

Index

NUMERICS

3DES (Triple DES), 40

32-bit and 64-bit issues, 306–308

A

Access

Excel source adapter for, 61–63

64-bit issues with, 307

ActiveX Script tasks, migrating to SSIS, 339, 351–352

administration of DTS packages, 335–339

administration of SSIS

command line utilities for, 299–301

firewall issues regarding, 299

in Management Studio

assigning package roles, 297–299

monitoring packages, 288–289

running packages, 290–297

performance counters, 308–309, 357

scheduling packages

with proxy accounts, 303–305

with SQL Server Agent, 301–303

troubleshooting, 303

when using central SSIS server, 284

tasks in, 6–7

ADO.NET Providers

for DB2, 331–332

for Oracle, 327, 328

ADO.NET sources, DataReader source adapter for, 60

Advanced Encryption Standard (AES), 40–44

Aggregate transformation

for data grain changes, 174

for maximum change identifier value, 77

for measure calculations, 160

AMO (Analysis Management Objects)

for partition management, 217, 218–220

processing SSAS objects using, 197, 216–217

Analysis Services Execute DDL Task

for partition management, 217–218

processing SSAS objects using, 197, 203–204

Analysis Services objects (SQL Server 2005

Analysis Services)

AMO (Analysis Management Objects) for, 197

attributes of, affected by ETL processing, 222–223

connection for, 198

control flow tasks for

Analysis Services Execute DDL Task, 197, 203–204

Analysis Services Processing Task, 197, 199–203

Analysis Services objects (SQL Server 2005 Analysis Services) (continued)

Analysis Services objects (SQL Server 2005 Analysis Services) (continued)

- Execute Command Task, 197
- cubes, processing, 192, 220–221
- data flow destinations for
 - Dimension Processing destination, 197, 205–207
 - Partition Processing destination, 197, 207–208
- data mining objects
 - data flow features for, 361
 - processing of, 192, 196
- database, 192
- dimensions, 192–194
- measure groups
 - definition of, 123
 - processing of, 192
- partitions
 - creating and modifying, 197–198
 - processing of, 194–195
- processing methods for, 197
- scheduling processing of, 191–192
- security for, 198
- types of objects processed, 192
- XMLA scripts for, 197, 209–211

Analysis Services Processing Task, 197, 199–203

Analysis Services tasks, migrating to SSIS, 339

archiving

- data staging used for, 66–67
- File System Task for, 8, 352

ASCMD.exe **executable**

- for dimension processing, 209–211
- for partition management, 217–218
- for partition processing, 211–214

assemblies

- adding to GAC, 30–35
- calling from Script Task, 33
- creating, 27–29
- referencing from local directory, 30
- strong name for, 30–31
- testing, 34–35

asymmetric encryption, 39

asynchronous transformation, 45, 360

attributes of dimensions, 92

Attunity Provider

- for DB2, 332
- for Oracle, 328
- for VSAM, 332

auditing

- audit table for, 68
- event handling for, 226–229
- log-based auditing, 68
- package logs used for, 237
- rows to be written, 13
- SCD transformation outputs for, 118–119
- SQL operations used for, 362
- staging table for, 171

auto-incrementing identifier, 68

B

backpressure, 365

backup and recovery

- Package Store location affecting, 287
- snapshots used for, 229

banding, data, 99

BI (business intelligence), 1

BIDS (Business Intelligence Development Studio)

- connection manager, creating, 7
- containers, creating, 19–20
- Control Flow tab, 8, 15–17
- cube designer in, 194–195
- Data Flow tab, 8–15
- dimension editor in, 193
- Migration Wizard, launching, 341
- not supported on Itanium 64-bit machines, 306
- package
 - creating, 7
 - executing, 17–18
 - password for, setting, 299
- PackageID, generating, 261
- source control integration using, 255–259
- user-defined variables, creating, 22–24

BizTalk, 2

BLOBTempStoragePath property, 370

blocking transformations, 363, 365**books**

- The Complete Guide to Dimension Modeling* (Kimball, Ross), 92
- Microsoft Data Warehouse Toolkit* (Wiley), 102
- Professional SQL Server 2005 Integration Services* (Wiley), 7, 21

Buffer Memory counter, 308**buffer spooling, 357****buffers**

- performance counters for, 308–309
- pipelines managed using, 363
- tuning, 370–371

Buffers In Use counter, 308**Buffers Spooled counter, 308–309**

`BufferTempStoragePath` property, 370

Bulk Insert Task

- location where executed, 377
- migrating to SSIS, 339

business intelligence (BI), 1

Business Intelligence Development Studio. See BIDS

business key (candidate key) of dimension table, 95, 144

C**Cache, Global Assembly (GAC), 30–35****cache size, managing, 366–367****cached Lookup transformation, 136–137**

candidate key (business key) of dimension table, 95, 144

change identifier columns

- capturing maximum change identifier value, 76–79
- definition of, 68
- flat files containing, 82–83
- process for using, 69
- targeting change records, 69–72
- updating package variables, 72–76

Change Type 0 (fixed attribute)

- definition of, 102
- implications on SSAS attributes, 222–223

Change Type 1 (changing attribute)

- definition of, 102, 103
- implications on SSAS attributes, 222–223

Change Type 2 (historical attribute)

- definition of, 102, 103–104
- with fact table ETL, 148
- implications on SSAS attributes, 222–223

change types of dimension

- in custom package, 137–141
- implications on SSAS attributes, 222–223
- inferred members, handling with, 104–106
- list of, 102–104
- specifying in SCD Wizard, 108–112
- supported by SCD Wizard, 107

changing attribute (Change Type 1)

- definition of, 102, 103
- implications on SSAS attributes, 222–223

Changing Attributes Updates Output, SCD Wizard, 114

Check constraints option, OLE DB destination, 375

checkpoint files

- definition of, 238
- dynamic names for, 240–241
- error handling for, 241–242
- implementing, 238–240
- using with transactions, 245

class libraries (assemblies)

