Index

A

acceptable use, 10, 13, 145–146, 150–152, 154–162, 172–176
Acxiom, 82, 259
Adaptive, 249
agents
  author–agent agreement, 83
Alberta Human Rights Act (Canada), 152
Alberta Personal Information Protection Act (Canada), 152
Amazon, 200, 242, 243, 259, 260, 261, 273, 286
  big data platform, 284–285
Amazon Web Services, 243, 259, 273, 284–285
American Association of Advertising Agencies, 160
American Civil Liberties Union, 150
American Medical Association, 148
Anti-Terrorism, Crime and Security Act (U.K.), 134–135
Apache, 3, 45, 68, 72, 73, 238, 239–241, 242, 243, 244, 249, 263, 272, 276, 286
App Engine (Google), 285
Apple, 159, 173
archiving, 22, 40, 137–138, 139–140, 224, 225–226, 256–257, 267, 281
Arkovi, 257
Article 29 Data Protection Working Party, 174–176, 179–181
ASG-Rochade, 249
Asia-Pacific Economic Cooperation, 98
asset management, 191
Association of British Insurers, 21, 203
Association of National Advertisers (U.S.), 160
Aster Data Systems, 283
Atlas eDiscovery Process Management, 267
Aster MapReduce Appliance, 251, 284
Attensity, 253
audit information logging and reporting, 30, 36
Australia, 69–70, 136
Australian Energy Regulator, 69–70
Autonomy (HP), 278
Autonomy Consolidated Archive, 257
Autonomy eDiscovery, 258
Autonomy Records Manager, 258

B

Backupify, 257
banking, 56, 82, 92, 131, 149, 194
  industry scenarios, 23
base station data, 171
behavioral advertising, 156, 160–162
best practices
  to align with MDM, 118–131
  for big data privacy, 83–99
  for big data quality, 102–112
  for business process integration, 113–116
  for governing the big data lifecycle, 133–140
  for governing biometric big data, 200–204
  for governing human-generated data, 206–208
  for governing M2M big data, 170–192
  for governing smart meter data, 219–226
  for governing web and social media big data, 144–167
  to improve organizational alignment, 49–65
  metadata, 68–78
big data. *See also*
the individual big data types
  ability to expose tensions across functions, 5
  analytics, 41–42, 71, 119, 213, 230, 252–255, 266, 271, 274, 277, 283

archiving of, 22, 40, 133, 137–138, 139–140, 224, 225–226, 256–257, 281
at rest, 137, 242
business case for, 37–42
business glossary definitions, 68–71
characteristics of, 3
custodian, 62
defined, 4–5
differences of, 6
discovery, 247–248, 265, 269, 280, 281
disposal of, 34, 92, 140, 223
examples of, 3
importance of reference data, 124–125
industries leveraging, 9
information governance disciplines applied to, 12–15
integration, 113–116, 244, 264, 269, 273, 277, 279, 281, 283
integration with master data, 117–131
lifecycle management, 5, 9, 14–15, 30, 34, 133–140, 184, 208, 223–224, 233, 255–258, 267, 274, 281
maturity assessment, 29–36
monetization of, 4, 231, 234
as part of an information governance program, 4
platforms, 261–286
privacy, 5, 79–99, 254–255, 266–267, 282
quality, 42, 101–112, 152–154, 181–184, 228, 232, 248–249, 265, 270, 280–281, 286
reference architecture, 237–260
reporting, 252–255, 266, 271, 274, 277
responsibilities, 60–65
retention of, 11, 12, 22, 23, 92, 133, 139–140, 180, 185–186
roles, 54–6074
security, 254–255, 266–267, 282
semi-structured, 107
sensitive, 73–74, 84–86
and the "three Vs," 3
types of, 9, 10–12
unstructured, 77, 107, 123, 126–127, 130
vendors, 261–286
Big Data Analytics (RainStor), 257
Big Data Appliance (Oracle), 268, 272
Big Data Connectors, 268
big data governance
financial benefits, 31, 39
organizing for, 49–65
roadmap, 43–46
roles and responsibilities, 54–64
stakeholders, 31, 102–103
symbiosis with BPM, 115
introduction to, 3–7
framework, 9–28
Big Data Retention (RainStor), 257
big transaction data. *See* transaction data (big)
BigIndex, 263
BigQuery, 285
BigSheets, 263
biometrics big data, 12, 16, 18, 21, 59
acceptable use of, 200–202
customers and employees, 200–202
defined, 199
governance of, 199–204
privacy and, 200–202
regulations, 202–204
BMC Remedy, 250
BP oil spill, 137
Brandeis, Louis, 13, 79
British Standards Institute, 191
browser fingerprinting, 84
bulk data movement, 245
Business Analytics (Pentaho), 285
business case for big data governance, 37–42
business glossary, 68–71, 191, 216, 249
vs. data dictionary, 68
manager, 77
Business Glossary (Informatica), 249, 281
Business Intelligence Tools (Oracle), 271
business outcomes, 29, 31–32
business process integration, 9, 14, 113–116
business process management (BPM), 113, 286
business rules, 111
Business Suite (SAP), 273
BusinessObjects, 252
BusinessObjects Business Intelligence, 274
BusinessObjects Data Management, 248
BusinessObjects Data Services, 247, 273
BusinessObjects Information Steward, 250, 273

