A
A/B testing, 251, 259
acquisition
of customers, 258
of employees, 260
act characteristics of data lake, 137
actionable analytics, 62
actionable dashboard, 68
actionable user experience (UEX), 61
action from analytics, 166–168
activation
of customers, 258
of employees, 260
advertisers
data for, 156
objective of analytics, 85
advocacy, of employees, 260
advocacy score, 101
airlines
analytics and data sources, 221
decisions about inflight performance, 217
maintenance, analytics and data sources, 222
metamorphosis phase, 213
“Travel Delight” business of, 29
airplane manufacturers, vision
statement and business model, 214
airplanes, design, analytics and data sources, 222
Alpine Miner, 95
Amazon, 7, 26
American Express, 7
analysis of variance (ANOVA), 96
analysis paralysis, 251
analytic algorithms, 116–128
association analysis, 119–120
cluster analysis, 116–117
cohorts analysis, 126–128
decision tree classifier analysis, 125–126
graph analysis, 121
Normal Curve Equivalent (NCE), 117–119
sentiment analysis, 123–124
summary, 129
text mining, 122–123
traverse pattern analysis, 124–125
analytic models, mapping cases to, 46–47
analytic profiles, 100, 102
analytics, 4
action from, 166–168
algorithms, 109
building models, 95–96
defining requirements in
metamorphosis phase, 216–223
for healthcare decisions, 225
identifying requirements, 218–220
mapping, 218–220
optimization from, 7
responsibility for, 4
analytics center of excellence (COE), CDMO and, 247
analytics chasm, data lake for crossing, 137–140
analytics dilemma, 139
analytics environment characteristics, 138
vs. data warehousing, 142
modernizing, 140–143
vs. reporting environments, 12
analytics sandbox, 141–142, 146–147
analyze characteristics of data lake, 137
Apache Hadoop, 11, 98, 98n, 142
data lake based on, 140–141
flat files to eliminate or reduce joins, 99
Apinions, 185
Apple, 7
asset
allocation recommendations, 79
data as, 32
association analysis, 119–120, 129
association rules
in analytic profile, 101
converting into segments, 120
AtScale, 149
attitudes, of data scientists, 90
attribute analytics, 255–257, 260
attribute score, action based on, 102

