

Index

Symbols

|| (Logical OR operator), 250

^ symbol, 176

A

ACE OLEDB 12.0 (Microsoft), 131

active FTP, 103

Adamson, C., 198

add process. See process add

ADO.NET

connection manager, 158, 168

dataset, 162

destination adapter, 414

**advanced file management (archiving system/
Script Task)**

design section, 360–363

accessing objects, 362

custom assemblies, 362

script languages, 361

scripting, 363

problem section, 359–360

scenario for, 360

solution section, 363–369

access external assembly, 366–367

archive the files, 367–369

create external assembly, 363–366

Adventure Works cube, 328, 332

Adventure Works DW 2008

catalog, 325

AdventureWorks2008 Connection

Manager, 213

AdventureWorks2008 transactional

system, 202

AdventureWorksDW2008 database

ADO.Net connection manager and, 168

AMO and, 325

business keys in, 200

DimCustomer table in, 168, 170

fact tables in, 199, 255

OLE DB connection to, 47

Aggregate transformation

as blocking transformation, 403, 408

call center fact table and, 290, 291, 292,

293, 381

defined, 261

Editor, 142, 292

measure calculations and, 271, 272, 273

Multicast and, 142, 143, 149

agile methods, 14–15

**AMO (Analysis Management Objects), 312,
321–323, 324, 325, 330**

Analysis Management Objects. See AMO

**Analysis Services. See SSAS processing
architecture**

**Analysis Services Execute DDL Task. See
Execute DDL Task**

**Analysis Services Processing Task. See
Processing Task**

**application factors (storage/deployment
methods), 77**

ApplicationConfiguration, 348

**archival and cleanup processes (scale-out ETL),
345–346**

data file archival process, 345

stage table clean-up process, 346

**archiving system (advanced file management/
Script Task)**

design section, 360–363

accessing objects, 362

custom assemblies, 362

script languages, 361

scripting, 363

problem section, 359–360

scenario for, 360

solution section, 363–369

access external assembly, 366–367

archive the files, 367–369

create external assembly, 363–366

ASCMD.exe, 312, 317–318

asynchronous outputs, 373

attributes

- dimensions and, 198
- File System Task and, 100–101

Attunity

- Oracle destination adapter, 418–419
- Oracle source adapter, 133
- Teradata destination adapter, 418
- Teradata source adapter, 133

audit tables, 122, 261, 289, 342, 403. See also logging/auditing mechanism

AuditDatetime, 292

Auto Shrink, 414

auto-incrementing change identifier, 122

automation

- data extraction and, 156
- deployment and, 95

Average Disk Queue Length, 395, 396, 397

B

backpressure, 408

Backup Operator, 82

backups

- drive size and, 23
- file system storage and, 82
- msdb database and, 82
- multiple environments and, 76
- SQL Server and, 39
- standardization and, 76

banding, data, 211, 212

batch logic modifications (management framework), 71

batches, 41, 49

BatchLog table, 41, 49

BatchLogID, 52

BI (business intelligence), 197. See also dimension table ETL; fact table ETL; scale-out ETL; SSAS

- data warehouse ETL and, 197
- dimension change tracking theory and, 204
- dimension table ETL and, 197
- fact table ETL and, 197
- NULLs and, 204
- scale-out ETL and, 335

- SSAS and, 197, 305, 306, 312, 323, 333
- surrogate keys and, 201

BIDemo database, 180, 181, 182, 190, 191

BIDS (Business Intelligence Development Studio), 43

- cube designer, 310
- deployment utility and, 83–88
- dimension editor, 309
- DTUtil.exe v., 90
- file system storage and, 83
- new package ID and, 43, 71, 80
- Package Configurations Organizer and, 45
- package execution and, 75
- Solution Explorer, 71, 125, 309
- task durations and, 390
- XML configuration file and, 44

BizTalk, 9

BLOBTempStoragePath, 412

blocking transformations, 403, 408. See also Aggregate transformation; Fuzzy Grouping transformation; Sort transformation

bottlenecks

- disk I/O, 7–8, 23, 395, 396, 401
- transformation/memory, 394–397

buffer spooling, 395

buffers

- defined, 372, 403
- tuning, 413

BufferTempStoragePath, 412

bulk INSERT operations, 402

BULK INSERT statement, 416

Bulk Insert Task, 24–25

business intelligence. See BI

Business Intelligence Development Studio. See BIDS

business keys

- as candidate keys, 17, 200, 258
- purpose of, 200

C

cache size management, 409–410

call center fact table (FactCallCenter), 289–294

- Aggregate transformation and, 290, 291, 292, 293, 381
- aggregation steps, 290

- audit data, 289
- data flow, 290
- data sampling, 289
- Derived Column transform and, 290, 291, 292, 293, 373, 375, 376, 385
- ETL objective, 289
- FactCallCenter.dtsx, 294
- IssueID and, 290, 292, 293
- LevelOneOperators, 290, 292
- LevelTwoOperators, 290, 292
- Multicast transformation and, 290, 291, 292
- package execution, 294
- row counts, 294
- Script Component, 294
 - component types, 371–372
 - design section, 371–374
 - entire script, 384
 - output type, 372–373
 - overridable methods, 373–374
 - package setup, 375–376
 - package wrap-up, 385–387
 - problem section, 370
 - process input, 383–384
 - reusable methods, 379–380
 - row-by-row processing, 380–382
 - solution section, 374–387
 - summary, 387
- as snapshot fact table, 256
- solution, 290
- Sort transformation and, 293
- values, data viewer and, 294
- CallCenterInput_ProcessInput, 383**
- CallCenterInput_ProcessInputRow, 380, 383, 384**
- Candidate Key profiles, 170–171**
- candidate keys, 17, 200, 258. See also business keys**
- CatalogExtract.txt file, 175**
- CDC (Change Data Capture), 30, 122, 147**
- central common services server (scale-out ETL process). See also scale-out processes servers**
 - database tables on, 347–348
 - packages on, 354–357
 - role of, 338
 - run/execute operations on, 347
 - SCOMain 01—Hourly Aggregator.dtsx, 355
 - SCOMain 02—Daily Aggregator.dtsx, 356
 - SCOMain 03—Cleanup.dtsx, 356–357
 - stored procedures on, 349–350
- cfgIncrementalExtraction table, 138**
- Change Data Capture (CDC), 30, 122, 147**
- change identifier columns, 122**
- change types, dimension, 204–209. See also Type 0 fixed attributes; Type 1 changing attributes; Type 2 historical attributes**
- Changing Attribute Updates Output, 220–222**
- changing attributes. See Type 1 changing attributes**
- changing metadata, 34–35**
- Chapter6—MultiFileType.dtsx, 185**
- Chapter6_OrderTables.sql, 184, 189**
- check constraints option, 417**
- CheckServerHourlyPreAggregateReadiness, 349**
- child packages, 93**
- cleansing data. See data cleansing**
- CLR (Common Language Runtime), 183**
- clustered index, 229, 399**
- clustering**
 - MsDtsServer100 and, 78
 - SANs and, 23
 - storage methods and, 91, 94
- collocation, 22**
- Column Length Distribution profiles, 171–172**
- Column Mappings tab, 221, 303**
- Column Null Ratio profiles, 172**
- Column Pattern profiles, 172–173**
- Column Statistics profiles, 173**
- Column Value Distribution profiles, 173–174**
- Combine Files data flow, 111, 113**
- Common Language Runtime (CLR), 183**
- common table expression (CTE), 391, 392**
- CommonConfigurations filter, 45, 46, 47**
- Conditional Split transformation**
 - defined, 261
 - dummy raw file and, 112
 - Fact_SalesOrderID and, 151, 152
 - flat files (with change identifier values) and, 148–149
 - Merge Join and, 149
 - XML source adapter and, 133
- Conditional Split Transformation Editor, importing multiple files (file preprocessor), 117**
- configuration management scheme, 35, 38–40, 43–47**
 - environment variable
 - configuration, 38–39
 - creation, 43
 - implementation, 43–47

configuration management scheme (*continued*)

