

Index

Numbers

- 1NF (First Normal Form), 106
- 2NF (Second Normal Form), 107
- 3NF (Third Normal Form), 107
- 4NF (Fourth Normal Form), 107
- 5NF (Fifth Normal Form), 107
- 12 rules of relational databases, 103-105

A

- Access (Microsoft) databases, 226-227
- access, role-based matrix, 206-207
- accounting, fair-value accounting, 4
- accuracy, validity rules, 56
- acquisitions, data integration process, 36-37
- adaptation
 - data quality improvement cycle, 69
 - integration process, 45

- administration
 - goal development costs, 299
 - UD (unstructured data), 283-284
- administrators
 - data governance group, 60
 - metadata management initiative, 77
- agendas
 - DBMS vendors, 252-253
 - team weekly meetings, 158
- analysts, BI (business intelligence), 13
- analytics, BI (business intelligence), 266
- APIs (application programming interfaces), 286
- applications
 - inventory, EA (enterprise architecture), 67
 - packages, 176-177, 232-233
 - measurements, performance monitoring, 193-194

- archeology, quality improvement practices, 57-58
 - archiving
 - data (tuning option), 197
 - metadata management, 95
 - UD (unstructured data), 284
 - assessments
 - Data Environment Assessment Questionnaire, 16-21
 - data quality improvement cycle, 68
 - organization teams, 156-157
 - assets
 - business data, 291-295
 - data integration, 31
 - organizations data, 4
 - atomic values, 102
 - attributes
 - business quality rules, 53-55
 - completeness, 56
 - auditing procedures, 211-212
 - availability
 - data strategy development costs, 296
 - establishing benchmark criteria and methodology, 173-174
- B**
- backups, metadata management, 95
 - balanced scorecard (BSC), 268-269
 - BAM (business activity monitoring), 30
 - BCNF (Boyce-Codd Normal Form), 107
 - benchmarks, capacity planning, 168-175
 - benefits of metric measurement
 - better decisions, 306-307
 - cash flow acceleration, 301
 - competitive effectiveness, 306
 - cost containment, 302-303
 - customer attrition control, 304
 - customer conversion rates, 303
 - customer service, 307
 - data mart consolidation, 305
 - demand management, 303
 - DW (Data Warehouse), 301
 - employee empowerment, 307
 - fraud reduction, 303
 - improved supplier relationships, 304
 - marketing campaign responses, 304
 - post implementation measurement, 307-308
 - productivity analysis, 301-302
 - public relations, 305-306
 - revenue enhancement, 301
 - Berkeley study, explosion of volume in data, 280
 - best practices, RFPs (requests for proposals), 242-245
 - BI (business intelligence), 7, 13-14, 259-261, 274
 - benefits, 262-263
 - BSC (balanced scorecard), 268-269
 - CRM, 263
 - data
 - cleansing, 265-266
 - mining, 267-268
 - presentation, 267
 - transformation, 265-266
 - visualization, 267
 - digital dashboards, 269
 - ERM, 263
 - history, 261-262
 - integration risks, 41
 - metadata repositories, 265
 - myths, 272-274
 - office politics, 263
 - OLAP tools and analytics, 266
 - pitfalls, 272-274
 - ROI, 262
 - rule-based analytics, 268
 - trends and technologies, 269
 - data mining, 270
 - RFID (Radio Frequency Identification), 271-272
 - big-bang effort, building enterprise logical data model, 109-110

- BLOBs (Binary Large Objects), 279
 - Boston Globe, report on cost of converting to digital media, 290
 - bottom-up logical data modeling, 112-115
 - Boyce-Codd Normal Form (BCNF), 107
 - BPM (business performance management), 30
 - break-even analysis, 309, 312
 - Brown, Robert, 101
 - BSC (balanced scorecard), 268-269
 - buffer pools (tuning option), 196
 - business activity monitoring (BAM), 30
 - Business Data Model, EA (enterprise architecture), 67
 - Business Function Model, EA (enterprise architecture), 67
 - business performance management (BPM), 30
 - Business Process Model, EA (enterprise architecture), 67
 - business-focused data analysis, data modeling, 106
- C**
- caching (tuning option), 196
 - calculations
 - cost template, 312-314
 - intangible benefits template, 315
 - ROI, 309
 - cost of capital, 309
 - risk, 310
 - California SB 1386 Identity Protection Bill, 210
 - call centers, data types, 292
 - Canada, security laws, 211
 - capability maturity model (CMM), 43
 - capacity planning (performance modeling), 166-168
 - benchmark teams, 169
 - communication of results, 175
 - costs, 171
 - criteria and methodology, 171-174
 - evaluation and measurement of results, 174-175
 - goals and objectives, 170
 - reasons for pursuing a benchmark, 168-169
 - standard benchmarks, 170-171
 - verification and reconciliation of results, 175
 - cardinality, business entity quality rules, 52
 - CASE (computer aided software engineering), 79, 94, 113
 - case studies, performance, 198-201
 - cash flow, strategic goal benefits metric, 301
 - categorization
 - DBMS vendor capabilities and functions, 254-255
 - data, 14-15, 296
 - central processing units (CPUs), 102
 - centralized metadata repositories, 85-86
 - certification data, 4
 - challenges, integrating data, 40-41
 - channels, business data value, 293-294
 - Character Large Objects (CLOBs), 279
 - Chen, Dr. Peter, 99, 101
 - chief information officer (CIOs), 133
 - chief operating officers (COOs), 269
 - chief technology officer (CTOs), 133
 - CIOs (chief information officer), 133
 - class words, business metadata names, 80
 - classes, choices, 139
 - cleansing
 - BI (business intelligence), 265-266
 - quality improvement practices, 58-59
 - CLOBs (Character Large Objects), 279

