

# Index

## A

Abstraction, 238, 243–244  
Academic impact, 78  
Accuracy of analysis, 154  
ACID. *See* Atomicity, consistency, isolation, and durability (ACID)  
Acquisition pipeline, 220  
Actionable information, 213  
Adaptive data, 238  
Adaptive data analysis, 159  
Added value, 158  
Advanced Message Queuing Protocol (AMQP), 42, 43  
Advertising, 149  
Algorithmic validation, 104  
AllegroGraph, 124  
Amazon, 126, 251  
Amazon Kinesis, 207  
Amazon Mechanical Turk, 95  
Amazon's Simple Storage Service, 127  
Amazon Virtual Private Cloud, 229  
Amazon Web Services, 229  
AMQP. *See* Advanced Message Queuing Protocol (AMQP)  
Analysis of industrial needs in the finance and insurance sectors, 213  
Analysis of industrial needs in the health sector, 180–182  
Analysis of industrial needs in the media and entertainment sectors, 248  
Analysis of industrial needs in the public sector, 199  
Analysis transparency, 154  
Analytic capabilities, 271  
Analytical databases, 133, 208, 272

Analytical insights, 175  
Analytical processing, 272  
Analytics, 172, 174, 175, 199, 238  
Analytics knowledge, 235  
Annotation, 265  
Annotation frameworks, 190  
Anonymization, 159, 191, 220, 231, 268, 269  
Anonymized data, 268  
Apache Cassandra, 66  
Apache Flume, 47  
Apache Hadoop, 70, 74  
Apache UIMA, 109  
Apache ZooKeeper, 47  
Apple, 251  
Application, 9, 290  
Approximate reasoning, 68  
AstraZeneca, 102  
Atomicity, consistency, isolation, and durability (ACID), 123  
Attribute-based encryption, 128  
Automated decision, 162  
Auto-tiering, 128  
Available Media and Entertainment Data Resources, 251

## B

Barriers, 173  
BDV cPPP. *See* Big Data Value contractual Public Private Partnership (BDV cPPP)  
BDVA. *See* Big Data Value Association (BDVA)  
BDVA SRIA, 291  
Behavioural patterns, 236

- Benchmarking, 68
- Big data, 30
  - analysis, 63
  - analytics, 40
  - architectures, 77
  - capabilities, 247
  - economy, 34
  - ecosystem, 8, 33–35, 280
  - processing pipeline, 40
  - public private forum, 24, 289
  - public private partnership, 24, 25
  - roadmap, 279
  - as a service, 160
  - skills, 173
  - value, 174, 280, 284, 291
  - value chain, 39, 63, 88, 119, 171, 288
  - value ecosystem, 288
- Big Data Value Association (BDVA), 24, 279, 289
- Big Data Value contractual Public Private Partnership (BDV cPPP), 25, 279, 289–290
- Big open data, 285, 287
- BIG project, 288, 292
- Broadband, 237
- Business, 10, 290
  - activity monitoring, 153
  - business-as-usual, 235
  - cases, 177, 185
  - decisions, 156
  - ecosystem, 8, 290
  - goals, 143
  - innovation, 172
  - intelligence, 175
  - model, 8, 143, 144, 150–151, 158, 176, 177, 185, 215, 227, 230–231, 237, 249, 269, 283
  - opportunities, 66
  - processes, 156, 158, 159
  - roadmap for big data, 282–284
  - rule engines, 153
  - rules, 272
  - users, 248
  - value, 173, 176, 177, 230, 233
- Business processes as a service (BPaaS), 160
- Business support systems (BSS), 175
  