SQL Server configurations, 39, 45–46

three-pass strategy, 38

XML configuration file, 39, 44

connAdventureWorksDW2008, 46, 47

connection manager configuration, 46–47

control flow stored procedures, 53–59

coordination, File Watcher Task and, 109

corporate political challenges, 8

CreateNewOutputRows method, 373

CTE (common table expression), 391, 392

cubes

cube designer, 310

processing, 30, 307, 308

currency rate, 283

currency rate fact table, 256, 266–267

Current Flag column, 222, 225

CurrentRowWhere property, 225

custom assemblies, Script Tasks and, 362

custom dimension ETL (without SCD transformation), 234–236, 246–252. See also SCD Wizard

data correlation approach, 235–236

data flow for customer dimension, 246

dimension changes, 248–250

dimension inserts/updates, 236, 250–252

source data joined to dimension data, 235,
247–248

customer dimension

custom dimension ETL and, 246–252

data flow for, 246

D

daily data aggregation process (scale-out ETL process), 345

DAS (direct attached storage), 23

“Data access mode” property, 133, 134, 136, 137, 191, 279, 415

data aggregation tasks (scale-out ETL process), 344–345

daily data aggregation process, 345

hourly data aggregation process, 345

hourly data pre-aggregation process, 344

data banding, 211, 212

data challenges, 8

data cleansing, 30, 167–195

Data Profiling Task, 30, 167, 168–174, 203

Candidate Key profiles, 170–171

Column Length Distribution profiles, 171–172

Column Null Ratio profiles, 172

Column Pattern profiles, 172–173

Column Statistics profiles, 173

Column Value Distribution profiles, 173–174

Functional Dependency profiles, 174

design section, 175–190

Fuzzy Grouping Transform and, 175,
177–180, 195

Fuzzy Lookup Transform and, 175,
180–183, 195

inCustomers table and, 181, 182, 190, 191

multiple record types in single file, 183–188

problem section, 168–174

Raw files and, 188–190

Script Transform and, 175–177, 193

solution section, 190–194

Data Conversion transformation

multiple record types in single file and, 186,
187, 188

non-Unicode and, 132

uses of, 210

data duplication. See duplicate data

data element documentation, 15–18, 31

data extraction, 29–30, 119–165. See also

incremental data extraction

automation of, 156

data lineage and, 123, 154–156

data staging method and, 152–154

defined, 119

design section, 124–156

file sources and, 124

importance of, 120

incremental. See incremental data extraction

metadata and, 123

metadata-driven extraction example, 156–165

objectives of, 120

problem section, 120–124

selection criteria, 120–121

separated from ETL, 119

solution section, 156–165

source system impact and, 121

staging database and, 123

data file archival process (scale-out ETL), 345

data file management (scale-out ETL process), 341–342

data flow

- advantages/disadvantages, 400–403
- backpressure, 408
- design example, 404–407
- destination issues, 392–394
- impact, 25
- package execution and, 25–28
- properties, setting of, 411–413
- source performance issues, 398–399
- /SQL operations balancing, 399–403
- tuning, 407–411
- unused columns and, 410–411

data flow design example, 404–407

- data flow–centric SSIS process, 406–407
- SQL–centric SSIS process, 404–406

data flow engine (pipeline architecture concepts), 403–404**Data Flow Tasks, 25, 106, 125, 141, 175, 177, 181, 185, 186, 189, 190, 191, 392, 405, 406, 411, 412****data flow–centric SSIS process (data flow design example), 406–407****data grain change, 255****data integration, 11–12****data lineage, 123**

- identifiers, 154–156

data nervous system. See also solution architecture

- agile methods and, 14
- organized, 10
- unorganized, 3–4, 8

data preparation

- dimension table ETL, 203–204, 210–214
- fact table ETL, 258, 260–264

data process. See process data**data processing**

- confusion, 4–5
- debilitating dependency chains and, 3, 11
- delayed data access and, 4
- fragile enterprise changes and, 4
- macro challenge, 2–4
- micro challenge, 4–5
- objectives, solution architecture and, 11
- tools, 9–10
- uncontrollable, 4
- unknown, 4

Data Profile Viewer, 168, 169, 171, 172, 174**Data Profiling Task, 30, 167, 168–174, 203**

- Candidate Key profiles, 170–171
- Column Length Distribution profiles, 171–172
- Column Null Ratio profiles, 172
- Column Pattern profiles, 172–173
- Column Statistics profiles, 173
- Column Value Distribution profiles, 173–174
- Functional Dependency profiles, 174

data scrubbing. See data cleansing**data silos, 11****data spaghetti, enterprise, 2–4****data staging method, 12, 30, 152–154****data tracking, 11****data tracking documentation, 15–18, 31****Data Warehouse Design Solutions (Adamson & Venerable), 198****data warehouse ETL, 12–13, 30, 197.**

- See also* ETL

- BI and, 197

- data processing architecture for, 13

data warehousing, 12–13

- dimension change tracking theory and, 204
- transactional systems v., 12

database growth, 414**database/destination optimization, 413–422****data–mining objects, 307, 311–312****The Data Warehouse Toolkit: The Complete Guide to Dimension Modeling, Second Edition (Kimball & Ross), 198****date dimension ETL**

- defined, 233

- processing, 243–246

Date dimension table, 200**DATEPART, 245, 293****datetime change indicators, 144****DATETIME COL, 421****dbo.sysssislog table, 391****dbo.sysssispackages table, 81****debilitating dependency chains, 3, 11****decision support, 12****de–duplication, 11, 119, 175, 177–180, 261****DefaultBufferMaxRows, 413****DefaultBufferSize, 413****degenerate dimensions, 156, 254****delayed data access, 4****deleted rows, incremental extraction and, 122****delimited files, 8, 29, 128, 133, 175, 185****dependency chains, source data and, 3, 11****deployment manifest file, 77, 96**

deployment methods, 29, 83–90, 94–96.