- adding to GAC, 30–35
- calling from Script Task, 33
- creating, 27–29
- referencing from local directory, 30
- strong name for, 30–31
- testing, 34–35

cleansing of data

- data profiling for, 48–52
- example of, 177–180
- transformations for, 97

clustering SSIS Service, 285–286

code page warnings from Oracle sources, 328

code reusability

- reusable packages, 259–261
- Script Tasks used for, 22

columns

- creating rows from, 311–315
- unused, removing from data flow, 367
- command line utilities for SSIS administration, 299–301**
- common table expression (CTE), 355**
The Complete Guide to Dimension Modeling (Kimball, Ross), 92
- Conditional Split transformation**
 - for change identifier columns, 82–83
 - for data grain changes, 174
 - for identifying fact table changes, 163, 165–166
 - for separating record types, 319
 - for separating valid and invalid data, 51
- configurable connections, 35–37**
- connection manager**
 - administration of, 291
 - creating, 7
 - Flat File Connection Manager
 - for file with multiple record types, 317, 318
 - migration to SSIS, 350–351
 - Local Temp DB Connection Manager, 346
- `Connection` **property, Analysis Services Execute DDL Task, 203**
- connections**
 - choosing when running packages, 291
 - creating, 57
 - delaying validation of, 64
 - dynamic and configurable, 35–37
 - included with Windows and SSIS, 56
 - for SSAS integration, 198
 - updating with property expressions, 64–66
- `ConnectionString` **property, package, 36**
- containers**
 - for SSAS objects, 192
 - types of, 19–20
- Control Flow tab**
 - precedence constraints, creating, 15–17
 - tasks, creating, 8
 - variables, creating, 8
- corrupt packages, handling, 301**
- Create, Read, Update, and Delete (CRUD) operations, 69**
- `CreateSymmetricKey()` **method, 40**

- CRUD (Create, Read, Update, and Delete) operations, 69**
- cryptographic key pair, 30–31**
- CTE (common table expression), 355**
- cubes (SSAS), 192, 220–221**
- `CurrentRowWhere` **property, SCD Wizard, 118**
- custom scripts**
 - Script Components
 - for data grain changes, 175
 - data profiling in, 49–52
 - encrypting data in, 39–48
 - language used in, 24
 - retrieving variables in, 26
 - Script Tasks
 - accessing and updating connections from, 36–37
 - calling assembly from, 33
 - code reusability with, 22
 - definition of, 22
 - language used in, 24
 - raising errors in, 37–39
 - retrieving variables in, 24–26
 - 32-bit and 64-bit issues regarding, 307
 - user-defined variables for, 22–24

D

- data archiving**
 - data staging used for, 66–67
 - File System Task for, 8, 352
- data banding, 99**
- data cleansing**
 - data profiling for, 48–52
 - example of, 177–180
 - transformations for, 97
- Data Conversion transformation, 97, 98–99, 319–321**
- Data Driven Query tasks, migrating to SSIS, 339**
- Data Encryption Standard (DES), 40**
- data extraction**
 - configuring in Data Flow task, 8–9
 - data lineage, 55
 - definition of, 55

incremental extraction
 approaches for, 68
 with audit table, 68
 with auto-incrementing identifier, 68
 with change identifier columns, 68
 correlating source with destination for, 83–86
 definition of, 55, 68
 with log-based auditing, 68
 with `timestamp` data type, 79–81
 objectives of, 55
 process of, 56

data flow destinations

correlating source with, for incremental extraction, 83–86
 Dimension Processing destination, 197, 205–207
 optimizing, 371–376
 Oracle destination, by Persistent Systems, 330–331
 Partition Processing destination
 for processing cubes, 221
 for SSAS processing, 197, 207–208
 for processing cubes, 220–221
 Raw File destination, 321–322
 raw files created by, 63
 for SSAS objects, 197, 204–208
 table locking in, 170

data flow optimization

data flow properties for, 369–371
 destination optimization, 371–376
 pipeline optimization, 365–369

data flow source adapters

creating connections for, 57
 DataReader source adapter, 60
 definition of, 57, 58
 delaying validation of, 64
 Excel source adapter, 61–63
 flat file source adapter
 change identifier columns in, 82–83
 using, 58–60
 OLE DB source adapter
 parameterized source queries, 70–71
 using, 60–61

raw file source adapter, 63
 XML source adapter, 63

Data Flow Task

memory utilization of, 357–360
 source adapter in, 57
 SQL operations compared to, 360–362

data flows

backpressure of, 365
 execution time of, measuring, 371
 execution trees for, 363

data grain changes

example of, 175–181
 forms of, 173
 transformations for, 173–175

data integration, 6

data lineage

approaches for, 86–87
 definition of, 55, 86
 sequence lineage number for, 87–88

data mining objects (SSAS)

data flow features for, 361
 processing of, 192, 196

data processing, tools for, 1–4. See also ETL (extraction, transformation, and loading)

data profiling, 48–52

data quality

data cleansing
 data profiling for, 48–52
 example of, 177–180
 transformations for, 97
 data validation
 data profiling for, 48–52
 data staging used for, 66–67

data sources

correlating to destination, for incremental extraction, 83–86
 DB2 sources, 331–332
 joining to dimension data, 133–136
 mapping to dimension and fact tables, 144
 multiple record types in, 316–323
 Oracle sources
 code page warnings from, 328
 Oracle client network components for, 326

data sources, Oracle sources (continued)

