

INDEX

- Acts and legislation. *See* Legislation and acts
- Advertising
ad targeting, 16, 102, 105
geospatial intelligence and, 105–106, 119–120
industry examples, 51–57, 102
- Affiliate marketing, 30–31
- Algorithm, definition, 83, 100
- Algorithmic trading, 40–44, 71, 91
- Alvarez, Chuck, 117–118
- Analysts. *See* Data analysts
- Analytics. *See also* Business intelligence (BI)
analytic applications, 125
being analytical (*see* Institutionalizing analytics)
business intelligence versus, 14, 49, 70
communication cycle, 108–109, 142
crowdsourcing analytics, 76–77
data discovery, 65–66
definition, 83, 103
delivering results, 101–102, 108–110 (*see also* Data scientists; Visualizing data)
description of, 5–6, 8
firewalls and, 73, 77–80
geospatial intelligence, 104–106, 119–120
health care analytics, 16, 46–47
holistic, 47–49, 139, 140–141
in-database analytics, 93, 95
institutionalizing (*see* Institutionalizing analytics)
marketing analytics, 44 (*see also* Marketing)
on mobile devices, 73–75
MPP appliance analytics, 93
new versus old, 7, 49, 64, 128–129
organizational structure and, 146–149
platforms (*see* Platforms)
predictive (*see* Predictive analytics)
quants for Big Data (*see* Data scientists)
risk analytics, 37–40, 42–43
self-service approach, 65–66, 73
software as a service (SaaS), 56, 72–73
spectrum of, 14, 99–100
tools for, 85, 124–125
value of, 12–15, 17–18, 102
volume and, 9
Web Analytics 2.0 (Kaushik), 20, 136
- Anonymized data, 77, 164
- Apache Hadoop. *See* Hadoop
- Appliances
data warehouse appliances, 90, 91
definition, 83
GPU appliances, 98
MPP appliances, 85, 92–93
- Argyros, Tasso, 69
- ASCII codes in binary, 86
- Avro, definition, 83
- Awadallah, Amr, 61–62
- Bandwidth, 97
- Batch, definition, 83
- Beard, Randal, 52–53, 54, 55, 56–57
- Behavioral data
advertising impact, 54–55
behavioral targeting, 101–102
credit scoring, 39
cross-channel marketing, 28–29, 31, 33
customer insights, 71, 77, 79, 117
emergence of, 25, 27
fraud prevention, 35, 37–38, 44
recommendation engines, 62, 71
technology and behavioral sciences, 129
- BI. *See* Business intelligence (BI)

- Big Data. *See also* Data
 analytics (*see* Analytics)
 business intelligence versus,
 5, 19–20
 definition, 5, 9, 83
 emergence of, xii–xiv, 1–2, 7
 interdependence of, 3–4 (*see also*
 Industry examples)
 processing of, 90–91, 96–97
 recommended resources, 175
 storage, 89–90, 92, 96 (*see also* HDFS)
- Big Insights, definition, 83
- Binary code, 86
- Bit, 86
- Blue Ocean Strategy*
 (Kim & Mauborgne), 71
- Botkin, David, 117
- Brocklebank, John, 72
- Burns, Nate, 50
- Business analytics. *See* Analytics;
 Predictive analytics
- Business intelligence (BI)
 advanced analytics versus, 14,
 49, 70
 Big Data versus, 5, 19–20
 “land and expand,” 65
- Bust out fraud case, 37
- Byte, 86
- C programming language, 98
- Cafarella, Mike, 61
- Cassandra, definition, 83
- Centralized analytics, 146, 148
- Champagne, David, 67, 128
- Children’s Online Protection Privacy
 Act (2000), 155
- CI (contextual integrity), 160. *See also*
 Privacy
- Closure, definition, 83
- Cloud computing
 Big Data and, 69–70
 definition, 83
 private cloud, 73
 public cloud, 92
 Whirr, 86
- Cloudera, 61–62, 67, 83
- Clusters. *See* Grids
- Collaboration
 consumption of analytics, 108
 convergence and, 131
 fear of, 77, 80, 147
 innovation management, 81
 interdependence, 3
 organizational structure, 146
 trans-firewall analytics, 77–80
- Collaborative planning, forecasting, and
 replenishment (CPFR), 77
- Columnar database, 83, 84, 89, 96
- Communication cycle, 108–109, 142
- Comparative effectiveness research
 (CER), 47
- Complex event processing (CEP),
 definition, 84
- Compute unified device architecture
 (CUDA), 38
- Conferences and tradeshows, 176
- Consumer Privacy Bill of Rights,
 154, 156
- Consumer Privacy Protection Act
 of 2011, 155
- Consumer products, 57–58
- Contextual integrity (CI), 160. *See also*
 Privacy
- Conway, Drew, 3, 134–135
- Cookies, and privacy, 157, 158, 159
- Cox, Donna, 112–113
- CPU bound, 97
- Credit card fraud, 34–37
- Credit risk analytics, 38–39, 43
- Credit risk management, 38–40
- Critical thinking, 137
- CRM (customer relationship
 management), 6, 10, 153–154
- Cross-channel lifecycle marketing,
 28–29, 31, 33
- Crowdsourcing analytics, 76–77
- Culture
 for data scientist development,
 144–146
 privacy and, 159–162
- Cunningham, Joe, 80–82
- Customer churn, 16, 44

