

Contents

Introduction	xxix
Chapter 1: Welcome to SQL Server Integration Services	1
SQL Server SSIS Historical Overview	2
What's New in SSIS	2
Getting Started	3
Import and Export Wizard	3
The Business Intelligence Development Studio	4
Architecture	5
Packages	7
Tasks	8
Data Source Elements	9
Data Source Views	10
Precedence Constraints	11
Constraint Value	11
Conditional Expressions	11
Containers	12
Variables	13
Data Flow Elements	13
Sources	14
Destinations	15
Transformations	15
Error Handling and Logging	17
Editions of SQL Server	19
Summary	20
Chapter 2: The SSIS Tools	21
Import and Export Wizard	21
Business Intelligence Development Studio	28
Creating Your First Package	30
The Solution Explorer Window	32
The Toolbox	33
The Properties Windows	34
Navigation Pane	35
Other Windows	36

Contents

The SSIS Package Designer	36
Control Flow	37
Connection Managers	40
Variables	41
Data Flow	42
Event Handlers	43
Package Explorer	44
Executing a Package	44
Package Installation Wizard	45
Management Studio	45
Summary	46
Chapter 3: SSIS Tasks	47
SSIS Task Objects	47
Using the Task Editor	48
The Task Editor Expressions Tab	49
Execution Results	50
Common Properties	51
Looping and Sequence Tasks	53
Scripting Tasks	53
ActiveX Script Task	53
Script Task (.NET)	55
Analysis Services Tasks	57
Analysis Services Execute DDL Task	57
Analysis Services Processing Task	58
Data Mining Query Task	59
Data Flow Task	61
Data Preparation Tasks	61
Data Profiler	62
File System Task	64
FTP Task	67
Web Service Task	69
XML Task	75
RDBMS Server Tasks	79
Bulk Insert Task	79
Execute SQL Task	85
Workflow Tasks	98
Execute Package Task	98
Execute Process Task	102
Message Queue Task	104
Send Mail Task	105

WMI Data Reader Task	106
WMI Event Watcher Task	108
SMO Administration Tasks	110
Transfer Database Task	110
Transfer Error Messages	112
Transfer Logins Task	112
Transfer Master Stored Procedures Task	113
Transfer Jobs Task	114
Transfer SQL Server Objects Task	115
Summary	116
Chapter 4: Containers	117
Task Host Containers	117
Sequence Containers	117
Groups	118
For Loop Container	119
Foreach Loop Container	122
Foreach File Enumerator Example	123
Foreach ADO Enumerator Example	125
Summary	129
Chapter 5: The Data Flow	131
The Data Flow	131
Data Viewers	132
Sources	132
OLE DB Source	132
Excel Source	135
Flat File Source	136
Raw File Source	142
XML Source	142
ADO.NET Source	142
Destinations	142
Data Mining Model Training	143
DataReader Destination	144
Dimension and Partition Processing	144
Excel Destination	144
Flat File Destination	145
OLE DB Destination	145
Raw File Destination	146
Recordset Destination	146
SQL Server and Mobile Destinations	146

Contents

Transformations	146
Synchronous versus Asynchronous Transformations	147
Aggregate	147
Audit	149
Cache Transform	150
Character Map	150
Conditional Split	151
Copy Column	153
Data Conversion	153
Data Mining Query	154
Derived Column	154
Export Column	155
Fuzzy Lookup	157
Fuzzy Grouping	163
Import Column	167
Lookup Transform	170
Merge Transform	170
Merge Join	170
Multicast	171
OLE DB Command	172
Percentage and Row Sampling	175
Pivot Transform	175
Unpivot	179
Row Count	182
Script Component	184
Slowly Changing Dimension	184
Sort	185
Term Extraction	186
Term Lookup	190
Union All	192
Data Flow Example	193
Summary	197
Chapter 6: Using Expressions and Variables	199
The Paradigm	199
Expression Overview	200
Variable Overview	201
Understanding Data Types	202
SSIS Data Types	202
Additional Date and Time Type Support	204
Wrong Data Types and Sizes Can Affect Performance	204