B
baggage management
analytics and data sources, 221
potential analytics, 219
bank checks, financial spending from, 76
bar charts, in R, 112
baseball team, 172
big data strategy document for, 52–57
Oakland A’s, 86–87
for World Series, 52–57
base data, visualizing variations with bloxplots, 112
Beane, Billy, 87
behaviors
groupings in analytic profile, 101
insights about, 22
benchmarks, 72
Berra, Yogi, 90
“best fit” trend line, 110
best practices, 72
“Beware the Data Lake Fallacy” (Gartner), 135
big data
business potential of, 231
business value from, 135–136
for data science, 85
economic potential, 8
economics of, 21
four Ms of, 10
integrating into business processes, 18–19
internal focus at start, 30–31
stories for, 255–267
strategy decomposition process, 36
use cases, 37
value drivers, 31
Big Data Business Model Maturity Index, 5–6, 17–34
basics, 18–29
lessons learned, 30–32
phase 1: Business Monitoring, 5, 19, 20–21
phase 2: Business Insights, 5, 19, 21–25
challenge of, 24–25
phase 4: Data Monetization, 5–6, 20, 27–28
phase 5: Business Metamorphosis, 6, 20, 28–29, 211, 212–213
articulating value propositions, 215–216
articulating vision, 214
defining data and analytic requirements, 216–223
understanding customers, 215
summary, 33–34
big data initiatives, biggest threat to success, 245
big data MBA, basics, 4–6
big data strategy document, 35–59
   basics, 37–51
   reasons for effectiveness, 38
   sections, 38
   for World Series, 52–57
Big Data Vision Workshop, 231–244
   data science in, 235–236
   deliverables, 239
   facilitated conversations in, 239–240
   functions, 232
   illustrative analytics, 236
   location for, 239–241
   pre-engagement research, 233–234
   prioritization matrix, 242
   process, 232–233
   stakeholder interviews, 234–235
   summarizing results, 238–239
   and “thinking like a data scientist,” 232
   time for, 240
   user experience mockup, 236
   workshop process, 236–239
   workshop setup, 239–241
   box-and-whisker plot, 112
   boxplots, 112–113, 129
BPM (Business Performance Management), 88
brainstorming, 220
   business decisions and questions in
   workshop, 237–238
   “By” analysis and, 172–174, 194
   creative thinking process in, 174
   data sources, 48
   healthcare industry data sources,
   225–226
   in metamorphosis phase, 216
   in monetization exercise, 203–204
   brand and category management analysis, 262
Business-as-a-Service, 6
   business case, stakeholder
   responsibility for positioning,
   242–243
   business-centric focus, role of, 4
   business decisions, 37, 43
   capturing, 161–162
   grouping into common themes, 238
   identifying in metamorphosis phase, 216–217
business entities, 37
business initiatives, 63
   brainstorming decisions and
   questions, 54
   characteristics, 157
   financial drivers for, 45–46
   focusing prioritization matrix on, 242
   identifying, 10, 39–41
   for baseball team, 52
   in key business initiatives, 157–158
   use cases for, 51
Business Insights phase, in Big Data
   Business Model Maturity Index, 5,
   19, 21–25
   challenge of, 24–25
business intelligence (BI), 20
   CDMO and, 246
   data modeling for, 97–98
   vs. data science, 11, 85–105
   questions, 88
business intelligence (BI) analysts
   vs. data scientists, 89
business intelligence (BI) tools
   dashboard, 67–68
   production environment
   characteristics, 137–138
   typical graphic options, 93
business mandate, for big data, 4
Business Metamorphosis phase, in
   Big Data Business Model Maturity
   Index, 6, 20, 28–29, 211, 212–213
   articulating value propositions,
   215–216
   articulating vision, 214
   defining data and analytic
   requirements, 216–223
   exercise, 211–227
   in healthcare industry, 223–226
   understanding customers, 215
business model, metamorphosis phase
   and shift in, 212
Business Monitoring phase, in Big Data
   Business Model Maturity
   Index, 5, 19, 20–21
Business Optimization phase, in Big Data
   Business Model Maturity
   Index, 5, 19–20, 25–26
Business Performance Management
   (BPM), 19, 88
business processes, integrating big data into, 18–19
business questions, evolution of, 13
business stakeholders, 14, 37, 63
brainstorming decisions with, 43–44
“By” analysis and, 172–173
and data scientists, 155
and decision governance team, 251
decisions of, 202
developing personas for, 158–160
envisioning by, 231
for fitness tracker manufacturer, 203
focus on, 156
interviews in Big Data Vision Workshop, 234–235
organizations need to understand, 62
personas development in monetization exercise, 201–203
responsibility for positioning of business case, 242–243
See also stakeholders
business strategy, 37
incorporating big data, 19
need for, 4
for professional baseball organization, 52
business-to-business (B2B) case study, 71–80
business transformation, 10
economic-driven, 7–9
business value
from big data, 135–136
of data sources, 49
BusinessWeek, 183
“butterfly effect,” 263
buyers, stakeholder personas for, 159–160
“By” analysis, 171–182, 190
basics, 172–174
basketball variables and metrics, 176–177
exercise, 174–178
for Foot Locker, 178–181
format, 173
reviewing variables and metrics for groupings, 192

