

Microsoft®

SQL Server™ 2005

SAMS

Integration Services

SQL Server 2005

Integration Services

标准指南 (中文版)

微软公司 Bill Baker 作序

[美] Kirk Haselden 著

马树奇 金 燕 译



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

<http://www.phei.com.cn>

Microsoft SQL Server 2005 Integration Services

SQL Server 2005 Integration Services 标准指南(中文版)

[美] Kirk Haselden 著
马树奇 金 燕 译

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

Integration Services 是 SQL Server 2005 的重要开发工具,是 DTS(数据转换服务)的后续与革新。SQL Server 2005 提供了最新的大型数据库管理系统。在多年来建立起各种数据库之后,几乎所有的应用单位都面临着如何将数据库中的数据根据企业应用的需要提取出来、转换成需要的形式和载入相关文件或者应用程序的问题,这就是所谓的 ETL。Integration Services 是 SQL Server 2005 提供的最新的 ETL 处理工具,它以 Visual Studio 为基础,提供了大量现成的组件,可供人们快速建立起运行稳定、性能出色的 ETL 程序。

本书适合具有一定 SQL Server 2005 应用经验,需要对数据库的企业应用进行二次开发的 IT 技术人员,以及希望了解 ETL 技术的有关人士使用。

SAMS

Authorized translation from the English language edition, entitled MICROSOFT SQL SERVER 2005 INTEGRATION SERVICES, 1st Edition, 0672327813 by HASELDEN, KIRK, published by Pearson Education, Inc, publishing as Sams, Copyright©2007 by Sams Publishing.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc. CHINESE SIMPLIFIED language edition published by BEIJING MEDIA ELECTRONIC INFORMATION CO. LTD & PUBLISHING HOUSE OF ELECTRONICS INDUSTRY, Copyright©2008.

本书英文版由 Sams 公司出版, Sams 公司已将本书简体中文版独家版权授予中国电子工业出版社及北京美迪亚电子信息有限公司。未经许可,不得以任何形式和手段复制或抄袭本书内容。

版权贸易合同登记号 图字:01-2007-2962

图书在版编目(CIP)数据

SQL Server 2005 Integration Services 标准指南:中文版/(美)哈瑟登(Haselden, K.)著;马树奇等译. 北京:电子工业出版社,2008.1

书名原文:Microsoft SQL Server 2005 Integration Services

ISBN 978-7-121-05259-0

I. S… II. ①哈…②马… III. 关系数据库—数据库管理系统, SQL Server 2005—指南 IV. TP311.138-62

中国版本图书馆 CIP 数据核字(2007)第 166414 号

责任编辑:吴 源

印 刷:北京天竺颖华印刷厂

装 订:三河市金马印装有限公司

出版发行:电子工业出版社

北京市海淀区万寿路 173 信箱 邮编:100036

北京市海淀区翠微东里甲 2 号 邮编:100036

开 本:787×1092 1/16 印张:39.75 字数:1010 千字

印 次:2008 年 1 月第 1 次印刷

定 价:79.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系。联系及邮购电话:(010)88254888。

质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010)88258888。

前 言

Microsoft SQL Server 7.0 和 SQL Server 2000 均包含名叫 Data Transformation Services(数据转换服务, DTS)的组件。如果读者还不知道 ETL 的意思是 Extract, Transform and Load(提取、转换和载入),那么本书可能不适合你。但如果人们确实很关心 ETL 技术并遇到了 ETL 方面的问题,而且使用的又是 SQL Server 2005,那么本书正好符合你的要求。本书介绍的是 SQL Server 2005 Integration Services(集成服务),它是 DTS 的后续版本。

SQL Server 2005 Integration Services 是由一个专门小组开发的。如果读者此前看过本人编写的关于 SQL Server 2000 的书,从中可知微软公司组建了一个约 30 人的开发队伍,他们全都热衷于 ETL 技术,那么一定对于其中的情况很熟悉了。那本书也是在 Kamal Hathi、Donald Farmer、Eduardo Alvarez Godinez 指导下由本书作者 Kirk Haselden 编写的,该书曾详细介绍了相关的情况。