Unicode and Non-Unicode Conversion Issues	205
Casting in SSIS Expressions	206
Using Variables	208
Defining Variables	208
Variable Data Types	209
Working with Expressions	211
C#-Like? Close, but Not Completely	211
The Expression Builder	213
Syntax Basics	214
Using Expressions in SSIS Packages	227
Summary	236
Chapter 7: Joining Data	237
<hr/>	
The Lookup Component	238
The Merge Join Component	239
Contrasting to the Relational Join	239
New Lookup Features	242
Building the Basic Package	242
Creating the Basic Package	244
Using a Relational Join in the Source	245
Using the Merge Join Component	247
Using the Lookup Component	252
Full-Cache Mode	252
No-Cache Mode	255
Partial-Cache Mode	257
Multiple Outputs	259
Expressionable Properties	264
Cascaded Lookup Operations	264
Cache Connection Manager and Transform	266
Summary	268
Chapter 8: Creating an End-to-End Package	271
<hr/>	
Basic Transformation Tutorial	271
Creating Connections	272
Creating the Tasks	274
Creating the Data Flow	275
Completing the Package	276
Saving the Package	277
Executing the Package	277

Contents

Typical Mainframe ETL with Data Scrubbing	278
Creating the Data Flow	280
Handling Dirty Data	280
Finalizing	284
Handling More Bad Data	285
Looping and the Dynamic Task	288
Looping	288
Making the Package Dynamic	289
Summary	291
Chapter 9: Scripting in SSIS	293
Scripting?	293
Getting Started in SSIS Scripting	295
Selecting the Scripting Language	296
Using the VSTA Scripting IDE	297
Example: Hello World	298
Adding Code and Classes	300
Using Managed Assemblies	302
Using the Script Task	306
Configuring the Script Task Editor	306
The Script Task Dts Object	307
Accessing Variables in the Script Task	308
Connecting to Data Sources in a Script Task	315
Raising an Event in a Script Task	322
Writing a Log Entry in a Script Task	329
Using the Script Component	330
Differences from a Script Task	330
Configuring the Script Component Editor	331
Accessing Variables in a Script Component	334
Connecting to Data Sources in a Script Component	335
Raising Events	335
Logging	336
Example: Data Validation	337
Essential Coding, Debugging, and Troubleshooting Techniques	347
Structured Exception Handling	347
Script Debugging and Troubleshooting	350
Summary	353

Chapter 10: Loading a Data Warehouse	355
Data Profiling	356
Initial Execution of the Data Profiling Task	356
Reviewing the Results of the Data Profiling Task	359
Turning Data Profile Results into Actionable ETL Steps	365
Data Extraction	366
Dimension Table Loading	366
Loading a Simple Dimension Table	367
Loading a Complex Dimension Table	375
Considerations and Alternates to the SCD Transformation	387
Fact Table Loading	388
SSAS Processing	400
Master ETL Package	408
Summary	411
Chapter 11: Using the Relational Engine	413
Data Extraction	414
SELECT * Is Bad	414
WHERE Is your Friend	416
Transform during Extract	417
Many ANDs Make Light Work	420
SORT in the Database	421
Modularize	423
SQL Server Does Text Files Too	425
Use Set-Based Logic	428
SQL Server 2008 Change Data Capture	430
Benefits of SQL Server 2008 CDC	431
Preparation	432
Capture Instance Tables	434
The CDC API	436
Using CDC from within SSIS	438
Data Loading	442
Database Snapshots	442
The MERGE Operator	444
Summary	448

Chapter 12: Accessing Heterogeneous Data	449
Excel and Access	450
Limited 64-Bit Support	450
Working with Excel Files	452
Access	452
Oracle	460
Oracle Client Setup	460
Importing Oracle Data	460
XML and Web Services	463
Configuring the Web Service Task	463
Working with XML Data as a Source	476
Flat Files	481
Loading Flat Files	481
Extracting Data from Flat Files	483
ODBC	486
Other Heterogeneous Sources	489
Summary	490
Chapter 13: Reliability and Scalability	491
Restarting Packages	491
Simple Control Flow	492
Containers within Containers and Checkpoints	496
Variations on a Theme	499
Inside the Checkpoint File	501
Package Transactions	502
Single Package, Single Transaction	503
Single Package, Multiple Transactions	506
Two Packages, One Transaction	508
Single Package Using a Native Transaction in SQL Server	509
Error Outputs	511
Scaling Out	514
Architectural Improvements	514
Scale Out Memory Pressures	515
Scale Out by Staging Data	515
Summary	520
Chapter 14: Understanding and Tuning the Data Flow Engine	521
The SSIS Engine	521
Understanding the SSIS Data Flow and Control Flow	522
Handling Workflows with the Control Flow	525