- parameterized queries with, 328–330
 - providers for, 327–328
 - reading from, 327–330
 - writing to, 330–331
 - preparing for dimension ETL, 97–101
 - raw files, 323–325
 - sharing among packages, 57
 - supported by SSIS, 56
 - VSAM sources, 332
- data staging, 66–67**
- data type of variable, 23**
- data validation**
- data profiling for, 48–52
 - data staging used for, 66–67
- data warehousing ETL. See also dimension ETL;**
- dimension tables; fact table ETL**
 - book about, 102
 - definition of, 4–6
 - NULL values in, 12
- database (SSAS), processing of, 192**
- database (SSIS)**
- administration of, 6–7
 - growth of, effect on performance, 371
 - msdb database
 - Package Store in, 287–288
 - roles in, used by package roles, 298
- DataDirect Providers**
- for DB2, 332
 - for Oracle, 328
- DataReader source adapter, 60**
- date dimension tables, with SCD Wizard, 128–132**
- db_dtsadmin **database role, 298**
 - db_dtsoperator **database role, 298**
- DB2 sources, 331–332**
- debug mode for package execution, 17**
- decision support, data warehousing focusing on, 4**
- Decrypt() **method, 40**
- DefaultBufferMaxRows **property, 370**
- DefaultBufferSize **property, 370**

degenerate dimension, 88

DelayValidation **property, connections, 64**

deployment of packages

- package installation wizard for, 275–279
- Package Store location affecting, 287–288
- utility for, creating, 274–275

derived calculations for measures, 159

Derived Column transformation

- for data grain changes, 174
- for measure calculations, 159, 160–161
- preparing source data for dimension ETL, 97, 99–100

DES (Data Encryption Standard), 40

destination server package execution, 379–380

destinations. See data flow destinations

dimension ETL

- definition of, 91–92, 96–97
- preparing source data for, 97–101
- transformations for, 97

Dimension Processing destination, 197, 205–207

dimension tables

- business key (candidate key) in, 95
- correspondence to source tables, 97
- definition of, 4
- indexes for, optimizing, 121
- loading of, updating inferred members during, 106–107
- mapping source files to, 144
- surrogate key for, 95, 143–144
- updates to, optimizing, 121–122

dimensional model, 4

dimensions (SSAS), processing of, 192–194, 209–211

dimensions (SSIS)

- attributes of, 92
- change types of
 - in custom package, 137–141
 - inferred members, handling with, 104–107
 - list of, 102–104
 - specifying in SCD Wizard, 108–112
 - supported by SCD Wizard, 107

definition of, 92–93
 fact tables relating to, 94–95
 hierarchy within
 definition of, 92–93
 ragged, 126
 unbalanced, 126
 inferred members of, 104–107
 slowly changing dimension, 102

disk input/output (I/O), 357–360

`Dispose()` **method**, `IDisposable` **interface**, **29**

distributed server package execution, 381–382

`Do...While` **loop**, in `For Loop` **container**, **19**

downloading files, over HTTP, 29–30

`DTEXec.exe` **utility**
 definition of, 300
 with distributed package execution, 382
 location where executed, 376–377
 for packages executed outside debug mode, 17
 32-bit and 64-bit versions of, 306–308

`DTEXecUI.exe` **utility**
 32-bit and 64-bit issues regarding, 306–308
 using DTS Migration Wizard, 290–297

DTS Migration Wizard, 339–340, 341–344

DTS (SQL Server 2000 DTS)
 migrating to SSIS
 DTS Migration Wizard for, 339–340, 341–344
 handling exceptions from, 344–352
 SQL Server Upgrade Advisor, 340–341
 success rates of, 339–340

packages
 administration of, 335–339
 editing, 336
 importing, 336
 running, 336–339
 scheduling, 336

`DTSER_SUCCESS` **code**, **17**

`DTSRun.exe` **command**, **336–337**

`DTSRunUI.exe` **command**, **337**

`DTUtil.exe` **utility**, **300–301**

dynamic connections, 35–37

Dynamic Properties tasks, migrating to SSIS, 339, 345–347

E

early binding approach, 26

`Encrypt()` **method**, **40**

encryption

algorithms for, 40

extending Script Component for

 capabilities for, 39

 decryption code, 48

 Rijndael encryption algorithm, implementing,
 40–44

 source and destination for, 45–47

of packages, 300–301

types of, 39

engine threads, 363

`EngineThreads` **property**, **370**

`EntryPoint` **property**, **Script Tasks**, **24**

error handling. See event handling

ETI Provider

for DB2, 332

for Oracle, 328

ETL (extraction, transformation, and loading).

See also data warehousing ETL; dimension ETL; fact table ETL

affecting SSAS attributes, 222–223

definition of, 1

scheduling of, 191–192

SSIS as appropriate tool for, 2–4

`EvaluateAsExpression` **property, variables**,
75–76

event handling

auditing through, 226–229

for checkpoint files, 241–242

disabling, 229

error events, raising in Script Tasks, 37–39

exception handling, when migrating DTS to SSIS,
 344–352

snapshots used with, 229–233

types of event handlers, 225–226

for variable value changing, 24

Excel

64-bit issues with, 307

source adapter for, 61–63

exception handling, when migrating DTS to SSIS, 344–352. See also event handling

Execute Command Task, 197

Execute DDL Task, 214–216

`execute` method, **382**

Execute Process Task

`ASCMD.exe` executable used in, 210
migrating to SSIS, 339

Execute SQL Task

for acquiring dimension surrogate key, 149
location where executed, 377
migrating to SSIS, 339

execution trees

breaking up, 367–368
definition of, 363

extending SSIS, 21–22. See also Script Components; Script Tasks

extract files, breaking up with raw files, 246–251

extraction of data

configuring in Data Flow task, 8–9
data lineage, 55
definition of, 55
incremental extraction
 approaches for, 68
 with audit table, 68
 with auto-incrementing identifier, 68
 with change identifier columns, 68
 correlating source with destination for, 83–86
 definition of, 55, 68
 with log-based auditing, 68
 with `timestamp` data type, 79–81
objectives of, 55
process of, 56

extraction, transformation, and loading (ETL).

