

Microsoft®

Microsoft® SQL Server® 2012 Integration Services

Wee-Hyong Tok
Rakesh Parida
Matt Masson
Xiaoning Ding
Karthik Sivashanmugam

Published with the authorization of Microsoft Corporation by:
O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, California 95472

Copyright © 2012 by Wee-Hyong Tok, Rakesh Parida, Matt Masson, Xiaoning Ding, Kaarthik Sivashanmugam.
All rights reserved. No part of the contents of this book may be reproduced or transmitted in any form or by any means without the written permission of the publisher.

ISBN: 978-0-7356-6585-9

2 3 4 5 6 7 8 9 10 LSI 7 6 5 4 3 2

Printed and bound in the United States of America.

Microsoft Press books are available through booksellers and distributors worldwide. If you need support related to this book, email Microsoft Press Book Support at mspinput@microsoft.com. Please tell us what you think of this book at <http://www.microsoft.com/learning/booksurvey>.

Microsoft and the trademarks listed at <http://www.microsoft.com/about/legal/en/us/IntellectualProperty/Trademarks/EN-US.aspx> are trademarks of the Microsoft group of companies. All other marks are property of their respective owners.

The example companies, organizations, products, domain names, email addresses, logos, people, places, and events depicted herein are fictitious. No association with any real company, organization, product, domain name, email address, logo, person, place, or event is intended or should be inferred.

This book expresses the authors' views and opinions. The information contained in this book is provided without any express, statutory, or implied warranties. Neither the authors, O'Reilly Media, Inc., Microsoft Corporation, nor its resellers, or distributors will be held liable for any damages caused or alleged to be caused either directly or indirectly by this book.

Acquisitions and Developmental Editor: Russell Jones

Production Editor: Melanie Yarbrough

Editorial Production: Stan Info Solutions

Technical Reviewer: boB Taylor

Copyeditor: Teresa Horton

Indexer: WordCo Indexing Services, Inc.

Cover Design: Twist Creative • Seattle

Cover Composition: ContentWorks, Inc.

Illustrator: Rebecca Demarest

[2012-10-19]

Dedicated to my wife, Juliet, and son, Nathaniel, for their love, support, and patience. And to my parents, Siak-Eng and Hwee-Tiang for shaping me into who I am today.

—WEE-HYONG TOK

I would like to dedicate this to my parents, Basanta and Sarmistha, and my soon-to-be-wife, Vijaya, for all their support and encouragement for making this happen.

—RAKESH PARIDA

Dedicated to my family and loving wife, whose patience and support made all this possible.

—MATT MASSON

I would like to dedicate this book to my parents, ChengXian and Dezhen, my wife Yanan, my cute daughter Ruiyi, and my lovely son Ruichen. I love all of you so much.

—XIAONING DING

I dedicate this book to my wife, Devi, and my son, Raghav, for their love and support.

—KAARTHIK SIVASHANMUGAM



Contents at a Glance

<i>Foreword</i>	<i>xxi</i>
<i>Introduction</i>	<i>xxiii</i>

PART I OVERVIEW

CHAPTER 1	SSIS Overview	3
CHAPTER 2	Understanding SSIS Concepts	27
CHAPTER 3	Upgrading to SSIS 2012	49

PART II DEVELOPMENT

CHAPTER 4	New SSIS Designer Features	83
CHAPTER 5	Team Development	115
CHAPTER 6	Developing an SSIS Solution	131
CHAPTER 7	Understanding SSIS Connectivity	161
CHAPTER 8	Working with Change Data Capture in SSIS 2012	195
CHAPTER 9	Data Cleansing Using SSIS	241

PART III DATABASE ADMIN

CHAPTER 10	Configuration in SSIS	265
CHAPTER 11	Running SSIS Packages	301
CHAPTER 12	SSIS T-SQL Magic	331
CHAPTER 13	SSIS PowerShell Magic	353
CHAPTER 14	SSIS Reports	365

PART IV DEEP-DIVE

CHAPTER 15	SSIS Engine Deep Dive	385
CHAPTER 16	SSIS Catalog Deep Dive	415
CHAPTER 17	SSIS Security	435
CHAPTER 18	Understanding SSIS Logging	463
CHAPTER 19	Automating SSIS	483

