Index

Access, single point, 480
Access tool data stores, 375
Accumulating snapshot fact table, 223
Activity metrics, 123
Activity monitoring services, 380–381
Additive facts, 193–194
Additivity, 144
Addresses:
  coding, for people, 175–179
  processing, 655–656
Ad hoc query tools, 21, 394
Affinity grouping, 403–404
Aggregate, 543–567
  administering, 556–555
  atomic data marts, 347
  awareness, 383–384
  benefits, 543
  boundaries, 554–555
  classifying records by, 638–639
  deciding what to, 311–312, 545–547
  family of fact tables, 211–212
  loading, 554
  processing, 549–552
  surrogate keys, 551
  using wrong technology, 550
Aggregate navigation, 544
  benefits, 566
Aggregate navigation algorithm, 560–566
  handling MIN, MAX, COUNT, and AVG, 565–566
  intersecting planning data, 564–565
  steps, 561
Aggregate navigation system, design goals, 555–560
  dimension tables attached to aggregate fact tables, 557–559
  family of schemas, 559–560
  separation into fact tables, 556–557
  SQL referring to base fact table, 560
Aggregate table:
  fact tables, 647–649
  plan, 547–549
  sum of, 552–553
Aggregation, 362, 437–438, 622
  designing for, 310–311
  goals and risks, 544–545
  premature, 333
Allocating, facts of differing granularity, 240–250
Alpha release, 706
Analysis-specific transformation, 362–363
Analyst, interviews, 117–120
Analytical cycle, 671–673
Analytical power, 526–527
Analytic culture, current, 45
ANSI SQL 92 support, 396
Application server, 388–389
  UNIX- or NT-based platforms, 424
Architectural requirements, document, 506–507
Architecture, 317–334
  as communications tool, 318–319
  component software, 452–457
  creating, 502–514
    flow chart, 503
    information resources, 512, 514
    project plan tasks, 536–541
    infrastructure, plan, 511–515
    technical architecture, plan, 505–511
  cross-check for project plan, 319
  data access. See Data access
  data stores, 375–378
  as documentation for system, 320
  flexibility and maintenance, 319
  framework, 320–327
    defining columns, 322–324
    defining level of detail, 324–326
    logical and physical models, 326–327
  front room, 373–375
  impact of Internet, 454–457
  infrastructure, 324
  models, level of detail, 325
  productivity and reuse, 320
  server, 451–452
  task force, 505
  technical. See Technical architecture
  value, 318–320
Architecture-related requirements, 505–506
Archiving, 368–369
  data staging area, 659
  planning, 369
ATM transactions, 206–207, 210
Atomic data mart, 427
Attributes, 165–166
  artificial, 285
  breaking off into separate dimension table, 184–187
  commercial customer address coding, 179–180
  detail descriptions, 282–285
  name and address coding, people, 175–179
  open-ended many-valued, 222
  time dimension, 175
  validating one-to-one and one-to-many relationships, 621
Audit, 24, 438
  data, 103
  statistics, fact tables, 631–632
  Audit dimension, building, 237–238
  Authentication, 380, 388
    Kerberos, 491–493
    tokens, 494–495
  Authorization, 388, 480
  Availability, 708
  AVG, 565–566
  Backup, 24, 367–368
  planning, 369
  Bandwidth, 432
  Basket analysis, 208, 260–263
  Bastion server, 478
  Behavior tracking, 254–260
  Benefits, 59–62
  Beta release, 706–708
  Beta team, 707–708
  Beverage store sales case study, 577, 581–583, 586, 592–594
  Biometric scanning, 494–495
  Bitmapped index, 588, 597
  snowflaking effect, 172
  Blocksize, 595
  Bottlenecks, 429
  Browser, importance, 170
  Browsing, 379
  data content, 303–304
  snowflaking effect, 171–172
  B-tree index, 587, 596
  Bulk loader, 621–623
  Bus architecture, 153–164
    conformed dimensions, 147, 156–157
    commitment to use, 158–159
    designing, 158
    when not needed, 163–164
    conformed fact definitions, establishing, 159–160
    data marts, 155
    granularity, 160–161
    multiple-source, 161–162
    planning crisis, 153–155
    rescuing stovepipes, 162–163
    standardized fact definitions, 156–157
  Business community, communication with, 92
  Business decision, 673
Business Dimensional Lifecycle, 31–40
  application, 36
  business manager, 117–120
  business requirements definition, 34
  data staging design and development, 35
  deployment, 37
  diagram, 33
  dimensional modeling, 34–35
  evolution, 31–33
  guidelines for using, 38–39
  maintenance and growth, 37
  mile markers and navigation aids, 39–40
  physical design, 35
  product selection and installation, 36
  project management, 37–38
  project planning, 33–34
  technical architecture design, 35–36
Business driver, 66
  management, 88–89
Business executive, interviews, 116–117
Business impact metrics, 123
Business interviewees, selecting, 102–103
Business justification, 58–64
  combine investments and returns to calculate ROI, 62–63
  financial investments and costs, 58–59
  financial returns and benefits, 59–62
Business management sponsor, 43–44
Business manager, Business Dimensional Lifecycle, 117–120
Business motivations, 44–45
Business policy, critical issues, 307–308
Business process:
  data mart, aggregated, 348–350
  definition, 17
  number of, 415
Business requirements, 95–136
  definition, 96–98
  facilitation, 98
    as alternative to interviewing, 132–133
    for confirmation, prioritization, and consensus, 132
  findings document, 130–131
  high-level analysis, 50
  interview team, 98–100
  key roles, 135
  level of detail, 324–325
pre-interview letter, 110
pre-interview research, 101–102
prepare and publish deliverables, 128–131
prioritization, 50–53
project plan tasks, 134–136
templates, 136
user kickoff meeting agenda, 108–109
your interview role, 111–112
see also Interviews
Business requirements definition, 34
Business sponsors, 65–66
  management, 88–89
Business systems analyst, 67–68
Business trends, determining, 672
Business user:
  comatose, 125
  community, 89–90
  core, working with, 292–293
  education, continuing, 718–720
  focus on, 718–720
  overzealous, 125
  present model to, 293–294
  support, continuing, 718–719
  uncooperative, 124–125
Case sensitivity, assuming, 573
Catalog mailings, accurate focusing, 62
Certificate authority, trusted, 487–489
Chains, 200–202
Changes, controlling, 84, 86–87
Charting, 398
Circles, 202–203
Classifying, 402
Class words, 571
Client/server ERP systems, 338–339
Clustering, 402
Code, generation, 358
Code tree, example, 576
Columnar databases, 589
Columns:
  defining, 322–324
  nullable, 579
Commercial customer, address coding, 179–180
Common Object Request Broker Architecture, 453
Communications:
ongoing, importance, 726–727
secure, 484, 486–487
with users, 702–703
Component software, architecture, 452–457
Computed fields, as inputs or targets, 639–640
Confirmation, 132
Conformed dimensions, 147, 156–157
commitment to use, 158–159
designing, 158
when not needed, 163–164
Conformed fact definitions, establishing, 159–160
Connectivity, 400, 432–434
Console, single, 481
Constraints, successive, 395
Consumer, multiple types, 390
Control documents, 188
Cookies, 495
Costs, 58–59
COUNT, 565–566
Coverage factless fact table, 214–216
Cross-footing, 658
Cross-platform support, 430
CS/MA/CD network, 472
Cube, 165
Currency, multinational, tracking, 248–250
Current rolling snapshot, 206, 209
Customer, improved acquisition capabilities, 61–62
Customer access, Web-based, 389
Customer matching, 302–303

