



Stichwortverzeichnis

Pavlo Baron

Big Data für IT-Entscheider

Riesige Datenmengen und moderne Technologien gewinnbringend nutzen

ISBN (Buch): 978-3-446-43339-7

ISBN (E-Book): 978-3-446-43392-2

Weitere Informationen oder Bestellungen unter

<http://www.hanser-fachbuch.de/978-3-446-43339-7>

sowie im Buchhandel.

# Stichwortverzeichnis

## A

- Abstraktion 43
  - Schicht 43
  - Service 43, 99, 104, 133, 185
- Analyse 9, 31–33, 40, 57, 70, 72, 75, 76, 86, 102, 111, 166, 178, 189, 194, 196, 198, 206
  - Aggregate 101
  - Aggregation 168, 199
  - Algorithmus 40, 50, 69, 94, 101, 115, 178, 199, 203
  - Effizienz 94, 135
  - False Negative 119
  - False Positive 119
  - Feature 73
  - Machine Learning 6, 38, 41, 45, 69, 149, 179, 200, 206
    - Accuracy 48, 70
    - Error Rate 48
    - Klassifizierung 41, 113
    - Supervised Machine Learning 41
    - Unsupervised Machine Learning 41
  - Mathematik 40–42, 48, 67, 73, 149
  - Natural Language Processing 6, 41, 45, 69, 112, 151, 178, 185
    - Document Classification 42
    - Sentiment Analysis 42, 114
    - Trigramm 112
  - Prognose 40, 41, 57, 70
  - Sample 103, 150
  - Statistik 38, 40, 64, 69, 149, 199, 206
  - Wahrscheinlichkeit 38, 40, 70
- Anbieter 8, 196, 207
  - 10gen 155
  - Akamai 32
  - Basho 153
  - BITKOM 207
  - Cloudera 164
  - Coursera 46, 206
  - DataSift 57
  - DataStax 153
  - EMC 160
  - Gnip 57
  - HortonWorks 164
  - Loggly 167
  - MapR 164
  - Neo Technology 158
  - Oracle 197, 199
  - Postgres 199
  - Splunk 167
  - VMware 157
- Anwendungsfall 29, 93, 109, 115, 158, 196, 200
  - Überwachung 33
  - CERN 101
  - Filterung 71
  - Fraud Detection 31
  - Klick-Zähler 52
  - Library of Congress 105
  - Logs 45, 50, 54, 158, 168, 169, 179, 200
  - Meinungsbildung 57
  - Routen-Optimierung 33
  - Spracherkennung 70, 112
  - Time-Series 157, 158, 168, 169
  - Tracking 61, 120, 199, 200

- Videostrom 33
- Zahlungsverkehr 95

## B

- Best Practices 72, 164
- Betriebssystem 43, 45, 146, 174
- Business Intelligence 2, 29, 59, 74, 186, 196, 198, 202
- DWH 187
- ETL 198
- Mining 187
- OLAP 187

## C

- Caching 68
- Cache-Miss 144
- Kohärenz 145
- Write-Behind-Cache 92, 144
- Write-Through-Cache 92, 144
- CAP 93, 138, 140
- AP System 141
- CP System 139
- Partition Tolerance 139
- Cloud 8, 28, 86, 94, 96, 154, 159, 170, 176, 188, 196, 205
- AWS 100
- CloudFront 99
- EC2 86, 100, 103
- Riak-CS 154, 196
- S3 86, 96, 154, 196
- CRM 26, 158, 183

## D

- Datenbank 38, 49, 64, 66, 72, 77, 78, 92, 104, 130, 139, 150, 152, 156, 161, 165, 166, 168, 178, 194, 195, 198, 202
- Backup 96, 121
- Chunk 164
- Column 154
- Graphendatenbank 195
- HDFS 115
- Infrastruktur 178
- Konsistenz 121, 139, 140
- NewSQL 160, 194, 198
- NoSQL 66, 72, 88, 123, 151, 159, 160, 168, 178, 184, 194, 198, 204, 207

