

Index

Note: Page numbers with “f” and “t” denote figures and tables, respectively.

A

- Access services, 312
- Accidental architecture. *See also* Analytical data architecture (ADA); Information architecture; Technical architecture.
 - addiction signs, 82
 - avoiding accidents, 81–83
 - BI, 115–116
 - recovering from, 83
- Accumulating fact tables, 218–219
- ACID. *See* Atomicity, consistency, isolation, and durability
- Ad hoc analysis, 150, 406
- ADA. *See* Analytical data architecture
- Additive facts, 202–203
- Adopters, 312
- Advanced analytics, 375b. *See also* Big Data analytics; Predictive analytics.
 - in action, 376–377
 - analytical hubs, 384–387
 - analytical sandboxes, 383–384
 - architecture, 381f
 - data mining, 377–383
 - data visualization, 401–402
 - overview and background, 375–377
 - past and predictive analytics, 376
 - predictive analytics, 377–383
 - roadblocks to success, 382–383
 - skills, 382t, 399t
 - techniques and examples, 381t
 - window to future, 376
- Advanced dimensional modeling, 237. *See also* Dimensional modeling.
 - approaches, 270
 - custom dimension groups, 271–272, 272f
 - hot swappable dimension, 270–271, 271f
 - causal dimension, 262–263, 263f
 - heterogeneous products, 269–270
 - hierarchies, 237, 238f
 - balanced, 238–240
 - ragged, 240–241, 240f
 - unbalanced, 241, 241f
 - variable-depth, 241–244
 - junk dimension, 265–267
 - multivalued dimension, 263–265
 - outrigger tables, 244, 245f
 - SCD, 245–262
 - too few dimensions, 272
 - too many dimensions, 272
 - value band reporting, 268–269, 268f
- Aggregation subservice, 323–324, 324f
- Agile methodology, 475–476
- AJAX. *See* Asynchronous JavaScript and XML
- Alternate key (AK), 187
- Analysts, 341–342
- Analytical data architecture (ADA), 120, 123, 124f.
 - See also* Information architecture; Technical architecture.
 - BI sources, 127
 - data marts, 125
 - data schemas, 125–126
 - EDW, 125
 - enhanced hub-and-spoke model, 124–125
 - physical data store combinations, 128
- Analytical hubs, 384–387, 386f
 - advice for, 394–395
 - architecture options, 388, 389f
 - advanced analytics layers, 388–391
 - business analytics, 388–391
 - data access and integration, 393–394
 - platforms, 391–392
 - predictive modeling with, 392–393
 - design principles, 387–388
- Analytical sandboxes, 383–384, 385f
 - advice for, 394–395
 - architecture options, 388, 390f
 - advanced analytics layers, 388–391
 - business analytics, 388–391
 - data access and integration, 393–394
 - platforms, 391–392
 - design principles, 387–388
- Analytics, 225
 - challenges, 7–8
 - deluge, 6–8
 - importance, 7
 - strategy, 8
- API. *See* Application programming interface
- Application development methodology. *See* Systems development life cycle (SDLC)
- Application programming interface (API), 109
- Application vendor, 101
- Architecture framework
 - accidental architecture, 81
 - addiction signs, 82
 - recovering from, 83
 - action plan, 84t
 - architectural blueprints, 65–66
 - BI framework, 66, 67f

Architecture framework (*Continued*)

- data architecture, 68–72
- information architecture, 67–68
- metadata management, 78–80
- privacy, 80–81
- product architecture, 78
- security, 80–81
- technical architecture, 72–78

“as is” analysis, 247, 249, 260

“as was” analysis, 247, 249, 260

Asynchronous JavaScript and XML (AJAX), 360

Atomic level, 220

Atomicity, consistency, isolation, and durability (ACID), 163

Audit services, 329

- data integration jobs, 329
- data quality metrics, 330
- error handling services, 330–331
- table and row updates, 330

Auditable data, 277

B

Balanced hierarchy, 238–240

Banded-value approach, 255

BASE. *See* Basically available, soft state, eventual consistency

Basically available, soft state, eventual consistency (BASE), 163

Battling project methodologies, 474

- agile methodology, 475–476
- waterfall methodology, 474–475

BI. *See* Business intelligence

BI COE. *See* Business Intelligence Centers of excellence

BI Competency Center (BICC), 493

Big Data, 3, 147, 163–164, 500

- database, 4
- projects, 396, 398
- refined, 149

Big Data analytics, 151, 395. *See also* Advanced analytics; Predictive analytics.

