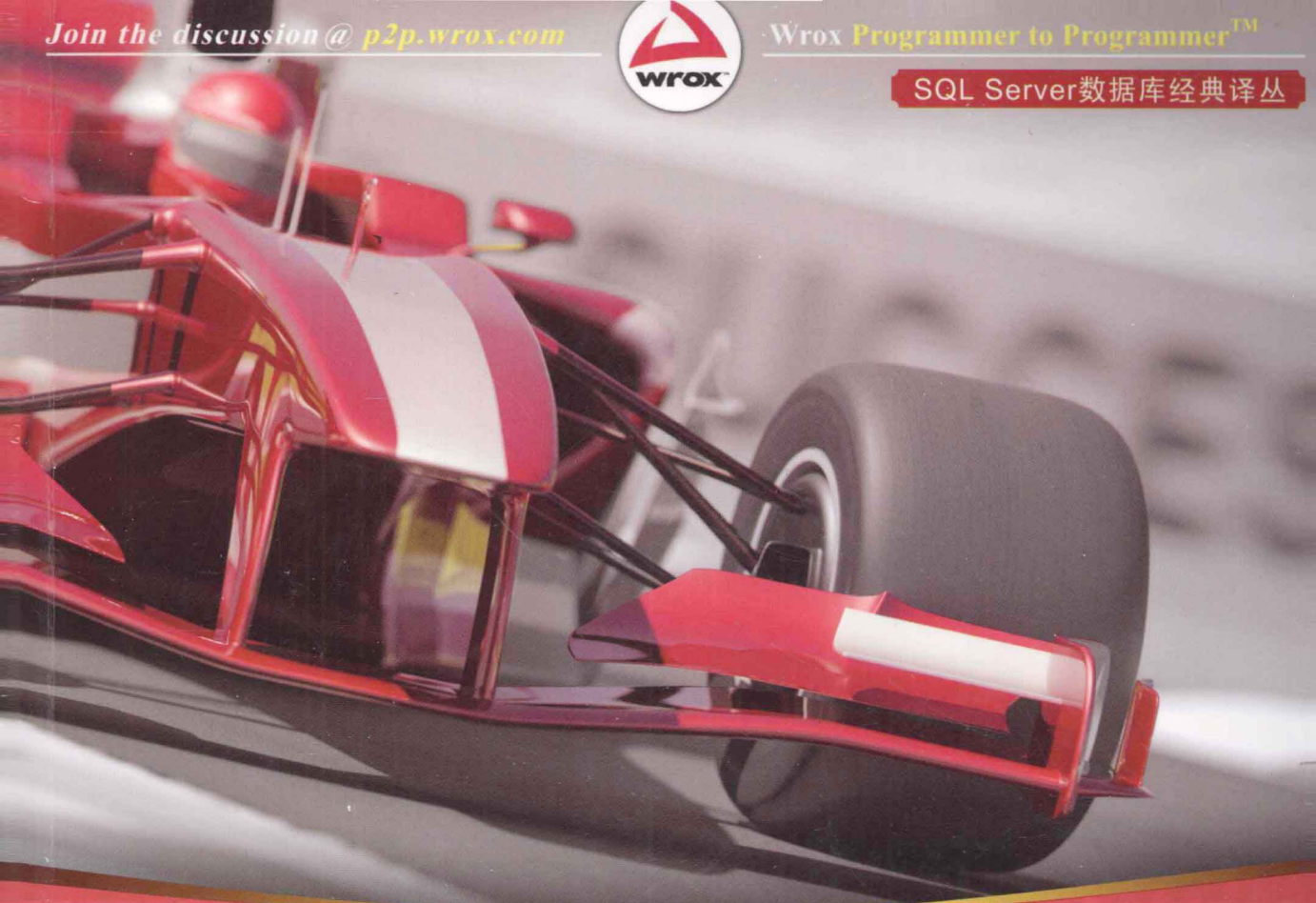




Professional Microsoft SQL Server 2012 Integration Services

SQL Server 2012 Integration Services 高级教程 (第2版)

Brian Knight Erik Veerman
[美] Jessica M. Moss Mike Davis 著
Chris Rock
王 净 谢连朋 译



清华大学出版社

SQL Server 数据库经典译丛

SQL Server 2012 Integration Services 高级教程

(第 2 版)

Brian Knight
Erik Veerman
[美] Jessica M. Moss 著
Mike Davis
Chris Rock
王 净 谢连朋 译

清华大学出版社

北 京

Brian Knight, Erik Veerman, Jessica M. Moss, Mike Davis, Chris Rock
Professional Microsoft SQL Server 2012 Integration Services
EISBN: 978-1-118-10112-4
Copyright © 2012 by Wiley Publishing, Inc., Indianapolis, Indiana
All Rights Reserved. This translation published under license.

本书中文简体字版由 Wiley Publishing, Inc. 授权清华大学出版社出版。未经出版者书面许可, 不得以任何方式复制或抄袭本书内容。

北京市版权局著作权合同登记号 图字: 01-2012-8056

Copies of this book sold without a Wiley Sticker on the cover are unauthorized and illegal.