- Document Store 72, 155, 159, 198
- Graphendatenbank 158, 179, 198
- In-Memory Store 68, 146, 156, 197, 198
- Key/Value Store 72, 152, 175, 184, 199
- Partitioning 91, 130
- RDBMS 67, 72, 89, 123, 126, 131, 147, 151, 155, 158, 161, 178, 194, 195, 198
- Read 44
- Relation 198
- Schlüssel 130, 152, 155
- Auto-Increment 122
- Hashwert 78, 94, 119, 122, 136, 199
- ID 78, 94
- UUID 119, 122
- Stored Procedure 90
- Version 126
- MVCC 130
- Siblings 137, 153
- Tombstone 137
- Write 44, 122, 137, 139, 153, 164

## E

- Enterprise-Architektur 91, 181

## F

- Framework 38, 43, 48, 70, 90, 195
- O/RM 38, 90

## G

- Graph 148, 157
- Knoten 148
- Pfad 148
- Traversieren 148

## H

- Hardware 28, 83, 85, 86, 170, 197
- Ausfall 94, 96
- Cluster 38, 83, 87, 93, 96, 118, 134, 138, 149, 163
- Commodity Hardware 28, 67, 83, 84, 170, 209
- CPU 85, 146, 171
- Firewall 107
- FPGA 175
- GPU 175

- Hauptspeicher 37, 71, 197, 198
- HDD 172
- I/O 171
- Infrastruktur 92, 134
  - DMZ 107
- Kapazität 37
- Knoten 93, 122, 124, 129, 137, 153
- Maschinerie 28
- Netzwerk 43, 85, 103, 158, 171, 173
  - Backpressure 118
- Physik 84
- Platte 37, 67, 71, 85, 119, 146, 171, 198
- RAID 170
- Rechnerarchitektur 43
- SSD 171
- Virtualisierung 86

**I**

- Industrie 194
- Öffentlicher Verkehr 32
- Ad-Tech 22
- Bank 8, 13, 31, 54, 95
- Gambling 32
- High Frequency Trading 22, 76
- Immobilien 12
- Krankenhaus 75
- Online-Spiele 31
- Partnervermittlung 31
- Telekommunikation 15, 79, 169
  - CDR 59, 169
- Traveling 76
- Versicherung 8, 10, 13, 31, 58, 60

**K**

- Kosten 28, 35, 83, 85, 198, 204
  - CapEx 35, 204
  - Investition 35, 162, 203
  - OpEx 35, 204
  - ROI 203
  - TCO 203

**L**

- Location 57, 59, 60, 112, 199
- IP-Adresse 56, 59, 87

**Locking 127**

- Optimistic Locking 127, 143
- Pessimistic Locking 127

**M**

- MapReduce 6, 162, 187, 196
  - In-Database-MapReduce 154, 162
  - Split 196
- Marketing 42, 73, 200
  - Cross-Selling 55
  - Werbung 26, 59
- Modell 64, 89, 92, 152
  - Relation 67, 68, 89
    - Join 68
  - Schema 68

**N**

- New Economy 43
  - Amazon 8, 22, 86, 99, 189
  - Boundary 18
  - Facebook 2, 13, 22, 26, 54, 66, 108
  - Foursquare 60
  - Google 11, 22, 66, 73, 77, 83, 88, 94, 106, 112, 154, 207
  - Instagram 14
  - LinkedIn 47, 207
  - Netflix 100, 190
  - Prismatic 17, 207
  - Qype 60
  - Simple 54
  - Square 16
  - Startup 14, 16, 33, 88, 103, 194
  - Twitter 6, 15, 22, 30, 55, 66, 88, 95, 110, 112, 206
  - Vimeo 99
  - Yahoo 88
  - YouTube 99