- C**
- Capability, 18
- Car-sharing, 233
- Cassandra, 126
- Causation, 154
- CDS. *See* Clinical decision support (CDS)
- CEP. *See* Complex event processing (CEP)
- Challenges, 172
- Chemspider, 101, 110
- Chevron, 229
- CIS. *See* Clinical information system (CIS)
- Cisco, 126
- Citizens, 199, 287
- Civil engagement, 287
- Classification, 152
- Cleaning, 267
- Clinical data, 180, 183, 185, 186
- Clinical decision support (CDS), 161, 183
- Clinical information system (CIS), 52
- Clinical operation intelligence, 183
- Cloud-based infrastructure, 160
- Cloud infrastructures, 269
- Cloud storage, 126, 127, 240
- Cloudera, 121, 122, 128, 133, 136, 229
- Clustering, 152
- Cockpits, 153
- Co-creation, 170
- Code Week for Europe, 286
- CO<sub>2</sub> emissions, 7
- Collaborative networks, 287
- Columnar stores, 124
- Communication and connectivity, 234
- Communities, 17, 24, 66, 77, 95
- Competitive advantage, 3, 29, 156, 280
- Competitiveness, 4, 15, 66, 279, 284
- Complex data types, 74
- Complex event processing (CEP), 71–72, 149, 153, 161, 162, 241, 272
- Complex exploration tools, 161
- Complex pattern matching, 272
- Complex systems, 230
- Confidentiality, 218, 233, 240
- Confidentiality preserving, 228
- Consortium, 16
- Constraints, 184–185, 201, 216
- Consumer, 6
  - awareness, 251
  - experience, 236
  - profile, 252
- Consumerization, 227
- Content, 247
- Context-awareness, 155
- Context information, 163, 191, 266
- Contextual, 163
- Continuous service improvement, 233
- Contractual information, 228
- Contractual PPP, 24
- Contractual PPP (cPPP) proposal, 279, 290
- Convolutional neural networks, 242
- Copyright, 131, 285

- Correlation, 154
- Cortana, 257
- Cost, 92, 172
- CPS. *See* Cyber-physical systems (CPS)
- Critical infrastructures, 174
- Cross-sector analysis, 264
- Cross-sectorial
  - business requirements, 263
  - consolidated requirements, 264–273
  - roadmap, 279
- Crowd sourced, 78
- CrowdDB, 96
- CrowdFlower, 95, 103, 114
- Crowds, 95
- Crowdsourced, 114, 170, 252
- Crowdsourcing, 78, 89, 94, 103, 111
- Cryptographically enforced access control, 128
- Cryptography, 128
- C-SPARQL, 71
- Cultural change, 285
- Culture, 216, 237
- Curation, 248, 268
  - models, 98
  - scale, 98
  - at source, 94
  - usage, 253
- Customer, 170
  - acceptance, 228
  - behaviour, 175
  - data, 171, 173–175, 216, 265, 268
  - experience, 174, 227, 230–233
  - insight, 213, 215
  - interactions, 272
  - loyalty, 173
  - relationship, 175
  - satisfaction, 171
- Customer relationship management (CRM), 81
- Cyber-physical infrastructure, 243
- Cyber-physical systems (CPS), 144, 145, 156, 174, 228, 230, 241, 242
  
- D**
- Dashboards, 153
- Data, 8, 75, 290
  - access, 231
  - access and sharing, 239–241
  - acquisition, 31, 40, 175, 206, 208, 218, 220, 266
  - actionability, 216
  - analysis, 32, 51, 151, 153, 158, 176, 205, 207, 208, 242, 259, 260
  - analytics, 158, 176, 180, 182, 185, 229–231, 238, 241
  - capture, 249
  - centres, 243
  - cleaning, 105
  - and code-literacy, 284
  - commons, 131
  - confidentially, 177
  - consumption, 91
  - curation, 32, 205, 208, 221
    - algorithms, 108
    - automation, 104
    - infrastructures, 103
    - pipelines, 103
    - platforms, 104
    - workflows, 90
  - digitalization, 187
  - discovery, 65
  - disruption, 173
  - ecosystem, 7, 8, 35, 238, 263
  - engineering, 284
  - engineers, 284
  - enrichment, 264, 274
  - exchange, 149
  - extraction, 218, 221–223
  - generation, 91
  - governance, 177
  - growth, 215
  - heterogeneity, 91
  - hub, 133, 136
  - improvement, 268, 274
  - integration, 159, 162, 163, 172, 175, 183, 187, 190–191, 204, 206–207, 264, 268, 272, 274
  - integration/sharing, 219, 223
  - interaction, 105
  - interpretability, 238
  - journalism, 250, 251
  - journalists, 77
  - lake, 238
  - lifecycle, 174, 247
  - literacy, 170
  - logistics, 238
  - management, 274
  - management engineering, 264
  - marketplace, 34, 238
  - markets, 143
  - mining, 65, 152, 238
  - model, 107
  - ownership, 235, 266, 285
  - portals, 56
  - privacy, 131, 145, 158, 159, 177, 182, 183, 185, 188, 191, 216, 219, 268, 285
  - privacy agency, 203
  - privacy and security, 223–224
  - processing, 267
  - protection, 145, 217, 224, 240
  - protection laws, 235

