

# Index

- $\theta$ -join, A-7, A-10
- n*-ary integration, 292
- , 147
- 2PC, *see* two-phase commit
- 2PL, *see* two-phase locking
- 3NF, A-5
- 3PC, *see* three-phase commit
  
- abort, 183, C-4, C-38
- abort list, C-46
- access control, 91, 102
- access frequency, 40
- access path, 23
- access path selector, 23
- ACID properties, 183, C-9
- ACID transaction, 513, 515, 517, 519
- action model, C-14
- activation queue, 371
- activity, C-19
- ad-hoc data delivery, 6
- AdaptCache, 80
- adaptive query processing, 172
- adaptive reaction, 173
- adaptive virtual partitioning, 379
- affix, 287
- after image, C-43
- aggregate assertion, 121
- aggregate constraint, 117
- algebraic query, B-9
- algebraic query *see* query
  - algebraic 127, 134, 136–138
- allocation, 33, 43, 49, 51, 67–69, 72, 73, 84, 85, 89
- Amazon Redshift Spectrum, 536
- Amazon SimpleDB, 506
- Ambari, 496
- analytical graph query, 473
- analytical graph workload, 473
- anomaly serializability, C-13
- AP, *see* asynchronous parallel
- Apache Flink, 471
- Apache Giraph, 484
- Apache Giraph, Giraph, 500
- Apache Ignite, 518
- Apache Storm, 458, 469, 499
- APPA, 399, 408, 410, 421, 423, 434
- application server, 21
- apprentice site, 163
- ArangoDB, 517
- archive, C-49
- ARIES, C-43
- ARTEMIS, 327
- AsterixDB, 509, 535, 536
- asynchronous parallel, 482
- at-least-once semantics, 471
- at-most-once semantics, 471
- atomic commitment, 208
- atomic operation, C-7
- atomicity, C-9
- attribute, A-1
- attribute affinity matrix, 55, 60
- attribute affinity measure, 54
- attribute usage value, 54
- Aurora, 458, 466
- AURORA data integration system, 281
- Aurora DSMS, 458, 461, 498, 499
- authorization, 91
- authorization matrix, 104
- Auto-Detect, 588
- autonomy, 18
  - communication, 303
  - design, 303
  - execution, 303
- Autoplex, 290

- availability, 15
- AVP, *see* adaptive virtual partitioning
- AWESOME, 536
- Azure HDInsight, 442
  
- B-tree index, 351, 354
- backlink, 543
- bandwidth, 11, D-7
- BAP, *see* barrierless asynchronous parallel, 488
- barrierless asynchronous parallel, 488
- base relation, 92
- basic Paxos, *see* Paxos
- BATON, 395, 415, 416
- BATON\*, 395
- before image, C-43
- behavioral conflict, 285
- behavioral constraint, 110
- Bell number, 53
- Best Position Algorithm, 406
- big data, 3, 16, 20, 437
- big data application, 437
- big data processing, 437
- big data processing system, 437
- BigchainDB, 432
- BigDAWG, 529, 533–536
- BigIntegrator, 521, 522, 533, 536
- BigTable, 533–535
- Bigtable, *see* Google Bigtable, 510, *see* Google Bigtable, 522
- binary integration, 291
- binary table, 581, 593
- Bio2RDF, 573
- bitcoin, 424, 427
- Bitcoin-NG, 431
- BitTorrent, 386, 388, 432
- block-based storage, 440
- block-centric graph model, 481
- BLOCKBENCH, 432
- blockchain, 424, 432
  - non-permissioned, 426
  - private, 426
  - public, 426
- Blockchain 2.0, 430
- blockchains
  - permissioned, 426
- Blogel, 490
- Bond Energy Algorithm, 56
- Borealis, 458, 499
- bottom-up design, 16, 277, 279
- Boyce-Codd normal form, A-5
- BPA, *see* Best Position Algorithm
- broadcast network, D-6
- BSP, *see* bulk synchronous parallel, 481, 482, 488
  
- bucket algorithm, 309
- buffer manager, C-20
- bulk synchronous parallel, 481
- bushy join tree, 148, B-17
- bushy query tree, 362
  
- cache manager, 24
- calculus query, 127
- CAN, 393, 418
- candidate key, A-2
- candidate set cover, 296
- canonical data model, 279
- CAP theorem, 502, 503, 534–536
- carrier sense medium access with collision
  - detection, D-5, D-12
- Cartesian product, A-7, A-9
- cascading abort, C-12, C-31
- Cassandra, 506, 510, 512
- catalog, 13, 83
- Catalyst, 530
- Causal clustering, 515
- Causal consistency, 515
- cell, D-6
- cellular network, D-6
- Ceph, 442
- ch-rep-s:eager, 384
- chained partitioning, 353
- chained query, 154
- chameleon-db, 582, 593
- checkpointing, C-48
  - action-consistent, C-49
  - fuzzy, C-49
  - transaction-consistent, C-49
- Chord, 394, 504
- chunk, 441
- circuit switching, D-7
- CLAMS, 594
- cleaning operator, 301
- client manager, 345
- client/server, 4, 17, 19–21, 23, 24
  - multiple client/multiple server, 20
  - multiple client/single server, 20
- cloud, 17, 443
- cloud computing, 27, 438
- CloudMdsQL, 520, 529, 532–536
- clustered affinity matrix, 56, 57, 60, 61, 63, 64, 87
- clustering, 56
- CockroachDB, 518
- column-store, 36
- COMA, 287
- commit, 183, C-4
- commit list, C-46
- commit protocol, 207