- club cards, business data value, 294
- CMM (capability maturity model), 43
- Codd, Dr. Edgar F., 102-105
- combining structured and unstructured data, 287
- commercial off-the-shelf (COTS), 176-177
- competencies, information steward, 155
- competitions, strategic goal benefits metric, 306
- completeness, validity rules, 55
- compliances
 - failures, reducing through data integration, 31
 - information legislation, 30
- Comprehensive Data Sublanguage rule (12 rules of relational databases), 104
- computer aided software engineering (CASE), 79, 94, 113
- conformance to measures of success, performance monitoring, 191
- consistencies, validity rules, 57
- consolidation, data integration, 42
- Constantine, Larry, 100
- content reusability, unified content strategy, 286
- contextual information, metadata, 74
 - analysis, 89
 - categories, 78-82
 - construction, 91-92
 - critical data strategy, 74-78
 - deployment, 92-93
 - design, 90-91
 - justification, 88
 - MME (Managed Metadata Environment), 93-97
 - planning, 88-89
 - repositories, 84-87
 - sources, 82-84
- COOs (chief operating officers), 269
- corporate assets, data integration, 31
- Corporate Data Stewardship Function*, 151
- correctness, validity rules, 56
- costs
 - benchmarks (capacity planning), 171
 - BI (business intelligence), 13
 - calculation template, 312-314
 - of capital, ROI calculation, 309
 - DBMSs, TCO (total cost of ownership), 228-232
 - integration risks, 40
 - justification process, risk, 310
 - reducing through data integration, 29
 - strategic goal development, 295-296
 - cost categories, 296-300
- COTS (commercial off-the-shelf), 176-177
- critical success factors (CSFs), 249
- CRM (Customer Relationship Management), 27, 263
 - BI, 263
 - cost containment, 303
 - promises *versus* realities, 27
- CSFs (critical success factors), 249
- CTOs (chief technology officer), 133
- cultures (company), influence on physical data model, 128
- currencies, integration risks, 41
- Customer Relationship Management. *See* CRM
- customers
 - attrition control, 304
 - BI (business intelligence), 13
 - call centers, 292
 - channel preferences, 293-294
 - click-stream data, 293
 - companies that sell data, 292
 - conversion rates, 303
 - demographics, 293
 - direct retailers, 294
 - internal information, 292
 - loyalty cards, 294
 - service integration, 30
 - travel data, 294-295

D

- DAM software (Digital Asset Management), 287-288
- DAs (data administrators), 10, 142
- data definition language (DDL), 114, 141
- Data Environment Assessment Questionnaire, 16-21
- data integration, 6-7
 - business case for, 31-32
 - business data
 - acquisitions, 36-37
 - data lineage, 37-38
 - knowing business entities, 35-36
 - mergers, 36-37
 - multiple DBMSs, 38
 - redundancy, 37
 - CMM (capability maturity model), 43
 - consolidating data, 42
 - CRM (Customer Relationship Management), 27
 - data modeling, 108
 - definitions, 23-24
 - disintegrated data, 24
 - DW (Data Warehousing), 26-27
 - EAI (Enterprise Application Integration), 28
 - ERP (Enterprise Resource Planning), 24-25
 - federating data, 42-43
 - implementation planning, 44-45
 - industry opportunities, 32-35
 - logical, 38
 - management support, 29-31
 - physical, 38
 - prioritizing data, 39-40
 - risks, 40-41
 - silver-bullet solutions, 24
- Data Mart (DM), 264
- data modeling, 9, 99-100
 - enterprise logical data model, 109
 - big-bang effort *versus* incremental, 109-112
 - top-down *versus* bottom-up, 112-115
 - logical data model, 105-108
 - process-independence, 105-106
 - origins, 100-101
 - physical data modeling, 115
 - database design, 117
 - database views, 122
 - denormalization, 117-120
 - dimensional model, 122-126
 - indexes, 121
 - influential factors, 126-130
 - partitioning, 121-122
 - process-dependence, 116
 - surrogate keys, 120-121
 - significance of, 102-105
- data ownership, 148-151
- data quality steward, 143-144
- Data Warehousing. *See* DW
- database administrators (DBAs), 10, 82, 141-142
- database management system. *See* DBMSs
- databases
 - 12 rules of relational databases, 103-105
 - Access, 226-227
 - controls, 213
 - design, physical data modeling, 117
 - metadata repositories, 77, 84
 - analysis, 89
 - building, 85
 - centralized, 85-86
 - construction, 91-92
 - deployment, 92-93
 - design, 90-91
 - distributed, 86-87
 - justification, 88
 - planning, 88-89
 - purchasing product, 84-85
 - XML-enabled, 87
 - security, 213
 - views, physical data modeling, 122
- Date, Christopher J., 102
- DB2 (IBM), 226