本书封面贴有 Wiley 公司防伪标签, 无标签者不得销售。

版权所有, 侵权必究。侵权举报电话: 010-62782989 13701121933

图书在版编目(CIP)数据

SQL Server 2012 Integration Services 高级教程/(美) 奈特(Knight, B.) 等著; 王净, 谢连朋 译. —2 版
—北京: 清华大学出版社, 2013.8

(SQL Server 数据库经典译丛)

书名原文: Professional Microsoft SQL Server 2012 Integration Services

ISBN 978-7-302-33284-8

I. ①S… II. ①奈… ②王… ③谢… III. ①关系数据库系统—教材 IV. TP311.138

中国版本图书馆 CIP 数据核字(2013)第 165691 号

责任编辑: 王 军 杨信明

装帧设计: 牛艳敏

责任校对: 成凤进

责任印制: 刘海龙

出版发行: 清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址: 北京清华大学学研大厦 A 座 邮 编: 100084

社 总 机: 010-62770175 邮 购: 010-62786544

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者: 清华大学印刷厂

经 销: 全国新华书店

开 本: 185mm×260mm 印 张: 52 字 数: 1395 千字

版 次: 2008 年 10 月第 1 版 2013 年 8 月第 2 版 印 次: 2013 年 8 月第 1 次印刷

印 数: 1~3000

定 价: 128.00 元

作者简介

Brian Knight 是 SQL Server MVP、MCITP、MCSE 以及 MCDBA，是 Pragmatic Works 的拥有者和创始人。同时，他还是 BIDN.com、SQLServerCentral.com 和 SQLShare.com 的共同创始人之一。他负责运行 Jacksonville(JSSUG)的本地 SQL Server 用户组。Brian 是几个技术杂志的专栏作家，也是许多 SQL Server 书籍的作者。他经常在一些非常重要的会议上演讲，比如 PASS、SQL Connections、TechEd、SQL Saturdays、Code Camps 以及其他会议。他的博客地址为 <http://www.bidn.com>。Brian 现居住在美国佛罗里达州的 Jacksonville，他喜欢他的孩子以及跑马拉松。

Erik Veerman 是 SolidQ 的指导者，专注于培训、指导和构建 SQL Server BI 平台上的解决方案。Erik 在广泛的业务领域内设计了许多 BI 解决方案——电信、营销、零售、商业地产、金融、供应链和信息技术。他的行业认可包括 Microsoft 的 Worldwide BI Solution of the Year 以及 *SQL Server Magazine* 的创新杯冠军。作为一名 OLAP 设计、ETL 处理以及三维建模方面的专家，Erik 是研究者、作者和导师。通过在 SSIS 方面的 Wrox 书籍，他帮助推动了行业 ETL 标准和 SSIS 的最佳实践，同时他还是 Microsoft 出版社 SQL Server BI 培训工具包的主要作者。Erick 现在和他的妻子以及 4 个精力充沛的孩子居住在美国佐治亚州的 Atlanta。

Jessica Moss 是 Microsoft SQL Server 商业智能领域著名的实践者、作者和演讲者。她为不同行业的不同公司创建了大量的数据仓库和商业智能解决方案，同时还进行关于 Integration Services、Reporting Services 和 Analysis Services 的培训课程。当她为一家大型的服装零售商工作时，曾经参与了 SQL Server 2005 TAP 计划，在该计划中，她为 Integration Services 开发了最佳的实现方法。Jessica 为多家杂志、网站和图书撰写技术资料，其中包括 Wrox 图书 *Microsoft SQL Server 2008 Integration Services: Problem-Design-Solution*，并且还在许多国际会议上发言，比如 PASS Community Summit、SharePoint Connections 以及 SQLTeach International Conference。作为一名创建用户到用户社区关系的强力支持者，Jessica 通过 PASS 组织为帮助教育他人贡献自己的力量。

Mike Davis 是 MCTS、MCITP，是 Pragmatic Works 的 Managing Project Lead。本书是他的第三本关于商业智能特别是 Integration Services 方面的书。他使用 SQL Server 近十年，并且和他的客户一起开发了许多成功的商业智能项目。Mike 是一名经验丰富的演讲者，在许多事件上都提出过自己的看法，比如 SQL Server 用户组、代码阵营、SQL Saturday 事件以及 PASS Summit。Mike 还是佛罗里达州 Jacksonville 的本地用户组(JSSUG)的活跃成员。在业余时间，他喜欢玩飞镖和弹吉他。读者可以通过 twitter@MikeDavisSQL 或者博客 MikeDavisSQL.com 和 BIDN.com 与他联系。

Chris Rock 是一名软件开发人员，同时也是 Pragmatic Works 的项目经理。他在 1998 年开始使用 VB 和 SQL Server 6.5 开发软件，自 SSIS 诞生以来就一直使用 SSIS。Chris 在多个本地 SQL Saturday 以及佛罗里达州的 Code Camp 中演讲。当他不编写代码时，Chris 喜欢训练他的猫说英语。他的博客为 <http://rocksthoughts.com>。

技术编辑简介

Wee-Hyong Tok 是一名来自 SQL Server 产品组的项目经理。近几年，Wee-Hyong 在 Integration Services 上投入了大量的时间。在加入 Microsoft 之前，Wee-Hyong 就是 Microsoft 的 MVP，同时还经常在各种亚太地区的会议上发表演讲。他非常热衷于研究如何使用商业智能工具来深入了解公司的数据块，并做出更好、更明智的决策。

Matthew Roche 是一名有将近 20 年经验的开发人员、培训师和导师，他主要关注 Microsoft SQL Server 平台。从 SQL Server 6.5 开始，他就设计、开发和部署了基于 SQL Server 的不同大小规模的应用程序，随着 SQL Server 2005 的发布，他又开始关注使用 SQL Server Integration Services 的 BI 和 ETL 解决方案。Matthew 在美国和欧洲众多用户组、Code Camps 以及技术会议上发表关于 SQL Server 主题的演讲，包括 TechEd 和 SQL Server Connections，并且作为作者以及技术评论者为多本 SQL Server 图书贡献了自己的力量。自从 1996 年开始，Matthew 就是 MCT，在 2008 年加入 Microsoft 之前他还是 SQL Server 方面的 MVP。目前，他是 SQL Server Integration Services 团队的高级项目经理。

