

INDEX

A

- Ab Initio, 43
- accuracy. *See* data quality dimensions
- ACORD, 275
- ActiveLinx, 50, 51, 294
- Activiti plug-in, 182, 248
- Axiom, 13, 94, 258
- ADABAS, 287
- Adaptive, 293
 - Business Glossary Manager, 10, 293
 - Metadata Manager, 10, 293
- addresses
 - data quality of, 93–95
 - profiling, 93
 - standardization, 89–92
- Adobe Acrobat/PDF/Reader, 17, 33, 34, 146, 202
- alerts, 105, 233, 236
- Amazon
 - DynamoDB, 12, 231
 - Elastic Compute Cloud (EC2), 209
 - Elastic MapReduce (EMR), 12, 209
 - Redshift, 290
 - Web Services, 209, 231, 250
- American National Standards Institute, 161
- analytics. *See also* analytics and reporting
 - big data, 11, 211, 240
 - in the EDM reference architecture, 11
 - provisioning a semantic lawyer for in Hadoop, 213–216
 - text, 239–243
 - tools, 11
 - pulling metadata from, 47
 - analytics and reporting, 177–180
 - integration points with data governance, 177–180
 - export data artifacts to a reporting tool for the visualization of data governance metrics, 178
 - export data profiling results to a reporting tool for further visual analysis, 177
 - integrate analytics and reporting tools with the business glossary for semantic context, 179
 - tools, 177
- Apache
 - Cassandra, 12, 231, 290
 - CouchDB, 12, 231
 - Flume, 12, 233
 - Hadoop. *See* Hadoop
 - Hadoop Distributed File System (HDFS), 51, 210, 211, 221, 222, 232, 233, 291
 - Hadoop User Experience (Hue), 210
 - HBase, 12, 51, 211, 231, 291
 - HCatalog, 51, 211, 291
 - Hive, 51, 210–211, 217, 218, 222, 223, 275, 276, 277, 291
 - HiveQL, 210–211
 - MapReduce, 12, 130, 209, 210, 211, 213, 214, 215, 222, 230, 275
 - Oozie, 51, 211
 - Pig, 211, 219, 220, 221, 275, 291
 - Pig Latin, 211
 - Sqoop, 211, 217, 218, 291
- archiving, 12, 201–202
- AS/400 RPF, 260
- ASG, 3, 259–261
 - ASG-becubic, 46, 260–261

- discovering data elements in a Visual Basic form, 260–261
- view encompassing Java programs, Cobol, and DB2, 46

ASG-metaGlossary, 10, 179, 259

ASG-MyInfoAssist, 179

- search results for business glossary term displayed in, 179–180

ASG-Rochade, 10, 46, 47, 259, 260

- detailed forward lineage report in, 53–54, 217–218, 260

- high-level forward lineage report in, 52, 217

- SAS scanner in, 47

- view encompassing Java programs, Cobol, and DB2, 46

- viewing metadata for Oracle table in, 216–217

ASN.1 HL7, 275

ASP, 260

ASP.net, 260

assets

- search for in metadata repository, 57–58

- types, custom, 26, 28

attributes

- custom, 22, 28

- OOTB, 22

audit

- capabilities, 296

- data, 194

- trail, 32–33

B

BackOffice Associates

- data quality/data migration for SAP, 294

Bank of England Two Letter codes, 137

big data

- analytics, 11, 211, 240

- captured by CDC technology, 169–170

- defined, 13

- examples, 13

- first step to governing of, 211

- governance tools for, 209–243, 247

- Hadoop and NoSQL for, 209–232

- impact on data governance, 3, 4

- and privacy requirements, 190

- pulling metadata from, 50–51

- stream computing and, 233–237

- text analytics, 239–243

- three Vs of, 13

- tools, 207–243, 267, 291

blog data, 226

BMC Remedy, 149

bulk data movement, 9, 169

bulk-loading business terms, 17–19

business glossary, 17–36

- business rules, documenting in, 67–68

- defined, 10, 17

- documenting information policies implemented by ILM tools in, 204

- in the EDM reference architecture, 10

- information policy support in, 146–147, 195

- integration with analytics and reporting tools, 179–180

- integration with data dictionary, 17

- integration with desktop applications, 33

- metrics, 256

tasks related to data governance, 17–36

- add custom attributes to business terms and other data artifacts, 22

- add custom relationships to business terms and other data artifacts, 23

- add custom roles to business terms and other data artifacts, 23–24

- allow business users to link to the glossary directly from reporting tools, 33–34

- automatically hyperlink embedded business terms, 21–22

- bulk-load business terms in Excel, CSV, or XML format, 17–19

- create categories of business terms, 20

- facilitate social collaboration, 20–21

- flag critical data elements, 28

- integrate business terms with associated unstructured data, 35–36

- link business terms and column names to the associated reference data, 24–25

- link business terms to technical metadata, 25–26

- provide OOTB and custom workflows to manage business terms and other data artifacts, 28–32

- review the history of changes to business terms and other data artifacts, 32–33

- search for business terms, 34–35

- support the creation of custom asset types, 26–27

- terms exported to analytics/reporting tools, 178

- tools, 10, 259, 263, 277, 283, 286, 295

- widgets providing access to business terms, 179–180

- business intelligence, 28, 44, 179, 242, 272
 - business process management (BPM), 181–188
 - defined, 11, 181
 - in the EDM reference architecture, 11
 - integration points with data governance, 181–188
 - data governance tools map to BPM tools, 187
 - data governance workflows leverage BPM capabilities, 181
 - master data workflows leverage BPM capabilities, 186
 - tools, 11, 181, 187, 292
 - Business Process Model and Notation (BPMN), 11, 181, 182, 248, 265
 - business processes, as part of information lifecycle management, 201
 - business rules, 20, 66–68, 71, 147–148, 273, 286
 - attribute, 69, 70
 - for data quality dimensions, 100–102
 - in IBM customer data governance dashboard, 272–273
 - managing, 147–148
 - business semantics, 294. *See also* business glossary, business terms
 - business terms. *See also* business glossary
 - consistent definitions for in text analytics, 242
 - defined by the MDM hub, 109–111
 - linked to business drivers and financial benefits in Global Data Excellence Data Excellence Management System, 295
 - linking with column names, 47–48
 - used by stream computing applications, governing, 236–237
 - workflows to manage, 294
- C**
- CA ERwin Data Modeler, 9, 37, 294
 - case management, 295
 - change data capture (CDC), 9, 169, 272
 - CICS, 260
 - Cisco, 231
 - Clarabridge, 13, 239
 - cloud
 - data governance platforms for, 13, 258, 278, 290, 292
 - in the EDM reference architecture, 13
 - as evaluation criterion for data governance tools, 258
 - Cloudera
 - Apache Hue landing page in, 210
 - Hadoop distribution, 12, 209
 - Impala, 211
 - MapReduce word-count program executing in, 213–216
 - Pig script to sort countries in descending order in, 219–220
 - word-count program executing within, 213–216
 - compiling the WordCount.java file, 213–214
 - creating the wordcount.jar file, 214
 - executing MapReduce code on the .csv file, 215
 - viewing the results of the MapReduce word-count program, 215
 - Cobol, 42, 43, 45, 46, 260, 287, 294
 - code lists, 11, 133, 283. *See also* reference data
 - code sets, 11, 133. *See also* reference data
 - code tables, 11, 133, 182. *See also* reference data
 - Collibra, 3, 263–265
 - Business Glossary, 10, 13, 258, 263–264
 - Data Governance Center, 11, 13, 110, 174, 235, 236, 258, 263–265
 - approval workflow in, 248–249
 - banking business terms in, 213
 - business rule for master data enrichment of address data in, 113
 - business rules display in, 68
 - business terms and data artifacts, categories of in, 20
 - business terms in, 216, 265
 - comparing reference data snapshots in, 140–141
 - country code mapping in, 137
 - custom asset types in, 26–27
 - data governance dashboard in, 256–257
 - data quality scorecard, viewing in, 103
 - data stewardship console in, 13, 265
 - definition of “Cash” in, 21
 - definition of “Customer” in, 110
 - definition of “Working Capital” in, 22
 - Excel Import wizard, 18–19
 - flagging Critical Data Elements in, 28
 - flagging sensitive data in, 192
 - hierarchy of code values in, 139–140
 - hyperlinked business terms in, 21–22
 - integration with key metadata repositories, 265
 - integration with Trillium Software TS Discovery for data profiling, 265
 - linking business terms to reference data in, 24–25
 - logging data issues in, 149–157
 - new business term approval process in, 181–186
 - OOTB approval workflow in, 182