- DBAs (database administrator), 10, 82, 141-142
- DBMSs (database management systems), 2, 223
 - application packages, 232-233
 - available choices, 226
 - capabilities/functions, 224-226
 - dictionaries, MME source, 94
 - ERPs, 232-233
 - multiple, integration process, 38
 - parameters (tuning option), 196
 - RFPs (requests for proposals), 242
 - best practices, 242-245
 - response formats, 246
 - selecting, 12
 - criteria, 233-234
 - process, 234-241
 - standardization, 12, 227-228
 - vendor evaluation, 246-249
 - early code, 250
 - financial capacity, 254
 - level of service, 250
 - performance, 249
 - personnel capacity, 253
 - rules of engagement, 250-252
 - selection matrix, 254-255
 - setting agenda for meetings and presentations, 252-253
- DDL (data definition language), 114, 141
- defect prevention, quality improvement practices, 59-60
- DeMarco, Tom, 100
- demographics, customer information, 293
- denormalization, physical data modeling, 117-120
- dependencies, data quality rules, 54-55
- designs
 - performance, 177-189
 - security, 213-214
- desired references, DBMS selection, 237-238
- Dessert logical data model, 117
- developer tools, MME source, 94
- development
 - data strategies, 15
 - quality disciplines methodology, 63
 - strategic goals, costs, 295-300
- dictionaries, DBMS, MME source, 94
- Digital Asset Management software (DAM software), 287-288
- digital dashboards, 269
- Digital rights management (DRM software), 102, 288, 290
- dimensional model, physical data modeling, 122-124
 - snowflake schema, 125
 - star schema, 124-125
 - starflake schema, 126
- direct retailers, business data value, 294
- dirty data
 - defect prevention, 59-60
 - enterprise quality disciplines, 65
 - quality improvement practices, 58-59
 - recognizing, 49-51
- disciplines, quality
 - development methodology, 63
 - dirty data handling, 65
 - manipulation reconciliation, 65
 - maturity levels, 61-62
 - metadata components, 63-64
 - metrics, 66
 - modeling, 64-65
 - naming and abbreviations
 - standards, 63
 - security, 66
 - standards and guidelines, 62-63
 - testing, 65
- Discovery data mining, 268
- disintegrated data, 24
- distributed metadata repositories, 86-87
- distributed organizations, 137
- Distribution Independence rule (12 rules of relational databases), 105

- DM (Data Mart), 264
- documentation, metadata
 - analysis, 89
 - categories, 78-82
 - construction, 91-92
 - critical data strategy, 74-78
 - deployment, 92-93
 - design, 90-91
 - justification, 88
 - MME (Managed Metadata Environment), 93-97
 - planning, 88-89
 - repositories, 84-87
 - sources, 82, 84
- Documentum™ (EMC), 283
- domains
 - business attribute data quality, 54
 - completeness, 56
 - data ownership, 148
- dormant data, measurements for
 - monitoring performance, 192-193
- DRM (Digital rights management)
 - software, 102, 288, 290
- DW (Data Warehousing), 26-27, 99, 264-265
 - DM (Data Mart), 264
 - EDW (Enterprise Data Warehouse), 264
 - integration risks, 41
 - ODS (Operational Data Store), 264
 - promises *versus* realities, 26-27
 - strategic goals
 - benefits metrics, 301
 - development costs, 296-298
- Dynamic On-Line Catalog Based on the Relational Model rule, 103
- E**
- E/R model (entity-relationship model), 99, 101, 105
 - business-focused data analysis, 106
 - data integration, 108
 - data quality, 109
 - process-independence, 105-106
- EA (enterprise architecture), 66-69
- EAI (Enterprise Application Integration), 28
- early code, DBMS vendors, 250
- ECMS (enterprise content management systems), 283, 287-288
- education
 - data quality improvement cycle, 69
 - integration planning, 44
- EDW (Enterprise Data Warehouse), 264
- EII (Enterprise information integration)
 - tools, 42
- electronic medical records (EMR
 - software), 290
- employee information, DBMS
 - vendors, 253
- EMR software (electronic medical records), 290
- encryption, 214
- English, Larry, CMM (capability maturity model) adaptation, 61
- Enterprise Application Integration (EAI), 28
- enterprise architecture (EA), 66
- enterprise content management systems (ECMS), 283, 287-288
- Enterprise Data Warehouse (EDW), 264
- Enterprise information integration (EII)
 - tools, 42
- enterprise logical data model, 109
 - big-bang effort *versus* incremental, 109-112
 - top-down *versus* bottom-up, 112-115
- Enterprise Resource Planning. *See* ERPs
- entity completeness, 55
- entity-relationship model. *See* E/R model

environments, metadata management, 96
 ERPs (Enterprise Resource Plannings), 2,
 24-25, 176-177, 232-233, 263
 BI, 263
 promises *versus* realities, 25
 errors, minimizing data errors, 7
 ETL (extract, transform, load), 8, 47,
 100, 142
 EU (European Union), 146, 211
 Europe, fair-value accounting, 4
 European Union (EU), 146
 evaluation
 data quality improvement cycle, 69
 DBMS vendors, 246-249
 early code, 250
 financial capacity, 254
 level of service, 250
 performance, 249
 personnel capacity, 253
 rules of engagement, 250-252
 selection matrix, 254-255
 setting agenda for meetings and
 presentations, 252-253
 results, benchmarks (capacity
 planning), 174-175
 executing, integration process, 45
 executives, quality incentive programs,
 69-71
 external data
 integration risks, 41
 security, 216
 external users, auditing procedures, 212
 extract/transform/load (ETL), 8, 47,
 100, 142

