Surajit Chaudhuri

List of Publications by Year
in descending order

76	6,630	331670	265206
papers	citations	21	42
		h-index	g-index

76	76	76	2566
all docs	docs citations	times ranked	citing authors

#	Article	IF	Citations
1	An overview of data warehousing and OLAP technology. SIGMOD Record, 1997, 26, 65-74.	1.2	1,669
2	Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals. Data Mining and Knowledge Discovery, 1997, 1, 29-53.	3.7	1,218
3	An overview of business intelligence technology. Communications of the ACM, 2011, 54, 88-98.	4.5	473
4	Robust and efficient fuzzy match for online data cleaning. , 2003, , .		301
5	Dynamic sample selection for approximate query processing. , 2003, , .		163
6	Optimized stratified sampling for approximate query processing. ACM Transactions on Database Systems, 2007, 32, 9.	2.8	150
7	On random sampling over joins. , 1999, , .		148
8	Random sampling for histogram construction. , 1998, , .		135
9	AutoAdmin â€œwhat-ifâ€•index analysis utility. , 1998, , .		121
10	Self-tuning histograms. , 1999, , .		113
11	An Online Approach to Physical Design Tuning. , 2007, , .		113
12	Automatic physical database tuning. , 2005, , .		102
13	Database Tuning Advisor for Microsoft SQL Server 2005. , 2004, , 1110-1121.		96
14	Selectivity estimation for range predicates using lightweight models. Proceedings of the VLDB Endowment, 2019, 12, 1044-1057.	3.8	93
15	Approximate Query Processing. , 2017, , .		89
16	Quickr. , 2016, , .		88
17	On random sampling over joins. SIGMOD Record, 1999, 28, 263-274.	1.2	87
18	STHoles. SIGMOD Record, 2001, 30, 211-222.	1.2	80

#	Article	IF	Citations
19	Exploiting statistics on query expressions for optimization. , 2002, , .		69
20	AutoAdmin â€œwhat-ifâ€•index analysis utility. SIGMOD Record, 1998, 27, 367-378.	1.2	65
21	Sample + Seek. , 2016, , .		65
22	A robust, optimization-based approach for approximate answering of aggregate queries. , 2001, , .		58
23	Database tuning advisor for microsoft SQL server 2005. , 2005, , .		58
24	The Claremont report on database research. Communications of the ACM, 2009, 52, 56-65.	4.5	58
25	AI Meets AI. , 2019, , .		57
26	Learning string transformations from examples. Proceedings of the VLDB Endowment, 2009, 2, 514-525.	3.8	55
27	Transformation-based Framework for Record Matching. , 2008, , .		53
28	ClusterJoin. Proceedings of the VLDB Endowment, 2014, 7, 1059-1070.	3.8	52
29	An efficient filter for approximate membership checking. , 2008, , .		51
30	Random sampling for histogram construction. SIGMOD Record, 1998, 27, 436-447.	1.2	45
31	The Seattle Report on Database Research. SIGMOD Record, 2020, 48, 44-53.	1.2	44
32	The Beckman Report on Database Research. SIGMOD Record, 2014, 43, 61-70.	1.2	41
33	Auto-join. Proceedings of the VLDB Endowment, 2017, 10, 1034-1045.	3.8	39
34	Exploiting web search to generate synonyms for entities. , 2009, , .		39
35	Automatically Indexing Millions of Databases in Microsoft Azure SQL Database. , 2019, , .		36
36	Query optimizers. , 2009, , .		31