C

California privacy protection, 94–95
California Public Utility Commission, decision 11-07-056180–181
call detail records (CDRs), 19, 109–110, 122, 124, 194–195, 228, 242
Canada, 152
case studies
advanced condition monitoring (railroad), 184–187
advanced condition monitoring (THSRC), 38
Austrian law student, 90
behavioral advertising principles, 160–162
big claims transaction governance, 51–54
big data analytics, 41
California privacy protection, 94–95
California smart meter privacy, 180–181
centralized insurance claims database, 195–196
claims data stewardship, 112
clickstream data governance, 167
customer matching, 229–230
data breach at health insurer, 39
demand signal repositories (DSRs),196–198
E.U. geolocation guidelines, 174–175

E.U. smart meter privacy, 179–180
electricity outages in Italy, 183–184
eliminating duplicate CDRs, 194–195
employee-related geolocation data, 176
employers requesting Facebook information, 150
executive sponsorship, 64
FDA and off-label information, 147–148
facial recognition and social media, 200–202
financial institution roadmap, 45–46
genetic data regulations, 202–204
geospatial data governance, 42
Google bypassing Safari privacy settings, 159
Hadoop/data warehousing implementation, 76–77
Hadoop for data archiving, 138
hybrid big data, master data, data warehouse strategy, 252
improving health and wellness programs, 207–208
implementing lifecycle governance, 40–41
integrating email with customer MDM, 127–131
leveraging social media, 115
Location Privacy Protection Act, 173–174
managing social media perceptions, 146
Mars Climate Orbiter, 6
masking sensitive voice data, 206–207
mobile phone tracking, 80
monitoring realtime network performance, 109
neonatal intensive care unit, 110–111
Netflix investigated by FTC, 85
network master data, 120
network monitoring system, 139
oil and gas sensor data, 189–191
pharmacy benefits roadmap, 43–44
pregnant teenager, 85–86
product master data quality, 123–124
production volumetrics, 57
public transport awareness solution, 182
reducing patient readmission rate, 213
securing the smart grid, 188
sensor data in a school building, 108–109
smart meter governance, 220
social listening, 106
social media for claims investigation, 145
Stuxnet worm, 187–188
T-Mobile big data, 230
telco big data and master data, 121–122
telco retention requirements, 134–136
Tweeting travelers detained, 80
Twitter data accuracy, 232
Twitter data quality, 103–104, 232
U.K. cookie regulations, 157–158
Web commerce policies, 162–165
Cassandra, 243, 286
Cassandra File System (CassandraFS), 240, 243
Change Data Capture (CDC), 245
chief data officer, 60
churn, 19, 118, 119
Cisco, 243
claims administration/investigation. *See* insurance
claims analytics, 16, 70, 111–112
Clarabridge, 246
classification and metadata, 30, 36
clickstream analytics, 70, 76, 166–167
cloud, 258–259, 278, 282
Cloud (Informatica), 259, 282
Cloud Preservation, 257
Cloud Storage (Google), 285
Cloudera, 72, 242, 262, 272
Cloudera CDH, 272
Cloudera Manager, 272
Cognos, 252, 253, 264, 265, 266, 278
Cognos Business Viewpoint, 265
Cognos Consumer Insight, 253, 264, 266
Collective Intellect, 253, 271
Comcast, 278
communications service providers. *See* telecommunications
complex event processing, 242
Compliance, Governance and Oversight Council, 140
Composite Software, 246
Compute Engine (Google), 285
confidence intervals, 152
consent, 158–159
Constant Contact, 243
consumer packaged goods, 58, 58–59, 114, 119, 124, 194, 197
industry scenario, 22
Content Analytics (IBM), 264
Content and Predictive Analytics for Healthcare (IBM), 264
Content Collector (IBM), 267
Content Manager, 267
Converged Cloud, 259, 278–279
cookies, 154–162
core disciplines, 31
Coremetrics, 253, 266
Couchbase, 243
CouchDB, 243
Council of Better Business Bureaus (U.S.), 160
Customer Hub, 270
customer experience analytics, 246
customer relationship management (CRM), 127
customer service, 39
function scenario, 24–5

