

# Pro SQL Server 2005 Integration Services



James Wightman

## **Pro SQL Server 2005 Integration Services**

**Copyright © 2008 by James Wightman**

All rights reserved. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without the prior written permission of the copyright owner and the publisher.

ISBN-13 (pbk): 978-1-59059-897-9

ISBN-10 (pbk): 1-59059-897-0

ISBN-13 (electronic): 978-1-4302-0447-3

ISBN-10 (electronic): 1-4302-0447-8

Printed and bound in the United States of America 9 8 7 6 5 4 3 2 1

Trademarked names may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, we use the names only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

Lead Editor: Ewan Buckingham

Technical Reviewer: Fabio Claudio Ferracchiati

Editorial Board: Steve Anglin, Ewan Buckingham, Tony Campbell, Gary Cornell, Jonathan Gennick,

Jason Gilmore, Kevin Goff, Jonathan Hassell, Matthew Moodie, Joseph Ottinger, Jeffrey Pepper,

Ben Renow-Clarke, Dominic Shakeshaft, Matt Wade, Tom Welsh

Senior Project Manager: Tracy Brown Collins

Copy Editor: Eileen Cohen

Assistant Production Director: Kari Brooks-Copony

Production Editor: Liz Berry

Compositor: Diana Van Winkle

Proofreader: Linda Seifert

Indexer: John Collin

Artist: Diana Van Winkle

Cover Designer: Kurt Krames

Manufacturing Director: Tom Debolski

Distributed to the book trade worldwide by Springer-Verlag New York, Inc., 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax 201-348-4505, e-mail [orders-ny@springer-sbm.com](mailto:orders-ny@springer-sbm.com), or visit <http://www.springeronline.com>.

For information on translations, please contact Apress directly at 2855 Telegraph Avenue, Suite 600, Berkeley, CA 94705. Phone 510-549-5930, fax 510-549-5939, e-mail [info@apress.com](mailto:info@apress.com), or visit <http://www.apress.com>.

The information in this book is distributed on an “as is” basis, without warranty. Although every precaution has been taken in the preparation of this work, neither the author(s) nor Apress shall have any liability to any person or entity with respect to any loss or damage caused or alleged to be caused directly or indirectly by the information contained in this work.

The source code for this book is available to readers at <http://www.apress.com>.

*Dedicated to my family: my wife Ellie and children Zakk and Layne.  
To Mollie, who is sadly no longer with us, and her brother Bill, for whom life has  
never been the same. Finally, to William Henry Gates III, for the opportunities.*



# Contents at a Glance

Foreword .....	xvii
About the Author .....	xix
About the Technical Reviewer .....	xxi
Acknowledgments .....	xxiii
Introduction .....	xxiv
■ <b>CHAPTER 1</b> Introducing SQL Server 2005 Integration Services .....	1
■ <b>CHAPTER 2</b> The Architecture of SSIS .....	13
■ <b>CHAPTER 3</b> From DTS to SSIS .....	29
■ <b>CHAPTER 4</b> The IDE and Extending Visual Studio .....	45
■ <b>CHAPTER 5</b> Package Hierarchy .....	73
■ <b>CHAPTER 6</b> Control Flow .....	91
■ <b>CHAPTER 7</b> Maintenance Plans .....	157
■ <b>CHAPTER 8</b> Data Flow Data Sources .....	171
■ <b>CHAPTER 9</b> Basic Data Flow Transformations .....	199
■ <b>CHAPTER 10</b> Advanced Data Flow Transformations .....	241
■ <b>CHAPTER 11</b> Data Flow Destinations .....	281
■ <b>CHAPTER 12</b> Optimization for SSIS .....	309
■ <b>CHAPTER 13</b> SSIS Security .....	337
■ <b>CHAPTER 14</b> Scripting, .NET Style .....	355
■ <b>CHAPTER 15</b> Custom Components .....	387
■ <b>APPENDIX A</b> Interview with Donald Farmer .....	459
■ <b>APPENDIX B</b> The Package Object .....	463
■ <b>APPENDIX C</b> The Control Flow Object .....	477
■ <b>APPENDIX D</b> Maintenance Plans Reference .....	509
■ <b>APPENDIX E</b> Data Flow Object Reference .....	519
■ <b>INDEX</b> .....	523



# Contents

Foreword .....	xvii
About the Author .....	xix
About the Technical Reviewer .....	xxi
Acknowledgments .....	xxiii
Introduction .....	xxiv
■ <b>CHAPTER 1</b> <b>Introducing SQL Server 2005 Integration Services</b> .....	1
A Little History .....	2
What Can SSIS Do? .....	3
The First Time Ever I Saw SSIS .....	6
I Have Such Delights to Show You .....	6
An Interview with Donald Farmer .....	7
Become a Knowledge Broker .....	10
Summary .....	12
■ <b>CHAPTER 2</b> <b>The Architecture of SSIS</b> .....	13
A Brief SQL Server Architecture Overview .....	13
SSIS and SQL Server Analysis Services .....	14
Common Language Runtime Integration .....	15
SSIS Architecture Overview .....	15
The Power of the Data Flow Architecture .....	21
A Brief Word on Terminology .....	22
Data Flow Transformations: Buffer-Usage Strategies .....	22
Execution Trees Explained .....	24
Summary .....	28
■ <b>CHAPTER 3</b> <b>From DTS to SSIS</b> .....	29
What's New? .....	29
Preparing to Migrate Legacy DTS Packages .....	30
Running SQL Server 2005 Upgrade Advisor .....	31
Understanding Upgrade Advisor Warnings .....	33
A Complete List of Possible Upgrade Advisor Messages .....	36

