

CONTENTS

| | |
|--|--------------|
| INTRODUCTION | xxvii |
| CHAPTER 1: WELCOME TO SQL SERVER INTEGRATION SERVICES | 1 |
| SQL Server SSIS Historical Overview | 2 |
| What's New in SSIS | 2 |
| Tools of the Trade | 3 |
| Import and Export Wizard | 3 |
| The SQL Server Data Tools Experience | 4 |
| SSIS Architecture | 5 |
| Packages | 5 |
| Control Flow | 5 |
| Data Flow | 9 |
| Variables | 14 |
| Parameters | 14 |
| Error Handling and Logging | 14 |
| Editions of SQL Server | 14 |
| Summary | 15 |
| CHAPTER 2: THE SSIS TOOLS | 17 |
| Import and Export Wizard | 17 |
| SQL Server Data Tools | 24 |
| The Solution Explorer Window | 26 |
| The SSIS Toolbox | 27 |
| The Properties Windows | 28 |
| The SSIS Package Designer | 28 |
| Control Flow | 29 |
| Connection Managers | 32 |
| Variables | 33 |
| Data Flow | 34 |
| Parameters | 35 |
| Event Handlers | 35 |
| Package Explorer | 36 |
| Executing a Package | 37 |
| Management Studio | 37 |
| Summary | 37 |

| | |
|--|-----------|
| CHAPTER 3: SSIS TASKS | 39 |
| SSIS Task Objects | 40 |
| Using the Task Editor | 40 |
| The Task Editor Expressions Tab | 41 |
| Looping and Sequence Tasks | 41 |
| Script Task (.NET) | 41 |
| Analysis Services Tasks | 44 |
| Analysis Services Execute DDL Task | 44 |
| Analysis Services Processing Task | 44 |
| Data Mining Query Task | 46 |
| Data Flow Task | 47 |
| Data Preparation Tasks | 48 |
| Data Profiler | 48 |
| File System Task | 50 |
| Archiving a File | 52 |
| FTP Task | 53 |
| Getting a File Using FTP | 54 |
| Web Service Task | 56 |
| Retrieving Data Using the Web Service Task and XML Source Component | 59 |
| XML Task | 62 |
| Validating an XML File | 64 |
| RDBMS Server Tasks | 66 |
| Bulk Insert Task | 66 |
| Using the Bulk Insert Task | 69 |
| Execute SQL Task | 71 |
| Workflow Tasks | 82 |
| Execute Package Task | 82 |
| Execute Process Task | 84 |
| Message Queue Task | 86 |
| Send Mail Task | 87 |
| WMI Data Reader Task | 88 |
| WMI Event Watcher Task | 91 |
| Polling a Directory for the Delivery of a File | 91 |
| SMO Administration Tasks | 92 |
| Transfer Database Task | 93 |
| Transfer Error Messages Task | 94 |
| Transfer Logins Task | 94 |
| Transfer Master Stored Procedures Task | 95 |
| Transfer Jobs Task | 96 |
| Transfer SQL Server Objects Task | 96 |
| Summary | 97 |

| | |
|---|-----------|
| CHAPTER 4: THE DATA FLOW | 99 |
| Understanding the Data Flow | 99 |
| Data Viewers | 100 |
| Sources | 101 |
| OLE DB Source | 102 |
| Excel Source | 104 |
| Flat File Source | 105 |
| Raw File Source | 110 |
| XML Source | 110 |
| ADO.NET Source | 111 |
| Destinations | 111 |
| Excel Destination | 112 |
| Flat File Destination | 112 |
| OLE DB Destination | 112 |
| Raw File Destination | 113 |
| Recordset Destination | 114 |
| Data Mining Model Training | 114 |
| DataReader Destination | 114 |
| Dimension and Partition Processing | 114 |
| Common Transformations | 115 |
| Synchronous versus Asynchronous Transformations | 115 |
| Aggregate | 115 |
| Conditional Split | 117 |
| Data Conversion | 118 |
| Derived Column | 119 |
| Lookup | 121 |
| Row Count | 121 |
| Script Component | 122 |
| Slowly Changing Dimension | 123 |
| Sort | 123 |
| Union All | 125 |
| Other Transformations | 126 |
| Audit | 126 |
| Character Map | 128 |
| Copy Column | 128 |
| Data Mining Query | 129 |
| DQS Cleansing | 130 |
| Export Column | 130 |
| Fuzzy Lookup | 132 |
| Fuzzy Grouping | 139 |
| Import Column | 143 |
| Merge | 145 |