- OOTB relations in, 23
 - OOTB workflows customizable in, 265
 - preferred view of Dodd-Frank business terms in, 19
 - providing the semantic layer for Hadoop analytics in Cloudera, 216
 - Reference Data Accelerator, 13, 258
 - reference data management in, 264
 - user interface, 263
 - Data Stewardship Manager, 13, 165, 258, 265
 - Reference Data Accelerator, 13, 258, 264
 - column
 - analysis, conducting, 59–61, 80–81, 272
 - dependencies, 81–84
 - length frequencies, discovering, 63–64
 - patterns distribution, discovering, 62–63
 - values distribution, discovering, 61–62
 - values with similar sounds in, discovering, 65–66
 - complete business objects, 199–200, 204–205
 - completeness. *See* data quality dimensions
 - complex event processing (CEP), 233
 - compliance, 11, 149, 190, 294
 - conceptual data model, 161
 - Concur, 278
 - conformity. *See* data quality dimensions
 - consistency. *See* data quality dimensions
 - console. *See* data steward, console/dashboard for
 - Constant Contact, 231
 - content analytics, 241
 - content management, 287
 - Couchbase, 12, 231
 - Critical Data Elements (CDEs), 28, 76, 256
 - CSV, 17
 - Current Procedural Terminology (CPT), 133
 - custom code, reflected in metadata tool, 45–46
- D**
- DAMA UK, 66
 - dashboard. *See* data steward, console/dashboard for
 - data access, policies around, 195–196
 - Data Advantage Group, 293–294
 - MetaCenter, 10
 - ActiveLinux scanner capabilities in, 294
 - business glossary in, 294
 - complex code mappings in, 137–139
 - data ownership display in, 149, 212–213
 - establishing group privileges in, 56–57
 - Hadoop ActiveLinux scanner in, 50–51
 - linking business terms to technical metadata, 25–26
 - managing business terms and other data artifacts with, 28–33
 - managing custom roles in, 24
 - metadata capabilities in, 294
 - reference data capabilities in, 294
 - reference data inventory in, 134–135
 - reference data relating to Pig script in, 221
 - reviewing business term history in, 32–33
 - workflow capabilities in, 294
 - data archiving, 195, 204, 272, 273, 275, 277, 287
 - data artifacts, hierarchy of, as evaluation criterion for data governance tools, 249
 - data brokers, 8, 13, 258
 - data cleansing, 13, 106, 108, 126, 258, 275, 286, 295
 - data consolidation projects, 163
 - data dictionaries, 17
 - integration with business glossary, 17
 - data discovery, 4, 10, 59, 163–166, 275. *See also* data profiling
 - performing natively in Hadoop, 223–224
 - tools, 61, 62, 64, 76, 80, 81, 84, 163, 167, 190, 199, 200, 295
 - creating a complete business object using, 199–200
 - discovering sensitive data using, 190–192
 - data encryption, 11, 189
 - data enrichment, 13, 112, 113, 126, 170, 258, 272, 291
 - data extraction, 172, 204
 - data federation, 170. *See also* data virtualization
 - data flows, conceptual view of, 46
 - data governance
 - cloud platforms for, 13
 - council, 146
 - defined, 3
 - EDM reference architecture categories related to, 14
 - impact of poor, example of, 5
 - introduction to, 3–6
 - in NoSQL databases, 231
 - prerequisites, 4–5
 - tool categories, 15–157
 - tools, 14, 76, 167, 247–297
 - mapping to BPM tools, 187–188
 - porting to Hadoop for improved performance, 229–230

- vendor landscape for, 247–297
- tools/platforms, evaluation criteria for, 247–258
 - approval workflows, 248–249
 - cloud, 258
 - data governance metrics, 256–257
 - data stewardship, 248
 - hierarchy of data artifacts, 249–256
 - total cost of ownership, 247–248
- data integration, 169–175
 - bulk data movement, 169
 - categories, 9, 169–170
 - data replication, 169–170
 - data virtualization, 170
 - in the EDM reference architecture, 9
 - integration points with data governance, 170–175
 - automate the production of data integration jobs by leveraging the metadata repository, 174–175
 - deploy data quality jobs in an integrated manner with data integration, 170–172
 - integrate data integration tools into the metadata repository, 174
 - leverage reference data for use by the data integration tool, 173–174
 - move data between the MDM or reference data hub and the source systems, 172–173
 - technologies, 169–170
 - tools, 10, 172, 272, 275
- data issues
 - managing, 149–157, 264, w286
 - metrics, 256
- data lineage
 - custom reporting of, 52–54
 - Hadoop data and, 50
 - providing detailed views on, 51
 - reports (high-level forward, detailed, and map view) in ASG-Rochade, 52–55, 217–218, 260, 286, 294
 - sample report in InfoSphere Metadata Workbench, 52
 - view encompassing Java programs, Cobol, and DB2, 46
 - viewing in and out of Hadoop, 216–218
- data management, 145, 161, 195, 287. *See also* enterprise data management (EDM), master data management (MDM), metadata management, reference data management, test data management
- data marts, in the EDM reference architecture, 11
- data masking, 11. *See also* data security, data security and privacy, sensitive data
 - based on primary key-foreign key relationships, 200
 - defined, 189
 - in Hadoop, 232
 - in near-production environments, 278
 - in non-production environments, 193–194, 274, 278
 - in production environments, 193, 278
 - tools, documenting policies implemented by, 195–199
- data matching, 34, 64, 95, 96, 97, 106, 109, 114, 115, 116, 119, 227, 229, 230, 235, 272, 275, 291
- data modeling, 161–167
 - defined, 161–162
 - in the EDM reference architecture, 9
 - evaluation criteria for, 247
 - integration points with data governance, 161–167
 - establish a data model to support master data management, 166–167
 - expose ontologies in the metadata repository, 163
 - integrate the logical and physical data models with the metadata repository, 162
 - prototype a unified schema across data domains using data discovery tools, 163–166
 - tools, 9, 37–39, 162, 263, 274, 280–282, 287, 294
- data overlaps across columns, discovering, 76–79
- data ownership, 5, 146, 149, 212, 213, 231, 212–213
- data privacy. *See also* data masking, data security, data security and privacy, sensitive data
 - determining obligations for, 190
 - in the EDM reference architecture, 11
 - requirements in healthcare, 243
 - subcategories, 11
 - tools, 11, 274, 278
- data profiling, 4, 59–88
 - defined, 59
 - in the EDM reference architecture, 10
 - exporting results to reporting tools, 177–178
 - performing natively in Hadoop, 221–222
 - stream data and, 234–235
 - tasks related to data governance, 59–88
 - agree on the data quality dimensions for the data governance program, 66–67
 - conduct column analysis, 59–61
 - create virtual joins or logical data objects that can be profiled, 86–88
 - develop business rules relating to the data quality dimensions, 67–68
 - discover data overlaps across columns, 76–79
 - discover data transformations, 84–86
 - discover dependencies, 81–84

