Kesheng Wu

List of Publications by Year
in descending order

196	4,697	279701	155592
papers	citations	23	55
		h-index	g-index

205	205	205	2877
all docs	docs citations	times ranked	citing authors

#	ARTICLE	IF	CITATIONS
1	Improving I/O Performance for Exascale Applications Through Online Data Layout Reorganization. IEEE Transactions on Parallel and Distributed Systems, 2022, 33, 878-890.	4.0	10
2	Improving nonnegative matrix factorization with advanced graph regularization. Information Sciences, 2022, 597, 125-143.	4.0	6
3	Organizing Large Data Sets for Efficient Analyses on HPC Systems. Journal of Physics: Conference Series, 2022, 2224, 012042.	0.3	2
4	The Imperial Valley Dark Fiber Project: Toward Seismic Studies Using DAS and Telecom Infrastructure for Geothermal Applications. Seismological Research Letters, 2022, 93, 2906-2919.	0.8	9
5	Access Trends of In-network Cache for Scientific Data. , 2022, , .		1
6	Predicting Slow Network Transfers in Scientific Computing. , 2022, , .		1
7	Real-time and post-hoc compression for data from Distributed Acoustic Sensing. Computers and Geosciences, 2022, 166, 105181.	2.0	4
8	FasTensor Programming Model. SpringerBriefs in Computer Science, 2021, , 9-22.	0.2	0
9	An empirical study of I/O separation for burst buffers in HPC systems. Journal of Parallel and Distributed Computing, 2021, 148, 96-108.	2.7	11
10	Adaptive Stochastic Gradient Descent for Deep Learning on Heterogeneous CPU+GPU Architectures. , 2021, , .		2
11	FasTensor in Real Scientific Applications. SpringerBriefs in Computer Science, 2021, , 73-84.	0.2	0
12	Network traffic performance analysis from passive measurements using gradient boosting machine learning. International Journal of Big Data Intelligence, 2021, 8, 13.	0.4	0
13	Network traffic performance analysis from passive measurements using gradient boosting machine learning. International Journal of Big Data Intelligence, 2021, 8, 13.	0.4	1
14	Asynchronous I/O Strategy for Large-Scale Deep Learning Applications. , 2021, , .		2
15	Performance of the Gold Standard and Machine Learning in Predicting Vehicle Transactions. , 2021, , .		1
16	An In-Depth I/O Pattern Analysis in HPC Systems. , 2021, , .		2
17	ADIOS 2: The Adaptable Input Output System. A framework for high-performance data management. SoftwareX, 2020, 12, 100561.	1.2	102
18	DASSA: Parallel DAS Data Storage and Analysis for Subsurface Event Detection. , 2020, , .		5

#	Article	IF	Citations
19	Predicting Resource Requirement in Intermediate Palomar Transient Factory Workflow. , 2020, , .		0
20	Clustering Life Course to Understand the Heterogeneous Effects of Life Events, Gender, and Generation on Habitual Travel Modes. IEEE Access, 2020, 8, 190964-190980.	2.6	5
21	Deep Learning for Surface Wave Identification in Distributed Acoustic Sensing Data. , 2020, , .		11
22	Feature Selection Improves Tree-based Classification for Wireless Intrusion Detection. , 2020, , .		13
23	Analyzing Scientific Data Sharing Patterns for In-network Data Caching. , 2020, , .		1
24	Transfer Learning Approach for Botnet Detection Based on Recurrent Variational Autoencoder. , 2020, , .		4
25	Access Patterns to Disk Cache for Large Scientific Archive. , 2020, , .		1
26	Towards HPC I/O Performance Prediction through Large-scale Log Analysis. , 2020, , .		17
27	GPU-based Classification for Wireless Intrusion Detection. , 2020, , .		3
28	HPC Workload Characterization Using Feature Selection and Clustering. , 2020, , .		8
29	Enhancing IoT Anomaly Detection Performance for Federated Learning. , 2020, , .		6
30	Botnet Detection Using Recurrent Variational Autoencoder. , 2020, , .		6
31	Effective Missing Value Imputation Methods for Building Monitoring Data. , 2020, , .		7
32	DCA-IO: A Dynamic I/O Control Scheme for Parallel and Distributed File Systems. , 2019, , .		10
33	Co-optimizing Latency and Energy for IoT services using HMP servers in Fog Clusters. , 2019, , .		3
34	Automatic Detection of Network Traffic Anomalies and Changes. , 2019, , .		6
35	SLOPE: Structural Locality-Aware Programming Model for Composing Array Data Analysis. Lecture Notes in Computer Science, 2019, , 61-80.	1.0	2
36	Multidimensional Compression with Pattern Matching. , 2019, , .		0