F

fair-value accounting, 4
 Family Educational Rights and Privacy
 Act (FERPA), 210
 Federal Bureau of Investigation (FBI), 270
 federation, data integration, 42-43

FERPA (Family Educational Rights and
 Privacy Act), 210
 Fifth Normal Form (5NF), 107
 financial capacity, DBMS vendors, 254
 First Normal Form (1NF), 106
 Flavin, Mat, 101
 foreign keys, 114
 Fourth Normal Form (4NF), 107
 fraud
 BI (business intelligence), 13
 detecting through data
 integration, 32
 strategic goal benefits metric, 303

G

Gane-Sarson, 100
 Gartner Group, report on BI, 262-263
 gathering references, DBMS selection, 237
 goals
 benchmarks (capacity planning), 170
 organizations data, 5-6
 ROI (return on investment), 295
 governors (tuning option), 197
 Gramm-Leach-Bliley Act, 210
 Guarantees Access rule (12 rules of
 relational databases), 103
 guidelines, quality disciplines, 62-63

H

Health Insurance Portability and
 Accountability Act (HIPAA), 210
 help desk/support, DBMS TCO (total
 cost of ownership), 231
 HIPAA (Health Insurance Portability and
 Accountability Act), 210
 history
 BI (business intelligence), 261-262
 data, quality, 8
 UD (unstructured data), 278, 280

HMOs, rule-based analytics, 268
 HOLAP (Hybrid OLAP), 266
 horizontal partitioning, 121
 hospitals, rule-based analytics, 268
 Hybrid OLAP (HOLAP), 266

I

IBM, DB2, 226
 IDUG (International DB2 User Group), 238
 IFS (Oracle), 283
 implementation
 data
 integration planning, 44-45
 quality improvement cycle, 69
 strategies, 15
 performance, 177, 180
 planning, 44-45
 improvement practices, quality
 cleansing dirty data, 58-59
 data profiling, 57-58
 defect prevention, 59-60
 inaccurate data, 49
 incentives, executive quality sponsorship, 69-71
 incomplete data, 50
 inconsistent data, 50
 incorrect data, 49
 incremental effort, building enterprise
 logical data model, 109-112
 indexes
 physical data modeling, 121
 tuning option, 196
 influential factors, physical data
 modeling, 126
 cultural influence, 128
 DBMS software, 127
 denormalization for short-term
 solutions, 127
 KISS principle, 130
 metric facts, 129-130
 modeling expertise, 128
 powerful servers, 127
 robust models, 126-127
 user-friendly structures, 129
 information
 consumers, 71
 legislation, compliance, 30
 rule (12 rules of relational
 databases), 103
 stewards, roles and responsibilities,
 151-155
 information resource management
 (IRM), 102
 inheritance, business attribute data
 quality, 53-54
 intangible benefits, template, 315
 integration (data), 6-7, 23
 business case, 31-32
 business data
 acquisitions, 36-37
 data lineage, 37-38
 knowing business entities, 35-36
 mergers, 36-37
 multiple DBMSs, 38
 redundancy, 37
 CMM (capability maturity
 model), 43
 consolidating data, 42
 CRM (Customer Relationship
 Management), 27
 definitions, 23-24
 disintegrated data, 24
 DW (Data Warehousing), 26-27
 EAI (Enterprise Application
 Integration), 28
 ERP (Enterprise Resource Planning),
 24-25
 federating data, 42-43
 implementation planning, 44-45
 industry opportunities, 32-35
 logical, 38
 management support, 29-31
 physical, 38

prioritizing data, 39-40
 risks, 40-41
 silver-bullet solutions, 24
 standardized DBMSs, 227
 integrity, DBMS vendors, 247
 Integrity Independence rule (12 rules
 of relational databases), 105
 intellectual capital, 4
 internal cost containment, 302
 internal rate of return (IRR), 311
 internal staff, DBMS TCO (total cost
 of ownership), 231
 International DB2 User Group
 (IDUG), 238
 international rules, security, 211
 IOUG (International Oracle User
 Group), 238
 IRM (information resource
 management), 102
 IRR (internal rate of return), 311

J-K

Jennings, Michael, *Universal Meta Data
 Models*, 93
 job scheduling, metadata management,
 96

 Kelvin, Lord, 11, 21
 Kimball, Ralph, 99
 KISS (keep it simple stupid)
 principle, 130

L

Large Objects (LOBs), 279
 legacy systems, retiring through
 integrated databases, 32
 legalities, prioritizing data, 40
 level of service, DBMS vendors, 250

levels, CMM (capability maturity model),
 43
 lineage
 data integration process, 37-38
 Y2K, 38
 load time, capacity planning, 174
 LOBs (Large Objects), 279
 Logical Data Independence rule (12 rules
 of relational databases), 104
 logical data integration, 38
 logical data model, 101, 105
 business-focused data analysis, 106
 data integration, 108
 data quality, 109
 enterprise logical data model, 109
 big-bang effort *versus* incremental,
 109-112
 top-down *versus* bottom-up,
 112-115
 enterprise quality discipline, 64
 process-independence, 105-106
 loyalty cards, business data value, 294