- Customer insight engines, 71
- Customer intent, 33–34
- Customer lifetime value, 44, 71, 103
- Customer relationship management (CRM), 6, 10, 153–154
- Customer relationships
 - consumer products, 58
 - credit risk management, 38
 - cross-channel consumers, 28–29
 - customer lifetime value, 44, 71
 - database marketing, 24–27
 - digital marketing, 22–23
 - health care, 47–48
 - online privacy, 151, 154, 155, 162, 167
- Cutting, Doug, 61, 62

- Dashboard example, 118–119
- Data. *See also* Big Data
 - anonymized data, 77, 164
 - classification, 159
 - cloud computing (*see* Cloud computing)
 - control of, 23–24
 - data discovery, 65–66
 - dirty data, 107
 - filter via influence scores, 12
 - Hadoop for management (*see* Hadoop)
 - latency, 98
 - samples versus all data, 121–122
 - semi-structured, 10, 11, 85
 - sensitive data, 158–159, 161, 162
 - signal versus noise, 106–108
 - size, 86–87
 - structured, 10, 11, 35, 86
 - terrorism and, 165
 - unstructured, 10, 11–12, 35, 86
 - variety (*see* Variety of data)
 - velocity (*see* Velocity of data)
 - visualizing (*see* Visualizing data)
 - volume (*see* Volume of data)
- Data analysts
 - data scientists versus, 134
 - “last mile” people, 101–102
 - recommended resources, 176
 - value of, 102
- Data artisans, 111, 113
- Data discovery, 65–66
- Data mining
 - in analytics spectrum, 14, 100
 - definition, 84
- Data privacy. *See* Privacy
- Data scientists
 - data analysts versus, 134
 - development of, 139–140, 142–146
 - emergence of, 128–132
 - investing in, 127–128, 136
 - “last mile” people, 101–102
 - recommended resources, 176
 - shortage of, 128
 - skills, 49, 100, 122, 133–136, 138, 143
- Data warehouse
 - for Big Data storage, 89–91, 96
 - Big Data versus, 19–20, 49
 - database marketing and, 153
 - MapReduce for, 94
- Database marketing, 24–27, 152–153
- Datakind, 3–4
- de Montcheuil, Yves, 67
- Decentralized analytics, 146, 148
- “Decision management,” 137–139
- Definitions of technology terms, 83–86
- Deighton, Anthony, 66
- Descriptive analytics
 - in analytics spectrum, 99–100
 - in holistic view, 140–141
- Digital marketing, 19–24, 25, 54–55, 101–102
- Disk bound, 96
- Disruptive innovations, 77, 79, 132
- Distributed processing, definition, 84.
 - See also* Parallel computing
- “Do Not Track,” 155
- Domo, 72, 73
- Dowlaty, Zubin, 135
- Doyle, Shaun, 24, 25–27
- Dremel, definition, 84
- Drilling down, 112
- Driscoll, Michael, 136
- Dumbill, Edd, 5
- Dweck, Carol S., 145