PART V	TROUBLESHOOTING	
CHAPTER 20	Troubleshooting SSIS Package Failures	517
CHAPTER 21	SSIS Performance Best Practices	539
CHAPTER 22	Troubleshooting SSIS Performance Issues	569
CHAPTER 23	Troubleshooting Data Issues	589
	<i>Index</i>	605
	<i>About the Authors</i>	637

Contents

<i>Foreword</i>	<i>xxi</i>
<i>Introduction</i>	<i>xxiii</i>

PART I OVERVIEW

Chapter 1 SSIS Overview	3
Common Usage Scenarios for SSIS	4
Consolidation of Data from Heterogeneous Data Sources	4
Movement of Data Between Systems.....	9
Loading a Data Warehouse	12
Cleaning, Formatting, or Standardization of Data	16
Identification, Capture, and Processing of Data Changes.....	17
Coordination of Data Maintenance, Processing, or Analysis	18
Evolution of SSIS	20
Setting Up SSIS	21
SQL Server Features Needed for Data Integration.....	22
SQL Server Editions and Integration Services Features.....	24
Summary.....	25
 Chapter 2 Understanding SSIS Concepts	 27
Control Flow.....	28
Tasks	28
Precedence Constraints.....	30
Variables and Expressions.....	31
Containers.....	32
Connection Managers.....	35

What do you think of this book? We want to hear from you!

Microsoft is interested in hearing your feedback so we can continually improve our books and learning resources for you. To participate in a brief online survey, please visit:

microsoft.com/learning/booksurvey

Packages and Projects	36
Parameters	37
Log Providers	38
Event Handlers	40
Data Flow	41
Source Adapters	41
Destination Adapters	42
Transforms	43
SSIS Catalog	44
Overview	45
Catalog	46
Folders	46
Environments	46
References	47
Summary	47

Chapter 3 Upgrading to SSIS 2012 49

What's New in SSIS 2012	49
Upgrade Considerations and Planning	50
Feature Changes in SSIS	50
Dependencies and Tools	52
Upgrade Requirements	52
Upgrade Scenarios	53
Unsupported Upgrade Scenarios	54
Upgrade Validation	55
Integration Services Upgrade	55
Upgrade Advisor	55
Performing Upgrade	61
Addressing Upgrade Issues and Manual Upgrade Steps	69
Conversion to Projects after Upgrade	71
Summary	79

Chapter 4 New SSIS Designer Features 83

- The Integration Services Designer 83
 - Visual Studio 83
 - Undo and Redo 84
 - Getting Started Window 85
 - Toolbox 85
 - Variables Window 87
 - Zoom Control 88
 - Autosave and Recovery 89
 - Status Icons 89
 - Annotations 90
- Configuration and Deployment 90
 - Solution Explorer Changes 90
 - Parameter Tab 92
 - Visual Studio Configurations 92
 - Project Compilation 93
 - Deployment Wizard 94
 - Project Conversion Wizard 95
 - Import Project Wizard 96
- New Tasks and Data Flow Components 96
 - Change Data Capture 96
 - Expression Task 99
 - DQS Cleansing Transform 100
 - ODBC Source and Destination 100
- Control Flow 100
 - Expression Adorners 100
 - Connection Managers 101
 - Execute SQL Task 101

Data Flow	102
Connection Assistants	102
Improved Column Mapping	103
Editing Components in an Error State	104
Grouping	104
Simplified Data Viewers	105
Row Count and Pivot Transform User Interfaces	105
Flat File Source Changes	106
Scripting	108
Visual Studio Tools for Applications	108
Script Component Debugging	109
.NET 4 Framework Support	111
Expressions	112
Removal of the Character Limit	112
New Expression Functions	112
Summary	113

Chapter 5 Team Development 115

Improvements in SQL Server 2012	115
Package Format Changes	115
Visual Studio Configurations	116
Using Source Control Management with SSIS	117
Connecting to Team Foundation Server	117
Adding an SSIS Project to Team Foundation Server	120
Change Management	124
Changes to the SSIS Visual Studio Project File	127
Best Practices	129
Using Small, Simple Packages	129
One Developer Per Package	129
Consistent Naming Conventions	130
Summary	130