See also data warehousing ETL; dimension ETL; fact table ETL

affecting SSAS attributes, 222–223
definition of, 1
scheduling of, 191–192
SSIS as appropriate tool for, 2–4

F

fact table ETL

acquiring dimension surrogate key
 with Execute SQL Task, 149
 Fuzzy Lookup transformation for, 149, 155–159
 for historical changes, 148
 for inferred members, 148, 154
 for large dimensions, 154
 for late-arriving facts, 148, 154
 Lookup transformation for, 149–154
 Merge Join transformation for, 149, 155
 for missing dimension records, 148
definition of, 147–148
for late-arriving facts, 148, 154, 187–190
measure calculations, 159–161
for missing dimension records, 148, 181–186

fact tables

changes to, handling
 approaches for, 161–163
 data deletions, 170
 data inserts, 170
 data updates, 170–172
 Lookup transformation for, 163–166
 Merge Join transformation for, 167–169
data grain changes
 example of, 175–181
 forms of, 173
 transformations for, 173–175
definition of, 4, 143–144
dimension surrogate keys in, 95, 143–144
loading of, adding inferred members during,
 105–106
mapping source files to, 144
measures in, 143, 145
metadata in, 145
relating to dimensions, 94–95
snapshot fact tables, 145–146
transaction fact tables, 145

`Failure()` **method, 37**

`FastParse` **property, flat file source adapter, 58–60**

file system, Package Store on, 287–288
file watcher task, building, 252–254
FireError() method, 37–39
firewall port settings, 299
fixed attribute (Change Type 0)
 definition of, 102
 implications on SSAS attributes, 222–223
Fixed Attribute Output, SCD Wizard, 120
Flat File Connection Manager
 for file with multiple record types, 317, 318
 migration to SSIS, 350–351
flat file source adapter
 change identifier columns in, 82–83
 using, 58–60
For Each Loop container, 19–20, 36, 129
For Loop container, 19, 129
full cache, 366
full outer join, Merge Join transformation, 155
full process, for SSAS dimensions, 193
Fuzzy Lookup transformation, for acquiring dimension surrogate key, 149, 155–159

G

GAC (Global Assembly Cache), 30–35
Gacutil.exe tool, 32
garbage collector, SuppressFinalize()
 method, 29
Global Assembly Cache (GAC), 30–35
Global Assembly Cache Tool (Gacutil.exe), 32
Global Variables tasks, migrating to SSIS, 339
grain, changes in
 example of, 175–181
 forms of, 173
 transformations for, 173–175

H

hierarchy within dimensions
 definition of, 92–93
 ragged, 126
 unbalanced, 126

historical attribute (Change Type 2)
 definition of, 102, 103–104
 with fact table ETL, 148
 implications on SSAS attributes, 222–223
Historical Attribute Inserts Output, SCD Wizard, 116–118
HIT Providers for DB2, 332
Host Integration Server, 2
HTTP, downloading files over, 29–30

I

IBM Providers for DB2, 331–332
IDisposable interface, 29
IDTSComponentEvents interface, 37
incremental extraction
 approaches for, 68
 with audit table, 68
 with auto-incrementing identifier, 68
 with change identifier columns, 68
 correlating source with destination for, 83–86
 definition of, 55, 68
 with log-based auditing, 68
 with timestamp data type, 79–81
incremental process, for SSAS partitions, 195
indexes, dropping and re-creating for optimization, 371–372
Inferred Member Updates Output, SCD Wizard, 114–115
inferred members of dimension
 adding during fact load, 105–106
 with fact table ETL, 148, 154
 handling in fact table ETL, 182
 implications on SSAS attributes, 222–223
 reasons for, 104–105
 specifying in SCD Wizard, 112–118
 updating during dimension load, 106–107
initial value of variable, 23
in-line data flow approach, for inferred member creation, 182
in-line transformation processing, 360
inner join, Merge Join transformation, 155
insertions, maximum commit size for, 375–376

installation wizard for packages, 275–279

integration of data, 6

Itanium 64-bit machines, 306

K

Keep identity option, OLE DB destination, 374

Keep nulls option, OLE DB destination, 374

key file, 30–31

Kimball, Ralph

The Complete Guide to Dimension Modeling, 92

dimension change tracking theory by, 102

L

late-arriving dimension scenario (inferred members)

adding during fact load, 105–106

with fact table ETL, 148, 154

handling in fact table ETL, 182

implications on SSAS attributes, 222–223

reasons for, 104–105

specifying in SCD Wizard, 112–118

updating during dimension load, 106–107

late-arriving fact, 148, 154, 187–190

late-binding approach, 26

left outer join, Merge Join transformation, 155

libraries (assemblies)

adding to GAC, 30–35

calling from Script Task, 33

creating, 27–29

referencing from local directory, 30

strong name for, 30–31

testing, 34–35

lineage, data

approaches for, 86–87

definition of, 55, 86

sequence lineage number for, 87–88

load method, 382

loading of data. See ETL (extraction, transformation, and loading)

Local Temp DB Connection Manager, 346

log-based auditing, 68

logging

definition of, 233–234

log file for, unique, 236–237

monitoring data flow of, 235–236

performance of, 368

provider for, creating, 234–235

reporting on, 237

Lookup transformation

for acquiring dimension surrogate key, 149–154

cached, 136–137

correlating data using, 133, 134

for data grain changes, 174

data preparation using, 101

definition of, 97

for identifying fact table changes, 163–166

optimizing, 366–367

loops

For Each Loop container, 19–20, 36, 129

For Loop container, 19, 129

M

Management Studio

accessing packages from, 288–289

administration of DTS packages using, 335

assigning package roles using, 297–299

running DTEXECUI.exe utility from, 290–297

Maximum insert commit size option, OLE DB destination, 375

MDAC (Microsoft Data Access Components)

providers for, 61

sources from, 56

32-bit and 64-bit issues regarding, 307

measure groups (SSAS)

definition of, 123

processing of, 192

measures, in fact tables

calculations for, 159–161

definition of, 143, 145

memory utilization, 357–360

Merge Join transformation

- for acquiring dimension surrogate key, 149, 155
- for data grain changes, 174
- for identifying fact table changes, 167–169
- joining source data to dimension data, 133–136