- committable state, 223
- communication cost, 132
- communication links, D-7
- communication time, 156
- complexity of relational algebra operators, 132, B-14
- composite matching, 291
- concurrency control, 183, 235, C-22
  - locking, 186, 187
  - optimistic, 186
  - pessimistic, 186
  - timestamp ordering, 186
- concurrency level, C-22
- conditional data delivery, 6
- conflict, C-23
  - read-write, C-23
  - write-read, C-23
  - write-write, C-23
- conflict equivalence, C-26
- conjunctive normal form, B-4
- conjunctive query, 306
- connection graph, 135, B-6
- consistency, C-10, C-22
  - degree 0, C-10
  - degree 1, C-10
  - degree 2, C-10
  - degree 3, C-10, C-26
  - strong, 244
  - weak, 244
- consistent hashing, 504
- constraint-based matching, 288
- containment edge, 289
- contingency task, C-20
- continuous processing model, 458
- continuous query, 6
- Continuous Query Language, 461, 498
- coordinator timeout, 217
- Cosmos DB, 517
- Cost Functions, 155, B-18
- Cost Model, B-18
- cost model, 63, 130–133, 167, 175, 176, 264, 312–317, 322, 328, 360, 363, 364, 367–369, 381, 382, 531–534, B-15, B-18
  - heterogeneous, 311, 317, 318, 328
  - mediator, 313, 338
- cost model, 155
- Couchbase, 509, 535
- CouchDB, 509
- COUGAR, 460, 499
- counting algorithm, 98
- CPU cost, 132
- CQL, 510
- crash recovery, C-9
- crawler, 542–544
  - focused, 545
  - incremental, 544
  - parallel, 545
- crawling, 563
- cryptocurrency, 424
- CSMA/CD, *see* carrier sense medium access
  - with collision detection
- cuclik query, 152
- cursor stability, C-12
- Cypher, 513, 515
  
- DaaS, *see* Database-as-a-Service, *see* Database-as-a-Service
- dark web, 5, 539
- DAS, *see* directly attached storage
- data center, 29
- data cleaning, *see also* data quality, 438, 567
  - instance-level, 301
  - schema-level, 301
  - web, 585
- data control, 14, 91
- data dictionary, 83
- data directory, 83
- data distribution, 13
- data encryption, 102
- data fusion, 587, 594
- data independence, 3
  - logical, 8
  - physical, 8, 9
- data integration, 279, 496, 497
  - web, 566
- data lake, 5, 277, 325, 493, 500, 567, 585
- data locality, 11, 33
- data localization, 33, 127, 134, 136, 138, 175, 325
- data partitioning, *see also* fragmentation, 8, *see also* fragmentation, 11, 34, 73, 80, 84, 85, 239, 345, 349, 351, 380, *see also* fragmentation, 475, 518, 519, 583
  - adaptive, 78, 79
  - workload-aware, 74, 78
- data processor, 23, 346
- data protection, 102
- data quality, 17, 496, 497
  - web, 562, *see also* data cleaning, 585
- data replication, 243
- data skew, 73, 366
- data spaces, 567
- data stream, 458, 460, 498
- data stream management system, 458
- data stream processing system, 458
- data transfer rate, D-7
- data translation, 299
- data veracity, 585