致 谢

感谢所有让本书得以出版的人。同样，也感谢我的妻子 Jenn 以及 4 个孩子：Colton、Liam、Camille 和 John，感谢他们每天忍受我到深夜。感谢 Kevin Kent 以及我的技术编辑 Matthew Roche 和 Wee-Hyong Tok，是他们始终让我保持清醒的头脑。此外，还要感谢 Guinness 的制造商，他们提供了帮助我完成本书的动力。感谢所有用户团队领导者如此努力地工作，帮助他人精通技术。最后感谢我的代理教练 Shawn Harrison，感谢他帮助我准备我的处女作。

—Brian Knight

有很多人需要感谢。首先感谢 Brian 和 Wrox 团队让我拥有了参与编写这本伟大图书的机会。您将会发现，我将这本书献给了我的父母。感谢他们为我购买了 Commodore 64，当然，还要感谢他们能够为我提供永久支持的机会(哪一个技术专家没有得到家庭的帮助支持呢？)。我永远忘不了将桌面图标从鼠标指针移开的病毒，或者去年打印机不工作的“问题”(插上它没有任何反应)。

我想要极力推荐 Discovery Channel 的纪录片 *How Beer Saved the World*。本书就是其理论的简介证据。同样，我还要感谢我的妻子 Amy 和孩子，感谢他们的耐心和宽容。最后感谢 John Calvin(我的哲学和神学崇拜者)以及改革者 Soli Deo Gloria。

—Erik Veerman

我要真诚感谢那些在我的职业生涯中帮助过我的人们。首先，要感谢这本书的合著者和编辑：Brian Knight、Erik Veerman、Mike Davis、Chris Rock、Matthew Roche、Wee-Hyong Tok、Kevin Kent、Bob Elliott 以及 Wrox 团队的其他人。养大一个孩子需要一个村庄，我们的“孩子”也不例外。同时，还要感谢我的家人以及我身边的朋友。没有你们，我也不会完成本书。谢谢！

—Jessica M. Moss

感谢我的 Pragmatic Works 团队对本书的支持。感谢 Brian Knight 为我提供这次机会，还要感谢 Adam Jorgensen 的推荐。感谢 Wiley 团队，特别是 Kevin 和 Bob。感谢技术编辑为本书所提供的大量帮助。此外，还要感谢我的母亲对我的培养，以及我的妻子和孩子。最后感谢 Flying Spaghetti Monster 为我提供了逻辑和推理的方法。

—Mike Davis

感谢所有帮助本书出版的人。首先，我想感谢我的妻子 Tammy，是她忍受了我多月来常说的一句话“我正在写书”。其次还要感谢我的两个女儿 Galista 和 Callie。没有你们，我也不会努力实现目前已完成的目标。此外，感谢 Matt Masson 在我编写本书期间所提及的相关问题。感谢 Brian Knight 给予我编写本书的机会。最后，感谢 Andy Warren 这些年来为我所指点的正确的专业发展道路。

—Chris Rock

前 言

在 Microsoft 的所有工具中，最重要的商业智能工具当属 SQL Server Integration Services(SSIS)。这是因为其他的工具不具有清洗和将数据转换为一种合适格式的功能。而 SSIS 却可以快速地完成对数据的提取、转换和加载(ETL)。2010 年所做的一项基准测试显示，通过使用 SSIS，在 1 小时内可以移动超过 1TB 的数据。如果你是初次接触 SSIS，那么可以选择一个领域来做深入研究。在当今技术就业市场所需的一项技能就是 ETL。如果某一公司想与另一家公司建立合作伙伴关系，那么这两家公司之间就需要来回地进行数据通信。如果您的公司想推出新产品，那么就需要一种方法将这些新产品集成到现有网站和目录中。而所有类型的任务都需要您将要开发以及从本书中学到的技能集。

那些以前从没有使用过 SQL Server 的公司现在正允许在它们的环境中使用该工具，因为 SSIS 是一种易于使用且经济高效的移动数据的方法。SSIS 与市场上那些大型的 ETL 工具相竞争，而价格却只占这些竞争元素一小部分。现在，SQL Server 2012 提供了更多使您的开发更加简单的组件，从而将其性能提升到一个前所未有的水平。

有关 SSIS 最好的事情就是价格标签：在购买 SQL Server 的同时免费附带了 SSIS 工具。而许多其他的 ETL 供应商则会收取数十万美元(甚至数百万美元)，本书将会介绍这些 ETL 工具。同时，SSIS 还是一个功能强大的平台，可以进行扩展和融入，而许多其他的 ETL 供应商则没有提供这些功能。一旦顺利地完成了最初的学习曲线，那么您会对该工具的强大功能感到吃惊，它可以将产品推向市场的周期缩短数周。本书的作者团队多年来培训了数以百计的人，您将会发现，相比于其他的竞争平台，SSIS 的学习曲线更加平缓。在 SQL Server 2012 中，该工具已经非常成熟，主要重点关注可扩展性、管理以及更高级的数据清洗功能。

本书读者对象

因为我们从 SQL Server 2005 的测试阶段就开始使用 SSIS，直到演变为目的的版本，所以编写这本书的想法是颇为引人注目的。如果您以前从没使用过 SSIS，那么本书的第 1 章所介绍的内容将会最大程度地降低您的学习曲线。当然，如果您使用过 SSIS，那么我们也会向您介绍一些 SQL Server 2012 所特有的新内容，从而将您的技能提升到一个新的水平。而如果您是 SQL Server 2005 或者 SQL Server 2008 的用户，那么庆幸的是 SQL Server 2012 的界面并没有太大的变化，所以您可以更专注地学习高级的数据清洗和管理功能(这些功能都经历了彻底的改革)。