D

data
amount of in digital universe, 3
architecture, 30, 35
federation, 246
lineage, 75
master, 117
personal, defined, 81

quality, 9, 12, 13–14, 167, 223. *See also* data quality management
reference, 123
replication, 245–246
trusted, 68, 283
types of, 3
virtualization, 24
data architect, 78
Data Archive (Informatica), 256, 280, 281
data brokers, 81–82, 93, 259
data collectors, 81
data dictionary, vs. business glossary, 68
data discovery tools, 86, 241, 247–248, 281
Data Explorer, 247, 280, 282
Data Governance Director, 250
Data Integration (Pentaho), 285–286
Data Integration Studio, 283
Data Integrator Application Adapter for Hadoop, 269
Data Integrator Enterprise Edition 11g, 269
data lineage administrator, 78
data profiling tools, 74, 247–248
data masking, 206–207, 255
Data Masking (Informatica), 255, 280, 282
Data Protection Commissioner (Ireland), 203
Data Quality (Informatica), 248, 280
data quality management, 3, 13, 30, 33–34, 41, 120, 232, 248–249
defined, 101
difference between traditional and big data programs, 101–102
Data Replication (Informatica), 246, 280
data risk management, 29, 33
data scientist, 59–60, 78
Data Services (Informatica), 246, 280
data steward, 78, 102, 111–112
asset, 62
big data, 54–59
big transaction, 58–59
biometric, 59
chief, 62
customer, 62
human-generated, 59
M2M, 56
materials, 62
reference, 125
web and social media, 55–56
Data Subset (Informatica), 258, 281
data users, 82
data warehousing, 24, 225, 241, 246, 251, 265, 271, 274, 277
Database 11g, 243, 251, 268, 271
Database Appliance (Oracle), 271
databases, 243–244, 255, 264, 268–269, 273, 276
DataFlux, 282
DataFlux Data Management Platform, 283
DataFlux Data Management Studio, 247, 248
DataFlux Marketplace, 259
DATAllegro, 277
DataStax, 243
DB2, 243, 251, 263, 264
DB2 Analytics Accelerator for z/OS, 244
DB2 for z/OS, 244, 264
de-identifying private data, 84
Dell Parallel Data Warehouse Appliance, 277
demand signal repositories (DSRs), 22, 58, 59, 196–198
Denodo, 246
Deutsch, Tom, 15
Direct Connector for HDFS, 244, 269
Direct Marketing Association (U.S.), 160
DNA mapping, 21, 202
DNA splicing, 21, 202
Documentum, 284
Documentum Records Manager, 258
Dynamic Data Masking (Informatica), 282
Dynamics CRM Online, 282
Dynamics GP, 282
Dynamics NAV, 282
DynamoDB, 243, 284–285

E

eBay, 72
E-Business Suite (Oracle), 282
eDiscovery, 140, 136, 137, 139–140, 256, 258, 267, 278
eDiscovery Analyzer, 264
eDiscovery Manager, 258
eDiscovery Solution, 267
education, 108–109
industry scenario, 24
Elastic Compute Cloud, 284–285
Elastic Map Reduce, 259, 284–285
EMC, 77, 242, 250, 251, 257, 258, 260, 261, 286
big data platform, 284
enablers, 31
EnCase eDiscovery, 258
encryption, 255
Endeca Information Discovery, 246, 269
Enterprise Data Quality family (Oracle), 248, 270
Enterprise Data Quality Match and Merge, 270
Enterprise Data Quality Parsing and Standardization, 270
Enterprise Data Quality Product Data Match and Merge, 270
Enterprise Data Quality Product Data Parsing and Standardization, 270
Enterprise Data Quality Profile and Audit, 247, 270
Enterprise Linux, 272
Enterprise Miner, 283
Enterprise Records, 258, 267
Enterprise Vault, 257, 258
Erado, 257
EURODAC, 199
European Convention on Human Rights, Article 8, 87
European Data Protection Authority, 97
European Data Protection Directive, 81, 87, 88, 96, 174–75
European Directive 2002/58/EC, 157–158
European Directive 2009/136/EC, 158
European Directive 95/46/EC, 158, 179–180