在我们制定“Yukon”(SQL Server 2005 的前身)书籍的编写计划时,就已经知道许多客户面临着前所未有的大规模的、十分棘手的 ETL 问题。几年前,数据仓库和智能化业务还只是“局部”问题,或者说是“局部问题的解决方案”。例如,销售部门会在本部门的数据仓库中寻找问题的答案,市场部门也会有自己的数据仓库。然而真正的性能管理、智能化业务的发展已经要求建立包含企业全局的视图。客户希望载入的数据大大超过了以往,而且他们提出这种需求的频繁程度也大为增加。

我们知道现在需要重新开发 ETL 工具了。人们希望系统的性能得到本质的提升,希望提供一流的易用性、无与伦比的程序兼容性以及高水平的成品功能。我们需要重新开发 ETL 的各项功能。考虑到业内是哪些用户真正知道如何对大量数据进行大规模的处理,因此我们从编译器开发队伍中寻找了一些程序员和工程师。很幸运,我们找到了 Mike Blaszcak,他曾经为 MFC 的开发做出了重要贡献,也是世界上最出色的程序员之一。Mike 还认识 Kirk,并且把他也引入我们早期的开发队伍中。在短短几年的时间里,Kirk 已经成为 Integration Services 的开发经理,并且由他所带领的开发队伍将该项技术推向了市场。

因此,Kirk 作为读者学习和使用 Integration Services 的引路人是最合适的。当然,他对于 SSIC(SQL Server Integration Services)的每一层面都有深入的理解,这也是我们对这位开发经理所期待的。不仅如此,他对于用户和客户怀有深厚的感情,具体表现在其清晰的写作风格和饶有趣味的内容中。在每一章中,Kirk 都会鼓励并教给读者只有“开发经理”才有的深入见解。请读者对各章的“注意”部分给予足够的重视,从这些内容中可以看到开发者是怎样想的。同时,还要注意每一章开篇的引言,我对于其中许多话都记忆深刻,因为当这些人说这些话的时候,许多情况下我都在场,深感其意义深刻,趣味无穷。

SSIS 对 DTS 进行的提高之一在于可管理性。“配置”(configuration)是 SSIS 中的一个新概念。ETL 开发者通过配置可以把其软件包中的一些元数据(metadata)放在软件包之外,这样系统管理员就可以在开发时对它们进行调整。第 14 章“解决方案的配置与部署”会详细讨论这些技术。第 23 章“数据流任务内幕和调整”介绍了相关的优化和调整技术。SSIS 可以满

是企业对 ETL 的需要,这一章提供了全面发挥 SSIS 软件包性能所需的深入理解的内容和技术细节。

DTS 和 SSIS 的另一个不同在于可扩展性。第 24 章“创建自定义任务”介绍了开发自定义任务的过程。第 25 章“创建自定义数据流组件”则介绍了如何开发自定义组件。从本书出版社的网站 www.sampublishing.com 可以下载本书包含的由 Kirk 在这些章节中开发的所有源代码。本人非常希望使用他的源代码组件从 JPEG 文件中提取 EXIF 信息,因为 Kirk 还是一位摄影发烧友,他在本书中使用的完全是来自现实生活中的实例。

能够与 Kirk 共事多年真是荣幸,希望以后我们仍然能够继续合作。他在本书的编写过程中十分努力,因为他热爱 SSIS 技术和它的客户。希望读者也能喜欢本书,并且从中获得裨益。