本书的主要读者是开发人员、DBA 以及那些希望使用 SSIS 来转换数据、创建 workflow 或者维护 SQL Server 的普通用户。本书是一本专业书，这意味着本书的作者将假设您已经掌握了如何查询 SQL Server 的基本知识和一些基本的编程技能。虽然本书并不需要太多的编程技能，但是这些技能将会帮助你学习。同时，本书也不需要 SSIS 以前版本的使用技能，但是我们重点介绍增强功能时会提及这些内容。

本书主要内容

不管你是初次接触 SSIS 还是一名有经验的 SSIS 开发人员,本书中都有您所需要的内容。本书将从 SSIS 的体系结构和基本概念开始介绍,然后讨论如何开发核心的 SSIS 解决方案来解决大多数业界常见的业务方案。同时本书还是一本基础教程,因而将通过一些简单的示例来讲解。

当阅读完本书后,您将会知道如何通过使用新的 SQL Server 2012 功能对使用了 SSIS 的数据库系统进行加载和保持同步。同时还将了解如何加载数据仓库(这是一项非常热门的专业技能)。即使在仓库方面,你也会在新的 2012 发布版本中找到相应的功能,此时你会惊奇如果没有 Data Quality Service 集成和 CDC 集成,那么开发将如何进行!

本书组织结构

在讨论完 SSIS 的体系结构后,接下来开始介绍 SSIS 的基本概念: Data Flow 和 Control Flow。然后介绍各种其他的功能(比如仓库和脚本)以及一些高级主题,比如对引擎进行编程和扩展。最后,我们以一个案例研究结束,通过该案例的学习,有助于将前面所需要的内容联系在一起。SSIS 是一个功能丰富的产品,本书用了相当多的章节来介绍该产品:

- 第 1 章“欢迎使用 SQL Server Integration Services”:主要介绍了本书后续章节中将要论述的一些概念。讨论了 SSIS 的体系结构,并简要概述了使用 SSIS 可以完成的工作。
- 第 2 章“SSIS 工具”:学习使用导入和导出向导导入和导出数据的快速方法,以及 SQL Server Data Tool(SSDT)。
- 第 3 章“SSIS 任务”:讨论 SSIS 中每一个可用的任务。这些任务是 SSIS 工作流的构建基块,类似于 LEGO 积木编程。
- 第 4 章“容器”:介绍了如何在 SSIS 中通过使用容器来实现循环,并描述了如何对每个基本转换进行配置。
- 第 5 章“数据流”:介绍了 SSIS 中的 Data Flow 组件。一般来说,当加载一个数据库时,ETL 开发人员 75%的时间都花费在这些组件上。
- 第 6 章“使用变量、参数和表达式”:通过演示一些示例,使用案例介绍如何使用 SSIS 中的模糊表达式语言,以及如何通过该语言解决这些使用案例。同时在本章中还将介绍包中参数和参数化的概念。
- 第 7 章“联接数据”:重点介绍如何将系统联接在一起,而不管这些系统是两个平面文件还是数据库平台。本章的大部分内容都将用来介绍 Lookup Transform。
- 到目前为止,您已经知道如何对大多数任务和转换进行配置,那么第 8 章“创建端到端的包”会将前面所介绍的内容融入一个大的示例中,从而让您体验一下 SSIS。
- 第 9 章“在 SSIS 中编写脚本”:介绍一些使用 SSIS 中脚本任务的方法。
- 第 10 章“加载数据仓库”:通过示例从头开始介绍如何加载数据仓库。现在,即使是小公司也逐步发现,为了提高竞争力,需要通过采用数据仓库的方式来使用它们的数据。本章重点介绍了如何加载维度表和事实表以及一些常见的问题。

- 第 11 章“SSIS 中高级的数据清洗”：介绍数据清洗中的常见模式以及如何在 SSIS 中实现这些模式。同时，还介绍了 Data Quality Services(DQS)以及如何将它与 SSIS 进行集成。
- 第 12 章“使用关系引擎”：讨论了 SQL Server 中可用来帮助构建健壮且高性能的 ETL 解决方案的相关功能。SQL Server 关系数据库引擎拥有许多用于数据加载的功能，通过该引擎和 SSIS，可以实现对数据的提取、加载和转换。
- 有时可能需要连接到其他系统而不是 SQL Server。第 13 章“访问异构数据”将会介绍如何连接到非 SQL Server 的系统上，比如 Excel、XML 以及 Web 访问。在 SQL Server 2012 发布版本中，许多都工作都由新的 ODBC 集成来完成。
- 第 14 章“可靠性和可扩展性”：论述了如何对 SSIS 进行扩展以及如何提高其可靠性。如果出现相关问题，您可以使用本章所介绍的功能来展示如何使包可重启。
- 第 15 章“理解和优化数据流引擎”：详细介绍了 SSIS 数据流引擎的体系结构以及如何对 SSIS 包进行调整，使其具有最大的效率。
- 第 16 章“SSIS 软件开发生命周期”：介绍了软件开发生命周期方法，以及如何将 SSIS 和 Visual Studio Team System 相集成。
- 第 17 章“错误和事件处理”：讨论如何使用错误和事件处理来处理相关的 SSIS 问题。
- 第 18 章“编程和扩展 SSIS”：介绍了 SSIS 对象模型以及如何使用该模型对 SSIS 进行扩展。同时教您如何创建自己的任务。然后第 19 章“向组件添加用户界面”将向第 18 章中的示例添加一个用户界面。
- 第 20 章“外部管理和 WMI 任务实现”：介绍如何创建一个与 SSIS 接口的、用来管理环境的应用程序。同时还将讨论 WMI 任务集。
- 第 21 章“通过外部应用程序使用 SSIS”：教您如何向其他应用程序(比如 InfoPath 和自己的 .NET 应用程序)公开 SSIS 数据流。
- 第 22 章“管理 SSIS”：介绍如何部署和管理所开发的包。在 SQL Server 2012 的 SSIS 中，大部分的工作都是用来修改部署和管理 SSIS 包的方法。
- 第 23 章“案例研究：一个编程示例”：介绍一个创建端到端项目的编程案例研究。