- EB (exabyte), 87
- Edsall, Robert M., 114
- Elastic Search, 35
- Electronic Communication Privacy Act (ECPA; 1986), 155
- Enterprise resource planning (ERP), 10
- Ethics, 162–164. *See also* Privacy
- Evidence-based medicine (EBM), 47
- Examples from industry. *See* Industry examples
- Fault tolerance
 - Hadoop, 62–63
 - MapReduce, 90, 91
 - Storm, 86
- Fayyad, Usama, 15–18, 101–102
- Federal Trade Commission (FTC), 155, 156, 157
- Federated model of analytics, 146, 148
- Financial Services Modernization Act (1999), 155
- Firewalls
 - analytics and, 77–80
 - software as a service and, 73
- Flume, definition, 84
- Fraud
 - Big Data and, 34–37, 44
 - database marketing tools for, 27
- FTC. *See* Federal Trade Commission (FTC)
- Fuzzy Logix, 40, 98
- GB (gigabyte), 87
- Geospatial intelligence, 104–106, 119–120
- Ghosh, Misha, 2, 11
- Glossary of technology terms, 83–86
- GNU General Public License (GPL), 67
- Goldbloom, Andrew, 76
- Golden, James, 46–49
- Google
 - data processed per day, 87
 - Hadoop and, 61
 - Percolator, 35
 - privacy and, 157, 160, 161
 - Google Analytics, 21, 33, 172–173
- GPGPU (general-purpose GPU), 98
- GPU (graphical processing units)
 - as emerging technologies, 98
 - for parallel computing, 40, 43, 44
- Grids
 - definition, 84
 - GPU grids, 98
 - as parallel computing platform, 92, 96
- Hadapt, definition, 84
- Hadoop. *See also* Cloudera; MapReduce
 - associated software, 83–86, 124–125
 - data accessibility, 7, 48, 69
 - for financial industry, 64
 - integrating into enterprise, 81–82
 - Mahout, 84, 95, 124
 - overview of, 61–63, 84
 - SQL on data, 69
- Hadoop Distributed File System (HDFS), 63, 84, 89–90
- Hammerbacher, Jeff, 62
- HANA, 8, 84
- Harahan, Pat, 65–66
- Harrower, Mark, 114
- Hbase, definition, 84
- HDFS (Hadoop Distributed File System), 63, 84, 89–90
- Health care
 - analytics, 16, 46–47
 - data in health care, 45–46
 - holistic value proposition, 47–49
 - insurance companies, 44, 47
 - quantitative pharmacology, 50
- Health Insurance Portability and Accountability Act (HIPAA; 1996), 155
- Heer, Jeffrey, 113–114
- Hive, 84, 94
- Holistic analytics
 - holistic education, 140
 - holistic framework, 140–141
 - holistic value proposition, 47–49
 - Mu Sigma vision, 139
- Horton Works, definition, 84
- Hougland, Curtis, 33