European Economic Area, 98
European privacy regulations, 87–90
European Union Article 29 Data Protection Working Party, 174–176
European Union Charter of Fundamental Rights, Article 8, 87
European Union General Data Protection Regulation, 203
evidence collection, 258
Exadata Database Machine, 271
Exalytics In-Memory Machine, 251, 271
Excel, 225, 258, 277, 286, 286
executive sponsorship, 63
experimentation, 15
Extract, Load, and Transform (ELT), 245, 269, 281
Extract, Transform, and Load (ETL), 245, 269, 273, 277, 280, 286

F

Facebook, 10, 13, 14, 17, 20, 26–27, 35, 55, 80, 82, 115, 116, 118, 131, 143, 145, 146, 149, 150–151, 152, 153, 200, 245, 250, 280, 281
  Statement of Rights and Responsibilities, 152
Facebook Platform Policies, 18, 125–126
facial recognition technology, 18, 200–202
Fast Clone, 246, 280
FileNet Content Manager, 267
financial services, 45–46, 81, 139, 208, 251–252, 263, 271. *See also* banking, insurance
fleet monitoring, 182
Flume, 242
Foursquare, 170
framework for big data governance, 9–28
France, 135
fraud analytics, 75, 252
function scenarios. *See* individual functions

G

General Data Protection Regulation (European Union), 88–90
General Parallel File System (GPFS), 240, 264
genetic data, 21,202–204
geolocation data, 11, 13, 20, 79, 82, 121–124, 170–172, 252
  acceptable use of, 172–176
  defined, 170
  regulations regarding use of, 173–175
  types of, 170–172
geospatial data, 21, 42, 73, 182
German Data Access and Digital Signature Authentication Law (Germany), 137
Germany, 38, 135, 137, 152
Global Data Synchronization (GDS), 197
Global Retention Policy (IBM), 255, 266, 267
global positioning system (GPS) data, 79, 80, 122, 169, 171, 182
glossary, business. *See* business glossary
GlusterFS, 240
goals, 31
GoldenGate, 246, 269
GoldMine, 282
Google, 82, 159, 160, 173, 239, 248, 259, 260, 261, 286
  big data platform, 285
Google Cloud Platform, 259, 285
Google Search Appliance, 248
government, 63–64
Graduate Management Admission Council, 92
Greenplum, 251, 286
Greenplum Chorus, 284
Greenplum Data Computing Appliance, 284
Greenplum Database, 284
Greenplum HD, 284
Greenplum Unified Analytics Platform, 284
Groupon, 278
Guidance Software, 258
Gzip, 138, 257

H

Hadoop, 3, 13, 33, 35, 68, 71–73, 74, 75–77, 78, 109, 137–138, 193, 233, 237, 238, 245, 246, 248, 249, 251, 252, 253, 257, 259, 262, 263, 264, 268, 273, 274, 277, 278, 279, 280, 281, 282, 284, 285, 286
  defined, 239
  distributions, 241–242, 243, 260, 262–263, 272, 276, 284
  security shortcoming, 254
  subprojects and technologies, 239–242
Hadoop Common, 239, 240
Hadoop Distributed File System (HDFS), 71–73, 239, 240, 254, 264, 269, 271, 280, 283
HANA, 243, 273, 273, 274, 275
Handbook of Data Intensive Computing, 254
Hanzo Archives, 257
Harvard Law Review, 13, 79
HBase, 240, 243, 263
HCatalog, 73, 249
HDFS READ, 244
healthcare/health plans, 58, 77, 98, 107, 110–111, 111–112, 114, 124, 194, 207–208, 263, 264
  big claims transaction data, 71
  big data governance, 213–217
  claim codes, 50–54
  industry scenarios, 15–17
  and RFID data, 178
  and social media, 146, 148
Higher Education Constituent Hub, 270
Hive, 240, 241, 249, 284, 286
Hive Open Database Connectivity driver (Microsoft), 277
HiveSQL, 241
HortonWorks, 242, 276
Howl. *See* HCatalog
HP, 251, 257, 258, 259, 260, 262, 273, 280, 281, 286
  big data platform, 278–279
  big data platform (diagram), 262
HParser, 244, 280, 281
human-generated big data, 12, 17, 24–25, 59

enriching MDM with, 208–
examples of, 205
governance of, 205–209
managing the lifecycle of, 207
masking policies, 206–207
using to improve quality of structured data, 207–208
human resources, 115
function scenario, 26–27
use of geolocation data, 175–176
use of social media, 150–152
Hyperion Data Relationship Management, 270
Hyperion Performance Management Applications, 271
HyperSQL Database, 244

