halbem Weg durch das Projekt kürzen	275
Fallstrick: Von einem Enddatum aus rückwärts arbeiten	276
Fallstrick: Das Data Warehouse so strukturieren, dass es die Quelldaten und nicht die Geschäftsbedürfnisse widerspiegelt	276
Fallstrick: Endbenutzern eine Lösung mit langen Reaktionszeiten oder anderen Performanceproblemen präsentieren	277

Fallstrick: Zu viel (oder zu wenig) Design Ihrer Datenarchitektur	277
Fallstrick: Mangelnde Kommunikation zwischen IT und Businessdomains	277
Tipps für den Erfolg	278
Knausern Sie nicht mit Ihren Investitionen	278
Benutzer und Benutzerinnen einbeziehen, ihnen Ergebnisse zeigen und sie begeistern	279
Mehrwert für neue Berichte und Dashboards	280
Die Endbenutzer bitten, einen Prototyp zu erstellen	280
Einen Projektchampion/Sponsor finden	281
Einen Projektplan erstellen, der auf 80% Effizienz abzielt	281
Zusammenfassung	282
16 Technologien	285
Eine Plattform auswählen	285
Open-Source-Lösungen	285
On-Premises-Lösungen	288
Cloud-Provider-Lösungen	290
Cloud-Service-Modelle	293
Große Cloud-Provider	295
Multi-Cloud-Lösungen	296
Software-Frameworks	299
Hadoop	300
Databricks	304
Snowflake	306
Zusammenfassung	307
Index	309