- HPC (high-performance computing),
84, 93, 96
- HStreaming, definition, 84
- Hughes, David, 20–21
- Incremental innovations, 132
- In-database analytics, 93, 95, 96
- Industry examples
advertising, 51–57, 102
algorithmic trading, 40–44
consumer products, 57–58
consumer relationships, 22–23
convergence, 131
data ethics, 162–164
data visualization, 111–112
database marketing, 24–27, 153
“decision management,” 137
definition of Big Data, 9
developing data scientists, 139
digital marketing, 19–24, 25, 54–55
emergence of Big Data, 2–8
ethics, 162–164
fraud detection, 34–37
geospatial intelligence, 103–106
health care (*see* Health care)
“last mile” people, 101–102
new school of marketing, 27–34
privacy, 157
risk management, 37–40
value of analytics, 12–13
visualizing data, 116, 117, 118–121
- Influencers, in marketing, 31–32
- Information management. *See also*
Technology
as analytic spectrum, 100
Big Data storage, 89–90, 92, 96
(*see also* HDFS)
bottlenecks, 64, 91, 95, 96–97, 124
emerging technologies, 97–98
parallel programming, 91, 93–96, 98
(*see also* Parallel computing)
platforms (*see* Platforms)
processing of data, 90–91, 93–97
- Innovation
in analytic technology, 80–81, 130, 132
Big Data for, 103
decision sciences for, 132
innovation ecosystem, 82
innovation engines, 71
- Inquisitive analytics, 140–141
- Institutionalizing analytics
being analytical, 15, 16, 108–110
culture creation, 144–146
data democracy, 169–170, 172, 174
data scientists (*see* Data scientists)
“decision management,” 137–138
executive sponsorship, 127, 137–138
holistic analytics, 47–49, 139, 140–141
investing in analytics, 127–128,
136–137
keys to success, 127, 129–132
OODA Loop, xv
organizational structure, 127, 146–149
value of analytics, 12–15, 17–18, 102
- Insurance companies, 44, 47
- Investing in analytics, 127–128,
136–137
- I/O bound, 97
- James, Josh, 72–73
- Jewett, Dan, 65
- Jonas, Jeff, 103–106, 151
- Kaggle, 76–77
- Kaushik, Avinash, 19–20, 21, 22, 23–24,
69, 136–137, 169–174
- KB (kilobyte), 87
- Kent, Paul, 2
- Kerzner, Dan, 74, 75
- Kim, W. Chan, 71
- “Land and expand,” 65
- Lander, Jared, 3
- “Last mile” people, 101–102
- Legislation and acts
Children’s Online Protection Privacy
Act (2000), 155
Consumer Privacy Bill of Rights,
154, 156
Consumer Privacy Protection Act
of 2011, 155
“Do Not Track,” 155

- Legislation and acts (*continued*)
- Electronic Communication Privacy Act (ECPA; 1986), 155
 - Financial Services Modernization Act (1999), 155
 - Health Insurance Portability and Accountability Act (HIPAA; 1996), 155
 - U.S. versus Europe, 159
- LinkedIn groups, 176
- Lucas, Steve, 7–8
- Lucene, 95, 124
- Machine learning
- in analytics spectrum, 100
 - definition, 84
 - Mahout, 84, 95, 124
 - model reevaluation, 122
 - as new analytics, 48, 49
- Mahout
- definition, 84
 - libraries of analytics, 95
 - as open source analytic tool, 124
- Maples, Creve, 114–116
- MapR, definition, 84
- MapReduce
- definition, 85
 - as Hadoop component, 61, 63, 84
 - as HDFS processing software, 90–91
 - for parallel programming, 93, 94–95, 96
 - SQL-MapReduce, 69
- Market risk, 38, 41–44
- Marketing
- advertising (*see* Advertising)
 - analytics, 44
 - Blue Ocean Strategy* (Kim & Mauborgne), 71
 - campaign optimization, 16
 - customer insight engines, 71
 - database marketing, 24–27, 152–153
 - digital marketing, 19–24, 25, 54–55, 101–102
 - marketing automation, 26–27
 - measurement of impact, 54–57
 - new school of, 27–34
 - prospect universes, 27
 - word-of-mouth, 30–31
- Marketing mixed modeling (MMM), 56
- Mark-to-market valuation, 42
- Mauborgne, Renée, 71
- MB (megabyte), 87
- Measurement of impact, 54–57
- Mehta, Abhishek, 64, 69–70
- Meister, John, 2
- Memory bound, 97
- Mennis, Jeremy L., 114
- Message passing interface (MPI), 93, 95–96, 125
- MicroStrategy, 74
- Mind-sets, fixed versus growth, 145
- Mobile devices, analytics on, 73–75
- Monte Carlo simulations, 14, 44, 97, 98
- Montessori, Maria, 145
- Moore's Law, definition, 1
- Motivation, 144, 145
- MPI (message passing interface), 93, 95–96, 125
- MPP (massively parallel processing), 85, 92–93, 96
- MPP appliances, 85, 92–93
- MPP database, definition, 85
- Mu Sigma, 56, 135, 139, 143, 144–146
- “Non-line world,” 20
- NoSQL databases, 84, 85
- Olson, Mike, 62
- Omniture, 21, 33, 56, 72–73
- OODA Loop, xv
- Oozie, definition, 85
- Open source technology
- commercial versus, 8, 67–69
 - definition, 67
 - Elastic Search, 35
 - GNU General Public License (GPL), 67
 - Hadoop (*see* Hadoop)
 - Mahout and Lucene, 95
 - R (analytics tool), 85, 124
 - Talend, 67
 - tools for analytics, 85, 124–125
 - visualization tools, 118

