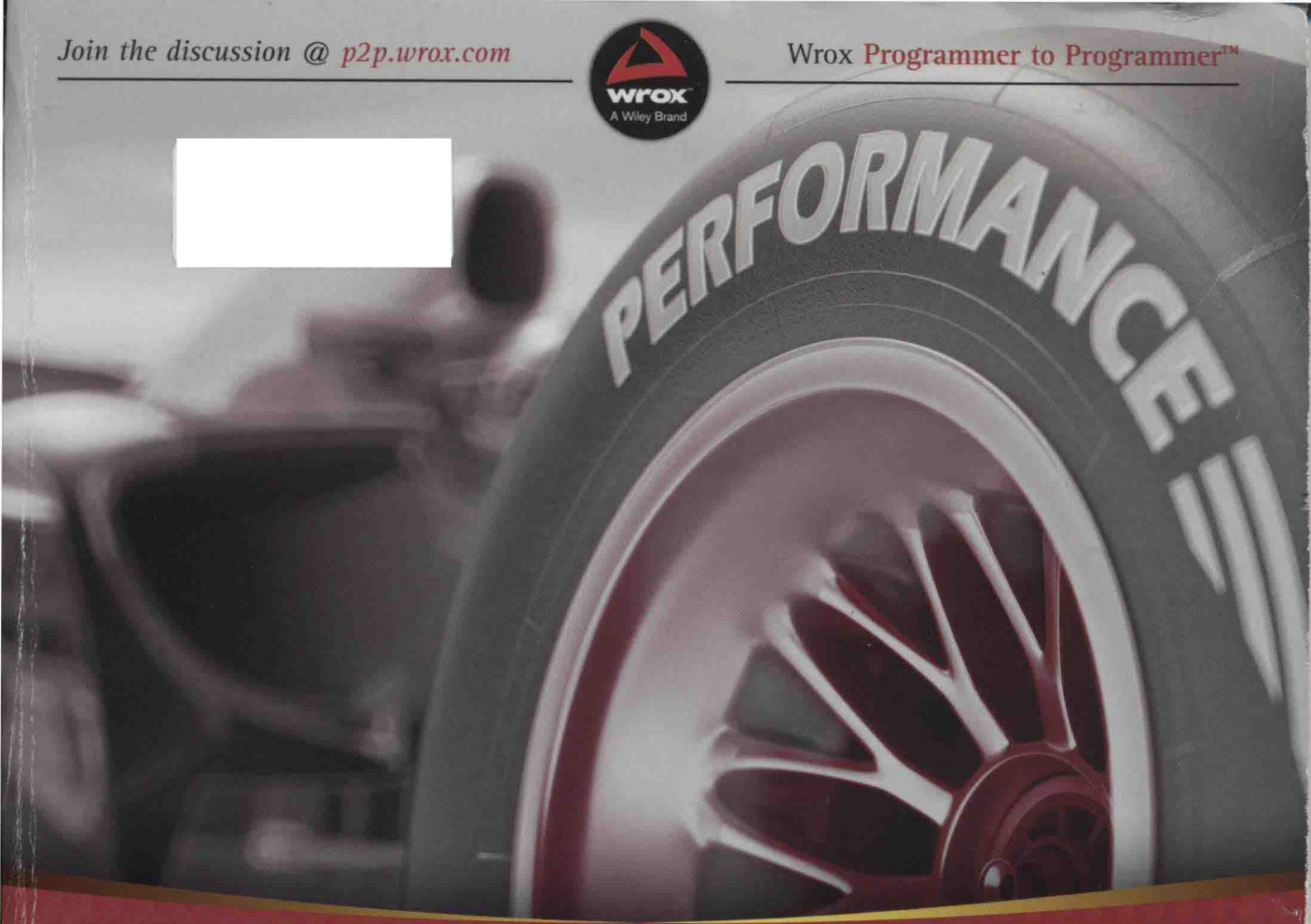
Join the discussion @ p2p.wrox.com

Wrox Programmer to Programmer™



Professional Microsoft® SQL Server® 2014 Integration Services

Brian Knight, Devin Knight, Jessica M. Moss, Mike Davis, Chris Rock

PROFESSIONAL

Microsoft® SQL Server® 2014 Integration Services

Brian Knight
Devin Knight
Jessica M. Moss
Mike Davis
Chris Rock

常州大学山书馆
藏书章

wrox™
A Wiley Brand

Professional Microsoft® SQL Server® 2014 Integration Services

Published by
John Wiley & Sons, Inc.
10475 Crosspoint Boulevard
Indianapolis, IN 46256
www.wiley.com

Copyright © 2014 by John Wiley & Sons, Inc., Indianapolis, Indiana

Published simultaneously in Canada

ISBN: 978-1-118-85087-9
ISBN: 978-1-118-85090-9 (ebk)
ISBN: 978-1-118-85085-5 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permissions.