- data warehouse, 278, 293, 301, 493
- database administrator, 91
- database buffer manager, 24
- database categorization, 565
- database cluster, 374
- database consistency, 15, 110, 181, 183, C-1
- database cracking, 81, 86
- database integration, 5, 16, 25, 277, 281
  - logical, 277, 278
  - physical, 277
- database integrity, 91
- database log, C-41
- database recovery, C-14
- database replication, 243
- database selection, 566
- database server, 21
- Database Statistics, 157, B-19
- database statistics, B-3
- DataGuide, 553
- Datalog, 305, 306, 400
- DB2 BigSQL, 495
- DBA, *see* database administrator
- DBPedia, 573
- deadlock, C-34
  - avoidance, 15, C-36
  - centralized detection, 192
  - detection, 15
  - detection and resolution, 191, C-37
  - distributed detection, 193
  - global, 191
  - hierarchical detection, 193
  - prevention, 15, C-35, C-37
- deadlock management, C-34
- decision tree, 290
- decomposition, B-25
- deep web, 539, 592
- degree distribution skew, 486
- deletion anomaly, A-4
- demand paging, C-40
- dependency conflict, 285
- dependency preservation, A-4
- detachment, 159, B-25
- deterministic search strategy, B-17
- differential file, C-44
- differential relation, 97
- DIKE, 290, 327
- DIPE, 327
- directly attached storage, 350
- dirty read, C-12
- disjointness, 37
- disjunctive normal form, B-4
- distributed computing system, 2
- distributed concurrency control, 10, 14
- distributed consensus, 228
- Distributed Cost Model, 155
- distributed database, 2
- distributed database design, 13
- distributed database management system, 2
- distributed database reliability, 15
- distributed deadlock, 191
- distributed directory, 91
- distributed execution monitor, 23
- distributed hash table
  - replica consistency, 417
- distributed join, 23
- distributed query, 33, 127
  - dynamic optimization, 128, 133, 134
  - execution, 134, 137
  - execution plan, 134
  - hybrid optimization, 128
  - optimization, 136
  - static optimization, 128, 163
- distributed query processing, 14
- distributed recovery protocols, 10
- distributed reliability, 10
- distributed storage system, 439
- distributed transaction, 34
- distributed transaction log, 231
- distributed transaction manager, 23
- division operator, A-7, A-14
- Document Type Definition, 571
- domain, A-1
- domain constraint, 112
- domain relational calculus, A-16, A-17
- domain variable, A-17
- DSMS, *see* data stream management system, 458
- DSPS, *see* data stream processing systems, 458
- DSS, 467, 471, *see* data stream system
- DTD, *see* Document Type Definition
- durability, C-14
- dynamic buffer allocation, C-40
- Dynamic distributed query optimization, 159
- dynamic programming, 133, B-17, B-27
- dynamic query optimization, B-2, B-3, B-24
- DynamoDB, 432, 503–505, 534, 535
- E-R model, 281, 329
- Eddy, 328
- eddy, 172–175, 177
- edge-centric asynchronous, 492
- edge-centric BSP, 492
- edge-centric graph model, 481
- edit distance, 287
- eDonkey, 432
- Edutella, 397, 400, 433
- elastic scalability, 443
- elasticity, 28

- element-level matching, 284, 285, 288
- elimination of redundancy, B-8
- entailment, 574
- entity-relationship data model, 281
- epidemic protocol, 391
- equijoin, A-11
- Esgyn, 518
- Estocada, 525, 528, 529, 533–535
- Ethernet, D-5
- Etherum, 428, 430
- ETL, *see* Extract-Transform-Load,
  - see* Extract-Transform-Load, *see* Extract-Transform-Load, *see* Extract-Transform-Load
  - see* Extract-Transform-Load, 494
- exactly-once semantics, 471
- exhaustive search, B-2
- external, B-28
- Extract-Transform-Load, 278
  
- F1, 518, 519, 534, 536
- failover, 343, 373–375, 441, 442
- failure, 15
  - communication, 206
  - hardware, C-38
  - media, C-39, C-49
  - site, 216, C-38
  - software, C-38
  - system, C-38
- failure transparency, 10
- failures of commission, 236
- failures of omission, 236
- federated database systems, 16
- fetch-as-needed, 164, 177
- file allocation, 67
- file storage, 440
- first normal form, A-5
- fix/flush, 231, C-48
- fix/no-fix decision, C-45
- fix/no-flush, C-47
- Flink, 458, 499
- FlumeJava, 448, 498
- flush/no-flush decision, C-45
- Flux, 469, 499
- force/no-force decision, C-45
- forcing a log, C-43
- foreign key constraint, 112
- Forward, 521, 523, 533, 536
- fragment, 13, 36, 37, 41–43, 45–50, 52, 53, 62–71, 73, 78, 84, 89
- fragment-and-replicate, 161, 356
- fragmentation, 8, 13, 33, 35, 37, 41–43, 48–52, 54, 55, 65, 66, 68, 72, 73, 84, 87, 89, 351
  - completeness, 36
  - derived horizontal, 37, 48
  - disjointness, 36
  - hash, 73
  - horizontal, 35
  - hybrid, 35
  - nested, 35
  - primary horizontal, 37, 40
  - range, 74
  - reconstructability, 36
  - round robin, 73
  - vertical, 35
- fragmentation predicate, 41
- Freenet, 388
- full reducer, 152
- fully decentralized top-k, 408
- fully duplicated, *see* fully replicated
- fully duplicated database, *see* fully replicated database
- fully replicated database, 13, 67
- functional dependency, A-4
- functional dependency constraint, 112
- fusion table, 567
- fuzzy read, C-13
  