使用本书所需条件

为了阅读本书，您只需要安装 SQL Server 2012 和 Integration Services 组件。同时还需要一台满足运行 SQL Server 2012 的最小硬件需求的计算机，以及安装 AdventureWorks 和 AdventureWorksDW 数据库。您可以从 Wrox 网站(www.wrox.com)上找到本书所使用的数据库版本。

源代码

在读者学习本书中的示例时，可以手工输入所有代码，也可以使用本书附带的源代码文件。本书使用的所有源代码都可以从本书合作站点 <http://www.wrox.com> 或 www.tupwk.com.cn/downpage 上下载。登录到站点 <http://www.wrox.com/>，使用 Search 工具或使用书名列表就可以找到本书。接着单击本书细目页面上的 Download Code 链接，就可以获得所有源代码。



由于许多图书的标题都很类似，所以按 ISBN 搜索是最简单的，本书英文版的 ISBN 是 978-1-118-10112-4。

在下载了代码后，只需用自己喜欢的解压缩软件对它进行解压缩即可。另外，也可以进入 <http://www.wrox.com/dynamic/books/download.aspx> 上的 Wrox 代码下载主页，查看本书和其他 Wrox 图书的所有代码。

勘误表

尽管我们已经尽了各种努力来保证文章或代码中不出现错误，但是错误总是难免的，如果您在本书中找到了错误，例如拼写错误或代码错误，请告诉我们，我们将非常感激。通过勘误表，可以让其他读者避免受挫，当然，这还有助于提供更高质量的信息。

请给 wkservice@vip.163.com 发电子邮件，我们就会检查您的反馈信息，如果是正确的，我们将在本书的后续版本中采用。在本书编辑过程中，我们接受了热心读者白爽针对第 4 版中文译著提出的一些修改意见，在此特向白爽表示衷心感谢。

要在网站上找到本书英文版的勘误表，可以登录 <http://www.wrox.com>，通过 Search 工具或书名列表查找本书，然后在本书的细目页面上，单击 Book Errata 链接。在这个页面上可以查看到 Wrox 编辑已提交和粘贴的所有勘误项。完整的图书列表还包括每本书的勘误表，网址是 www.wrox.com/misc-pages/booklist.shtml。

p2p.wrox.com

要与作者和同行讨论，请加入 p2p.wrox.com 上的 P2P 论坛。这个论坛是一个基于 Web 的系统，便于您张贴与 Wrox 图书相关的消息和相关技术，与其他读者和技术用户交流心得。该论坛提供了订阅功能，当论坛上有新的消息时，它可以给您传送感兴趣的论题。Wrox 作者、编辑和其他业界专家和读者都会到这个论坛上来探讨问题。

在 <http://p2p.wrox.com> 上，有许多不同的论坛，它们不仅有助于阅读本书，还有助于开发自己的应用程序。要加入论坛，可以遵循下面的步骤：

- (1) 进入 p2p.wrox.com，单击 Register 链接。
- (2) 阅读使用协议，并单击 Agree 按钮。
- (3) 填写加入该论坛所需要的信息和自己希望提供的其他信息，单击 Submit 按钮。
- (4) 您会收到一封电子邮件，其中的信息描述了如何验证账户，完成加入过程。



不加入 P2P 也可以阅读论坛上的消息，但要张贴自己的消息，就必须加入该论坛。

加入论坛后，就可以张贴新消息，响应其他用户张贴的消息。可以随时在 Web 上阅读消息。如果要让该网站给自己发送特定论坛中的消息，可以单击论坛列表中该论坛名旁边的 **Subscribe to this Forum** 图标。

关于使用 Wrox P2P 的更多信息，可阅读 **P2P FAQ**，了解论坛软件的工作情况以及 P2P 和 Wrox 图书的许多常见问题。要阅读 **FAQ**，可以在任意 **P2P** 页面上单击 **FAQ** 链接。