See also storage methods

application factors and, 77
automation, 95
deployment utility, 83–88
design section, 83–90
DTUtil.exe and, 71, 82, 90, 95
flowchart, 94–96
manual deployment, 88–90
multiple environments and, 76–77
problem section, 75–77
solution architecture and, 29
solution section, 94–96
standardization and, 76
/storage methods relationship, 96

deployment needs (deployment method flowchart), 95–96

deployment utility (SSIS Deployment Wizard), 83–88, 94, 95

Derived Column transformation

call center fact table and, 290, 291, 292, 293, 373, 375, 376, 385
currency rate fact table process and, 266–267
data flow–centric SSIS process and, 407
date dimension ETL and, 244, 245
defined, 261
Editor, 223
Fuzzy Grouping Transform and, 191, 192
Historical Attribute Inserts Output and, 223
measure calculations and, 271, 272, 288
missing dimension lookups and, 295
Product dimension table and, 211, 212, 214
SCD Wizard and, 224
Script Transform and, 376
SUBSTRING() and, 184
surrogate keys and, 265
uses of, 210

descending order sort, 107, 111, 112, 114, 192

destination adapter

OLE DB, 414–418
SQL, 26, 27, 414

Destination component, 27, 156, 372, 373

Destination Connection properties (File System Task), 99–100

destination data dictionary, 16–18

destination issues, data flow, 391–394

destination server package execution, 26–27

destination tables, indexes and, 414

destination/database optimization, 413–422

destinationDirectory variable, 367

development servers, 22

Dickinson, G., 37

different development methods, 34

DimCustomer table, 168, 170

DimCustomerProfile, 240, 241

dimension change tracking theory, 204

dimension change types, 204–209. See also

Type 0 fixed attributes; Type 1 changing attributes; Type 2 historical attributes

dimension editor, 309

dimension key lookups, 268–271, 286, 303

Dimension Processing destination, 312, 319–320

dimension surrogate keys. See surrogate keys

dimension table ETL, 197–252. See also

SCD Wizard

complexities in, 197, 202, 228
data preparation, 203–204, 210–214
design section
 advanced dimension ETL, 229–236
 fundamental dimension ETL, 203–209
dimension change types, 204–209
fact table ETL v., 257
inferred members, 207–209
junk dimensions, 228, 229
 defined, 234
 ETL, 234
 processing, 240–243
metadata and, 259
parent–child dimensions, 200, 201
 ETL, 231–233
 processing, 238–240
problem section
 advanced dimension ETL, 227–228
 fundamental dimension ETL, 197, 202–203
profile dimensions
 customer, 234, 241, 242
 defined, 234
 ETL, 234
 processing, 240–243
SCD Wizard. See SCD Wizard
snowflake dimensions, 200
 defined, 231
 processing, 237–238
solution section
 advanced dimension ETL, 236–252
 fundamental dimension ETL, 209–227

dimensional modeling, 30, 201, 211

dimensions, 30. See also dimension table ETL

attributes and, 198
basics, 198–201

Data Warehouse Design Solutions, 198
The Data Warehouse Toolkit: The Complete Guide to Dimension Modeling, 198

defined, 198
 degenerate, 156, 254
 fact tables and, 199–201
 junk, 228, 229, 234, 240–243
 parent–child, 200, 201, 231–233, 238–240
 profile, 234, 241, 242, 243
 snowflake, 200, 231, 237–238
 SSAS, 308–309
 star, 231

direct attached storage (DAS), 23

disk I/O. See I/O

Disk Queue Length, 396

disk volumes, 23

documentation

data element/data tracking, 15–18, 31
 proactive caching and, 306
 standardization and, 76

drive configuration, 23

DropDailyData, 350

DropHourlyData, 350

Dts object, 362

Dts.VariableDispenser.LockOneForRead
 method, 362

DTUtil.exe command–line tool, 71, 82, 90, 95

dummy Raw file, 112

duplicate data (de–duplication), 11, 119, 175, 177–180, 261

dynamic source queries, 134–137

OLE DB parameterized source query, 134–136
 variable bound SQL source approach, 136–137

E

element tracking, 16, 17

Employee dimension table, 200

parent–child relationships, 232–233
 SSAS solution, 308

encryption

DTUtil.exe and, 90
 msdb database and, 82
 Raw files and, 188
 SQL Server and, 39, 85
 un–encryption, 325

EndBatchAudit, 52

EndDate, 207, 208, 209, 223, 224, 225, 286

EngineThreads property, 404, 412

enterprise data spaghetti, 2–4

entity tracking, 16, 17

EntryPoint, 363

environment variable configuration, 38–39

ETL (extraction, transformation, loading), 1. See also data extraction; deployment methods; dimension table ETL; fact table ETL; management framework; scale-out ETL

collocation, 22

data extraction and. See data extraction

data warehouse, 12–13, 30, 197

BI and, 197

data processing architecture for, 13

transactional systems and, 23, 121

ETL process framework, 72

EvaluateAsExpression property, 140, 141

event handler stored procedures, 59–69

Event Watcher Task, WMI, 110

Excel Connection Manager, 126, 181

Excel connections, 126, 128, 131, 181

Excel source adapter, 131–133

ExclusionGroup property, 114, 115

Execute DDL Task

Analysis Services, 312, 316–317

with ASCMD.exe, 317–318

Execute Package Task, 20, 21

Execute SQL Tasks

Bulk Insert Task and, 24–25

control flow stored procedures and, 53, 55, 58
 Editor, 55, 58, 63, 65, 68

event handler stored procedures and, 59, 61, 62, 63, 65, 66

partition management and, 420

scale–out ETL and, 352

vdtLastModifiedDatetime and, 138

ExecutePreAggregatorPackageJob, 349, 351

execution trees, 404

existing infrastructure

deployment method flowchart, 95

storage method flowchart, 92

Expression Builder, 128, 140

external assembly (archiving system)

access to, 366–367

creation of, 363–366

Extract Source Data Task, 158, 161

Extract to Staging (task), 392

extraction. See data extraction

extraction, transformation, loading. See ETL

F

fact table ETL, 253–303. See also call center fact table; sales fact table package

- advanced concepts, 295–303
 - handling late-arriving facts, 300–303
 - handling missing dimension lookups, 295–299
- data preparation, 258, 260–264
- design section, 260–280
- dimension ETL v., 257
- dimension surrogate keys, 258, 265–271
- inferred members and, 295–299
- inserts, 259
 - MERGE statement and, 277–278
 - optimizing, 278–280
- measure calculations, 258–259, 271–273
- metadata and, 259
- problem section, 253, 257–259
- solution section, 280–303. *See also* call center fact table; sales fact table package
- updates, 259
 - MERGE statement and, 277–278

fact tables, 30

- in AdventureWorksDW2008 database, 199, 255
- basics, 254–257
- currency rate, 256
- defined, 198
- degenerate dimensions and, 156, 254
- dimensions and, 199–201
- finance, 256, 257
- Internet sales, 254, 255
- metadata and, 259
- snapshot, 255–257
- transaction, 255