- gap recovery, *see* at-most-once semantics
- Garlic, 319
- GAS, *see* gather-apply-scatter, *see* gather-apply-scatter
- gather-apply-scatter, 481, 482
  - asynchronous, 483
  - edge-centric, 492
  - partition-centric, 491
  - vertex-centric, 489
- GAV, *see* global-as-view, *see* global-as-view, *see* global-as-view
- GCS, *see* global conceptual schema, *see* global conceptual schema, *see* global conceptual schema, *see* global conceptual schema, *see* global conceptual schema, *see* global conceptual schema, *see* global conceptual schema, *see* global conceptual schema, *see* global conceptual schema
- general constraint, 112
- geo-distributed DBMS, *see* geographically distributed DBMS
- geographically distributed DBMS, 2
- GFS, *see* Google File System, *see* Google File System, *see* Google File System
- GFS2, *see* Global File System 2
- Gigascop, 458, 499
- Giraph, *see* Apache Giraph
- Giraph++, 490
- GiraphUC, 488
- GLAV, *see* global-local-as-view



- Hyperledger, 430
  - Fabric, 431
  - Iroha, 431
- hypernym, 286, 287
- I/O cost, 132
- IaaS, *see* Infrastructure-as-a-Service, *see*
  - Infrastructure-as-a-Service, *see*
  - Infrastructure-as-a-Service
- IBM DB2RDF, 579, 593
- ICQ, 386
- idempotency rules, B-8
- IEEE 802 Standard, D-12
- iMAP, 291
- in-place updating, C-41
- inclusion dependency, 110
- independent parallelism, 12
- independent recovery protocol, 208, 216
- individual constraint, 117, 119
- Infinite Graph, 516
- information integration, 279
- Infrastructure-as-a-Service, 5
- INGRES, 4, 31, 67, 92, 123, 439
  - distributed, 159, 177, 257, 273
- INGRES, 439
- inner join, A-10
- insertion anomaly, A-3
- instance matching, 284
- instance-based matching, 284–286
- integrity constraint, 181, C-1
- internal relation, B-28
- Internet, D-2
- Internet layer protocol, D-9
- Internet of Things, 430
- interoperability, 277
- interoperator load balancing, 368
- interoperator parallelism, 11, 12, 360
- interquery parallelism, 11, 33
- interschema rules, 287
- intersection operator, A-7, A-9
- intranet, D-2
- intraoperator load balancing, 367
- intraoperator parallelism, 11, 354, 360
- intraquery load balancing, 369
- intraquery parallelism, 11, 33, 36
- intraschema rules, 287
- inverse rule algorithm, 309
- IoT, *see* Internet of Things, 430
- isolation, C-11, C-22
- isolation level, 185, 186, 200, 233, 248, 250
- iterative improvement, B-2
- JAQL, 448, 498
- JDBC/ODBC, 448
- JEN, 533, 536
- Jena, 579, 581, 593
- job tracker, 446
- join, A-4
- join graph, 38, 40, 48, 49, 52, 87, 135, 149, 150, 152, 154, 174, 178–180, B-6
  - partitioned, 49
  - simple, 49, 52
- join implementation on MapReduce, 449
- Join Ordering, 148
- join ordering, 137, 146
  - distributed, 128, 146, 175
- join predicate, A-11
- join selectivity factor, B-19
- join tree, 61, 147
- join trees, B-15
- JSON, 506, 507, 509, 519, 523, 524, 535
  - binary, 507
- JXTA, 397, 433
- k-means algorithm, 454
- Kazaa, 386, 388, 417
- key, A-2
  - candidate, *see* candidate key
  - primary, *see* primary key
- key conflict, 285
- key-splitting, 468
- key-value store, 503
- KiVi, 520
- label propagation, 477
- LAN, *see* local area network
- landmark window, window
  - landmark, 460
- latency, 11
- LAV, 279, *see* local-as-view, *see* local-as-view, *see* local-as-view
- LCS, *see* local conceptual schema, *see* local conceptual schema, *see* local conceptual schema, *see* local conceptual schema, *see* local conceptual schema, *see* local conceptual schema, *see* local conceptual schema, *see* local conceptual schema
- LeanXcale, 233, 518–520, 534, 536
- learning-based matching, 290
- least recently used algorithm, C-40
- left linear join tree, 148
- left-deep tree, 362
- Lewenstein metric, 287
- LFGraph, 484, 500
- linear join tree, 148, B-17
- linguistic matching, 286
- link analysis, 547

- Linked Open Data, 567, 568, 584, 593
- load balancing, 364
- local area network, D-3, D-4
- local conceptual schema, 23, 277
- local directory/dictionary, 83
- local history, C-27
- local query, 137
- local query optimizer, 23
- local recovery manager, 23, 184
- Local Relational Model, 398
- local wait-for graph, 191, 192
- local-as-view, 279, 305, 306, 308, 326, 399
- lock, C-28
  - logical, C-28
  - manager, C-30
  - mode, C-29, C-32
  - point, C-31
  - unit, C-28
- lock mode, 187
- lock table, 187
- locking, 15
- locking algorithm, C-28, C-29
- locking granularity, C-28
- LOD, *see* Linked Open Data, *see* Linked Open Data, *see* Linked Open Data
- log, 183
  - stable, 184
  - volatile, 184
- log buffer, C-43
- logical link control layer, D-12
- Lorel, 550, 593
- lossless decomposition, 37, A-4
- lost update, C-12, C-23
- LRM, *see* Local Relational Model
- LSD, 290, 291
- Lucene, 517
  