Data:
accessibility, 300
accuracy, 300
challenges, 526
cleaning issues, 656–658
content
audit, 362
browsing, 303–304
simplification, 382
critical issues, 307
decompression, 360
division into training, test, and evaluation sets, 639
feedback, reverse, 24
improvement, 654–658
informal, 296
issues, 78
lineage audit, 362
loading services, 363–364
longevity of feed, 300
management processes, 722–723
manual examination, 658–659
not balancing, 684
private encryption, 482
problems, 654
quality, characteristics, 653
quality assurance, 658–659
reconciliation, support, 701–702
restructuring, fact tables, 636–637
scarsity, 310
shared secret encryption, 482
size, 414
sparsity, 548–549
striping, 599
structure, physical, design, 578–583
transformations, 619
services, 360–363
types, 579
conversion, 361
verification, 683–684
Data access, 378–409
access and security services, 380
activity monitoring services, 380–381
desktop services, 389–400
data access tool capabilities, 392–400
multiple consumer types, 390
multiple information needs, 391–392
direct connect, 406–407
future services, 388–389
modeling applications and data mining, 400–404
MOLAP, application layer, 408–409
query services, 381–385
locations, 385–386
ROLAP, application layer, 407
standard reporting services, 386–388
warehouse browsing, 379
Web implications, 404–405
Data access tool, 331
architecture approaches, 406–409
capabilities, 392–400
ad hoc query tools, 394
analysis and presentation capabilities, 396–398
push-button access, 392–393
query formulation, 394–396
standard reports, 393–394
technical features, 399–400
user experience, 398–399
evaluation, 532–535
metadata definition, 681
selection, 515
Data acquisition. See Technical architecture, back room
Data administrator, 68
Data architecture area, 322–323
Data audit, 103
interview, 103–104
Database:
build scripts, 597–598
connectivity, 433
logs, 642
parallel, 645
performance tuning, 722
relational versus multidimensional, 425
temp, 601
Database management system:
metadata, 438
platform
evaluation, 531
selection, 515
Database object, naming standards, 571–573
Database vendor, development focus, 424
Data cube, 165
Data Encryption Standard algorithm, 482
Data extraction tools, mainframe-based, 356
Data map, source-to-target, 305–307
Data mart, 26–27, 154
aggregated business process, 348–350
atomic, 347–348, 427
bus architecture, with, 155
choosing, 194–195, 273
definition, 17–18
disposable, 377
granularity, 160–161
list, 268–270
multiple-source, 161–162, 267
overlapping another data mart, 267
personal, 376–377
stovepipe, 154–155
rescuing, 162–163
Data mining:
origins, 401
systems, 377
transformations, fact tables, 637–640
unusual case, 640
Data model:
atomic data mart, 348
data staging, 345–346
Data modeler, 68
Data modeling tool, 578
selection, 533
using, 309–312
Data providers, 299
Data push, 371–372
Data sources:
candidate, 297–298
data stores, 336, 338–344
definitions, 298
extracting data, 357
flow to user desktop, 330–331
mapping to target, 305–307
multiple, 358
ownership, 299
selection criteria, 299–302
Data staging, 22–24, 344–350, 609–661
aggregated business process data marts, 348–350
atomic data marts, 347–348
data improvement, 654–658
data models, 345–346
data quality assurance, 658–659
design and development, 35
disk space management, 660
effective planning, 611–617
detailed plan, 615–617
high-level plan, 612–613
tools, 612–615
job control services, 364–366
metadata, 437
metadata driven, 352–355, 441–443
overview, 610
preliminary work, 611
presentation servers, 346–347
process validation, 659
requirements, 351–355
source system rollback segments, 660
Data staging (cont.)
storage types, 344–345
tool selection, 515
warehouse operation, 650–652
see also Dimension tables; Facts, atomic-level; Technical architecture, back room
Data staging area:
archiving, 659
definition, 15
organizing, 615–617
Data staging programmer, 70
Data staging system designer, 69
Data staging tools, 612–615
evaluation, 531–532
Data steward, 70–71
Data stores, 330, 375–378
access tools, 375
client/server ERP systems, 338–339
data marts, 376–377
data mining systems, 377
data sources, 336, 338–344
downstream systems, 378
moving into application layer, 378
standard reporting, 375–376
Data warehouse, 26–27
database administrator, 68
definition, 18
distributed versus centralized, 27
dysfunctional, 716–717
educator, 69
goals, 9–10
manager, 450
modeling, 26
quality assurance analyst, 71
Date awareness, 384
Decision making, supporting, 10
Decisions, made using data warehouse, 724–725
De-duplicating, 302–303