M

Managed Metadata Environment.
See MME
 management
 data, 15
 integration support, 29-31, 40
 planning integration, 44
Managing Enterprise Content, 290
 many-to-many cardinality, 52
 many-to-one cardinality, 52
 Marco, David, *Universal Meta Data
 Models*, 93
 marketing
 business data value, 294
 response rates, 304
Mastering Data Warehouse Design, 122
 maturity levels, data quality, 61-62

- means of measurement, performance monitoring, 193
- measurements
 - integration planning, 44
 - performance monitoring, 190
 - conformance to measures of success, 191
 - dormant data, 192-193
 - means of measurement, 193
 - reporting results to management, 194-195
 - resource utilization, 192
 - response time, 191
 - responsibility for measurement, 193
 - ROI (return on investment), 194
 - usage metrics, 191
 - use of measurement, 193-194
 - user satisfaction, 192
 - results, benchmarks (capacity planning), 174-175
- meetings, DBMS vendors, 252-253
- membership cards, business data value, 294
- mergers, data integration process, 36-37
- metadata, 8-9, 74
 - administrator
 - data governance group, 60
 - roles and responsibilities, 142
 - categories, 78-79
 - business, 79-81
 - process, 81-82
 - technical, 81
 - usage, 82
 - critical data strategy, 74
 - business intelligence keystone, 74-75
 - management initiative, 76-78
 - required support, 75
 - data strategy development costs, 296
 - enterprise quality disciplines, 63-64
 - MME (Managed Metadata Environment), 93-94
 - communication, 97
 - delivery, 97
 - integration, 95
 - management, 95-96
 - metadata marts, 96
 - selling, 97
 - sources, 94
 - repository, 77, 84, 265
 - analysis, 89
 - building, 85
 - centralized, 85-86
 - construction, 91-92
 - deployment, 92-93
 - design, 90-91
 - distributed, 86-87
 - EA (enterprise architecture), 68
 - justification, 88
 - planning, 88-89
 - purchasing product, 84-85
 - XML-enabled, 87
 - sources, 82, 84
- methodology, benchmarks (capacity planning), 171
 - actual test data and queries, 172
 - availability, 173-174
 - data volume, 172
 - load time, 174
 - success criteria, 172-173
 - system configuration, 172
- metrics
 - enterprise quality disciplines, 66
 - facts, influence on physical data model, 129-130
 - monitoring performance, 190
 - conformance to measures of success, 191
 - dormant data, 192-193
 - means of measurement, 193
 - reporting results to management, 194-195
 - resource utilization, 192
 - response time, 191
 - responsibility for measurement, 193
 - ROI (return on investment), 194
 - usage, 191

- use of measurement, 193-194
 - user satisfaction, 192
 - strategic goal benefits
 - benefit decisions, 306-307
 - cash flow acceleration, 301
 - competitive effectiveness, 306
 - cost containment, 302-303
 - customer attrition control, 304
 - customer conversion rates, 303
 - customer service, 307
 - data mart consolidation, 305
 - demand management, 303
 - employee empowerment, 307
 - fraud reduction, 303
 - improved supplier relationships, 304
 - marketing campaign
 - responses, 304
 - post implementation
 - measurement, 307-308
 - productivity analysis, 301-302
 - public relations, 305-306
 - revenue enhancement, 301
 - Microsoft, 226
 - The Mind Manipulators: A Non-Fiction Account*, 205
 - mining (data), 267
 - Discovery, 268
 - Predictive, 267
 - trends, 270
 - MME (Managed Metadata Environment), 93-96
 - modeling, 9
 - data modeling, 99-100
 - enterprise logical data model, 109-115
 - logical data model, 105-109
 - origins, 100-101
 - physical data modeling, 115-130
 - significance of, 102-105
 - data strategy development costs, 296
 - enterprise quality discipline, 64-65
 - expertise, influence on physical data model, 128
 - performance, capacity planning, 166-175
 - requirements, 166
 - modules, ERP (Enterprise Resource Planning), 24-25
 - MOLAP (Multidimensional OLAP), 266
 - monitoring
 - measurements, 190
 - conformance to measures of success, 191
 - dormant data, 192-193
 - means of measurement, 193
 - reporting results to management, 194-195
 - resource utilization, 192
 - response time, 191
 - responsibility for measurement, 193
 - ROI (return on investment), 194
 - usage metrics, 191
 - use of measurement, 193-194
 - user satisfaction, 192
 - security policies, 217
 - Multidimensional OLAP (MOLAP), 266
 - MySQL, 226
 - myths, BI (business intelligence), 272-274
- N**
- naming, quality standards, 63
 - Napster™, 289
 - NCR, Teradata, 226
 - NDA (nondisclosure agreements), 242
 - near real-time data
 - integration risks, 41
 - versus* real time, 150
 - net present value (NPV), 309-311
 - network usage, DBMS TCO (total cost of ownership), 230

nondisclosure agreements (NDAs), 242
 nonintegrated data, 50
 Nonsubversion rule (12 rules of relational databases), 105
 normalization rules, 106, 111, 122
 NPV (net present value), 309-311

O

ODS (operational data store), 147, 264
 office politics, BI, 263
 OLAP (online analytical processing), 143, 262, 266
 OLTP (online transaction processing), 161, 165, 261
 one-to-many cardinality, 52
 one-to-one cardinality, 52
 one-to-one optionality, 52
 one-to-zero optionality, 53
 online analytical processing (OLAP), 143, 262, 266
 online transaction processing (OLTP), 161, 165, 261
 operational data