Chapter 6 Developing an SSIS Solution 131

SSIS Project Deployment Models131
 Package Deployment Model131
 Project Deployment Model133
Develop an Integration Services Project.....136
 Creating an SSIS Project136
 Designing an Integration Services Data Flow147
 Using Parameters and the *ForEach* Container.....152
 Using the Execute Package Task156
 Building and Deploying an Integration Services Project159
Summary.....160

Chapter 7 Understanding SSIS Connectivity 161

Previous Connectivity Options in SSIS.....161
 Providers for Connectivity Technology162
 OLE DB, ADO.NET, and ODBC164
New Connectivity Options in SSIS 2012165
 Introducing ODBC166
 ODBC Components for SSIS.....168
 ODBC Source169
 ODBC Destination174
Connectivity Considerations for SSIS.....177
 64-Bit and SSIS177
 SSIS Tools on 64-Bit Architecture178
Connectivity to Other Sources and Destinations184
 Connecting to Microsoft Excel and Access184
 Connecting to Oracle186
 Creating Custom Components189
 Using Script Components.....191
Summary.....193

Chapter 8	Working with Change Data Capture in SSIS 2012	195
	CDC in SQL Server.....	195
	Using CDC in SQL Server	196
	CDC Scenarios in ETLs.....	197
	Stages in CDC	198
	CDC in SSIS 2012.....	202
	CDC State.....	202
	CDC Control Task.....	205
	Data Flow Component: CDC Source	211
	CDC Splitter Component	215
	CDC for Oracle.....	217
	Introduction	217
	Components for Creating CDC for Oracle.....	219
	CDC Service Configuration MMC	219
	Oracle CDC Designer MMC	221
	MSXDBCDC Database	233
	Oracle CDC Service Executable (xdbcscsvc.exe).....	235
	Data Type Handling	238
	SSIS CDC Components	240
	Summary.....	240
Chapter 9	Data Cleansing Using SSIS	241
	Data Profiling Task	241
	Fuzzy Lookup Transformation	246
	Fuzzy Grouping Transformation.....	251
	Data Quality Services Cleansing Transform	254
	Summary.....	261

Chapter 10 Configuration in SSIS 265

- Configuration Basics266
 - How Configurations Are Applied266
 - What to Configure266
 - Changes in SSIS 2012267
- Configuration in SSIS 2012267
 - Parameters268
 - Creating Package Parameters268
 - Creating Project Parameters271
 - API for Creating Parameters273
 - Using Parameters274
 - Configuring Parameters on the SSIS Catalog281
 - Configuring, Validating, and Executing Packages and Projects . . .281
 - Configuration Through SSMS281
 - Configuration Using SQL Agent, DTEXec, and T-SQL285
 - SSIS Environments287
 - Evaluation Order of Parameters291
- Package Deployment Model and Backward Compatibility291
 - Package Deployment Model292
- Best Practices for Configuring SSIS295
 - Best Practices with Package Deployment Model295
 - Best Practices with Project Deployment Model299
- Summary300

Chapter 11 Running SSIS Packages 301

- Ways to Run SSIS Packages301
 - Package Locations303
 - Configuring Packages306
 - Error Dumps308
 - Logging Options309

Running Packages in the SSIS Catalog	311
Prepare Executions.	312
Starting SSIS Package Executions	314
View Executions	318
Executions with T-SQL.	318
Running Packages from SQL Agent.	320
Create an SSIS Job Step.	320
Execute Packages from the SSIS Catalog.	322
Running Packages via PowerShell	325
Creating and Running SSIS Packages Programmatically	325
Summary.	330

Chapter 12 SSIS T-SQL Magic 331

Overview of SSIS Stored Procedures and Views	331
Integration Services Catalog.	332
SSIS Catalog Properties.	332
Querying the SSIS Catalog Properties	333
Setting SSIS Catalog Properties.	333
SSIS Projects and Packages	334
Deploy an SSIS Project to the SSIS Catalog.	334
Learning About the SSIS Projects Deployed to the SSIS Catalog .	335
Configuring SSIS Projects	336
Managing SSIS Projects in the SSIS Catalog	339
Running SSIS Packages in the SSIS Catalog.	341
SSIS Environments.	345
Creating SSIS Environments	346
Creating SSIS Environment Variables	346
Configuring SSIS Projects Using SSIS Environments	347
Configuring SSIS Projects Using Reference Values.	348
Package Execution Using SSIS Environments	349
Managing SSIS Environment and Environment Variables.	349
Summary.	351