目 录

第 1 章 欢迎使用 SQL Server Integration Services.....	1	2.7 小结.....	34
1.1 SQL Server SSIS 历史概述.....	2	第 3 章 SSIS 任务.....	35
1.2 SSIS 中的新内容.....	2	3.1 SSIS 任务对象.....	35
1.3 商业工具.....	2	3.1.1 使用任务编辑器.....	36
1.3.1 导入和导出向导.....	2	3.1.2 任务编辑器的“表达式”选项卡.....	36
1.3.2 SQL Server Data Tools 体验.....	3	3.1.3 SDDTCommon 属性.....	37
1.4 SSIS 体系结构.....	4	3.2 循环和序列任务.....	38
1.4.1 包.....	4	3.3 脚本任务(.NET).....	39
1.4.2 控制流.....	5	3.4 Analysis Services 任务.....	41
1.4.3 数据流.....	8	3.4.1 Analysis Services 执行 DDL 任务.....	41
1.4.4 变量.....	12	3.4.2 Analysis Services 处理任务.....	42
1.4.5 参数.....	12	3.4.3 数据挖掘查询任务.....	42
1.4.6 错误处理和日志记录.....	12	3.5 数据流任务.....	43
1.5 SQL Server 版本.....	12	3.6 数据准备任务.....	44
1.6 小结.....	13	3.6.1 数据事件探查器.....	44
第 2 章 SSIS 工具.....	15	3.6.2 文件系统任务.....	46
2.1 导入和导出向导.....	15	3.6.3 对文件进行归档.....	47
2.1 SQL Server Data Tools.....	21	3.6.4 FTP 任务.....	48
2.3 创建第一个包.....	23	3.6.5 使用 FTP 获取文件.....	49
2.4 解决方案资源管理器窗口.....	24	3.6.6 Web 服务任务.....	50
2.4.1 工具箱.....	25	3.6.7 使用 Web 服务任务和 XML 源组件来 获取数据.....	52
2.4.2 属性窗口.....	25	3.6.8 XML 任务.....	55
2.5 SSIS 包设计器.....	25	3.6.9 验证 XML 文件.....	56
2.5.1 控制流.....	26	3.7 RDBMS 服务器任务.....	59
2.5.2 连接管理器.....	30	3.7.1 大容量插入任务.....	59
2.5.3 变量.....	30	3.7.2 使用大容量插入任务.....	61
2.5.4 数据流.....	31	3.7.3 执行 SQL 任务.....	62
2.5.5 参数.....	32	3.8 工作流任务.....	72
2.5.6 事件处理程序.....	32	3.8.1 执行包任务.....	72
2.5.7 包资源管理器.....	33	3.8.2 执行进程任务.....	73
2.5.8 执行包.....	33	3.8.3 消息队列任务.....	74
2.6 Management Studio.....	34	3.8.4 发送邮件任务.....	75

3.8.5	WMI 数据读取器任务	76	5.4.8	DataReader 目标	110
3.8.6	WMI 事件观察器任务	79	5.4.9	处理维度目标和处理分区目标	110
3.8.7	为了文件传输而轮询目录	79	5.5	常用转换	110
3.9	SMO 管理任务	80	5.5.1	同步转换和异步转换	111
3.9.1	传输数据库任务	80	5.5.2	聚合转换	111
3.9.2	传输错误消息任务	81	5.5.3	有条件拆分转换	112
3.9.3	传输登录名任务	82	5.5.4	数据转换	113
3.9.4	传输主存储过程任务	82	5.5.5	派生列转换	114
3.9.5	传输作业任务	83	5.5.6	查找转换	115
3.9.6	传输 SQL Server 对象任务	83	5.5.7	缓存转换	116
3.10	小结	84	5.5.8	行计数转换	116
第4章	容器	85	5.5.9	脚本组件	117
4.1	任务宿主容器	85	5.5.10	渐变维度转换	117
4.2	序列容器	85	5.5.11	排序转换	118
4.3	组	86	5.5.12	Union All 转换	119
4.4	For 循环容器	87	5.6	其他转换	120
4.5	Foreach 循环容器	89	5.6.1	审核转换	120
4.5.1	Foreach 文件枚举器示例	90	5.6.2	字符映射表转换	121
4.5.2	Foreach ADO 枚举器示例	92	5.6.3	复制列转换	122
4.6	小结	95	5.6.4	数据挖掘查询转换	122
第5章	数据流	97	5.6.5	DQS 清除转换	123
5.1	理解数据流	97	5.6.6	导出列转换	123
5.2	数据查看器	98	5.6.7	模糊查找转换	124
5.3	源	98	5.6.8	模糊分组转换	130
5.3.1	OLE DB 源	99	5.6.9	导入列转换	133
5.3.2	Excel 源	101	5.6.10	合并转换	135
5.3.3	平面文件源	102	5.6.11	合并联接转换	136
5.3.4	原始文件源	106	5.6.12	多播转换	137
5.3.5	XML 源	107	5.6.13	OLE DB 命令转换	137
5.3.6	ADO.NET 源	107	5.6.14	百分比抽样转换和行抽样转换	138
5.4	目标	107	5.6.15	透视转换	138
5.4.1	Excel 目标	108	5.6.16	逆透视转换	141
5.4.2	平面文件目标	108	5.6.17	字词提取转换	142
5.4.3	OLE DB 目标	108	5.6.18	字词查找转换	146
5.4.4	原始文件目标	109	5.7	数据流示例	147
5.4.5	记录集目标	109	5.8	小结	150
5.4.6	SQL Server 和移动目标	109	第6章	使用变量、参数和表达式	151
5.4.7	数据挖掘模型定型目标	110	6.1	动态包对象	151