Data (*cont.*)

provenance, 157, 240, 266, 270  
 quality, 87, 89, 91, 94, 146, 155, 157–159,  
 162, 177, 182, 183, 188, 192, 218,  
 221, 257, 259, 268, 274  
 reuse, 107  
 science, 4, 91, 284  
 science skills, 177  
 scientists, 152, 157, 161, 242, 243, 248, 284  
 security, 158, 159, 182, 183, 185, 188, 191,  
 204, 207, 268  
 security and privacy, 274  
 as a service, 175, 233  
 sharing, 172, 183, 187, 190–191, 197, 204,  
 206–207, 231, 238, 240, 264, 274  
 sharing and integration, 266  
 silos, 158, 185, 231, 232, 285  
 skills, 286  
 storage, 32, 176, 205, 207–209  
 strategy, 172, 212  
 streaming, 229  
 stream management, 220  
 stream processing, 147  
 streams, 260, 267  
 tamer, 96  
 tracing, 131  
 type agnostic architectures, 244  
 usage, 33, 209, 272  
 value chain, 17, 31, 147, 239, 248  
 variety, 52, 87, 91, 223  
 visualization, 146, 260, 270, 274  
 Datacentre, 147  
 Data-driven, 171, 173  
   business, 234, 235  
   documents, 105  
   economy, 7, 288  
   innovation, 172, 173, 177, 258, 259  
   service, 237  
   society, 14, 122  
 DBpedia, 109  
 De-anonymization, 238  
 Decentralized data generation, 87  
 Decision processes, 146  
 Decision support, 147, 149, 153, 158, 162, 223  
 Decision support and automation, 230  
 Decision support systems (DSS), 219  
 Decision-makers, 93  
 Decision-making, 143, 161, 170, 242  
 Deep data analytics, 270, 274  
 Deep learning, 242  
 De-identification algorithms, 191  
 Dell, 126  
 Descriptive analytics, 260, 271  
 Differential privacy, 241

Digital agenda, 5, 279  
 Digital rights, 90, 107, 235, 237  
 Digital services, 159  
 Digital single market, 7, 8, 34, 177  
 Digital technologies, 248  
 Digital transformation, 173  
 Digitalized, 228  
 Digitally divided, 235  
 Digitization, 227, 231, 235, 237, 242  
 Digitization and automation, 234  
 Directive 2003/98/EC, 202  
 Disruptive, 176  
 Distributed computing, 229, 243  
 Distributed data analytics, 243  
 Distributed file systems, 123  
 Distributed stream computing, 241  
 Distributional semantic models, 68, 109  
 Document databases, 124  
 Domain experts, 103  
 Domain know-how, 235  
 Do-Not-Track, 269  
 Dremel, 147, 148  
 Drill, 126, 207  
 Drivers, 184, 201, 215  
 Drivers and constraints for big data in media  
   and entertainment sectors, 249–251  
 DSS. *See* Decision support systems (DSS)  
 Dublin, 229  
 Dynamic bandwidth, 162  
 Dynamic semantic publishing, 250  
 Dynamic visualizations, 153

**E**

Ease of use, 229  
 eBay, 113  
 Economic impact, 102  
 Economic models, 98  
 Economy, 170  
 Ecosystem, 8, 14, 15, 33, 66, 170, 287  
 Edge computing, 243  
 Education, 177  
   and skills, 284, 286  
   of workforce, 283  
 Electronic Health Record (EHR), 185  
 Effectiveness, 173, 176  
 Efficiency(ies), 155, 161, 169, 173, 198,  
 234, 239  
   of care, 182  
   of healthcare, 180  
   increase, 234  
 Efficient data structures, 242  
 EHR. *See* Electronic health record (EHR)  
 Elastic computing, 223