Bill Baker
Redmond, WA

作者简介

Kirk Haselden, Integration Services 开发组的开发经理。Kirk 在微软公司已经工作了 10 年,曾经在多个部门中工作,其中包括硬件部门、eHome 部门,在 SQL Server 部门的 Integration Services 组中已经工作了 5 年。Kirk 主要负责 Integration Services 的运行设计和实施。在进入微软公司之前,Kirk 曾经在许多中、小型软件企业中工作,编写过教育软件、牙科软件以及网络在线软件。

谨以此书献给:

Christine、Lindsey、Tyson、Amanda、Jenny 和 Sarah。

致 谢

多年以来,我可能已经看过了数百本书中的致谢,就像接下来的内容一样。但只有现在我才真正理解了那些作者想要表达的真情实感。编写一部技术参考书是一项极富挑战性的工作,需要付出大量的时间、耐心,要持之以恒,而且会影响到作者身边的和最亲近的人们。

首先我要感谢我深爱的、给予我不断支持的、耐心的新娘——Christine,她为这部书做了太多的工作。感谢她把每一顿饭端到我的办公桌上,还有她那充满支持的话语以及无限的耐心。你是我永恒的伴侣。没有你,就没有这一切。感谢我的孩子们,他们是 Lindsey、Tyson、Amanda、Jennifer 和 Sarah,他们牺牲了 6 个月本来可以与父亲玩耍的宝贵时间,因为我要编写这本书。希望我能够找回这些本应由父亲与儿子、女儿共享的时光。

感谢本书的技术编辑 Matt David,感谢他阅读了本书的每一行文字并且审查了每一段代码。他真是一位有才干的专业开发者。感谢 Sergei Ivanov,他在本书的编写以及 Integration Services 的开发过程中都有突出的贡献。感谢他提供的巧妙构思,并且编写了配置文件转换示例。感谢 Mike Blaszcak,感谢你的友谊和指导,以及提供的深入见解和对如何编写技术著作的指导。感谢 Ashvini Sharma,感谢你不断给我支持和鼓励,特别是在编写的最后阶段。此外还要感谢 Integration Services 开发组的其他所有人员,没有这么好的产品,也就不会有这本书的出现。

感谢 Scott Turner、Mark Durley 和 Gunther Beersaerts 对本书进行的审阅和提出的宝贵意见。

关于 Sams 公司的工作人员,我想表达如下谢意:感谢 Neil Rowe、Mark Renfrow 和 Karen Annett 耐心和努力的工作;感谢 Karen 完成了本书的审稿,你提出的更改建议总是那么准确,我对此非常钦佩;感谢 Kamal Hathi、Bill Baker 和 Tom Casey,我的经理和总经理,感谢你们给予我的支持并且为我提供了优越的空间;感谢 Euan Garden 为编写本书所主持的白板讨论工作;感谢 Richard Waymire 的友谊、指导,是他邀请我编写了本书。

感谢 Scott Stout 的友谊和支持;感谢 Craig Osterloh 改变了我原来的税务记账生涯。

最后,感谢我的父母给予我的支持。

我们想听听你的意见！

作为本书的读者,你是我们最主要的评论员。我们很重视你的意见,并且想知道哪些是做得好的,如何才能做得更好,你在本书中想看到哪些内容,不愿意看到哪些内容。

作为 Sams 出版社的总编,我欢迎你提出批评和建议。可以通过电子邮件或直接写信给我,告诉我是否喜欢这本书以及如何才能使本书做得更好。

请注意,我无法对你提出的与本书相关的技术问题给出回答。但我们有一个用户服务组,他们会给出本书相关技术问题的解答。

你在写信时,一定要写清这本书的题目、作者、你的姓名、Email 地址以及电话号码。我会认真阅读你的意见,并将其转达给本书的作者和编辑。

Email: feedback@sampublishing.com

Mail:

Paul Boger

Publisher

Sams Publishing

800 East 96th street

Indianapolis, IN 46240 USA

至于这本书更多的信息或 Sams 出版社的其他书目,请访问我们的网站 www.sampublishing.com。在搜索字段中输入 ISBN 0672327813 或书名来搜索你要找的那一页。