Degenerate dimensions, 188–189
Denormalization, 361
Deployment, 37, 691–713
definition, 691
desktop installation readiness, 692–693
end user education, 693–698
readiness assessment, 709–710
release framework, 705–710
alpha release, 706
beta release, 706–708
production release, 708–710
strategy, 710–711
DES algorithm, 482
Design issues, 526
Desktop, 429–431, 512
installation readiness, 692–693
operating system, 430
Desktop tool, 406–409
Detail diagram:
dimension table, 281–285
fact table, 280–281
Detailed models, level of detail, 325–326
Development cycle, iterative, 78
Diagnosis, multiple, 218–220
Dimensional model, 265–314
definition, 16–17
matrix method, 266–276
building matrix, 266–272
design fact tables, 272–276
Dimensional modeling, 34–35, 139–263
aggregates, 149–150
definition, 144–146
extended designs, 218–240
building audit dimension, 237–238
many-to-many dimensions, 218–222
many-to-one-to-many traps, 222–223
organization and parts hierarchies, 226–231
role-playing dimensions, 223–226
time stamping changes in a large
dimension, 233–237
too few and too many dimensions, 238–240
unpredictably deep hierarchies, 231–232
extensibility, 148–149, 160–161
importance of good browser, 170
managing, 276–312
aggregate, deciding what to, 311–312
aggregation, designing for, 310–311
data modeling tool, using, 309–312
dimension table, detail, 281–285
estimating number of rows, 309–310
fact table
detail, 280–281
diagram, 277–279
identifying sources for fact and
dimension tables, 296–307
browsing data content, 303–304
candidate data sources, 297–298
customer matching and householding,
302–303
data providers, 299
mapping data from source to target,
305–307
selection criteria, 299–302
source data ownership, 299
initial draft creation, 286–287
issues, 294–296
IS team input, 292
matrix, 277
preparing for future, 308
track base facts and derived facts,
287–292
when are you done, 307–308
work with core business users, 292–293
myths, 150–151
permissible snowflaking, 170–173
predictable, standard framework, 147–148
relationship with entity-relation modeling,
146–147
standard approaches for common modeling
situations, 149
star join schema framework, 148
strengths, 147–153
see also Dimension tables; Fact table
Dimension hierarchies, intersection of
planning and budgeting, 564–565
Dimension maintenance, slowly changing,
360
Dimension record, creating new, 627
Dimensions:
audit, building, 237–238
changes in large, time stamping, 233–237
choosing, 197–199, 274–275
granularity, 198–199
list, 270
many-to-many relationships, 218–222, 284
many-to-one-to-many traps, 222–223
role-playing, 223–226
slowly changing, 284
too few and too many, 238–240
Dimension tables, 16–17, 144–146, 165–166,
167–191
active source system participation, 371
additive, semiadditive, and nonadditive
facts, 193–194
archive and retrieval, 368–369
attached to aggregate fact tables, 557–559
backup and recovery, 367–368
planning, 369
combining from separate sources, 620–621
data push, 371–372
degenerate dimensions, 188–189
deleted items, 626
detail, 281–285
dimension change logic, 625–629
drilling down, 167–169
drilling up, 169
effective and end dates, 181
evolution, 333–334
extracts, 625–626
flexible services layers, 332–333
flow from source system to user desktop,
330–331
foreign keys, 191–193
high-level elements, 328–330
high-quality verbose attributes, 173–180
identifying sources, 296–307
indexing, 591
index management, 623
junk dimensions, 189–191
large dimensions, 183–184
load, 621–623
maintaining, 624–625
management, 437
metadata driven, 331–332
object-oriented systems, 372
populate remaining dimensions, 629–630
primary keys, 191–193
processing slowly changing, 626–629
rapidly changing dimensions, 182–188
security issues, extract and load, 369–370
services and data stores, 330
shrunken, 557–559
simultaneous hierarchies, 169
slowly changing dimensions, 180–182
smart keys, 192–193
staging, 617–630
static, 617–619
static transformation, 619–621
surrogate key, 191–193
Dimension tables (cont.)
  assignment, 619–620
  technical architecture, 327–334
  transformation and load, 629
  truncating before full refresh, 623
  see also Conformed dimensions
Direct connect, 406–407
Directory server, 479–481
Directory services, 434
Disk, 429
  access, 595–693
  space management, 660
Disk drives, 422
Disposable data marts, 377
Distributed Component Object Model, 452
Document, enhancement requests, 86
Documentation, 438
  requirements analysis process, 128–131
  support, 703–705
Double-counting, 556
Downstream systems, 378
Draft, initial, 286–287