- Electric vehicles, 228
  - Electricity industry, 228
  - Electronic health data, 184
  - Electronic health record (EHR), 52, 53, 190
  - Electronic on board recorders (EOBRs), 229
  - Embedded analytics, 243
  - Embedded systems, 156
  - Emergency response, 80
  - Emerging paradigms for big data analysis, 77
  - Emerging paradigms for big data curation, 97
  - Emerging paradigms for big data storage, 132
  - Encrypted data storage, 240
  - Encrypted storage, 207, 269
  - End user acceptance, 235
  - Energy, 41, 134, 137, 172, 176
    - consumption, 57
    - and transport, 7, 173
    - usage data, 176
  - Engaged citizen, 287
  - Enterprise data, 170
  - Enterprise resource planning (ERP), 55
  - Entertainment, 172
  - Entity linking, 272
  - Entity matching, 266
  - Entity recognition and linking, 109
  - Entity summarization, 73
  - Entrepreneurial atmosphere, 285
  - Entrepreneurial spirit, 283
  - Entrepreneurs, 34, 280, 287
  - Entrepreneurship, 287
  - Environmental, 98
  - EOBRs. *See* Electronic on board recorders (EOBRs)
  - ERP. *See* Enterprise resource planning (ERP)
  - Error tolerance, 156–157
  - Ethical aspects, 236
  - ETL. *See* Extract, transform, load (ETL)
  - EU data protection, 285
  - EU data protection directive, 219
  - European big data economy, 288
  - European big data ecosystem, 8, 34
  - European big data roadmap, 287
  - European big data value partnership, 202
  - European commission, 31, 279, 290
  - European data economy, 7
  - European data ecosystem, 7
  - European data forum (EDF), 17
  - European digital single data market, 287
  - European digital single market, 285
  - European market, 281
  - European strategy, 14
  - European Union, 191, 235
  - Evolutionary, 176
  - Execution engines, 147
  - Exploration, 147, 151
  - Explorative analysis, 157
  - Explorative data analysis, 158
  - eXtensible Access Control Markup Language (XACML), 129
  - eXtensible Business Reporting Language (XBRL), 219
  - Extract, transform, load (ETL), 96, 104, 136, 238
- F**
- Facebook, 122, 237
  - Finance, 82, 172, 251
  - Finance and insurance, 6, 59, 173
  - Financial, 98
  - Financial impact, 179
  - Flume, 48, 49, 208
  - FoldIt, 101, 112
  - Format, semantics, and quality, 254
  - Fraud detection, 162, 213
  - Freebase, 106
  - Fully homomorphic encryption, 240
  - Functional models, 163
  - Funding, 285
  - Future requirements and emerging trends for big data acquisition, 51–52
  - Future requirements and emerging trends for big data analysis, 74
  - Future requirements and emerging trends for big data curation, 97
  - Future requirements and emerging trends for big data usage, 154–161
  - Future requirements for big data analysis, 74
  - Future requirements for big data storage, 130–132
- G**
- GalaxyZoo, 101
  - Games, 112
  - Gamification, 112
  - Gap analysis, 20
  - GE, 150, 156
  - General data protection regulation (GDPR), 131
  - GeoConnections Canada, 102
  - Geo-location data, 236
  - Geo-spatial, 153
  - Geospatial data, 102
  - Geo-tagged data, 236
  - Global competitiveness, 173
  - Google, 122, 126, 251
  - Google data flow, 207
  - Google Pregel, 152
  - Google's file system (GFS), 147

- Government, 98
- Government, public, non-profit, 56
- GPS, 229
- Graph-based, 121
- Graph-based models, 132
- Graph data, 132
- Graph databases, 124
- Graph search, 132
- GraphLab, 152
- GSK, 102
  
- H**
- Hadoop, 47, 49, 66, 75, 122, 126, 133, 136, 229
- Hadoop Distributed File System (HDFS), 40, 49, 120, 123, 125, 133
- Hadoop MapReduce, 50
- Hash algorithms, 191, 269
- HBase, 126, 133, 136
- HDFS. *See* Hadoop Distributed File System
- Health, 80–81, 98, 110, 132, 172
  - data, 159, 182, 184, 191
  - data analysis, 186
  - data management, 180
  - data on the web, 186
  - sector, 52–54, 135, 264
- Healthcare, 5, 176
- Healthcare analytics, 179
- Heterogeneity, 64, 89
- Heterogeneous, 265
  - data, 180, 181, 254
  - data sources, 177
  - sources, 260
- Hewlett-Packard, 212
- High-level requirements, 264
- High-speed broadband, 251
- Hive, 125, 147
- Horizon 2020, 15, 279, 289
- Horizontal IT landscape, 236
- Hortonworks, 121
- Human-algorithmic data curation, 89
- Human computation, 90, 103, 112, 114
- Human–computer interaction, 143, 144, 153
- Human–data interaction, 90, 98, 192, 268
- Human-in-the-loop, 149
- Human-powered, 96
  