目 录

第一部分 入 门

第 1 章 欢迎使用 SQL Server Integration Services	2	帮助资源	13
什么是 SQL Server Integration Services		SQL Server 2005 安装和升级论坛	14
数据导入/导出向导	2	联机丛书	14
ETL 工具	2	SQLIS.COM	14
控制流引擎	2	MSDN	14
应用程序平台	3	SQL Server Integration Services 论坛	14
高性能的数据转换数据管线	3	小结	14
Integration Services 简史	3	第 3 章 从 DTS 迁移	15
有哪些新技术(或者说,这并不是上一代的 DTS!)	4	影响的范围	15
本书的组织方式	4	SSIS 不是 DTS	15
项目示例	5	区别的范围	16
第 2 章 设置 Integration Services	7	好消息	16
安装 Integration Services(IS)	7	可用迁移资源	17
安装 Integration Services	7	升级顾问	17
外围应用配置	10	Execute DTS 2000 Package Task(执行 DTS 2000 包任务)	21
服务和连接的外围应用配置器	10	迁移向导	21
功能的外围应用配置器	11	关于迁移工作的一些考虑	24
设置丛书示例	12	SQL Server 实例和企业管理器	24
各项功能都在哪里	12	Meta Data Services	25
商业智能开发工作室	12	迁移带密码的包	25
管理工作室	12	小结	27
导入和导出向导	13	第 4 章 建立一个包——快速入门	28
迁移向导	13	快速入门	28
DTEExec 和 DTUtil	13	情景举例	28
DTEExecUI	13	配置数据流	31
		小结	40

第二部分 Integration Services 基础和概念

第 5 章 Business Intelligence Development Studio	42	入门	43
		“起始页”和“社区”菜单	43

新建解决方案	44	容器的概念	84
Integration Services 的选项	45	组织	84
SSIS 菜单	45	降低复杂性	85
工具箱	46	作用域划分	86
设计器	47	事务的作用域	87
Visual Studio 的标准选项	61	配置的作用域	87
解决方案资源管理器	62	变量的作用域	87
“属性”窗口	68	执行的作用域	87
“变量”窗口	68	连接的作用域	88
“文件”菜单	68	日志记录的作用域	88
“选项”对话框	69	事件处理器的作用域	88
“调试”窗口	70	容器公共属性	88
“调用堆栈”窗口	72	“标识”	88
小结	75	“执行”	89
第6章 Integration Services 的基本		“强制执行”	90
构件	76	“事务”	90
Integration Services 的基本构件	76	其他类别中的属性	91
包和其他对象	76	设计器分组——非容器	91
任务	77	任务和 Taskhost 容器	91
变量	77	属性集	92
连接管理器	78	持续性	92
日志提供程序	79	包的路径和配置	92
容器	79	调试功能	93
优先约束	79	禁用控制流	93
数据流任务、适配器和转换	80	自定义注册	94
数据流适配器	80	联系信息和良好加载故障恢复	95
将数据移入或移出数据流缓冲区	81	隔离	96
数据流转换	81	对敏感数据的保护	97
数据流与工作流——基本区别	81	简单的“序列容器”	97
执行模型	82	循环容器	99
连接线	83	“包”容器	99
功能范围	83	包的重要属性	101
小结	83	小结	102
第7章 用容器组合控制流	84		

第三部分 控制流服务

第8章 存储区任务	104	标准选项卡	106
普通任务约定	104	任务组	107
任务用户界面	104	工作流任务	107
访问方法	106	执行 DTS 2000 包任务	107