- Big Data team, 398–400
- hybrid architecture, 397–398, 398t
- program, 397
- scope, 396–397
- structured vs. unstructured data, 395t
- worst practices, 400–401

Bogged-down feedback loop, 428

“Bottom-up” approach, 451–452

Bridge table, 243–244, 243f

- for multivalued dimensions, 264–265
- navigation, 244f

Bring your own device (BYOD), 113

Bring your own results (BYOR), 113

BSA. *See* Business systems analyst

Bulk loading utilities, 312

Business

- analyst, 439–440
- analytics, 8
- benefits documentation, 27, 28f
- case, 24–28
- initiatives, 25
- metadata, 79
- processes, 25–27
- sponsorship, 25–26
- transformation services, 321
- value determination, 27–28

Business intelligence (BI), 3, 143, 375

accidental architecture, 115–116

“in actionable” information, 10

analysis phase, 463–465

analytics, 149–151, 149f

appliances, 164–165

architectural plan, 470–472

assessment, 460–465, 460f

balancing, 457–459

building, 455–457

business involvement, 454f

business participation, 455

business sponsors, 452

COE, 433

content specifications, 337–339

data capture vs. information analysis, 10

data warehousing, 11–12

operational BI, 11

operational system, 11

data design for self-service, 348

inconsistency, 349–350

last data preparation step, 348–349

OLAP cubes and in-memory columnar databases, 350–351

data integration activities, 301

data modeling use, 173

design, 359

BI application, 362–367

privacy, security and access standards, 361–362

UI standards, 359–361

visual design methods, 362f

Web and BI development challenges, 360t

design layout, 343–348

dashboard example, 337–339, 347f

focus on purpose, 343

development, 367

application development tasks, 370–372, 371f

BI application testing, 372–374, 373f

prototyping lifecycle, 369–370, 369f

discovery phase, 461–463

distribution analysis, 357

- DW and DI, 15f
 - enterprise and BI applications, 472t
 - enterprise business strategy, 451–452
 - and business initiatives, 450f
 - and plans drive business intelligence program, 451f
 - establishment, 450–460
 - evolution, 144
 - data management, 146–147
 - enterprise applications, 145–146
 - technology platforms, 145
 - internal marketing, 459–460
 - matching types of analysis to visualization, 351, 352f
 - comparative analysis, 351
 - contribution analysis, 354
 - correlation analysis, 352
 - geographic data, 337–339s
 - time-series or trending analysis, 351–352
 - personas, 340
 - analysts, 341–342
 - casual consumers, 341
 - data scientists, 342–343
 - power users, 342
 - portfolio, 452–453
 - program
 - composition, 453f
 - governance, 454–455
 - project, 472–473
 - management, 449
 - schedule, 484–492
 - project phases, 479–484, 479f
 - analysis and definition phase, 481–482
 - architect and design phase, 482
 - build, test, and refine phase, 482–483
 - deploy and roll-out phase, 484
 - implementation phase, 483
 - scope and plan phase, 480–481
 - recommendations phase, 465
 - revising BI applications list, 339–340
 - solution, 43
 - sources, 127
 - sponsorship, 454
 - swim lane, 487–488
 - targets, 152–153
 - technology and terms, 14–19
 - trade-offs, 457f
 - working committee, 433
 - Business Intelligence Centers of excellence (BI COE), 493–501. *See also* Centers of excellence (COEs); Data Integration Centers of excellence (DI COE).
 - building business case, 496
 - alignment of resources improvement, 498
 - business and IT objectives, 497
 - cost savings, 496–497
 - monitor BI performance, 497–498
 - value of BI, 496
 - business needs, 494
 - business preferences, 495
 - organizational structures drives reporting silos, 494–495
 - deliverables, 498–499
 - interactions, 499f
 - organization and funding, 501
 - skills, 499–500
 - Business requirements, 46, 49–51, 282–284
 - documenting requirements, 60–61
 - interviewing, 56
 - conducting interviews, 58
 - interview follow-ups, 59–60
 - preparation for, 57–58
 - reviewing interview content, 58–59
 - primary deliverable, 45–47
 - primary goals, 44–45
 - purpose of, 43–44
 - roles, 47
 - BI team participants, 47–48
 - business participants, 48
 - IT participants, 49
 - people vs., 48b
 - stepwise refinement, 47f
 - workflow, 49, 50f
 - compliance requirements, 53
 - data quality requirements, 51–52
 - data requirements, 51–52
 - functional requirements, 52–53
 - IT all together, 55
 - prioritizing requirements, 55–56
 - regulatory requirements, 53
 - reverse engineering, 54–55
 - technical requirements, 54
 - Business systems analyst (BSA), 484
 - BYOD. *See* Bring your own device
 - BYOR. *See* Bring your own results
- ## C
- Calendar dimensions. *See* Date dimensions
 - Cardinality, 181, 182t
 - Cascading style sheets (CSS), 360
 - Causal dimension, 262–263, 263f
 - CDC. *See* Change data capture
 - CDI. *See* Customer data integration
 - CDW. *See* Central data warehouse
 - Centers of excellence (COEs), 493
 - data-driven enterprise, 511–512
 - DI COE, 501–511
 - purpose of, 493
 - Central data warehouse (CDW), 109

