

# Contents

<b>Acknowledgments</b>	<b>ix</b>
<b>Introduction</b>	<b>xvii</b>
<b>Chapter 1: Getting Started</b>	<b>1</b>
<b>Choosing the Right Tool for the Job</b>	<b>1</b>
Data Warehousing ETL	4
Data Integration	6
SSIS Administration	6
<b>SSIS Review</b>	<b>7</b>
Creating a Connection Manager	7
Using the Control Flow	8
Working in the Data Flow	8
Precedence Constraints	15
Package Execution	17
Containers	19
Review Conclusion	20
<b>Summary</b>	<b>20</b>
<b>Chapter 2: Extending Scripts in SSIS</b>	<b>21</b>
<b>Script Tasks and Custom Libraries</b>	<b>22</b>
User-Defined Variables	22
Retrieving Variables with Code	24
Accessing Variables in the Data Flow	26
<b>Building a Custom Assembly</b>	<b>27</b>
Downloading Files over HTTP	29
Adding Assemblies to the GAC	30
<b>Making Connections Configurable and Dynamic</b>	<b>35</b>
<b>Raising Error Events</b>	<b>37</b>
<b>Encrypting Data Through Script Components</b>	<b>39</b>
<b>Data Profiling</b>	<b>48</b>
<b>Summary</b>	<b>52</b>

<b>Chapter 3: Data Extraction</b>	<b>55</b>
<b>Package Connections and Data Flow Sources</b>	<b>56</b>
Source Adapters	58
Advanced Features and Concepts	63
Optimal Data Staging Method	66
<b>Incremental Data Extraction</b>	<b>68</b>
Incremental Extraction Using a Change Identifier Value	69
Incremental Extraction from SQL Server Without a Trigger	79
Using SSIS to Handle All Aspects of an Incremental Extraction	81
<b>Tracking Data Lineage Identifiers</b>	<b>86</b>
<b>Summary</b>	<b>89</b>
<b>Chapter 4: Dimension ETL with SSIS</b>	<b>91</b>
<b>Dimension ETL Overview</b>	<b>91</b>
<b>Dimensions: The Basics</b>	<b>92</b>
<b>Dimension ETL: The Challenge</b>	<b>96</b>
Preparing Your Source Data for Dimension ETL	97
Dimension Change Types	102
<b>SSIS Slowly Changing Dimension Wizard</b>	<b>107</b>
Advanced Properties and Additional Outputs of the SCD	118
Slowly Changing Dimension Wizard Advantages and Disadvantages	120
Optimizing the Built-in Slowly Changing Dimension Support	121
Handling Advanced Dimension Processing with the Slowly Changing Dimension Support	122
<b>Creating a Custom Slowly Changing Package</b>	<b>133</b>
Joining Source Data to Dimension Data	133
Determining Dimension Changes	137
Handling Dimension Inserts and Updates	138
<b>Summary</b>	<b>141</b>
<b>Chapter 5: Fact Table ETL</b>	<b>143</b>
<b>Fact Table Overview</b>	<b>143</b>
Mapping Dimension Keys	144
Calculating Measures	145
Adding Metadata	145
Fact Table Types	145
<b>Fact Table ETL</b>	<b>147</b>
<b>Fact Table ETL Challenge</b>	<b>147</b>

<b>Fact Table ETL Basics</b>	<b>148</b>
Acquiring the Dimension Surrogate Key	148
Measure Calculations	159
Managing Fact Table Changes	161
<b>Advanced Fact Table ETL Concepts</b>	<b>172</b>
Managing Fact Table Grain	173
Grain Change SSIS Example	175
Handling Missing Dimension Lookups	181
Handling Late-Arriving Facts	187
Advanced Fact Table Loading	190
<b>Summary</b>	<b>190</b>
<b>Chapter 6: Processing Analysis Services Objects with SSIS</b>	<b>191</b>
<hr/>	
<b>SSAS ETL Processing and Management Overview</b>	<b>191</b>
SSAS Objects and Processing Basics	192
Methods to Process SSAS Objects Through SSIS	197
Creating and Modifying Partitions	197
<b>SSAS Integration Basics Within SSIS</b>	<b>198</b>
Control Flow Tasks for SSAS	199
Data Flow Destinations for SSAS Objects	204
<b>Advanced Processing and Partition Management Examples</b>	<b>208</b>
Dimension Processing	209
Partition Creation and Processing	211
Partition Management	217
Processing SSAS Cubes from Non-SQL Server Sources	220
Implications of Type 1, Type 2, and Inferred Members on Attribute Relationships	222
<b>Summary</b>	<b>223</b>
<b>Chapter 7: Package Reliability</b>	<b>225</b>
<hr/>	
<b>Error and Event Handling</b>	<b>225</b>
Types of Event Handlers	225
Auditing Through Event Handlers	226
Disabling Event Handling	229
Integrating Snapshots into SSIS	229
<b>Logging</b>	<b>233</b>
Creating a Unique Log File	236
Reporting on the Logging Providers	237