Merge transformation, for data grain changes, 174**metadata**

- definition of, 86
- in fact tables, 145

Microsoft Data Access Components (MDAC)

- providers for, 61
- sources from, 56
- 32-bit and 64-bit issues regarding, 307

Microsoft Data Warehouse Toolkit (Wiley), 102**Microsoft Distributed Transaction Coordinator (MSDTC), 243****Microsoft Providers**

- for DB2, 331
- for Oracle, 327, 328

mining models (SSAS), 196. See also data mining objects (SSAS)**mining structures (SSAS), 196. See also data mining objects (SSAS)****msdb database**

- Package Store in, 287–288
- roles in, used by package roles, 298

MSDTC (Microsoft Distributed Transaction Coordinator), 243

`MsDtsServer` **service**. See **SSIS Service**

`MsDTSSrvr.exe` **executable**, 354

`MsDtsSrvr.ini.xml` **file**, 284, 285

Multicast transformation

- change identifier columns using, 77
- data staging using, 66–67, 368

N**namespace for variable, 24****New Output, SCD Wizard, 116****O****object-oriented programming (OOP), 259****ODBC Providers for Oracle, 328****OLE DB destination, optimizing, 372–375****OLE DB Providers**

- for DB2, 331–332
- for Oracle, 327, 328
- for VSAM, 332

OLE DB source adapter

- parameterized source queries, 70–71
- using, 60–61

OnError event, 225–226**OnInformation event, 226****OnPostExecute event, 355****OnPreExecute event, 355****OnVariableValueChanged event, 24****OnWarning event, 225****OOP (object-oriented programming), 259****optimizing SSIS. See also scalability**

- data flow properties for, 369–371
- destination optimization, 371–376
- performance
 - database growth affecting, 371
 - of logging, 368
 - memory utilization, 357–360
 - partition tables affecting, 372
 - SQL operations compared to Data Flow Tasks, 360–362
 - SSIS Service status, checking, 354
 - task duration, checking, 354
 - of transactions, 368–369
 - of writing data to Oracle, 330–331
- pipeline optimization, 365–369

Oracle destination, by Persistent Systems, 330–331**Oracle sources**

- code page warnings from, 328
- Oracle client network components for, 326
- parameterized queries with, 328–330
- providers for, 327–328
- reading from, 327–330
- writing to, 330–331

P

package configurations

for dynamic connections, 35–36, 64–65

repository for

creating, 266–269

reading variables from, 269–272

reasons to use, 266

setting variables, 272–274

SSIS Package Configuration option for

creating, 262–263

definition of, 261

managing, 265

multiple, 264–265

package execution strategies

Bulk Insert Task, 377

data flows, 378

destination server package execution, 379–380

distributed server package execution, 381–382

Execute SQL Task, 377

source server package execution, 378–379

stand-alone server package execution, 380–381

storage and execution locations, 376–377

package installation wizard, 275–279

package roles, 297–299

Package Store

definition of, 281

location of

determining, 283

on file system, 287–288

in `msdb` database, 287–288

SSIS Service for, 281–283

packages (DTS)

administration of, 335–339

editing, 336

importing, 336

running, 336–339

scheduling, 336

packages (SSIS)

administration of

command line utilities for, 299–301

firewall issues regarding, 299

in Management Studio, 288–299

performance counters, 308–309, 357

scheduling packages, 284, 301–305

tasks in, 6–7

availability of, Package Store location affecting, 288

containers for, creating, 19–20

corrupt, handling, 301

creating, 7

deployment of

package installation wizard for, 275–279

Package Store location affecting, 287–288

utility for, creating, 274–275

executing

in BIDS, 17–18

with `DTEXEC .exe` utility, 300

with `DTEXECUI .exe` utility, 290–297

with SSIS Service, 282

monitoring execution of

in Management Studio, 288–289

with `MsDtsServer` service, 282–283

with performance counters, 308–309, 357

precedence constraints for, 15–17

re-encrypting, 300

reliability of

auditing for, 226–229

checkpoint files for, 238–242, 245

event handling for, 225–233, 241–242

logging for, 233–237

raw files for, 245–251

snapshots for, 229–233

transactions for, 242–245

WMI integration for, 251–254

reusable, 259–261

scheduling of

with proxy accounts, 303–305

with SQL Server Agent, 301–303

troubleshooting, 303

when using central SSIS server, 284

security for

firewall port settings, 299

package roles for, 297–299

Package Store location affecting, 287, 299

- passwords for packages, setting, 299, 301
- re-encrypting packages, 300
- on separate, central server, 283–284
- sharing data sources among, 57
- source control integration for, 255–259
- stopping with SSIS Service, 282
- tasks, configuring, 8–14
- team development of, 255–259
- templates for, 260–261
- parameterized queries with Oracle sources, 328–330**
- parameterized source queries, 70–71**
- parent-child dimension tables**
 - definition of, 94
 - with SCD Wizard, 125–128
- partial cache, 366**
- partially blocking transformations, 363**
- Partition Processing destination**
 - for processing cubes, 221
 - for SSAS processing, 197, 207–208
- partition tables, performance affected by, 372**
- partitions (SSAS)**
 - managing (creating, modifying, deleting), 197–198, 217–220
 - processing dynamically
 - with AMO, 216–217
 - with ASCMD executable, 211–214
 - with Execute DDL Task, 214–216
 - processing of, 194–195
- performance**
 - database growth affecting, 371
 - of logging, 368
 - memory utilization, 357–360
 - partition tables affecting, 372
 - SQL operations compared to Data Flow Tasks, 360–362
 - SSIS Service status, checking, 354
 - task duration, checking, 354
 - of transactions, 368–369
 - of writing data to Oracle, 330–331
- performance counters, 308–309, 357**
- Persistent Systems, Oracle destination by, 330–331**
- pipeline**
 - architecture of, 363–365
 - definition of, 101
 - optimization, 365–369
- Pivot transformation**
 - for data grain changes, 174
 - null values handled by, 314
- post-data flow process, for inferred member creation, 182**
- `PostExecute` event, 225
- `PostExecute()` subroutine, 26
- precedence constraints, 15–17**
- pre-data flow check, for inferred member creation, 182**
- `PreExecute` event, 225
- process add, for SSAS dimensions, 194**
- process clear, for SSAS mining structures and models, 196**
- process data**
 - for SSAS dimensions, 194
 - for SSAS partitions, 195
- process full**
 - for SSAS mining structures and models, 196
 - for SSAS partitions, 195
- process index**
 - for SSAS dimensions, 194
 - for SSAS partitions, 195
- process structure, for SSAS mining structures and models, 196**
- process threads, 363**
- process update, for SSAS dimensions, 193**
- Professional SQL Server 2005 Integration Services (Wiley), 7, 21***
- profiling data, 48–52**
- properties. See also specific properties**
 - for data flow optimization, 369–371
 - overriding when running packages, 293
- property expressions**
 - for dynamic connections, 36
 - updating connections using, 64–66
- proxy accounts, scheduling packages using, 303–305**
- publications. See books**