Change data capture (CDC), 287, 292, 312–316
 Chief information officer (CIO), 144, 451
 CIF. *See* Corporate information factory
 CIO. *See* Chief information officer
 Clean, consistent, conformed, current, and comprehensive (Five Cs), 12–14
 Clustered configuration, 162
 COEs. *See* Centers of excellence
 Column analysis, 319–320
 Column-family, 164
 Columnar databases, 160–161
 Comparative analysis, 351
 Compliance requirements, 46, 53
 Conceptual data model, 174–175
 Conforming dimensions, 220
 Conforming facts, 220–221
 Connectors, 312
 Consolidated fact tables, 233–234, 233f
 Contribution analysis, 352
 Corporate information factory (CIF), 120, 121f
 Corporate performance management (CPM), 139, 146
 Correlation analysis, 352, 354
 heat map, 355f
 scatter plot, 356f
 CPM. *See* Corporate performance management
 CRM. *See* Customer relationship management
 Cross-reference dimensions, 266b
 Cross-table analysis, 320
 CSS. *See* Cascading style sheets
 Cubes, 159
 CurrencyKey, 199
 Custom data loaders, 312
 Custom dimension groups, 271–272, 272f
 Customer data integration (CDI), 104, 157, 506
 Customer relationship management (CRM), 14, 68, 146, 502
 Customer resource management. *See* Customer relationship management (CRM)
 CustomerKey, 199

D

Dashboards, 150
 Data, 3–4
 architect, 437
 auditability story, 131b
 capture, 119, 167
 cleansing, 89–90
 conversion services, 320–321
 data capture vs. information analysis, 10
 BI operational system, 11
 data warehousing, 11–12
 operational BI, 11

deluge, 4–6
 discovery, 151
 federation, 154
 5 Cs, 12–14
 franchising, 91
 data preparation vs., 95
 franchising process, 133
 need for, 91–93
 process, 93–95, 94f
 gathering, 409
 information vs., 8–10
 mining, 377
 architecture for, 380
 setting up, 377–378
 techniques for, 381
 preparation, 88, 89f
 cleansing data, 90–91
 consolidating data, 90–91
 data franchising vs., 95
 definitional work, 90
 steps, 88–90
 profiling services, 317
 column analysis, 319–320
 cross-table analysis, 320
 primary key analysis, 320
 relationship analysis, 320
 source systems analysis, 318–320
 source-to-target mapping, 320
 table analysis, 319
 and quality, 32–33
 reporting and analysis, 410
 requirements, 46, 51–52
 restructuring process, 133
 schemas, 125–126
 scientists, 7, 377
 source system, 68
 requirements, 284–285
 structure transformation services, 322
 aggregation subservice, 323–324
 dimension table subservice, 322–323
 fact table subservice, 323, 323f
 too much data, too little information, 8–10
 too much information, 5f
 transformation services, 320, 410
 business transformation services, 321
 data conversion services, 320–321
 data integration workflow, 321f
 data structure transformation services, 322–324
 transport services, 332–333
 variety, 4–6
 velocity, 4–6

- virtualization, 154
- visualization, 359b, 401–402
- volume, 4–6
- workflow, 410f
- Data Analysis eXpressions (DAX), 153
- Data architecture, 68, 70f, 107–108, 469. *See also* Analytical data architecture (ADA); Information architecture; Technical architecture.
 - choices, 118
 - ADA, 123–128, 124f
 - data architecture selection, 119–122
 - data categories, 119
 - enterprise data bus architecture, 121f
 - Inmon vs. Kimball architecture comparison, 122t–126t
 - 3NF, 122–123
 - data integration workflow, 128–135
 - EDW, 136–137
 - data modeling vs., 107b
 - data systems, 71f
 - data warehousing, 69–72
 - EDW, 68–69, 69f
 - history
 - in beginning, 108–109
 - BI accidental architecture, 115–116
 - data mart, 110–112, 111f
 - DW evolution, 108–110
 - EDW, 110f
 - federated DWs, 114–115, 115f
 - hub-and-spoke, 116–117, 117f
 - multiple data marts, 112–113, 112f
 - ODS, 113–114, 113f
 - ODS, 137–142
- Data governance, 444–446
 - board, 446
 - business or IT, 445
 - organization, 446
 - task force, 446, 447t
 - on track, 447–448
 - working with COEs, 446, 447f
- Data ingestion services, 312
 - CDC, 313–316
 - audit columns, 313–314
 - database log or message queue scanners, 315
 - row difference comparisons, 315–316
 - table or event triggers, 315
 - timed extracts, 314–315
 - reference lookups, 316–317
 - SCD, 316, 316t
- Data integration (DI), 14, 275, 410
 - architecture, 277–280
 - design, 285
 - conceptual data integration process, 285, 286f
 - designing data, 289–290
 - logical data integration modeling, 286–287, 287f
 - physical model, 287–288, 288f
 - source to target mapping, 288–289
 - efficient method, 277
 - ETL, 301, 302f
 - holistic approach, 275–276
 - incremental approach, 276–277
 - iterative approach, 276–277
 - jobs, 329
 - loading historical data, 295–296
 - historical data, 297–298
 - same as old, 296–297
 - manual coding vs. tool-based, 301–309
 - hand code, 302–304, 304f
 - tool selection, 305–308
 - using tools, 304–305
 - two tools, 309
 - productive method, 277
 - prototyping, 298
 - requirements, 280
 - business, 282–284
 - data sourcing system, 284–285
 - designing data models, 282f
 - prerequisites, 280–282
 - services, 309
 - access and delivery services, 312
 - breakdown of suite–services, 311f
 - data ingestion services, 312–317
 - data profiling services, 317–320
 - data quality services, 324–328
 - data transformation services, 320–324
 - data transport services, 332–333
 - extract, transform, load, and manage, 310f
 - operations management services, 331–332
 - process management services, 328–331
 - suite–services, 311f
 - standards, 290
 - development project standards, 291–292
 - development standards, 293–295
 - reusable components, 292–293
 - software development mindset, 290–291
 - swim lane, 489–490
 - testing, 298–299
 - workflow, 128–129, 128f, 321f
 - EDW, 136–137
 - hub-and-spoke, 129, 129f
 - ODS, 137–142
 - SOA, 133–135
 - SOI, 130–133
- workflow, 278f–279f, 280t

