Index

A
Abbreviation, standards for, 72
Acceptance testing
  application development, 294, 296
  ETL development, 272, 277
  meta data repository and, 326–327, 332
Access interface, meta data repository, 325
Access requirements
  analyzing for prototype, 163
  database design for, 204, 206–207
Acronyms, standards for, 72
Activities
  application development, 295–297
  application prototyping, 163–164
  business case assessment, 45–48
  data analysis, 141–143
  data mining, 313–315
  database design, 204–207
  enterprise infrastructure evaluation, nontechnical, 75–76
  enterprise infrastructure evaluation, technical, 61–62
  ETL design, 231–232
  ETL development, 276–277
  implementation, 352–354
  meta data repository analysis, 186–188
  meta data repository design, 254–255
  meta data repository development, 331–332
  project planning, 90–92, 98–100
  project requirements definition, 118–121
  release evaluation, 369–370
  Activity Dependency Matrix, 405–410, xxvii
Additional reading. see Bibliography/additional reading
Administrative support, DBMS features, 60
Agenda, post-implementation review, 367
Aggregation
  database design and, 204, 206
  defining, 286
  managing growth in data volume with, 350
  OLAP tools providing, 285
  source data transformation and, 262
Alerts, 347
Amount counts, ETL reconciliation, 228, 265–266
Analysis tools. see also data mining
  application development considerations, 282
  multidimensional, 287–289
  multivariate, 289
  OLAP, 283–287, 291
  Analytical data model, 314–315
  APIs (application programming interfaces), 58, 239, 244, 292
Application design document, 297
Application developers
  application development, 298
  book chapters of interest, xxxi
  Human Resource Allocation Matrix, 383–385
  implementation, 355
  responsibilities of, 24
Application development, 281–300
  activities, 295–297
  Activity Dependency Matrix, 408
  additional reading, 299
  considerations about, 282–283
  construction stage, 13–14
  deliverables, 297
  Entry & Exit Criteria and Deliverables Matrix, 399
  environments, 292–295
  Human Resource Allocation Matrix, 383
Application development continued
multidimensional analysis factors, 287–289
OLAP architecture and, 289–292
OLAP tools and, 283–287
overview of, 281
Practical Guidelines Matrix, 481–482
risks of not performing, 299
roles, 298–299
Task/Subtask Matrix, 442–444
Application inventory, 70–71
Application lead developer
application development, 298
application prototyping, 165
book chapters of interest, xxxi
database design, 208
Human Resource Allocation Matrix, 380–385
implementation, 355
project planning, 101
project requirements definition, 121–122
release evaluation, 371
responsibilities of, 22
Application program library, 297
application programming interfaces (APIs), 58,
239, 244, 292
Application prototyping, 149–168
activities, 163–164
Activity Dependency Matrix, 407
additional reading, 167
best practices, 153–155
business analysis stage, 12
considerations about, 150–151
deliverables, 165
Entry & Exit Criteria and Deliverables
Matrix, 395
environment, 293–294
Human Resource Allocation Matrix, 381
overview of, 149
Practical Guidelines Matrix, 467–469
purposes of, 151–152
risks of not performing, 166–167
roles, 165–166
successful, 159–163
Task/Subtask Matrix, 427–430
types of, 156–159
Application release concept, 361–363
Application requirements document
contents of, 113, 114–116
revising for prototype, 165
writing, 121
Application test plans, 297
Application Track, 17–18
Applications
meta data repository design, 255
meta data repository development, 331–332
security measures for, 340–341
Arbitration boards, 26
Architecture
enterprise, 68–71
OLAP, 289–292
security and, 342
three-tier computing, 56
Associations discovery technique, 306, 307–308
Assumptions, project planning and, 87–88
Authentication, defined, 345
Automation silos (stovepipe systems)
data discrepancies of, 130
effects of, 65–66
Availability, application requirements
document, 116
B
Back end, 17
Backups
overview of, 345–347
physical database design, 203
Bandwidth, 348
BI infrastructure architect
book chapters of interest, xxxi
Human Resource Allocation Matrix, 379, 382, 385
meta data repository design, 256
nontechnical enterprise infrastructure
evaluation, 77
release evaluation, 371
responsibilities of, 22
technical enterprise infrastructure
evaluation, 63
BI target databases. see also Database design
building, 206–207
designing, 205–206
designing load programs, 223–224
ETL implementation strategies, 213–215
operational databases vs., 194
overview of, 196
physical design of, 207–208
as sources for data mining, 306–307
Bibliography/additional reading
application development, 299
application prototyping, 167
Index

business case assessments, 50
data analysis, 146–147
data mining, 317
database design, 209–210
development steps, 27
enterprise infrastructure evaluation, 78–80
ETL design, 234–235
ETL development, 279
implementation, 356–357
meta data repository analysis, 190
meta data repository design, 257–258
meta data repository development, 335
project planning, 103
project requirements definition, 123–124
release evaluation, 375
Bottom-up source data analysis, 133–136
Brainstorming, 267–268
Budget, 84–89, 114, 156, 364, 460, 468
Business
application prototyping and, 150–151, 155
architecture models, 69–70
Business analysis stage
defined, 6
development steps in, 12
issues, 35–37
nontechnical infrastructure activities, 67
parallel development tracks, 19
stovepipe systems and, 65–66
Business case assessment, 29–50
activities, 45–48
Activity Dependency Matrix, 405
additional reading, 50
business analysis issues, 35–37
business drivers, 33–35
business justification, 31–33
considerations about, 30–31
cost-benefit analysis, 37–40
deliverable, 48
Entry & Exit Criteria and Deliverables Matrix, 387
Human Resource Allocation Matrix, 379
justification stage, 11
overview of, 29
Practical Guidelines Matrix, 455–456
risk assessment, 40–45
risks of not performing, 49–50
roles, 49
Task/Subtask Matrix, 411–413
Business-focused data analysis
data domain rules, 134–135
data integrity rules, 135–136
meta data, 131–133, 143
overview of, 127
Business representative
application development, 298
application prototyping, 166
application release concept, 361
application release concept guidelines, 362–364
book chapters of interest, xxix
business case assessment, 49
data analysis, 144
data mining, 316
ETL development, 278
ETL integration testing, 269
guidelines for project planning, 460
Human Resource Allocation Matrix, 379–381, 383, 385
justifying BI decision, 32–33
logical data modeling, 131
meta data repository development, 333
nontechnical infrastructure needs, 67
project planning, 82, 101
project requirements definition, 114, 122
release evaluation, 371
responsibilities of, 22
Business sponsor
book chapters of interest, xxix
business case assessment, 49
Human Resource Allocation Matrix, 379, 385
post-implementation review, 368
project-specific requirements of, 114
release evaluation, 372
responsibilities of, 24