# Contents

---

<b>Checkpoint Files</b>	<b>238</b>
Dynamic Checkpoint File Name	240
Testing Your Error Logic	241
<b>Transactions</b>	<b>242</b>
<b>Raw Files</b>	<b>245</b>
<b>Proactive WMI Integration</b>	<b>251</b>
<b>Building a File Watcher Task</b>	<b>252</b>
<b>Summary</b>	<b>254</b>
<b>Chapter 8: Deployment</b>	<b>255</b>
<b>Working with a Team in SSIS</b>	<b>255</b>
Source Control Integration	255
Adding a New Project	256
<b>Reusable Packages</b>	<b>259</b>
<b>Creating a Package Template</b>	<b>260</b>
<b>Package Configuration</b>	<b>261</b>
SSIS Package Configurations	261
Configuration Repository	266
<b>Deployment Utility</b>	<b>274</b>
<b>The Package Installation Wizard</b>	<b>275</b>
<b>Summary</b>	<b>279</b>
<b>Chapter 9: Managing SSIS</b>	<b>281</b>
<b>The Package Store</b>	<b>281</b>
Creating a Central SSIS Server	283
Clustering SSIS	285
File System or the msdb Deployment	287
<b>Management Studio</b>	<b>288</b>
<b>Running Packages with DTEXECUI</b>	<b>290</b>
<b>Security</b>	<b>297</b>
<b>Firewall Issues</b>	<b>299</b>
<b>Command Line Utilities</b>	<b>299</b>
DTEXEC.EXE	300
DTUTIL.EXE	300
<b>Scheduling a Package</b>	<b>301</b>
<b>Proxy Accounts</b>	<b>303</b>
<b>64-Bit Issues</b>	<b>306</b>
<b>Performance Counters</b>	<b>308</b>
<b>Summary</b>	<b>309</b>

---

<b>Chapter 10: Handling Heterogeneous and Unusual Data</b>	<b>311</b>
<b>Unusual Data Flow Scenarios</b>	<b>311</b>
Creating Rows from Columns	311
Multiple Record Types in a Single File	316
Using the Raw File	323
<b>Oracle</b>	<b>326</b>
Reading from Oracle	327
Writing Data to Oracle	330
<b>Other Data Sources</b>	<b>331</b>
DB2	331
VSAM	332
<b>Summary</b>	<b>333</b>
<b>Chapter 11: Migrating from DTS to SSIS</b>	<b>335</b>
<b>SQL Server 2005 DTS Backward Compatibility</b>	<b>335</b>
<b>Administering and Editing a DTS Package</b>	<b>335</b>
<b>Upgrading from DTS</b>	<b>339</b>
The Upgrade Advisor	340
The Migration Wizard	341
<b>Handling Exceptions</b>	<b>344</b>
Dynamic Properties Task	345
Complex Transform Data Tasks	348
Flat File Connection Managers	350
ActiveX Script Task	351
<b>Summary</b>	<b>352</b>
<b>Chapter 12: Scaling SSIS</b>	<b>353</b>
<b>SSIS Scalability Foundations</b>	<b>354</b>
SSIS Service Status	354
Identifying Task Durations	355
Memory Utilization	357
Balancing SQL Operations with the Data Flow	360
<b>Data Flow Optimization</b>	<b>362</b>
Pipeline Architecture Review	363
General Pipeline Optimization	365
Data Flow Properties	369
Optimizing Destinations	371

# Contents

---

<b>Package Execution Principles</b>	<b>376</b>
Package Storage Location versus Execution Location	376
Execute SQL Task and Bulk Insert Task Execution	377
Package Execution and the Data Flow	378
Packages Executed on the Source or Destination Servers	378
Stand-alone SSIS Servers	380
Distributed Package Execution	381
<b>Summary</b>	<b>383</b>
<b>Index</b>	<b>385</b>