- cleansing dirty data, 58-59
- defect prevention, 59-60

 operational data store (ODS), 147, 264
 operational transactions, 165
 opportunities, integration support, 32-35
 optimizer tweeking (tuning option), 196
 optionality, business entity quality rules, 52-53
 options, tuning, 196-197
 Opton, Edward M. Jr., *The Mind Manipulators: A Non-Fiction Account*, 205
 Oracle, 226, 283
 organizations

- responsibilities, 10
- roles, 10

security, 11-12
 strategic goals, ROI (return on investment), 295
 teams

- assessment exercise, 156-157
- building, 134
- change resistance, 134-135
- data ownership, 148-151
- information stewards, 151-155
- roles and responsibilities, 140-148
- structure, 135-138
- training, 138-140
- weekly meeting agenda, 158
- worst practices, 156

 UD (unstructured data), 282
 unstructured data, 14
 vision and goals, 4-6
 origins, data modeling, 100-101
 outsourced personnel, 137-138
 ownership, data, 148-150

P

Page-Jones, Meilir, 100
 partitioning, physical data modeling, 121-122
 payback period, 309
 performance-guiding principles, 162-163
 personnel, goal development costs, 298
 personnel capacity, DBMS vendors, 253
 Physical Data Independence rule (12 rules of relational databases), 104
 physical data integration, 38
 physical data modeling, 115-129
 pitfalls, BI (business intelligence), 272-274
 planning, data

- integration, 44-45
- quality improvement cycle, 69

 policies

- metadata management initiative, 77
- security, 217, 218

politics, prioritizing data, 39
 practices, quality improvement
 cleansing dirty data, 58-59
 data profiling, 57-58
 defect prevention, 59-60
 precision, validity rules, 56
 Predictive data mining, 267
 preferred savings cards, business data
 value, 294
 presentations, DBMS vendors, 252-253
 prevention, quality improvement
 practices, 59-60
 prime words, business metadata
 names, 80
 principles, performance-guiding, 162-163
 prioritizing
 data integration, 39-40
 planning integration, 44
 privacy
 auditing procedures, 211-212
 common practices, 218-219
 data, 11-12
 ownership, 148-149
 sensitivity exercise, 219-220
 strategy development costs, 296
 warehouse, 215. *See also* DW
 design, 213-214
 policies, 217-218
 regulatory laws, 210-211
 role-based access matrix, 206-207
 staff roles and responsibilities,
 208-210
 vendors
 external data, 216
 software, 215-216
 procedures, metadata management
 initiative, 77
 process for selection, DBMSs, 234-241
 process-dependence, physical data
 modeling, 116
 process-independence, data modeling,
 105-106

processes
 improvement through data
 integration, 31
 integration, 35-41
 metadata, 81-82
 enterprise quality disciplines, 63
 sources, 83
 reference checking, DBMS selection,
 238-239
 product development time, increasing
 efficiency through data integration, 29
 production data, security, 213
 productivity, strategic goal benefits
 metric, 301-302
 professional employee information,
 DBMS vendors, 253
 profiling, quality improvement practices,
 57-58
 public relations, strategic goal benefits
 metric, 305-306
 purging, metadata management, 95

Q

qualifiers, business metadata names, 80
 queries, 161
 design reviews, 185-186
 establishing benchmark criteria
 and methodology (capacity
 planning), 172
 questionnaires,
 Data Environment Assessment, 16-21
 reference checking, DBMS selection,
 239-241

R

radio frequency identification (RFID),
 29, 271
 rate of return, 309-311
 real-time data
 integration risks, 41
 versus near real time, 150

- reconciliation
 - enterprise quality discipline, 65
 - results, benchmarks (capacity planning), 175
- recoveries, metadata management, 96
- recruiting, planning integration, 44
- redundancy (data)
 - integration process, 37
 - minimizing, 7-9
- reference checking, DBMS selection, 236
 - alternatives to reference checking, 236-237
 - desired references, 237-238
 - process, 238-239
 - questions to ask, 239-241
 - selecting and gathering references, 237
- referential integrity (RI), 197
- regulations, prioritizing data, 40
- regulatory laws, security, 210-211
- relational database management system.
See RDBMS
- relational databases, 12 rules of relational databases, 103-105
- relational model, 102
- “A Relational Model of Data for Large Shared Data Banks” (Codd), 102
- Relational OLAP (ROLAP), 266
- relationship completeness, 56
- reporting
 - design reviews, 185-186
 - metadata marts, 96
- repositories, metadata, 77, 84
 - analysis, 89
 - building, 85
 - centralized, 85-86
 - construction, 91-92
 - deployment, 92-93
 - design, 90-91
 - distributed, 86-87
 - justification, 88
 - planning, 88-89
 - purchasing product, 84-85
 - XML-enabled, 87
- requests for information (RFIs), 242
- requests for proposals. *See* RFPs
- requests for quotes (RFQs), 242
- requirements
 - effective modeling, 166
 - performance, 163-164
- research, planning integration, 44
- resource utilization, measurements for monitoring performance, 192
- response time
 - measurements for monitoring performance, 191
 - SLAs (service level agreements), 165-166
- responsibilities, 140
 - consultants, 145
 - contractors, 145
 - data administrator, 142
 - data ownership, 148-151
 - data quality steward, 143-144
 - data strategist, 140-141
 - DBA (database administrator), 141-142
 - information stewards, 151-155
 - metadata administrator, 142
 - organizational, 10
 - security, 145-146, 208-210
 - sharing data, 146-147
 - strategic data architect, 147
 - technical services, 147-148
 - worst practices, 156
- responsibility for measurement, performance monitoring, 193
- retail, cost containment, 302
- retailers, 294
- retention, UD (unstructured data), 285-286
- return on investment. *See* ROI
- revenues
 - BI (business intelligence), 13