C
Caching, DBMS, 60
Calling card fraud, 312
Campaigns, marketing, 312
CASE tools
defined, 241
nontechnical infrastructure and, 64, 75, 459
populating meta data repository using, 322
Centralized meta data repository, 242–244
Change-control procedures
application release guidelines, 363
post-implementation review, 366
practical guidelines for, 460
project planning, 88–89
Chaos, minimizing hardware, 54–55
Cipher text, 345
Classification, data mining, 309, 310–311
Classifications, meta data, 176–182
Cleansing data. see Data cleansing
Clustering
data mining technique, 309
physical database design decisions, 202
Competition, business case assessment, 47
Compilation, unit testing, 269
Complexity risk, 41
Computer utilization, 347
Connectivity paths, 343
Constraints, project
overview of, 86–88
planning considerations, 84, 418
Construction stage
defined, 6
development steps in, 13–14
parallel development tracks and, 19
Contingency plan, 86
Core team members
nontechnical enterprise infrastructure
evaluation, 65
post-implementation review, 367
release evaluation, 489
responsibilities of, 21–23
Cost-benefit analysis
business issues, 37–40
determining ROI, 47
ETL tools and, 230
project planning and, 82
Costs
considerations about, 30
constraints, 84
determining or revising, 99
justifying, 31–33
post-implementation review of budget, 364
Practical Guidelines Matrix, 455–456
prototyping and, 151
risk and, 42
standardizing of data and, 37
CPM (critical path method), 95–96, 101
Credit card fraud, 312
Critical path method (CPM), 95–96, 101
Critical success factors, 99
CRM (customer relationship management), 9, 39
Cross-organizational integration
immature attempts to organize meta data, 240
need for nontechnical infrastructure, 66–68
overview of, 8–11
project-specific vs., 16
CRUD (create, read, update, delete) access, 207, 331
customer relationship management (CRM), 9, 39
Customers
market management of, 312
planning incremental rollouts, 339
Customization
meta data repository design, 242–243
meta data repository development, 320
D
Data
accuracy, 140
application prototyping and, 153
application requirements document and, 115
business analysis issues, 35–37
conversion rules, 134
domain rules, 134–135
formats, 140
integrity rules, 135
managing growth in, 349–350
names, 132
naming standards, 72
policy, 133
precision, 140
project requirements, 106
reliability, 140
resolving discrepancies, 143
sharing, 60
transformation activities, 261–263
values, 221–222
Data administrator
book chapters of interest, xxxi
data analysis, 144
database design, 208
Human Resource Allocation Matrix, 379–382, 385
logical data modeling, 131
meta data repository analysis, 188
meta data repository design, 256
nontechnical enterprise infrastructure
evaluation, 77
project planning, 101
project requirements definition, 122
release evaluation, 372
responsibilities of, 22
Data analysis, 125–148
activities, 141–143
Activity Dependency Matrix, 406
additional reading, 146–147
bottom-up source data analysis, 133–136
business analysis stage, 12
business-focused vs. traditional, 127
considerations about, 126–127
data cleansing, 136–141
deliverables, 143–144
Entry & Exit Criteria and Deliverables Matrix, 392
Human Resource Allocation Matrix, 381
overview of, 125
Practical Guidelines Matrix, 464–466
risks of not performing, 145–146
roles, 144–145
Task/Subtask Matrix, 425–426
top-down logical data modeling, 128–133
Data cleansing, 136–141
application requirements document and, 115
data analysis and, 126
data mining and, 313
estimates for, 461
ETL design and, 215–216
ETL development and, 262
overview of, 136–141
tools for, 241
writing specifications for, 143
Data content, 133. see also Domain
Data control language. see DCL (data control language)
Data definition, 132
Data definition language. see DDL (data definition language)
Data dictionaries, 61, 239–240
Data identifier, 132–133
Data-in philosophy, 195
Data integration
ETL transformation programs, 221
logical data modeling for, 127
normalization rules, 262
prototyping best practices, 155
Data integrity
business rules for, 135–136
source data selection and, 139
Data length
defined, 133
technical data conversion rules for, 134
Data management middleware, 57
Data marts
approach, 351
implementation strategies, 213
staging area, 232, 475
Data mining, 301–318
activities, 313–315
Activity Dependency Matrix, 409
additional reading, 317
applications of, 311–313
considerations about, 302–303
construction stage, 13–14
defining, 303–307
deliverables, 315
Entry & Exit Criteria and Deliverables Matrix, 400
Human Resource Allocation Matrix, 383
operations, 310–311
overview of, 301
Practical Guidelines Matrix, 483–484
risks of not performing, 316
roles, 316
Task/Subtask Matrix, 445–446
techniques, 307–310
tools, 241, 302
Data mining expert
book chapters of interest, xxxi
data mining, 316
Human Resource Allocation Matrix, 383–385
implementation, 355
release evaluation, 372
responsibilities of, 22
Data mining operations, 310–311
Data mining techniques, 307–310
Data-out philosophy, 195, 196