“发送邮件任务”	112	连接管理器的设置	168
“FTP 任务”	114	HTTP 连接管理器	168
“消息队列任务”	117	MSMQ 连接管理器	170
“WMI 任务”	121	微软分析服务连接管理器	170
WMI 数据读取器任务	123	MultiFile 连接管理器	171
WMI 事件观察器任务	126	MultiFlatFile 连接管理器	172
Web 服务任务	128	ODBC 连接管理器	173
执行包任务	130	OleDb 连接管理器	174
执行进程任务	133	SMOServer 连接管理器	176
执行 SQL 任务	135	SMTP 连接管理器	176
脚本任务	139	WMI 连接管理器	177
ActiveX 脚本任务	140	数据源和数据源视图	178
数据处理任务	140	“数据源”和“数据源视图”简介	178
大容量插入任务	140	小结	183
XML 任务	141	第 11 章 日志记录和存储区日志提供	
数据流任务	143	程序	184
小结	143	基础知识	184
第 9 章 表达式的用法	144	快速入门	185
表达式基础知识	144	包日志记录的配置	185
表达式函数	145	存储区日志提供程序	188
列	146	Windows 事件日志提供程序	188
属性表达式	146	文本文件日志提供程序	188
简单举例	146	XML 日志提供程序	188
数据流属性表达式	149	SQL Server 日志提供程序	189
添加数据流属性表达式	150	SQL 探查器日志提供程序	189
变量表达式	151	设计器日志事件查看器——“虚拟的日	
优先约束表达式	152	志提供程序”	189
For 循环表达式	153	日志记录提供程序和日志记录	
小结	153	选项	190
第 10 章 存储区连接管理器	155	日志提供程序	190
连接管理器	155	日志记录选项	190
创建连接管理器	157	日志记录架构	191
存储区连接管理器	158	日志事件	192
ADO 连接管理器	158	自定义日志记录	194
ADO.NET 连接管理器	159	使用“脚本任务”和“脚本注释”	194
Excel 连接管理器	160	使用“SQL 任务”	194
文件连接管理器	162	小结	196
平面文件连接管理器	163	第 12 章 古老的变量	197
连接管理器的设置:高级	166	变量的重要概念——属性、作用域、	
FTP 连接管理器	168	类型和名称空间	198

变量的属性	198	For 循环	208
变量的作用域	199	While 循环	210
变量类型	202	Foreach 循环	210
变量的名称空间	203	存储区 Foreach 枚举器	213
系统变量	204	Foreach 文件枚举器	213
创建和使用变量	205	Foreach Item	214
运行时行为	205	Foreach ADO	215
关于变量的提示和技巧	206	Foreach ADO.NET 架构行集	217
在“执行 SQL 任务”中使用带参数的		Foreach 源变量	220
“EvaluateAsExpression”	206	Foreach NodeList	220
在子包中隐藏父包中的变量	206	Foreach SMO	223
在变量中存储对象	207	小结	224
小结	207		
第 13 章 循环和 Foreach 枚举器	208		

第四部分 管理服务

第 14 章 解决方案的配置与部署	226	“脚本任务”的使用	256
包的配置	226	Dts 对象	256
包配置的工作方式	227	建议实践	258
包配置组织程序	230	脚本任务示例	262
包配置类型	232	在 SSIS 对象模型上工作	262
XML 配置文件	232	使用 HTTP 下载文件	263
配置编辑器	233	把文本文件的内容装入一个变量	264
环境变量配置	234	简单的自定义日志记录	265
注册表项配置	236	异常处理	265
父包变量配置	237	安全消息框	266
SQL Server 表配置	238	小结	267
包配置小结	240	第 16 章 源代码管理的使用	268
包的部署	241	什么是源代码管理	268
用于部署的实用程序	242	源代码库	269
配置和部署的方法	247	同步	269
与位置相关的和驻留在计算机上的		签出/签入	269
配置	247	对比	269
移动的数据库对象	247	添加标签	269
配置替换	248	分支	269
配置连接管理器	249	合并	270
小结	249	BIDS 设计器中的源代码管理	270
第 15 章 “脚本任务”的使用	250	设置	270
脚本任务环境	250	使用源代码管理程序	273
“脚本任务”编辑器	251	实践	277