- discover hidden relationships between columns, 80–81
- discover hidden sensitive data, 64–65
- discover length frequencies of a column, 63–64
- discover patterns distribution of a column, 62–63
- discover values distribution of a column, 61–62
- discover values with similar sounds in a column, 65–66
- profile data relating to accuracy dimension of data quality, 75–76
- profile data relating to completeness dimension of data quality, 69
- profile data relating to conformity dimension of data quality, 69–70
- profile data relating to consistency dimension of data quality, 71
- profile data relating to synchronization dimension of data quality, 71–72
- profile data relating to timeliness dimension of data quality, 74–75
- profile data relating to uniqueness dimension of data quality, 73–74
- tools, 10, 265, 267, 268, 272, 275–277, 286, 291, 295, 296
- data quality. *See also* data quality management, data quality scorecard
 - of addresses, 93–95
 - considerations for stream computing applications, 235
 - dashboard, 123
 - dimension, defined, 66
 - dimensions for data governance program
 - accuracy, 67, 75–76, 104
 - agreeing on, 66–67
 - business rules for, 68, 100–102
 - completeness, 66, 69, 100–102, 104
 - conformity, 66, 69–70
 - consistency, 67, 71, 99, 103, 269
 - defined, 66–67
 - developing business rules relating to, 67–68
 - measuring, 99–100
 - synchronization, 66, 71–72
 - sample business rules for, 66–67
 - timeliness, 66, 74–75
 - uniqueness, 66, 73–74
 - weights assigned to, 99–100
 - in the EDM reference architecture, 10
 - exceptions, 106–108
 - financial impact of poor, 104
 - history, 104–106
 - linked to business drivers and financial benefits in Global Data Excellence Data Excellence Management System, 295
 - metrics, 20, 124, 178, 248
 - process for rating attributes at an insurance carrier, 249–250
 - rules, 48
 - executing natively in Hadoop, 224–226
 - scorecards, 68, 98–103, 256
 - thresholds, 98
 - tools, 10, 267, 272, 275–277, 286, 291, 294, 295
 - pulling metadata from, 48–50
- data quality management, 4, 10, 89–108, 267
 - defined, 89
 - tasks related to data governance, 89–108
 - conduct time series analysis, 104–106
 - create a data quality scorecard, 98–102
 - highlight the financial impact associated with poor data quality, 104
 - improve the quality of address data, 93–95
 - manage data quality exceptions, 106–108
 - match and merge duplicate records, 95–97
 - transform data into a standardized format, 89–92
 - view the data quality scorecard, 103
- data quality scorecard
 - creating, 98–102
 - assign weights to each business rule in a given data quality dimension, 101–102
 - bind the business rules to the relevant columns, 102
 - define the acceptable thresholds of data quality, 98
 - select the business rules for each data quality dimension, 100–101
 - select the data domain or entity, 98
 - select the data quality dimensions to be measured for the specific data domain or entity, 66, 99
 - select the weights for each data quality dimension, 99–100
 - tools, 256, 269, 273
 - viewing, 103
- data replication
 - defined, 9, 169–170
 - tools, 9, 170, 272
- data security. *See also* data security and privacy
 - in the EDM reference architecture, 11
 - subcategories, 11
 - tools, 11, 274, 278

- data security and privacy, 189–200. *See also* data masking, data security, sensitive data
 - integration points with data governance, 189–200
 - create a complete business object using data discovery tools that can be acted upon by data masking tools, 199–200
 - determine privacy obligations, 190
 - discover sensitive data using data discovery tools, 190–192
 - document information policies implemented by data masking and database monitoring tools, 195–199
 - flag sensitive data in the metadata repository, 191–192
 - mask sensitive data in non-production environments, 193–194
 - mask sensitive data in production environments, 193
 - monitor database access by privileged users, 194
 - tools, 11, 274, 278
- data sources
 - in the EDM reference architecture, 8
 - examples, 8
 - internal and external, 8
 - legacy, 42–43
- data space, 127
- data standardization, 92, 106, 170, 272, 291, 295, 296
- data steward. *See also* data stewardship
 - activities of, 248
 - adding and deleting, 248
 - assignment of, 265
 - console/dashboard for, 55, 56, 98, 103, 106, 118, 121–123, 126, 178, 191, 228, 229, 248, 256–257, 269, 272, 277, 282, 290, 295, 296
 - group, 56–57
 - outstanding tasks of, 117–118
- data stewardship
 - support for, as evaluation criterion for data governance tools, 248–256
- data tokenization, 11
 - defined, 189
- data transformations, discovering, 84–86
- data validation, 13, 75, 76, 113–114, 147–148, 252–255, 258, 272, 291
 - rules, 113–114
- data verification, 94, 103, 272, 295, 296
- data virtualization
 - defined, 9, 170
 - tools, 9, 170, 272
- data warehousing/data marts, 9, 268, 290. *See also* Hive appliances, 11, 41
 - in the EDM reference architecture, 11
 - pulling metadata from, 41–42
- database monitoring, 11, 194, 195
 - defined, 189–190
 - tools
 - documenting policies implemented by, 195–199
 - enforcing separation of duties, 194
- databases
 - access/activity monitoring, 194, 199, 274
 - columnar, 11
 - in the EDM reference architecture, 8–9
 - encryption, 11
 - in-memory, 8
 - legacy, 9
 - monitoring, 11
 - relational, 8–9, 11, 40, 268, 287, 294
- de-duplication, 97, 106, 114, 118, 120, 126, 129, 130, 170, 295
- Denodo, 9, 170
- dependencies, discovering, 81–84
- Diaku
 - Axon
 - business glossary functionality in, 294
 - capabilities for business semantics, 294
 - capabilities for data lineage, 294
 - capabilities for data quality, 294
 - capabilities for process mapping, 294
 - capabilities for stakeholder analysis, 294
 - KYC activities mapped to business glossary in, 188
 - KYC business process reflected in, 187–188
 - specialization in regulatory compliance for financial services, 294
- dimensions. *See* data quality dimensions
- disposal of information, 11, 12, 202
- document retention, 202
- Dodd-Frank banking legislation, 17
 - business terms, 18, 19, 213–215
- Dun & Bradstreet (D&B), 111, 122
- duplicate records
 - managing, 282
 - matching and merging, 95–97
 - terminology, 96