Data owners
data analysis, 145
data quality, 137
logical data modeling, 131
Practical Guidelines Matrix, 464
responsibilities of, 25
Data quality
data analysis and, 126, 137, 142
general business requirements for, 110–111
normalization, 195
planning project for, 461
standards, 73
Data quality analyst
book chapters of interest, xxxi
Data quality analyst continued
business case assessments, 49
data analysis, 144
ETL design, 233
Human Resource Allocation Matrix, 379–382, 385
nontechnical enterprise infrastructure evaluation, 77
project planning, 102
project requirements definition, 122
release evaluation, 372
responsibilities of, 22
Data relationship, 132
Data rule, 133
Data types
defined, 133
technical data conversion rules for, 134
Database administrator
application development, 298
application prototyping, 153, 166
book chapters of interest, xxxi
data mining, 316
database design, 193, 208–209
ETL design, 233
ETL development, 278
Human Resource Allocation Matrix, 379–385 implementation, 355
meta data repository development, 334
personnel utilization and, 348–349
project planning, 102
release evaluation, 372
responsibilities of, 22
technical enterprise infrastructure evaluation, 63
Database design, 191–210
activities, 204–207
Activity Dependency Matrix, 407
additional reading, 209–210
considerations about, 192–193
deliverables, 207–208
design stage, 13
Entry & Exit Criteria and Deliverables Matrix, 395
Human Resource Allocation Matrix, 382
logical, 197–200
overview of, 191
philosophies, 193–196
physical, 201–204
Practical Guidelines Matrix, 472–474
risks of not performing, 209
roles, 208–209
Task/Subtask Matrix, 433–435
Database management system. see DBMS (database management system)
Database segmentation, data mining, 311
Databases
backup and recovery, 345–347
data mining, 311, 315
drivers, 58
loading production, 353
maintaining, 207–208, 338
meta data repository, 331, 332–333
OLAP architecture support of, 292
Dataset placement, 201–202
DBMS (database management system)
considerations about, 53–54
criteria for selecting, 59–61
design considerations, 192
dictionaries, 322–323
gateways, 57–59
meta data from tools, 241
meta data in, 240
Practical Guidelines Matrix, 458
query resolution, 199
scalability, 55
security, 343
technical enterprise infrastructure evaluation activities, 61–62
DCL (data control language)
database security and, 206
defined, 208
meta data repository design and, 256
DDL (data definition language)
building BI target databases and, 206
defined, 208
meta data repository design and, 255
Deadline management, 155
Decentralized meta data repository, 244–245
Decision-support initiatives
assessing business drivers, 33–35
dynamics of, 10
justifying, 31–33
overview of, 5
Decision-support system (DSS) solutions, 46–47
Deleted records, processing, 218–219
 Deliverables. see also Entry & Exit Criteria and Deliverables Matrix
application development, 297
application prototyping, 150, 155, 165
business case assessment, 48
data analysis, 143–144
data mining, 315
database design, 207–208
enterprise infrastructure evaluation, nontechnical, 76–77
enterprise infrastructure evaluation, technical, 62–63
ETL design, 233
ETL development, 277
implementation, 354
meta data repository analysis, 188
meta data repository design, 255–256
meta data repository development, 332–333
project planning, 100–101
project requirements definition, 121
release evaluation, 371
Demo prototype, 158
Demonstrations, prototype, 164
Deployment stage
defined, 6
development steps in, 14
parallel development tracks and, 19
Derivation process, ETL development, 262
Design stage
defined, 6
development steps in, 13
parallel development tracks and, 19
Developers, 373, 385
Development approaches
cross-organizational, 8–11
overview of, 5–6
post-implementation review on, 366
traditional, 6–8
Development environment, 293–294
Development steps
Activity Dependency Matrix, 405–410
additional reading, 27
arbitration boards, 26
cross-organizational integration, 8–11
decision-support initiatives, 5
defining, 4–5
development approaches, 5–6
engineering stages and, 11–16
Entry & Exit Criteria and Deliverables Matrix, 387–403
establishing standards, 72
Human Resource Allocation Matrix, 379–385
justification for using, 26–27
parallel development tracks, 17–20
Practical Guidelines Matrix, 455–490
project team structures, 20–25
Task/Subtask Matrix, 411–454
traditional development approach, 6–8
Deviation detection, data mining, 311
Dictionaries, data
DBMS, 61, 322–323
overview of, 239–240
Dimension tables, star schema, 198–199
Directory, meta data repository, 330
Display, OLAP tools, 287
Distributed logic middleware, defined, 57
Distributed XML-enabled meta data solution, 245–247
Distribution, data mining, 313
Documentation
meta data repository, 333
production, 354
Domain. see also Data content
business meta data and, 143, 177
enforcements, 228
violations, 135, 479
Domain counts, ETL reconciliation, 228, 265–266
Double maintenance, defined, 246
Drivers, database, 38
Drivers, of business, 30, 32, 33–35