Drilling down, 167–169
Drilling up, 169
Drives, configuration, 600–601
Economic unit, 656
Education, 693–698
  content, 693–695
  continuing, business users, 719–720
  customized, 694
  delivery, 697–698
  development, 696–697
Employee Transaction dimension, 234–237
Encryption, 481–495
  authentication tokens, 494–495
  cookies, 495
  Kerberos authentication, 491–493
  public key, 483–487
  sandbox software, 490
  secure sockets layer, 490
  signed software certificates, 489–490
  symmetric key, 481–483
  trusted certificate authorities, 487–489
  virtual private networks, 491
End user applications, 665–688
  ad hoc, 666
  definition, 20–21
  developers, 69
  development, 36, 678, 680–685
    approach selection, 680–681
    data access tool definition, 681
    document and roll out, 684
    prerun report selection, 682–683
    pretending to be professional software development manager, 684–685
    template and navigation, 681–685
    template maintenance, 685
    testing and data verification, 683–684
  education content, 695
  Lifecycle timing, 669–670
  read and write access, 669
  role, 666–670
  specification, 36
    detailed template specification, 677
    format, 677, 679
    initial template set, 670–674
    template navigation approach, 674–675
    template standards, 675–677
    user review, 677–678
  support, 702
  template, 667–668
End user data access tool, definition, 21
End user education, 693–698
  content, 693–695
  delivery, 697–698
  development, 696–697
End users:
  problems with high-end enterprise resource planning, 142
  support strategy, 699–705
  application support, 702
  communication and feedback, 702–703
  data reconciliation support, 701–702
  documentation, 703–705
  organization structure, 699–701
Enhancements, minor, prioritizing, 730–731
Enterprise Requirements Planning systems, client/server, 338–339
Entity-relation modeling, 26, 140–143
  relationship with dimensional modeling, 146–147
  use, 152–153
  weaknesses, 152
Error handling, 366
Estimating, 402
Event-tracking table, 212–214
Evolution, of warehouse, 727–733
managing using Lifecycle, 731–733
prioritizing opportunities, 729–731
Exception handling, 365–366
Executive information system, 392–393
Executive management, communication with, 90
Export, to multiple file types, 399
External consultants, role on project team, 72
External support, 519
Extract:
dimension tables, 625–626
maintenance, 722
metadata, 651
security issues, 369–370
Extracting, 22
Extract services, 357–360
Facilitation:
as alternative to interviewing, 132–133
for confirmation, prioritization, and consensus, 132
Fact definition, 165
conformed, 159–160
standardized, 156–157
Fact record:
changed status, flagging, 638
leaving flags and attributes in, 189–190
Facts:
atomic-level, 631–640
audit statistics, 631–632
historic extracts, 631–632
base, 287–292
choosing, 275–276
derived, 287–292
Fact tables, 16, 144, 165–166
accumulating snapshot, 223
aggregate, 647–649
dimension tables attached to, 557–559
computation and loading techniques, 648–649
conversion from textual to numeric or numeral category, 640
core, 203–204
data mining transformations, 637–640
data restructuring, 636–637
defining grain, 273–274
design method, 194–199
choosing data mart, 194–195
choosing dimensions, 197–199
choosing facts, 199
defining grain, 195–197
detail, 280–281
diagram, 277–279
communications purposes, 278–279
extended designs, 240–250
differing granularity and allocating, 240–250
multinational currency tracking, 248–250
multiple units of measure, 247–248
time of day, 245–246
value-band reporting, 251–252
factless, 212–216
facts at wrong grain, 243–244
families, 200–212, 559–560
aggregates, 211–212
chains and circles, 200–203
heterogeneous product schemas, 203–205
snapshot schema, 209–211
transaction schema, 206–209
four-step method to design, 272–276
identifying sources, 296–307
improving content, 636
incremental extracts, 641–643
incremental load, 643
incremental staging, 641–647
indexing, 590–591
line-of-business view, 203–205
nulls, 635–636
spanning multiple time zones, 246
speeding up load cycle, 643–647
star join, 156
surrogate key lookup, 632–635
Fault tolerance, 599–600
Feasibility, 46
Feedback, 702–703
reverse, 24
Files:
partitioned, 643–644
physical locations, 574–575
pre-sorting, 622
File transfer protocols, 433
Financial resources, 417
Financial returns, 59–60
Firewall:
  packet filter, 475–479, 497–498
  screened subnet, 476–479
Flags, 637–638
Flexibility, 424–425
Forecasting, 669
Foreign keys, 191–193
Formatting, complex, 397
Future, 25