**P**

- Prozess 39, 44, 181
  - Configuration Management 87
  - Continuous Delivery 43, 87
  - DevOps 43, 185
  - Manipulationssicherheit 43
  - Qualitätssicherung 43

- Rollentrennung 43
- Vieraugenprinzip 43

## Q

- Query 160, 195
  - CQL 195
  - CRUD 64, 88
  - Cypher 195
  - SQL 35, 195
- Queue 102, 104, 111, 118, 175
  - JMS 111
  - Publish/Subscribe 102

## S

- Schmerzen 11, 12, 22, 50, 57, 83, 89
  - Alarm 101
  - Anomalie 101, 103
  - Aufbereitung 97
  - Batch 102, 168, 169, 179, 189, 196, 197, 200, 202
  - Bottleneck 104, 128, 133
  - Datenmenge 37, 80, 91, 97, 176, 196
  - Datensicherheit 105, 107
    - Anonymous 107
    - Lulzssec 107
  - Datenverlust 126
  - Echtzeit 100, 169, 173, 194
  - Fast-Echtzeit 71, 72, 102, 166, 168, 172, 174, 179, 187, 197, 202
  - Latenz 95, 139, 161
  - Performance 45, 85, 90, 103, 150
  - Single Point of Error 128
  - Single Point of Failure 119, 133, 164
  - Social Hacking 107
  - Symptome 50
  - Transformation 103
  - Variety 23, 25
  - Velocity 23
  - Verfügbarkeit 95, 99, 139, 161
  - Volume 23, 24
- Schnittstelle 196
  - JDBC 195
  - Protocol Buffers 154, 184, 195
  - REST 104, 154, 158, 184, 195
  - Service 62

- Spring Data 195
- Thrift 184, 195
- SOA 62, 185
- Soziale Netzwerke 12, 25, 26, 31, 55, 70, 88, 110, 200, 207
- Staff
  - Data Scientist 45, 48, 202
  - DBA 93, 130, 162, 177
  - Team 35
- Strategie 2, 21, 47, 61, 87, 196
  - Business 25
  - Business-IT-Alignment 205
  - Buy 204, 209
  - Closed Source 88
  - Community 27, 88, 204
  - Crowdsourcing 58, 60, 104
  - Datenschutz 56, 59, 105, 108
  - Datenzentriertheit 2, 10, 14
  - Empfehlungssysteme 2, 9, 79, 208
  - Entscheidungsunterstützungssysteme 2, 34, 193, 208
  - Expertise 41
  - Forschung 46, 59
  - Full-Stack 45, 205
  - Gamification 14, 53, 108, 201
  - Human Business 203
  - Information 1, 21, 31, 35, 37, 41, 50, 57, 59, 62, 69, 70, 74, 198, 200, 202, 203
  - Innovation 2, 11, 105, 199, 204
  - Make 204, 209
  - Marktvorteil 36, 40, 84
  - Master Data Management 64, 182
  - Mechanical Sympathy 44
  - Open Source 38, 47, 87, 112, 153, 175, 204
  - Perspektive 27, 89, 207
  - Policy 39, 96, 203
  - Polyglott 49, 151, 167, 190
  - Qualitätssteigerung 46
  - Scale-Up 161
  - Technologie 20, 28, 37, 43, 44, 49, 68, 72, 94, 110, 117, 166, 176, 181, 185, 197, 198
  - Transparenz 54
  - Trial-And-Error 40
  - User Experience 53

- Visualisierung 80, 151, 167, 180, 188, 191, 206
  - Dashboard 102, 183
- Wissenschaft 12, 40, 41, 47, 69, 150, 206
- Suchfunktion 77, 90
- Secondary Index 154, 175
- Trefferliste 77