- Data Integration Centers of excellence (DI COE), 493, 501. *See also* Business Intelligence Centers of excellence (BI COE).
- building business case, 504
 - data-integration investment portfolio, 505–507
 - enlist sponsorship, 504–505
- businesses need, 502
 - data sources and silos exploding, 502
 - data-integration problems, 503–504
 - myriad integration approaches, 502–503
 - silos, 503
- deliverables, 509–510
- organizational model, 507
 - centralized services, 509
 - services, 508–509
 - sharing best practices, 507–508
 - technology, 508
- skills, 510–511
- Data integration framework (DIF), 86, 505
 - building blocks, 87t
 - information architecture, 86–100, 87f
 - BI and analytics, 95–97
 - data franchising, 91–95
 - data management, 98, 99f
 - data preparation, 88–91
 - metadata management, 98–100
- Data management, 98, 99f. *See also* Master data management (MDM).
- Data mart, 110–112, 111f, 125
- Data Mining eXtensions (DMX), 153
- Data model, 173–174
 - data modeling vs., 173–174
 - entity vs. dimensional, 178f
 - hybrid dimensional-normalized, 126, 128–130, 135, 144
 - levels, 174, 174f
 - conceptual data model, 175
 - logical data model, 175–176
 - physical data model, 176–177
 - usage, 178–179
- Data modeling, 173–174
 - data architecture vs., 107b
 - data model vs., 173–174
 - ER modeling, 173, 179
 - attributes types, 180–181, 181f
 - building blocks, 179–180, 180f
 - cardinality, 181
 - entities types, 180–181, 181f
 - keys, 187–188, 187f
 - referential integrity, 188–189, 189f
 - types, 181–187
 - workflow, 177–178, 177f
- Data quality, 132
 - metrics, 330
 - requirements, 51–52
 - services, 324
 - building, 327–328
 - implementation, 328
 - misconceptions, 325–327
- Data shadow system, 54, 419f
 - benefits, 413–414
 - BI silos, 404f
 - damages, 412–413
 - evolution, 408
 - data workflow, 410f
 - filling in gap, 408–410
 - IT group response to, 411
 - rise and expansion, 411
 - misguided attempts to replace, 417–418
 - BI tools, 418
 - set of reports, 418
- moving beyond, 414
 - business and IT, 415–416
 - changing approach, 416–417
 - changing culture, 417
 - choices assessment, 416t
 - stopping blame game, 415
- organization, 405–406
- problem, 403–405
- renovation, 418–421
 - analytical process considerations, 421
 - balanced priorities, 421
 - considerations, 419–421
- triage, 407, 408t
 - mild and moderate, 407
 - serious, 407–408
 - very serious, 408
- types, 406, 407t
- Data source system. *See* Systems of record (SOR)
- Data warehousing (DW), 9, 24, 65, 144, 147, 308
 - applications, 173
 - BI and DI, 15f
 - industry terms, 16t–19t
 - technology and terms, 14–19
- Data-and-database-task, 488
- Data-driven enterprise, 511–512
- Data-integration investment portfolio, 505–507
- Database Administrator (DBA), 438
- Database log, 315
- Database management system (DBMS), 156–157
- Date dimension, 219, 221, 223f, 239f
 - benefits, 224–225
 - in BI, 222
 - designing, 222–224
 - time dimension vs., 228–229
- DateKey, 199
- DAX. *See* Data Analysis eXpressions
- DBA. *See* Database Administrator
- DBMS. *See* Database management system

Degenerative dimensions, 230–232, 231f
 Delivery services, 312
 Developer unit tests, 372
 DHTML. *See* Dynamic hypertext markup language
 DI. *See* Data integration
 DI COE. *See* Data Integration Centers of excellence
 DIF. *See* Data integration framework
 Dimension, 203, 204f
 hierarchy, 204–205, 205f
 keys, 205, 205f
 benefits, 208
 not null values, 207–208
 smart, 207
 surrogate, 206
 schemas, 208
 multidimensional, 211–212, 212f
 multifactor star models, 212–213, 212f
 snowflake, 210–211, 210f
 star, 208–209, 209f
 table, 265
 subservice, 322–323, 322f
 Dimensional modeling, 216. *See also* Advanced dimensional modeling; Entity-relationship modeling (ER modeling).
 achieving consistency, 220–221
 consolidated fact tables, 233–234, 233f
 date dimension, 221–225, 228–229
 degenerative dimensions, 230–232, 231f
 ER modeling vs., 213, 214f, 216f
 comparing approaches, 213–215
 structures, 215–216
 event tables, 232–233, 232f
 facts, 198–203
 types, 202–203
 fact tables, 218–219, 219t
 keys, 199–201, 199f
 measures, 201–202
 granularity, 202f
 high-level view, 198
 logical design technique, 197
 mapping
 to business report, 217, 217f
 to OLAP analysis, 217–218, 218f
 role playing dimensions, 229–230, 230f
 time dimension, 225–229
 Distribution analysis, 357
 DMX. *See* Data Mining eXtensions
 Documentation, 277
 Documenting requirements, 60–61
 DW. *See* Data warehousing
 Dynamic hypertext markup language (DHTML), 360