I

IBM, 29, 225, 240, 242, 243, 244, 245, 246, 247, 248, 249, 250, 249, 251, 252, 253, 255, 256, 257, 258, 260, 261, 273, 286
big data platform, 262–267
big data platform (diagram), 262
IBM Information Governance Council Maturity Model, 29–31
ICD healthcare codes, 50–54, 124
impact analysis, 75–77
Imperva, 255
in-memory databases, 243
Incubator, 73, 249
India, 95, 136
industry scenarios. *See* individual industry entries
Informatica, 245, 246, 247, 248, 249, 250, 251, 255, 256, 258, 259, 260, 261, 278, 286
big data platform, 279–282
big data platform (diagram), 279
Information and Privacy Commissioner of Ontario (Canada),180–181
information governance
charter, 32
council,
disciplines, 9, 13–15
maturity assessment, 29–26
maturity categories, 29–31
maturity model, 29–31
officer, 60–6161
responsibilities, 60–65
roadmap, 43
roles, 54–60
working group, 61
Information Lifecycle Governance Leader Reference Guide, 258
information lifecycle management, 5, 9, 14–15, 30, 34, 133–140, 184, 208, 223–224, 233, 255–258, 267, 274, 281
Information Management (SAS), 282, 283
Information Management System (IMS), 244, 264
information policy management, 249–250
information security and privacy, 30, 35
function scenario, 27
information services/technology, 40, 124, 133, 137, 166–167
function scenario, 25
Information Technology Act of 2009 (India), 136
Information Technology Rules (India), 95
Informix, 264
InfoSphere BigInsights, 244, 262, 263, 264
InfoSphere Business Glossary, 249, 265
InfoSphere Data Replication, 246, 264
InfoSphere DataStage, 263, 264
InfoSphere Discovery, 247, 264, 265
InfoSphere Federation Server, 246, 264
InfoSphere Guardium, 255, 266, 267
InfoSphere Information Analyzer, 247, 264, 265
InfoSphere Master Data Management, 250, 265
InfoSphere Master Data Management Custom Domain Stand Alone, 265
InfoSphere Master Data Management Reference Data Management Hub, 265
InfoSphere Metadata Manager, 249
InfoSphere Metadata Workbench, 265
InfoSphere Optim Data Growth Solution, 256
InfoSphere Optim Data Masking Solution, 255, 266
InfoSphere Optim Test Data Management Solution, 258, 267
InfoSphere QualityStage, 248, 265
InfoSphere Streams, 242, 263, 264
InfoSphere Warehouse, 251, 263, 265
Initiate, 265
Institute of Asset Management, 191
insurance, 77, 114, 124, 131, 181, 194, 195–196, 203
big claims transaction data, 58, 70
claims investigation using social media, 20, 145
industry scenarios, 20–21
and social media, 145–146
underwriting (pricing), 69, 145–146
Integrated Automated Fingerprint Identification System (IAFIS), 199
intelligence, 59, 199, 242
International Mobile Equipment Identity (IMEI), 134
International Mobile Subscriber Identity (IMSI), 134
"internet of things," 169
Internet service providers, 94
Ireland, 202–203
ISO/IEC 11179-4, 69
Italian Electrical Authority, 70, 182
Italy, 70, 135, 183–184, 255

J

Jaql, 263
JAQL_SUBMIT, 244
Java Virtual Machine, 272
JD Edwards, 282
job candidates, screening, 26–27

K

Kalido, 249
Klout, 277

L
law enforcement, 59, 134, 180, 199
legacy databases, 244
legal, 136, 137, 149–150, 199, 202, 258
License Agreement for Provision of Internet Services (India), 136
life sciences, 55, 147–148
lifecycle management. *See* information lifecycle management
LinkedIn, 10, 26, 145, 149, 151, 245, 280
Lithium, 253
Loader for Hadoop, 244, 268
location-based services, 119, 120
longitudinal data warehouses, 24
loyalty programs, 17, 19, 201, 202
Lucene, 263
LZO, 138, 257