E
E-R (entity-relationship) meta data repository, 247–248, 254–255
E-R (entity-relationship) modeling. see Logical data models
Ease of use, defining, 469
Edits, ETL unit testing, 269
EIS (executive information system), 283, 305–306
Encryption, 344, 345
Engineering stages, 11–16
Business analysis stage, 12
Construction stage, 13–14
dependencies, 15
Deployment stage, 14
Design stage, 13
development approach, 5–6
Justification stage, 11
parallel development tracks and, 17–20
Planning stage, 11
project-specific vs. cross-organizational steps, 16
Enterprise architecture, 68–71

Enterprise infrastructure evaluation, non-technical activities, 75–76
Activity Dependency Matrix, 405
additional reading, 79–80
architecture components, 68–71
considerations about, 64–65
deliverable, 76–77
Entry & Exit Criteria and Deliverable Matrix, 389
Human Resource Allocation Matrix, 379
need for, 66–68
overview of, 52
planning stage, 11
Practical Guidelines Matrix, 459
risks of not performing, 78
roles, 77–78
standards for, 71–75
stovepipe development effects and, 65–66
Task/Subtask Matrix, 416–417
Enterprise infrastructure evaluation, technical activities, 61–62
Activity Dependency Matrix, 405
additional reading, 78–79
considerations about, 53–54
DBMS requirements, 58–61
deliverables, 62–63
Entry & Exit Criteria and Deliverables Matrix, 388
hardware platform scalability, 54–57
Human Resource Allocation Matrix, 379
middleware platform, 57–58
overview of, 51
planning stage, 11
Practical Guidelines Matrix, 457–458
risks of not performing, 63
roles, 63
Task/Subtask Matrix, 414–415
Enterprise logical data model, 129–130, 142–144
Entity-relationship (E-R) meta data repository, 247–248, 254–255
Entity-relationship (E-R) modeling. see Logical data models
Entry & Exit Criteria and Deliverables Matrix, 387–403
application development, 399
application prototyping, 393
business case assessments, 387
data analysis, 392
data mining, 400
database design, 395
defined, xxvii
tecture infrastructure evaluation, 388–389
ETL design, 396
ETL development, 398
implementation, 402
meta data repository analysis, 394
meta data repository design, 397
meta data repository development, 401
project planning, 390
project requirements definition, 391
release evaluation, 403
Environments
application development, 292–295
multi-tier, 341–344
production, 293–294, 353
setting up implementation, 353
Error-handling, ETL process flow design, 228
Estimating techniques, 92–93
ETL developers
book chapters of interest, xxxi
ETL development, 278
Human Resource Allocation Matrix, 383–385
implementation, 355
release evaluation, 373
responsibilities of, 24
ETL (extract/transform/load) design, 211–236
activities, 231–232
Activity Dependency Matrix, 408
additional reading, 234–235
considerations about, 212–213
deliverables, 233
design stage, 13
Entry & Exit Criteria and Deliverables Matrix, 396
evaluating tools for, 229–231
extract programs, 219–220
Human Resource Allocation Matrix, 382
implementation strategies, 213–215
load programs, 223–224
overview of, 211
Practical Guidelines Matrix, 475–476
preparing for process, 215–219
process flow, 225–229
risks of not performing, 234
roles, 233–234
## Index

Task/Subtask Matrix, 436–437  
transformation programs, 221–223  
ETL (extract/transform/load) development, 259–280  
activities, 276–277  
Activity Dependency Matrix, 408  
considerations about, 260–261  
construction stage, 13–14  
deliverables, 277  
Entry & Exit Criteria and Deliverables Matrix, 398  
Human Resource Allocation Matrix, 383  
overview of, 259  
peer reviews, 267  
Practical Guidelines Matrix, 479–480  
reconciliation, 263–267  
risks of not performing, 279  
roles, 278  
source data transformation, 261–263  
Task/Subtask Matrix, 440–441  
Testing, 268–273  
ETL lead developer  
book chapters of interest, xxxi  
data analysis, 144  
database design, 209  
ETL design, 225–228, 234  
ETL development, 278  
Human Resource Allocation Matrix, 380–385  
implementation, 355–356  
project planning, 102  
release evaluation, 373  
risk of not performing, 279  
resolutions of differing, 222  
ETL process flow, 225–229  
ETL program library, 277  
ETL test plan, 273–275, 277  
ETL Track, 17  
Evaluation  
nontechnical enterprise infrastructure. see Enterprise infrastructure evaluation, nontechnical  
technical enterprise infrastructure. see Enterprise infrastructure evaluation, technical  
Executive information system (EIS), 283, 305–306  
Exit criteria. see Entry & Exit Criteria and Deliverables Matrix  
Extended team members  
post-implementation review, 367  
release evaluation, 489  
risk assessment, 23–25  
External data  
analyzing sources of, 141–142  
resolutions of differing, 222  
Extract programs, 219–220, 225  
Extract/transform/load (ETL) design. see ETL (extract/transform/load) design  
F  
Facilitator  
Human Resource Allocation Matrix, 385  
post-implementation review, 368–369  
release evaluation, 373  
risk assessment, 25  
Fact tables, star schema, 197–198  
Financial investment risk, 42  
Financial services, data mining, 312–313  
Finish to Finish, task dependencies, 94–95  
Finish to Start, task dependencies, 94–95  
Forecasting technique, data mining, 309–310  
Formats  
preparing for ETL process, 215  
resolution of differing, 222  
Formula-based estimations, 92–93  
Fraud detection, data mining, 312  
Front end, 17–20  
Functional project requirements, 106, 115  
Functionality, application prototyping, 467  
G  
Gantt charts, 100, 101  
Gap analysis security matrix, 343–344  
Gateways, 57–59, 245–247, 485  
General business requirements, 108–112  
data quality requirements, 110–111  
interviewees for, 109–110  
overview of, 108–109  
project-specific requirements vs., 107  
report, 111–112  
Goals  
data mining, 313  
project, 85  
Group IDs (group identifiers), 207  
Growth management, 349–351  
GUI (graphical user interface) components, prototypes, 161
Guidelines. see also Practical Guidelines Matrix
  nontechnical enterprise infrastructure evaluation, 64–65
  physical database design, 201–204
  prototyping, 161–162
  using application release concept, 362–363

H
  Hardware
  data and usage growth and, 351
  platform requirements, 54–57
  Practical Guidelines Matrix, 458
  technical enterprise infrastructure evaluation, 53, 61–62

  Help desk (support)
  implementation and, 338–339, 353
  role of extended team members, 24–25

  Help function. see Online help function

  Historical load process, ETL, 217

History
  basing project estimates on, 92–93
  data requirements and, 106
  Homonyms, 222
  Human Resource Allocation Matrix, 379–385, xxvii