campaign marketing effectiveness, 259
capacity planning, in network and operational analytics, 264
case study
data science, set up, 107–109
enabling frontline employees, 66–71
improving customer engagement, 64–66
See also Chipotle Mexican Restaurants case study; Fairy-Tale Theme Parks; Foot Locker casino, association rules about, 124–125
cause and effect, quantifying, 89
CDMO. See Chief Data Monetization Officer (CDMO)
CDO (Chief Data Officer), 245
cell phone provider, subscriber e-mail on data limits, 62–64
central repository, data lake as, 136
chaos theory, “butterfly effect,” 263
Chief Data Monetization Officer (CDMO), 245–248
analytics center of excellence (COE), 247
leadership, 248
organizational structure, 246–247
responsibilities, 246
Chief Data Officer (CDO), 245
children, rating families as risk, 183
Chipotle Mexican Restaurants case study, 39, 57–58
analytic insights to capture business entities, 43
customers, 42
data sources, 48–51
data strategy document, 45
final big data strategy document, 47
identifying financial drivers, 45–47
identifying key business entities and decisions, 41–45
identifying key business initiatives, 39–41
implementation feasibility assessment chart, 50–51
mapping cases to analytic models, 46–47
potential value of data sources, 49
President’s Letter to Shareholders, 40
prioritizing data sources, 48–51
prioritizing use cases, 51
classifications, in analytic profile, 101
client discretionary spending of, 78
financial status on dashboard, 75–76
personal information on dashboard, 74, 75
Cloudera Impala, 149
cluster analysis, 116–117, 129
Cognos, 92
cohorts analysis, 126–128, 130, 259
competitive analysis, 69–70
competitive differentiation
Big Data to drive, 6–9
technology to power, 7
competitive parity, 6
compound trend analysis, 111
computers, economic impact, 9
confidence levels of metrics, 90, 103
conformed dimensions, 145
confusion matrix, 95
Consumer Comments data, 50–51
correspondence, facilitated in workshop, 239–240
Corporate Information Factory, 145
corporate mission, 37
cost
of data warehousing, 133
of IT infrastructure, reducing, 134
cost of living metrics, 80
creative thinking process
in brainstorming, 174
fueling, 232–241
stimulating, 237
in workshop, 241
creativity, organizational, 251–253
credit cards
data governance, 148
financial spending from, 76
credit history, FICO score and, 185
credit mix, in FICO credit score, 187–188
credit utilization, in FICO credit score, 187
CRM (Customer Relationship Management), 6
Cross Industry Standard Process for Data Mining (CRIPS) model, 90
“crossing analytics chasm” moment, 25
data lake, 137–140
current state of business, monitoring, 12–13
customer lifetime value (CLTV), 195–196, 259
customer loyalty programs, 28, 249–250
evaluating, 261
customer promotional decisions by stakeholders, 161
customer purchase behaviors, influence, 26
Customer Relationship Management (CRM), 6
customers
acquisition, 258
activation, 258
advocacy, 259
analytics, 257–261
in Business Metamorphosis phase, 215
cross-sell and up-sell, 258
decision governance by organization, 250–251
fraud, 259
improving engagement, 64–66
loyalty program, 42
privacy of, 248–251
retention, 258
sentiment, 259
stakeholder personas for, 159
customer usage behaviors, for monetarization exercise, 199
cyclical component, in trend analysis, 115
D
“dark” data, analysis, 231
dashboard
  for financial advisor, 73–74
  informational sections of, 74–76
  recommendations section, 77–80
for retail store manager, 67–68
use cases added to, 70–71
data, 37
  as asset, 32
  defining requirements in
    metamorphosis phase, 216–223
for FICO score, 186
governance as lifecycle, 147–148
governance procedures for accuracy, 146
identifying requirements, 220–223
responsibility for, 4
storing structured and unstructured
  as-is, 134
three Vs of, 10
data acquisition, 5
data architecture, CIO responsibility
  for, 247
data assets, expanding, 19
data gathering, in data science, 94
data governance, vs. decision
  governance, 250–251
data integration, 31
data lake, 11–12, 94, 133–151
  advantages, 140
  basics, 134–136
  characteristics of business-ready,
    136–137
  CIO responsibility for, 247
  for crossing analytics chasm, 137–140
  data types, 140
  in hub and spoke analytics
    environment, 143–145
  performing ETL work in, 142
  relation to data warehouse, 148–149
  single vs. multiple, 146–147
data mart, 145
data modeling, 94
  for business intelligence (BI), 97–98
  for data science, 98–100
  in data warehouse, vs. data
    scientists’ use, 96
  pre-built, 91–92
Data Monetization phase, in Big Data
  Business Model Maturity Index,
    5–6, 20, 27–28
  creating new opportunities, 31–32
DataRPM, 95
data science, 4, 107–131
  analytic profiles from, 100
  basics, 86–89
  in Big Data Vision Workshop,
    235–236
  vs. Business Intelligence (BI), 11,
    85–105
  and “By” analysis, 171
  case study, set up, 107–109
CDMO and, 246
data modeling for, 98–100
  engagement process, 147
  “fail fast” exploration approach, 141
  goal of, 157
  hypothesis in, 94
  questions, 88–89
  testing and validating, 166–168
See also analytics
data scientists, 155–170
  engagement process, 93–96
data silos, 14
  eliminating, 134
  sharing data across, 143
data sources
  business value of, 49
  identifying and prioritizing, 48–51
  potential for San Francisco Giants, 56
  support in monetization exercise,
    204–206
data-to-analytics mapping, 220–223
data visualization tools, 95
data warehousing, 5, 8, 19
  vs. analytics environments, 142
  CIO responsibility for, 247
  cost of, 133
  data lake relation to, 136–137, 148–149
  ETL processing within, 142–143
  lessons from limitations, 145–149
  production environment
    characteristics, 137–138
  reducing workloads, 135
  “schema-on-load” approach, 92
  structured data for, 23
  “day in the life” description, for
    stakeholder persona, 159
decisions
data to guide, 32
governance by organization, 250–251
by HIPPO, 14
identifying, 41–45
of key users or business stakeholders, 202
optimizing, 5
decision support, 137
decision tree classifier analysis, 125–126, 130
demand forecasting, in network and operational analytics, 264
demarcation, data lake line of, 139
demographic information, in analytic profile, 100
descriptive analytics, 11, 13
business intelligence focus on, 88
for Foot Locker, 179, 180
product promotional questions, 165
design, decisions about, 218
device analytics, 261–263
digital media, attribution analytics for, 256–257
dimensional modeling, 97
dimension tables, 98
direct marketing effectiveness, 259
discretionary spending, of client, 78
distribution recommendations, 26
dot charts, in R, 112
downtime, unplanned, in network and operational analytics, 264