小结	278	重新启动服务器	290
第17章 SQL Server Management Studio	279	服务器配置	291
SQL Server Management Studio		小结	294
综述	279	第18章 包的安全保护	296
对象资源管理器	279	控制对包的访问	297
查看 Windows 事件日志	282	包存储在 Integration Services 服务器	297
连接到 Integration Services 服务器	282	存储在 SQL Server 中	298
启动“Integration Services 服务器”	282	标识敏感数据	300
连接到本地服务器	284	保护包及其他文件	301
连接到远程 SSIS 服务器	284	包的保护级别	302
包的管理	284	保护非包文件	304
包的运行	284	与“SQL Server 代理”集成	305
已存储的包	285	设置“代理”(Agent)作业来运行 SSIS 包	306
DTS 2000 包的管理	287	代理子系统的路径	310
服务器缓冲存储	287	检测对包的修改	311
使用“SQL Server 代理”程序执行包	288	安全场景设置	312
配置 Integration Services 服务器	290	三层操作结构	313
		小结	314

第五部分 数据流任务

第19章 “数据流任务”介绍	316	源适配器	330
管道思想 101	316	DataReader 源	330
低附加开销转换	317	“OLE DB 源”和“OLE DB 目标”适配器	337
高性能——没有复制,不开玩笑	317	“原始文件源”和“原始文件目标”适配器	339
数据沙箱	317	“XML 源”适配器	339
扩展性	317	流控制转换	341
数据流术语	317	条件性拆分	341
数据流属性	319	派生列	343
可使用表达式的组件属性	319	合并	346
存储路径	320	合并联接	347
缓冲区大小属性	321	多播	349
引擎线程	321	Union All	349
RunInOptimizedMode	321	“查找”转换	351
小结	322	“导入列”转换	355
第20章 存储区中的“数据流”组件	323	“导出列”转换	358
普通组件约定	323		
自定义组件编辑器和高级编辑器	323		

审核转换	359	渐变维度向导	393
“审核”转换	360	生成的数据流	398
“百分比抽样”和“行抽样”转换	361	“渐变维度”组件如何处理数据行	400
“行计数”转换	362	性能方面的考虑	400
数据管理转换	363	数据净化组件	401
“聚合”转换	363	模糊查找	402
“字符映射表”转换	365	“模糊分组”转换	405
“复制列”转换	367	设置“模糊分组”转换组件	406
“数据转换”组件	368	文本挖掘和文档仓库组件	409
“OLE DB 命令”转换	368	文本存在的问题	410
“排序”转换	371	“字词提取”转换	411
目标适配器	372	字词提取转换编辑器	411
“DataReader 目标”适配器	372	“字词查找”转换	412
“记录集目标”适配器	372	小结	414
“SQL Server 目标”适配器	374	第 23 章 数据流任务内幕和调整	415
小结	375	“数据流任务”揭秘	415
第 21 章 使用“脚本组件”	376	基本概念	415
“脚本组件”介绍	376	布局子系统	416
“脚本组件”具有高性能的特点	376	缓存	416
建立概念验证模块	377	缓存子系统	418
访问不常见的或者不支持的数 据源	377	运行引擎	419
创建源适配器	377	用于提速的总体设置	423
创建组件	378	消除不必要的处理	424
创建目标适配器	381	数据流优化	425
创建组件	383	插入性能	429
创建转换	385	“数据流”优化方法	430
创建组件	385	约束理论	430
异步输出	388	资源约束的识别和优先级确定	430
让脚本输出成为异步输出	388	确定如何利用约束	434
小结	389	根据前面的决策对其他部分进行协调 和同步	434
第 22 章 高级数据流转换	390	提升瓶颈因素的性能	434
渐变维度向导程序和转换	390	如果在上述任何步骤中约束发生了改 变,那么回到第一步	435
典型的渐变维度	391	小结	435
其他渐变维度类型	392		

