## Index

### A
- Accessing data, 345  
  paths, 345
- Ad hoc queries, 245
- Adelman, Sid (biographical sketch), 391
- Aggregating data values  
  definition, 345  
  rolled up, 363
- Analytics applications, 345
- Analyzing  
  campaigns, 348  
  data  
    analytics applications, 345  
    data mining, 351  
    definition, 345, 351  
    gap analysis, 356  
    users’ fear of multidimensional analysis tools, 121–123  
    project costs/benefits, 350  
    justification of projects, 357
- Anomalies, 346
- Application service providers (ASPs)  
  definition, 346  
  outsourcing data warehouse activities, 171–173
- Architecture, data warehouses, 209  
  backdated transactions, effect on values, 318–319  
  basics, 303–305  
  blurred visions, 37–39  
  checklist, 215  
  definition, 346, 351  
  developing simultaneously with operational system, 321–325  
  inadequate architecture, 305–309  
  meta data integration across multiple products, 326–330  
  problems of click-stream data warehouses, 319–320  
  reconciling UPC code changes, 330–332  
  role of reporting system, 325–326  
  stovepipes impeding integration, 309–317  
  time-variant analysis, 320–321
- Atomic data, 346
- Attributes of entities, 346
- Availability of system, 346

### B
- Back-end processes, 346
- Base tables  
  definition, 346  
  security, 242
- Batch windows, 347
- Best of breed (tools), 347
- Best practices  
  definition, 347  
  preventing data warehouse failure, 8  
  Beta release, 347
- BI (business intelligence)  
  definition, 347  
  federated architecture, 214  
  standards for data warehouse planning, 212–215
- Bischoff, Joyce (biographical sketch), 391–392
- Brinkman Method, 140
- Budgeting/justification of data warehouse projects  
  basics, 51–53  
  cost allocations, 62–66  
  definition, 51, 357
Budgeting/justification of data warehouse projects continued
historical data, 66–69
identifying infrastructure benefits, 57–59
need by retailers, 59–62
prototype money unavailable, 69–73
substantiation of estimates required, 172–180

Business intelligence (BI)
definition, 347
federated architecture, 214
standards for data warehouse planning, 212–215

Business lines
definition, 357
effectiveness of building data warehouses, 88–93

Business process engineering, 347

Business rules
data warehouse design, 32
definition, 348

Business sponsors
definition, 348
prioritizing data for data warehouses, 112–116
real-time data warehouse updates, 289–292

Business timestamps, 348

Business users. See Users, data warehouses

C
Caching reports, 348
Campaign analysis, 348

Capability Maturity Model (CMM)
chain of command for data warehouse team reporting, 81
definition, 348
recognition of data quality value, 265

CASE Data Interchange Format (CDIF)
definition, 348
integrating meta data across multiple products, 328

CASE tools, 349
CDIF. See CASE Data Interchange Format (CDIF)

CEO (chief executive officer)
chain of command for reporting, 79, 80
data warehouses
role in building, 10
role in data sharing, 19, 21, 22
role in encouraging data, 19
definition, 349

Certified Information Systems Security Professional (CISSP) certification, 247
CFO (chief financial officer)
chain of command for reporting, 79
data warehouses
needs of retailers, 59–62
project cost justification, 51
role in data sharing, 22
definition, 349
Champions, 349
Channels, products, 349
Charge backs
definition, 349
importance, 64
Check totals, 349
CIF (corporate information factory), 350
CIO (chief information officer)
chain of command for reporting, 78–79, 81
data warehouses
decisions required when consultants are running show, 153–157
decisions required when consultants offer to solve problems, 150–153
ensuring success, 9, 10
needs of retailers, 59–62
role in data sharing, 20–22
undermining success, 14–16
definition, 349
CKO (chief knowledge officer)
chain of command for reporting, 80
definition, 349
Classes (objects), 349, 359
instances, 356
Click streams, 349
problems of click-stream data warehouses, 319–320
CMM (Capability Maturity Model)
chain of command for data warehouse team reporting, 81
definition, 348
recognition of data quality value, 265
Communication Profile (DISC), 140
Computer Aided Software Engineering. See CASE
Conformed dimensions, 349
Consultants and contractors
challenges in training team members, 161–165
definition, 349, 350
effective use, 165–171
infringing on team responsibilities, 153–157
lack of warehouse support by business,
116–119
leaving poor data warehouse behind, 157–161
offering to revamp outdated data warehouse,
150–153
Control totals, 350
COO (chief operating officer)
chain of command for reporting, 79
data warehouses
role in data sharing; 22
definition, 350
Corporate information factory (CIF), 350
Cost/benefit analysis
definition, 350
justification of projects, 51, 357
CREATE TABLE command, 345
Critical success factors, 350
CRM (customer relationship management)
definition, 350
designing click-stream data warehouse,
319–320
outsourcing data warehouse activities, 172–173
Cross-organizational, 350
Cross-selling, 350
CTO (chief technology officer)
data warehouse and operational systems inte-
gration, 287–289
data warehouse needs, 61
definition, 350
Customer relationship management (CRM)
definition, 350
designing click-stream data warehouse, 319–320
outsourcing data warehouse activities, 172–173
Customer segmentation, 350
"Cybersecurity Today and Tomorrow: Pay Now
or Pay Later," 247