M
machine-to-machine (M2M) big data, 11–12, 15, 17, 18, 20, 21, 22, 23, 25, 26–27, 56–57, 253, 275
  defined, 11–12, 169
  governance of, 169–192
  metadata for, 181–184
  quality of, 181–184
  retention policies, 184
MapR, 242
MapReduce, 36, 233, 239, 240, 241, 244, 251, 268, 271, 273, 276, 281, 283, 284
marketing
  function scenario, 25–26
Mars Climate Orbiter, 6
master data, 33, 35, 117–131, 216, 229–231
  domain, 118
  integration, 14, 117–131
  network, 119–120
  quality, 119–120, 121–124, 184–187
Master Data Governance (SAP), 273
master data management (MDM), 9, 40, 55, 60, 113, 115, 116, 117, 118, 127–131, 250–251, 265, 270–271, 273, 281, 286
Master Data Management (Informatica), 250, 281
Master Data Management Suite (Oracle), 250, 270
maturity assessment, 29–36
maturity model, 29–31
MDM for Product Information Management (IBM), 265
Medium Access Control (MAC) address, 171
metadata, 4, 9, 10, 13, 30, 67–78, 86, 167, 184, 265, 281
  administrator, 77–78
  best practices for big data governance, 68–78
  for big data, 249–250
  big data challenge for, 67
  defined, 67
  as enabler for Hadoop, 71–73
  and M2M data, 181–184
Metadata Manager (Informatica), 249, 281
Metadata Server (SAS), 283
metrics, 111–112, 165–167
Microsoft, 77, 127, 225, 241, 242, 243, 244, 256, 260, 262, 267, 282, 284, 285, 286
  big data platform, 275–278
  big data platform (diagram), 276
Microsoft System Center, 276
MicroStrategy, 278
MongoDB, 243, 286
Monster.com, 149
Morgan-Stanley and Co., 208
MySQL, 244, 285

N
Netezza, 244, 251, 263, 264, 265, 286
Netflix, 85, 243
NetFlow, 139
NetSuite, 282
NetWeaver Business Warehouse, 274
NetWeaver Information Lifecycle Management, 274
NetWeaver Master Data Management, 250, 273, 273
NetWeaver Process Orchestration, 274
network master data, 119–120, 229–230
network monitoring application, 139
New York Times, 85
Nielsen, 82, 182
North American Electricity Reliability Corporation-Critical Infrastructure Protection (NERC-CIP), 188
Northeastern University, 213
NoSQL, 35, 78, 137, 240, 243, 251, 285, 286
NoSQL Database (Oracle), 269
NoSQL Database Community Edition (Oracle), 272
NoSQL databases, 243
Nuix eDiscovery, 258

O
Obama administration, 94
Office of the Information and Privacy Commission (Canada), 152
oil and gas, 42, 57, 115, 137, 187
  governance of sensor data, 189–191
  industry scenarios, 21–22
Open Studio, 286
OpenPages, 250, 265
OpenText, 256
OpenText Records Management, 258
optimization, 4
Oracle, 243, 244, 246, 247, 248, 250, 251, 253, 256, 260, 261, 282, 285, 286
  big data platform, 268–72
  big data platform (diagram), 268
organization, 9, 13, 49–65
organizational structures and awareness, 29, 32
outsourcing, 96

P
P20 data warehouses, 24
PageFreezer, 257
ParAccel, 251
Payment Card Industry Data Security Standard, 207
PCI Security Standards Council, 207
Pentaho, 245, 252, 260, 261

big data platform, 285–286
performance management, 119–120
Persistent Data Masking (Informatica), 282
Personal Data Protection Act (Singapore), 95
personally identifiable information (PII), 73, 84, 86, 165, 176, 231, 233, 247
pharmaceuticals, 77, 147–148, 186, 278
pharmacy, 43–44, 194
Pig, 240, 241, 278, 281, 283, 284, 286
Pig Latin, 73, 241
plant operation
function scenario, 26
policies, 4, 10, 12, 14, 29, 33, 116, 133, 206–207, 214, 265
information policy management, 249–250
social media, 150–152, 162–165
Point of Sale Transaction Logs (POS TLogs), 22, 123, 197
PostreSQL, 244
Power View,
PowerCenter, 245, 280, 281
PowerExchange, 245, 280
PowerPivot, 277
Prediction API (Google), 285
predictive analytics, 16, 19, 34, 246, 252–253, 252–253
predictive maintenance models, 119
preventive maintenance, 114, 119, 184–187
privacy, 5, 9, 13, 15, 19, 20, 21, 23, 26–27, 35, 73, 121–124, 266–267, 282
best practices for big data, 83–99
big data, 79–99
of biometric data, 200–202
European regulations, 86–90
FTC proposed framework for, 84
India regulations, 95–96
laws and regulations, 86–96, 216–217, 254–255
location, 173–175
network data, 231
RFID, 176–179
right of, 13, 79
Singapore regulations, 95
smart meter, 179–181
and social media, 125–126, 146, 148, 149, 151, 152, 156, 157, 159, 160, 161, 165, 167
South Africa regulations, 95
United States regulations, 90–95
Privacy Act (Australia), 203
privacy impact assessment, 84, 176, 203
privileged users, 98–99
process map, 115
Product Hub, 270
Professional Petroleum Data Management Association, 191
Protection of Personal Information Act (South Africa), 95
Publicly Available Specification (PAS) 55, 191