- machine learning, 438, 473
- MADMAN, C-36
- MAN, *see* metropolitan area network
- map function, 443
- map-only join, 451
- mapping creation, 293, 294
- mapping maintenance, 293, 299
- MapReduce, 439, 442–456, 479, 484, 494, 497, 499, 509, 524–529, 534, 535, 583, 592
- mashup, 567
- master site, 159, 163
- materialization program, 136
- materialized view, 96, 98–100, 102, 121–123, 309, 461, 525, 529, 536
  - maintenance, 96, 123, 278
- Maveric, 300
- maximally-contained query, 309
  
- MDBS, *see* multidatabase system, *see* multidatabase system, *see* multidatabase system, *see* multidatabase system, *see* multidatabase system
- mediated schema, 25, 277, 278, 281, 293
- mediator, 25, 303
- mediator/wrapper architecture, 25
- mediator/wrapper architecture, 303, 304, 316, 325, 328
- medium access control layer, D-12
- Memcached, 506
- MemSQL, 518
- merge-join, B-29
- metadata, 83
- metasearch, 548, 564–566, 592
- METIS, 476–478
- metropolitan area network, D-3, D-4
- MillWheel, 458, 471
- MinCon algorithm, 309, 310
- minterm fragment, 42, 45
- minterm predicate, 39–47, 52, 53
- minterm selectivity, 40
- MISO, 536
- mixed fragmentation, 66
- Mizan, 484, 500
- MonetDB, 36
- MongoDB, 507, 509, 534, 535
- monotonic query, 461
- Mulder, 559, 593
- multi-point network, D-6
- multi-tenant, 29
- multicast, D-6
- multidatabase, 17, 312, 319
- multidatabase query optimization, 311
- multidatabase query processing, 302
- multidatabase system, 16, 25, 277, 279, 302, 329
- multigraph, 472
- multiquery optimization, 466
- multivalued dependency, A-4
- multiversion concurrency control, 200
- mutual consistency, 14
- MVCC, *see* multiversion concurrency control
  
- n-gram, 287
- n-way partitioning, 63
- N1QL, 509
- NAS, *see* network-attached storage
- natural join, A-7, A-11
- negative tuple, 465
- Neo4J, 534, 535
- Neo4j, 513–516
- nested fragmentation, 66





- parallel associative join, 413
- parallel DBMS, 3, 16
- parallel hash join, 355, 357, 413
- parallel merge sort join, 355
- parallel nested loop join, 355
- parallel query optimization, 360
- partial function evaluation, 584
- partial key grouping, 468, 499
- partial redo, C-46
- partial undo, C-46
- partially duplicated, *see* partially replicated
- partially duplicated database, *see* partially replicated database
- partially replicated database, 13, 67
- participant timeout, 218
- partition-centric asynchronous, 491
- partition-centric BSP, 490
- partition-centric graph model, 481
- partitioned database, 13, 67
- partitioning, 33, 62
- path expression, 550
- Paxos, 227, 429
  - basic, 229
- pay-as-you-go integration, 567
- peer-to-peer, 17, 385
  - data management, 385
  - hierarchical structured, 433
  - pure, 388
  - replication, 417
  - structured, 387, 388, 391, 392, 414, 422, 435
  - superpeer, 395, 432, 435
  - unstructured, 387, 388, 391, 400, 408, 432–435
- peer-to-peer computing, 16
- peer-to-peer DBMS, 22
- peer-to-peer system, 4, 19, 23, 24
- peer-to-peer systems, 284
- PeerDB, 400
- Pentaho, 495
- periodic data delivery, 6
- persistent query, 458
- pessimistic concurrency control, 14, C-28
- PGrid, 398
- phantom, 240, C-6, C-13
- phantom read, *see* phantom
- PHJ, *see* parallel hash join
- PHORIZONTAL, 44
- PHT, 395
- physical layer, D-12
- Piazza, 398
- PIER, 413
- PIERjoin, 413
- Pig Latin, 448, 498
- pipeline parallelism, 12
- pipelined symmetric hash join, 463
- PIW, *see* publicly indexable web
- PKG, *see* partial key grouping, *see* partial key grouping
- PlanetP, 408
- planning function, 319
- Platfora, 495
- Platform-as-a-Service, 5
- PNL, *see* parallel nested loop join
- point-to-point network, D-6
- Polybase, 525, 526, 533, 535, 536
- Polystore, 501
- polystore, 502, 519–521, 524, 528, 533–536
  - hybrid, 529, 536
  - loosely-coupled, 521, 536
  - tightly-coupled, 525, 536
- posttest, 114
- PoW, *see* Proof of Work, 430
- Power BI, 495
- PowerLyra, 479
- precise recovery, *see* exactly-once semantics
- precondition constraint, 112
- predefined constraint, 112
- predicate calculus, A-7
- prefix hash tree, 415
- Pregel, 483, 484, 500
- Pregelix, 484, 500
- pretest, 114, 115
- primary key, A-2
- prime attribute, A-2
- projection operator, A-7, A-8
- projection-join dependency, A-4
- Proof of Work, 429
- property graph, 472
- property table, 579, 593
- protocol, D-9
- publicly indexable web, 539, 562
- publish/subscribe system, 460
- punctuation, 463, 498
- push-based system, 5, 6
- QoX, 524, 525, 534, 536
- Qox, 521
- query
  - algebraic, 127
  - distributed, 95
  - execution plan, 131
- query analysis, B-5
- query decomposition, 134
- query execution, 305, 322
- query execution plan, B-14
- query graph, 135, B-6
- query modification, 93
- query normalization, B-3