FactCallCenter.dtsx, 294. See also call center fact table

FactInternetAndResellerSales_Load.dtsx, 288. See also sales fact table package

FactInternetAndResellerSales_Stage.dtsx, 284

FactResellerSales, 199, 200, 255

Fact_SalesOrderID, 151, 152, 282

FactSalesQuota, 277. See also sales quota fact table package

fast response, File Watcher Task and, 109

FastParse property, 129–130

file criteria, 98

File enumerator, 104, 105

file iteration, 98

file management. See also file-handling tasks

- advanced file management (archiving system) design section, 360–363
- problem section, 359–360
- solution section, 363–369

file management tasks (scale-out ETL process), 341–344

- data file management, 341–342
- scale-out source file process, 343
- work allocation process, 342
- work reassignment process, 343–344

file notification, 99

file preprocessor. See importing multiple files

file processor scale-out processes servers. See scale-out processes servers

file system storage, 79–80, 81–83

File System Task, 99–101

- attributes and, 100–101
- Destination Connection properties of, 99–100
- Foreach Loop and, 101, 106
- FTP Task v., 102
- multiple files and, 101, 106
- OverwriteDestination and, 101
- Raw files and, 111, 118, 190
- Source Connection properties of, 99–100

File Watcher Task, 109–110

- Event Watcher Task v., 110
- SQLIS.com and, 109, 110

file-handling tasks, 29, 97–118. See also file management; File System Task; Foreach Loop; FTP Task

- design section, 99–110
- disk I/O and, 8
- extraction and, 124
- File Watcher Task, 109–110
- importing multiple files (file preprocessor), 110–118
- problem section, 98–99
- remote, 98
- simple, 98
- solution section, 110–118
- WMI Event Watcher Task, 110

FileLog, 348

FileLogArchive, 348

FileManagement class, 364–365

FileManagement.dll, 366

Filtered Cache Lookup, 236
finance fact table, 256, 257
firewall, 103
Fixed and Changing Attribute Options screen, 217–218
Fixed Attribute Output, 226–227
fixed attributes. *See* **Type 0 fixed attributes**
flag column, 208, 209, 218, 222, 225
Flat File Connection Manager, 129, 177, 185, 186
Flat File Source, 114, 175, 177, 186
flat files
 with change identifier values, 147–149
 types of, 128
flat-file source adapter, 128–130
flowcharts
 deployment methods, 94–96
 storage methods, 91–94
Foreach Loop, 104–109
 advanced criteria applied to, 106–109
 enumerators, 104, 107, 108
 File enumerator, 104, 105
 File System Task and, 101, 106
fragile enterprise changes, data processing and, 4
framework. *See* **management framework**
FrequentValueThreshold property, 174
FTP Connection, 102–103
FTP Task, 101–104, 106, 392
 File System Task v., 102
 FTP Connection configuration, 102–103
 HTTP Connection, 103–104
full cache, 409
full outer join, 261, 268
full process. *See* **process full**
Fully Cached Lookup, 236
Functional Dependency profiles, 174
Fuzzy Grouping transformation
 as blocking transformation, 403, 408
 de-duplicating data and, 175, 177–180, 195
 Derived Column transformation and, 191, 192
 Fuzzy Lookup v., 180
 TempDB database and, 180
Fuzzy Lookup Transform
 data cleansing and, 175, 180–183, 195
 Fuzzy Grouping v., 180

 pulling dimension keys (based on similarity) and, 268–271
 surrogate keys and, 265
FuzzyGroupingExample.txt, 177

G

GAC (Global Assembly Cache), 366
Gacutil.exe (Global Assembly Cache Tool), 366
General page, 102, 169, 181, 191, 192
Get Source Columns Task, 160
Get Sources Task, 158
GetDailyAggregatorWorkLoad, 350
GetFileList, 349, 352
GetFilesToBeDropped, 350
GetHourlyAggregatorWorkLoad, 350
GetHourlyPreAggregateDateAndHour, 349
GetPreAggServerListForHour, 350
GetRawTablesToBeDropped, 350
Global Assembly Cache (GAC), 366
Global Assembly Cache Tool (Gacutil.exe), 366
global variables, 381
globally unique identifier (GUID), 200
GUID (globally unique identifier), 200

H

Herbold, D., 37
high availability, 21, 31, 333. *See also* **scale-out ETL**
Hinson, D., 37, 128, 147
Historical Attribute Inserts Output, 222, 223, 224
Historical Attribute Options screen, 218–219
historical attributes. *See* **Type 2 historical attributes**
Host Integration Server, 9
hourly data aggregation process (scale-out ETL process), 345
hourly data pre-aggregation process (scale-out ETL process), 344
HourlyProcessLog, 348
HourlyServerProcessLog, 348
HTTP (Hypertext Transfer Protocol), 103

HTTP Connection, 103–104
Hypertext Transfer Protocol. See HTTP

I

“ignore fixed attribute changes” option, 227
importing multiple files (file preprocessor), 110–118

Combine Files data flow, 113
Conditional Split Transformation Editor, 117
Merge Join transformation, 116
package control flow, 111
sample data, 111

incremental data extraction, 121–122, 133–152

capturing maximum change identifier value, 141–144
CDC and, 30, 122, 147
change identifier values, 134
flat files and, 147–149
change identifier values, capturing maximum
change identifier values, 141–144
changed source records
dynamic source queries, 134–137
identification of, 122
defined, 121, 133–134
SSIS and, 147–152
without trigger, 144–147

incremental process (processing partitions), 311

IncrementalDateMax, 161

inCustomers table, 181, 182, 190, 191

inCustomersClean table, 194

index optimizations (SCD transformation), 229

index process. See process index

indexes, destination tables and, 414

Inferred Dimension Members screen, 219–220

Inferred Member Boolean column, 219, 220

Inferred Member Flag column, 208, 209, 222

Inferred Member Updates Output, 222

inferred members, 207–209

fact table ETL and, 295–299
Script Component and, 299
support option, 219–220

infrastructure challenges, 7–8

InitialDatetime, 292

in-line data flow approach, 295, 296

inner join, 230, 261, 263, 268, 409

InputName_ ProcessInputRow, 374

InputName_ ProcessInput, 374

INSERT statements

bulk, 402
data flow destinations and, 392
SSIS data flow v., 401
SSISConfigurations table and, 47

Integration Services. See SSIS

Internet sales fact table, 254, 255. See also sales fact table package

I/O

disk
bottlenecks, 7–8, 23, 395, 396, 401
data staging and, 154
importance of, 7
incremental extraction and, 121
SSIS and, 335
memory, 401
network, 26, 28