- increasing through data
 - integration, 29
 - strategic goal benefits metric, 301
 - reviews, design reviews, 180-187
 - RFID (radio frequency identification), 29, 271-272
 - RFIs (requests for information), 242
 - RFPs (requests for proposals), 242
 - DBMSs, 242
 - best practices, 242-245
 - response formats, 246
 - RFQs (requests for quotes), 242
 - RI (referential integrity), 197
 - risks
 - integrating data, 40-41
 - ROI calculation, 310
 - robust models, influence on physical data model, 126-127
 - Rockley, Ann, unified content strategy, 282
 - ROI (return on investment), 194, 295
 - BI, 262
 - break-even analysis, 309, 312
 - calculations, 309
 - cost of capital, 309
 - example, 310-312
 - risk, 310
 - net present value, 309-311
 - performance monitoring, 194
 - rate of return, 309-311
 - ROLAP (Relational OLAP), 266
 - role-based access matrix, 206-207
 - roles, 140
 - assessment exercise, 156-157
 - consultants, 145
 - contractors, 145
 - data administrator, 142
 - data ownership, 148-151
 - data quality steward, 143-144
 - data strategist, 140-141
 - DBA (database administrator), 141-142
 - information stewards, 151-155
 - metadata administrator, 142
 - organizational, 10
 - security, 145-146
 - security officer, 208-209
 - system administrators, 209-210
 - sharing data, 146-147
 - strategic data architect, 147
 - technical services, 147-148
 - worst practices, 156
 - rule-based analytics, 268
 - rules
 - 12 rules of relational databases, 103-105
 - data quality
 - business attributes, 53-54
 - business entities, 51-53
 - dependency, 54-55
 - validity, 55-57
 - of engagement, DBMS vendors, 250-252
 - normalization, 106, 111, 122
- S**
- Sarbanes-Oxley Act of 2002, 30, 210, 281
 - satisfaction surveys, measurements for monitoring performance, 192
 - Schefflin, Alan W., *The Mind Manipulators: A Non-Fiction Account*, 205
 - SCI (supply chain intelligence), 29
 - Scofield, Michael, *Corporate Data Stewardship Function*, 151
 - SDLC (system development life cycle), 77, 100
 - searchability, UD (unstructured data), 286-287
 - Second Normal Form (2NF), 107
 - security
 - auditing procedures, 211-212
 - common practices, 218-219

- data, 11-12
 - warehouse, 215. *See also* DW
 - ownership, 148-149
 - sensitivity exercise, 219-220
 - strategy development costs, 296
- databases, 213
- design, 213-214
- enterprise quality disciplines, 66
- policies, 217-218
- prioritizing data, 39
- regulatory laws, 210-211
- role-based access matrix, 206-207
- roles and responsibilities, 145-146, 208-210
- vendors, 215-216
- selection
 - criteria, DBMSs, 233-234
 - matrix, DBMS vendors, 254-255
 - process, DBMSs, 234-241
- senior management, 15
- service level agreements. *See* SLAs
- set placement of data (tuning option), 196
- shared data, 146-147
- simple object access protocol (SOAP), 87
- Single Version of the Truth, data modeling, 9
- Six Sigma, 270
- skills, information steward, 155
- SLAs (service level agreements), 140, 164
 - data strategists responsibilities, 140
 - data warehouse, queries, 166
 - metrics, 165
 - online transactions, 166
 - response time, 165-166
- smart keys (tuning option), 196
- snowflake schema (dimensional model), 125
- SOAP (simple object access protocol), 87
- software
 - DAM (Digital Asset Management), 287-288
 - DRM (Digital rights management), 288, 290
 - EMR (electronic medical records), 290
 - expense, standardized DBMSs, 228
 - goal development costs, 297-298
 - security, 215-216
- sources, metadata, 82-84
- sponsorship, integration planning, 44
- spreadsheets, MME source, 94
- SQL Server (Microsoft), 226
- stability, DBMS vendors, 246
- staff
 - assessment exercise, 156-157
 - data ownership, 148-151
 - expenses, standardized DBMSs, 228
 - goal development costs, 298
 - information stewards, 151-155
 - responsibilities, 140-148
 - structure, 135-136
 - distributed organizations, 137
 - outsourced personnel, 137-138
 - training, 138-140
 - weekly meeting agenda, 158
 - worst practices, 156
- standard benchmarks (capacity planning), 170-171
- standardization, DBMSs, 12, 227
 - integration, 227
 - reduced staff expense, 228
 - software expense, 228
- standards
 - quality disciplines, 62-63
 - resistance to, 135
 - security, 214
 - XML-enabled metadata repositories, 87
- star schema (dimensional model), 124-125
- starflake schema (dimensional model), 126

storage, UD (unstructured data), 283-284
 strategic data architect, 147
 strategic goals
 benefits metric measurement, 301-308
 development costs, 295-300
 ROI (return on investment), 295
 stewards, roles and responsibilities, 151-155
 success criteria, establishing benchmark criteria and methodology, 172-173
 summary tables (tuning option), 196
 suppliers, improved relationships, 304
 supply chain intelligence (SCI), 29
 supply chains, improving through data integration, 29
 support, DBMS vendors, 246
 surrogate keys, physical data modeling, 120-121
 Sybase, 226
 system development lifecycle (SDLC), 77, 100
 Systematic Treatment of Null Values rule (12 rules of relational databases), 103