E

- Eclipse, 11, 181, 182, 292
- EDI, 275
- eDiscovery, 12, 202
- EDM reference architecture, 7–14
- Eloqua, 278
- Embarcadero Technologies
 - CONNECT, 10, 21, 294
 - business glossary in, 294
 - data model integration in, 294
 - Facebook-style user interface in, 20–21, 294
 - social collaboration features in, 21, 294
 - data modeling capabilities of, 294
 - ER/Studio, 9, 25–26, 37, 294
 - ERwin Data Modeler, 9, 37, 294
- EMC Pivotal, 12, 290
 - Documentum Records Manager, 202
 - Greenplum, 11
 - Global IDs Database Crawler for, 41–42
 - Greenplum Database, 41
 - Greenplum HD, 12, 209
 - RSA Archer GRC, 10, 146
 - SourceOne, 202
- encryption, 11
- enrichment rules, 112–113. *See also* data enrichment
- enterprise data management (EDM)
 - categories, 8–13
 - reference architecture for, 7–14
 - three pillars of, 8
- entity relationships, managing, 111–112
- exact matching function, 96
- Experian, 94
- Extract, Load, and Transform (ELT), 169
- Extract, Transform, and Load (ETL), 9, 53, 169, 211, 217, 275, 294
 - tools, pulling metadata from, 43–44

F

- Facebook, 21, 125, 126, 212, 226, 277, 290, 295
- Financial Industry Business Ontology (FIBO), 163
- FIX, 275
- Force.com, 278

G

- Global Data Excellence, 295
 - business glossary functionality in, 295
 - business terms and data quality functionality in, 295
 - Data Excellence Management System, 295
- Global IDs, 267–270
 - Cobol Copybook Scanner, 42–43
 - data profiling capabilities of, 268–269
 - data quality capabilities of, 269
 - data quality dashboard in, 269
 - Database Crawler for Greenplum, 42
 - Graph Database Profiler, 231
 - HDFS File Profiler, 221–222
 - Hive Profiler, 268–269
 - metadata capabilities of, 270
 - metadata repository, searching for assets in, 57–58
 - product suite, 268
 - Profiler, discovering hidden sensitive data in, 64–65
 - remediating data in, 269
 - scanner for Financial Industry Business Ontology (FIBO), 163
 - ultra-large implementations and, 267
 - Web Ontology Language (OWL) ontology scanner, 270
- golden record, 96, 116, 119, 120, 121
- Google, 12, 209
 - BigQuery, 290
 - Cloud Platform, 290
 - Maps Enricher, 112
 - Search Appliance, 35
- Governance, Risk, and Compliance (GRC) platforms/
 - tools, 10, 146
- graph database, 231
- Guidance Software EnCase eDiscovery, 202

H

- Hadoop, 10, 209–232, 236, 268, 270, 275
 - ActiveLinux in MetaCenter, 50, 294
 - big data and, 13, 50
 - “data lake,” 211
 - defined, 12
 - distributions, 12, 209
 - in the EDM reference architecture, 12
 - extracting metadata from, 50, 51
 - intersection with data governance tools, 211–232
 - assign ownership for data in Hadoop, 212–213
 - conduct an inventory of data in Hadoop, 211–212

- discover data natively in Hadoop, 223–224
 - execute data quality rules natively in Hadoop, 224–226
 - govern data in NoSQL databases, 231
 - integrate Hadoop with master data management, 226–229
 - manage reference data for Hadoop, 219–221
 - mask sensitive data in Hadoop, 232
 - port data governance tools to Hadoop for improved performance, 229–230
 - profile data natively in Hadoop, 221–222
 - provision a semantic layer for analytics in Hadoop, 213–216
 - view the lineage of data in and out of Hadoop, 216–218
 - MDM hub integration with, 129–130
 - technologies, 210–211
 - Hadoop Distributed File System (HDFS), 210, 211, 291
 - Harte-Hanks Trillium Software. *See* Trillium Software
 - Health Insurance Portability and Accountability Act (HIPAA), 243, 275
 - hidden data, 223
 - hidden relationships between columns, 80–81
 - hidden sensitive data. *See* sensitive data
 - hierarchy management, 122–123, 291
 - HIPAA, 243, 275
 - Hive, 51, 210–211, 217, 218, 222, 223, 275, 276, 277, 291
 - HiveQL, 210–211
 - Hortonworks, 12, 209, 290
 - HP, 12
 - Consolidated Archive, 202
 - eDiscovery, 202
 - Quality Center, 149
 - Records Manager, 202
 - Vertica, 11, 290
 - hyperlinking business terms, 21–22
- I**
- IBM, 3, 11, 12, 109, 197, 198, 199, 204, 232, 271–274
 - Business Process Manager, 11, 181
 - Cognos, 11, 33, 177, 272, 273
 - Cognos BI Reporting, 44–45
 - data integration capabilities of, 272
 - data lifecycle management capabilities of, 273
 - data profiling capabilities of, 272
 - data quality capabilities of, 272
 - data quality scorecard
 - based on Cognos, InfoSphere Information Analyzer, and InfoSphere Business Glossary, 272–273
 - steps for defining, 273
 - DB2, 9, 11, 40, 45, 46, 172, 193, 201, 260, 273, 274, 290
 - Domino, 201
 - eDiscovery Manager, 202
 - Enterprise Records, 202
 - Global Retention Policy and Schedule Management, 190
 - Information Integration and Governance platform, 271
 - Information Management System (IMS), 9
 - InfoSphere BigInsights, 12, 209, 210
 - creating an unstructured entity from social media in, 226–227
 - integration with InfoSphere Master Data Management, 227
 - probabilistic matching engine (“Big Match”) in, 229–230
 - InfoSphere Blueprint Director, 272
 - InfoSphere Business Glossary, 10, 55, 110, 175, 236, 254, 272, 273
 - administrators, 56
 - business term linked to information governance rule for data validation in, 253
 - creating custom attributes in, 22
 - data governance policy theoretically integrated with InfoSphere Guardium Data Activity Monitor, 198–199
 - data policy relating to data quality in, 250–251
 - data validation rule for U.S. State in, 252
 - information governance rule related to data quality in, 48–50
 - information policy display in, 147
 - integrating business terms with associated unstructured data in, 35–36
 - linking to business terms from, 33–34
 - NCOA data policy in, 253
 - NCOA information governance policy in, 251–252
 - policy-based integration with InfoSphere Optim Data Masking Solution, 197–198
 - published glossary permissions in, 56
 - smart meter data governance policies in, 196–199
 - InfoSphere Business Glossary Anywhere, 33–34
 - InfoSphere Data Architect, 9, 37
 - InfoSphere Data Explorer, 35–36
 - InfoSphere Data Replication, 9, 170, 272