Data Processing in the Data Flow	526
Memory Buffer Architecture	527
Types of Transformations	529
Advanced Data Flow Execution Concepts	538
SSIS Data Flow Design and Tuning	548
Data Flow Design Practices	548
Optimizing Package Processing	555
Troubleshooting Data Flow Performance Bottlenecks	560
Pipeline Performance Monitoring	562
Summary	565
Chapter 15: Source Control and Software Development Life Cycle	567
Introduction to Software Development Life Cycles	568
Software Development Life Cycles: A Brief History	568
Types of Software Development Life Cycles	569
Versioning and Source Code Control	570
Microsoft Visual SourceSafe	570
Team Foundation Server, Team System, and SSIS	585
MSF Agile and SSIS	590
The Project Portal	594
Putting It to Work	594
Code Deployment and Promotion from Development to Test to Production	605
The Deployment Wizard	605
Import a Package	606
Summary	608
Chapter 16: DTS 2000 Migration	609
Managing DTS 2000 Packages within SQL Server Management Studio	609
Running DTS 2000 Packages under SSIS	610
Migrating DTS 2000 Packages to SSIS	611
Using the Package Migration Wizard	614
Third-Party Migration Solution	620
Summary	621
Chapter 17: Error and Event Handling	623
Precedence Constraint	623
Precedence Constraint Basics	624
Advanced Precedence Constraints and Expressions	625

Contents

Event Handling	633
Events	634
Inventory Example	635
Event Handler Inheritance	643
Breakpoints	645
Error Rows	648
Logging	653
Logging Providers	654
Log Events	654
Summary	659
Chapter 18: Programming and Extending SSIS	661
The Sample Components	662
Component 1: Source Adapter	662
Component 2: Transformation	663
Component 3: Destination Adapter	663
The Pipeline Component Methods	664
Design-Time Functionality	664
Runtime	668
Connection Time	670
Building the Components	670
Preparation	671
Building the Source Component	678
Building the Transform Component	690
Building the Destination Adapter	702
Using the Components	709
Installing the Components	710
Debugging Components	710
Design-Time	710
Building the Complete Package	712
Runtime Debugging	713
Upgrading to SQL 2008	716
Summary	716
Chapter 19: Adding a User Interface to Your Component	717
Three Key Steps	717
Building the User Interface	718
Adding the Project	719
Implementing IDtsComponentUI	722
Setting the UITypeName	725
Building the Form	727

Further Development	732
Runtime Connections	732
Component Properties	735
Handling Errors and Warnings	737
Column Properties	739
Other Considerations	740
Summary	740
Chapter 20: External Management and WMI Task Implementation	741
External Management of SSIS with Managed Code	741
Setting up a Test SSIS Package for Demonstration Purposes	742
The DTS Runtime Managed Code Library	743
Application Object Maintenance Operations	744
Package Maintenance Operations	745
Package Folder Maintenance	750
Package Role Maintenance	752
Package Monitoring	753
Package Listing	755
A Package Management Example	756
Package Log Providers	765
Specifying Events to Log	767
Programming to Log Providers	768
Package Configurations	771
Creating a Configuration	772
Programming the Configuration Object	773
Configuration Object	774
Windows Management Instrumentation Tasks	775
WMI Reader Task Explained	776
WMI Data Reader Example	777
WMI Event Watcher Task	782
WMI Event Watcher Task Example	783
Summary	786
Chapter 21: Using SSIS with External Applications	787
InfoPath Documents	788
ASP.NET Applications	797
Winform .NET Applications	802
Summary	808

Chapter 22: Administering SSIS	809
Package Configuration	809
Deployment Utility	814
Creating the Deployment Manifest	814
The Package Deployment Wizard	815
The Package Store	818
Creating a Central SSIS Server	820
Clustering SSIS	822
File System or the MSDB Deployment	824
Management Studio	825
Running Packages with DTEXECUI	826
Security	832
Command-Line Utilities	834
DTEXEC	834
DTUTIL	835
Scheduling a Package	836
Proxy Accounts	838
64-Bit Issues	839
Performance Counters	840
Summary	841
Chapter 23: Case Study: A Programmatic Example	843
What You Will Take Away	843
Background	844
Business Problem	845
Solution Summary	845
Solution Architecture	846
Naming Conventions and Tips	849
Additional SSIS Tips Before You Start a Large Project	850
Data Architecture	850
File Storage Location Setup	851
Bank ACH Payments	851
Lockbox Files	852
PayPal or Direct Credits to Corporate Account	853
Case Study Database Model	853
Database Setup	854
Case Study Load Packages	863
Bank File Load Package	864
ACH Load Package	894
Email Load Package	913
Testing	919

- Case Study Invoice Matching Process** **919**
- Matching Process Control Flow 920
- Matching Process High-Confidence Data Flow 923
- Matching Process Medium-Confidence Data Flow 928
- Interpreting the Results 932
- Creating a Parent Driver Package** **934**
- Driver Package Setup 934
- Driver Package Deployment 935
- Summary** **935**

- Index** **937**