| | |
|--|------------|
| Merge Join | 146 |
| Multicast | 146 |
| OLE DB Command | 147 |
| Percentage and Row Sampling | 148 |
| Pivot Transform | 149 |
| Unpivot | 152 |
| Term Extraction | 154 |
| Term Lookup | 158 |
| Data Flow Example | 160 |
| Summary | 164 |
| CHAPTER 5: USING VARIABLES, PARAMETERS, AND EXPRESSIONS | 165 |
| <hr/> | |
| Dynamic Package Objects | 166 |
| Variable Overview | 166 |
| Parameter Overview | 166 |
| Expression Overview | 167 |
| Understanding Data Types | 168 |
| SSIS Data Types | 168 |
| Date and Time Type Support | 170 |
| How Wrong Data Types and Sizes Can Affect Performance | 171 |
| Unicode and Non-Unicode Conversion Issues | 171 |
| Casting in SSIS Expressions | 173 |
| Using Variables and Parameters | 174 |
| Defining Variables | 174 |
| Defining Parameters | 175 |
| Variable and Parameter Data Types | 176 |
| Working with Expressions | 177 |
| C#-Like? Close, but Not Completely | 178 |
| The Expression Builder | 179 |
| Syntax Basics | 180 |
| Using Expressions in SSIS Packages | 194 |
| Summary | 204 |
| CHAPTER 6: CONTAINERS | 205 |
| <hr/> | |
| Task Host Containers | 205 |
| Sequence Containers | 206 |
| Groups | 207 |
| For Loop Container | 207 |
| Foreach Loop Container | 210 |

| | |
|--|------------|
| Foreach File Enumerator Example | 211 |
| Foreach ADO Enumerator Example | 213 |
| Summary | 218 |
| CHAPTER 7: JOINING DATA | 219 |
| <hr/> | |
| The Lookup Transformation | 220 |
| Using the Merge Join Transformation | 221 |
| Contrasting SSIS and the Relational Join | 222 |
| Lookup Features | 224 |
| Building the Basic Package | 225 |
| Using a Relational Join in the Source | 227 |
| Using the Merge Join Transformation | 230 |
| Using the Lookup Transformation | 235 |
| Full-Cache Mode | 235 |
| No-Cache Mode | 239 |
| Partial-Cache Mode | 240 |
| Multiple Outputs | 243 |
| Expressionable Properties | 246 |
| Cascaded Lookup Operations | 247 |
| Cache Connection Manager and Cache Transform | 249 |
| Summary | 252 |
| CHAPTER 8: CREATING AN END-TO-END PACKAGE | 253 |
| <hr/> | |
| Basic Transformation Tutorial | 253 |
| Creating Connections | 254 |
| Creating the Control Flow | 257 |
| Creating the Data Flow | 257 |
| Completing the Package | 259 |
| Saving the Package | 260 |
| Executing the Package | 260 |
| Typical Mainframe ETL with Data Scrubbing | 261 |
| Creating the Data Flow | 263 |
| Handling Dirty Data | 263 |
| Finalizing | 268 |
| Handling More Bad Data | 269 |
| Looping and the Dynamic Tasks | 271 |
| Looping | 271 |
| Making the Package Dynamic | 272 |
| Summary | 274 |