fraud, 261
retention, 260
sentiment, 261
empowerment, 14
empowerment cycle, 252
end customer, adding value for, 6
energy companies
“Home Energy Optimization” business, 29
metamorphosis phase, 213
Enterprise Data Warehouse (EDW), 145–146
Enterprise Resource Planning (ERP), 6
ERP effect, 7
environment, three dimensions in two-dimensional, 207
envisoning, 231–244
error matrix, 95. See also confusion matrix
ETL (Extract, Transform, and Load) processes, off-loading from data warehouse, 135, 142–143
EventBrite, 51
executives, technology-focused, 4
experimentation, 251, 252
external unstructured data, access to, 22–23
extrapolations, for time series, 112

F
Facebook, 7
fact tables, 96–98
Fair, Isaac, and Company, 185
Fairy-Tale Theme Parks
analytic algorithms, 109
boxplots use, 112
cluster analysis, 117
cohorts analysis, 128
data science goals, 108
decision tree classifier analysis, 126
geographical trend analysis, 114
graph analysis, 121
market basket analysis, 120
mobile app, 108
NCE technique, 118–119
pairs plot analysis, 115
sentiment score for attraction, 124
summary, 128–130
text mining, 123

economic-driven business transformation, 7–9
economics
of big data, 21
CDMO background in, 246
EDW (Enterprise Data Warehouse), 145–146
effectiveness, targeting, 259
EMC. See Big Data Vision Workshop
emotions, classifying, 124
employee lifetime value, 261
employees
acquisition, 260
activation, 260
analytics, 257–261
development, 260
groupings
decisions into common themes, 238
See also cluster analysis