D
Data. See Atomic data; Historical data; Meta data;
real-time data
Data access, 345
access paths, 345
Data administrators (DAs), 351
Data analysis
analytics applications, 345
data mining, 351
definition, 345, 351
Data definition language (DDL)
challenges of changes
operational system, 30
source systems, 33
definition, 351
Data derivations, 353
Data dimensions, 353
Data integration, 356
Data loading, 351
Data mapping, 351
Data marts
architecture, 312
definition, 351
effectiveness of single department control,
88–93
Data mining
data warehouses standards, 210–212
definition, 351
Data models
logical, 357
subject oriented, 35
Data owners, 351
Data quality
basics, 249–251
data-cleansing tools, 254
definition, 352
eliminating redundant data, 256–259
estimating dirty data, 259–263
ETL process partial failure, 269–272
obsession with quality by data warehouse
architect, 266–269
sampling, 251–256
source data errors causing massive updates,
272–273
value not recognized by management, 263–266
Data stewardship, 352
Data warehouse projects
3 Ss (scope, staff, schedule), 187
budgeting and justification, 357
basics, 51–53
cost allocations, 62–66
historical data, 66–69
identifying infrastructure benefits, 57–59
need by retailers, 59–62
prototype money unavailable, 69–73
substantiation of estimates, 177–180
user productivity justification not allowed,
53–57
Data warehouse projects continued
business sponsor rotation, 85–88
deadlines
  definition, 352
  unrealistic, 180–184
deliverables, 353
design reviews, 353
effectiveness of business lines building warehouses, 88–93
goals and objectives, 28–29
iterations, 357
iterative development, 357
lack of business support, 116–119
management offices, 361
milestones, 358
objectives, 359
PERT charts, 360
project and scope agreements, 176, 184–188, 361
project planning, 176
  scheduling, 177
  underestimating, 192–195
sign-off, 364
stakeholders, 365
team members, 77
  chain of command for reporting, 78–81
  consultants and contractors, 133–134
  consultants and contractors, challenges in training team members, 161–165
  consultants and contractors, effective use, 165–171
  consultants and contractors, infringing on team responsibilities, 153–157
  consultants and contractors, leaving poor data warehouse behind, 157–161
  consultants and contractors, offering to revamp outdated data warehouse, 150–153
dishonesty by members in making progress reports, 147–150
dissension, caused by prima donnas, 144–146
dissension, caused by single member, 134–137
dysfunctional members, 137–141
FAQs (frequently asked questions), 76
  goals, 132–133
  help desk staffs, 76–77
  incentives, 133
ineffectiveness of multiple DBAs assigned to project, 81–85
issues, 131–132
lack of dedicated staff, 93–99
management wanting to outsource warehouse activities, 171–173
requirement of consensus by all members, 141–144
responsibilities, 175
unrealistic expectations by users, 188–192
variables, 28
  blurred vision, 37–39
  changing of operational systems, 29–32
  changing of source systems, 32–37
  management not recognizing success, 48–450
  objectives improperly identified leaving no method for evaluation, 40–43
  prototypes become production too early, 43–48
Data warehouses
  architecture, 209
  backdated transactions, effect on values, 318–319
  basics, 303–305
  blurred visions, 37–39
  checklist, 215
  definition, 346, 351, 352
  developing simultaneously with operational system, 321–325