- Optimization engines, 71, 133
- Organizational structure, 127, 144–146, 146–149
- Ownership, of data
organizational structure and, 148
privacy and, 165
- Parallel computing
Clojure, 83
description of, 90
GPUs for, 40, 43, 44
Hadoop, 63
MapReduce, 85, 90–91, 94–95
MPP (massively parallel processing), 85, 92–93, 96
as new analytics, 64
Pig, 85
platforms for, 92–93 (*see also* Platforms)
- Parallel programming
description of, 91
for GPUs, 98
on parallel computing platforms, 93–96
- PB (petabyte), 87
- Peled, Ori, 38
- Percolator, 35–36
- Personal information (PI), 158
- Personally identifiable information (PII), 159, 164
- Pig, 85, 94
- Pink, Daniel H., 144
- Platforms
analytics platform definition, 83
Google Analytics, 21, 33, 172–173
Hadoop (*see* Hadoop)
in holistic value proposition, 48
MPP appliances, 85, 92–93
Mu Sigma SaaS, 56
as new analytics, 64
Omniture, 21, 33, 56, 72–73
for parallel computing, 92–93
QlikTech International, 65–66
Revolution Analytics, 67
SAP HANA, 8, 84
SAS (*see* SAS)
- Tableau Software, 65–66, 116, 117, 118
- Tracx, 31, 33
- Unica, 56
- Porway, Jake, 3–4
- Predictive analytics. *See also* Analytics
in analytics spectrum, 14, 99–100
applying, 71
business intelligence versus, 49, 70
data volume and, 9
in database marketing, 25–26
in holistic view, 140–141
machine learning, 14
origin of term, 99
risk management, 43
- Prescriptive analytics
in analytics spectrum, 99–100
in holistic view, 141
- Privacy
anonymized data, 77, 164
bill of rights, 154, 156
cloud computing, 70
contexts of, 159–162, 163–164
definition, 157
legislation, 155
middleware layer, 165–166
online agreements, 151, 154
personal information (PI), 158
personally identifiable information (PII), 159, 164
privacy landscape, 152
privacy policies, 157
Safe Harbor, 155–156, 161
sensitive data, 158–159, 161, 162
terrorism and, 165
transparency of data, 11, 34, 151, 156
- Prospect universes, 27
- QlikTech International, 65–66
- Quants. *See* Data scientists
- R (analytics tool), 85, 124
- Rajaram, Dhiraj, 139
- Ramanathan, Murali, 50

- Real-time processing
 - algorithmic trading, 91
 - data into action, 123
 - definition, 85
 - geospatial intelligence, 104
 - streaming analytics, 106
- Recommendation engines, 62, 71
- Reiskind, Andrew, 157–158, 159, 161, 163–164
- Relational databases, 85, 96
- Revolution Analytics, 67, 124, 128
- Risk management, 16, 37–40, 41–44, 71
- Rockefeller, John D. (Jay), 154–155

- SaaS (software as a service), 56, 72–73
- Safe Harbor, 155–156, 161
- Sample, of data, 121
- SAP HANA, 8, 84
- SAS
 - as advanced analytics platform, 8, 56
 - analytics on-demand center, 72
 - descriptive analytics and, 99
 - Social Network Analysis (SNA), 37
- Scalability
 - cloud computing, 70
 - data democracy, 169–170
 - Dremel, 84
 - Hadoop, 62–63
 - limitations, 96–97
 - Mahout, 84
 - as new analytics, 26–27, 48, 64
 - NoSQL databases, 85
 - parallel computing platforms, 92–93
 - Percolator, 35
- Schema, 89
- Scoring, definition, 85
- Security. *See also* Privacy
 - cloud computing, 70
 - mobile analytics, 75
 - as privacy principle, 156
- Self-service approach, 65–66, 73
- Semi-structured data, 10, 11, 85
- Sen, Partha, 40, 41–42, 43, 44
- Sensitive data, 158–159, 161, 162
- Sensor data, 104
- Shneiderman, Ben, 113–114

