

Contents

Foreword	17
Preface	19

PART I SAP's Enterprise Information Management Strategy and Portfolio

1 Introducing Enterprise Information Management 27

1.1	Defining Enterprise Information Management	28
1.1.1	Example of Information Flow Through a Company	30
1.1.2	Types of Information Included in Enterprise Information Management	33
1.2	Common Use Cases for EIM	36
1.2.1	Operational EIM	37
1.2.2	Analytical EIM	39
1.2.3	Information Governance	40
1.3	Common Drivers for EIM	40
1.3.1	Operational Efficiency as a Driver of EIM	41
1.3.2	Information as an Organizational Asset	43
1.3.3	Compliance as a Driver of EIM	44
1.4	Impact of Big Data on EIM	45
1.5	SAP's Strategy for EIM	48
1.6	Typical User Roles in EIM	50
1.7	Example Company: NeedsEIM, Inc.	51
1.7.1	CFO Issues	52
1.7.2	Purchasing Issues	52
1.7.3	Sales Issues	52
1.7.4	Engineering and Contracts Issues	52
1.7.5	Information Management Challenges Facing NeedsEIM, Inc.	53
1.8	Summary	53

2 Introducing Information Governance 55

2.1	Introduction to Information Governance	56
-----	--	----

2.2	Evaluating and Developing Your Information Governance	
	Needs and Resources	58
2.2.1	Evaluating Information Governance	58
2.2.2	Developing Information Governance	62
2.3	Optimizing Existing Infrastructure and Resources	63
2.4	Establishing an Information Governance Process: Examples	65
2.4.1	Example 1: Creating a New Reseller	67
2.4.2	Example 2: Supplier Registration	68
2.4.3	Example 3: Data Migration	71
2.5	Rounding Out Your Information Governance Process	75
2.5.1	The Impact of Missing Data	75
2.5.2	Gathering Metrics and KPIs to Show Success	76
2.5.3	Establish a Before-and-After View	80
2.6	Summary	80

3 Big Data with SAP HANA and Hadoop 81

3.1	SAP HANA	81
3.1.1	Basics of SAP HANA	82
3.1.2	How SAP HANA Works	83
3.1.3	SAP HANA for Analytics and Business Intelligence	85
3.1.4	SAP HANA as an Application Platform	85
3.2	SAP HANA and EIM	86
3.2.1	Using SAP Data Services to Load Data into SAP HANA ...	87
3.2.2	EIM Solutions and SAP HANA	88
3.3	Big Data and Hadoop	88
3.3.1	The Rise of Parallel Processing Architectures	88
3.3.2	NoSQL: A New Kind of Database	89
3.3.3	Hadoop Primer	90
3.3.4	Hadoop: Enterprise Use Cases	95
3.4	SAP HANA and Hadoop	99
3.4.1	The V's: Volume, Variety, Velocity	99
3.4.2	SAP HANA: Designed for Enterprises	100
3.4.3	Hadoop as an SAP HANA Extension	100
3.5	EIM and Hadoop	100
3.5.1	ETL—Data Services and the Information Design Tool	101
3.5.2	Unsupported: Information Governance and Information Lifecycle Management	101
3.6	Summary	102

4 SAP's Solutions for Enterprise Information Management 103

4.1	SAP Data Services as a Data Foundation	106
4.1.1	Basics of SAP Data Services	106
4.1.2	SAP Data Services Integration with SAP Applications	108
4.1.3	SAP Data Services Integration with Non-SAP Applications	111
4.1.4	Data Cleansing and Data Validation with SAP Data Services	112
4.1.5	Text Data Processing in SAP Data Services	115
4.2	SAP Information Steward	118
4.2.1	Data Profiling and Data Quality Monitoring	119
4.2.2	Metadata Analysis	122
4.2.3	Cleansing Rules	123
4.2.4	Business Glossary	125
4.3	SAP NetWeaver Master Data Management and SAP Master Data Governance	126
4.3.1	SAP NetWeaver Master Data Management	127
4.3.2	SAP Master Data Governance	127
4.4	SAP Solutions for Enterprise Content Management	130
4.4.1	Overview of SAP's ECM Solutions	133
4.4.2	SAP Extended Enterprise Content Management by OpenText	137
4.4.3	SAP Document Access by OpenText and Archiving by OpenText	140
4.5	SAP NetWeaver Information Lifecycle Management	142
4.5.1	Retention Management	146
4.5.2	System Decommissioning	147
4.6	SAP Solutions for Data Migration	150
4.6.1	SAP Data Migration Software	151
4.6.2	Predefined Content and Rules for Data Migration	153
4.6.3	Systems Migration	155
4.6.4	SAP Rapid Marts	155
4.7	Information Governance in SAP	156
4.7.1	Information Governance Use Scenario Phasing	157
4.7.2	Technology Enablers for Information Governance	159
4.8	NeedsEIM, Inc. and SAP's Solutions for EIM	162
4.9	Summary	163