6.1.1 变量概述	152	第 8 章 创建端到端的包	213
6.1.2 参数概述	152	8.1 基本转换教程	213
6.1.3 表达式概述	153	8.1.1 创建连接	214
6.2 理解数据类型	153	8.1.2 创建控制流	216
6.2.1 SSIS 数据类型	153	8.1.3 创建数据流	217
6.2.2 日期和时间类型支持	155	8.1.4 完成包	218
6.2.3 错误的数据类型和大小如何影响 性能	155	8.1.5 保存包	219
6.2.4 Unicode 和非 Unicode 转换问题	156	8.1.6 执行包	219
6.2.5 SSIS 表达式中的强制转换	158	8.2 典型的带有数据清理的大型机 ETL	220
6.3 使用变量和参数	159	8.2.1 创建数据流	221
6.3.1 定义变量	159	8.2.2 处理脏数据	222
6.3.2 定义参数	160	8.2.3 最终敲定	225
6.3.3 变量和参数数据类型	161	8.2.4 处理更多的不良数据	226
6.4 使用表达式	162	8.2.5 循环和动态任务	228
6.4.1 像 C# 吗? 接近但不完全是	162	8.2.6 循环	228
6.4.2 表达式生成器	163	8.2.7 使包具有动态性	229
6.4.3 基本语法	164	8.3 小结	232
6.4.4 在 SSIS 包中使用表达式	176	第 9 章 在 SSIS 中编写脚本	233
6.5 小结	184	9.1 SSIS 脚本介绍	233
第 7 章 联接数据	185	9.2 SSIS 脚本使用入门	234
7.1 查找转换	186	9.2.1 选择脚本语言	235
7.2 使用合并联接转换	186	9.2.2 使用 VSTA Scripting IDE	236
7.3 对比 SSIS 以及关系联接	187	9.2.3 示例: Hello World	237
7.4 查找功能	189	9.2.4 添加代码和类	238
7.5 生成基本包	190	9.2.5 使用托管程序集	240
7.5.1 在源中使用关系联接	192	9.2.6 示例: 使用自定义 .NET 程序集	241
7.5.2 使用合并联接转换	194	9.3 使用脚本任务	243
7.6 使用查找转换	198	9.3.1 配置脚本任务编辑器	244
7.6.1 完全缓存模式	199	9.3.2 脚本任务 Dts 对象	245
7.6.2 无缓存模式	201	9.3.3 在脚本任务中访问变量	245
7.6.3 部分缓存模式	202	9.3.4 在脚本任务中连接数据源	251
7.6.4 多个输出	204	9.3.5 在脚本任务中触发事件	259
7.6.5 表示式属性	207	9.3.6 在脚本任务中写入日志条目	264
7.6.6 级联查找操作	208	9.4 使用脚本组件	266
7.7 缓存连接管理器和缓存转换	210	9.4.1 与脚本任务的区别	266
7.8 小结	212	9.4.2 配置脚本组件编辑器	266
		9.4.3 在脚本组件中访问变量	269
		9.4.4 在脚本组件中连接数据源	270

9.4.5	触发事件	270	12.1.2	最佳方法是什么呢?	353
9.4.6	日志记录	271	12.1.3	提取期间的转换	354
9.4.7	同步与异步	280	12.1.4	许多 AND 使工作更加轻松	357
9.5	基本代码、调试和故障诊断		12.1.5	数据库中的排序	357
技术		282	12.1.6	模块化	359
9.5.1	结构化异常处理	282	12.1.7	SQL Server 也处理文本	360
9.5.2	脚本调试和故障诊断	285	12.1.8	使用基于集的逻辑	364
9.6	小结	288	12.2	SQL Server 变更数据捕获	365
第 10 章	加载数据仓库	289	12.2.1	SQL Server CDC 的好处	366
10.1	数据分析	290	12.2.2	准备 CDC	367
10.1.1	初次执行数据事件探查任务	291	12.2.3	捕获实例表	369
10.1.2	检查数据事件探查任务的结果	293	12.2.4	CDC API	370
10.1.3	将数据分析的结果转化为可操作的 ETL 步骤	296	12.2.5	使用新的 SSIS CDC 工具	372
10.2	数据提取和清洗	296	12.2.6	在 SSIS 中查询 CDC	376
10.3	维度表加载	297	12.3	数据加载	379
10.3.1	加载简单的维度表	298	12.3.1	数据库快照	379
10.3.2	加载复杂维度表	302	12.3.2	MERGE 运算符	381
10.3.3	SCD 转换的注意事项和替代方案	311	12.4	小结	384
10.4	事实表加载	312	第 13 章	访问异构数据	385
10.5	SSAS 处理	322	13.1	Excel 和 Access	386
10.6	使用主 ELT 包	327	13.1.1	64 位支持	387
10.7	小结	328	13.1.2	使用 Excel 文件	388
第 11 章	SSIS 中高级的数据清洗	329	13.1.3	使用 Access	392
11.1	高级派生列的使用	330	13.2	从 Oracle 中导入数据	399
11.2	高级模糊查找和模糊分组	333	13.2.1	Oracle 客户端设置	399
11.2.1	模糊查找转换	333	13.2.2	导入 Oracle 数据	400
11.2.2	模糊分组转换	339	13.3	使用 XML 和 Web 服务	402
11.3	DQS 清除	342	13.3.1	配置 Web 服务任务	402
11.3.1	数据质量服务	343	13.3.2	使用 XML 数据作为源	411
11.3.2	DQS 清除转换	346	13.4	平面文件	414
11.4	小结	349	13.4.1	加载平面文件	414
第 12 章	使用关系引擎	351	13.4.2	从平面文件中提取数据	416
12.1	数据提取	351	13.5	ODBC	418
12.1.1	SELECT * 的缺点	352	13.6	其他的异构源	420
			13.7	小结	421
			第 14 章	可靠性和可扩展性	423
			14.1	重新启动包	423
			14.1.1	简单的控制流	424