IsDate function, 175

IsDestinationPathVariable, 100

ISNULL, 143, 192, 250, 276, 282

IsSorted property, 116, 246, 247

IsSourcePathVariable, 100

“Issue Aggregates,” 293

“Issue Stats,” 292, 293

IssueID, 290, 292, 293

IsTransferAscii property, 102

J

JET provider, 131

junk dimensions, 228, 229

defined, 234
ETL, 234
processing, 240–243
profile dimensions v., 234

K

keep identity option, 416

keep nulls option, 416

“keep nulls” option, 416, 417

KeyStrengthThreshold property, 171

Kimball, R., 198, 204
Kimball Group, Inc., 204
Knight, B., 37

L

large files, File Watcher Task and, 109
Late Fact, 301
late-arriving dimension scenario, 207
late-arriving facts, 300–303
left outer join, 150, 151, 235, 248, 261, 268, 281, 282
LevelOneDistinctOperator, 292
LevelOneOperators, 290, 292
LevelTwoDistinctOperator, 292
LevelTwoOperators, 290, 292
limiting database growth, 414
Load Detail (task), 392
Load Header (task), 392
log sequence number (LSN), 147
log-based auditing, 122
logging
 defined, 40
 impact of, 411
 troubleshooting and, 411
logging/auditing mechanism, 36, 40–42, 48–69
 audit tables, 122, 261, 289, 342, 403
 stored procedures in, 41–42, 52–69
 tables in, 40–41, 48–52
Logical OR operator (||), 250
LogPackageEnd stored procedure, 42, 57–59
LogPackageError store procedure, 42, 60–62
LogPackageStart stored procedure, 42, 53–57
LogTaskPostExecute stored procedure, 42, 64–66
LogTaskPreExecute stored procedure, 42, 63–64
LogVariableValueChanged stored procedure, 42, 66–69
Lookup cache, 409, 410
Lookup transformations
 currency rate fact table process and, 266–267
 data correlation with the data flow, 274
 data flow-centric SSIS process and, 407
 defined, 261
 DimCustomerProfile and, 241
 incremental data extraction and, 149, 152
 optimizing, 409–410

sales quota fact table and, 274–277
 surrogate keys and, 265
 uses of, 210

LSN (log sequence number), 147

M

Main subroutine, 363, 373
management framework, 29, 33–73
 architecture, 37
 batch logic modifications, 71
 benefits, 36–37
 configuration management scheme, 35, 38–40, 43–47
 customizations, 71–72
 design section, 37–43
 ETL process framework and, 72
 implementation guidelines, 43
 logging/auditing mechanism, 36, 40–42, 48–69
 multiple environments on server, 72
 not having, 33–35
 package requirements and, 33
 problem section, 33–37
 solution section, 43–72
 template package, 36, 42–43, 69–71
 value, 36–37
Management Studio. See SSMS
manual deployment, 88–90
many-to-many relationships, 202, 401
master data management service, 12
master dimension management, 12
master packages (parent packages), 19–21
 example, 20
master-child packages, 93
maximum insert commit size option, 417–418
Maximum operation, 292, 293
MDAC (Microsoft Data Access Component) providers, 131
MDX (Multidimensional Expressions) script, 308, 312, 317, 318
measure calculations
 fact table ETL, 258–259, 271–273
 sales fact table package, 272
measure group partitions, 309–311. See also partitions

measure groups, 231, 306, 307, 308, 309, 310, 313, 315

memory

- bottlenecks, 394–397
- cache size management, 409–410
- I/O operations, 401
- management, performance and, 397
- server hardware and, 21

“Memory: Available MBytes,” 395, 396, 397

Merge Join transformation

- Conditional Split and, 149
- data correlation with the data flow, 274
- data flow–centric SSIS process and, 407
- defined, 261
- importing multiple files (file preprocessor), 116
- with Presorted Input, 235
- with Sort, 235
- surrogate keys and, 265

MERGE statement

- inserts/updates and, 273, 277–278
- sales quota fact table and, 277–278

metadata

- changing, 34–35
- data lineage and, 123
- fact tables/dimension tables and, 259
- Process Partitions package and, 328–329
- tables, 157

metadata–driven extraction example, 156–165

- control flow, 157–161
- metadata tables in, 157
- script, 161–165

Microsoft ACE OLEDB 12.0, 131

Microsoft Data Access Component (MDAC) providers, 131

Microsoft Data Warehouse Toolkit (Kimball Group), 204

Microsoft SQL Server 2008

- Analysis Services. See SSAS
- Integration Services. See SSIS

Minimum operation, 292

Minimum Similarity column, 178

mining models, 311

mining structures, 311–312

missing dimension lookups, 295–299

missing records. See inferred members

“Modify the SQL statement” checkbox, 302

modular packages, 5, 18–19

- example, 19

MOLAP (Multidimensional Online Analytical Process) storage, 310, 311

msdb database, 81–83

MsDtsServer100 service, 78–79

MsDtsSrvr.ini.xml file, 79, 80

Multicast transformation

- Aggregate transformation and, 142, 143, 149
- call center fact table and, 290, 291, 292
- data flow–centric SSIS process and, 407
- data staging and, 153
- DimCustomerProfile and, 241

multicore processors, 21

Multidimensional Expressions (MDX) script, 308, 312, 317, 318

Multidimensional Online Analytical Process (MOLAP) storage, 310, 311

multiple environments

- on server (management framework), 72
- storage/deployment methods and, 76–77

multiple files

- File System Task and, 101, 106
- importing, file preprocessor and, 110–118

multiple record types in single file (data cleansing), 183–188

N

NAT (Network Address Translation) device, 103

.NET application deployments, 90

.NET framework, scripting and, 370

.NET languages, soundex routine and, 177

.NET libraries, 162

Network Address Translation (NAT) device, 103

network I/O, 26, 28

New Output, 222, 223, 224

new package ID, 43, 71, 80

New Sales Facts, 282

No Match Output, 116, 181, 182, 214, 241, 269, 270, 283, 297

NT File System (NTFS), 82

NTFS (NT File System), 82

NULLs

- business key and, 249
- Candidate Key profile request and, 170
- Column Null Ratio profile request and, 172
- Conditional Split and, 186
- currency rate and, 283
- data preparation (dimension ETL) and, 203, 204
- Data Profiling Task and, 168

Derived Column and, 210
 dimension foreign keys and, 295
 EndDate and, 207, 208, 209, 223, 224, 225, 286
 Fact_SalesOrderID and, 151, 152
 Inferred Member Boolean column and, 219,
 220
 ISNULL and, 143, 192, 250, 276, 282
 “keep nulls” option, 416, 417
 Row.ProductNumber_IsNull statement and, 176
 Unknown and, 213

O

ODBC (Open Database Connectivity) driver, 124
Object Explorer window, 79
OLAP (Online Analytical Processing), 119
 cube structures, 305
 Multidimensional, 310, 311
 Relational, 306, 310, 311
 XMLA and, 309
OLE DB Command transformation, 143, 220,
222, 223, 228, 229, 236, 277, 297, 298,
329, 402, 408
OLE DB connection, to AdventureWorksDW2008
database, 47
OLE DB destination adapter, 414–418
OLE DB parameterized source query, 134–136
OLE DB provider, 124, 136, 221, 418
OLE DB source adapter, 130–131
 one-to-many relationships, 202
 one-to-one relationships, 258, 271, 401
 Online Analytical Processing. *See* OLAP
Open Database Connectivity (ODBC) driver, 124
Oracle Attunity destination adapters, 418–419
Oracle Attunity source adapters, 133
ORDER BY statement, 246, 263
OrderDateAlternateKey, 301, 302, 303
Output Alias, 186, 187, 292
 overridable methods, 373–374
OverwriteDestination, 101