Q

quality of data

- data cleansing
 - data profiling for, 48–52
 - example of, 177–180
 - transformations for, 97
- data validation
 - data profiling for, 48–52
 - data staging used for, 66–67

queries

- in Data Flow task, 8–11
- parameterized source queries, 70–71
- variable bound SQL source, 71–72

queuing technologies, 382

R

ragged hierarchy, 126

Raw File destination, 321–322

Raw File source, 323–325

raw file source adapter, 63

raw files

- breaking up extract files using, 246–251, 316–323
- configuring, 248–249
- consuming, 323–325
- definition of, 245
- uses of, 245–246

Reader Role, 297–298

`ReadOnlyVariables` property, **Script Tasks, 24**

`ReadWriteVariables` property, **Script Tasks, 24**

reliability of packages

- auditing for, 226–229
- checkpoint files for, 238–242, 245
- event handling for, 225–233, 241–242
- logging for, 233–237
- raw files for, 245–251
- snapshots for, 229–233
- transactions for, 242–245
- WMI integration for, 251–254

replicating data (data integration), 6

reporting on logging, 237

resources. *See* books

Restating History (Change Type 1)

- definition of, 102, 103
- implications on SSAS attributes, 222–223

reusability

- reusable packages, 259–261
- Script Tasks used for, 22

Rijndael encryption algorithm, 40–44

Ross, Margy (*The Complete Guide to Dimension Modeling*), 92

Row Count transform, auditing rows to be written, 13

row transformations, 363

row-based operations, 366

row-by-row operations, 366

rows, creating from columns, 311–315

Rows per batch option, OLE DB destination, 375

Rows Read counter, 308

Rows Written counter, 308

`RunInOptimizedMode` property, 370

S

scalability

- data flow optimization
 - data flow properties for, 369–371
 - destination optimization, 371–376
 - pipeline optimization, 365–369
- package execution strategies for
 - Bulk Insert Task, 377
 - data flows, 378
 - destination server package execution, 379–380
 - distributed server package execution, 381–382
 - Execute SQL Task, 377
 - source server package execution, 378–379
 - stand-alone server package execution, 380–381
 - storage and execution locations, 376–377
- performance evaluations for
 - memory utilization, 357–360
 - SQL operations compared to Data Flow Tasks, 360–362

SSIS Service status, checking, 354
 task duration, checking, 355–356
 of SCD (Slowly Changing Dimension) Wizard,
 120, 133

SCD (Slowly Changing Dimension) Wizard

advantages of, 120
 Changing Attributes Updates Output, 114
 CurrentRowWhere property, 118
 date dimension tables with, 128–132
 dimension changes, specifying, 108–112
 dimension table and keys, specifying, 107–108
 Fixed Attribute Output, 120
 Historical Attribute Inserts Output, 116–118
 index optimizations for, 121
 inferred members, specifying, 112–118
 invoking, 107
 limitations of, 120–121
 locking contentions with, 121
 New Output, 116
 parent-child dimension tables with, 125–128
 scalability of, 120, 133
 snowflake dimension tables with, 123–125
 SQLCommand property, 118
 Unchanged Output, 119
 update optimizations for, 121–122

scheduling DTS packages, 336

scheduling packages

with proxy accounts, 303–305
 with SQL Server Agent, 301–303
 troubleshooting, 303
 when using central SSIS server, 284

scope of variable, 23

Script Component Source, 312

Script Components

for data grain changes, 175
 data profiling in, 49–52
 encrypting data in, 39–48
 language used in, 24
 retrieving variables in, 26

Script Tasks

accessing and updating connections from, 36–37
 calling assembly from, 33
 code reusability with, 22

definition of, 22
 language used in, 24
 raising errors in, 37–39
 retrieving variables in, 24–26
 32-bit and 64-bit issues regarding, 307
 user-defined variables for, 22–24

ScriptMain class, 24

security

firewall port settings, 299
 package roles for, 297–299
 Package Store location affecting, 287, 299
 passwords for packages, setting, 299, 301
 re-encrypting packages, 300
 for SSAS integration, 198

Send Mail tasks, migrating to SSIS, 339

Sequence container, 19

Simple Load (UCASE).dts package, 341

Simple Load.dts package, 341

slowly changing dimension. See also SCD (Slowly Changing Dimension) Wizard

custom package for
 correlating source with destination, 133–137
 dimension changes for, 137–141
 definition of, 102

SN utility, creating key file using, 30–31

snapshot fact tables, 145–146

snapshots, 229–233

snowflake dimension tables

definition of, 94
 with SCD Wizard, 123–125

Sort transformation

alternative to, 366
 for data grain changes, 174

source adapters

creating connections for, 57
 DataReader source adapter, 60
 definition of, 57, 58
 delaying validation of, 64
 Excel source adapter, 61–63
 flat file source adapter
 change identifier columns in, 82–83
 using, 58–60

source adapters (continued)