Gateways, 432–433
Graphical flat-file transformation systems, 356
Graphs, 398
Growth, of warehouse, 727–733
  managing using Lifecycle, 731–733
  prioritizing opportunities, 729–731
  steering committee, 728–729

Hardware:
  considerations for, 417–418
  factors for platform, 424–425
  performance boosters, 422–423
  secondary storage, 423–424
  server, 511–512
  service level agreements, 423
Hardware platform:
  evaluation, 530–531
  selection, 515
Hash index, 588–589
Hierarchy:
  parts, 226–231
  unpredictably deep, 231–232
Highlighting, 395
Householding, 302–303, 656
HTTP, browser-server session, 454–456
Human resources data mart, 234–237
Hypercube, 165

Implementation:
  level of detail, 326
  metrics, 123
  Implementation team, cross-functional, 78
Incremental loads, 358–359
Index:
  analyzing after load, 592
  bitmapped, 588, 597
  B-tree, 587, 596
  dimension tables, 623
  hash, 588–589
  other types, 589
  partitioned, 643–644
  size, 598
Indexing, 23
  dimension tables, 591
  fact tables, 590–591
  for loads, 592
Index plan, initial, 585, 587–592
Individual transactions, 196
Information:
  access control, 10
  accessibility, 9
  adaptive and resilient source, 10
  assets, vulnerabilities, 459, 462–468
  consistency, 9–10
  modification, 462, 468
  needs, multiple, 391–392
Infrastructure, 412–434
  application server, 429
  connectivity and networking factors, 432–434
  desktop, 429–431
  drivers, 412–413
  evolution, 413–414
  hardware and operating system platforms, 417–418
  parallel processing architectures, 418
  plan, 511–515
  secondary storage, 423–424
  service level agreements, 423
  technical, managing, 721–722
Infrastructure architecture area, 324
Initiatives, major, prioritizing, 731
Insert statements, implementing, 731
Installation:
  readiness, desktop, 692–693
  software products, 36, 535–536
Integration, 360
  seamless, 399
Interface, custom-coded, 681
Internet security:
immediate tactical measures, 497–498
managing, 495–496
size of problem, 458
solutions, 472–495
directory server, 479–481
encryption, 481–495
routers and firewalls, 472–479
strategic measures, 498, 500
vulnerabilities, 457–472
ability to conduct business, 467, 470
information assets, 459, 462–468
network threats, 467–468, 471–472
physical assets, 459–462
software assets, 467, 469
Intersections, marking, 270–272
Interviewees:
preparing, 108–111
selection, 102–104
Interview questions, architecture-related, 506
Interviews, 97
analyst, 117–120
avoid burnout, 113
avoiding obstacles, 124–125
business executive, 116–117
deﬁne terminology, 112
establish peer basis, 113
establish time and place, 106–108
facilitation as alternative, 132–133
IS data audit, 120–122
maintain schedule ﬂexibility, 113
manage user expectations, 114
on users’ turf, 107
pre-interview
letter, 110
research, 101–102
project success criteria, 122–126
questionnaires, 104–105
reviewing results, 126–128
scheduling, 104, 106–108
sequencing, 106
starting, 114–116
verifying, what interviewee is saying, 112
write-ups, 129–130
Interview team, 98–100
IS/business partnership, 45
IS data audit interviews, 120–122
interviewee selection, 103–104
IS organization, communication with, 90
Issues, tracking, 84–85
IS team, getting input from, 292
Job:
definition, 364
scheduling, 364
Job control, approaches, 650–651
Job logs, 438
Job schedule, 652
Junk dimensions, 189–191
Kerberos authentication, 491–493
Keys, duplicate, 683
LDAP, 479–480
Lead interviewer, 99
Legacy system, 14
Letter, pre-interview, 110
Lifecycle:
end user applications, 669–670
managing growth and evolution, 731–733
project plan, 737–746
Lightweight Directory Access Protocol, 479–480
Line items, 196–197
Load:
dimension tables, 621–623
incremental, fact tables, 643
security issues, 369–370
Load cycle, speeding up, fact tables, 643
Loading, 23
aggregates, 554
Logging, 365
turning off, 622
Logical models, 326–327, 575, 577
Mainframe, 417–418
Maintenance:
application template, 685
business users, focus on, 718–720
source system-based, dimension tables, 624–625
warehouse-based, dimension tables, 624
Manageable projects, 55
Market, reducing time to, 61
Market basket analysis, 260–263
Marketing, 603
data warehouse, 725
Market research, product selection, 519
Massively parallel processing, 420–421
Matrix, 277, 346
Matrix method, 266–276
building matrix, 266–272
data marts
choosing, 273
list, 268–270
defining grain, 273–274
design fact tables, 266–276
design tips, 276
dimensions
choosing, 274–275
list, 270
facts, choosing, 275–276
intersection marking, 270–272
MAX, 565–566
Meaningful projects, 54–55
Measures, multiple units, 247–248
Memory, 422–423, 429, 595
on desktop, 431
Merchandise, improved seasonal stocking, 62
Metadata:
access, 398
active example, 439–445
content simplification, 382
database management systems, 438
data staging, 437
definition, 22, 435–436
data access tool, 681
development steps, 440–441
driving warehouse, 331–332
extraction, 651
front room, 438–439
management processes, 722–723
operations, 652
role in driving
data staging, 441–443
front-end tools, 443–444
source system, 436–437
Metadata catalog, 332, 350
maintenance, 445–447
Metadata catalog tools, 446
Metadata driven, 400
data staging, 352–355
Metaphor computer systems, 31–32
MIN, 565–566
Modeling applications, definition, 21
MOLAP:
application layer, 408–409
definition, 20
MOLAP engines, 427–428, 649
Monitoring, 364–365
activity, 380–381
Multidimensional database cube, 408
Multidimensional OLAP. See MOLAP
Multitasking, 399
Names:
coding, for people, 175–179
objects, 676
processing, 655–656
National Security Agency, encryption, 482–483
Navigation, development, 681–682
Navigation bridge table, 229–230
Network, 512
vulnerability, 467–468, 471–472
Networking, 432–434
New hire trainers, 720
Newsletter, 720
Nonadditive facts, 193–194
Non-uniform memory architecture, 421
Notification, 366
NT servers, 418
Nulls, fact tables, 635–636
Null values, 363
uniform treatment, 638
Object-oriented systems, 372
Objects, naming and placement, 676
Observers, 100
ODS. See Operational data store
OLAP, definition, 20
One-to-many relationship, validating, 621
One-to-one relationship, validating, 621
On-line analytic processing, 20
Open system servers, 418
Operating system platforms, 417–418
Operational data store, 339–344
INDEX