P

Package Configurations Organizer, 45, 46
 package connections, 124–128

package control flow
 importing multiple files (file
 preprocessor), 111
 index management, 280
 modular package, 19
package design
 patterns, 18–21
 performance and, 404–407
package execution
 BIDS and, 75
 data flow and, 25–28
 location, 24–28
 problems, 5–7
package ID, new, 43, 71, 80
package makeup (storage method flowchart), 93
package performance. *See* performance
package scaling. *See* scalability
Package Store, 78, 82, 86, 90
Package table, 40, 48, 49
PackageErrorLog table, 41, 50–51
PackageLog table, 40, 50
PackageLogID, 52
packages. *See also* master packages;
modular packages
 deployment. *See* deployment methods
 management framework, 33–73
 storage. *See* storage methods
 unmanageable, 4–5
PackageTaskLog table, 41, 51
PackageVariableLog table, 41, 52
PackageVersion table, 40, 49
 parameterized source query, OLE DB, 134–136
 parent packages. *See* master packages
 parent-child dimensions, 200, 201
 ETL, 231–233
 processing, 238–240
partial cache, 409
partially blocking transformations, 403
Partition Processing destination, 312,
320–321
partitioned table example, 420–422
 complete code/solution for, 421
 control flow for, 420
 DATETIME COL and, 421
partitioned tables, 419–422
PartitionRangeDateKey, 421
partitions
 creating/managing, 323
 measure group, 309–311
passive FTP, 103

pattern matching, 176. See also regular expressions

Perfmon. See Performance Monitor performance, 31. See also scalability

- bottlenecks and, 394–397
 - disk I/O, 7–8, 23, 395, 396, 401
 - transformation/memory, 394–397
- data flow source performance issues, 398–399
- memory management and, 397
- package design and, 404–407
- scalability and, 389–422
- troubleshooting, 335

Performance Monitor (Perfmon), 395

Performance Monitor counters, 394, 395–396

“Physical Disk: Average Disk Queue Length,” 395, 396, 397

pipeline, 214

“Pipeline: Buffers spooled” counter, 395, 396

pipeline architecture concepts, 403–404

PipelineExecutionPlan, 404

PipelineExecutionTrees, 404

Pivot transformation, 261

post-data flow approach, 295, 296

PostExecute method, 374

pre-aggregation scale-out processes servers. See scale-out processes servers

pre-data flow approach, 295, 296

PostExecute method, 374

pre-sort sources, Sort transformation v., 235, 246, 247, 263, 274, 408

proactive caching, 310

process add (processing dimensions), 309

process clear (mining structures), 312

process data

- processing dimensions, 309
- processing partitions, 311

Process Dimensions package, 324, 325–327

- execution, 333
- parallel XMLA, 326–327
- /Process Partitions and overall solution, 332–333
- Process Task, 325–326

process full

- mining structures/mining models, 312
- processing dimensions, 309
- processing partitions, 310

process index

- processing dimensions, 309
- processing partitions, 311

process input (Script Component/call center fact table), 383–384

Process Partitions package, 324, 327–332

- execution, 333
- /Process Dimensions and overall solution, 332–333
- SSAS processing, 329–332
- storing/loading metadata, 328–329

process update (processing dimensions), 309

Processing Task (SSAS), 312, 313–316

ProcessInput (·InputName·ProcessInput), 374

ProcessInputRow, 175, 176, 193, 374

ProcessInputRow (·InputName·_ProcessInputRow), 374

Product dimension table, 200

- Derived Column transformation and, 211, 212, 214
- inferred members and, 208–209
- pre-processed source data, 203–204
- table structure, 202
- type 2 attribute and, 206–207

ProductAlternateKey, 206, 208, 209, 299, 302

ProductKey_inferred, 298, 299

Professional Microsoft SQL Server 2008

Integration Services (Knight, Veerman, Dickinson, Hinson, and Herbold), 37, 128, 147

profile dimensions

- customer, 234, 241, 242
- ETL, 234
- junk dimensions v., 234
- Multicast transformation and, 241
- processing, 240–243

profile requests

- Candidate Key, 170–171
- Column Length Distribution, 171–172
- Column Null Ratio, 172
- Column Pattern, 172–173
- Column Statistics, 173
- Column Value Distribution, 173–174
- Functional Dependency, 174

Profile Viewer, 171, 174. See also Data Profile Viewer

project planning, 14–18

Property Expressions, 126–128

Property Expressions Editor, 47, 127

Q

Quick Profile option, 168, 170

R**Raw files**

- data cleansing and, 188–190
- defined, 111
- dummy, 112
- encryption and, 188
- File System Task and, 111, 118, 190

RDBMSs (relational database management systems), 29, 40, 91, 92, 93, 235, 274, 401

RecordFile, 349

Reference Table tab, 182, 183, 213

RegEx. See regular expressions

regular expressions (RegEx), 172, 176, 195.

See also Script Transform

regulation, of transactions, 411

relational database management systems. See RDBMSs

Relational OLAP (ROLAP), 306, 310, 311

remote file operations, 98

Reseller dimension table, 200, 205

reseller sales fact table, 255. See also sales fact table package

ResolutionDatetime, 292

restating history, 205

reusability, 35

- custom components and, 362

reusable methods (Script Component/call center fact table), 379–380

Reuse Existing button, 45, 46

‘right information to the right people at the right time,’ 35

ROLAP (Relational OLAP), 306, 310, 311

rollback layers, 11

Ross, M., 198

Row Count transformation, 136, 142, 226, 395

row transformations, 403

row-by-row processing

- limiting of, 408
- Script Component/call center fact table and, 380–382

Row.ProductNumber_IsNull statement, 176

rows per batch option, 417

rowversion data type, 145, 146

runaway stored procedures, 4

RunInOptimizedMode, 413

Running Packages folder, 80

S

sales fact table package (Internet and Reseller sales fact tables), 280–288

data flow, 296

extraction and transform process, 281–284

FactInternetAndResellerSales_Load.dtsx, 288

FactInternetAndResellerSales_Stage.dtsx, 284

load process, 285–288

measure calculations, 272

overall fact ETL process, 280–281

sales header table/sales detail table

consolidation, 262–263, 280

sales order line number, 263, 264, 280

Script Component and, 263, 264

transaction fact tables and, 255

Union All transformation and, 283

sales quota fact table package

data flow for, 268

FactSalesQuota and, 277

Lookup transformation and, 274–277

MERGE statement and, 277–278

Sales.Customer table, 145, 146

SalesOrderLineNumber, 264, 280

SANs (storage area networks), 23, 406

scalability (package scaling), 31, 389–422.