E

EAI. *See* Enterprise application integration
 EDW. *See* Enterprise data warehouse
 8/80 rule, 467
 EII. *See* Enterprise information integration
 ELT. *See* Extract, load &; transform
 EMS. *See* Enterprise message services
 End users, 380
 Enterprise
 application stack vendors, 166
 data evolution, 144f
 standard, 132
 Enterprise application integration (EAI), 12, 153–154, 503
 Enterprise data warehouse (EDW), 69f, 87–88, 110f, 148, 302, 338
 rise of, 68–69
 schemas, 125
 Enterprise information integration (EII), 12, 154, 303, 503, 506
 Enterprise message services (EMS), 154
 Enterprise resource planning (ERP), 14, 68, 123, 175, 472–473, 494
 Enterprise service bus (ESB), 82, 153–154, 312, 503
 Entity, 179
 Entity Relationship Diagrams (ERD), 437
 Entity-relationship modeling (ER modeling), 173, 178–179
 building blocks, 179–180, 180f
 cardinality, 181
 dimensional modeling vs., 213, 214f, 216f
 comparing approaches, 213–215
 structures, 215–216
 entities and attribute types, 180–181, 181f
 example, 185–187, 186f
 keys, 187–188, 187f
 referential integrity, 188–189, 189f
 types, 181
 identifying relationship, 182, 183f
 many-to-many relationships, 184, 184f
 nonidentifying mandatory relationship, 182–183, 183f
 nonidentifying optional relationship, 183–184, 184f
 recursive relationships, 185, 185f
 ERD. *See* Entity Relationship Diagrams
 ERP. *See* Enterprise resource planning
 Error handling services, 330–331
 ESB. *See* Enterprise service bus
 ETL. *See* Extract, transform, and load
 Event tables, 232–233, 232f. *See also* Fact tables.
 Extended project team, 440–441
 Extensible markup language (XML), 153, 360
 Extensible markup language for Analysis (XML/A), 153
 Extract, load & transform (ELT), 154, 156f, 393
 ETL vs., 155–157

Extract, transform, and load (ETL), 10, 66, 145
 architecture, 156f
 tools, 86, 301, 302f
 workflows, 82

F

Fact table, 264
 in dimensional modeling, 218–219, 219t
 granularity, 202f
 keys, 199–201, 199f
 measures, 201–202
 subservice, 323, 323f
 Facts, 198–203
 additive, 202–203
 nonadditive, 203
 semiadditive, 203
 Fifth normal form (5NF), 190
 First normal form (1NF), 190–191
 Fiscal calendar, 239f
 Five Cs. *See* Clean, consistent, conformed, current, and comprehensive
 Fixed hierarchy. *See* Balanced hierarchy
 Foreign key (FK), 188
 Fourth normal form (4NF), 190
 Full-fledged analytical application, 406
 Functional decomposition, 46

G

Garbage dimension. *See* Junk dimension
 Garbage in, garbage out (GIGO), 13, 382
 Goldilocks syndrome, 45
 “Green bar”, 74
 Greenwich Mean Time (GMT), 229

H

Health care, 252, 263
 Health Insurance Portability and Accountability Act (HIPAA), 53
 High-tech titans, 166
 Hot swappable dimension, 270–271, 271f
 Hub-and-spoke architecture, 116–117, 117f
 100-percent rule, 467
 Hybrid BI project methodology, 477–478
 Hybrid dimensional-normalized model, 126, 128–134, 135, 144
 Hybrid online analytical processing (HOLAP), 159
 Hypertext markup language (HTML), 360

I

ICC. *See* Integration Competency Center
 Identifying relationship, 182, 183f
 In-database analytics, 161–162
 In-memory
 columnar databases, 350–351
 databases, 162

Incremental approach, 276–277

Information

backbone, 10
 data capture vs. information analysis, 10
 BI operational system, 11
 data warehousing, 11–12
 data vs., 8–10

Information architecture, 67–68, 85–86, 469. *See also*
 Analytical data architecture (ADA); Data architecture;
 Information architecture; Technical architecture.