H
H2O, 95
Hadoop, 11, 98, 142
data lake based on, 140–141
flat files to eliminate or reduce joins, 99
Hadoop Distributed File Systems
(HDFS), 11, 140
Hadoop/HDFS architecture, 134
advantages, 149–150
Harvey Balls, 49
HDFS (Hadoop Distributed File
Systems), 11, 140
healthcare industry
brainstorming data sources, 225–226
business metamorphosis in, 223–226
friction between providers and
payers, 223
key business stakeholders, 235
patient actionable analytic profile, 224
preventive care, 224
providers’ decisions, 224–225
highly governed data, 148
HIPPO (Highest Paid Person’s
Opinion), 14, 32
hiring, 260
histograms, in R, 112
historical data, accuracy, 90
“Home Energy Optimization”
business, of energy companies, 29
home value stability score, 188–189
Hortonworks Stinger, 149
hub and spoke analytics architecture,
143–145
hypothesis
in data science, 94
in empowerment cycle, 252
testing, 102–103

I
icebreakers, for workshop, 240
implementation, probability of
success, 241
implementation feasibility assessment
chart, for Chipotle, 50–51
individual in group, impact of,
126–128
industry best practices, 6
industry trends, 72
inertia, 245
inflight performance, potential
analytics, 219
informational sections, of financial
advisor dashboard, 74–76
ingest characteristics of data lake, 137
Inmon, Bill, 145
insights, 178
instrumentation, 251, 252
intellectual property, protecting
analytics as, 32
intelligence quotient (IQ), 184
interchangeable quotient (IQ), 184
interchangeable parts, economic
impact, 9
internal combustion engine, economic
impact, 9
internal process optimization, 30
internal unstructured data, access to,
22–23
Internet, economic impact, 9
“internet of things” (IoT), 263
interstate highway system, economic
impact, 9
interviews of stakeholders, in Big Data
Vision Workshop, 234–235
inventory recommendations, 26
irregular component, in trend
analysis, 115
IT infrastructure, reducing costs, 134

J
James, LeBron, 174–178, 189–193
“By” analysis of shooting, 190
Recommendations Worksheet for
defending against, 191
shooting effectiveness, 175
shooting percentages, 177
jet engine manufacturer,
metamorphosis phase, 212
job security score, 188

K
key business decisions
identifying, 63
See also scores
key business entities
analytic insights to capture, 42–43
identifying, 41–45
key business initiatives, 63
brainstorming decisions and
questions, 54
characteristics, 157
financial drivers for, 45–46
focusing prioritization matrix on, 242
identifying, 10, 39–41
for baseball team, 52
in key business initiatives, 157–158
use cases for, 51
key performance indicators (KPIs),
dashboard for, 67–68
keyword searches, of customers, 249
kick off workshop, 240
Kimball, Ralph, 96, 145
Kolmogorov-Smirnov test, 96
KPIs (key performance indicators),
dashboard for, 67–68
length of credit history, in FICO credit
score, 187
Lewis, Michael, Moneyball: The Art of
Winning an Unfair Game, 86–87
linear compute scalability, 140
line charts, in R, 112
load balancing, in network and
operational analytics, 264
loans, borrower’s ability to repay, 185
local events calendar, impact on sales,
70
local events data, 51
location, for Big Data Vision
Workshop, 239–241
London, steam locomotive and, 8–9
loyalty score, 101
machines, analytics, 261
MADlib, 95
Mahout, 95
maintenance, repair, and operations
inventory optimization, 261–262
maintenance recommendations, 26
maintenance scheduling optimization,
261
“Make Me More Money” (4 Ms), 32
manufacturing effectiveness, 262
MapMyRun, 27
markdown management optimization,
260
market basket analysis, 119, 260
Market Basket data, 50
marketing analytics, 259–260
mathematical models, for trend lines,
111
MaxPreps, 51
medical data, governance, 148
merchandise managers, stakeholder
personas for, 159
merchandising, 158
metadata, 122, 135
metamorphosis. See Business
Metamorphosis phase
Metaphor Computers, 97
metrics
identifying predictors of
performance, 172
See also scores
MicroStrategy, 92
Mint.com, 76
mission of organization, 37
model goodness of fit, 96
moderately governed data, 148
“Mom” test, 250
monetization
customer recommendation on
opportunities, 66
exercise, 199–210
fitness tracker, 200–209
See also Data Monetization phase
monetization exercise, 200–209
process steps, 199–200
step 1: product usage understanding,
200–201
step 2: stakeholder personas
development, 201–203
step 3: brainstorming potential
recommendations, 203–204
step 4: support of data sources,
204–206
step 5: prioritizing opportunities,
206–207
step 6: plan development, 208–209
“A Moneyball Approach to Helping
Troubled Kids” BusinessWeek, 183
Moneyball: The Art of Winning an Unfair Game (Lewis), 86–87
monitoring current state of business, 12–13
motivation, from stories, 265
“Motivation” score, 192