Q

quality. *See* big data, quality; data quality management
quality requirements matrix, 103–104
QlikView, 252

R

R, 252, 272
R Connector for Hadoop, 271
RACI matrix, 49–54
railroads
industry scenario, 23–24
RainStor, 138, 257
Recommind Axcelerate eDiscovery Suite, 258
Records and Retention Management (IBM), 267
records management, 257
Reed Elsevier, 82, 259
reference architecture for big data, 237–260
reference data, 124–125, 217, 251, 265
regulatory requirements, 34, 86–96, 145–150
Relational Database Service (Amazon RDS), 284
relational databases, 243–244
retail, 41–42, 85, 106, 115, 116, 118, 178, 274
industry scenarios, 17–19
retention policies and schedules, 11, 12, 22, 23, 92, 133, 134–136, 139–140, 180, 184, 185–186, 191, 202, 257–258, 267
RFID, 11, 18, 56, 84, 169, 171, 176–179, 181
Riak, 243
risk, 23, 75. *See also* data risk management
roles and responsibilities, big data governance, 54–64
RSA Archer eGRC, 250
RulePoint, 280

S

Safe Harbor Framework, 97–98
Safe Harbor Principles, 96–98
SalesForce, 259, 282
SalesForce Radian6, 253
San Diego Gas and Electric, 180–181
SAP, 243, 247, 248, 250, 282, 286
big data platform, 272–275
big data platform (diagram), 272
SAS, 246, 247, 248, 252, 253, 259, 260, 261, 286
big data platform, 282–283
SAS/Access Interface to Hadoop, 283
scalability, 72
Schedule Management (IBM), 255, 266, 267
search indexes, 77
security breach notification laws, 94
Security Information and Event Management (SEIM), 27, 255
seismic data, 21
sensitive big data
flagging in the metadata repository, 86
identifying, 84–86
and privileged users, 98–99
tagging in the business glossary, 73–74
sensor data, 14, 20, 22, 23–24, 38–29, 56, 108–109, 115, 124–125, 137, 138, 169, 185, 189–191

sentiment analysis, 15, 25–26, 105, 119, 123, 250
Service Target Performance Incentive Scheme (Australia), 69–70
Siemens, 187–188
Simple Storage Service, 284–285
single point of failure, 72
Singapore, 95
Site Hub, 270
"skip tracking," 149
Smart Analytics System (IBM), 251, 263, 265
Smart Archive (IBM), 257, 267
Smart Meter Analytics, 275
smart meters, 40–41, 56, 69–70, 137–138, 179–181, 219–226
"social graph," 172
social listening, 105–106, 232, 253
"social MDM," 250
Social Media Analytics (SAS), 253
social media big data, 10–11, 13, 15, 17, 18, 19, 20, 23, 24, 25–26, 26–27, 40, 55–56, 86, 88, 92, 94, 115, 118, 125–126, 131, 139–140, 143–167, 152, 170, 200, 202, 229, 230, 232, 250, 253, 256, 257, 266
  acceptable use of data from, 145–146, 150–152
  governance of, 143–167
  and privacy, 125–126, 146, 148, 149, 151, 152, 156, 157, 159, 160, 161, 165, 167
  quality of, 152–154
solidDB, 243, 264
Song-Beverly Credit Card Act (California), 95
SourceOne, 257, 284
South Africa, 95
Spitz, Malte, 80
Splunk, 253
SPSS, 252, 266
SPSS Text Analytics for Surveys, 264
SQL Server, 243, 244, 276, 285
SQL Server Analysis Services, 277–278
SQL Server Integration Services, 277
SQL Server Parallel Data Warehouse, 277
SQL Server Reporting Services, 277
Sqoop, 243
"stalking apps," 174
steward. *See* data steward
stewardship, 29, 32. *See also* data stewards
StreamBase, 243
streaming analytics, 102, 107–111, 111–112, 124–125, 138, 182, 195, 242, 248, 252, 263, 273
Stuxnet worm, 187–188
Supervisory Control and Data Acquisition (SCADA), 12, 187–191
Supplier Hub, 270
supporting disciplines, 31
Sybase, 273
Sybase Adaptive Server Enterprise, 273
Sybase Event, 243
Sybase Event Stream Processor, 273
Sybase IQ, 251, 273
Sybase SQL Anywhere, 273
Symantec, 257, 258