- I**
- IBM, 40, 126, 212
- IBM Watson, 105
- Images, 189, 190, 265
- Impala, 126
- Incentive models, 95
- Incentive regulation, 235
- Incentives, 89, 92, 98, 184, 285
- Incentivize, 170
- Increasing research and development, 284
- Industrial community, 14
- Industrial demand-side management, 233
- Industrial needs, 229–231
- Industrial requisites, 21
- Industry 4.0, 144, 145, 156, 159, 175, 269
- Industry adoption, 4
- In-field, 174
- In-field analytics, 238, 243
- Information aggregation, 151
- Information extraction, 151, 190, 192, 240, 265, 272
- Information logistics, 143
- Information visualization, 270
- In-memory, 273
- In-memory databases, 158
- Innovation, 170, 171, 237
- Insight, 250
- In-store data, 175
- In-stream processing, 209
- Insurance, 172
- Integrate structured, 109
- Integrated roadmap, 287
- Integrating the healthcare enterprise (IHE), 54
- Integration, 108, 122, 146, 149, 155, 158, 180, 232
- Intellectual property, 269
- Intelligent electronic devices, 234
- Intelligent infrastructure, 228
- Intelligent processing, 159
- Interactive exploration, 146
- Interactivity, 153
- Intermediary business models, 233
- Internet of services, 159
- Internet of things (IoT), 7, 55, 134, 144, 159, 170, 198, 258, 267, 280, 286
- Interoperability, 43, 88, 90, 159, 173, 283, 290
- Interoperable, 101
- Interpretability, 239, 240
- Interpretation, 235
- Investments, 177, 185, 186
- IoT. *See* Internet of things (IoT)
- i-Spaces, 291
- Iterative analysis, 151–152
- Iterative data streams, 156–157
  
- J**
- Jaql, 147
- Java Message Service (JMS), 44

**K**

Kafka, 47, 208  
 Kaggle, 95  
 Karma, 103  
 Key insights for big data analysis, 64–67  
 Key insights for big data curation, 88  
 Key insights for big data storage, 121  
 Key-value stores, 124, 266  
 Knowledge, 171
 

- acquisition, 149
- economy, 31
- graph, 73, 132
- models, 266
- as a service, 146, 160
- workers, 161

**L**

Laboratory information system (LIS), 52  
 Labour market, 251  
 Lambda architecture, 208, 209  
 Language(s), 77
 

- detection, 218
- modelling, 221

 Large-scale, reasoning, benchmarking, and machine learning, 67  
 LarKC, 68  
 Law of information recovery, 241  
 LDBC. *See* Linked data benchmark council (LDBC)  
 Legal, 9  
 Legal and policy, 290  
 Legislation, 177, 184  
 Legislative, 173  
 Liability, 218  
 Liberalization, 232  
 Liberalized markets, 227  
 Life sciences, 110  
 Lighthouse projects, 291  
 Linked data, 64, 72, 75, 107, 158, 207, 219, 240, 241, 243, 244, 252, 272  
 Linked data benchmark council (LDBC), 68  
 Linked graphs, 153  
 Linked open data cloud, 96  
 Logistics, 82  
 Logistics chains, 176  
 Long tail of data variety, 92, 109  
 Lucene, 147

**M**

Machine learning (ML), 69–70, 80, 90, 104, 152, 163, 176, 222, 223, 242, 243, 272  
 Machine readable, 189, 243  
 Magna carta, 177

Mahout, 69  
 Manufacturing, 7, 172
 

- retail, transport, 54–55
- sector, 175

 Map/Reduce, 151  
 MapR, 121  
 MapReduce, 46, 69, 75, 147  
 Market impact of big data, 198, 212  
 Market manipulation, 214  
 Market trading analysis, 213  
 Market value, 5  
 Marketing, 155, 171, 175  
 Marketplaces, 144  
 Master data management (MDM), 94  
 Materialization, 156  
 Matlab, 147  
 Maturity and gaps, 18  
 Media, 93, 172  
 Media and Entertainment, 58, 112, 174  
 Medical data, 189  
 Meritocratic model, 110  
 Message queuing, 41  
 Metadata, 112, 159, 240  
 Methodology, 19  
 Microcharts, 153  
 Microsoft, 126, 212  
 Microsoft Azure, 127  
 Microsoft's Dryad, 147  
 Mind-set, 235  
 Modelling and simulation, 205, 209, 270–272, 274  
 MongoDB, 126  
 Monitoring energy consumption, 137  
 Multi-dimensional analytics, 241  
 Multimodal interfaces, 272

**N**

Named entity recognition, 272  
 Natural language analytics, 204, 208, 271, 272, 274  
 Natural language interfaces, 192  
 Natural language processing (NLP), 53, 90, 176, 192, 270  
 Natural language processing pipelines, 109  
 Near real-time visualization, 153  
 NESSI European Technology Platform, 24, 279, 289  
 New business models, 174, 231, 233–234  
 New York Times (NYT), 101, 107, 109, 113  
 NewSQL databases, 123, 125  
 Non-social, 266  
 Non-technical issues, 280  
 Non-technical requirements, 186–187, 203, 217, 253