  inadequate architecture, 305–309
  meta data integration across multiple products, 326–330
  problems of click-stream data warehouses, 319–320
  reconciling UPC code changes, 330–332
  role of reporting system, 325–326
  stovepipes impeding integration, 309–317
time-variant analysis, 320–321
business sponsors, 348
  rotating sponsors for warehousing projects, 85–88
  champions, 349
  consultants, 349
data loading, 351
data staging, 352
  definition, 352
executive sponsors, 75–76
integrated with operational data/data sources
  avoiding building of stovepipe systems, 292–296
business sponsor wants real-time updates, 289–292
centralization or decentralization, 287–289
delay by enterprise model development, 281–286
reports from warehouse and operational systems don’t match, 297–299
responsibility for fixing inadequate operational systems, 299–301
from separate management systems, 276–281
organization’s lack of readiness, 104–108
performance basics, 333–334
fact table loading, 340–343
software not performing properly, 334–335
warehouse growth exceeds source data growth, 336–340
security basics, 239–240
designating responsible person, 243–245
information resources, 245–247
security plan, implementing, 245–247
security plan, lack of, 241–243
standards basics, 199–200
business intelligence environment planning, 212–215
companies with no previous methodology experience, 200–202
data mining, 210–212
DBAs wanting OLTP rather than OLAP design, 203–205
employees misusing data warehouse terminology, 206–210
tie and foot validation, 366
tools based on OO (object-oriented) environment, 223–225
basics, 217–218
performance criteria, 228–231
rejected vendors not taking no for answer, 233–235
RFPs (requests for proposals), 218–220
selection by IT rather than committee, 225–228
selection influenced unduly by vendors, 231–233
users dissatisfaction with tools, 220–223
vendors’s acquiring company providing poor support, 235–237
VLDBs (very large databases), 218
virtual enterprise data warehouse, 367
visions/visionaries, 367
blurred vision, 37–39
Data warehouses and management best practices to prevent failure, 8
criteria for project success, 5
definition, 352, 361
dissension between management positions, 99–102
documentation of project’s definition, scope and requirements, 13
inability to recruit employees, 102–104
ineffectiveness of multiple DBAs assigned to warehousing project, 81–85
IT role failure in its responsibilities, 9–11
lack of understanding of projects, 3–4
undermining project success, 14–16
lack of understanding of projects/details, 3–4, 22–26
management’s constant change, 11–14
organization’s lack of readiness for project, 104–108
QA expectations, unrealistic for pilot, 16–19
reluctance to accept warehouses fear of sharing data, 4, 19–22
previous failures, 5–9
Database administrators (DBAs) chain of command for reporting, 77
definition, 352
ineffectiveness of multiple DBAs assigned to warehousing project, 81–85
Database design schemas. See Physical data model
Database management systems (DBMSs), 352
Databases multidimensional, 358
very large databases, 367
DBAs (database administrators) chain of command for reporting, 77
ineffectiveness of multiple DBAs assigned to warehousing project, 81–85
using OLTP instead of OLAP for data warehouses, 202–205
DBMSs (database management systems), 352
DDL (data definition language) challenges of changes operational system, 30 source system, 33 definition, 351