#	Article	IF	Citations
37	Transform-data-by-example (TDE). Proceedings of the VLDB Endowment, 2018, 11, 1165-1177.	3.8	31
38	On the Equivalence of Recursive and Nonrecursive Datalog Programs. Journal of Computer and System Sciences, 1997, 54, 61-78.	1.2	30
39	Sharing buffer pool memory in multi-tenant relational database-as-a-service. Proceedings of the VLDB Endowment, 2015, 8, 726-737.	3.8	30
40	Mining document collections to facilitate accurate approximate entity matching. Proceedings of the VLDB Endowment, 2009, 2, 395-406.	3.8	26
41	Constrained physical design tuning. Proceedings of the VLDB Endowment, 2008, 1, 4-15.	3.8	24
42	A robust, optimization-based approach for approximate answering of aggregate queries. SIGMOD Record, 2001, 30, 295-306.	1.2	21
43	Fast foreign-key detection in Microsoft SQL server PowerPivot for Excel. Proceedings of the VLDB Endowment, 2014, 7, 1417-1428.	3.8	21
44	Efficiently approximating selectivity functions using low overhead regression models. Proceedings of the VLDB Endowment, 2020, 13, 2215-2228.	3.8	21
45	Data warehousing and OLAP for decision support. SIGMOD Record, 1997, 26, 507-508.	1.2	20
46	Scalable ad-hoc entity extraction from text collections. Proceedings of the VLDB Endowment, 2008, 1, 945-957.	3.8	20
47	Pushing data-induced predicates through joins in big-data clusters. Proceedings of the VLDB Endowment, 2019, 13, 252-265.	3.8	18
48	A pay-as-you-go framework for query execution feedback. Proceedings of the VLDB Endowment, 2008, 1, 1141-1152.	3.8	16
49	Exact cardinality query optimization for optimizer testing. Proceedings of the VLDB Endowment, 2009, 2, 994-1005.	3.8	16
50	On the complexity of equivalence between recursive and nonrecursive Datalog programs. , 1994, , .		15
51	Plan stitch. Proceedings of the VLDB Endowment, 2018, 11, 1123-1136.	3.8	12
52	Physical Design Refinement: The â€œMerge-Reduceâ€•Approach. Lecture Notes in Computer Science, 2006, , 386-404.	1.3	11
53	DSB. Proceedings of the VLDB Endowment, 2021, 14, 3376-3388.	3.8	11
54	Materialized view and index selection tool for Microsoft SQL server 2000. SIGMOD Record, 2001, 30, 608.	1.2	10

#	Article	IF	Citations
55	Finding nonrecursive envelopes for Datalog predicate. , 1993, , .		9
56	Transform-Data-by-Example (TDE). , 2018, , .		9
57	Auto-FuzzyJoin. , 2021, , .		9
58	Experiences with approximating queries in Microsoft's production big-data clusters. Proceedings of the VLDB Endowment, 2019, 12, 2131-2142.	3.8	9
59	Auto-pipeline. Proceedings of the VLDB Endowment, 2021, 14, 2563-2575.	3.8	9
60	Budget-aware Index Tuning with Reinforcement Learning. , 2022, , .		9
61	Leveraging Re-costing for Online Optimization of Parameterized Queries with Guarantees. , 2017, , .		8
62	Bitvector-aware Query Optimization for Decision Support Queries. , 2020, , .		8
63	Efficient creation of statistics over query expressions. , 0, , .		7
64	The Next 5 Years: What Opportunities Should the Database Community Seize to Maximize its Impact?. , 2020, , .		6
65	Constrained physical design tuning. VLDB Journal, 2010, 19, 21-44.	4.1	5
66	Efficient Identification of Approximate Best Configuration of Training in Large Datasets. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33, 3862-3869.	4.9	5
67	New frontiers in business intelligence. Proceedings of the VLDB Endowment, 2011, 4, 1502-1503.	3.8	5
68	Can Datalog Be Approximated?. Journal of Computer and System Sciences, 1997, 55, 355-369.	1.2	4
69	Extracting predicates from mining models for efficient query evaluation. ACM Transactions on Database Systems, 2004, 29, 508-544.	2.8	4
70	Customizable and scalable fuzzy join for big data. Proceedings of the VLDB Endowment, 2019, 12, 2106-2117.	3.8	4
71	Multi-Tenant Cloud Data Services: State-of-the-Art, Challenges and Opportunities. , 2022, , .		4
72	Data Management Technology for Decision Support Systems. Advances in Computers, 2004, 62, 293-326.	1.6	2

#	Article	IF	Citations
73	Leveraging query logs and machine learning for parametric query optimization. Proceedings of the VLDB Endowment, 2021, 15, 401-413.	3.8	2
74	ISUM: Efficiently Compressing Large and Complex Workloads for Scalable Index Tuning. , 2022, , .		2
75	Data warehousing and OLAP for decision support. Lecture Notes in Computer Science, 1997, , 33-34.	1.3	1
76	Query portals. , 2010, , .		1