- Normative analyses, 149
- NoSQL, 40, 121–124, 127, 130, 132, 134, 224, 229, 283
- Novartis, 102
- O**
- Obama administration, 198
- Obamacare, 184
- Oil and gas industry, 228–229
- Online analytical processing frameworks, 244
- Online transactional processing (OLTP), 123
- Ontologies, 176
- Ontology alignment, 107
- Open and interoperable data policies, 100
- Open data, 6, 55, 76, 87, 88, 109, 122, 155, 158, 170, 197, 199, 201, 234, 236–238, 240, 254, 266, 285, 287
- Open datasets, 76, 252, 255
- Open energy information, 237
- Open government, 197
- Open refine, 103, 106
- Open sensor, 170
- Open Street Map, 234
- Open Weather Map, 234
- OpenCorporates, 106
- Operational efficiency, 174, 184, 227, 231, 232, 272
- Operations, 171
- Oracle, 40, 212
- Organizational level, 171
- organizational factors, 173
- Organizational silos, 185
- Ownership, 145, 146, 158, 177, 228
- P**
- Parallel faceted browsing, 161
- Patient engagement applications, 183, 184
- Patient monitoring data, 185
- Pattern discovery, 204, 205, 271, 274
- Peer Energy Cloud (PEC), 137
- Permission management, 107, 192
- Personal data, 159, 188, 251, 256, 269
- Personalization, 155
- Personalized recommendation, 171
- Pfizer, 102
- Pharmaceutical data, 180
- Pig, 147
- Pipelines, 152
- Pipelining, 156
- Piracy, 251
- Pistoia Alliance, 102
- Planning and forecasting, 149
- Policy roadmap for big data, 284–286
- Policymakers, 177
- Political willingness, 201
- Pre-competitive, 89
  - collaboration, 102
  - partnerships, 102
- Predictions, 152
- Predictive analysis, 144, 147, 150–151, 162, 163, 232
- Predictive analytics, 145, 150, 162, 204, 208–209, 271, 272, 274
- Predictive and real-time analysis, 232
- Predictive lifecycle management, 233
- Predictive maintenance, 145, 150, 176, 272
- Predictive models, 150, 239, 242
- Predictive modelling, 230, 243
- Prescriptive analysis, 162
- Prescriptive analytics, 242–243, 271, 272, 274
- Press Association (PA), 109, 112
- Privacy, 8, 60, 67, 76, 81, 122, 129–131, 146, 155, 197, 203, 204, 207, 217, 220, 228, 233, 235, 238, 240, 269
  - and confidentiality, 174, 239
  - by design, 129, 224, 270, 283
  - enhancing methods, 188
  - and legal, 285
  - preserving analytics, 237, 241
  - protection, 207, 240, 269
  - and security, 127, 177
  - and security issues, 173
  - and trust, 287
- Privacy-aware, 173
- Problem solving, 244
- Processes, 170, 171
- Product(s), 170, 171
  - data, 175
  - development, 250
  - innovation, 233
- Productivity, 172
- Proprietary APIs, 220
- Prosumer, 233
- Protection, 146
- Protein Data Bank (PDB), 107, 111
- Protocols, 42–44
- Provenance, 90, 106, 129, 131, 159, 207
  - management, 106, 192, 268
  - metadata, 131
- Pseudo-identifiers, 191
- Pseudonymization, 159, 191, 231
- Public, 172
  - consultation, 280
  - health analytics, 182
  - private partnership, 290
  - sector, 6, 79, 170, 172
- Public-private hybrid environments, 224
- Public-private partnership (PPP), 7, 89, 102



Public sector information (PSI), 202  
 Publish/subscribe, 41

## Q

Quality, 182  
 Quality of Experience (QoE), 272  
 Query interfaces, 130  
 Query interfaces and languages, 147  
 Querying, 106  
 Querying platforms, 123  
 Qurk, 96