N
National Basketball Association, 174–178, 189
Real Plus-Minus (RPM) metric, 126–127
shooting effectiveness, 175
NCE (Normal Curve Equivalent), 115, 117–119, 129
Netflix, 7, 26
business model, 161
network and operational analytics, 263–265
network security, in network and operational analytics, 264
new customers, business potential to reach, 215
new products, introductions, 262
nodes, in graph analysis, 121
Normal Curve Equivalent (NCE), 115, 117–119, 129

O
Oakland A’s, 87
obsolescence, risk of, 4
OLTP (Online Transaction Processing), 145
omni-channel marketing analysis, 260
online digital marketplaces, 28
Online Transaction Processing (OLTP), 145
open source software, 8
operational data, access to all, 22
operational decisions, identifying in metamorphosis phase, 216–217
operational effectiveness, 6
operational systems, integrating analytic insights in, 103
optimization from analytics, 7
Business Optimization phase, in Big Data Business Model Maturity Index, 5, 19–20, 25–26
internal process, 30
of process, 101
organizational creativity, 251–253
organizational impediments, 245
organizational structure, and decision governance, 250–251
outliers, identifying, 112

P
packaging data, 27
pairs plot analysis, 113, 114–115, 129
Pandora, 26
parking lot flipchart, 240
Parks case study. See Fairy-Tale Theme Parks
passenger management analytics and data sources, 221
potential analytics, 219
patient actionable analytic profile, 224
payment history, for FICO score, 186
Pearson’s Chi-squared test, 96
Pentaho, 92
percentile rank analysis, normalizing data set for, 117–118
performance, metrics as predictors, 87, 172
personally identifiable information, governance, 148
personas for stakeholders developing, 158–160
development in monetization exercise, 201–203
pie chart, 117
Pivotal HAWQ, 149
planning, in monetization exercise, 208–209
Point-of-Sales (POS) transactional data, 22, 50
“points per shot” metric, 174
portfolio mix, 72
predictions, by data scientist, 90
predictive analytics, 5, 11, 13
adding to dashboard, 68
applying, 31
in data science, 89
developing, 25
for financial advisor, 77
for Foot Locker, 179, 180
Foot Locker business questions, 163–164
integrating, 24
product promotional questions, 165
in stakeholders’ expanded thinking, 163
predictive maintenance, in network and operational analytics, 264
predictors, of business performance, 88–89
prescriptive analytics, 5, 11, 13, 79
adding to dashboard, 68
in data science, 89
developing, 25
for Foot Locker, 179, 180–181
Foot Locker business questions, 163–164, 165
product promotional questions, 165
in stakeholders’ expanded thinking, 163
preventive maintenance, 261
price inflection points, 118–119
pricing
analytics and data sources, 220
decisions about, 216–217
potential analytics, 218
pricing and yield optimization, 260
printing press, economic impact, 9
prioritization matrix, 36, 51–52, 205–206
in monetization exercise, 206–207
template, 242
in workshop, 241–243
prioritizing
data sources, 48–51
decision governance and, 251
privacy, 248–251
and trust, 249–250
probabilities, 90
process, optimization of, 101
Procter & Gamble, 7
product analytics, 261–263
product pricing recommendations, 26
product promotional decisions by stakeholders, 162
product promotional questions, 165
for Foot Locker, 180
products
performance optimization, 262
rationalization/retirement, 262
testing and QA effectiveness, 262
usage patterns, for monetarization exercise, 199, 200–201
usage score, 101
promotional effectiveness, 259
p-values, 90