Deadlines definition, 352 unrealistic, 180–184

Decision support systems (DSSs), 353, 354 Deliverables, projects, 353 Delta, 353 Demos, 353 Denormalization of data, 353 Departmental systems, 353 Derivations, data, 353 Derived data, 353 Design reviews, projects, 353 Dimensional hierarchy, 354 Dimensions, 353 Direct access storage device (DASD), 354 Dirty data, 249 estimating, 259–263 DISC (Communication Profile), 140 Domains, 354 Dotted-line (responsibility), 354 DSADs (direct access storage devices), 354 DSSs (decision support systems), 353, 354 Dyché, Jill (biographical sketch), 392


Fact tables, 355 granularity, 356 performance of data warehouses, 340–343 star schemas, 365 FAQs (frequently asked questions) definition, 355 help desk staff, 76 Federated database systems data flow, 310 definition, 355
Index

File transfer programs (FTP), 355
Foreign keys, 355
  referential integrity, 362
Frequently asked questions (FAQs)
  definition, 355
  help desk staff, 76
Front-end processes, 356
FTP (file transfer programs), 355

G
Gap analysis, 356
Global 2000, 356
Goals, data warehouse projects, 356
  misunderstandings, 40–43
  team members, 132–133
Granularity, fact tables, 356
Guidelines, 356

H
Hackney, Douglas (biographical sketch), 392–393
Hard-dollar benefits, 356
Help desk staffs, 76–77
Hierarchy, dimensional, 354
Historical data
  data warehouses
    architecture, 304
    justification, 66–69
    definition, 356

I
I/Os (inputs/outputs), 356
Information centers, 356
Information Systems Audit and Control
  Association Web site, 246
Information technology (IT)
  attitude required for project success, 3
  data warehouses
    lack of responsiveness for resources and
    information, 9–11
    definition, 356
    undermining project success, 14–16
  definition, 356
  iterations, projects, 357
  iterative development, 357
Ivoghli, Sean (biographical sketch), 393

J
Joining tables, 357
Justification of data warehouse projects, 51, 357
  basics, 51–53
  cases where user productivity justification not
  allowed, 53–57
  cost allocations, 62–66
  historical data, 66–69
  identifying infrastructure benefits, 57–59
  need by retailers, 59–62
  prototype money unavailable, 69–73
  substantiation of estimates required, 172–180

K
Kelley, Chuck (biographical sketch), 393–394
Knowledge transfers
  definition, 357
  inability of consultants to transfer knowledge
to team members, 161–165

L
Legacy systems, 357
Libraries of queries and reports, 175, 357
Lines of business, 357
Logical data models, 357
Logical database design, 360

M
Management and data warehouses
  best practices to prevent failure, 8
  criteria for project success, 5
  definition, 352, 361
  dissension among management positions,
  99–102
  documentation of project’s definition, scope
  and requirements, 13
Management and data warehouses continued

- inability to recruit employees, 102–104
- IT role
  - failure in its responsibilities, 9–11
  - lack of understanding of projects, 3–4
  - undermining project success, 14–16
  - lack of understanding of projects/details, 3–4, 22–26
  - management’s constant change, 11–14
  - multiple DBAs assigned to projects, ineffec-
    tiveness, 81–85
  - organization’s lack of readiness for project,
    104–108
  - QA expectations, unrealistic for pilot, 16–19
  - reluctance to accept warehouses
  - fear of sharing data, 4, 19–22
  - previous failures, 5–9

-Mapping data, 351
-Marko, David (biographical sketch), 394–395
-Market penetration, 358
-Massively parallel processing (MPP), 358
-Matrix management, 358
-MBTI (Myers-Briggs Type Indicator), 140
-Measurements
  - metrics, 358
  - success, 52
-Mentors, 358
-Meta data, 175, 199, 358
  - CASE Data Interchange Format (CDIF), 348
  - data warehouse architecture, 304
  - eliminating redundant data, 257, 258
  - integration across multiple products, 326–330
-Meta Group Web site, 246
-Methodologies, data warehouse standards,
  200–201
  - definition, 358
-Project Management Methodology (PMM), 202
-Metrics, 358
-Milestones, 358
-MIS Training Institute, 246
-Missions, 358
-Moss, Larissa T. (biographical sketch), 395
-MPP (massively parallel processing), 358
-Multidimensional databases, 358
-Myers-Briggs Type Indicator (MBTI), 140

 Networks of systems, 359
 Normalization of data
  - denormalization, 353
  - third normal form, 366

