

Inhalt

Seite

Vorwort	9
1 Einführung	11
1.1 Warum Open Source Business Intelligence?	12
1.1.1 Business Intelligence	12
1.1.2 Open Source Business Intelligence	13
1.1.3 Zum Aufbau des Buches	15
1.1.4 Ziele und Zielgruppen des Buches	16
1.2 Business Intelligence: Überblick und aktuelle Trends	18
1.2.1 Der technische BI-Begriff als Sammelbezeichnung für multidimensional speichernde Informationssysteme und deren Werkzeuge	18
1.2.2 Werkzeuge der Datensammlung und -verteilung	20
1.2.3 Werkzeuge der Informationsaufbereitung und -bereitstellung	23
1.2.4 Werkzeuge der Analyse und Präsentation der Informationen	30
1.2.5 Metadatenmanagement	37
1.2.6 Die Referenzarchitektur als Zusammenfassung technischer BI-Elemente	37
1.2.7 Die betriebswirtschaftliche BI-Anwendung als kontinuierlicher Anreiz zur technischen Innovation	39
1.2.8 BI-Architekturen und ihre technische Evolution	40
1.2.9 BI-Industrialisierung durch allgemeinen IT-Fortschritt	43
1.3 Open Source	45
1.3.1 Historische Wurzeln	45
1.3.2 Lizenzmodelle und ihre Grenzen	47
1.3.3 Open Content	50
1.3.4 Open-Source-Software als Innovationsmotor	51
1.3.5 Sicherheit von Open-Source-Software	52
1.3.6 Wirtschaftliche Bedeutung von Open Source	53
1.3.7 Quellen für Open-Source-Software	55
1.3.8 Open Source im kommerziellen Einsatz	57
2 Architekturen und Kosten	59
2.1 BI-Strategie	60
2.1.1 Aspekte der BI-Strategie	60
2.1.2 Open Source als Teil der IT- bzw. BI-Strategie	65
2.2 BI-Architektur: Anforderungen und Ausprägungen	67
2.2.1 „BI Light“: Berichte und Analysen ohne Data Warehouse	69
2.2.2 Data Warehouse und Enterprise Data Warehouse	71
2.2.3 Sonderfall Planungs- und Budgetierungsprozesse	76
2.2.4 BI und Enterprise Content Management	77
2.2.5 Serviceorientierte Architekturen (SOA) und BI	80
2.2.6 Cloud Computing und OSBI	82
2.2.7 Architektur als Erfolgsfaktor	82

	Seite
2.3 Vorgehensmodelle	83
2.3.1 Projektphasen	83
2.3.2 Top-down- vs. Bottom-up-Vorgehen	87
2.3.3 Vorgehensmodelle des Software Engineering	88
2.3.4 Spezielle BI-Vorgehensmodelle	94
2.3.5 Aktueller Trend: Agile BI	95
2.3.6 Auswahl eines Vorgehensmodells	98
2.4 TCO – Kosten für Skalierbarkeit und Performance bei OSBI	99
2.4.1 TCO-Berechnung in BI-Projekten und ihre Grenzen	100
2.4.2 Kostenvergleichskalkulation OSBI-Stack vs. Closed-Source-Stack.....	104
2.4.3 Checkliste: Wann lohnt sich der Umstieg?	109
2.4.4 Zusammenfassende Empfehlungen	111
 3 OSBI-Lösungen in der Praxis	 113
3.1 Open-Source-BI-Tools	114
3.1.1 ETL – Extract, Transform, Load	115
3.1.2 Relationale Open-Source-Datenbanksysteme	117
3.1.3 Reporting	120
3.1.4 On-Line Analytical Processing (OLAP)	122
3.1.5 Data Mining	124
3.1.6 All in One – BI-Suiten	126
3.1.7 OSBI-Angebote heute	128
3.2 Talend Open Studio	129
3.2.1 Über Talend Open Studio	129
3.2.2 Installation und erster Start	129
3.2.3 Der Anwendungsfall: Kaffeehaus Espresso	130
3.2.4 Der Arbeitsbereich	131
3.2.5 Der erste ETL-Job	133
3.2.6 Typische Fehlerquellen und Fehlerbehandlung	138
3.2.7 Kontexte für unterschiedliche Zielsysteme	139
3.2.8 Export	139
3.2.9 Dokumentation	140
3.2.10 Modellierung von Business-Prozessen	140
3.2.11 Talend on Demand	141
3.2.12 Community und Hilfe im Internet	141
3.2.13 Ausblick auf die kommerziellen Versionen und Services	142
3.3 OSBI-Suiten im Vergleich – eine strukturelle Analyse	147
3.3.1 Installation	147
3.3.2 Architektur der OSBI-Suiten	154
3.3.3 Informationsangebote	155
3.3.4 Kostenpflichtiger Support	156
3.3.5 Feature-Vergleich	158
3.3.6 Zusammenfassender Vergleich	159