Q
Qlikview, 92
query
construction, 146
for data model, 92
questions
brainstorming, 162–166
evolution of, 13

R
R (programming language), 95, 109
association analysis, 120
for boxplots, 112
building decision trees with, 126
cohorts analysis, 128
Paired Plot options, 115
for plots and graphs, 112
Sentiment Analysis, 124
Social Network Analysis in, 121
text mining, 123
time series decomposition, 116
z-scores to normalize data, 119
RDBMS (Relational Data Base Management Systems), 23, 149
Real Plus-Minus (RPM) metric, NBA use of, 126–127
real-time analytics, 23, 31
Realtor.com, 80
recommendations
asset allocation, 79
data requirements, 205–206
scheduling, 26
spending analysis, 78
store manager ability to modify, 68
Recommendations Worksheet
for defending against L. James, 191, 193
template, 167
relational data base management systems (RDMBS), 23, 149
relationships, graph analysis to evaluate strength, 121
repackaging insights, 28
report
for data model, 92
from data warehouse, from SQL commands, 92
reporting environments, vs. analytics environments, 12
research, pre-engagement, in Big Data Visions Workshop, 233–234
“Reset on Big Data” (Forrester study), 4
resource scheduling recommendations, 26
retailers
big data to transform, 67
dashboard for store manager, 67–68
“Shopping Optimization” business of, 29
retention
of customers, 258
of employees, 260
retirement readiness score, 188
retrospective reporting, 12–13
return on investment (ROI), 30–31
for big data, 25
revenue, creating new sources, 20
“right-time” analytics, 23
“right” use case, determining, 242

S
sabermetrics, 87
sales and marketing
analytics and data sources, 221
decisions about, 217
potential analytics, 219
Sales Force Automation (SFA), 6
S.A.M. test (Strategic, Actionable, Material), 24
criteria, 178
San Francisco Giants, 52–57
big data strategy document, 55
SAS, 95
SAS Miner, 95
scheduling recommendations, 26
“schema-on-load” approach, for data warehousing, 11, 92, 93
“schema-on-query” process, 11, 95
scores, 183–197
for decision-making, 101
definition of, 184–185
FICO example, 185–188
other industry examples, 188–189
screen scraping, 51
seasonal component, in trend analysis, 115
security, in network and operational analytics, 264
segmentation effectiveness, 259
sensitivity analysis, 103
sentiment analysis, 123–124, 130
of customers, 259
of employees, 261
shared storage platform, 134
“Shopping Optimization” business, of retailers, 29
silos, 14
eliminating, 134
sharing data across, 143
simulations, 103
“slice and dice” technique, 25
social media data, 51
in analytic profile, 100–101
of customers, 249
social network analysis (SNA), 121
spatial (geographical) analysis, 113–114, 129
SP Business Objects, 92
special-purpose data lakes, 146
spending analysis, recommendations, 78
split testing, 251n
Spotfire, 95
spreadmarts, 135n
SQL-based query, 92–93
“SQL on Hadoop” products, 149
stakeholders, 14, 37, 63
brainstorming decisions with, 43–44
“By” analysis and, 172–173
and data scientists, 155
and decision governance team, 251
decisions of, 202
developing personas for, 158–160
evening by, 231
for fitness tracker manufacturer, 203
focus on, 156
interviews in Big Data Vision Workshop, 234–235
organizations need to understand, 62
personas development in monetization exercise, 201–203
responsibility for positioning of business case, 242–243
star schemas, 97, 99, 145
statistical analysis, 110–116
boxplots, 112–113
geographical (spatial) analysis, 113–114
pairs plot analysis, 114–115
time series decomposition, 115–116
trend analysis, 110–112
statistical model, goodness of fit, 95
steam engine, 8
economic impact, 9
Stephenson, George, 8
storage platform, shared, 134
store characteristics of data lake, 137
Store Demographics data, 50
store managers, stakeholder personas for, 159
stories, 255–267
characteristics of good, 265–266
customer and employee analytics, 257–261
digital media attribution analysis, 256–257
network and operational analytics, 263–265
product and device analytics, 261–263
strategic business initiatives, 37
strategic nouns, 37, 41
identifying, 160–161
strategy, 37
incorporating big data, 19
need for, 4
for professional baseball organization, 52
structured data, storing as-is, 134
supplier decommits/recommits analytics, 262
supplier network analytics, 262
supplier performance analytics, 262
supply chain optimization, 262
surface characteristics of data lake, 137