physical storage structure, 598–601
Pivoting, 397
Planning, 603–604
usage monitoring, 603–604
see also Data staging
Platform sharing, 429
Post-deployment assessment, 723
Predicting, 402
Pre-interview letter, 110
Presentation servers, 346–347
definition, 15–16
Primary keys, 191–193
Prime words, 571
Prioritization, 132
Private network, virtual, 491
Process management, 651
Process measurement, 651–652
Process validation, 659
Product:
selection, 36, 514–535
check references, 521
data access tool, 515, 532–535
data staging tool, 515, 531–532
DBMS platform, 515, 531
develop prototype, 524–528
evaluate options, 521–522
evaluation matrix, 516–519
evaluation shortcuts, 528, 530
hardware platform, 515, 530–531
keeping business focus, 514–515
market research, 519
narrow options, 520–521
recommending, 528
vendor, evaluation criteria, 517, 519
installation, 36
schemas, heterogeneous, 203–205
Production keys, 14
Productivity, 320
support, 351
Project, defining:
business justification, 58–64
preliminary scope, 53–57
defining, 54
project magnitude, 56
sample, 57
see also Readiness, assessing
Project identity, 64
Project management, 37–38, 77–92
  business user community, 89–90
  communications plan to manage expectations, 86, 88–92
  communication with other interested parties, 90, 92
  key roles, 93
  maintain project plan and documentation, 82–83
  project status
    meetings, 80–81
    reports, 81–82
  project team, 88
    kickoff meeting, 78–79
  scope, 83–86
    controlling changes, 84, 86–87
    document enhancement requests, 86
    tracking issues, 84
  sponsor and driver briefings, 88–89
  templates, 94
see also Dimensional modeling, managing