- InfoSphere DataStage, 43, 172, 174, 272
 - with Rule Stage from InfoSphere Information Analyzer, 171–172
 - stages to retrieve data from InfoSphere Master Data Management, 172–173
 - InfoSphere Discovery
 - column analysis in, 61, 80, 164–165
 - column analysis view with sensitive data classifications in, 192
 - creating a complete business object in, 205
 - defining a unified schema in, 164
 - discovering data transformations in, 84–85
 - discovering relationships in, 81
 - drilling down into credit card number pattern in, 191
 - hidden business rule inferred in, 149–150
 - length frequencies display in, 63–64
 - mapping sources to target column names in, 165
 - match-and-merge analysis in, 165–166
 - overlap analysis in, 76–79, 164–165
 - sensitive data classifications in, 190
 - uncovering primary key-foreign key relationships to support data masking in, 200
 - InfoSphere FastTrack
 - automating data integration jobs with, 174–175
 - InfoSphere Federation Server, 9, 170, 272
 - InfoSphere Guardium, 11, 274
 - masking sensitive data in production databases in, 193
 - InfoSphere Guardium Data Activity Monitor, 194
 - data governance policy theoretically integrated with, 198–199
 - InfoSphere Information Analyzer, 48, 171–172, 234, 272, 273
 - business rule trend display in, 105
 - data quality alert report in, 105
 - data rules stored as objects in, 104
 - drilling down into data rule from, 49–50
 - virtual table definition pane in, 86–87
 - InfoSphere Information Analyzer Database, 273
 - InfoSphere Information Server
 - administration console, 55
 - assigning roles in, 55
 - metadata repository underlying, 174
 - InfoSphere Master Data Management, 122, 235, 273
 - data extraction from 172–173
 - data quality dashboard for data from, 123
 - data quality metrics exposed in InfoTrellis Veriscope, 124
 - data stewardship console displaying social media data, 228
 - integration with InfoSphere BigInsights, 227
 - largest bulk entity resolution in, 229
 - probabilistic matching engine in, 229
 - Reference Data Management Hub, 11, 235, 254–255
 - InfoSphere Metadata Asset Manager, 55
 - InfoSphere Metadata Workbench, 10, 37–41, 49–50, 55, 174, 197, 272
 - data lineage report in, 52
 - graph view of physical model in, 39
 - imported transaction report in, 44–45
 - logical data model in, 37–38
 - technical metadata for Address in, 255–256
 - view of database schema in, 41
 - views of metadata in, 40, 43–44, 44–45
 - physical data model in, 38–39
 - InfoSphere Optim Data Growth Solution, 201, 204, 273
 - InfoSphere Optim Data Masking Solution, 197, 198
 - masking office phone numbers in, 193–194
 - InfoSphere Optim Data Privacy Solution, 11, 274
 - masking sensitive data in Hadoop with, 232
 - InfoSphere Optim Designer, 198
 - InfoSphere Optim Test Data Management, 203, 273, 274
 - InfoSphere QualityStage, 10, 272
 - InfoSphere QualityStage Designer, 171–173
 - InfoSphere Streams, 12, 233, 234, 236–237
 - composite flow graph to handle POS transactions in, 236–237
 - streaming application flow graph in, 234
 - master data management capabilities of, 273
 - metadata capabilities of, 272
 - Notes, 201
 - OpenPages GRC, 10, 146
 - privacy and security capabilities of, 274
 - PureData System for Analytics, 11, 41
 - solidDB, 8
 - SPSS, 11, 177, 240
 - Streams Processing Language, 236
 - support for data masking in Hadoop, 232
 - text analytics, 13, 239
 - Value-Based Archiving Solutions, 202
- ICS-9 reference data, 242, 243
 - impact analysis, 50, 286
 - Imperva, 11
 - Database Activity Monitoring, 194

- Informatica, 3, 11, 12, 20, 109, 232, 275–278
 - AddressDoctor, address correction in, 94–95
 - archiving, policy-based, in, 277
 - Business Glossary, 10
 - business glossary capabilities of, 277
 - Cloud, 13, 258, 278
 - cloud capabilities of, 278
 - Data Archive, 201, 275, 277
 - Data Masking, 11, 193, 275
 - data masking capabilities of, 278
 - data profiling capabilities of, 275–277
 - Data Quality, 10, 234, 275
 - Data Governance Edition, 277
 - data stewardship interface in, 108
 - duplicate triage in, 107–108
 - standardizing data in, 90–92
 - workflow of human and automated tasks, 106–107
 - data quality capabilities of, 275–277
 - Data Replication, 9, 170
 - Data Services, 9, 170
 - Data Subset, 203, 277
 - Developer
 - Customer logical data object in, 88
 - logical data object (LDO) concept in, 87–88
 - Dynamic Data Masking, 278
 - Facebook-style user interface of, 20–21, 277
 - Fast Clone, 9, 170
 - HPParser, 275
 - Informatica 9.6, 275
 - information lifecycle management capabilities of, 277
 - integration of MDM with social media data in, 125
 - Master Data Management, 125
 - 360-degree view of customer relationships in
 - Salesforce, 278
 - customer attributes in, 125–126
 - data stewardship console in, 126
 - fuzzy matching in a Hadoop MapReduce cluster, 230
 - integration with social media data, 277
 - multi-domain implementations and, 277
 - social media profile in, 125–126
 - master data management capabilities of, 277
 - metadata capabilities of, 277
 - Metadata Manager, 10, 110, 236
 - PowerCenter, 9, 43, 170, 275
 - mapping for country code lookup, 173–174
 - testing accuracy dimension in, 75–76
 - testing timeliness dimension in, 74–75
 - PowerCenter Big Data Edition
 - complex data parsing with, 225–226, 276
 - data domain discovery within Apache Hive in, 223–224
 - data profiling, ETL, and complex data parsing on Hadoop in, 275–276
 - discovering hidden data natively in Hadoop with, 223–224
 - mappings in, 225
 - monitoring console of, 277
 - profiling the WebLogs Logical Data Object in Hive in, 276
 - security and privacy capabilities of, 278
 - social collaboration features of, 21
 - support for data masking in Hadoop, 232
- information
 - archiving of, 201–202
 - disposal of, 11
 - leveraging, optimizing, and securing, 4
 - lifecycle of. *See* information lifecycle management (ILM)
 - policies, 145–157
 - processes, 145, 146
 - standards, 145, 146
 - treating as an enterprise asset, 4
- information governance policy. *See* policies
- information lifecycle management (ILM), 201–205
 - compliance and, 11
 - defined, 201
 - in the EDM reference architecture, 11–12
 - information archiving in, 201–202
 - intersection points with data governance
 - document information policies in the business glossary that are implemented by ILM tools, 204
 - discover complete business objects that can be acted on efficiently by ILM tools, 204
 - legal holds and evidence collection (eDiscovery) in, 202
 - platform components, 201–203
 - records and retention management in, 202
 - subcategories, 12
 - test data management in, 202–203, 277
 - tools, 272, 273, 277, 287
 - vendors, 12
- information policy management, 145–157. *See also* policies
 - in the EDM reference architecture, 10