See also performance

data flow design example, 404–407

data flow-centric SSIS process, 406–407

SQL-centric SSIS process, 404–406

design section, 399–407

data flow design example, 404–407

SQL operations/data flow balance, 399–403

SSIS pipeline architecture concepts, 403–404

problem section, 390–399

data flow destination issues, 392–394

data flow source performance issues, 398–399

memory bottlenecks, 394–397

scalability (package scaling) (continued)

- task durations, 390–392
- transformation bottlenecks, 394–397
- solution section, 407–422
 - database/destination optimization, 413–422
 - setting data flow properties, 411–413
 - tuning data flow, 407–411
- scale-out ETL, 31, 335–357. See also central common services server; scale-out processes servers**
 - archival and cleanup processes, 345–346
 - data file archival process, 345
 - stage table clean-up process, 346
 - data aggregation tasks, 344–345
 - daily data aggregation process, 345
 - hourly data aggregation process, 345
 - hourly data pre-aggregation process, 344
 - design section, 336, 337–346
 - file management tasks, 341–344
 - data file management, 341–342
 - scale-out source file process, 343
 - work allocation process, 342
 - work reassignment process, 343–344
 - problem section, 335, 336–337
 - solution section, 346–357
 - code, 346
- scale-out processes servers (file processor/pre-aggregation/staging). See also central common services server**
 - advantages of, 346
 - packages for, 351–354
 - redundant tasks of, 347
 - role of, 338–340
 - SCOSTage 01—File Processor.dtsx, 351–354
 - SCOSTage 02—Hourly PreAggregator.dtsx, 354
 - stored procedure for, 351
- scale-out source file process (scale-out ETL process), 343**
- SCD (Slowly Changing Dimension) Wizard, 198, 214–229. See also custom dimension ETL**
 - advanced properties
 - CurrentRowWhere property, 225
 - SQLCommand property, 220, 222, 223, 225
 - advantages, 214, 227–228, 229
 - change types identification, 216
 - column matches defined, 215
 - as Conditional Split, 220
 - Derived Column transform and, 224
 - disadvantages, 228
 - optimization, 229–230

- index optimizations, 229
- update optimizations, 229–230
- outputs, 220
 - Changing Attribute Updates Output, 220–222
 - Fixed Attribute Output, 226–227
 - Historical Attribute Inserts Output, 222, 223, 224
 - Inferred Member Updates Output, 222
 - New Output, 222, 223, 224
 - Unchanged Output, 226
- screens
 - final, 220
 - first, 215
 - Fixed and Changing Attribute Options screen, 217–218
 - Historical Attribute Options screen, 218–219
 - Inferred Dimension Members screen, 219–220
- SCOMain 01—Hourly Aggregator.dtsx, 355**
- SCOMain 02—Daily Aggregator.dtsx, 356**
- SCOMain 03—Cleanup.dtsx, 356–357**
- scope creep, 8**
- scope management, agile and, 14**
- SCOSTage 01—File Processor.dtsx, 351–354**
- SCOSTage 02—Hourly PreAggregator.dtsx, 354**
- Script Component Transformation Editor, 376–377**
- Script Components, 31, 294**
 - call center fact table
 - design section, 371–374
 - problem section, 370
 - solution section, 374–387
 - date dimension ETL and, 244
 - defined, 261
 - extraction data flow and, 154
 - features of, 264, 272
 - inferred members and, 299
 - sales fact table and, 263, 264
 - Select Script Component Type dialog box, 371–372
 - unneded, 363
- script languages, for Script Tasks, 361**
- Script Tasks, 31, 59, 101, 106**
 - accessing objects, 362
 - advanced file management
 - design section, 360–363
 - problem section, 359–360
 - solution section, 363–369

- custom assemblies and, 362
 - reasons for usage, 360
 - script languages for, 361
 - unnneeded, 363
 - Script Transform**
 - data cleansing and, 175–177, 193, 195
 - Derived Column Transform and, 376
 - scripting design patterns, 31, 359–387**
 - call center fact table, 370–387
 - design section
 - advanced file management, 360–363
 - call center fact table, 371–374
 - problem section
 - advanced file management, 359–360
 - call center fact table, 370
 - solution section
 - advanced file management, 363–369
 - call center fact table, 374–387
 - scrubbing data. See data cleansing**
 - security questions (storage method flowchart), 92–93**
 - SELECT *, 411**
 - Select Script Component Type dialog box, 371–372**
 - SELECT statement, 55, 145, 411**
 - SellStartDate column, 175**
 - server hardware, 21**
 - service level agreements (SLAs), 8, 229**
 - Service Level Calculation Derived Column transformation, 385**
 - ServiceGrade, 293**
 - shadow table, 147**
 - shared central server. See central common services server**
 - “Similarity threshold,” 179, 182, 183, 191, 270**
 - simple file operations, 98**
 - 64-bit servers, 21, 131, 236, 267, 335**
 - SLAs (service level agreements), 8, 229**
 - slowly changing dimension phrase, 204**
 - Slowly Changing Dimension Wizard.**
 - See SCD Wizard
 - snapshot fact tables, 255–257. See also call center fact table**
 - snowflake dimensions, 200**
 - defined, 231
 - processing, 237–238
 - solution architecture, 1–31**
 - deployment and, 29. See also deployment methods
 - design section, 8–28
 - data processing objectives, 11
 - data processing tools, 9–10
 - development servers, 22
 - ETL collocation, 22
 - project planning, 14–18
 - server hardware, 21
 - storage hardware, 22–23
 - storage methods and, 29. See also storage methods
 - team development, 14–18
 - test servers, 22
 - management framework and, 29. See also management framework
 - problem section, 2–8
 - corporate political challenges, 8
 - data challenges, 8
 - data processing confusion, 4–5
 - enterprise data spaghetti, 2–4
 - infrastructure challenges, 7–8
 - macro challenge, 2–4
 - micro challenge, 4–5
 - package execution problems, 5–7
 - solution section, 28–31
 - source data and, 29–30
- Solution Explorer, 71, 125, 309**
- Sort transformation**
 - BIDemo database and, 192, 193
 - as blocking transformation, 403, 408
 - call center fact table and, 293
 - data flow impact and, 25
 - data flow–centric SSIS process and, 407
 - defined, 261
 - DimCustomerProfile and, 241
 - Merge Join with, 235
 - ORDER BY statement and, 246, 263
 - pre-sort sources v., 235, 246, 247, 263, 274, 408
- sorting**
 - ascending order, 107
 - descending order, 107, 111, 112, 114, 192
- SortKeyPosition, 247**
- soundex routine, 177**
- source adapters, 124, 128–133**
- Source component, 372, 373**
- Source Connection properties (File System Task), 99–100**
- source data, 29–30. See also data cleansing; data extraction; dimension table ETL; fact table ETL; file management**
 - dependency chains and, 3, 11
 - dictionary, 15–16