I
  Implementation, 337–358
  activities, 352–354
  Activity Dependency Matrix, 409
  additional reading, 356–357
  considerations about, 338–339
  data backup and recovery, 345–347
  deliverables, 354
  deployment stage, 14
  Entry & Exit Criteria and Deliverables
    Matrix, 402
  growth management, 349–351
  Human Resource Allocation Matrix, 384
  incremental rollout, 339
  monitoring utilization of resources, 347–349
  overview of, 337
  Practical Guidelines Matrix, 487–488
  risks of not performing, 356
  roles, 355–356
  security management, 340–345
  Task/Subtask Matrix, 450–452
  Incremental load process, ETL, 217–219

Indexing
  approaches to, 473–474
  managing growth in data through, 349
  physical database design and, 202–203
  poor database performance and, 224

Information
  accessibility of, 30
  needs, 35
  Information architecture, 67–68, 70, 73
  Infrastructure architect role. see BI infrastructure architect
  Initial load programs, ETL, 216–217
  Instruction guides, 329, 353
  Insurance fraud, data mining, 312
  Integration risk, 41
  Integration testing
    application development, 296
    ETL development, 270, 276
    meta data repository development, 326–327, 332
  Interface
    meta data repository design, 238
    meta data repository development, 324–325
  Internet access, security, 344–345
  Interviewing
    general business requirements, 109–110
    overview of process, 116–118
    project-specific requirements, 113–114
  Intuitive estimating technique, 92–93
  Inventory
    application, 70–71
    controlling through data mining, 313
  Issues log
    application prototyping using, 165
    guidelines for project planning, 460
    management procedures using, 90–91
  IT (information technology) managers
    application release concept, 361, 362–364
    justifying BI decision, 32–33
    logical data modeling, 131
    responsibilities of, 24–25
  Iterative releases. see also Application prototyping; Release concept; Release evaluation
    building meta data repository projects in, 242
    comparing BI applications with traditional, 7–8
    cross-organizational integration and, 9–11
    guidelines for, 467–469
    implementation planning for, 339
Index

J
Justification stage
assessing business case, 31–33, 45–46
assessing business drivers, 33–35
defined, 6
development step in, 11
parallel development tracks and, 19

K
Knowledge discovery, 305, 400, 483

L
Libraries
application program, 354
ETL program, 277, 354
meta data repository program, 328, 333, 354
Licensing (buying), meta data repository, 242–245, 250–254
Link analysis, data mining, 311
Load process, 60
Load programs, ETL design, 223–224
Load sequence, ETL design, 228
Log data, Web, 294–295
Logical data identifiers, 132–133
Logical data models
business meta data components of, 131–133
enterprise, 129–130
nontechnical enterprise infrastructure evaluation and, 64
overview of, 128
participants in, 131
project-specific, 128–129
refining, 142
requirements for expanding, 120
standards for, 73
Logical database design
overview of, 207
snowflake schema, 200
star schema, 197–200
Logical meta model, 184–186
Logon, 341

M
Management
of data growth, 349–350
data mining applications, 312–313
Manuals, meta data repository, 329
Market basket analysis
defined, 308
using data mining tools for, 306–307
Market management, data mining, 312
Marketing
data mining considerations, 302
multidimensional analysis factors, 288
MDC (Meta Data Coalition), 325
Merging utility, 225, 228
Meta data
capture, 73
immature attempts to organize, 239–240
importance of, 172–173
migration process, 331
populating meta data repository, 321–324
sources of, 240–242
Meta data administrator
book chapters of interest, xxxi
data analysis, 145
Human Resource Allocation Matrix, 379–385
implementation, 356
meta data repository analysis, 189
meta data repository design, 256
meta data repository development, 323, 334
nontechnical enterprise infrastructure evaluation, 78
project planning, 102
project requirements definition, 122
release evaluation, 373
responsibilities of, 22
Meta data classifications, 176–182
Meta Data Coalition (MDC), 325
Meta data repository
implementation of, 354
nontechnical infrastructure considerations, 64, 67
overview of, 71
standards for, 73
Meta data repository analysis, 169–190
activities, 186–188
Activity Dependency Matrix, 407
additional reading, 190
business analysis stage, 12
challenges of implementing, 182–184
considerations about, 170–171
deliverables, 188
Entry & Exit Criteria and Deliverables Matrix, 394
Human Resource Allocation Matrix, 381
importance of meta data, 172–173
logical meta model, 184–186
meta data classifications, 176–182
as navigation tool, 174–176
Index

Meta data repository analysis continued
overview of, 169
Practical Guidelines Matrix, 470–471
risks of not performing, 189
roles, 188–189
Task/Subtask Matrix, 431–432
Meta data repository design, 237–258
activities, 254–255
Activity Dependency Matrix, 408
additional reading, 257–258
centralized, 242–244
considerations about, 238–239
decentralized, 244–245
deliverables, 255–256
design stage, 13
distributed XML-enabled, 245–247
entity-relationship, 247–248, 254–255
Entry & Exit Criteria and Deliverables Matrix, 397
Human Resource Allocation Matrix, 382
licensing, 250–254
meta data silos, 239–242
object-oriented, 248–250, 254–255
overview of, 237
Practical Guidelines Matrix, 477–478
risks of not performing, 257
roles, 256
Task/Subtask Matrix, 438–439
Meta data repository developers book chapters of interest, xxxi
Human Resource Allocation Matrix, 383–385
implementation, 356
meta data repository development, 334
release evaluation, 373
responsibilities of, 24
Meta data repository development, 319–336
activities, 331–332
Activity Dependency Matrix, 409
additional reading, 335
considerations about, 320–321
construction stage, 13–14
deliverables, 332–333
directory, 330
Entry & Exit Criterias and Deliverables Matrix, 401
Human Resource Allocation Matrix, 383
interface processes, 324–326
overview of, 319
populating, 321–324
Practical Guidelines Matrix, 485–486
preparing for rollout, 327–330
risks of not performing, 334
roles, 333–334
Task/Subtask Matrix, 447–449
testing, 326–327
Meta data repository program library, 328, 333, 354
Meta data repository test plan, 333
Meta Data Repository Track, 18
Meta-meta data, 186
Middleware. see also Gateways
platform, 57–58
technical enterprise infrastructure evaluation of, 53, 61–62
Migration process, meta data, 331
Mitigation plan, 86
Mnemonics, 223
Mock-up prototype, 156–157
MOLAP (multidimensional OLAP) tools, 292
Monitoring
database performance, 207
resource utilization, 347–348
Multi-tier environment, security, 341–344
Multidimensional database design
OLAP tools, 284–289, 292
overview of, 195–196
Multidimensional OLAP (MOLAP) tools, 292
Multivariate analysis, 288
N
Navigation tool, meta data, 174–176
Negotiation skills, 365
Networks
DBMS selection criteria for, 60
evaluating technical infrastructure, 53
monitoring utilization of, 347–348
services staff, 24–25
technical enterprise infrastructure evaluation activities, 62
Nontechnical enterprise infrastructure evaluation.
see Enterprise infrastructure evaluation, nontechnical
Normalization rules
E-R modeling based on, 128
ETL development, 262
operational databases and, 194–195
Note-taking, interviewing process, 118
Index