5 Practical Examples of EIM 165

5.1	EIM Architecture Recommendations and Experiences by Procter and Gamble	165
5.1.1	Principles of an EIM Architecture	166
5.1.2	Scope of an EIM Enterprise Architecture	168
5.1.3	Structured Data	169
5.1.4	The Dual Database Approach	170
5.1.5	Typical Information Lifecycle	171
5.1.6	Data Standards	176
5.1.7	Unstructured Data	177
5.1.8	Governance	179
5.1.9	Role of the Enterprise Information Architecture Organization	184
5.2	Managing Data Migration Projects to Support Mergers and Acquisitions	184
5.2.1	Scoping for a Data Migration Project	185
5.2.2	Data Migration Process Flow	187
5.2.3	Enrich the Data Using Dun and Bradstreet (D&B) with Data Services	191
5.3	Evolution of SAP Data Services at National Vision	192
5.3.1	Phase 1: The Enterprise Data Warehouse	192
5.3.2	Phase 2: Enterprise Information Architecture—Consolidating Source Data	193
5.3.3	Phase 3: Data Quality and the Customer Hub	195
5.3.4	Phase 4: Application Integration and Data Migration	197
5.3.5	Phase 5: Next Steps with Data Services	198
5.4	Recommendations for a Master Data Program	198
5.4.1	Common Enterprise Vision and Goals	198
5.4.2	Master Data Strategy	199
5.4.3	Roadmap and Operational Phases	199
5.4.4	Business Process Redesign and Change Management	199
5.4.5	Governance	200
5.4.6	Technology Selection	200
5.5	Recommendations for Using SAP NetWeaver Process Integration and SAP Data Services	201
5.5.1	A Common Data Integration Problem	201
5.5.2	A Data Integration Analogy	202

5.5.3	Creating Prescriptive Guidance to Help Choose the Proper Tool	203
5.5.4	Complex Examples in the Enterprise	204
5.5.5	When All Else Fails...	206
5.6	Ensuring a Successful Enterprise Content Management Project by Belgian Railways	206
5.6.1	Building the Business Case	206
5.6.2	Key Success Factors for Your SAP Extended Enterprise Content Management by OpenText Project	213
5.7	Recommendations for Creating an Archiving Strategy	216
5.7.1	What Drives a Company into Starting a Data Archiving Project?	216
5.7.2	Who Initiates a Data Archiving Project?	217
5.7.3	Project Sponsorship	218
5.8	Summary	220

PART II Working with SAP's Enterprise Information Management Solutions

6 Getting to Know Your Data with SAP Information Steward ... 223

6.1	Cataloging Data Assets and Their Relationships	224
6.1.1	Configuring a Metadata Integrator Source	225
6.1.2	Executing or Scheduling Execution of Metadata Integration	227
6.2	Profiling Data	228
6.2.1	Configuration and Setup of Connections and Projects	229
6.2.2	Getting Basic Statistical Information about the Data Content	232
6.2.3	Identifying Cross-Field or Cross-Column Data Relationships	236
6.3	Monitoring and Assessing the Quality of Your Data	239
6.3.1	Defining Validation Rules Representing Business Requirements	241
6.3.2	Binding Rules to Data Sources for Data Quality Assessment	244
6.3.3	Executing Rule Tasks and Viewing Results	246
6.4	Summary	250