- query optimization, 127
  - dynamic, 133, 134, 163
  - static, 133
- query processing, 127
- query processor, 127, 345, B-1
- query rewrite, 304
  - using views, 309
- query rewriting, B-9
- query translation, 305, 322
- question answering system, 558
- quorum, 227
- quorum-based voting protocol, 270
  
- R\*, 177, 190
- Raft, 536
- randomized search strategy, B-17
- randomized strategy, B-2
- range partitioning, 512
- range query on P2P systems, 414
- ranking, 543, 547
- RavenDB, 509
- RDD, *see* resilient distributed dataset, *see* resilient distributed dataset
- RDF, 432, *see* Resource Description Framework, 573, 592–594
- RDF triple, 574
- RDF-3X, 579, 593
- reachability query, 473
- read lock, C-29
- read quorum, 270
- read-one/write-all available protocol, 268
  - distributed, 269
- read-one/write-all protocol, 249, 269, 270
- reconstruction, 37
- recovery, 15, 182, C-1
- recovery manager, C-20
- recovery protocol, 207, 221
- Redis, 506
- redo/no-undo, C-47
- reduce function, 443
- reducer, 149, 151
- reduction technique, 138
- referential edge, 290
- referential integrity, 52
- relation, A-1
  - cardinality, A-2
  - degree, A-2
  - instance, A-2
  - schema, *see* schema
- relational algebra, A-6, A-7
- relational calculus, A-6, A-16
- relational database, A-1
- relevant simple predicate, 43
- reliability, 10, 15
  - reliability protocol, C-39
- repartition join, 451
- repetition anomaly, A-3
- replicated database, 13
- replication, 8, 10, 15, 28, 29, 442, 471
  - P2P, 417
- Research Description Framework, 472
- resiliency, 182, C-1
- resilient distributed dataset, 455
- Resource Description Framework
  - graph, 472
  - schema, 569, 573
- resource description framework
  - graph, 574, 575, 577, 582–584
- response time, 132, 155, 156, B-18
- revision tuple, 460
- Riak, 506
- right-deep tree, 362
- ring network, D-5
- rollback, C-4
- rollback recovery, *see* at-least-once semantics
- round-robin partitioning, 468
- routing, D-6
- ROWA, *see* read-one/write-all protocol, *see* read-one/write-all protocol, *see* read-one/write-all protocol, *see* read-one/write-all protocol
- ROWA-A, *see* read-one/write-all available protocol, *see* read-one/write-all available protocol
- run-time support processor, 24
  
- SaaS, *see* Software-as-a-Service, *see* Software-as-a-Service, *see* Software-as-a-Service, *see* Software-as-a-Service
- saga, C-15
- SAN, *see* storage area network, *see* storage-area network
- SAP HANA, 518
- Sawtooth, 431
- Sawzall, 448, 498
- scale-out architecture, 13
- scale-up, 11
- schedule, *see* history
- scheduler, 184
- schema, 2, A-1, A-2
  - heterogeneity, 285
    - adaptation, 299
    - generation, 279
  - heterogeneity, 283
  - integration, 281, 291
  - integration, *n*-ary, 291
  - integration, binary, 291
  - mapping, 281, 283, 293