- subcategories, 12
- tasks related to data governance, 145–157
 - leverage data governance tools to monitor and report on compliance, 149
 - manage business rules, 147–148
 - manage data issues, 149–157
 - manage information policies, standards, and processes within the business glossary, 147
- vendors, 12

InfoTrellis Veriscope

- data quality dashboard, 122–123
- data quality metrics exposed in, 124

ISO 3166-1, 137, 181, 281

ISO 3166-2, 135, 249, 252, 281

ISO/IEC Metadata registry standard, 110

J

Jaro–Winkler matching function, 96

Java, 45, 46, 182, 210, 211, 260

Java Message Service (JMS), 234

Java Server Pages (JSPs), 46

JavaScript Object Notation (JSON), 275

JCL, 260

JIRA, 149

JSON, 275

K

key performance indicators (KPIs), 256

Know Your Customer (KYC) business process, 187–188

L

legal requirements, 11, 12, 190, 202

length frequencies, 63–64

lists of values, 11, 133. *See also* reference data

logical data model

- defined, 161

- integrating with metadata repository, 162

- pulling from data modeling tools, 37–38

logical data objects, 86–88, 223, 276

M

MapR, 12, 209, 290

MapReduce, 130, 211, 222, 230, 275

- defined, 210

- word-count program, 213–216

Marketo, 290

Mars Climate Orbiter, 6

master data. *See also* master data management

- change history, 128–129

- consistency in text analytics, 242–243

- duplicate management, 119–121

- duplicate suspect processing, 186

- enrichment rules, 112–113

- hierarchy management, 122–123

- metrics, 256, 257

- quality of, improving, 123–125

- record consolidation rules, 116–117

- record matching rules, 114–116

- snapshots, 127–128

- validation rules, 113–114

- workflows, 126–127, 186–187

master data management, 14, 109–131

- defined, 10, 109

- in the EDM reference architecture, 10–11

- integrating Hadoop with, 226–229

- moving data between hub and source systems, 172–173

- tasks related to data governance, 109–131

- compare snapshots of master data, 127–128

- define business terms consumed by the MDM hub, 109–111

- improve the quality of master data, 122–125

- integrate social media with MDM, 125–126

- manage duplicates, 119–121

- manage entity relationships, 111–112

- manage hierarchies, 122–123

- manage master data enrichment rules, 112–113

- manage master data validation rules, 113–114

- manage master data workflows, 126–127

- manage record consolidation rules, 116–117

- manage record matching rules, 114–116

- offload MDM tasks to Hadoop for faster processing, 129–130

- provide a history of changes to master data, 128–129

- view a list of outstanding data stewardship tasks, 117–118

- view the data stewardship dashboard, 121–122

tools, 267, 273, 277, 280, 287

- evaluation criteria for, 247

vendors, 11, 109

match and merge, 95–97, 165–166, 296

- matching attribute, 96
 - matching function, 96
 - MDM hubs, best-in-class, 166
 - metadata. *See also* metadata management
 - asset search support in repository, 57–58
 - business, 37
 - in column analysis, 59–61
 - customizing data lineage reporting in tool, 52–55
 - defined, 37
 - in the EDM reference architecture, 10
 - ingesting from relational databases, 40
 - integrating from legacy data sources, 42–43
 - linking business terms with column names in repository, 47–48
 - managing permissions in repository, 55–57
 - permissions, 56
 - providing detailed data lineage views from, 51–52
 - pulling from analytics tools, 47
 - pulling from big data sources, 50–51
 - pulling from data quality tools, 48–50
 - pulling from data warehouse appliances, 41–42
 - pulling from ETL tools, 43–44
 - pulling from reporting tools, 44–45
 - reflecting custom code in tool, 45–47
 - repository, 26, 43, 48, 55, 57, 162, 163, 174, 191, 198, 260, 265
 - flagging sensitive data in, 192
 - technical, 37
 - types of, 37
 - metadata management, 37–58
 - tasks related to data governance, 37–58
 - customize data lineage reporting, 52–54
 - ingest metadata from relational databases, 40–41
 - integrate metadata from legacy data sources, 42–43
 - link business terms with column names, 48
 - manage permissions in the metadata repository, 55–57
 - provide detailed views on data lineage, 51–52
 - pull in metadata from data warehouse appliances, 41–42
 - pull logical models from data modeling tools, 37–38
 - pull metadata from analytics tools, 47
 - pull metadata from big data sources, 50–51
 - pull metadata from data quality tools, 48–50
 - pull metadata from ETL tools, 43–44
 - pull metadata from reporting tools, 44–45
 - pull physical models from data modeling tools, 38–39
 - reflect custom code in the metadata tool, 45–46
 - support the search for assets in the metadata repository, 57–58
 - tools, 10, 259, 267, 270, 272, 277, 286, 291
 - Metaphone, 65
 - metrics
 - data quality, 20, 124, 178, 248
 - as evaluation criterion for data governance tools, 256–257
 - MDM hub, 121
 - visualization in reporting tool, 177, 178–179
 - MFS, 260
 - Microsoft
 - Azure, 290
 - Dynamics, 290
 - Excel, 17, 18, 33, 174, 202
 - Exchange, 201
 - Outlook, 33
 - SharePoint, 146, 147, 201
 - SQL Server, 9, 45
 - Visual Basic, 45, 211, 260, 261, 312
 - Word, 33, 146
 - migration projects, 163
 - models
 - enterprise, 287
 - logical. *See* logical data model
 - physical. *See* physical data model
 - MongoDB, 12, 231, 290
 - MQSeries, 260
 - MySQL, 193, 290
- ## N
- NACHA, 275
 - National Change of Address (NCOA), 94, 249, 251, 253
 - data quality policy related to, 249–256
 - natural-language processing technology, 242
 - Netezza, 41, 290
 - Netflix, 231
 - NetSuite, 278
 - non-youthful drivers example, 149–157
 - North American Industry Classification System (NAICS), 110, 111, 133, 140, 141, 142, 143, 178, 272
 - NoSQL (“Not Only SQL”), 12, 51, 209, 270, 290

- as alternative to relational databases, 231
- and big data, 13, 231
- database offerings, 12
- databases, governing data in, 231
- defined, 12
- in the EDM reference architecture, 12

Nuix eDiscovery, 202

null values, 69

O

ontologies, exposing in the metadata repository, 163

OpenText, 12

- enterprise content management capabilities for SAP, 287
- Records Management, 202
- solutions for archiving SAP data, 202

Oracle, 8, 11, 109, 193, 201, 216, 217, 290, 296

- CRM On Demand, 278
- customer hierarchy management in, 296
- Customer Hub, 296
- data governance capabilities of, 296
- Data Integrator, 43
- Data Relationship Governance, 296
- Data Relationship Management, 11, 296
 - customer hierarchy management in, 297
- Database 12c, 9, 11
- EBS, 278
- Endeca Information Discovery, 13, 239–240
- Enterprise Data Quality, 9, 296
 - address verification capabilities of, 296
 - audit capabilities of, 296
 - case management capabilities of, 296
 - dashboard capabilities of, 296
 - data profiling capabilities of, 296
 - data standardization capabilities of, 296
 - match and merge capabilities of, 296
- Exalytics In-Memory Machine, 11
- GoldenGate, 9, 170
- Higher Education Constituent Hub, 296
- master data management, analytical, 296
- PL/SQL, 260
- Product Hub, 296
- single view capabilities, 296
- Site Hub, 296
- Supplier Hub, 296
- TimesTen In-Memory Database, 8