## R

Rackspace, 126  
 Raw data, 63, 137, 146, 228, 230, 260  
 RDF. *See* Resource description framework  
 RDF data stream, 70–71  
 Real-time, 112, 135, 155, 162, 175, 204, 229, 230, 232, 235, 241, 243, 267  
   analysis, 60  
   analytics, 238, 242  
   data, 208, 228  
   data stream, 272  
   data transmission, 264, 274  
   decision automation, 242  
   insights, 204, 207, 219, 271, 274  
   integration, 219  
   processing, 75  
 Reasoning, 67, 272  
 Reduce costs, 174  
 Regulation, 146, 175, 176, 231, 233, 237, 269  
 Regulators, 215  
 Regulatory, 121, 144, 145, 158, 159, 174, 217, 228, 234, 235  
 Regulatory obligations, 231  
 Relational database, 216  
 Relational Database Management Systems (RDBMS), 127, 147  
 Renewable energy sources, 234  
 Requirements, 238, 273  
 Research data, 185  
 Resource-centric infrastructure, 239  
 Resource description framework (RDF), 68, 72, 107, 125  
 Resource efficiency, 173  
 RESTful, 51  
 Retail, 6, 81, 113, 170, 172  
 Retail sector, 175  
 Revenue, 173  
 Reverse transparency, 231  
 Revolutionary, 176  
 Roadmaps, 14, 15, 22, 24, 192, 212, 263, 280  
 Roadmapping, 22

Robotic curation, 111  
 Rolls Royce, 150

## S

S4 (simply scalable streaming system), 46  
 Sandboxing, 131  
 SAP HANA, 133  
 SAS Analytics, 147  
 SAS Teragram, 113  
 Sawzall, 147  
 Scalability, 92  
 Scalable advanced analytics, 230  
 Scale-out, 123  
 Schema mapping, 107, 108  
 Schema-agnostic query, 192  
 Schema-on-read, 238  
 Science, 98  
 Scientific efficiency, 101  
 Secondary usage of health data, 183  
 Secondary use, 231, 233  
 Sector, 172  
 Sectorial forums, 17, 21  
 Secure communication, 128  
 Secure data exchange, 191, 269  
 Secure data storage, 128  
 Security, 122, 127, 129–131, 216, 219  
   architecture, 224  
   by design, 224  
 Semantic approximation, 68, 106  
 Semantic data, 266  
   enrichment, 182, 187, 189–190, 257, 259  
   models, 131, 190, 266  
 Semantic event-matching, 72  
 Semantic interoperability, 191  
 Semantic knowledge models, 190  
 Semantic linkage, 244  
 Semantic pattern, 272  
 Semantic systems, 266  
 Semantic techniques, 64, 73, 241  
 Semantic technologies, 160, 176  
 Semantic validation, 268  
 Semantically annotated, 266  
 Semantically enriched, 182  
 Semantics, 153, 157–159, 230  
 Semi-structured data, 123, 217  
 Semi-structured knowledge, 109  
 Sensor(s), 80, 170, 175, 228, 267  
   data, 228, 230, 268  
   networks, 70  
 Sentiment, 223, 254  
   analysis, 197, 272  
   classification, 218  
 Servers, 147  
 Service integration, 145–146

- Services, 171
  - Sharding schemes, 125
  - Sharing, 155, 266
    - data, 6
    - of health data, 182
  - Sheer curation, 94
  - Shell, 229
  - Short-term, 98
  - Siloed data ownership, 240
  - Silos, 199, 235
  - Simplicity, 75
  - Simulation, 144, 147, 154, 176, 243
  - Siri, 257
  - Situation adaptivity, 155
  - Situational awareness, 267
  - Skilled, 243
  - Skills, 9, 121, 201, 203, 216, 234–235, 238, 290
  - Skillset, 229, 249
  - Small and Medium-sized Enterprises (SMEs), 251
  - Smart appliances, 228
  - Smart buildings, 7
  - Smart cards, 229
  - Smart cities, 7, 34, 79, 146, 198, 204, 232, 267, 271
  - Smart data, 144–146, 157–159, 174, 227, 230
  - Smart devices, 228
  - Smart gas, 232
  - Smart grids, 135, 137, 146, 232, 238
  - Smart machinery, 176
  - Smart meter, 7, 137, 233
  - Smart metering, 41, 57, 230, 232, 240
  - Smart oil, 232
  - Smart products, 159, 175
  - Social, 10, 265, 290
    - and economic impact for big data usage, 146
    - engagement mechanisms, 98, 101
    - incentives and engagement mechanisms, 100–101
    - media, 135, 197
    - media analysis, 249, 250
    - network, 155
    - web, 74
  - Social and economic impact of big data, 41
    - curation, 92
    - storage, 122
  - Societal challenges, 170
  - Society roadmap for big data, 286–287
  - Software as a service, 160
  - Solr, 147
  - Spark, 126, 151, 207
  - SPARQL, 72, 130
  - Spatiotemporal databases, 209
  - Spring framework, 47
  - SPSS, 147
  - SQL, 148, 244
  - Stakeholder(s), 17, 34, 170, 172, 173, 177, 230, 233, 263, 273, 288
    - business cases, 187
    - groups, 181
    - interact, 182
  - Standardization, 175, 237, 283, 290
  - Standardization and interoperability, 98, 107–108
  - Standards, 121
  - Start-ups, 34, 237, 280, 287
  - Storm, 45, 208
  - Strategic business decisions, 155
  - Strategic research & innovation agenda (SRIA), 24, 280
  - Strategic research & innovation agenda (SRIA) on big data value, 290
  - Stratosphere, 147, 152
  - Stream data mining, 70
  - Stream data processing, 70–72
  - Stream pattern matching, 205
  - Stream-based data mining, 223
  - Streams from sensors, 254
  - Structured data, 93, 105, 123, 136, 216
  - Subjunctive interfaces, 161
  - Sub-level requirements, 264
  - Summarization, 105
  - Supervisory Control and Data Acquisition (SCADA), 228
  - System models, 238, 243
  - SystemT, 147
- T**
- Tableau, 105, 147
  - Taxonomy managers, 113
  - Technical, 9
    - infrastructure, 177, 286
    - papers, 280
    - Requirements, 187–189, 204, 218, 253
    - working groups, 17, 20
  - Technology, 290
    - roadmap, 189–192, 205–209, 220–224, 231, 239–244, 257, 273
    - roadmap for big data, 281
  - Telecom, 172
  - Telecom sector, 175
  - Temporal alignment, 23
  - Tensor modelling, 242
  - Text analysis, 189
  - Third-party data, 236
  - Thomson Reuters, 109
  - Time-stamped, 236