	Seite
3.4 Jaspersoft	160
3.4.1 Schnittstellen und Aufteilung der Komponenten	160
3.4.2 Beschreibung der Komponenten	161
3.4.3 Datenaustausch zwischen Komponenten	163
3.4.4 ETL-Prozess	165
3.4.5 OLAP-Analyse und Reporting	166
3.4.6 Multi-User und Benutzergruppen	173
3.4.7 Entwicklerkomponenten, Open Source, Erweiterbarkeit	174
3.4.8 Service, Support, Dokumentation	175
3.4.9 Zusammenfassende Beurteilung	175
3.5 Pentaho	177
3.5.1 Das Unternehmen und seine Produkte	177
3.5.2 Komponenten der Pentaho Community Edition	179
3.5.3 Voraussetzungen und Installation	180
3.5.4 BI Platform	181
3.5.5 ETL mit Pentaho Data Integration	182
3.5.6 Metadaten	184
3.5.7 Analyse	185
3.5.8 Berichtswesen	189
3.5.9 Bewertung	192
3.6 Palo Suite	194
3.6.1 Komponenten: Beschreibung und Aufteilung	194
3.6.2 Datenmodellierung	198
3.6.3 Dateneingabe/-import	207
3.6.4 OLAP-Analyse und Reporting	210
3.6.5 Administration	211
3.6.6 Service, Support, Dokumentation	211
3.6.7 Zusammenfassende Beurteilung	211
3.7 Data Mining mit dem RapidMiner	214
3.7.1 Was ist Data Mining?	214
3.7.2 Ziele	214
3.7.3 Typische Anwendungsgebiete	215
3.7.4 Wichtige Typen von Data-Mining-Aufgabenstellungen	216
3.7.5 Data-Mining-Verfahren	218
3.7.6 Fallstricke im Data Mining	218
3.7.7 Betrieblicher Einstieg ins Data Mining	220
3.7.8 Data Mining mit Werkzeugen von Rapid-I	221
3.7.9 Zusammenfassende Empfehlungen	231
3.8 Anwendungsszenarien	232
3.8.1 Szenario 1: OLAP-Analyse bestehender Tabelle	233
3.8.2 Szenario 2: Unternehmensplanung mit Palo	237
3.8.3 Szenario 3: Auslesen von Daten aus SAP-Systemen	239
3.8.4 Szenario 4: Ablösung einer programmierten Reporting-Lösung	241
3.8.5 Szenario 5: Datenqualität	243
3.8.6 Zusammenfassung	246

	Seite
4 Der Markt und seine Trends	247
4.1 Marktphänomen Open-Source-BI – ernstzunehmende Alternative zu traditionellen Angeboten?	248
4.1.1 Hauptakteure des OSBI-Marktes	248
4.1.2 Abdeckung von BI-Aufgaben durch OSBI-Produkte	252
4.1.3 Vor- und Nachteile des Einsatzes von OSBI-Produkten	256
4.1.4 Die aktuelle Marktsituation für OSBI	258
4.2 Geschäftsmodelle von Open-Source-BI-Anbietern	259
4.2.1 Struktur von Geschäftsmodellen	259
4.2.2 Open-Source-Geschäftsmodelle	260
4.2.3 Geschäftsmodelle der Anbieter von Open-Source-BI-Software	263
4.2.4 Entwicklungen und Trends	269
4.3 Einsatzszenarien für OSBI	271
4.3.1 Community versus Enterprise oder Professional: Open Source oder Commercial Open Source?	272
4.3.2 Das Dilemma der Anbieter: Ohne Lizenzen keine Kontrolle?	273
4.3.3 Das Problem und das Potenzial der Download-Zahlen	274
4.3.4 KMU als Zielgruppe für OSBI	275
4.3.5 Mögliche Szenarien für den Einsatz von OSBI	279
4.3.6 Resümee	283
4.4 OSBI trifft Forschung: GPUs als parallele Coprozessoren für OLAP-Aggregationen	285
4.4.1 Multidimensionale Aggregation	285
4.4.2 GPGPU	288
4.4.3 GPU-Computing im Palo OLAP Server	290
4.4.4 Innovation und OSBI	295
 Anhang	
Die Autoren	296
Danksagung	298
Abkürzungsverzeichnis	299
Literaturverzeichnis	300
Register	315