7 Introducing SAP Data Services 251

- 7.1 Data Integration Scenarios 252
- 7.2 SAP Data Services Platform Architecture 253
 - 7.2.1 User Interface Tier 255
 - 7.2.2 Server Tier 257
- 7.3 SAP Data Services Designer Overview 259
- 7.4 Creating Data Sources and Targets 263
 - 7.4.1 Connectivity Options for SAP Data Services 263
 - 7.4.2 Connecting to SAP 265
 - 7.4.3 Connecting to Hadoop 268
- 7.5 Creating Your First Job 268
 - 7.5.1 Create the Data Flow 268
 - 7.5.2 Add a Source to the Data Flow 269
 - 7.5.3 Add a Query Transform to the Data Flow 269
 - 7.5.4 Add a Target to the Data Flow 270
 - 7.5.5 Map the Source Data to the Target by Configuring
the Query Transform 270
 - 7.5.6 Create the Job and Add the Data Flow to the Job 271
- 7.6 Basic Transformations Using the Query Transform and Functions ... 272
- 7.7 Overview of Complex Transformations 275
 - 7.7.1 Platform Transformations 275
 - 7.7.2 Data Integrator Transforms 277
- 7.8 Executing and Debugging Your Job 280
- 7.9 Exposing a Real-Time Service 281
 - 7.9.1 Create a Real-Time Job 282
 - 7.9.2 Create a Real-Time Service 284
 - 7.9.3 Expose the Real-Time Service as a Web Service 286
- 7.10 SAP Data Services and SAP HANA 287
- 7.11 Summary 288

8 Managing Data Quality with SAP Data Services 289

- 8.1 Data Cleansing 291
 - 8.1.1 Data Parsing 292
 - 8.1.2 Data Standardization 303
 - 8.1.3 Data Correction 309
 - 8.1.4 Data Validation 311

8.2	Data Enhancement	313
8.3	Data Matching	315
8.3.1	Rule-Based Matching Method	321
8.3.2	Weighted Scoring Matching Method	322
8.3.3	Combination Matching Method	323
8.3.4	Data Consolidation	330
8.4	Other Important Data Quality Capabilities	333
8.4.1	SAP Data Quality Management, Version for SAP Solutions	333
8.4.2	Data Quality Blueprints	335
8.4.3	Using Data Quality Beyond Customer Data	336
8.5	Summary	337

9 Monitoring Data Quality with Scorecards in SAP Information Steward 339

9.1	Components of a Data Quality Scorecard	340
9.2	Defining and Setting Up a Data Quality Scorecard	343
9.2.1	Modeling the High-Level Key Data Domains	343
9.2.2	Defining Data Quality Dimensions for the Key Data Domains	344
9.2.3	Adding the Validation Rules and Quality Dimensions for a Key Data Domain	346
9.2.4	Binding the Data Sources to the Data Quality Scorecard	348
9.3	Viewing the Data Quality Scorecard	349
9.3.1	Drilling Down into Data Quality Scorecard Details	350
9.3.2	Analyze Failed Record Sample Data	351
9.4	Identifying Data Quality Issue Impact or Data Quality Root Causes in the System Landscape	352
9.4.1	Reviewing Data Quality Issue Impact on Downstream Processes or Reports	353
9.4.2	Analyzing Data Quality Root Causes	355
9.5	Summary	355

10 Processing Text Data with SAP Data Services 357

10.1	Introduction to Text Data Processing Capabilities in SAP Data Services	359
10.1.1	Entity Extraction Transform Overview	361
10.1.2	How Extraction Works	362
10.2	Text Data Processing and NeedsEIM, Inc.	364
10.2.1	NeedsEIM, Inc. Pain Points	364
10.2.2	Using the Entity Extraction Transform	366
10.3	Summary	372