| | |
|---|------------|
| CHAPTER 9: SCRIPTING IN SSIS | 275 |
| Introducing SSIS Scripting | 276 |
| Getting Started in SSIS Scripting | 277 |
| Selecting the Scripting Language | 277 |
| Using the VSTA Scripting IDE | 278 |
| Example: Hello World | 279 |
| Adding Code and Classes | 281 |
| Using Managed Assemblies | 282 |
| Example: Using Custom .NET Assemblies | 283 |
| Using the Script Task | 286 |
| Configuring the Script Task Editor | 287 |
| The Script Task Dts Object | 288 |
| Accessing Variables in the Script Task | 289 |
| Connecting to Data Sources in a Script Task | 293 |
| Raising an Event in a Script Task | 303 |
| Writing a Log Entry in a Script Task | 309 |
| Using the Script Component | 310 |
| Differences from a Script Task | 310 |
| Configuring the Script Component Editor | 311 |
| Accessing Variables in a Script Component | 313 |
| Connecting to Data Sources in a Script Component | 314 |
| Raising Events | 314 |
| Logging | 315 |
| Example: Data Validation | 316 |
| Synchronous versus Asynchronous | 324 |
| Essential Coding, Debugging, and Troubleshooting Techniques | 327 |
| Structured Exception Handling | 327 |
| Script Debugging and Troubleshooting | 330 |
| Summary | 333 |
| CHAPTER 10: ADVANCED DATA CLEANSING IN SSIS | 335 |
| Advanced Derived Column Use | 336 |
| Text Parsing Example | 338 |
| Advanced Fuzzy Lookup and Fuzzy Grouping | 340 |
| Fuzzy Lookup | 340 |
| Fuzzy Grouping | 347 |
| DQS Cleansing | 350 |
| Data Quality Services | 351 |

| | |
|--|------------|
| DQS Cleansing Transformation | 355 |
| Master Data Management | 358 |
| Master Data Services | 359 |
| Summary | 362 |
| CHAPTER 11: INCREMENTAL LOADS IN SSIS | 363 |
| <hr/> | |
| Control Table Pattern | 363 |
| Querying the Control Table | 364 |
| Querying the Source Table | 366 |
| Updating the Control Table | 366 |
| SQL Server Change Data Capture | 367 |
| Benefits of SQL Server CDC | 368 |
| Preparing CDC | 369 |
| Capture Instance Tables | 371 |
| The CDC API | 372 |
| Using the SSIS CDC Tools | 374 |
| Summary | 379 |
| CHAPTER 12: LOADING A DATA WAREHOUSE | 381 |
| <hr/> | |
| Data Profiling | 383 |
| Initial Execution of the Data Profiling Task | 383 |
| Reviewing the Results of the Data Profiling Task | 386 |
| Turning Data Profile Results into Actionable ETL Steps | 390 |
| Data Extraction and Cleansing | 391 |
| Dimension Table Loading | 391 |
| Loading a Simple Dimension Table | 392 |
| Loading a Complex Dimension Table | 397 |
| Considerations and Alternatives to the SCD Transformation | 408 |
| Fact Table Loading | 409 |
| SSAS Processing | 421 |
| Using a Master ETL Package | 426 |
| Summary | 428 |
| CHAPTER 13: USING THE RELATIONAL ENGINE | 429 |
| <hr/> | |
| Data Extraction | 430 |
| SELECT * Is Bad | 430 |
| WHERE Is Your Friend | 432 |
| Transform during Extract | 433 |