Chapter 13 SSIS PowerShell Magic 353

PowerShell Refresher353

PowerShell and SQL Server354

Managing SSIS with PowerShell357

 SSIS Management Object Model357

 PowerShell with SSIS Management Object Model358

 PowerShell and SSIS Using T-SQL362

Advantages of Using PowerShell with SSIS364

Summary364

Chapter 14 SSIS Reports 365

Getting Started with SSIS Reports365

Data Preparation367

Monitoring SSIS Package Execution368

 Integration Services Dashboard368

 All Executions Report370

 All Validations and All Operations Reports371

Using SSIS Reports to Troubleshoot SSIS Package Execution373

Using the Execution Performance Report to Identify
 Performance Trends378

Summary381

PART IV DEEP-DIVE

Chapter 15 SSIS Engine Deep Dive 385

The Control Flow Engine385

 Overview385

 Load386

 Apply Parameters388

What do you think of this book? We want to hear from you!

Microsoft is interested in hearing your feedback so we can continually improve our books and learning resources for you. To participate in a brief online survey, please visit:

microsoft.com/learning/booksurvey

Validate	388
Execute.....	390
The Data Flow Engine.....	397
Overview	398
Execution Control.....	401
Backpressure.....	408
Engine Tuning.....	411
Summary.....	414

Chapter 16 SSIS Catalog Deep Dive 415

SSIS Catalog Deep Dive.....	415
Creating the SSIS Catalog.....	415
Unit of Deployment to the SSIS Catalog	417
What Is Inside SSISDB?	418
SQL Server Instance Starts Up.....	420
SSIS Catalog and Logging Levels.....	422
Understanding the SSIS Package Execution Life Cycle.....	423
Stopping SSIS Package Executions	426
Using the Windows Application Event Log.....	426
SSIS Catalog Maintenance and SQL Server Agent Jobs.....	427
Backup and Restore of the SSIS Catalog.....	430
Back Up SSISDB	431
Restore SSISDB	432
Summary.....	434

Chapter 17 SSIS Security 435

Protect Your Package.....	435
Control Package Access	435
Package Encryption.....	439
Sensitive Variables and Parameters.....	441
Package Signing.....	442

Security in the SSIS Catalog	443
Security Overview	444
Manage Permissions	446
DDL Trigger.....	453
Running SSIS with SQL Agent	454
Requirements	454
Create Credentials	454
Create Proxy Accounts	456
Create SQL Agent Jobs	459
Summary.....	461

Chapter 18 Understanding SSIS Logging 463

Configure Logging Options	463
Choose Containers.....	464
Select Events.....	466
Add Log Providers	468
Log Providers.....	471
Text Files	471
SQL Server	471
SQL Server Profiler.....	472
Windows Event Log	472
XML Files	473
Logging in the SSIS Catalog	474
Logging Levels	474
Event Logs	476
Event Context Information.....	477
Advanced Logging Topics.....	478
Customizing Logging Fields.....	478
Logging with dtexec Utility	479
Developing Custom Log Providers.....	479
Summary.....	481

Chapter 19 Automating SSIS	483
Introduction to SSIS Automation	483
Programmatic Generation of SSIS Packages	483
Metadata-Driven Package Execution	484
Dynamic Package Generation	485
Handling Design-Time Events	486
Samples	488
Metadata-Based Execution	497
Custom Package Runner	498
Using PowerShell with the SSIS Management Object Model	502
Using PowerShell with SQL Agent	505
Alternative Solutions and Samples	508
Samples on Codeplex	508
Third-Party Solutions	509
Summary	513

PART V TROUBLESHOOTING

Chapter 20 Troubleshooting SSIS Package Failures	517
Getting Started with Troubleshooting	517
Data Preparation	519
Troubleshooting Failures of SSIS Package Executions	520
Three Key Steps Toward Troubleshooting Failures of SSIS Package Executions	522
Execution Path	526
Finding the Root Cause of Failure	526
Troubleshooting the Execute Package Task and Child Package Executions	529
DiagnosticEx Events	531
Execute Package Task and Execution Path	532
Troubleshooting SSIS Package Execution Failures Scheduled with SQL Agent	534