Limit of Liability/Disclaimer of Warranty: The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation warranties of fitness for a particular purpose. No warranty may be created or extended by sales or promotional materials. The advice and strategies contained herein may not be suitable for every situation. This work is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If professional assistance is required, the services of a competent professional person should be sought. Neither the publisher nor the author shall be liable for damages arising herefrom. The fact that an organization or Web site is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or Web site may provide or recommendations it may make. Further, readers should be aware that Internet Web sites listed in this work may have changed or disappeared between when this work was written and when it is read.

For general information on our other products and services please contact our Customer Care Department within the United States at (877) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at http://booksupport.wiley.com. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2014930406

Trademarks: Wiley, Wrox, the Wrox logo, Programmer to Programmer, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates, in the United States and other countries, and may not be used without written permission. Microsoft and SQL Server are registered trademarks of Microsoft Corporation. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

PROFESSIONAL

Microsoft® SQL Server® 2014 Integration Services

此为试读,需要完整PDF请访问: www.ertongbook.com

To my great team and work family at Pragmatic Works

—Brian Knight

Proverbs 3:5-6

—Devin Knight

For Rich: Thank you for your support.

—Jessica M. Moss

To my kids, Gabriel and Sydney — they are the light of my life.

—Mike Davis

To the three girls in my life that mean more than anything to me, Tammy, Calista, and Callie

—Chris Rock

ABOUT THE AUTHORS

BRIAN KNIGHT, SQL Server MVP, MCITP, MCSE, MCDBA, is the owner and founder of Pragmatic Works. He is also the cofounder of BIDN.com, SQLServerCentral.com, and SQLShare.com. He runs the local SQL Server users group in Jacksonville (JSSUG). Brian is a contributing columnist at several technical magazines. He is the author of a dozen SQL Server books. Brian has spoken at conferences like PASS, SQL Connections, and TechEd, SQL Saturdays, Code Camps, and many pyramid scheme motivational sessions. His blog can be found at `http://www.bidn.com`. Brian lives in Jacksonville, Florida, where he enjoys his kids and running marathons.

DEVIN KNIGHT, SQL Server MVP, is the Training Director at Pragmatic Works. This is the sixth SQL Server book that he has authored. Devin has spoken at past conferences like PASS Summit, PASS Business Analytics Conference, SQL Saturdays, and Code Camps and is a contributing member to the PASS Business Intelligence Virtual Chapter. Making his home in Jacksonville, Florida, Devin is the Vice President of the local users' group (JSSUG). His blog can be found at `http://www.devinknightsql.com`.

JESSICA M. MOSS is a well-known practitioner, author, and speaker of Microsoft SQL Server business intelligence and has received a Microsoft MVP award for the past 6 years. She has created numerous data warehouse and business intelligence solutions for companies in different industries and has delivered training courses on Integration Services, Reporting Services, and Analysis Services. Jessica has authored technical content for multiple magazines, websites, and books, including the Wrox book *Professional Microsoft SQL Server 2012 Integration Services*, and has spoken internationally at conferences such as the PASS Community Summit, SharePoint Connections, and the SQLTeach International Conference. As a strong proponent of developing user-to-user community relations, Jessica actively participates in local user groups and code camps in central Virginia. In addition, Jessica volunteers her time to help educate people through the PASS organization.

MIKE DAVIS, MCTS, MCITP, is the Managing Project Lead at Pragmatic Works. This book is his fifth on the subject of Business Intelligence and specifically Integration Services. He has worked with SQL Server for over a decade and has led many successful Business Intelligence projects with his clients. Mike is an experienced speaker and has presented at many events such as several SQL Server User Groups, Code Camps, SQL Saturday events, and the PASS Summit. Mike is an active member at his local user group (JSSUG) in Jacksonville, Florida. In his spare time he likes to play darts and guitar. You can also find him on twitter `@MikeDavisSQL`, and his blog on `MikeDavisSQL.com`.