-Object-oriented (OO) environment, 359
-Objectives, data warehouse projects, 28, 359
  - misunderstandings, 40–43
  - instances, 356
-OCM (organizational change management)
  - definition, 359–360
  - justification of projects, 56–57
  - outsourcing data warehouse activities, 173
-ODSs (operational data stores), 359
  - data warehouse and operational systems inte-
    gration, 291
-OLAP (online analytical processing)
  - definition, 359
  - OLTP standards inappropriate for data ware-
    houses, 202–210
-OLTP (online transaction processing), 359
  - operational system changes, 30
  - role with IT in building successful data ware-
    house
  - reasons for lack of cooperation, 10
  - ways to solve problems, 11
  - standards inappropriate for data warehouses,
    202–210
-Online analytical processing (OLAP), 359
-Online transaction processing (OLTP), 359
-OO (object-oriented) environment, 359
  - data warehouses tools, 223–225
-Operational data/data sources
  - challenges to data warehouse projects
  - source changes, 32–37
  - system changes after project’s beginning,
    29–32
  - definition, 359
-ETL processes, 355
  - role in building successful data warehouses, 9
  - integrating with data warehouses
  - avoiding building of stovepipe systems,
    292–296
  - business sponsor wants real-time updates,
    289–292
  - centralization or decentralization, 287–289

-Near-line storage, 358
-Networking (people), 359
Index

- data flow, 317
- delay by enterprise model development, 281–286
- reports from warehouse and operational systems don’t match, 297–299
- responsibility for fixing inadequate operational systems, 299–301
- from separate management systems, 276–281
- stovepipes impeding integration, 309–317
- Operational data stores (ODSs), 209, 359
- data warehouse and operational systems integration, 291
- Organizational change management (OCM) definition, 359–360
- justification of projects, 56–57
- Outsourcing definition, 360
- management wanting to outsource warehouse activities, 171–173
- Ownership, data warehouses, 360
- total cost, 367

P
- Pain, 360
- Parallelism, 360
- Partitioning tables, 360
- Performance of data warehouses basics, 333–334
- slow fact table loading, 340–343
- software not performing properly, 334–335
- warehouse growth exceeds source data growth, 336–340
- Periodicity, 360
- Persistent storage, base tables, 345
- Personal Interests, Attitudes, and Values inventory (PIAV), 140
- PERT (Program Evaluation and Review Technique) charts definition, 360
- dishonesty in progress reports, 147
- Phasing of data warehouse delivery, 360
- Physical data model, 360
- PIAV (Personal Interests, Attitudes, and Values) inventory, 140
- Pilot data warehouses definition, 361
- unrealistic QA expectations, 16–19
- Platforms, 361
- Political agendas, 361
- Primary keys, 361
- referential integrity, 362
- Products channels, 349
cross-selling, 350
integrating meta data across multiple products, 326–330
suites, 365
UPCs (universal product codes) definition, 367
reconciling changes, 330–332
Program Evaluation and Review Technique (PERT) charts definition, 360
dishonesty in progress reports, 147
Project Management Methodology (PMM), 202
Proof-of-concept, 361–362
unrealistic QA expectations for pilot data warehouses, 16–19
Prototypes, data warehouses cost justification, 69–73
definition, 362
delays, 29
unable to meet production standards, 43–48
users inability to articulate needs, 128–129

Q
- QA (quality assurance) definition, 362
unrealistic expectations for pilot data warehouses, 16–19
users’ expectations of high data quality, 123–127
Queries ad hoc, 245
libraries, 175, 357
tables foreign keys, 355
joining, 357
tools, 220–223

R
- Rapid application development (RAD), 115
definition, 362
RDBMS (relational database management systems) challenges of changes operational system, 30
source system, 33
definition, 362
Real-time data
  data warehouse updates, 289–292
definition, 362
Recursive, 362
Referential integrity, 362
Rehm, Clay (biographical sketch), 395–396
Relational database management systems (RDBMS)
  challenges of changes
    operational system, 30
    source system, 33
definition, 362
Relational databases
  base tables, 345
  RDBMS, 362
Release concepts, 363
Reports
  caching, 348
  data warehouse architecture, 313, 325–326
  libraries, 357
  team members, data warehouse projects
    chain of command, 78–81
    dishonesty in making progress reports, 147–150
    tools, 220–223
Return on investment (ROI), 5, 363
RFPs (requests for proposals), 363
data warehouse tools, 218–220
ROI (return on investment), 5, 52, 363
Rolled up values, 363