11 Migrating Data with SAP Data Services 373

11.1	Introduction to Data Migration	373
11.2	Data Migration Content Delivered by SAP Best Practices and Rapid Deployment Solutions	375
11.2.1	Data Migration Content	376
11.2.2	Architecture of SAP Best Practices and Data Services for Data Migration to SAP Applications	377
11.3	Migrating Customer Data to SAP ERP	379
11.3.1	Step 1: Download the Migration Content	379
11.3.2	Step 2: Install the Software	380
11.3.3	Step 3: Review the Jobs in Data Services	381
11.4	Systems Migration	383
11.5	Summary	384

12 Introducing SAP Master Data Governance 385

12.1	SAP Master Data Governance Overview	386
12.1.1	Deployment Options	388
12.1.2	Change Request and Staging	389
12.1.3	Process Flow in SAP Master Data Governance	391
12.2	Getting Started with SAP Master Data Governance	392
12.2.1	Data Modeling	393
12.2.2	User Interface (UI) Modeling	394
12.2.3	Data Quality and Search	395
12.2.4	Process Modeling	396
12.2.5	Data Replication	397

12.2.6	Key and Value Mapping	397
12.2.7	Data Transfer	399
12.2.8	Activities Beyond Customizing	399
12.3	Governance for Custom-Defined Objects: Example	400
12.3.1	Define New Models for Airline Entity and Currency Entity	401
12.3.2	Define Relationship Between the Airline and Currency Entity Types	402
12.3.3	Define User Interface	403
12.3.4	Link the User Interface Configuration to the Data Model	405
12.3.5	Create a Change Request Process	406
12.3.6	Assign Processors to Workflow	407
12.3.7	Test the New Airline Change Request	407
12.4	Rule-Based Workflows in SAP Master Data Governance	409
12.4.1	Classic SAP Business Workflow and Rule-Based Workflow using SAP Business Workflow and BRFplus	410
12.4.2	Designing Your First Rule-Based Workflow in SAP Master Data Governance	417
12.5	NeedsEIM, Inc.: Master Data Remediation	420
12.6	Summary	422

13 Introducing SAP NetWeaver Information Lifecycle Management **423**

13.1	The Basics of Information Lifecycle Management	425
13.1.1	What Does ILM Mean?	425
13.1.2	Why Is ILM Needed?	425
13.2	Overview of SAP NetWeaver Information Lifecycle Management	426
13.2.1	Cornerstones of SAP NetWeaver ILM	428
13.2.2	Data Archiving Basics	429
13.2.3	ILM-Aware Storage	432
13.2.4	Architecture—What You Need to Run SAP NetWeaver ILM	434
13.3	Managing the Lifecycle of Information in Live Systems	436
13.3.1	Audit Area	437
13.3.2	Data Destruction	439

- 13.3.3 Legal Hold Management 440
- 13.4 Managing the Lifecycle of Information from Legacy Systems 441
 - 13.4.1 Preliminary Steps 442
 - 13.4.2 Steps Performed in the Legacy System 444
 - 13.4.3 Steps Performed in the Retention Warehouse System 444
 - 13.4.4 Handling Data from Non-SAP Systems During Decommissioning 446
- 13.5 System Decommissioning—Detailed Example 447
 - 13.5.1 Data Extraction 448
 - 13.5.2 Data Transfer and Conversion 452
 - 13.5.3 Reporting 459
 - 13.5.4 Data Destruction 463
- 13.6 Summary 466

14 Introducing SAP Extended Enterprise Content Management by OpenText 467

- 14.1 Capabilities of SAP Extended ECM 469
 - 14.1.1 Data and Document Archiving 470
 - 14.1.2 Records Management 471
 - 14.1.3 Content Access 472
 - 14.1.4 Document-Centric Workflow 472
 - 14.1.5 Document Management 472
 - 14.1.6 Capture 473
 - 14.1.7 Collaboration and Social Media 473
- 14.2 How SAP Extended ECM Works with the SAP Business Suite 474
- 14.3 Integration Content for SAP Business Suite and SAP Extended ECM 476
 - 14.3.1 SAP ArchiveLink 476
 - 14.3.2 Content Management Interoperability Standard (CMIS) and SAP ECM Integration Layer 477
 - 14.3.3 SAP Extended ECM Workspaces 479
- 14.4 Summary 485
- The Authors 487
- Index 495