第六部分 Integration Services 编程

第 24 章 创建自定义任务	438	决定该编写什么	440
自定义运行时组件	438	启动 SSIS 自定义组件项目	441
创建自定义组件的要求	439	创建解决方案	442

添加强名称和密钥文件	442	JPG 文件 EXIF 读取器源适配器	491
引用 SSIS 及其他程序集	443	数据配置转换	492
定义类	444	SQL 客户和 ODBC 目标适配器	493
编译和安装	446	示例转换组件的用户界面	493
运行环境如何找到安装的任务	448	设计时方法	493
任务的开发和调试	449	ProvideComponentProperties	494
让任务具有一个身份	449	ReinitializeMetaData	494
添加针对具体功能的属性和函数	450	Validate	494
添加 SSIS 的专用属性及功能	453	运行时方法	495
创建任务用户界面	467	编写源适配器	496
创建示例任务的用户界面(简单)	468	设置和设计时方法	496
创建复杂的任务用户界面(复杂)	472	编写目标适配器	506
自定义任务的最佳实践	484	目标适配器的基本处理方式	506
约定和约束	484	“ReinitializeMetaData”方法	508
任务用户界面	485	“PreExecute”方法	510
通用原则	486	“ProcessInput”方法	514
小结	486	从输入缓存获取数据	514
第 25 章 创建自定义数据流组件	487	编写转换组件	515
自定义组件的概念	487	小结	519
设计时	488	附录 A 错误、警告和信息事件	520
运行时	490	附录 B 表达式求值器参考表	618
示例组件	491		

第一部分 入门

第 1 章 欢迎使用 SQL Server Integration Services

第 2 章 设置 Integration Services

第 3 章 从 DTS 迁移

第 4 章 建立一个包——快速入门

第 1 章 欢迎使用 SQL Server Integration Services

本章要点:

- 什么是 SQL Server Integration Services
- Integration Services 简史
- 有哪些新技术? (或者说,这并不是上一代的 DTS!)
- 本书的组织方式
- 项目示例

“我用自己的脸省去了一张照片!”——Mark Durley

本章将简要介绍 Integration Services(集成服务)、它的起源、简单的历史,并且从较高的层次说明什么是 Integration Services 以及如何使用。

什么是 SQL Server Integration Services

针对这个问题,不同的人会有不同的答案。有人会说数据导入/导出向导程序,有人会说 ETL 工具,有人会说控制流引擎,有人会说应用程序平台,或者是一个高性能的数据转换管线。这些答案都是正确的,因为 Integration Services 就是由一组实用程序、应用程序、设计程序、组件和服务组成的,它们都打包在一个功能强大的软件应用程序集中。SQL Server Integration Services(SQL Server 集成服务,SSIS)对于不同的人有不同的含义。

数据导入/导出向导

Integration Services 中最常使用的一项功能就是“导入/导出向导”程序,它可以方便地把数据从一个源位置,例如一个平面文件或者数据库表,转移到另一个平面文件、表或者其他目标位置。“导入/导出向导”程序是早在 SQL Server 7.0 时就已经开发出来的第一个实用程序,作为数据库管理员工具箱中的重要实用程序一直沿用到了现在。

ETL 工具

ETL 是 Extract(提取)、Transform(转换)和 Load(载入)的首字母缩写,指的是在数据仓库环境中从某个来源事务系统中提取数据,进行转换、清理、消除冗余,再对数据进行整合,最后将数据载入 Cube 或其他分析系统中。尽管 Integration Services 的前身[即数据转换服务(Data Transformation Services,DTS)]也是进行 ETL 处理的出色工具,但 Integration Services 才是 SQL Server 中提供的真正的企业级 ETL。

控制流引擎

将数据从一个位置转移到另一个位置的同时对它们进行转换,这种处理并不仅限于数据