DIF, 86–100, 87f

BI and analytics, 95–97
 data franchising, 91–95
 data management, 98, 99f
 data preparation, 88–91
 metadata management, 98–100

MDM, 103–106

operational BI vs. analytical BI, 100
 application-specific environment, 101–102
 blend application-specific BI environment,
 102–103
 DW-based BI environment, 102–103

Information technology (IT), 144, 377

Infrastructure swim lane, 490–491

Integration Competency Center (ICC), 493

Integration testing, 299

Intellectual property (IP), 146–147

Internet of things (IoT), 4, 145

IP. *See* Intellectual property

IT. *See* Information technology

Iterative approach, 276–277

J

Java database connectivity (JDBC), 153, 312

JavaScript Object Notation (JSON), 154

JDBC. *See* Java database connectivity

Job control services, 328, 329f

JSON. *See* JavaScript Object Notation

Junk dimension, 265

alternatives, 266t
 example, 267f
 misguided attempts, 266
 recommended solution, 266–267

Justification process, 23

approval obtainment, 40

BI road map creation, 35

building business case, 24–28

business

benefits documentation, 27, 28f
 initiatives, 25
 processes, 25–27
 value determination, 27–28

pitfalls, 40–41

project

- budget, 37–38
- plan, 36–37
- scope, 36
- readiness assessment, 32
 - cultural change, 34
 - data and data quality, 32–33
 - expertise and experience, 33–34
 - financial commitment, 34
 - organization, 34
 - resource commitment, 34
- ROI calculation and benefits, 39–40
- sponsorship, 25–26
- stakeholders, 26
- technical case, 28
 - convincing business people, 30–31
 - convincing technologists, 31–32
 - product short lists, 28–30

K

- Key performance indicators (KPIs), 12, 26–27, 49, 321, 338, 445, 487
- Key-value, 164

L

- Landing area. *See* Staging area
- Logical data model, 175–176, 437
- Logical data warehouse (LDW), 148

M

- Many-to-many relationships, 184, 184f
- Mappings, 289
- Markets in Financial Instruments Directive (MiFID), 53
- Massively parallel processing (MPP), 146–147, 162
- Master data management (MDM), 33, 46, 68, 103, 146, 451–452
 - identification, 104
 - problem areas finding, 104–105
 - solutions, 105–106, 139
- MDX. *See* Multidimensional eXpressions
- Mega-application vendor, 101
- Message queue scanners, 315
- Metadata management, 78–80, 98
 - categories, 98–99
 - value, 99–100
- MiFID. *See* Markets in Financial Instruments Directive
- Mini-dimension approach, 254
 - multiple, 254–255
 - outrigger needed, 256–257
- Mock-ups, 367, 368f
- MOLAP. *See* Multidimensional online analytical processing
- MPP. *See* Massively parallel processing
- Multidimensional eXpressions (MDX), 153
- Multidimensional online analytical processing (MOLAP), 159

- Multidimensional schema, 211–212, 212f
- Multifact star models, 212–213, 212f
- Multiple approach—multiple mini-dimension, 255
- Multiple data marts, 112–113, 112f
- Multivalued dimension, 263–265

N

- Natural key standard, 294f
- Nonadditive facts, 203
- Nonidentifying mandatory relationship, 182–183, 183f
- Nonidentifying optional relationship, 183–184, 184f
- Normalization, 189. *See also* Entity-relationship modeling (ER modeling)
 - ER model, 193–194
 - levels, 190
 - limits and purpose, 194–195
 - normalizing entity, 190
 - 1NF, 190–191
 - 3NF, 192–193
 - 2NF, 191–192
- Normalized modeling. *See* Entity-relationship modeling (ER modeling)
- NoSQL database, 163–164
- “Not Only” SQL database. *See* NoSQL database
- NULL values, 207–208

O

- ODBC. *See* Open database connectivity
- ODS. *See* Operational data store
- OLAP. *See* Online analytical processing
- OLTP. *See* Online transactional processing
- 1NF. *See* First normal form
- One-off reports, 406
- Online analytical processing (OLAP), 65, 94f, 159
 - analysis, 150
 - BI tool, 95
 - cubes, 91, 117, 148, 211–212
 - MDX language, 159–160
 - tools, 243
 - types, 159
- Online transactional processing (OLTP), 213
- Open database connectivity (ODBC), 153, 312
- Open source software (OSS), 145
- Operational data store (ODS), 65, 87–88, 113–114, 146
 - designs, 137–142
 - pragmatic approach, 142
 - rationale for, 138
 - data hub establishing for updates, 139–140
 - integrated reporting, 138–139
 - SOR for dimensions, 139
 - update data, 140
 - reexamining, 140
 - data hub establishing for updates, 141