O
Object Management Group (OMG), 325
Object-oriented (OO) meta data repository, 248–250, 254–255
Objectives
project, 47, 85
prototype, 160, 467
ODBC (Open Database Connectivity), 58
OLAP (online analytical processing) tools
advantages of, 284–285
architecture, 289–292
data mining vs., 305
features of, 285–287
meta data, 241
multidimensional analysis factors, 287–289
overview of, 283
populating meta data repository and, 323
post-implementation review, 365
Practical Guidelines Matrix, 457
OMG (Object Management Group), 325
Online analytical processing. see OLAP (online analytical processing) tools
Online help function
application development, 296–297
meta data repository, 255–256, 327–328, 332, 477
OO (object-oriented) meta data repository, 248–250, 254–255
Open Database Connectivity (ODBC), 58
Operational data
assessing business cases, 47
business analysis issues, 36–37
Operational databases
embedded process logic and, 222
overview of, 193
Operational prototype, 159
Operational systems
cleansing, 141
data mining, 306–307
Operations staff role, 24
Organization risk, 42
Peer reviews, 67, 267
Performance
application requirements document content on, 115
database design for, 192
E-R meta data design for, 248
ETL design for, 212
ETL development testing for, 271–272, 276–277
indexing schemes and, 224
load, 225
monitoring database, 207
project requirements and, 107
selecting DBMS based on, 60
Performance testing
application development, 296
ETL development, 271–272, 276
with prototypes for, 152
Personnel utilization, 348
Pert charts, 100, 101
Physical database design, 201–204, 206
Placement, dataset, 201–202
Plain text, defined, 345
Planning stage
defined, 6
development steps in, 11
parallel development tracks and, 19
Point-to-point gateways, defined, 58
Policies, standards, 75
Portfolio management, 313
Post-implementation reviews, 364–369
activities, 369–370
organizing, 366–368
release evaluation considerations, 360
session flow, 368–369
topics for, 364–366
"Power users", 18, 110, 114, 173, 325
Practical Guidelines Matrix, 455–490
application development, 481–482
application prototyping, 467–469
business case assessment, 455–456
data analysis, 464–466
data mining, 483–484
database design, 472–474
defined, xxvii
enterprise infrastructure evaluation, nontechnical, 459
time enterprise infrastructure evaluation, technical, 457–458

P
Parallel development tracks, 17–20
Parallel query execution, 204
Partitioning, database design, 202
Passwords
gap analysis matrix and, 343
security management and, 341
Practical Guidelines Matrix continued
ETL design, 475–476
ETL development, 479–480
implementation, 487–488
meta data repository analysis, 470–471
meta data repository design, 477–478
meta data repository development, 485–486
project planning, 460–461
project requirements definition, 462–463
release evaluation, 489–490
Predictive modeling, data mining, 310–311
Presentation services, OLAP, 290–291
Primary keys
logical data identifiers vs., 133
resolution of duplicate, 221
Private data, 36–37
Procedures, 64–65, 75
Process flow, ETL
contents of, 233
designing, 212, 225–229, 232
integration testing and, 269
preparing for, 215–219
Process independence, 128–129
Product evaluation, 251–252
Production
environment, 293–294, 353
meta data repository development, 327–330, 332
planning implementation, 338
Profit, cost-benefit analysis, 39
Programming specifications
ETL design, 233
meta data repository design, 256
Programs
application, 296
ETL, 232–233, 277
meta data repository, 332, 333
Project charter, 99, 100–101
Project core team members, 20–23
Project manager
book chapters of interest, xxx
business case assessment, 49
Human Resource Allocation Matrix, 379–380, 385
peer reviews, 267
post-implementation review, 368
project planning, 82, 83–84, 102
release evaluation, 373
responsibilities of, 22
Project planning, 81–103
activities, 98–100
Activity Dependency Matrix, 406
additional reading, 103
considerations about, 82–83
critical path method, 95–96
defining the project, 84–90
deliverables, 100–101
Entry & Exit Criteria and Deliverables Matrix, 390
estimating techniques, 92–93
Human Resource Allocation Matrix, 380
managing the project, 83–84
overview of, 81
planning stage, 11
planning the project, 90–98
Practical Guidelines Matrix, 460–461
resource assignments and, 93–94
resource dependencies and, 95
risks of not performing, 103
roles, 101–102
scheduling, 96–98
task dependencies, 94–95
Task/Subtask Matrix, 418–421
Project requirements definition, 105–124
activities, 118–121
Activity Dependency Matrix, 406
additional reading, 123–124
considerations about, 106–107
deliverable, 121
Entry & Exit Criteria and Deliverables Matrix, 391
general business, 108–112
Human Resource Allocation Matrix, 380
interviewing process and, 116–118
overview of, 105
Practical Guidelines Matrix, 462–463
project-specific, 112–116
risks of not performing, 122–123
roles, 121–122
Task/Subtask Matrix, 422–424
Project-specific requirements
application requirements document and, 114–116
cross-organizational developmental steps vs., 16
general business requirements vs., 107
interviewees for, 113–114
logical data models and, 128–129
overview of, 112–113
Index