T

tables, design reviews, 184-185
 tasks, performance, 201-202
 TCO (total cost of ownership), 228, 299
 DBMSs, 228
 actual DBMS, 230
 consultants and contractors, 231
 hardware, 230
 help desk/support, 231
 internal staff, 231
 IT training, 232
 network usage, 230
 operations and system administration, 232
 goal development costs, 299-300

teams
 assessment exercise, 156-157
 building, 134
 change resistance
 existing staying same, 134-135
 nonacceptance to standards, 135
 reasons for, 135
 data strategy, 15
 information stewards, 151-155
 responsibilities, 140-148
 structure, 135-136
 distributed organizations, 137
 outsourced personnel, 137-138
 training, 138
 choices for classes, 139
 employees attendance, 138-139
 required mindset, 139
 timing, 140
 weekly meeting agenda, 158
 worst practices, 156
 technical metadata, 81
 enterprise quality disciplines, 63
 sources, 83
 technical segment, metadata repositories, 265
 technical services, roles and responsibilities, 147-148
 techniques, anticipating performance, 167
 technologies
 BI (business intelligence), 269
 data mining, 270
 RFID (Radio Frequency Identification), 271-272
 UD (unstructured data), 287
 DAM software, 287-288
 DRM (Digital rights management) software, 288-290
 EMR (electronic medical records) software, 290
 Teradata (NCR), 226
 testing
 data
 establishing benchmark criteria and methodology, 172

- security, 213
 - enterprise quality discipline, 65
 - information, design reviews, 187
 - Third Normal Form (3NF), 107
 - time, real *versus* near real time, 150
 - Title VIII (Sarbanes-Oxley Act of 2002), 281
 - top-down logical data modeling, 112
 - total cost of ownership. *See* TCO
 - Total Quality Management, 270
 - training
 - goal development costs, 299
 - IT, DBMS TCO (total cost of ownership), 232
 - security policies, 217
 - teams, 138
 - choices for classes, 139
 - employees attendance, 138-139
 - required mindset, 139
 - timing, 140
 - transactions, 161, 165
 - travel data, business value, 294-295
 - trends, BI (business intelligence), 269
 - data mining, 270
 - RFID (Radio Frequency Identification), 271-272
 - triaging data, 65
 - tuning
 - databases, metadata management, 96
 - performance, 195
 - options, 196-197
 - reporting performance results, 197-198
 - selling management on, 198
- U**
- UD (unstructured data), 277-278, 290
 - central strategy, 282-287
 - current state in organizations, 282
 - emerging technologies, 287
 - DAM software, 287-288
 - DRM (Digital rights management) software, 288-290
 - EMR (electronic medical records) software, 290
 - focus on, 280-282
 - history, 278-280
 - unified content strategy, dealing with UD (unstructured data), 282-283
 - archiving UD, 284
 - combining structured and unstructured data, 287
 - content reusability, 286
 - retention, 285-286
 - search and delivery, 286-287
 - storage and administration, 283-284
 - uniquenesses
 - business entity quality rules, 51
 - validity rules, 56-57
 - United States, fair-value accounting, 4
 - Universal Meta Data Models*, 93
 - Universal Product Code (UPC), 271
 - University of California at Berkeley study, explosion of volume in data, 280
 - UPC (Universal Product Code), 271
 - usage
 - measurements for monitoring performance, 191
 - metadata, 82
 - enterprise quality disciplines, 63
 - sources, 84
 - segment, metadata repositories, 265
 - standards, security, 214
 - user-friendly structures, influence on physical data model, 129
 - users
 - expectations, performance, 189-190
 - role-based access matrix, 207
 - satisfaction, measurements for monitoring performance, 192

V

- validity
 - integration risks, 41
 - quality rules, 55-57
- value
 - business data, 291-292
 - call centers, 292
 - channel preferences, 293-294
 - click-stream data, 293
 - demographics, 293
 - direct retailers, 294
 - internal customer
 - information, 292
 - loyalty cards, 294
 - selling customer data, 292
 - travel data, 294-295
- strategic goals
 - benefits metric measurement, 301-308
 - development costs, 295-300
 - ROI (return on investment), 295
- vendors
 - DBMSs, evaluation, 246-255
 - security
 - external data, 216
 - software, 215-216
- verification, results, benchmarks (capacity planning), 175
- versioning, metadata management, 96
- vertical partitioning, 121
- View Updating rule (12 rules of relational databases), 104
- vision
 - data strategy, 4
 - organizations data, 5-6
- visualization, BI (business intelligence), 267

W

- Wall Street, data value to, 4
- websites, click-stream data, 293
- wisdom, CMM (capability maturity model), 62
- word processing files, MME source, 94

X-Y-Z

- XML-enabled metadata repositories, 87

- Y2K, data lineage, 38
- yield, 309-311
- Yourdon, Ed, 100

- zero-to-one optionality, 53
- zero-to-zero optionality, 53



