- Operational data store (ODS) (*Continued*)
 - integrated reporting, 140
 - SOR for dimensions, 140–141
 - update data, 141
- Operational systems, 12, 178
 - bundled with reporting capabilities, 30
 - implementation in 3NF, 190
 - role, 11
 - trends, 123
- Operations management services, 331–332. *See also* Process management services.
- Organizational performance measure. *See* Fact
- OSS. *See* Open source software
- Outtrigger tables, 244, 245f
- P**
- Payment Card Industry (PCI), 53
- People, process and politics (Three P's), 425
 - building BI team, 431, 432f
 - extended project team, 440–441
 - project development team, 435–440
 - project management team, 434–435
 - project sponsorship and governance, 431–433
- business and IT relationship, 427
 - in charge, 427–428
 - communication shortcomings, 428–429
 - users or customers, 427
- under control, 426–427
- data governance, 444–448
- meeting expectations, 425–426
- roles and responsibilities, 429
 - in back office, 430
 - in front office, 429–430
- technology trap, 425–427
- training, 441
 - business group, 442–443
 - IT group, 442
 - methods, 443–444
 - types of, 441–442
- Performance management (PM), 82
- Performance testing, 299
- Periodic fact tables, 218
- Phases, 474
- Physical data model, 175–177, 438
- PIM. *See* Product information management
- PK. *See* Primary key
- PM. *See* Performance management
- PMO. *See* Program Management Office; Project management office
- PNG. *See* Portable network graphics
- POC. *See* Proofs of concept
- Portable network graphics (PNG), 360
- Portfolio approach, 505–506
- Power users, 342, 427
- Predictive analytics, 151, 377. *See also* Advanced analytics; Big Data analytics.
 - architecture for, 380
 - project methodology, 378f
 - resources and skills, 382
 - setting up, 377–378
 - tasks for, 378–380
 - techniques for, 381
 - tool selection, 380
- Primary key (PK), 187
 - analysis, 320
- Prioritizing requirements, 55–56
- Process management services, 328
 - audit services, 329
 - data integration jobs, 329
 - data quality metrics, 330
 - error handling services, 330–331
 - table and row updates, 330
 - job control services, 328, 329f
- Product, 143
 - architecture, 78
 - migration, 168–169
- Product evaluation, 165, 167
 - BI product vendors, 165–166
 - dazed and confused, 166–167
 - product migration, 168–169
 - technology and product evaluation, 167–168
- Product information management (PIM), 157
- Productive method, 277
- Productivity, 224
- ProductKey, 199
- Profile tables. *See* Hot swappable dimension
- Program Management Office (PMO), 433
- Project architecture, 470
- Project development team, 435
 - BI
 - developer, 439
 - leader, 439
 - business analysis leader, 440
 - business analyst, 439–440
 - core functions, 435f
 - data architect, 437
 - data modeler or designer, 437–438
 - DBA, 438
 - DI developer, 438–439
 - DI leader, 439
 - principal architect, 436–437
 - source data analyst, 440
 - sub-teams, 436t
- Project management

BI program, 450–460
 enterprise business strategy and business initiatives, 450f
 project methodologies, 473–478
 role, 449–450
 team, 434
 BI/DW project advisor, 434–435
 business advisor, 434
 project development manager, 434
 Project management office (PMO), 491
 Project methodologies, 473–478. *See also* Battling project methodologies.
 false choice, 476–477
 Project phases, 479–484, 479f
 analysis and definition phase, 481–482
 architect and design phase, 482
 build, test, and refine phase, 482–483
 deploy and roll-out phase, 484
 implementation phase, 483
 scope and plan phase, 480–481
 PromotionKey, 199
 Proofs of concept (POC), 88, 167, 455

Q

Query
 processing, 160
 speed, 224
 types, 265

R

RAD. *See* Rapid application development
 Radio Frequency Identification (RFID), 4
 Ragged hierarchy, 240–241, 240f
 Rapid application development (RAD), 455
 Rapidly-changing dimension, 254
 Real-time business intelligence, 11
 Recursive
 pointer, 242, 242f
 relationships, 185, 185f
 Reference lookups, 312, 316–317
 Referential integrity, 132, 188–189, 189f
 Reformat data, 89
 Regulatory requirements, 46, 53
 Relational database, 159
 technology, 163
 Relational modeling. *See* Entity-relationship modeling (ER modeling)
 Relational online analytical processing (ROLAP), 75, 159
 Relationship analysis, 320
 Representational state transfer (REST), 312
 Request for proposal (RFP), 480
 REST. *See* Representational state transfer
 Return on investment (ROI), 23, 74, 303, 341

 calculation and benefits, 39–40
 SOA benefit, 135
 Reusing data, 277
 Reverse engineering, 54–55, 419
 RFID. *See* Radio Frequency Identification
 RFP. *See* Request for proposal
 ROI. *See* Return on investment
 ROLAP. *See* Relational online analytical processing
 Role playing dimensions, 229–230, 230f

S

Scalable vector graphics (SVG), 360
 SCD. *See* Slowly changing dimension
 SCM. *See* Supply chain management
 Scop creep, 481
 Scope-priorities-budget task, 488
 Scorecards, 150
 SDLC. *See* Systems development life cycle
 Second normal form (2NF), 190–192
 Self-referencing relationship. *See* Recursive relationships
 Semiadditive facts, 203
 Senior vice president (SVP), 452
 Service level agreement (SLA), 327, 457
 Service oriented architecture (SOA), 12, 145, 153–154, 303, 503
 Shared dimensions, 212–213
 “Shared everything” system, 162
 “Shared nothing” system, 162
 SI. *See* Systems integrator
 Single mini-dimension, 255
 Sketches, 363, 363f
 SLA. *See* Service level agreement
 Slowly changing dimension (SCD), 237, 245–262, 283, 316
 effective date standard, 295f
 type 0 technique, 246
 type 1 technique, 246–249
 type 2 technique, 249–251
 type 3 technique, 251–253
 type 4 technique, 254–257
 type 5 technique, 257–259
 type 6 technique, 259–260
 type 7 technique, 260–262
 types, 316t
 Small-to medium-size businesses (SMB), 301
 Smart keys, 206f, 207
 SMB. *See* Small-to medium-size businesses
 SME. *See* Subject matter expert
 SMP. *See* Symmetric Multiprocessing
 Snowflake schema, 210–211, 210f
 SOA. *See* Service oriented architecture; Systems of analytics
 Software prototyping, 369
 SOI. *See* Systems of integration
 SOR. *See* Systems of record