| | |
|---|------------|
| Many ANDs Make Light Work | 437 |
| SORT in the Database | 437 |
| Modularize | 439 |
| SQL Server Does Text Files Too | 440 |
| Using Set-Based Logic | 444 |
| Data Loading | 446 |
| Database Snapshots | 446 |
| The MERGE Operator | 448 |
| Summary | 452 |
| CHAPTER 14: ACCESSING HETEROGENEOUS DATA | 453 |
| <hr/> | |
| Excel and Access | 455 |
| 64-Bit Support | 455 |
| Working with Excel Files | 457 |
| Working with Access | 462 |
| Importing from Oracle | 469 |
| Oracle Client Setup | 469 |
| Importing Oracle Data | 470 |
| Using XML and Web Services | 472 |
| Configuring the Web Service Task | 472 |
| Working with XML Data as a Source | 483 |
| Flat Files | 486 |
| Loading Flat Files | 487 |
| Extracting Data from Flat Files | 489 |
| ODBC | 491 |
| Other Heterogeneous Sources | 494 |
| Summary | 495 |
| CHAPTER 15: RELIABILITY AND SCALABILITY | 497 |
| <hr/> | |
| Restarting Packages | 498 |
| Simple Control Flow | 499 |
| Containers within Containers and Checkpoints | 501 |
| Variations on a Theme | 503 |
| Inside the Checkpoint File | 505 |
| Package Transactions | 507 |
| Single Package, Single Transaction | 508 |
| Single Package, Multiple Transactions | 509 |
| Two Packages, One Transaction | 511 |
| Single Package Using a Native Transaction in SQL Server | 512 |

| | |
|--|------------|
| Error Outputs | 513 |
| Scaling Out | 516 |
| Architectural Features | 516 |
| Scaling Out Memory Pressures | 517 |
| Scaling Out by Staging Data | 517 |
| Scaling Out with Parallel Loading | 522 |
| Summary | 528 |
| CHAPTER 16: UNDERSTANDING AND TUNING THE DATA FLOW ENGINE | 529 |
| <hr/> | |
| The SSIS Engine | 530 |
| Understanding the SSIS Data Flow and Control Flow | 530 |
| Handling Workflows with the Control Flow | 533 |
| Data Processing in the Data Flow | 533 |
| Memory Buffer Architecture | 534 |
| Types of Transformations | 534 |
| Advanced Data Flow Execution Concepts | 543 |
| SSIS Data Flow Design and Tuning | 549 |
| Data Flow Design Practices | 550 |
| Optimizing Package Processing | 555 |
| Troubleshooting Data Flow Performance Bottlenecks | 558 |
| Pipeline Performance Monitoring | 559 |
| Summary | 562 |
| CHAPTER 17: SSIS SOFTWARE DEVELOPMENT LIFE CYCLE | 563 |
| <hr/> | |
| Introduction to Software Development Life Cycles | 565 |
| SDLCs: A Brief History | 566 |
| Types of Software Development Life Cycles | 566 |
| Versioning and Source Code Control | 567 |
| Subversion (SVN) | 568 |
| Team Foundation Server, Team System, and SSIS | 573 |
| Summary | 590 |
| CHAPTER 18: ERROR AND EVENT HANDLING | 591 |
| <hr/> | |
| Using Precedence Constraints | 592 |
| Precedence Constraint Basics | 592 |
| Advanced Precedence Constraints and Expressions | 593 |
| Event Handling | 601 |
| Events | 602 |

| | |
|--|------------|
| Using Event Handlers | 603 |
| Event Handler Inheritance | 611 |
| Breakpoints | 612 |
| Error Rows | 616 |
| Logging | 622 |
| Logging Providers | 622 |
| Log Events | 623 |
| Catalog Logging | 627 |
| Summary | 629 |
| CHAPTER 19: PROGRAMMING AND EXTENDING SSIS | 631 |
| <hr/> | |
| The Sample Components | 632 |
| Component 1: Source Adapter | 632 |
| Component 2: Transform | 633 |
| Component 3: Destination Adapter | 634 |
| The Pipeline Component Methods | 634 |
| Design-Time Functionality | 635 |
| Runtime | 639 |
| Connection Time | 640 |
| Building the Components | 642 |
| Preparation | 642 |
| Building the Source Component | 648 |
| Building the Transformation Component | 660 |
| Building the Destination Adapter | 671 |
| Using the Components | 679 |
| Installing the Components | 679 |
| Debugging Components | 680 |
| Design Time | 680 |
| Building the Complete Package | 682 |
| Runtime Debugging | 682 |
| Upgrading to SQL Server 2014 | 687 |
| Summary | 687 |
| CHAPTER 20: ADDING A USER INTERFACE TO YOUR COMPONENT | 689 |
| <hr/> | |
| Three Key Steps for Designing the UI: An Overview | 690 |
| Building the User Interface | 690 |
| Adding the Project | 691 |
| Implementing IDtsComponentUI | 693 |
| Setting the UITypeName | 697 |
| Building the Form | 699 |