- Trade-Offs, 148
- Traffic, 79
- Traffic management systems, 230
- Transaction Logs, 128
- Transactional guarantees, 120
- Transfer of technology, 158
- Transform, 121
- Transformational change, 144
- Transformations, 227
- Transparency, 146, 155
- Transport, 172
- Treato, 135
- Trinity, 132
- Triple stores, 266
- Trust, 98, 106–107, 235, 269
- Twitter, 122, 237, 257
  
- U**
- UIMA, 190
- Unilever, 114
- Unmanned aerial vehicles (UAV), 257
- Unstructured and structured data
  - integration, 109
- Unstructured data, 93, 105, 109, 112, 123, 136, 182, 211, 212, 217, 254, 272
- Unstructured-structured integration, 98
- Unstructured text, 189
- Urban, 227, 230
- US Healthcare Reform, 184
- Usage analytics, 271, 274
- Usage data and patterns, 236
- Usage rights, 159
- User acceptance, 241
- User adaptivity, 153
- User experience, 270, 274
- User-generated, 170
  
- V**
- Value, 8, 29, 30, 35, 39, 59, 93, 135, 143, 151, 170–172, 174, 176, 197, 211, 227, 242, 260
  - chains, 15, 30, 131, 143, 159, 171, 175, 176, 232, 235
  - of big data, 180
  - generation, 280
  - Value-based healthcare delivery, 184
  - Value-based incentives, 187
  - Value-based system, 182
  - Variety, 30, 64, 120, 122, 136, 146, 153, 177, 234, 238, 240, 241, 244, 252, 266
  - Variety of health data, 180
  - Velocity, 30, 40, 64, 70, 120, 153, 177, 211, 217, 227, 230, 234, 238, 241, 252, 267
  - Veracity, 30, 230
  - Vertica (HP), 126, 147
  - Virtual representation, 228
  - Virtual value chain, 171
  - Virtualization, 131
  - Virtualized service infrastructure, 160
  - Visual analytics, 146, 153, 157, 162, 270
  - Visual perception, 157
  - Visual queries, 153, 270
  - Visualization, 105, 147, 152–153, 157, 270, 273
  - Vivisimo, 40
  - Vocabularies, 107
  - VoltDB, 125
  - Volumes, 30, 64, 120, 136, 137, 153, 170, 177, 211, 227–230, 234, 238, 241, 252
  - Volume of structured data, 180
  - Volume, variety, velocity, and veracity, 252
  
- W**
- W3C, 56
- W3C PROV, 106, 108
- Web scale reasoning, 68
- Web standards, 240
- Wiki, 95
- Wikidata/DBpedia, 252
- Wikipedia, 95, 101, 104
- Wikipedia bots, 96
- Wisdom of crowds, 95
- Workflows, 108
  
- Y**
- YouTube data warehouse (YTDW), 148
  
- Z**
- ZenCrowd, 96