OLE DB source adapter

parameterized source queries, 70–71
using, 60–61

raw file source adapter, 63

XML source adapter, 63

source control integration for packages, 255–259

source queries

parameterized, 70–71

variable bound SQL source, 71–72

source server package execution, 378–379

sources, data

correlating to destination, for incremental
extraction, 83–86

DB2 sources, 331–332

joining to dimension data, 133–136

mapping to dimension and fact tables, 144

multiple record types in, 316–323

Oracle sources

code page warnings from, 328

Oracle client network components for, 326

parameterized queries with, 328–330

providers for, 327–328

reading from, 327–330

writing to, 330–331

preparing for dimension ETL, 97–101

raw files, 323–325

sharing among packages, 57

supported by SSIS, 56

VSAM sources, 332

`SourceType` property, **Analysis Services
Execute DDL Task, 203**

`sp_start_job` stored procedure, **381**

SQL

Data Flow Tasks compared to, 360–362
queries

in Data Flow task, 8–11

parameterized source queries, 70–71

variable bound SQL source, 71–72

SQL Native Client (SQLNCLI), 61

SQL Server Agent

distributed package execution using, 381

scheduling packages using, 284, 301–303

SQL Server Integration Services Report Pack, 237

SQL Server 2000 DTS

migrating to SSIS

DTS Migration Wizard for, 339–340, 341–344

handling exceptions from, 344–352

SQL Server Upgrade Advisor, 340–341

success rates of, 339–340

packages

administration of, 335–339

editing, 336

importing, 336

running, 336–339

scheduling, 336

SQL Server 2005 Analysis Services objects. See SSAS objects

SQL Server 2005 Integration Services (SSIS)

book about, 7, 21

extending, approaches for, 21–22

uses of, 4

SQL Server 2005 Reporting Service (SSRS), 356

`SqlCommand` property, **SCD Wizard, 118**

`SqlConnection` object, **36–37**

`sqlldr` command, **330–331**

SQLNCLI (SQL Native Client), 61

SQLServer: SSIS Pipeline, 308–309

SSAS (SQL Server 2005 Analysis Services) objects

AMO (Analysis Management Objects) for, 197

attributes of, affected by ETL processing,
222–223

connection for, 198

control flow tasks for

Analysis Services Execute DDL Task, 197,
203–204

Analysis Services Processing Task, 197,
199–203

Execute Command Task, 197

cubes, processing, 192, 220–221

data flow destinations for

Dimension Processing destination, 197,
205–207

Partition Processing destination, 197, 207–208

- data mining objects
 - data flow features for, 361
 - processing of, 192, 196
 - database, 192
 - dimensions, 192–194
 - measure groups
 - definition of, 123
 - processing of, 192
 - partitions
 - creating and modifying, 197–198
 - processing of, 194–195
 - processing methods for, 197
 - scheduling processing of, 191–192
 - security for, 198
 - types of objects processed, 192
 - XMLA scripts for, 197, 209–211
 - SSIS Package Store**
 - definition of, 281
 - location of
 - determining, 283
 - on file system, 287–288
 - in `msdb` database, 287–288
 - SSIS Service for, 281–283
 - SSIS Service**
 - clustering, 285–286
 - configuring, 281–283
 - highly available, benefits of, 285
 - required for Package Store, 281
 - running packages with, 282
 - on separate, central server, 283–284
 - status of, checking, 354
 - stopping packages with, 282
 - SSIS (SQL Server 2005 Integration Services)**
 - book about, 7, 21
 - extending, approaches for, 21–22
 - uses of, 4
 - SSRS (SQL Server 2005 Reporting Service), 356**
 - staging data, 66–67**
 - staging table, 171**
 - stand-alone server package execution, 380–381**
 - standard calculations for measures, 159**
 - star dimension, 123**
 - star schema, 4**
 - streaming transformations, 363**
 - strong name for assembly, 30–31**
 - strong typing, 29**
 - `SuppressFinalize()` **method, garbage collector, 29**
 - surrogate key of dimension table**
 - acquiring for fact table ETL
 - with Execute SQL Task, 149
 - Fuzzy Lookup transformation for, 149, 155–159
 - for historical changes, 148
 - for inferred members, 148, 154
 - for large dimensions, 154
 - for late-arriving facts, 148, 154
 - Lookup transformation for, 149–154
 - Merge Join transformation for, 149, 155
 - for missing dimension records, 148
 - definition of, 95
 - in fact table, 143–144
 - symmetric encryption, 39**
 - synchronizing data (data integration), 6**
 - synchronous transformation, 45**
 - `sysdtslog90` **table, 234**
- T**
- Table lock option, OLE DB destination, 375**
 - tables. See also fact tables**
 - audit table, 68
 - date dimension tables, with SCD Wizard, 128–132
 - dimension tables
 - business key (candidate key) in, 95
 - correspondence to source tables, 97
 - definition of, 4
 - indexes for, optimizing, 121
 - loading of, updating inferred members during, 106–107
 - mapping source files to, 144
 - surrogate key for, 95, 143–144
 - updates to, optimizing, 121–122
 - multiple, in one file, 316–323
 - partition tables, performance affected by, 372
 - staging table, 171
 - Task Scheduler, scheduling packages using, 284**

TaskHost **container, 19**

tasks

- Analysis Services Execute DDL Task
 - for partition management, 217–218
 - processing SSAS objects using, 197, 203–204
 - Analysis Services Processing Task, 197, 199–203
 - Bulk Insert Task
 - location where executed, 377
 - migrating to SSIS, 339
 - configuring, 8–14
 - containers for, 19–20
 - creating, 8
 - Data Flow Task
 - memory utilization of, 357–360
 - source adapter in, 57
 - SQL operations compared to, 360–362
 - duration of, checking, 355–356
 - Execute Command Task, 197
 - Execute DDL Task, 214–216
 - Execute Process Task
 - ASCMD.exe executable used in, 210
 - migrating to SSIS, 339
 - Execute SQL Task
 - for acquiring dimension surrogate key, 149
 - location where executed, 377
 - migrating to SSIS, 339
 - precedence constraints for, 15–17
 - Script Tasks
 - accessing and updating connections from, 36–37
 - calling assembly from, 33
 - code reusability with, 22
 - definition of, 22
 - language used in, 24
 - raising errors in, 37–39
 - retrieving variables in, 24–26
 - 32-bit and 64-bit issues regarding, 307
 - user-defined variables for, 22–24
 - WMI Data Reader Task, 251
 - WMI Data Watcher Task, 251
- team development of packages, 255–259**