- Silos, of data
 - in database marketing, 25
 - social data as, 32
- Singer, Niv, 31–32, 33
- Smith, David, 5, 67–68
- SNA (social network analysis), 37
- Social media
 - consumer relationships, 22–23
 - intelligence software, 31–34
 - as system monitors, 12
- Social network analysis (SNA), 37
- Software as a service (SaaS), 56, 72–73
- Solid-state drives (SSD), 97–98
- Spark, 85
- Springer, Dan, 27–28
- SQL (structured query language)
 - in analytics spectrum, 14
 - definition, 85
 - Hbase, 84
 - Hive, 34, 94
 - NoSQL database, 84, 85
 - Pig, 85, 94
 - SQL-H, 69
 - SQL-MapReduce, 69
 - Sqoop, 86
 - user-defined extensions, 95
- Sqoop, 86
- Stolte, Chris, 65
- Stomppff, Guido, 111
- Storm, 86
- Structured data, 10, 11, 35, 86
- Supply chain optimization, 10, 16, 71

- Tableau Software, 65–66, 116, 117, 118
- Tags
 - as data structure, 11
 - in digital advertising, 54–55
- Tal, Marcia, 137–139
- TB (terabyte), 87
- Technology. *See also* Information management
 - cloud computing (*see* Cloud computing)
 - emerging technologies, 97–98
 - evolution of, xii–xiv, 1–3, 5–7, 64
 - Hadoop (*see* Hadoop)

- open architecture, 32–33
- open source (*see* Open source technology)
- outsourcing, 23
- parallel computing (*see* Parallel computing)
- people/process first, 6
- Percolator, 35–36
- term definitions, 83–86
- “10/90 rule” for investing, 136
- Teradata Aster, 69
- Thampi, Arun, 162–163
- Thorp, Jer, 114
- Total cost of ownership (TCO)
 - grids, 92
 - Hadoop, 62
 - institutionalizing analytics, 132
 - software as a service, 72
- Tracx, 31, 33
- Tradeshows and conferences, 176
- Translational medicine (TM), 47

- Unica, 56
- U.S. Library of Congress, data collection of, 5
- Unstructured data, 10, 11–12, 35, 86
- User-defined aggregate (UDA), 95
- User-defined extensions, 95
- User-defined functions (UDF), 95
- User-defined table functions (UDTF), 95

- Value of analytics, 12–15, 17–18, 102
- Variables, in analytics, 17–18, 41, 121
- Variety of data
 - advertising and, 57
 - as Big Data, 1, 9–10, 83
 - for fraud detection, 35
 - online data, 20
- Vector processing, 98

- Velocity of data
 - advertising and, 57
 - as Big Data, 1, 9, 10, 83
 - for fraud detection, 35
- Visiphors, 112
- Visualizing data
 - data artisans, 111, 113
 - definition, 110
 - examples of, 118–121
 - intangibles visualized, 111–112
 - interactivity, 113–114, 116–117
 - recommended resources, 175
 - talented producers of, 134
 - tools for, 118
 - visiphors, 112
 - visualization designer, 114
 - word clouds, 114–116
- VMI (vendor managed inventory), 77
- Volume of data
 - advertising and, 56–57
 - analytics and, 9
 - annual growth projected, 5
 - as Big Data, 1, 9, 83
 - for fraud detection, 35
 - Vs of Big Data, 1, 9, 35, 56–57, 83

- Web Analytics* (Kaushik), 136
- Web Analytics 2.0* (Kaushik), 20, 136
- Websites
 - ads on, 54–55
 - control over data, 23–24
 - digital marketing, 19–23, 54–55
- Whirr, 86
- Word clouds, 114–116
- Word-of-mouth marketing, 30–31

- YARN, 86
- Yau, Nathan, 134
- YB (yottabyte), 87

- ZB (zettabyte), 87
- Zeitlin, Michael, 111–112

<http://www.pbookshop.com>

<http://www.pbookshop.com>

<http://www.pbookshop.com>

<http://www.pbookshop.com>

<http://www.pbookshop.com>

<http://www.pbookshop.com>

<http://www.pbookshop.com>

<http://www.pbookshop.com>

<http://www.pbookshop.com>

<http://www.pbookshop.com>

<http://www.pbookshop.com>