Using the Package Migration Wizard .....	39
Summary .....	43
 <b>CHAPTER 4 The IDE and Extending Visual Studio .....</b>	<b>45</b>
A First Look at BIDS .....	45
Projects and Solutions .....	46
Solution Explorer .....	47
Creating and Saving a Project .....	48
Populating a Solution .....	49
Project Dependencies and Build Order .....	50
Properties Explorer .....	51
Compilation, Running, and Debugging .....	53
References .....	54
Using a Reference in Code .....	58
The GAC .....	58
The Toolbox .....	60
Adding and Removing Toolbox Components .....	60
Class Explorer .....	61
Project Templates .....	62
The SSIS Menu .....	63
Package Explorer .....	69
Execution Results .....	70
Summary .....	71
 <b>CHAPTER 5 Package Hierarchy .....</b>	<b>73</b>
Vocabulary .....	74
Containers .....	74
Packages .....	75
Package Properties .....	75
Package Methods .....	75
Control Flow .....	75
Data Flow .....	75
Checkpoints .....	75
Variables .....	76
Variable Scope .....	77
Event Handlers .....	78
Adding Components .....	82
Using Executable Designers .....	82
Precedence Constraints .....	82



Flow Between Executables .....	84
Directing Errors .....	88
Summary .....	90

<b>CHAPTER 6 Control Flow .....</b>	<b>91</b>
Where Does the Control Flow Fit? .....	92
A Note About Using Variables .....	92
Transactions .....	92
Common Control Flow Properties .....	93
Control Flow Tasks .....	93
For Loop Container .....	93
Properties .....	95
Foreach Loop Container .....	95
Foreach ADO Enumerator .....	96
Foreach ADO.NET Schema Rowset Enumerator .....	106
Foreach File Enumerator .....	108
Foreach From Variable Enumerator .....	109
Foreach Item Enumerator .....	110
Foreach Nodelist Enumerator .....	110
Foreach SMO Enumerator .....	112
Sequence .....	112
ActiveX Script Task .....	113
Analysis Services Execute DDL Task .....	113
DDL Configuration .....	114
Analysis Services Processing Task .....	115
Bulk Insert Task .....	116
Data Mining Query Task .....	120
Execute DTS 2000 Package Task .....	121
Execute Package Task .....	122
Execute Process Task .....	123
Execute SQL Task .....	126
File System Task .....	128
FTP Task .....	129
Message Queue Task .....	130
Script Task .....	131
Send Mail Task .....	133
Transfer Database Task .....	134
Transfer Error Messages Task .....	134
Transfer Jobs Task .....	135
Transfer Logins Task .....	136

	Transfer Master Stored Procedures Task .....	136
	Transfer SQL Server Objects Task .....	137
	Web Service Task .....	138
	WMI Data Reader Task .....	146
	WMI Event Watcher Task .....	146
	XML Task .....	146
	Summary .....	155
<b>CHAPTER 7</b>	<b>Maintenance Plans .....</b>	<b>157</b>
	Back Up Database .....	157
	Check Database Integrity .....	160
	Execute SQL Server Agent Job .....	161
	Execute T-SQL Statement .....	162
	History Cleanup .....	162
	Maintenance Cleanup .....	164
	Notify Operator .....	165
	Rebuild Index .....	166
	Reorganize Index .....	167
	Shrink Database .....	168
	Update Statistics .....	168
	Summary .....	169
<b>CHAPTER 8</b>	<b>Data Flow Data Sources .....</b>	<b>171</b>
	DataReader Source .....	171
	Excel Source .....	178
	Flat File Source .....	182
	OleDb Source .....	191
	Raw File Source .....	194
	XML Source .....	195
	Summary .....	198
<b>CHAPTER 9</b>	<b>Basic Data Flow Transformations .....</b>	<b>199</b>
	SSIS Data Types .....	199
	Data Viewers .....	200
	Audit .....	204
	Character Map .....	204
	Conditional Split .....	205
	Copy Column .....	212
	Data Conversion .....	214