## T

Tool 38, 109, 117, 150, 177

- Akka 174
- Appliance 43, 176, 197, 201, 205
- Bash 49
- BerkleyDB 159
- C10K-Server 174
- Cascalog 206
- Cassandra 153, 195
- Clojure 48, 166
- Corpora 70, 112
- Couchbase 156
- CouchDB 155, 184
- D3 180
- Disco 6, 115, 162
- DiscoDB 164
- ElasticSearch 176
- Erlang 49, 115, 155, 166, 169, 174, 185
- ESB 36, 43, 98, 104
- F# 48
- Funktionale Programmierung 48
- Greenplum 160
- Hadoop 2, 6, 49, 88, 102, 115, 116, 151, 162, 164, 168, 184, 191, 196, 201
- HBase 164
- HDFS 164, 184, 196
- Hive 165, 168
- IPython 178, 180
- Java 48, 49, 111, 150, 158, 166, 175
- JavaScript 49, 81, 104, 111, 166, 167
- JSON 104, 111, 115
- Julia 150
- jVectorMap 81
- JVM 49, 115, 155, 158, 174, 196, 197
- LingPipe 151
- Lucene 175
- Mahout 151
- MATLAB 150

- memcached 174
- MongoDB 2, 6, 49, 155, 165, 184, 191, 198, 201
- MySQL 49, 178
- Neo4j 6, 158, 169, 195
- Nginx 6, 174
- NLTK 112, 151
- Node.js 167, 174
- NuoDB 160
- Octave 150
- OpenNLP 151
- Oracle Loader for Hadoop 88, 196
- OrientDB 159
- Perl 49
- Pig 165, 168
- Python 48, 49, 110, 150, 166, 180, 185
- R 48, 49, 150, 166, 178
- RabbitMQ 6, 155, 185
- Redis 6, 157, 169, 174
- Riak 6, 153, 159, 184, 196
- Ruby 166
- SAP HANA 159, 197
- Scala 48, 49, 166
- Sensor 75
- Solr 176
- Storm 168, 174, 206
- Tool-Chain 109
- Tweepy 110
- VoltDB 160
- XML 174
- ZeroMQ 175
- Transaktion 91, 118, 140
  - ACID 91, 140, 160, 161
  - BASE 140
  - Zweiphasen-Commit 91
- Trend 102, 106, 194

## U

- Uhrzeit 94, 118, 153
- Atomuhr 128
  - NTP 128
  - Zeitstempel 48, 91, 94, 127, 168

## V

Version 155

- Verteilte Systeme 6, 44, 48, 49, 95, 102, 117, 122, 198
  - Anti-Entropy 137
  - Asynchrone Kommunikation 118
  - Callback 95
  - Clock 128
  - Complex Event Processing 101, 167, 168, 174
  - Consens 119
  - Consistent Hashing 134, 136, 147
  - Content Delivery Network 99, 175
  - Data Locality 86, 115, 196
  - Dynamo 153
  - Event 95, 101, 168
  - Eventual Consistency 95, 141, 154
  - Fault-Tolerance 44
  - Garbage Collection 137
  - Gossip 138, 153
  - Heartbeat 119
  - Hinted Handoff 139
  - Hinted Handoff 124, 140, 153
  - Kompensation 95, 118
  - Konsistenz 9, 95
  - Merkle-Tree 120, 153
  - Message 95
  - Parallelität 44, 49, 145
  - Polling 95
  - Prozessor 97, 115, 163
  - Quorum 119, 125
    - Sloppy Quorum 125, 136
  - Redundanz 147
  - Rehashing 133
  - Replikation 96, 121, 122, 147, 202
    - Lazy Replication 122
    - Master/Master 148, 161
    - Master/Slave 122, 148, 157, 158, 161
    - Operation Transfer 121
    - State Transfer 121
  - Ring 134, 153
  - Session 142
  - Sharding 91, 130, 136, 147, 149, 156
  - Snapshot 96, 187
  - Split-Brain-Effekt 119
  - Timeout 118
  - Vector Clock 129, 136, 153
  - Voting 119
  - Worker 163