Table of Contents

INTRODUCTION xxix

PART I: INTRODUCTION TO ENTERPRISE CONTENT MANAGEMENT

CHAPTER 1: WHAT IS ENTERPRISE CONTENT MANAGEMENT? 3

Introduction to ECM 4
A Historical Perspective 4
Document Imaging 4
Electronic Documents 6
COLD/Enterprise Report Management 6
Business Process Management/Workflow 6
ECM Components 7
Capture 7
Paper 7
Office Documents 8
E-mail 8
Reports 9
Electronic Forms 9
Other Sources 9
Store and Preserve 9
Software 10
Hardware and Media Technologies 10
Cloud 11
Management Components 12
Document Management 12
Web Content Management 12
Business Process Management and Workflow 12
Records Management 12
Collaboration 13
Delivery 13
Search 13
Viewing 14
Transformation 14
Security 16
Summary 16
CHAPTER 2: THE SHAREPOINT 2010 PLATFORM 17

A Brief History of SharePoint 18
SharePoint 2010 18
Capability Categories 19
Sites 19
Composites 19
Insights 20
Communities 20
Content 21
Search 21
SharePoint Concepts 21
Architecture 23
Development Concepts 26
ECM in SharePoint 2010 27
Managed Metadata 27
Ratings 28
The Content Type Hub 28
Search 28
Workflow 28
Document Sets 28
Document IDs 29
Content Organizer 29
Records Management 29
Digital Asset Management 29
Web Content Management 29
Summary 30

PART II: PILLARS OF SHAREPOINT ECM

CHAPTER 3: DOCUMENT MANAGEMENT 33

What Is Document Management? 34
Microsoft SharePoint As a Document Management System 35
Document Taxonomy 35
Document Libraries 36
The Document Library Programming Model 36
Columns 39
The Column Programming Model 41
Content Types 44
The Content Type Programming Model 45
Managed Metadata 47
CHAPTER 4: WORKFLOW 87

Workflow and ECM 87

Windows Workflow Foundation 88
CHAPTER 8: RECORDS MANAGEMENT 243

What Is Records Management? 244
Why Records Management Is Important 244
Microsoft SharePoint as a Records Management System 245
CHAPTER 10: DOCUMENT IMAGING 293

What Is Document Imaging? 294
SharePoint as a Document Imaging Platform 295
Setting Up the Scenario 295
Solution Data Flow Diagram 295
Model-View-ViewModel Primer 296
Creating a Simple WPF Capture Application 298
Architecture and Design 299
Implementation 299
Building the MVVM Infrastructure 299
Building the Target Dialog 300
Building the Main Window 301
Deployment 322
Creating a Simple Silverlight Viewer Web Part 322
Architecture and Design 323
Implementation 323
Building the Image Loader 323
Building the Imaging Web Service 324
Building the Imaging HTTP Handler 325
Building the Viewer Web Part 325
Making the Application Accessible from JavaScript 326
Deployment 327
Deploying the Imaging Services 327
Deploying the Viewer Application as a Web Part 327
Log Files 392
Crawl Databases 393
Content Databases 395
Property Databases 396
Service Application Databases 397
Management Databases 400
Prioritizing Disk I/O 400
Index Partition Storage 401
Storage Tuning and Optimization 401
Storage Performance Monitoring 401
Database Data File Management 402
Remote BLOB Storage 403
When to Implement an RBS Solution 405
RBS Provider Options 406
Backup and Restore Considerations 407
SQL Server Licensing Considerations 407
SharePoint 2010 Scalable Topology Design 408
Knowing the Users, the Corpus, and the Processes 408
Farm Size Definitions 409
The Case for Additional Web Servers 412
The Case for Additional Application Servers 412
The Case for Additional SQL Servers 412
Scalable Taxonomy Factors 413
Content Organization and Scalable Taxonomy 414
An Exercise in Scalable Taxonomy Design 415
Content Database Size Supported Limits 416
Performance and Resource Throttling 417
Summary 418

PART III: SHAREPOINT ECM SUPPORT CONCEPTS

CHAPTER 13: ECM FILE FORMATS 421

It's Alive — Your Document, That Is 422
Microsoft Office Formats 422
Microsoft Office Binary 422
Office Open XML 423
Viewing and Editing Microsoft Office Formats with Office Web Apps 425
Word Automation Services 428
Open Document Format 437
Archive Formats 437
CHAPTER 14: THE SHAREPOINT ECM ECOSYSTEM 451

The Microsoft Partner Ecosystem 451
Becoming a Partner 452
ISV/Software Competency 452
The SharePoint Ecosystem 453
Technical Community 453
ISV Solutions 454
ABBYY 455
AvePoint 458
GimmalSoft 460
KnowledgeLake 462
Nintex 465
Summary 467

CHAPTER 15: GUIDANCE FOR SUCCESSFUL ECM PROJECTS 469

Migrating to SharePoint 2010 470
Identifying Content for Migration 470
Extracting Content from the Source System 470
File Shares 470
Internally Developed Document Management Solutions 471
Other Legacy Document Management Solutions 472
Preparing Content for Importing 472
Setting the Content Type 472
Metadata Merge 472
Controlling the Import Process 473
General Metadata Cleanup 473
Importing Content into SharePoint 473
Protocols for Importing Content 474
Web Services 474
SharePoint Server Object Model 477
FrontPage Remote Procedure Calls (FPRPC) 479
Protocols for Updating SharePoint Content 480
SharePoint Server Object Model 480
SharePoint Client Object Model 483
Mapping Legacy ECM Features to SharePoint Solutions 486
Document Versions 487
Metadata-based Security 487
Document Protection and Rights Management 488
Annotations and Redaction 488
Search 488
Scanning and Indexing 488
Records Retention and Disposition 489
Workflow 489
Avoiding the Pitfalls 489
Capacity Planning 489
Illegal Characters in Metadata 489
Missing Required Data 490
Content Database Transaction Log Growth 490
Managing Upload File Size Restrictions 490
Upgrading a SharePoint ECM Farm to SharePoint 2010 491
Know Your Farm 491
SharePoint Portal Server 2003 and WSS v2.0 491
Microsoft Office SharePoint Server 2007 and WSS v3.0 492
Imaging or Archive-Only Farm with No Customization 492
Collaboration Farm with Customizations 492
Collaboration Farm with Large Imaging or Archive Site Collections 492
Summary 493
INDEX 495