- matching, 281, 283
  - translation, 279, 337
- schema-based matching, 284–286
- schema-level matching, 288
- schema-on-read, 493, 494
- schema-on-write, 493
- Schism, 74, 85
- SDD-1, 176
- search engine, 539, 542
- Search Space, B-15
- search space, 360, B-14
- Search Strategy, B-17
- search strategy, 133, 364, B-15, B-17
- searchspace, 131
- selection operator, A-7, A-8
- selection predicate, A-8
- selection selectivity, B-20
- selectivity factor, 158, B-20
- semantic data control, 14
- semantic data controller, 23
- semantic heterogeneity, 286
- semantic integrity constraint, 91, 110
- semantic integrity control, 91, 110
- semantic translation, 299
- semantic web, 399, 568, 573, 593
- semiautonomous system, 18
- semijoin, 151
- semijoin operator, A-7, A-13
- semijoin program, 152, 153, 176
- semijoin selectivity, B-22
- SEMINT, 290
- semistructured data, 549
- serial history, C-25, C-26
- serializability, 183, 185, C-13, C-22, C-24
  - conflict-based, C-26
  - one-copy, 248
- serializable history, C-26
- service level agreement, 28
- service oriented architecture, 27
- Sesame, 578, 593
- set difference operator, A-7, A-9
- set-oriented constraint, 117, 118, 120
- SETI@home, 386
- shadow page, C-44
- shadowing, C-44
- sharding, 36
- shared-disk, 348
- shared-memory, 346
- shared-nothing, 350
- ship-whole, 164
- shuffle, 444
- shuffle partitioning, 468
- SI, *see* snapshot isolation, *see* snapshot isolation, *see* snapshot isolation, *see* snapshot isolation
- similarity flooding, 327
- similarity value, 283
- simple predicate, 38–40, 42, 43, 45, 46, 53, B-4
  - completeness, 42
  - minimality, 42
- simple virtual partitioning, 377
- simplification, 115
- simulated annealing, B-2
- single location DBMS, 2
- single-source shortest path, 473
- skip graph, 395
- SkipNet, 395
- snapshot database, 457
- snapshot isolation, 182, 183, 186, 200, 203, 204, 232–234, 237, 273, C-13
  - strong, 250
- Software-as-a-Service, 5
- sort merge join, 356
- soundex code, 288
- source accuracy, 589
- source dependency, 590
- source freshness, 591
- source schema, 281
- Spanner, 233, 519
- Spark, 439, 443, 453–456, 473, 480, 494, 496–498, 509, 525, 529–532, 534, 535, 594
- Sparksee, 516
- SPARQL, 72, 549, 575–579, 582–585, 593
  - distributed, 584
  - endpoint, 584
- SPARQL, 593
- Splice Machine, 518
- Splitting, 53
- SQL, A-17
- SQL++, 509, 523, 524, 535
- stable database, C-20, C-22
- stable log, C-43
- stable storage, C-22
- star network, D-5
- Start system, 559, 593
- state logging, C-43
- static optimization, 137
- static query optimization, B-2, B-27
- steal/force, C-47
- steal/no-steal decision, C-45
- storage area network, 349
- storage-area network, 349
- STREAM, 458, 499
- stream data, 439
- StreaQuel, 461, 498

- strict history, C-12
- strongly connected component query, 475
- structural conflict, 285
- structural constraint, 110
- structural similarity, 289
- structure index, 545
- structure-based matching, 289
- structure-level matching, 284, 285
- StruQL, 554, 593
- subgraph matching, 473
- substitution, B-25
- superkey, A-2
- superpeer system, 388
- superstep, 481
- SVP, *see* simple virtual partitioning
- SW-Store, 593
- switching, D-6
- SWORD, 76, 80, 85
- symmetric hash join, 359
- synonym, 286, 287
- synonyms, 286
- System R, 92, C-44
- System R\*, 190, 194
- SystemML, 448, 498
  
- TA, *see* Threshold Algorithm, 402, *see*  
Threshold Algorithm
- Tableau, 495
- tablet, 512
- Tapestry, 393, 418
- target schema, 281
- TCP/IP, D-8, D-9, D-11
- TelegraphCQ, 458, 499
- Tenzing, 448, 498
- termination protocol, 207, 217
  - non-blocking, 208, 216
- text index, 546
- think-like-a-vertex, 481
- Three Phase Uniform Threshold Algorithm, 403
- three-phase commit, 223
- Threshold Algorithm, 401
- Threshold algorithm, 402
- tight integration, 18
- time travel query, 200
- timeout, 206
- timestamp, 194–196, 199, 200, 203, C-33
  - read, 196, C-34
  - write, 196, C-34
- timestamp order, 186
- timestamp ordering, 195, 199, C-28, C-33
  - basic, 195, C-28
  - conservative, 198–200, C-28
  - multiversion, C-28
- timestamping, 15
  