Project team risk, 42
Promotions, 312
Proof-of-concept prototype, 157
Prototype charter
  contents of, 165
  overview of, 160–161
  preparing, 164
Prototyping, see Application prototyping

QA (Quality assurance)
  analyst, 24
  environment, 293–294
QA testing
  application development, 296
  ETL development, 272, 277
  meta data repository development, 327
Quality, project constraints and, 87
Queries
  bandwidth and, 348
  database design for, 192
  DBMS selection criteria for, 60
  designing prototype for, 164
  monitoring database performance of, 207
  OLAP and, 286, 291
  parallel execution design for, 204
Questionnaires, interview, 117

Readiness assessment, interview, 117
Reconciliation, ETL, 263–267
  activities of, 228
  calculating totals, 264–266
  designing, 213
  overview of, 263–264
  preparing for, 215
  standards, 74
  storing statistics, 266–267
Record counts, ETL reconciliation, 228, 264–265
Records, processing deleted, 218–219
Recovery
  overview of, 345–347
  physical database design, 203–204
Reference guides. see Instruction guides
Referential integrity (RI), 224, 473, 476
Regression analysis, data mining, 309
Regression testing
  application development, 296
  ETL development, 271–272, 276
  meta data repository development, 326–327, 332
Relational OLAP (ROLAP) tools, 292
Release concept, 5, 14, 361–364
Release evaluation, 359–376
  activities, 369–370
  Activity Dependency Matrix, 410
  additional reading, 375
  application release concept, 361–364
  considerations about, 360
  deliverables, 371
  deployment stage, 14
  Entry & Exit Criteria and Deliverables Matrix, 403
  Human Resource Allocation Matrix, 385
  overview of, 359
  post-implementation reviews, 364–369
  Practical Guidelines Matrix, 489–490
  risks of not performing, 374
  roles, 371–374
  Task/Subtask Matrix, 453–454
Reorganization, physical database design, 203
Reports
  business case assessment, 48
  database design, 192
  defining requirements for, 120
  early meta data, 239
  general business requirements, 111–112
  nontechnical infrastructure assessment, 76
  OLAP, 291
  operational system, 195
  prototype design, 164
  technical infrastructure assessment, 62
  Request for information (RFI), 63
  Request for proposal (RFP), 63
  Research, interviewing process, 117
  Resource dependencies, 95
  Resources. see Bibliography/additional reading
Return on investment. see ROI (return on investment)
Revenue, 39
Reviews, post-implementation, 364–369
RFI (request for information), 63
RFP (request for proposal), 63
RI (referential integrity), 224, 473, 476
Risk management, data mining, 312
Risk, of omitting
application development, 299
application prototyping, 166–167
business case assessment, 49–50
data analysis, 145–146
data mining, 316
database design, 209
ETL design, 234
ETL development, 279
implementation, 356
meta data repository analysis, 189
meta data repository design, 257
meta data repository development, 334
nontechnical enterprise infrastructure
evaluation, 78
project planning, 103
project requirements definition, 122–123
release evaluation, 374
technical enterprise infrastructure evaluation, 63
Risks
assessment, 40–45
business case assessment and, 456
common project, 85–86
considerations about, 30
data mining and, 312
justifying, 33
project planning and, 99, 460
security standards minimizing, 74
ROI (return on investment)
activities for determining, 45–47
assessing risk on, 42
considerations about, 30
cost-benefit analysis of, 37–40
hardware platforms and, 55
post-implementation review of, 365
ROLAP (relational OLAP) tools, 292
Roles
application development, 298
application prototyping, 165–166
BI arbitration board, 26
business case assessment, 49
core team members, 21–23
data analysis, 144–145
data mining, 316
database design, 208–209
ETL design, 233–234
ETL development, 278
extended team members, 23–25
Human Resource Allocation Matrix, 379–385
implementation, 355–356
limited, 25
meta data repository analysis, 188–189
meta data repository design, 256
meta data repository development, 333–334
nontechnical enterprise infrastructure
evaluation, 77–78
project planning, 101–102
project requirements definition, 121–122
release evaluation, 371–374
technical enterprise infrastructure evaluation, 63
S
Savings, cost-benefit analysis, 39
Scalability
DBMS selection criteria for, 60
hardware platforms and, 54–57
Practical Guidelines Matrix, 457
Schedules
deadline management, 155
interview, 117
meta data repository rollout, 329
post-implementation review, 364, 367
project, 96–97
prototype activities, 152–153
setting up production, 353
Scope
application development and, 282
defining requirements for, 120
post-implementation review, 365
Scope, application prototyping
building slimware to limit, 155
considerations about, 150
determining, 163–164
limiting, 153
system analysis and, 159–160
Scope, project planning
change-control and, 88–89
considerations about, 82
determining requirements for, 98–99
overview of, 85
Scribe
Human Resource Allocation Matrix, 385
post-implementation review, 369
release evaluation, 373
responsibilities of, 25
Security, 340–345
application requirements document and, 115–116
database, 206
Index