Project manager, 66–67

Project planning, 33–34, 64–77
  develop plan, 73–77
  growth and evolution, 732
  key roles, 93
  project identity, 64
  staffing. See Staffing templates, 94

Project plan tasks:
  architecture creation, 540
  dimensional model, 314
  business requirements, 136
  data staging, 663
  deployment, 713
  end user applications, 688
  implementing security, 500
  Lifecycle, 737–746
  maintaining and growing warehouse, 735
  physical design, 607
  selecting products, 540

Project scheduling, 300–301

Project team:
  core, 67–69
  kickoff meeting, 78–79
  management, 88
  role of external consultants, 72
  special, 70–71

Proof of concept, 53

Prototype:
  define completion criteria, 525
  developing, 522–528
  development speed, 527–528
  guidelines, 529
  never move into production, 525
  scale, 523–524

Proxy servers, 478

Public key:
  revocation, 487, 489
  secure retrieval, 487–488

Public key encryption, 483–487

Publishing, 23
  Push-button access, 392–393
  Push technology, 371–372

Quadrant analysis, requirement prioritization, 51–52

Qualifiers, 571

Quality assurance, data, 658–659

Quality assurance analyst, 71

Quality assurance checking, 23

Queries, 23
  ad hoc tools, 21, 394
  architecture, disparate conformed data sources, 253–254
  behavioral, 254–260
  cancel, 399
  categories, 255–256
  direct, 400
  drill-across, with multiple technologies, 252–254
  embedded, 399
  formulation, 394–396
  governing, 384–385
  performance and memory, 595
  reformulation, 382–383
  retargeting, 383
  self referencing, 254–260
  services, 381–385
    locations, 385–386