- TimeStream, 458
- Titan, 516
- top-down database design, 13
- top-k query, 401
- total cost optimization, 132
- total isolation, 19
- total time, 155, B-18, B-19
- TPUT, *see* Three Phase Uniform Threshold  
Algorithm
- transaction, 181, C-1
  - abort, *see* abort
  - atomicity, 183, *see* atomicity
  - base set, 183, C-6
  - batch, C-14
  - closed, C-15
  - closed nested, 235, C-16
  - compensating, C-17
  - consistency, 181, 245, 247, 248, 272, *see*  
consistency
  - conversational, C-14
  - distributed, 10
  - durability, 183, 536, *see* durability
  - failure, *see* transaction failure
  - flat, 235, C-15
  - formal definition, C-7
  - global undo, *see* global undo
  - isolation, 183, *see* isolation
  - long-life, C-14
  - manager, 184
  - nested, 235, C-16
  - online, C-14
  - open nested, 235, C-15, C-17
  - partial undo, *see* partial undo
  - properties, C-9
  - read set, 183, C-6
  - read-before-write, C-14
  - recovery, *see* transaction recovery
  - redo, *see* transaction redo
  - restricted, C-14
  - restricted two-step, C-14
  - short-life, C-14
  - split, 235, C-17
  - two-step, C-14
  - types, C-14
  - undo, *see* transaction undo
  - workflow, *see* workflow
  - write set, 183, C-6
- transaction consistency, C-1
- transaction failure, C-38
- transaction log, *see* log
- transaction recovery, C-9
- transaction redo, C-42, C-43
- transaction undo, C-42, C-43, C-46
- transformation rule, B-10

- transition constraint, 113
- transitive closure, A-7
- transparency, 2, 7, 20
  - concurrency, 10
  - distribution, 9
  - failurfe, 10
  - fragmentation, 9
  - location, 9
  - naming, 9
  - network, 9
  - replication, 10
- transport layer protocol, D-9
- tree query, 152
- Tribeca, 460, 499
- Trinity, 484, 500, 516
- triple, *see* RDF triple
- Tritus, 559, 593
- tuple, A-1
  - variable, A-16
- tuple relational calculus, A-16
- tuple substitution, B-26
- two-phase commit, 10, 208, 239
  - centralized, 211
  - distributed, 212
  - linear, 211
  - nested, 211
  - presumed abort, 214
  - presumed commit, 216
- two-phase locking, 187, C-31
  - centralized, 187
  - distributed, 189
  - primary copy, 257
  - primary site, 187
  - strict, 196, C-31
- type
  - conflict, 285
- UDF, *see* user-defined function
- UMA, 347
- undo/no-redo, C-47
- unfolding, 307
- unicast network, D-6
- uniform memory access, 347
- unilateral abort, 208
- union operator, A-7, A-9
- Uniprot RDF, 573
- unique key constraint, 112
- UnQL, 593
- update anomaly, A-3
- user interface handler, 23
- user processor, 23
- user-defined function, 445
- VBI-tree, 395
- veracity, 438, *see also* data quality, 497
- vertex-centric asynchronous, 487
- vertex-centric BSP, 483, 484
- vertex-centric gather-apply-scatter, 489
- vertex-centric graph model, 481
- Vertica, 36
- Vertical Fragmentation, 142
- view, 91, 92, 306, 307
  - definition, 92, 305
  - management, 91, 92
  - materialization, 92
  - materialized, 92
- view management, 91
- virtual machines, 28
- virtual relation, 92
- volatile database, C-20, C-22
- VoltDB, 518
- voting-based protocol, 227
- W3QL, 554, 593
- WAIT-DIE algorithm, C-36
- wait-for graph, C-35, C-37
- WAL, *see* write-ahead logging
- WAN, *see* wide area network
- WCC, 487, 490
- weakly connected component, 475, 484
- web
  - crawling, 543
  - data fusion, 587
  - data management, 539
  - graph, 540
  - querying, 548
  - search, 542
- web data integration
  - seedata integration
    - web, 566
- web graph, 592
- web indexing, 593
- web portal, 567
- web service, 27
- web table, 567
- WebLog, 554, 593
- WebOQL, 554, 556, 593
- WebQA, 559, 593
- WebSQL, 554, 556, 593
- wide area network, D-3, D-4
- Wide column store, 510
- window, 459–463
  - aggregate, 466
  - count-based, 460, 464
  - fixed, 460
  - join, 463, 465
  - model, 460
  - partitioned, 460

- predicate, 460
- query, 459, 461
- session, 461
- sliding, 460
- time-based, 460, 464
- user-defined, 461
- windowed execution, 458, 461, 463, 464
- wireless broadband network, D-6
- wireless LAN, *see* wireless local area network
- wireless local area network, D-6
- wireless network, D-6
- workflow, 236, C-16, C-18
  - human-oriented, C-18
  - system-oriented, C-18
  - transactional, C-18
- working-set algorithm, C-40
- World Wide Web, 4, 16, 539
- WOUND-WAIT algorithm, C-36
- wrapper, 293, 303
- wrapper schema, 305
- write lock, C-29
- write quorum, 270
- write-ahead logging, C-43
- WWW, *see* World Wide Web
- X-Stream, 492, 500
- xLM, 524
- XML, 432, 506, 525, 528, 539, 569, 592
  - document tree, 571
  - XML schema graph, 571
  - XMLSchema, 571
  - XPath, 573
  - XQuery, 573
- Yago, 573
- YAML, 506
- zigzag tree, 362
- Zookeeper, 496