14.1.2 容器内的容器和检查点	427	第 17 章 错误和事件处理	513
14.1.3 关于主题的变化	429	17.1 使用优先约束	513
14.2 包事务	433	17.1.1 优先约束基础知识	514
14.2.1 单个包, 单个事务	433	17.1.2 高级优先约束和表达式	515
14.2.2 单个包, 多个事务	435	17.2 事件处理	521
14.2.3 两个包, 单个事务	437	17.2.1 事件	522
14.2.4 使用 SQL Server 中本机事务 的单一包	438	17.2.2 使用事件处理程序	523
14.3 错误输出	439	17.2.3 事件处理程序继承	530
14.4 向外扩展	442	17.3 断点	531
14.4.1 体系结构功能	442	17.4 错误行	534
14.4.2 向外扩展内存压力	443	17.5 日志记录	538
14.4.3 通过暂存数据向外扩展	443	17.5.1 日志记录提供程序	538
14.4.4 使用并行加载来向外扩展	447	17.5.2 日志事件	539
14.5 小结	454	17.5.3 目录日志记录	542
第 15 章 理解和优化数据流引擎	455	17.6 小结	544
15.1 SSIS 引擎	456	第 18 章 编程和扩展 SSIS	545
15.1.1 理解 SSIS 数据流和控制流	456	18.1 示例组件	546
15.1.2 使用控制流来处理工作流	458	18.1.1 组件 1: 源适配器	546
15.1.3 数据流中的数据处理	459	18.1.2 组件 2: 转换	547
15.1.4 内存缓冲区结构	459	18.1.3 组件 3: 目标适配器	547
15.1.5 转换类型	460	18.2 管道组件方法	547
15.1.6 高级数据流执行概念	468	18.2.1 设计时功能	548
15.2 SSIS 数据流设计和优化	474	18.2.2 运行时	551
15.2.1 数据流设计实践	474	18.2.3 连接时间	553
15.2.2 优化包处理	479	18.3 构建组件	554
15.2.3 排除数据流性能瓶颈	481	18.3.1 准备	554
15.3 管道性能监视	483	18.3.2 构建源组件	559
15.4 小结	485	18.3.3 构建转换组件	570
第 16 章 SSIS 软件开发生命周期	487	18.3.4 构建目标适配器	580
16.1 软件开发声明周期简介	489	18.4 使用组件	587
16.1.1 SDLC 的简要历史	489	18.4.1 安装组件	587
16.1.2 SDLC 的类型	489	18.4.2 调试组件	588
16.2 版本和源代码控制	490	18.4.3 设计时	588
16.2.1 SVN	491	18.4.4 构建完整的包	589
16.2.2 Team Foundation Server、Team System 和 SSIS	496	18.4.5 运行时调试	590
16.3 小结	511	18.5 升级到 SQL Server 2012	593
		18.6 小结	593

第 19 章 向组件添加用户界面	595
19.1 设计 UI 的三个关键步骤:	
概述.....	595
19.2 构建 UI.....	596
19.2.1 添加项目.....	596
19.2.2 实现 IDtsComponentUI.....	599
19.2.3 设置 UITypeName.....	602
19.2.4 构建窗体.....	604
19.3 扩展 UI.....	608
19.3.1 运行时连接.....	609
19.3.2 组件属性.....	611
19.3.3 处理错误和警告.....	613
19.3.4 列属性.....	615
19.4 其他 UI 的注意事项.....	616
19.5 小结.....	616
第 20 章 外部管理和 WMI 任务实现	619
20.1 使用托管代码进行 SSIS 的外部管理.....	619
20.1.1 设置一个用来演示的测试	
SSIS 包.....	620
20.1.2 托管对象模型代码库.....	620
20.1.3 目录管理.....	622
20.1.4 文件夹管理.....	622
20.1.5 环境.....	624
20.1.6 DTS 运行时托管代码库.....	625
20.1.7 SSIS 部署项目.....	626
20.1.8 参数对象.....	627
20.1.9 服务器部署.....	628
20.1.10 执行 SSIS 包部署到 SSIS	
目录.....	629
20.1.11 EnvironmentReferences.....	630
20.2 包操作.....	631
20.3 应用程序对象维护操作.....	631
20.3.1 包操作.....	631
20.3.2 包监视.....	634
20.3.3 项目、文件夹和报列表.....	636
20.3.4 包管理示例.....	637
20.4 包日志提供程序.....	645
20.4.1 指定需要记录的事件.....	647
20.4.2 对日志提供程序进行编程.....	647
20.4.3 SQL Server 2012 操作日志.....	649
20.5 包配置.....	651
20.5.1 创建配置.....	651
20.5.2 对配置对象进行编程.....	653
20.5.3 配置对象.....	653
20.6 WMI.....	654
20.6.1 WMI 数据读取器任务说明.....	654
20.6.2 WMI 数据阅读器示例.....	655
20.6.3 WMI 事件观察器任务.....	660
20.6.4 WMI 事件观察器任务示例.....	661
20.7 小结.....	663
第 21 章 通过外部应用程序使用 SSIS	665
21.1 InfoPath 文档.....	666
21.2 ASP.NET 应用程序.....	671
21.3 WinForm .NET 应用程序.....	676
21.4 小结.....	684
第 22 章 管理 SSIS	685
22.1 使用 SSIS 目录.....	685
22.1.1 设置 SSIS 目录属性.....	686
22.1.2 SSISDB.....	689
22.2 部署模型.....	689
22.2.1 项目部署模型.....	690
22.2.2 包部署模型.....	692
22.3 在 SSIS 中使用 T-SQL.....	698
22.3.1 执行包.....	698
22.3.2 使用参数.....	700
22.3.3 查询表以获取参数值.....	700
22.3.4 使用环境.....	701
22.3.5 使用数据分流.....	706
22.4 创建中央 SSIS 服务器.....	707
22.5 聚类 SSIS.....	708
22.6 包配置.....	710
22.7 命令行实用工具.....	714
22.7.1 DTExec.....	714
22.7.2 DTExecUI.....	715
22.7.3 DTUtil.....	719