Quan Chen

List of Publications by Year
in descending order

55 papers	**1,104** citations	933447 **10** h-index	839539 **18** g-index
55 all docs	**55** docs citations	**55** times ranked	**566** citing authors

#	Article	IF	Citations
1	Blockchain-Aided Edge Computing Market: Smart Contract and Consensus Mechanisms. IEEE Transactions on Mobile Computing, 2023, 22, 3193-3208.	5.8	9
2	Toward QoS-Awareness and Improved Utilization of Spatial Multitasking GPUs. IEEE Transactions on Computers, 2022, 71, 866-879.	3.4	14
3	Special Issue on Programming Models and Applications for Multicores and Manycores 2020. Concurrency Computation Practice and Experience, 2022, 34, e6457.	2.2	0
4	Special issue on programming models and applications for multicores and manycores 2019â€"2020. Concurrency Computation Practice and Experience, 2022, 34, e6677.	2.2	0
5	Adaptive Resource Efficient Microservice Deployment in Cloud-Edge Continuum. IEEE Transactions on Parallel and Distributed Systems, 2022, 33, 1825-1840.	5.6	40
6	The Serverless Computing Survey: A Technical Primer for Design Architecture. ACM Computing Surveys, 2022, 54, 1-34.	23.0	34
7	Online Thread Auto-Tuning for Performance Improvement and Resource Saving. IEEE Transactions on Parallel and Distributed Systems, 2022, , 1-1.	5.6	2
8	Tacker: Tensor-CUDA Core Kernel Fusion for Improving the GPU Utilization while Ensuring QoS. , 2022, , .		8
9	Efficient and Secure Deep Learning Inference in Trusted Processor Enabled Edge Clouds. IEEE Transactions on Parallel and Distributed Systems, 2022, 33, 4311-4325.	5.6	3
10	CSC: Collaborative System Configuration for I/O-Intensive Applications in Multi-Tenant Clouds. , 2022, , .		1
11	Exploring Efficient Microservice Level Parallelism. , 2022, , .		3
12	Adaptive Preference-Aware Co-Location for Improving Resource Utilization of Power Constrained Datacenters. IEEE Transactions on Parallel and Distributed Systems, 2021, 32, 441-456.	5.6	7
13	AlphaR: Learning-Powered Resource Management for Irregular, Dynamic Microservice Graph. , 2021, , .		15
14	E^{2}bird: _E_nhanced _E_lastic _B_atch for _I_mproving _R_esponsiveness and Throughput of _D_eep Learning Services. IEEE Transactions on Parallel and Distributed Systems, 2021, 32, 1307-1321.	5.6	14
15	Lasagna. , 2021, , .		7
16	Enable simultaneous DNN services based on deterministic operator overlap and precise latency prediction. , 2021, , .		25
17	Exploiting Intra-SM Parallelism in GPUs via Persistent and Elastic Blocks. , 2021, , .		6
18	Amoeba: QoS-Awareness and Reduced Resource Usage of Microservices with Serverless Computing. , 2020, , .		16

#	Article	IF	Citations
19	Sturgeon: Preference-aware Co-location for Improving Utilization of Power Constrained Computers. , 2020, , .		5
20	Probabilistic robust regression with adaptive weights â€" a case study on face recognition. Frontiers of Computer Science, 2020, 14, 1.	2.4	2
21	URSA: Precise Capacity Planning and Fair Scheduling based on Low-level Statistics for Public Clouds. , 2020, , .		11
22	Task Offloading in Trusted Execution Environment empowered Edge Computing. , 2020, , .		3
23	CODA: Improving Resource Utilization by Slimming and Co-locating DNN and CPU Jobs. , 2020, , .		8
24	DR Refresh: Releasing DRAM Potential by Enabling Read Accesses Under Refresh. IEEE Transactions on Computers, 2019, 68, 1584-1596.	3.4	3
25	Avalon. , 2019, , .		16
26	Laius. , 2019, , .		18
27	Themis: Predicting and Reining in Application-Level Slowdown on Spatial Multitasking GPUs. , 2019, , .		20
28	When Power Oversubscription Meets Traffic Flood Attack. , 2019, , .		2
29	Mobile Sampling Strategy for Environment Information Reconstruction from View of Cloud. , 2019, , .		0
30	Optimizing the Aggregated Throughput of GPUs in Public Clouds Based on Adaptive Kernel Reordering. , 2019, , .		0
31	Characterizing and Balancing the Workloads of Semi-Containerized Clouds. , 2019, , .		3
32	Ebird: Elastic Batch for Improving Responsiveness and Throughput of Deep Learning Services. , 2019, , .		18
33	Contention and Locality-Aware Work-Stealing for Iterative Applications in Multi-Socket Computers. IEEE Transactions on Computers, 2018, 67, 784-798.	3.4	8
34	Power Grab in Aggressively Provisioned Data Centers: What is the Risk and What Can Be Done About It. , 2018, , .		11
35	DR DRAM: Accelerating Memory-Read-Intensive Applications. , 2018, , .		3
36	In-growth test for monolithic 3D integrated SRAM. , 2018, , .		3

#	Article	IF	Citations
37	KSM: Online Application-Level Performance Slowdown Prediction for Spatial Multitasking GPGPU. IEEE Computer Architecture Letters, 2018, 17, 187-191.	1.5	12
38	Bandwidth and Locality Aware Task-stealing for Manycore Architectures with Bandwidth-Asymmetric Memory. Transactions on Architecture and Code Optimization, 2018, 15, 1-26.	2.0	8
39	Prophet. Operating Systems Review (ACM), 2017, 51, 17-32.	1.9	12
40	PowerChief. , 2017, , .		29
41	Prophet. , 2017, , .		76
42	Preemption-Aware Kernel Scheduling for GPUs. , 2017, , .		4
43	Prophet. Computer Architecture News, 2017, 45, 17-32.	2.5	6
44	Baymax. , 2016, , .		76
45	Baymax. ACM SIGPLAN Notices, 2016, 51, 681-696.	0.2	27
46	Baymax. Computer Architecture News, 2016, 44, 681-696.	2.5	11
47	DjiNN and Tonic. , 2015, , .		121
48	EEWA: Energy-Efficient Workload-Aware Task Scheduling in Multi-core Architectures. , 2014, , .		15
49	LAWS. , 2014, , .		31
50	Adaptive workload-aware task scheduling for single-ISA asymmetric multicore architectures. Transactions on Architecture and Code Optimization, 2014, 11, 1-25.	2.0	76
51	HAT: history-based auto-tuning MapReduce in heterogeneous environments. Journal of Supercomputing, 2013, 64, 1038-1054.	3.6	24
52	CATS. , 2012, , .		43
53	WATS: Workload-Aware Task Scheduling in Asymmetric Multi-core Architectures. , 2012, , .		31
54	CAB: Cache Aware Bi-tier Task-Stealing in Multi-socket Multi-core Architecture. , 2011, , .		24

#	Article	IF	Citations
55	SAMR: A Self-adaptive MapReduce Scheduling Algorithm in Heterogeneous Environment. , 2010, , .		140