CHRIS ROCK is a software developer and program manager for Pragmatic Works. He started developing software using VB6 and SQL Server 6.5 in 1998 and has been using SSIS since its inception. Chris has spoken at many local SQL Saturday and Code Camp events in Florida. When he's not writing code, Chris enjoys training cats to speak English. He blogs at `http://rocksthoughts.com`.

ABOUT THE TECHNICAL EDITOR

MICHAEL MCINTYRE recently progressed into the role of Consulting Sales Engineer with Pragmatic Works, coordinating solutions and engagements that utilize the Microsoft BI stack for customers with a variety of BI needs. Previously with Pragmatic Works, he was a BI Software Support Engineer focusing on ETL within SQL Server Integration Services. This transition provided the opportunity to relocate to Salem, New Hampshire, from Jacksonville, Florida. Michael spends his free time with his wife and 3-year-old daughter and enjoys keeping up with and "tinkering" with mobile technologies

CREDITS

EXECUTIVE EDITOR
Robert Elliot

SENIOR PROJECT EDITOR
Kevin Kent

TECHNICAL EDITOR
Michael McIntyre

PRODUCTION EDITOR
Christine Mugnolo

EDITORIAL MANAGER
Mary Beth Wakefield

ASSOCIATE DIRECTOR OF MARKETING
David Mayhew

MARKETING MANAGER
Ashley Zurcher

BUSINESS MANAGER
Amy Knies

VICE PRESIDENT AND EXECUTIVE GROUP PUBLISHER
Richard Swadley

ASSOCIATE PUBLISHER
Jim Minatel

PROJECT COORDINATOR, COVER
Todd Klemme

PROOFREADER
Nancy Carrasco

TECHNICAL PROOFREADERS
Steve Wake
Stephen Wynkoop

INDEXER
Johnna VanHoose Dinse

COVER DESIGNER
Wiley

COVER IMAGE
©Mark Evans/iStockphoto.com

此为试读,需要完整PDF请访问: www.ertongbook.com

ACKNOWLEDGMENTS

THANKS to everyone who made this book possible. As always, I owe a huge debt to my wife, Jenn, for putting up with my late nights, and to my children, Colton, Liam, Camille, and John, for being so patient with their tired dad who has always overextended. Thanks to Kevin Kent and my tech editor Michael McIntyre for keeping me in my place. Thanks also to the makers of Guinness for providing my special juice that helped me power through the book. Thanks for all the user group leaders out there who work so hard to help others become proficient in technology. You make a huge difference!

—Brian Knight

I MUST GIVE THANKS to God; without God in my life, I would not have such blessings. Thanks to my wife, Erin, who has had amazing patience during the late nights of writing and editing. To our three children, Collin, Justin, and Lana, who have sacrificed time away from daddy. Finally, I would like to thank my Fierljeppen mentor, Bart Helmholt, for showing me the extreme sport of Canal Jumping. The sport of Canal Jumping has always been a dream of mine, and I look forward to accomplishing my ultimate goal of Dutch Champion Canal Jumper.

—Devin Knight

THANK you to everyone for your support throughout the book writing process. Specifically, I would like to thank my coauthors and editors: Brian Knight, Devin Knight, Mike Davis, Chris Rock, Kevin Kent, and Bob Elliott. Your passion and dedication still astound me on a regular basis. Thank you to Rich, who helped me through the late nights, and my family, for putting up with my regular disappearances to write. Also, thank you to the many others that have helped me in my career, but are too numerous to name.

—Jessica M. Moss

THANKS to my Pragmatic Works Team for their support in this book. Thank you to Brian Knight for giving me the opportunity of a lifetime. Thank you to Adam Jorgensen for pushing me. Thank you to the Wiley team, especially Kevin and Bob. Thank you to the technical editor and technical proofreaders for their help in making this book great. Thank you to my mother for raising me to be the man I am today. Thank you to my kids for being awesome and putting up with all the writing time. And finally, thank you to the Flying Spaghetti Monster for showing me the way of logic and reasoning.

—Mike Davis

THANKS to everyone who made this book possible. First and foremost, I want to thank my colleague Tyler Ryan. His help made parts of this book possible, and I couldn't have completed the book without his help. I want to thank my wife, Tammy, and my two girls, Calista and Callie. Without you I wouldn't have tried so hard to achieve the goals I've reached in my life. Finally, thanks to Brian Knight again for giving me the opportunity to contribute to this book.