T

T-Mobile, 230
Tableau, 252, 278
Taiwan High Speed Rail Corporation (THSRC), 37–39
Talend, 245, 260, 261
  big data platform, 286
Talend Enterprise, 286
Talend Unified Platform, 286
telecommunications, 56, 58, 59, 73, 79, 80, 98, 109–110, 119–120, 121–124, 135, 181, 194–195, 242
  big data governance for, 227–233
  big data lifecycle management for, 233
  big data quality for, 232
  big data types, 228–229
  data privacy for, 231–232
  industry scenarios, 19–20
  integrating big data with master data, 229–231
telematics, 20, 181
Teradata, 251, 260, 261, 286
  big data platform, 283–284
test data management, 258, 267
text analytics, 23, 25, 52, 127–131, 246, 263, 264
Text Analytics (SAS), 246
Thailand, 136
Thomson Reuters, 82, 259
TimesTen In-Memory Database, 243, 269, 272
Title VII, U.S. Civil Rights Act, 151
Tivoli Security Operations Manager, 255, 266, 267
traffic management, 182, 232
transaction data (big), 12, 16, 19, 21, 22, 49–54, 58–59, 70–71, 193–198, 257
  examples of, 193–194
  governance of, 193–198
transportation, 23–24, 75–77, 98, 182
triangulation, 122, 171
Trillium Software, 247, 259
TS Discovery, 247
TS Quality, 248
TS Quality on Demand, 259
Twitter, 10, 13, 25–26, 35, 40, 80, 82, 102–107, 108, 118, 131, 145, 146, 149, 151, 153, 232, 243, 245, 251, 253, 269, 278, 280

U

U.K. Department of Culture, Media and Sport, 156–157
U.K. Financial Services Authority, 208
U.K. Home Office, 134
U.K. Information Commissioner, 158–159
U.K. Privacy and Electronic Communications Regulations 2003, 157–158
U.S. Age Discrimination in Employment Act, 151
U.S. Americans with Disabilities Act, 152
U.S. Cable Communications Act, 91, 173
U.S. Children's Online Privacy Protection Act, 91

U.S. Civil Rights Act, 151
U.S. Consumer Privacy Bill of Rights, 94
U.S. Department of Commerce, 94, 96, 97
U.S. Department of Homeland Security, 80
U.S. Department of Justice, 173
U.S. Electronic Communications Privacy Act, 91, 173
U.S. Equal Employment Opportunity Commission, 141
U.S. Fair Credit Reporting Act, 23, 91, 149
U.S. Fair Debt Collection Practices Act, 23, 149
U.S. Family Educational Rights and Privacy Act, 91
U.S. Federal Bureau of Investigation, 199
U.S. Federal Deposit Insurance Corporation, 23
U.S. Federal Railroad Administration, 186
U.S. Federal Trade Commission, 18, 84–85, 91, 92, 161, 201
U.S. Federal Trade Commission Act, 91
U.S. Financial Industry Regulatory Authority, 139, 256
U.S. Food and Drug Administration, 55, 147–148
U.S. Genetic Information Non-Discrimination Act, 152, 203
U.S. Gramm-Leach-Bliley Act, 91
U.S. Health Insurance Portability and Accountability Act (HIPAA), 39, 91, 146, 148, 214, 216
U.S. Location Privacy Protection Act of 2011, 173–74
U.S. Pregnancy Discrimination Act, 152
U.S. Privacy Act, 91
U.S. privacy regulations, 90–95
U.S. Securities and Exchange Commission, 23, 208
U.S. Stored Communications Act, 139
U.S. Telephone Consumer Protection Act, 91
U.S. Video Privacy Protection Act, 91
Ultra Messaging, 280
Unica, 266
"unique visitor," 74
United Nations General Assembly, 86
Universal Declaration of Human Rights, Article 12, 86–87
Universal Product Codes (UPCs), 123
unstructured data, leveraging, 207–208, 214–215
utilities, 40–41, 56, 69–70, 99, 137–138, 169, 181–182, 183–184, 187–188, 194
  big data governance, 219–226
  industry scenario, 17

V
value, 3, 193
variety, 3, 193, 217
velocity, 3, 193
Vertica, 251, 278
visualization and reporting tools, 252
Vivisimo, 248
voice analytics, 206
volume, 3, 193

W
Warren, Samuel, 13, 79
Watson, 263, 266
web and social media big data, 10–11, 13, 15, 17, 18, 19, 20, 23, 24, 25–26, 26–27, 40, 55–56, 92, 123–124
  governance of, 143–167
web analytics, 70, 154–155, 166–167
web tracking devices, 154–162
Williams-Sonoma, 94–95
Windows Azure, 276, 278
wireless networks (WiFi), 169, 171

X
xDRs, 109–110

Y
Yahoo!, 239
Yelp, 200
YouTube, 146

Z
ZyLAB eDiscovery and Protection System, 258
Zynga, 278