- source performance issues, data flow, 398–399**
- Source Safe, 82, 83**
- source server package execution, 25–26**
- source system impact, data extraction and, 121**
- sourceDirectory variable, 367**
- Sources object variable, 158, 159**
- sp_configure, 22**
- SQL Destination adapter, 26, 27, 414**
- SQL INSERT statements. See INSERT statements**
- SQL Native Client (SQLNCI), 130**
- SQL operations/data flow balancing, 399–403**
- SQL Server**
 - configurations (configuration management scheme), 39, 45–46
 - encryption and, 39, 85
 - storage, 81–83
- SQL Server 2008**
 - Analysis Services. See SSAS processing architecture
 - Integration Services. See SSIS Management Studio. See SSMS
- SQL Server Integration Services 10.0, 78**
- “SQL Server SSIS Pipeline 10: Buffers spooled,” 395**
- “SQL Server SSIS Pipeline 10: Rows read,” 395**
- SQL Server TempDB, 7**
- SQL-centric SSIS process (data flow design example), 404–409**
- SQLCommand property, 220, 222, 223, 225**
- SQLIS.com, 109, 110**
- SQLNCI (SQL Native Client), 130**
- SSAS dimensions, 308–309**
- SSAS (SQL Server 2008 Analysis Services) processing architecture, 30, 305–333**
 - basics, 307–312
 - BI and, 197, 305, 306, 312, 323, 333
 - design section, 307–323
 - problem section, 305–307
 - solution section, 323–333
 - transactional systems and, 308
- SSIS (SQL Server 2008 Integration Services), 2**
 - benefits of, 92, 335
 - incremental data extraction. See incremental data extraction
 - solution architecture, 1–31
 - tool selection and, 9–10
 - uses of, 11–12
- SSIS Deployment Wizard. See deployment utility**
- SSIS Service, 24, 78–79**
- SSISConfigurations table, 45, 47, 69**
- .SSISDeploymentManifest file, 83, 84, 87**
- SSIS_PDS_CONFIG_FILE, 43, 45**
- SSIS_PDS_Template, 45, 52**
- SSIS_SQL_Design.dtsx, 392**
- SSMS (SQL Server Management Studio), 79, 81, 82, 83, 87, 88, 90, 309**
- stage table clean-up process (scale-out ETL), 346**
- staging databases, 7, 118, 121. See also data staging method**
 - benefits of, 123
- staging scale-out server. See scale-out processes servers**
- standalone SSIS servers, 27–28**
- standardization, 76. See also deployment methods; storage methods**
- star dimension, 231**
- star schema, 231**
- storage area networks (SANs), 23, 406**
- storage hardware, 22–23**
- storage methods, 29, 77–83, 91–94. See also deployment methods**
 - application factors and, 77
 - comparison, 81–83
 - /deployment methods relationship, 96
 - design section, 77–83
 - flowchart, 91–94
 - multiple environments and, 76–77
 - questions (deployment method flowchart), 95
 - solution section, 91–94
 - standardization and, 76
 - troubleshooting and, 82–83
- stored procedures. See also specific stored procedures**
 - on central common services server, 349–350
 - control flow, 53–59
 - event handler, 59–69
 - logging/auditing mechanism and, 41–42, 52–69
 - runaway, 4
 - on staging servers, 351. See also scale-out processes servers
 - TSQL and, 297
- streaming transformations, 403, 404**
- strong name, 365**
- strVariableValue, 52**
- SUBSTRING(), 184**

surrogate keys

- advantages of, 201
- defined, 200, 201
- fact table ETL and, 258, 265–271

SuspectRowFlag, 175**synchronous outputs, 372–373****System::StartTime variable, 219, 223, 224****System.Text.RegularExpressions namespace, 176****T****table lock, 250, 416, 417****table trigger, 134, 146****tables. See also specific tables**

- logging/auditing mechanism and, 40–41, 48–52
- partitioned, 419–422

task durations (scalability), 390–392**team development, 5, 14–18****TempDB databases**

- Fuzzy Grouping and, 180
- Fuzzy Lookup and, 270
- SQL Server, 7
- staging database and, 23

template package

- development, 69–70
- installation, 71
- management framework and, 36, 42–43, 69–71

temporary storage locations, optimizing, 412**Teradata Attunity destination adapters, 418****Teradata Attunity source adapter, 133****test servers, 22****timestamps, 16, 110, 114, 134, 390****“Token delimiters,” 179****tracking history, 205****Transact SQL. See TSQL****transaction fact tables, 255. See also sales fact table package****transactional systems**

- data warehousing v., 12
- ETL and, 23, 121
- SSAS and, 308

transactions, regulation of, 411**transformation bottlenecks, 394–397****Transformation component, 372****transformations, 403–404. See also specific transformations****triggers, 122, 134, 145, 146, 156, 183, 344, 394****troubleshooting**

- agile methods and, 14
- data tracking and, 11
- logging and, 411
- management framework and, 29
- modular packages and, 18
- package performance and, 335
- solution architecture and, 5–7
- storage methods and, 82–83
- unmanageable packages and, 4, 5

Truncate Staging (task), 392**TSQL (Transact SQL), 9**

- dbo.sysssislog table and, 391
- sp_configure, 22
- stored procedure, 297
- update optimizations (SCD) and, 230

tuning buffers, 413**tuning data flow, 407–411****Type 0 fixed attributes, 204, 214, 216, 252****Type 1 changing attributes, 205–206, 214, 216, 252****Type 2 historical attributes, 205, 206–207, 214, 216, 252****U****UnassignFileProcessorWorkLoad, 350****Unchanged Output, 226****uncontrollable data processing, 4****un-encryption, 325****Unicode data types, 131, 133, 175, 210, 378****UNION ALL operation, 261****Union All transformation, 299**

- defined, 261
- Fuzzy Lookup and, 183
- Historical Attribute Inserts Output/New Output and, 222
- pipeline and, 214
- sales fact table package and, 283
- sales quota fact table package and, 270

unique identifier, 200**Unknown, NULLs and, 213****unknown data processing, 4**

unmanageable packages, 4–5

Un-pivot transformation, 261

unprocess

mining structures/mining models, 312

processing dimensions, 309

processing partitions, 311

unused columns, data flow and, 410–411

Update Attributes (task), 392

**update optimizations (SCD transformation),
229–230**

update process. See process update

**UPDATE statement, 143, 220, 221, 222, 223,
228, 230, 238, 252, 402, 405, 408**

UpdateHourlyAggregatedFileStatus, 350

UpdateMultipleFileStatusStaged, 349

USCustomers.xls, 180, 181

V

ValueDistributionOption, 174

**variable bound SQL source approach,
136–137**

Variables object, 362

**vdtLastModifiedDate, 135, 136, 137, 138,
139, 141, 148, 159**

Veerman, E., 37, 128, 147

Venerable, M., 198

Visual Studio Tools for Applications, 361, 367

vsExtractionSQL, 136, 138, 139, 141

W

wildcards, 101, 104, 106, 110

Windows Management Instrumentation (WMI)

Event Watcher Task, 110

WMI (Windows Management Instrumentation)

Event Watcher Task, 110

**work allocation process (scale-out ETL
process), 342**

**work reassignment process (scale-out ETL
process), 343–344**

X

XCOPY command, 90

XML configuration file, 39, 44

XML for Analysis. See XMLA

XML Schema Definition (XSD), 133

XML source adapter, 133

XML Task, 168, 169

XMLA (XML for Analysis), 309

OLAP structures and, 309

process add and, 309

Process Dimensions package and,
326–327

XPath operation, 168

XSD (XML Schema Definition), 133