Derived Column .....	215
Export Column .....	218
Import Column .....	220
Lookup .....	221
Merge .....	225
Merge Join .....	232
Multicast .....	234
OLE DB Command .....	235
Percentage Sampling .....	236
Row Count .....	237
Row Sampling .....	238
Sort .....	238
Union All .....	239
Summary .....	240
 <b>CHAPTER 10    Advanced Data Flow Transformations</b> .....	 241
Aggregate .....	242
Data Mining Query .....	245
Fuzzy Grouping .....	245
Fuzzy Lookup .....	248
Pivot .....	254
Script Component .....	258
Slowly Changing Dimension .....	263
Term Extraction .....	268
How It Works .....	272
Term Lookup .....	273
Unpivot .....	276
Summary .....	280
 <b>CHAPTER 11    Data Flow Destinations</b> .....	 281
Data Mining Model Training .....	281
DataReader Destination .....	283
Dimension Processing .....	284
Excel Destination .....	287
Flat File Destination .....	291
OLE DB Destination .....	292
Partition Processing .....	296
Raw File Destination .....	300
Recordset Destination .....	302

SQL Server Destination .....	303
SQL Server Mobile Destination .....	306
Summary .....	308
<b>CHAPTER 12 Optimization for SSIS .....</b>	<b>309</b>
Decisions: The Basics .....	309
Real-time Processing vs. Preprocessing .....	310
Normalization and Denormalization .....	310
SQL Server Partitioning .....	311
Identifying Performance Issues .....	313
Performance Metrics .....	313
The Hacker's Toolkit .....	317
SQL Server Tweaks .....	330
Platform Tweaks .....	330
SSIS-Specific Optimizations .....	331
Component and Component-Query Tweaks .....	332
Bulk Inserts/Text Files .....	333
A Note on Buffers and Memory Usage .....	333
Data Flow Parallelization .....	334
Summary .....	335
<b>CHAPTER 13 SSIS Security .....</b>	<b>337</b>
SSIS Security Options .....	338
Digitally Signing Packages (Or, Do You Need to Know If a Package Has Been Changed?) .....	338
Package Encryption (Or, Does a Package or a Part of the Package Need to Be Encrypted?) .....	340
Package Access (Or, Who Should Access Your Packages?) .....	343
Custom Security Options .....	351
Custom Launcher .....	351
Jetico BestCrypt .....	351
Summary .....	353
<b>CHAPTER 14 Scripting, .NET Style .....</b>	<b>355</b>
Development Environments .....	355
OOP .....	356
The Benefits of OOP .....	357
OOP 101 .....	357
Further Reading .....	362

Design Patterns .....	362
Further Reading .....	363
.NET Concepts .....	363
Everything You Need to Know About .NET but Were Afraid to Ask ..	364
Classes in the Script Component .....	365
Script Component Crash Course .....	369
Preparation .....	369
Building the Data Flow .....	371
Summary .....	385
 <b>■ CHAPTER 15 Custom Components .....</b>	<b>387</b>
What Flavor Would You Like? .....	388
Connection Managers .....	388
Data Flow Components (Source, Transform, or Destination) .....	388
Foreach Enumerators .....	388
Control Flow Tasks .....	389
Log Providers .....	389
What Makes a Custom Component? .....	389
SSIS Assemblies Supporting Components .....	390
Class Attributes .....	390
Deriving from and Overriding SSIS Classes .....	392
Custom Control Flow Tasks .....	392
Custom Data Flow components .....	395
Putting It All Together .....	399
The Spellcheck Example, Part II .....	399
The Class Library .....	400
The Custom Data Flow Transform Component .....	411
The SSIS Data Flow .....	445
Designers .....	452
Control Flow UIs .....	452
Data Flow Component UIs .....	452
Summary .....	456
 <b>■ APPENDIX A Interview with Donald Farmer .....</b>	<b>459</b>
 <b>■ APPENDIX B The Package Object .....</b>	<b>463</b>
Package Properties .....	463
Package Methods .....	473

<b>■ APPENDIX C    The Control Flow Object</b>	477
Control Flow Common Properties	477
Control Flow Task Properties	480
Analysis Services Processing Task	480
Bulk Insert Task	482
Data Mining Query Task	484
Execute Package Task	485
Execute Process Task	485
Execute SQL Task	486
File System Task	487
FTP Task	488
Message Queue Task	490
Script Task	491
Send Mail Task	492
Transfer Database Task	493
Transfer Error Messages	494
Transfer Jobs Task	495
Transfer Logins Task	496
Transfer Master Stored Procedures Task	496
Transfer SQL Server Objects Task	497
Web Service Task	501
WMI Data Reader Task	502
WMI Event Watcher Task	503
XML Task	504
<b>■ APPENDIX D    Maintenance Plans Reference</b>	509
Maintenance-Plan Tasks Common Properties	509
Properties That Inherit from DatabaseMaintenanceBaseTask	509
Properties That Inherit from ExecuteSQLTask	511
Properties Whose Implementations Are Enforced by the	
IContextSelection Interface	512
Maintenance-Plan Tasks	513
Backup Database	513
Check Database Integrity	515
Execute SQL Server Agent Job	515
History Cleanup Task	515
Maintenance Cleanup Task	516
Notify Operator Task	517
Rebuild Index Task	517

Reorganize Index Task .....	518
Shrink Database Task .....	518
Execute T-SQL Task .....	518
Update Statistics Task .....	518
<b>■ APPENDIX E Data Flow Object Reference .....</b>	<b>519</b>
<b>■ INDEX .....</b>	<b>523</b>