DBMS selection criteria for, 61
implementation and, 338
Internet access, 344–345
measures for, 340–341
meta data repository, 328
multi-tier environments and, 341–344
officers, 24
project requirements and, 106
standards, 74
Web, 294
Security gap analysis matrix, 343–344
Sequential pattern discovery technique, 308
Server platforms, 328, 351
Service-level agreements. see SLAs (service-level agreements)
Services, OLAP, 291
Show-and-tell prototype, 156
“Single-swim-lane” development practices, 7–8, 9
Skills survey, 162–163
SLAs (service-level agreements)
  defining preliminary, 120–121
  establishing standards, 74–75
  nontechnical enterprise infrastructure evaluation, 76
Slimware, 155
Snowflake schema, 200
Software, application release concept, 361–362
Sorting utilities, ETL process, 225, 228
Source data
  analyzing quality of, 142
  bottom-up analysis of, 133–136
  creating inventory of, 67
  data analysis considerations of, 126
  for data mining, 306–307
  defining requirements for, 120
  determining condition of, 99
  ETL development, 260, 261–263
  guidelines for analyzing, 465–466
  selection process for, 137–139
  transformation problems with, 221–222
Source-to-target mapping document
  contents of, 233
  creating, 231–232
  overview of, 225–226
Spreadsheets, 322
SQL (Structured Query Language)
  E-R meta data repository design and, 248
  gateways using, 58
  object-oriented meta data repository design and, 249
Staffing
  data mining, 302
  database design, 193
  meta data repository analysis, 171, 183
  meta data repository design, 239
  meta data repository development, 320–321
  parallel development tracks, 17–18
  post-implementation review, 365
  project core team, 21
  project planning, 82
  resource dependencies, 95
  step core team, 21
Staging area, ETL
  contents of, 233
  design considerations, 212
  designing for, 228–229
  setting up, 232
Stakeholders
  application prototyping, 166
  data analysis, 145
  Human Resource Allocation Matrix, 381, 385
  project-specific requirements, 114
  release evaluation, 374
  responsibilities of, 24
Stand-alone systems, 7–8
Standards
  nontechnical, 64–65, 67, 71–75
  technical, 54
Star schema, 197–200
Start to Finish, task dependencies, 94–95
Start to Start, task dependencies, 94–95
Statistical analysis, 303–305
Step core team members, 20–23
Stock performance, 313
Stored procedures, 60
Stovepipe systems (automation silos), 65–66, 130
Strategic architect, 24
Stress testing, 271–272
Structured Query Language (SQL)
  E-R meta data repository design and, 248
  gateways using, 58
  object-oriented meta data repository design and, 249
Subject matter expert
  application development, 298
  application prototyping, 166
  book chapters of interest, xxxi
Subject matter expert continued
  business case assessment, 49
data analysis, 145
data mining, 316
ETL design, 234
ETL development, 278
ETL integration testing, 269
Human Resource Allocation Matrix, 379–383, 385
meta data repository analysis, 189
project planning, 102
project requirements definition, 114, 122
release evaluation, 374
responsibilities of, 22
Subtask matrix. see Task/Subtask Matrix
Success, measuring, 360
Summarization
database design, 204, 206
defining, 285
ETL development, 262
managing growth in data with, 350
OLAP tools providing, 285
Support (help desk)
implementation and, 338–339, 353
role of extended team members, 24
Synonyms, 222
SYSTABLES, 323

TCP/IP (Transmission Control Protocol/Internet Protocol), 351
Team members
  Application track, 18
  assessing risk to, 42
  book chapters of interest, xxxi
core team. see Core team members
database design, 193
ETL track, 17
extended team. see Extended team members
Meta Data Repository track, 18
post-implementation review, 367
project planning, 82
prototyping, 154, 468
resource assignment and, 93–94
Technical data conversion rules, 134
Technical enterprise infrastructure evaluation. see Enterprise infrastructure evaluation, technical
Test logs, 274–275, 297, 327, 333
Testers
  application development, 298
  book chapters of interest, xxxi
  ETL development, 278
  Human Resource Allocation Matrix, 383
  meta data repository development, 326–327, 334
  responsibilities of, 25
Testing
  acceptance, 268, 272, 277, 296, 326–327, 334
  application development, 296–297
  application prototypes, 154
  integration (system), 268, 270, 276, 296, 326–327, 332
  meta data repository development, 326–327, 331–333
  nontechnical enterprise infrastructure evaluation, 65, 76
  performance (stress), 268, 271–272, 276, 296
  quality (QA), 268, 272, 277, 296, 327
  regression, 268, 271–272, 276, 296, 326–327, 332
  standards for, 73–74
  unit, 268–269, 276, 296, 326, 331
Testing, ETL development, 268–275
  considerations about, 260
  peer reviews and, 267–268
  plan for, 273–275, 277
Time
  boxing, 152–153
  constraints, 84
  sequence discovery, data mining, 309–310
Index

Tool administrators, 25
Tool interface process, meta data repository, 324–325
Tools, see also OLAP (online analytical processing) tools
  access and analysis, 282
  application prototyping, 151, 164
  data cleansing, 139
  meta data sources, 240–242
Tools, ETL
  considerations about, 212
  evaluating, 229–231
  test functionality of, 232
Top-down logical data modeling
  define business rules, 465
  overview of, 128–133
Training
  application development, 296–297
  materials, 297, 333
  meta data repository, 329–330, 332, 333
  planning for, 338–339
  post-implementation review, 365
Transformation programs
  designing, 221–223
  designing ETL process flow, 228
  ETL development, 261–263
Triggers, 60, 86

U
Unit testing
  application development, 296
  ETL development, 269, 276
  meta data repository development, 326, 331
Usage
  DBMS selection criteria for, 60
  managing growth in, 350–351
  Utilization, of resources, 347–348

V
Validation, 151
Vendor evaluation, 252–254

W
Waterfall deployment, 7–8
Web
  application development considerations, 283
  environment, 294–295
  security, 344–345
Web developers
  application development, 297
  book chapters of interest, xxxi
  Human Resource Allocation Matrix, 381, 383–385
  implementation, 356
  release evaluation, 374
  responsibilities of, 25
Web master
  application development, 297
  application prototyping, 166
  book chapters of interest, xxxi
  Human Resource Allocation Matrix, 381, 383–385
  implementation, 356
  release evaluation, 374
  responsibilities of, 25
Word processing files, 321–322

X
XML-enabled meta data solution, 245–247. see also Distributed XML-enabled meta data solution.