- templates for packages, 260–261**
- Term Extraction transformation, for data grain changes, 174**
- Term Lookup transformation, for data grain changes, 175**
- 32-bit and 64-bit issues, 306–308**
- thread mapping model, 363**
- 3DES (Triple DES), 40**
- timestamp datatype, incremental extraction using, 79–81**
- tnsnames.ora file, 326**
- Tracking History (Change Type 2)**
 - definition of, 102, 103–104
 - with fact table ETL, 148
 - implications on SSAS attributes, 222–223
- Transact SQL (TSQL), 2**
- transaction fact tables, 145**
- transactional system, 4**
- transactions**
 - definition of, 242–243
 - enabling, 243
 - implementing, 243–245
 - performance of, 368–369
 - using with checkpoint files, 245
- Transform Data tasks, migrating to SSIS, 339, 348–350**
- transformations. See also ETL (extraction, transformation, and loading); SCD (Slowly Changing Dimension) Wizard**
 - Aggregate transformation
 - for data grain changes, 174
 - for maximum change identifier value, 77
 - for measure calculations, 160
 - asynchronous transformation, 45, 360
 - blocking transformations, 363, 365
 - Conditional Split transformation
 - for change identifier columns, 82–83
 - for data grain changes, 174
 - for identifying fact table changes, 163, 165–166
 - for separating record types, 319
 - for separating valid and invalid data, 51

configuring in Data Flow task, 9–13
 Data Conversion transformation, 97, 98–99, 319–321
 Derived Column transformation
 for data grain changes, 174
 for measure calculations, 159, 160–161
 preparing source data for dimension ETL, 97, 99–100
 Fuzzy Lookup transformation, for acquiring dimension surrogate key, 149, 155–159
 in-line transformation processing, 360
 Lookup transformation
 for acquiring dimension surrogate key, 149–154
 cached, 136–137
 correlating data using, 133, 134
 for data grain changes, 174
 data preparation using, 101
 definition of, 97
 for identifying fact table changes, 163–166
 optimizing, 366–367
 Merge Join transformation
 for acquiring dimension surrogate key, 149, 155
 for data grain changes, 174
 for identifying fact table changes, 167–169
 joining source data to dimension data, 133–136
 Merge transformation, for data grain changes, 174
 Multicast transformation
 change identifier columns using, 77
 data staging using, 66–67, 368
 partially blocking transformations, 363
 Pivot transformation
 for data grain changes, 174
 null values handled by, 314
 row transformations, 363
 Sort transformation
 alternative to, 366
 for data grain changes, 174
 streaming transformations, 363
 synchronous transformation, 45

Term Extraction transformation, for data grain changes, 174
 Term Lookup transformation, for data grain changes, 175
 Union All transformation
 for breaking up execution tree, 367–368
 for data grain changes, 174
 Unpivot transformation
 for creating rows from columns, 312–313
 for data grain changes, 174

Triple DES (3DES), 40

troubleshooting

code page warnings from Oracle sources, 328
 event handling
 auditing through, 226–229
 for checkpoint files, 241–242
 disabling, 229
 error events, raising in Script Tasks, 37–39
 exception handling, when migrating DTS to SSIS, 344–352
 snapshots used with, 229–233
 types of event handlers, 225–226
 for variable value changing, 24
 Package Store location affecting, 288
 scheduling packages, 303

TSQL (Transact SQL), 2

U

unbalanced hierarchy, 126

Unchanged Output, SCD Wizard, 119

Union All transformation

 for breaking up execution tree, 367–368
 for data grain changes, 174

Unpivot transformation

 for creating rows from columns, 312–313
 for data grain changes, 174

unprocess

 for SSAS dimensions, 194
 for SSAS mining structures and models, 196
 for SSAS partitions, 195

Upgrade Advisor, 340–341

user-defined variables

- creating, 22–24
- retrieving in Script Components, 26
- retrieving in Script Tasks, 24–26

V

ValidateExternalMetadata property, source adapter, 64

validation of data

- data profiling for, 48–52
- data staging used for, 66–67

VariableDispenser object, 24, 26

variables

- definition of, 22
- user-defined variables
 - creating, 22–24
 - retrieving in Script Components, 26
 - retrieving in Script Tasks, 24–26

Variables object, 24, 26

Variables window, 22–24

Visual Basic .NET, used in Script Tasks and Components, 24

Visual Source Safe (VSS), 255–259

Visual Studio Team Systems, 255

Visual Studio 2005 source-control integration, 255

VSAM sources, 332

VSS (Visual Source Safe), 255–259

W

warehousing ETL. See also dimension ETL; dimension tables; fact table ETL

- book about, 102
- definition of, 4–6

WebClient object, 29–30

Wiley Publishing

- Microsoft Data Warehouse Toolkit*, 102
- Professional SQL Server 2005 Integration Services*, 7, 21

Windows Query Language (WQL), 251

WMI Data Reader Task, 251

WMI Data Watcher Task, 251

WMI (Windows Management Instrumentation)

- definition of, 251
- file watcher task using, 252–254
- integration with SSIS, 251

workflow. See tasks

WQL (Windows Query Language), 251

Writer Role, 297–298

X

XML source adapter, 63

XMLA scripts, for SSAS processing, 197, 209–211

xp_cmdshell procedure, 382

XSD schema, using with XML source adapter, 63