—Chris Rock

INTRODUCTION

THE MOST IMPORTANT BUSINESS INTELLIGENCE TOOL in the Microsoft Swiss Army knife of tools is SQL Server Integration Services (SSIS). This is because the other tools would be nothing without the cleansing and movement of data into a presentable format. The product can extract, transform, and load (ETL) data astonishingly fast. A 2010 benchmark showed movement of more than a terabyte an hour with SSIS! If you're new to SSIS, you've picked a fantastic field to become involved in. The one consistent skill needed in today's technical job market is ETL. If a company wants to establish a partnership with another company, it'll need to communicate data back and forth between the two companies. If your company wants to launch new products, it'll need a way to integrate those products into its website and catalog. All of these types of tasks are going to require the skill set you are developing and will learn in this book.

Companies that had never used SQL Server before are now allowing it in their environment because SSIS is such an easy-to-use and cost-effective way to move data. SSIS competes with the largest ETL tools on the market like Informatica, DataStage, and Ab Initio at a tiny fraction of the price. SQL Server 2014 now offers more components that you use to make your life even easier and the performance scales to a level never seen on the SQL Server platform.

The best thing about SSIS is its price tag: free with your SQL Server purchase. Many ETL vendors charge hundreds of thousands of dollars, if not millions, for what you will see in this book. SSIS is also a great platform for you to expand and integrate into, which many ETL vendors do not offer. Once you get past the initial learning curve, you'll be amazed with the power of the tool, and it can take weeks off your time to market. This author team has trained hundreds of people over the years, and you'll find that the learning curve of SSIS is shallow relative to competing platforms. In SQL Server 2012, the product matured to its third major envisioning. In that release the focus was on scalability, management, and more advanced data cleansing. In SQL Server 2014 an incremental change has happened so new content focuses on patterns in SSIS.

WHO THIS BOOK IS FOR

Because we have used SSIS since the beta stages of SQL Server 2005 and through its evolution into its current form, the idea of writing this book was quite compelling. If you've never used SSIS before, we spend the first chapters focusing on lowering your learning curve on this product. If you've used SSIS in the past, we've added more patterns and best practices to this SQL Server 2014 version of the book to take your skills to the next level. If you're an SSIS 2005 or 2008 user, luckily the interface has not drastically changed, so you'll want to focus on advanced data cleansing and administration, which has gone through a drastic overhaul.

This book is intended for developers, DBAs, and casual users who hope to use SSIS for transforming data, creating a workflow, or maintaining their SQL Server. This book is a professional book, meaning that the authors assume that you know the basics of how to query a SQL Server and have

some rudimentary programming skills. Not many programming skills will be needed or assumed, but it will help with your advancement. No skills in the prior release of SSIS are required, but we do reference it throughout the book when we call attention to feature enhancements.

WHAT THIS BOOK COVERS

Whether you're new to SSIS or an experienced SSIS developer, there's something for you in this book. This book takes you from the architecture and basics of SSIS all the way through to developing hard-core SSIS solutions to solve many of the industry's common business scenarios. The book is tutorial based, meaning that it teaches you through simple examples.

By the time you complete this book, you'll know how to load and synchronize database systems using SSIS by using some of the new SQL Server 2014 features. You'll also know how to load data warehouses, which is a very hot and specialized skill. Even in warehousing, you'll find features in the new 2014 release that you'll wonder how you lived without like the Data Quality Services integration and CDC integration!

HOW THIS BOOK IS STRUCTURED

After discussing the architecture of SSIS, we'll start with the basics by introducing the fundamental concepts of SSIS: the Data Flow and Control Flow. We'll then build through the various other features, including the warehousing and scripting, and proceed to advanced topics like programming and extending the engine. SSIS is a very feature-rich product, and it took a lot to cover the product:

Chapter 1, "Welcome to SQL Server Integration Services," introduces the concepts that we're going to discuss throughout the remainder of this book. We talk about the SSIS architecture and give a brief overview of what you can do with SSIS.

Chapter 2, "The SSIS Tools," shows you how to quickly learn how to import and export data by using the Import and Export Wizard and then takes you on a tour of the SQL Server Data Tools (SSDT).

Chapter 3, "SSIS Tasks," goes into each of the tasks that are available to you in SSIS. These tasks are the building blocks for your SSIS workflow and are much like LEGO block programming.