Sorting loading utilities, 312

Source

- data analyst, 440
- source-to-target mapping, 320
- systems analysis, 318–320, 319f
- tables, 289

SP. *See* Stored procedures

Spreadmart. *See* Data shadow system

Spreadsheets, 403

- integration, 150–151

Sprints, 475

Staging area, 131–132

Stakeholders, 26

Standard calendar, 239f

Star schema, 208–209, 209f

Stepwise refinement, 46, 47f

Stored procedures (SP), 303

StoreKey, 199

Storyboards, 365, 366f

Subject matter expert (SME), 44, 49, 59, 281, 458, 498

Supply chain management (SCM), 6, 68, 146, 502

Surrogate keys, 206

SVG. *See* Scalable vector graphics

SVP. *See* Senior vice president

Swim lanes, 479, 487–488

- data and database, 488–489
- data integration, 489–490
- infrastructure, 490–491
- program and project management, 491–492

Symmetric Multiprocessing (SMP), 162

System testing, 299

Systems development life cycle (SDLC), 473

Systems integrator (SI), 157

Systems of analytics (SOA), 71, 124

- data workflow, 133
- benefits, 135
- need for, 134–135

Systems of integration (SOI), 71

- data workflow, 130
- data auditability story, 131b
- data stores, 130f
- splitting EDW, 132–133
- staging area, 131–132

Systems of record (SOR), 51, 68, 71, 124, 338

- data integration architecture, 277
- dimension, 294f
- SMEs, 498

T

Table analysis, 319

Target tables, 289

TCO. *See* Total cost of ownership

Technical architecture, 28, 72

BI, 72f–73f, 74–75

- analytical styles, 74f
- data-related technology evolution, 77–78, 77f
- DW and data stores, 75–76, 75f
- convincing business people, 30–31
- convincing technologists, 31–32
- data integration, 76, 76f
- product short lists, 28–30
- source systems, 76–77, 77f

Technical metadata, 79

Technology architecture, 147, 148f, 470

BI analytics, 149–151, 149f

BI information access and data integration, 151–152

- BI targets, 152–153
- data access APIs, 153
- data integration, 157
- data integration suites, 154–155
- ETL vs. ELT, 155–157
- integration applications, 155
- integration services, 153–154

databases, 158, 158f

BI appliances, 164–165

Big Data, 163–164

OLAP comparisons, 160f

relational, 159

relational alternatives for BI, 159–162

technology and product evaluation, 167–168

Technology evaluation, 165

BI product vendors, 165–166

dazed and confused, 166–167

product migration, 168–169

technology and product evaluation, 167–168

Third normal form (3NF), 122, 179b, 189–190, 192–193

database development, 190

no longer necessary, 122–123

Three P's. *See* People, process and politics

3 Vs. *See* Volumes, varieties and velocity

3NF. *See* Third normal form

Time boxing, 475

Time dimension, 225

date dimension vs., 228–229

time periods, 227–228

time-of-day

- as dimension, 225–227, 226f
- as fact, 225, 226f

Time zones, 229f

date vs. time dimension, 228–229

Time-series or trending analysis, 351–352

bar chart, 355f

line graph, 353f

“Top-down” approach, 452

Total cost of ownership (TCO), 34, 76, 135, 154–155, 305
 Training, 441. *See also* Business intelligence (BI).
 business group, 442–443
 IT group, 442
 methods, 443–444
 types, 441–442
 Transaction fact tables, 218
 Transactional processing. *See* Operational systems
 Transform data, 89
 2NF. *See* Second normal form

U

UAT. *See* User acceptance testing
 UI. *See* User interface
 UML. *See* Unified modeling language
 Unbalanced hierarchy, 241, 241f
 Unified modeling language (UML), 477
 Unit testing, 299
 User acceptance testing (UAT), 455
 User acceptance testing, 299
 User interface (UI), 150–151
 standards, 359–361
 User unit tests, 374

V

Variable-depth hierarchy, 241–244
 Velocity, 5
 Vendor, 143
 Volatile changing dimension, 254
 Volume, 4
 Volumes, varieties and velocity (3 Vs), 375

W

Waterfall methodology, 474–475
 Wireframes, 363–365, 364f
 Work breakdown structure (WBS), 450, 465, 467. *See also*
 Project management.
 for BI, 468
 architecture WBS, 469–470
 program WBS, 468
 project WBS, 468–469
 design principles, 467–468

X

XML. *See* Extensible markup language
 XML/A. *See* Extensible markup language for Analysis
 XQuery, 153