Tableau, 95
targeting effectiveness, 259
technology, leveraging to power competitive differentiation, 7
technology-focused executives, 4
technology innovations
business opportunities from, 9
and economic change, 8
telephone, economic impact, 9
test cases, for hypothesis testing, 253
testing
data science, 166–168
hypothesis, 102–103
text mining, 122–123, 130
theft and revenue protection, in network and operational analytics, 264
thinking differently, 21, 98
importance of, 10–14
in organization, 245
stories for, 255
“thinking like a data scientist” framework, 155–156
action from analytics, 166–168
Big Data Vision Workshop and, 232
brainstorming business questions, 162–166
capturing business decisions, 161–162
decomposition process, 162, 170
developing business stakeholder persona, 158–160
identifying key business initiatives, 157–158
identifying strategic nouns, 160–161
scores, 197
working backwards in, 199
time
limits for business initiatives, 157
for prioritization process, 52
for workshop, 240
timeline, for Big Data Vision Workshop, 233
time series decomposition, 115–116, 129
Traackr, 185
trade promotion effectiveness, 260
transactional data
access to all, 22
granularity level of, 31
transactional metrics, in analytic profile, 100
“Travel Delight” business, of airlines, 29
traverse pattern analysis, 124–125, 130
trend analysis, 110–112, 128
“trickle feeding” data, 23
trust, and privacy, 249–250
t-tests, 90
two-dimensional environment, three dimensions in, 207

U
Uber, 223n1
business model, 161
ungoverned data, 148
United States, economic cycles, 133–134
unplanned downtime reduction, in network and operational analytics, 264
unstructured data
access to, 22–23
storing as-is, 134
urgency, creating sense of, 39
use cases, 37
“By” analysis, for Foot Locker, 178–181
competitive analysis, 69–70
determining “right,” 242
grouping decisions into, 45–47, 55
in hub and spoke analytics environment, 144–145
for key business initiative, 51
user decisions, supporting, 63–64
user experience, 61–81
unintelligent, 62–64

V
validating data science, 166–168
value propositions, in Big Data Business Model Maturity Index, 215–216
variables
data science team responsibility for quantifying, 173
in pairs plot analysis, 114
quantifying relationships between, 119
and scores, 184–185
vertices, in graph analysis, 121
vision statement, 214
visualizing data, in data science, 95

W
Walmart, 7
Walt Disney Company, mission, 37
wearable computing, 263
weather forecast, integration with dashboard, 70–71
whiskers, in boxplots, 112
Wikipedia, 109
workshop. See Big Data Vision Workshop
World Series, big data strategy document for, 52–57

X
Xplain.io, 149

Y
Yahoo, 7

Z
Zillow, 80
z-scores, to normalize data with R, 119