Orchestra Networks, 11, 109, 279–283

EBX, 11, 235, 279–283

- AutoStoreDataModel, 166
- Business Glossary, 110–111, 283
- business glossary capabilities in, 110–111, 283
- collaborative workflow for master data on-boarding and approval in, 126–127
- comparing reference data snapshots in, 141–143
- data model for Third Parties entity in, 280–281
- data model’s Country attribute linking to reference data in, 282
- data modeling in, 110–111, 166, 280–282
- data stewardship interface in, 248, 282
- dynamic lookup from MDM domain to business glossary in, 283
- generating the data authoring interface in, 167
- legal entity hierarchy in, 122–123
- master data change history in, 128–129
- master data management capabilities in, 110–111, 282
- master data quality and execution time metrics in, 257
- matching policy in, 115–116
- multi-domain MDM in, 280, 282
- reference data management capabilities in, 110, 283
- workflows in, 279–280

MDM

- comparing master data snapshots in, 127–128

out-of-the-box (OOTB) attributes, 22

overlap analysis, 76–79, 163–164

P

pattern matching, 64

patterns distribution, 63–64

Pega Business Process Management, 11, 181

Pentaho, 11, 177

permissions

- managing in the metadata repository, 55–57
- metadata, in MetaCenter, 56

phone number standardization, 92

PHP, 260

physical data model

- defined, 162
- integrating with metadata repository, 162
- pulling from data modeling tools, 38–39

PitneyBowes, 94

- Pivotal. *See* EMC Pivotal
- policy, 20. *See also* information policy management
 data quality, 249–253
 defined, 145
 formulating, 4
 implemented by data masking and database monitoring tools, 195–199
 implemented by information lifecycle management tools, discovering, 204
 information governance, 250
 National Change of Address (NCOA), 249–253
 nomenclature, 145
 as part of information lifecycle management, 201
- predictive analytics, 12, 239
- predictive model, 241
- predictive variables, 242
- privacy. *See* data masking, data privacy, data security and privacy, sensitive data
- privileged users, monitoring database access by, 194
- privileges. *See* permissions
- process mapping, 294
- profiling. *See* data profiling
- protected health information, 243
- Protegrity, 11
- Python, 211, 312
- Q**
- quality. *See* data quality
- QlikView, 11, 177
- R**
- R, 11, 177
- Recommind Axcelerate eDiscovery Suite, 202
- record consolidation rules, 116–117
- record matching rules, 114–116
- records management, 12, 202
- redundancy, 72
- Reed Elsevier, 13, 258
- reference architecture, EDM, 7–14
- reference data. *See also* reference data management
 considerations for stream computing applications, 235
 defined, 11, 133
 examples of, 11, 133–134
 hierarchies, 139–140
 inventory of, 134–135
 leveraging for use by data integration tool, 173–174
 managing for Hadoop, 219–221
 mappings, 137–139
 master list of values for, 135–136
 metrics, 256
 moving data between hub and source systems, 172–173
 other terms for, 11
 snapshots of, 140–143
 text analytics, managing reference data with respect to, 243
 used by stream computing application, governing, 235–236
 workflows to manage, 294
- reference data management, 133–143
 in the EDM reference architecture, 11
 tasks related data governance to, 133–143
 agree on the master list of values for each code table, 135–136
 build an inventory of code tables, 134–135
 build and compare snapshots of reference data, 140–141
 build complex mappings between code values, 137–139
 build simple mappings between master values and related code tables, 137
 manage hierarchies of code values, 139–140
 visualize inter-temporal crosswalks between reference data snapshots, 141–143
 tools, 11, 264, 269, 280, 283, 294, 295, 296
- regulatory requirements, 11, 12
- relational database management systems (RDBMSs), 8–9, 40
- relationships
 custom, 23
 OOTB, 23
- replication. *See* data replication
- reporting. *See also* analytics and reporting
 in the EDM reference architecture, 11
 tools, 11, 294
 linking to business glossary from, 33–34
 pulling metadata from, 44–45
- Representational State Transfer (REST) API, 265, 273, 292
- retention management, 12, 202, 287
- Riak, 12, 231
- Riversand, 11, 109

roles

- assigning in InfoSphere Information Server, 55
- custom, 23–24
- OOTB, 23

rule, 112, 113, 114, 116. *See also* business rules

- business, 250
- data, 250
- data validation, 113–114
- information governance, 250

S

Sage, 290

Salesforce, 8, 13, 258, 278, 290

- CRM, 278

SAP, 8, 11, 109, 201, 278, 285–287, 290, 293

- BackOffice Associates data quality/data migration products for, 294

Business Suite, 287

BusinessObjects, 11, 177, 179

data quality capabilities of, 286

Data Services, 93, 285, 287

- Address Transform function, 93
- data connectivity in, 287

Enterprise MDM, 285

ERP Central Component, 48, 81

Extended ECM, 285

HANA, 11

- as centerpiece of SAP big data platform, 286
- collecting metadata from HANA, 286
- connectivity to Data Services for, 287
- developing data models for, 287
- in-memory database, 8, 286
- pushing data profiling operations to, 286

Information Steward, 10, 285

- address profiling in, 93–94
- bidirectional sharing of data models, 287
- business glossary capabilities in, 286
- business rule capabilities in, 286
- business term associated with column names in, 47–48