Chapter 4, "The Data Flow," dives into the Data Flow components in SSIS. These components are where typical ETL developers will spend 75 percent of their time when loading a database.

Chapter 5, "Using Variables, Parameters, and Expressions," instructs you how to use the obscure expression language in SSIS by showing you many example use cases and how to solve them through the language. We also cover in the chapter a new concept of parameters and parameterization of a package.

Chapter 6, "Containers," covers how to use containers to do looping in SSIS and describes how to configure each of the basic transforms.

Chapter 7, "Joining Data," focuses on how to join systems together, whether those systems are two flat files or database platforms. Much of the chapter is spent showing the Lookup Transform.

Now that you know how to configure most of the tasks and transforms, **Chapter 8, "Creating an End-to-End Package,"** puts it all together with a large example that lets you try out your SSIS experience.

Chapter 9, "Scripting in SSIS," shows you some of the ways you can use the Script task in SSIS.

Chapter 10, "Advanced Data Cleansing in SSIS," walks through common patterns in data cleansing and how to accomplish them in SSIS. The chapter also covers Data Quality Services (DQS) and how to integrate it into SSIS.

Chapter 11, "Incremental Loads in SSIS," shows you common patterns you can use to incrementally refresh your data. It also shows you how to use and integrate Change Data Capture (CDC) into SSIS.

Chapter 12, "Loading a Data Warehouse," covers how to load a data warehouse from the ground up through example. Even smaller companies now are finding that to compete they need to make their data work for them by employing a data warehouse. In this chapter we show how to load dimension and fact tables and some of the common issues.

Chapter 13, "Using the Relational Engine," discusses other features in the SQL Server arsenal that can help you build robust and high-performance ETL solutions. The SQL Server relational database engine has many features that were designed with data loading in mind, and as such the engine and SSIS form a perfect marriage to extract, load, and transform your data.

Chapter 14, "Accessing Heterogenous Data," shows you how to connect to systems other than SQL Server like Excel, XML, and web services.

Chapter 15, "Reliability and Scalability," demonstrates how to scale SSIS and make it more reliable. You can use the features in this chapter to show you how to make the package restartable if a problem occurs.

Chapter 16, "Understanding and Tuning the Data Flow Engine," explains the architecture of the SSIS Data Flow engine in detail and how to tune your SSIS packages for maximum efficiency.

Chapter 17, "SSIS Software Development Life Cycle," introduces a software development life cycle methodology to you. It speaks to how SSIS can integrate with Visual Studio Team System.

Chapter 18, "Error and Event Handling," discusses how to handle problems with SSIS with error and event handling.

Chapter 19, "Programming and Extending SSIS," shows the SSIS object model and how to use it to extend SSIS. The chapter goes through creating your own task, and then . . .

Chapter 20, "Adding a User Interface to Your Component," adds a user interface to the discussion.

Chapter 21, "External Management and WMI Task Implementation," walks through creating an application that interfaces with the SSIS to manage the environment. It also discusses the WMI set of tasks.

Chapter 22, "Administering SSIS," shows you how to deploy and administer the packages that you've worked so hard to develop. Much of the work in SSIS in SQL Server 2014 has gone into modifying the way you deploy and administer SSIS packages.

Appendix A, "SSIS Crib Notes," provides solutions to common problems and a cheat sheet for the SSIS expression language.

Appendix B, "SSIS Internal Views and Stored Procedures," gives you a listing and way to use common stored procedures in the SSIS catalog.

Appendix C, "Interviewing for an ETL Developer Position, "is your interview coach to help you ace an interview where SSIS is part of your job description.

WHAT YOU NEED TO USE THIS BOOK

To follow this book, you will only need to have SQL Server 2014 and the Integration Services component installed. You'll need a machine that can support the minimum hardware requirements to run SQL Server 2014. You'll also want to have the AdventureWorks and AdventureWorksDW databases installed. You can find the versions of these databases we used for this book on the Wrox website (`www.wrox.com`). If you download the AdventureWorks databases from CodePlex, they will have a version number at the end of the database names. Make sure that you rename the database to take this version off the name.

CONVENTIONS

To help you get the most from the text and keep track of what's happening, we've used a number of conventions throughout the book.

> **WARNING** *Warnings hold important, not-to-be-forgotten information that is directly relevant to the surrounding text.*

> **NOTE** *Notes indicate notes, tips, hints, tricks, and asides to the current discussion.*

此为试读,需要完整PDF请访问: www.ertongbook.com