S
Scalability
  data warehouse architecture, 314
definition, 363
Scope, 363
Scope creep, 27, 188, 363
Scope gallop, 27
Security for data warehouses
  basics, 239–240
  designating responsible person, 243–245
  information resources, 245–247
  security plan
    implementing, 245–247
    lack of plan, 241–243
Semantic layers, 363
Service-level agreement (SLA), 364
Shelfware, 364
Sign-off, 364
Silos and siloizing, 364
Single versions of the truth, 364
SLA (service-level agreement), 364
SMP (symmetrical multiprocessing), 366
Snowflake structure, 364
SOP (standard operating procedure), 365
Source data, 364
targets, 366
tie and foot validation, 366
Source systems
  constant changes, 32–37
data warehouse design, 32
definition, 364
Sponsors, 364
  business
definition, 348
  prioritizing data for data warehouses, 112–116
  real-time data warehouse updates, 289–292
  executive, 75–76
Staging areas, 365
Stakeholders, 365
Standard operating procedure (SOP), 365
Star schemas
  data warehouse and operational systems integration, 277–278
definition, 365
Stovepipes
definition, 365
  impeding data warehouses integration, 309–317
  integrating data warehouses with operational data/data sources, 292–296
Strategic data warehouse planning, 365
  blurred visions, 37–39
Subject areas, 365
Suite of products, 365
Supply chains, 365
SWAT teams, 365
Symmetrical multiprocessing (SMP), 366
System Security Certified Practitioner (SSCP) certification, 247
System timestamps, 366
Systems Administration, Networking and Security Institute Web site, 246
Systems integration, 366
**Index**

**T**
- Tables. See also Base tables; Fact tables
  - foreign keys, 355
  - joining, 357
  - partitioning, 360
  - primary keys, 361
  - referential integrity, 362
- Tactical approaches
  - blurred data warehouse visions, 37–39
  - definition, 366
- Targets, 366
- Team members, data warehouse projects
  - chain of command for reporting, 78–81
  - consultants and contractors, 133–134
  - challenges in training team members, 161–165
  - effective use, 165–171
  - infringing on team responsibilities, 153–157
  - offering to revamp outdated data warehouse, 150–153
- DBAs (database managers), 77
- dishonesty by members in status reporting, 147–150
- dissension
  - caused by prima donnas, 144–146
  - caused by single member, 134–137
- dysfunctional members, 137–141
- goals, 132–133
- help desk staff, 76
- incentives, 133
- issues, 131–132
- lack of dedicated staff, 93–99
- management wanting to outsource warehouse activities, 171–173
- requirement of consensus by all members, 141–144
- responsibilities, 175
- Terabytes, 366
- Third normal form, 366
- Tie and foot validation, 366
- Time dimensions, 366
- Time variances
  - analysis of data warehouses, 320–321
  - definition, 366
- Timeliness, 366
- Timestamps
  - business, 348
  - system, 366
- Tokenizing of sensitive material, 241
- Topology, 366
- Total cost of ownership, 367
- TQM (total quality management), recognition of data quality value, 265
- Triage, 367
- Trickle feed, 367
- Truth, single versions, 364

**U**
- UPCs (universal product codes)
  - definition, 367
  - reconciling changes, 330–332
- Users, data warehouses
  - definition, 348, 367
  - dissatisfaction with tools, 220–223
  - fear of data sharing, 4, 19–22
  - issues, 109–111
  - complexity of multidimensional analysis tools, 121–123
  - expectations of high data quality, 123–127
  - inability to articulate needs, 127–130
  - lack of business support for projects, 116–119
  - lack of support for Web-based application, 119–121
  - prioritizing data for data warehouses, 112–116
  - questionnaires on opinions, 23–24
  - semantic layers, 364
  - use cases, warehouse design, 32
  - warehousing projects
    - commitment to ensure staff dedication, 94
    - problems with rotating sponsors, 85–88

**V**
- Value added, 367
- Very large databases (VLDBs), 209
  - data warehouse and operational systems integration, 291
  - data warehouse tools, 218
  - definition, 367
- Virtual enterprise data warehouses, 367
- Visionaries, 367
- Visions for data warehouses
  - blurred visions, 37–39
  - definition, 367
Visualization, 368
VLDNs (very large databases), 209
- data warehouse and operational systems integration, 291
- data warehouse tools, 218
- definition, 367

W
- WBS (work breakdown structure), 368
- Web-based applications, lack of support by users, 119–121
- Web sites, click streams, 349
- Work breakdown structure (WBS), 368
- Workloads, 368