Cleansing Package module, 286

creating data quality scorecard in, 98–102

Data Insight module, 71, 73, 81, 93–94, 98–102, 286

data issue management capabilities in, 286

data profiling capabilities in, 286

Data Validation Advisor, 147–148

Define Redundancy Task in, 71–72

discovering dependencies in, 81–84

financial impact analysis of data quality, 104

metadata management capabilities in, 286

Metadata Management module, 286

Metapedia module, 34–35, 48, 286

modules in, 286

My Worklist module, 286

Redundancy view in, 72

Rule Editor, 148

searching for business terms in, 34–35

uniqueness profiling in, 73–74

Master Data Governance, 287

metadata management capabilities of, 286

NetWeaver Information Lifecycle Management, 287

NetWeaver Master Data Management, 287

OpenText content management capabilities for, 287

OpenText solutions for archiving data from, 202

PowerDesigner, 9, 37, 285, 287

business glossary in, 162

enterprise modeling capabilities of, 287

logical model in, 162, 287

Solutions for Enterprise Information Management, 285

Sybase, 193

Sybase Event Stream Processor, 12, 233

Sybase IQ, 11

SAS, 11, 109, 296

Analytics, 11, 177

analytics space market share, 296

Business Intelligence, 11, 177

Data Management, 10, 296

business glossary capabilities of, 296

data integration capabilities of, 296

data monitoring capabilities of, 296

data profiling capabilities of, 296

data quality capabilities of, 296

entity resolution capabilities of, 296

master data management capabilities of, 296

metadata capabilities of, 296

reference data management capabilities of, 296

DataFlux division, 296

enterprise data management tooling, 296

scanner in ASG-Rochade for, 47

Text Miner, 13, 239

scanners, 42, 43, 47, 50, 163, 174, 260, 270, 294

- scorecard. *See* data quality scorecard; data steward, console/dashboard for
- security. *See* data security, data security and privacy
- Semarchy, 11, 109
- Convergence for MDM
 - address data validation rule in, 116–117
 - customer record consolidation rule in, 116–117
 - data stewardship dashboard in, 121–122
 - definition of “Customer” in, 109–110
 - duplicate management workflow in, 119
 - entity relationships diagram in, 111–112
 - golden record based on consolidation rules in, 119–120
 - golden record created manually in, 120
 - master data enrichment rule in, 112–113
 - pending data stewardship tasks in, 117–118
 - record matching rules for customer data in, 114–115
- sensor data, 8, 235–236, 290
- sensitive data, 64–65. *See also* data masking, data security and privacy
 - discovering using data discovery tools, 190–192
 - examples, 193
 - flagging in the metadata repository, 192
 - hidden, 64–65, 190, 275
 - masking
 - in Hadoop, 232
 - in production environments, 193
 - in non-production environments, 193–194
- SEPA, 275
- single version of the truth, 10, 109, 289, 291
- single view capabilities, 296
- smart meter data, governance of, 170, 195–199, 204
- snapshots, 127–128, 140–141, 204, 264, 283
- social collaboration, 20–21
- social media, 8, 228, 239. *See also* blogs, Facebook, Twitter
 - integration with MDM, 125–126
- Software-as-a-Service, 290
- soundexes distribution, 65–66
- space values, 69
- SQL, 9, 10, 12, 59, 170, 193, 210, 211, 231, 260
- stakeholder analysis, 294
- standards, 20
- Stibo Systems, 11, 109
- stream computing, 233–237
 - aspects supported by data governance tools
 - govern business terms to be used by the stream computing application, 236
 - govern reference data to be used by the stream computing application, 235
 - use data profiling tools to understand a sample set of input data, 234
 - defined, 12, 233
 - in the EDM reference architecture, 12
 - rate of arrival of source data, 235
 - sinks/destinations, 234
 - sources of data, 234
 - temporal alignment/temporal offset of source data, 235
 - tools, 12, 233
- structured data
 - extracting additional relevant predictive variables not available in, 242
 - leveraging unstructured data to improve quality of sparsely populated, 241
- survivorship, 291
- SWIFT, 275
- Symantec, 12
 - Enterprise Vault, 202
- synchronization. *See* data quality dimensions
- ## T
- Tableau, 11, 33, 109, 177
 - Trillium’s data assurance dashboard exposed in, 178
 - visualizing business glossary categories in, 178–179
- Talend, 3, 11, 289–292
 - big data environment, 291
 - BPM, 181, 291, 292
 - collaborative stewardship enabled in, 186–187
 - detailed BPM workflows in, 186–187
 - Data Integration, 43, 170, 289, 291
 - business modeler in, 291
 - job designer in, 291
 - metadata manager in, 291
 - data management platform, 289
 - Data Quality, 97, 289, 291
 - data enrichment capabilities of, 291
 - data matching capabilities of, 291
 - data profiling capabilities of, 291
 - data standardization capabilities of, 291
 - postal validation capabilities of, 291
 - profiling data natively in Hive with, 222
 - survivorship capabilities of, 291

- data services platform, 289
 - Enterprise, 289
 - enterprise service bus (ESB), 289, 291
 - extracting metadata from Hadoop data sources in, 51
 - MDM, 289, 291
 - automatic generation of MapReduce code to support de-duplication, 130
 - data steward web user interface, 124–125
 - Data Stewardship Console, 118
 - de-duplication job executing natively in Hadoop from, 129–130
 - hierarchy browser in, 291–292
 - manual de-duplication (golden record) process in, 120–121
 - single version of the truth support in, 289, 291
 - viewing pending data stewardship tasks in, 118
 - Open Studio, 289
 - Studio
 - designing/refining record-matching strategy in, 95–97
 - Match Rule Settings feature, 97
 - unified environment for data integration, data quality, and MDM, 170–171
 - Unified Platform, 289–290
 - telematics, 290
 - Teradata, 41, 290
 - Aster, 11
 - test data management, 12, 202–203, 205, 272, 273–274, 277
 - text analytics, 239–243
 - and big data, 13
 - defined, 12, 239
 - in the EDM reference architecture, 12–13
 - example of data governance issues related to, 240–243
 - privacy requirements, adherence to in, 243
 - processes related to data governance, 241–243
 - adhere to privacy requirements, 243
 - define consistent definitions for key business terms, 242
 - ensure consistency in patient master data across facilities, 242–243
 - extract additional relevant predictive variables not available in structured data, 242
 - leverage unstructured data to improve the quality of sparsely populated structured data, 241
 - manage reference data, 243
 - tools, 13
 - Thomson Reuters, 13, 258
 - TIBCO StreamBase, 12, 233
 - time series analysis, 104–106
 - timeliness. *See* data quality dimensions
 - token, 189
 - total cost of ownership (TCO), 247–248
 - Trillium Software, 295
 - business glossary and reference data capabilities of, 295
 - data quality capabilities of, 295
 - TS Discovery
 - completeness dimension display in, 69
 - data assurance dashboard, 178
 - data discovery/data profiling capabilities of, 295
 - data quality scorecard and, 103
 - integration with Collibra Data Governance Center for data profiling, 265
 - Patterns Check attribute business rule in, 69–70
 - patterns distribution display in, 62–63
 - soundexes distribution display in, 65–66
 - time series analysis in, 106
 - Validate Shipped Date entity business rule in, 71
 - values distribution display in, 61–62
 - TS Quality, 10, 295
 - TS Quality on Demand, 13, 258
 - Twitter, 212, 226, 228, 231, 234, 290
- ## U
- unified schema prototyping, 163–166
 - uniqueness. *See* data quality dimensions
 - unstructured data, leveraging to improve quality of sparsely populated structured data, 241
 - U.S. Federal Reserve, 26, 137
- ## V
- values distribution, 61–62
 - variety, and big data, 13
 - velocity, and big data, 13
 - virtual joins that can be profiled, creating, 86–88
 - virtualization. *See* data virtualization
 - volume, and big data, 13
- ## W
- Waterline Data Science, Hadoop Data Inventory file browser in, 211–212

WebLogs Logical Data Object, 223, 276

workflows

- approval, 181–182, 248–249

- BPM, 186–187

- custom, 28, 149, 248, 265

- data issue, 149–157

- as evaluation criterion for data governance tools,
148–249

- of human and automated tasks, 106–107

- integrated with data quality and policy management,
296

- leveraging BPM capabilities, 181

- managing business terms, reference data, and other
artifacts with, 28–32, 294

- master data, leveraging BPM capabilities, 186–187

- for master data on-boarding and approval, 126–127

- OOTB, 28, 149, 248, 265

- with task related to suspicious records, 279–280

World Wide Web Consortium, 270

X

XML, 9, 17, 18, 47, 170, 275, 287

Y

Yahoo!, 12, 209

Z

ZyLAB eDiscovery and Production System, 202