| | |
|--|------------|
| Extending the User Interface | 704 |
| Runtime Connections | 704 |
| Component Properties | 707 |
| Handling Errors and Warnings | 708 |
| Column Properties | 711 |
| Other UI Considerations | 712 |
| Summary | 712 |
| | |
| CHAPTER 21: EXTERNAL MANAGEMENT AND WMI TASK IMPLEMENTATION | 715 |
| <hr/> | |
| External Management of SSIS with Managed Code | 716 |
| Setting Up a Test SSIS Package for Demonstration Purposes | 716 |
| The Managed Object Model Code Library | 717 |
| Catalog Management | 718 |
| Folder Management | 719 |
| Environments | 720 |
| The DTS Runtime Managed Code Library | 722 |
| SSIS Deployment Projects | 722 |
| Parameter Objects | 723 |
| Server Deployment | 725 |
| Executing SSIS Packages Deployed to the SSIS Catalog | 726 |
| Environment References | 727 |
| Package Operations | 728 |
| Application Object Maintenance Operations | 729 |
| Package Operations | 729 |
| Package Monitoring | 732 |
| Project, Folder, and Package Listing | 734 |
| A Package Management Example | 735 |
| Package Log Providers | 745 |
| Specifying Events to Log | 747 |
| Programming to Log Providers | 748 |
| SQL Server 2014 Operation Logs | 749 |
| Package Configurations | 751 |
| Creating a Configuration | 752 |
| Programming the Configuration Object | 753 |
| Configuration Object | 754 |
| Windows Management Instrumentation Tasks | 755 |
| WMI Reader Task Explained | 755 |
| WMI Data Reader Example | 756 |
| WMI Event Watcher Task | 762 |
| WMI Event Watcher Task Example | 763 |
| Summary | 766 |

| | |
|---------------------------------------|------------|
| CHAPTER 22: ADMINISTERING SSIS | 767 |
| Using the SSIS Catalog | 768 |
| Setting the SSIS Catalog Properties | 768 |
| SSISDB | 771 |
| Deployment Models | 772 |
| Project Deployment Model | 773 |
| Package Deployment Model | 775 |
| Using T-SQL with SSIS | 781 |
| Executing Packages | 781 |
| Using Parameters | 782 |
| Querying Tables for Parameter Values | 783 |
| Using Environments | 784 |
| Using Data Taps | 789 |
| Creating a Central SSIS Server | 790 |
| Clustering SSIS | 792 |
| Package Configuration | 794 |
| Command-Line Utilities | 798 |
| DTEXec | 798 |
| DTEXecUI | 799 |
| DTUtil | 804 |
| Security | 806 |
| Securing the SSIS Catalog | 806 |
| Legacy Security | 809 |
| Scheduling Packages | 811 |
| SQL Server Agent | 811 |
| Proxy Accounts | 813 |
| 64-Bit Issues | 814 |
| Monitoring Package Executions | 815 |
| Built-in Reporting | 815 |
| Custom Reporting | 819 |
| Performance Counters | 819 |
| Summary | 820 |
| APPENDIX A: SSIS CRIB NOTES | 821 |
| When to Use Control Flow Tasks | 821 |
| When to Use Data Flow Transforms | 822 |
| Common Expressions and Scripts | 824 |

| | |
|---|------------|
| APPENDIX B: SSIS INTERNAL VIEWS AND STORED PROCEDURES | 829 |
| Views | 829 |
| Stored Procedures | 830 |
| APPENDIX C: INTERVIEWING FOR AN ETL DEVELOPER POSITION | 833 |
| Questions | 833 |
| Answers | 834 |
| INDEX | 839 |