RAID, 599–601

Random number, 638

Readiness:
  assessing, 43–53, 709–710
  business requirements
high-level analysis, 50
prioritization, 50–53
compelling business motivation, 44–45
current analytic culture, 45
evaluate combined readiness factors, 46, 49
feasibility, 46
IS/business partnership, 45
“litmus test”, 46–48
proof of concept, 53
strong business management sponsor, 43–44
techniques to address shortfall, 49–53
desktop installation, 692–693
technical, 416
Recovering, 24
Recovery, 367–368
drill, 369
Redundant array of inexpensive disks, 599–601
References, checking vendor’s, 521
Referential integrity, 632–635
checking, 360–361
Refresh, full, 359
Relational database:
history, 141
versus multidimensional, 425
Relational database management systems:
columnar, 589
star schema, 589
Relational engines, characteristics, 426
Relational OLAP. See ROLAP
Release, 23
Remote access, 432
Renormalization, 361
Replication, 359, 642–643
Report:
output look and feel, 676–677
selection, prerun, 682–683
standard, 393–394
Report candidates, identifying, 671
Reporting, standard, 386–388
Reporting instance, 339
Report templates:
initial set, 670–674
prioritizing, 673–674
Request for Information, 520
Research, pre-interview, 101–102
Results set, calculations, 396–397
Retinal pattern biometric scanning, 495
Return on investment, 62–63
ROLAP:
application layer, 407
behavior tracking, 254–260
definition, 20
drill-across queries with multiple
technologies, 252–254
market basket analysis, 260–263
self referencing queries, 254–260
sequential subsetting, 254–260
Role, 111–112
Role-playing dimensions, 223–226
Rollback segments, 660
Routers, 474–475
Row headers:
adding, 167–168
subtracting, 169
Rows, defining, 324–326
Sandbox software, 490
Scalability, 428
Scaling, difficulty in, 493–494
Scheduling, 400
Scope:
defining, 54
management, 83–86
preliminary, 53–57
Scribe, 99–100
Scripting, 399
Secure sockets layer, 490
Securing, 24
Security, 369–370, 400
blindness to, 473–474
configuration, 477–478
data access, 380
see also Internet security
Segment, 660
Self referencing queries, 254–260
Semiadditive facts, 193–194, 210
Semiadditive summations, 395–396
Server:
architecture, 451–452
hardware, 511–512
requirements, 414–417
Service level
agreements, 423
metrics, 123
Service metrics, 723–724
Services, 330
    build versus buy, 356–357
data loading, 363–364
data staging job control, 364–366
data transformation, 360–363
    extract, 357–360
    layers, 332–333
    transaction, processing support, 370–371
see also Data staging
Signed software certificates, 489–490
Size estimates, initial, 583–586
Smart cards, 494
Smart keys, 192–193
Snapshots, 196
Snapshot schema, 209–211
Sniffing, 476–477
Snowflaking, 151, 580
    effect on
    bitmap indexes, 172
    browsing, 171–172
    permissible, 170–173
Software:
    administration, 400
    availability, 416–417
    component, 452–457
distribution, 430–431
    installation, 36, 535–536
    theft, vulnerabilities, 467, 469
Software certificates, signed, 489–490
Software development manager, thinking like, 684–685
Sorting, 397
Sort order, wild card, 553–554
Source system:
    active participation, 371
definition, 14
    metadata, 436–437
    methods for getting data from, 618
    rollback segments, 660
Sparsity failure, 553
SQL:
    direct entry, 396
    multipass, 383, 394–395
SQL92, 396
Staffing, 64–73
    business project lead, 67
    business systems analyst, 67–68
core project team, 67–69
database administrator, 68
data modeler, 68
data staging programmer, 70
data staging system designer, 69
data steward, 70–71
educator, 69
end user application developers, 69
project manager, 66–67
quality assurance analyst, 71
special teams, 70
sponsors and drivers, 65–66
technical/security architect, 70
technical support specialists, 70
Standards, 571–575
Star join, 144, 146
Star schema, optimization, 589–590
Statistical analysis, 401
Status meeting, 80–81
Status report, 81–82
Steering committee, 66, 728–729
Storage, secondary, 423–424
Store, facilitate movement to new format, 61
Stovepipe, 17
Stream extract, 618
Subsetting queries, sequential, 254–260
Success criteria, determining, 122–126
Success, monitoring, 723–724
Support team, 699–701
Surrogate key, 191–193
    for aggregates, 551
    assignment, 619–620
    lookups, fact tables, 632–635
Symmetric key encryption, 481–483
Symmetric multiprocessing, 419–420
SYN bit, sending, 471
Synonym, 574
Tables:
    analyzing after load, 592
dropping and creating, 353
    metadata, 355
    parallel, 645–647
    size, 598
    synonyms for, 574
Tablespace, 660
Tape recorders, 100