Using CallerInfo to Determine SSIS Package Executions That Are Executed by SQL Agent	537
Using SQL Agent History Tables to Determine the SSIS Job Steps That Failed	537
Summary	538

Chapter 21 SSIS Performance Best Practices 539

Creating a Performance Strategy	540
OVAL Technique	540
Measuring SSIS Performance	542
Measuring System Performance	542
Measuring Performance of Data Flow Tasks	546
Designing for Performance	552
Parallelize Your Design	552
Using SQL Server Optimization Techniques	556
Bulk Loading Your Data	558
Keeping SSIS Operations in Memory	561
Optimizing SSIS Lookup Caching	562
Optimizing SSIS Infrastructure	566
Summary	568

Chapter 22 Troubleshooting SSIS Performance Issues 569

Performance Profiling	569
Troubleshooting Performance Issues	570
Data Preparation	571
Understanding SSIS Package Execution Performance	572
SSIS Package Execution Duration	572
Time Spent at Each Task in the SSIS Package	573
Time Spent at Each Phase of the Data Flow Component	573
Elapsed Time for Data Flow Component Phases (Active Time vs. Total Time)	574
Monitoring SSIS Package Execution Performance	576

Per-Execution Performance Counters	578
Interactive Analysis of Performance Data	579
Summary	588

Chapter 23 Troubleshooting Data Issues 589

Troubleshooting in the Design Environment	589
Row Count Values	589
Data Viewers	590
Data in Error Output	592
Breakpoints and Debug Windows	593
Troubleshooting in the Execution Environment	593
Execution Data Statistics	593
Data Tap	596
Error Dumps	600
Summary	603
<i>Index</i>	605
<i>About the Authors</i>	637

What do you think of this book? We want to hear from you!

Microsoft is interested in hearing your feedback so we can continually improve our books and learning resources for you. To participate in a brief online survey, please visit:

microsoft.com/learning/booksurvey

Foreword

In 1989, when we were all much younger, I had a bizarre weekend job: During the week, I was an engineer at Microrim Incorporated, the makers of R:Base—the second most popular desktop database in the world. But on Saturday mornings I would sit completely alone in our headquarters building in Redmond and rebuild the database that ran our call center. This involved getting the latest registered licenses from accounting, the up-to-date employee list from human resources, the spreadsheets from marketing that tracked our independent software vendors, and of course all of the previous phone call history from the log files, and then mashing it all together. Of course none of these systems had consistent formats or numbering schemes or storage. It took me six hours—unless I messed up a step. The process was all scripted out on a sheet of paper. There wasn't a name for it at the time, but I was building a data warehouse.

Anyone who's done this work knows in their heart the message we hear again and again from customers: Getting the right data into the right shape and to the right place at the right time is 80 percent of the effort for any data project. Data integration is the behind-the-wall plumbing that makes a beautiful fountain work flawlessly. Often the fountains get all the attention, but on the SSIS team at Microsoft, we are proud to build that plumbing.

The authors of this book are at the core of that proud team. For as long as I have known him, Kaarthik has been an ardent advocate for this simple truth: You can understand the quality of a product only if you first deeply understand the customers that use it. As the first employee for SSIS in China, Xiaoning blazed a trail. He is one of those quiet geniuses, who, when he speaks, everyone stops to listen to, because what he says will be deep and important. One of my best professional decisions was overriding my manager's advice to hire Matt. You see, he didn't quite fit our mold. Yes, he could write code well, but there was something that just didn't match our expectations. He cared way too deeply about the real world and about building end-to-end solutions to solve business problems; he wouldn't stop talking about it! We made it work. Don't tell Wee Hyong I said this, but he is probably overqualified for his job. His background as a lecturer in academia, and his history as a SQL MVP (our most valuable partners) made him a perfect candidate to be one of the public faces of SSIS. And finally, Rakesh. At the end of his first week on the job, he decided to create a community event for our customers attending a trade show that just happened to be running nearby. He coerced his teammates into helping, found a room at the convention center, and sent out the invite to our customers. In all the authors, you can see a passion for customers and solutions. You are in great hands.