#	Article	IF	Citations
37	Enabling SQL-Query Processing for Ethereum-based Blockchain Systems. , 2019, , .		10
38	Understanding Parallel I/O Performance Trends Under Various HPC Configurations. , 2019, , .		4
39	Terabyte-scale Particle Data Analysis. , 2019, , .		4
40	Performance Prediction for Data Transfers in LCLS Workflow. , 2019, , .		3
41	Similarity-based Compression with Multidimensional Pattern Matching. , 2019, , .		0
42	Parallel membership queries on very large scientific data sets using bitmap indexes. Concurrency Computation Practice and Experience, 2019, 31, e5157.	1.4	15
43	Special issue on scientific and statistical data management. Distributed and Parallel Databases, 2019, 37, 1-3.	1.0	2
44	Evaluating the Effects of Missing Values and Mixed Data Types on Social Sequence Clustering Using t-SNE Visualization. Journal of Data and Information Quality, 2019, 11, 1-22.	1.5	7
45	An Assessment of the Prediction Quality of VPIN. , 2019, , .		0
46	Federated Wireless Network Intrusion Detection. , 2019, , .		27
47	Spatiotemporal Real-Time Anomaly Detection for Supercomputing Systems. , 2019, , .		1
48	Machine Learning for Prediction of Mid to Long Term Habitual Transportation Mode Use. , 2019, , .		6
49	Understanding Data Similarity in Large-Scale Scientific Datasets. , 2019, , .		0
50	Analysis and Prediction of Data Transfer Throughput for Data-Intensive Workloads. , 2019, , .		2
51	A Reinforcement Learning Based Network Scheduler for Deadline-Driven Data Transfers. , 2019, , .		10
52	Incremental nonnegative matrix factorization based on correlation and graph regularization for matrix completion. International Journal of Machine Learning and Cybernetics, 2019, 10, 1259-1268.	2.3	7
53	Extracting Signals from High-Frequency Trading with Digital Signal Processing Tools. The Journal of Financial Data Science, 2019, 1, 124-138.	0.9	1
54	HDF5 as a vehicle for in transit data movement. , 2019, , .		2

#	Article	IF	Citations
55	Consensus Ensemble System for Traffic Flow Prediction. IEEE Transactions on Intelligent Transportation Systems, 2018, 19, 3903-3914.	4.7	32
56	ARCHIE: Data Analysis Acceleration with Array Caching in Hierarchical Storage. , 2018, , .		8
57	Predicting Network Traffic Using TCP Anomalies. , 2018, , .		1
58	Automated Parallel Data Processing Engine with Application to Large-Scale Feature Extraction. , 2018, , .		3
59	Automatic identification and classification of Palomar Transient Factory astrophysical objects in GLADE. International Journal of Computational Science and Engineering, 2018, 16, 337.	0.4	0
60	Efficient Online Hyperparameter Learning for Traffic Flow Prediction. , 2018, , .		1
61	Detecting Anomalies in the LCLS Workflow. , 2018, , .		0
62	ArrayBridge: Interweaving Declarative Array Processing in SciDB with Imperative HDF5-Based Programs. , 2018, , .		5
63	Auto-Tuned Publisher in a Pub/Sub System: Design and Performance Evaluation. , 2018, , .		2
64	Distributed caching for processing raw arrays. , 2018, , .		4
65	Modeling Data Transfers: Change Point and Anomaly Detection. , 2018, , .		3
66	Predicting baseline for analysis of electricity pricing. International Journal of Big Data Intelligence, 2018, 5, 3.	0.4	2
67	Identifying Anomalous File Transfer Events in LCLS Workflow. , 2018, , .		4
68	Towards Autonomic Science Infrastructure. , 2018, , .		8
69	Querying Large Scientific Data Sets with Adaptable IO System ADIOS. Lecture Notes in Computer Science, 2018, , 51-69.	1.0	14
70	Predicting baseline for analysis of electricity pricing. International Journal of Big Data Intelligence, 2018, 5, 3.	0.4	0
71	Automatic identification and classification of Palomar Transient Factory astrophysical objects in GLADE. International Journal of Computational Science and Engineering, 2018, 16, 337.	0.4	0
72	Expanding Statistical Similarity Based Data Reduction to Capture Diverse Patterns. , 2017, , .		2

#	Article	IF	Citations
73	Incremental View Maintenance over Array Data. , 2017, , .		12
74	Statistical data reduction for streaming data. , 2017, , .		6
75	Parallel Variable Selection for Effective Performance Prediction. , 2017, , .		5
76	Feature Engineering and Classification Models for Partial Discharge Events in Power Transformers. , 2017, , .		0
77	Optimizing the query performance of block index through data analysis and I/O modeling. , 2017, , .		7
78	UoI-NMF Cluster: A Robust Nonnegative Matrix Factorization Algorithm for Improved Parts-Based Decomposition and Reconstruction of Noisy Data. , 2017, , .		6
79	Improving Statistical Similarity Based Data Reduction for Non-Stationary Data. , 2017, , .		3
80	ArrayUDF. , 2017, , .		28
81	Accurate signal timing from high frequency streaming data. , 2017, , .		1
82	Convolutional Filtering for Accurate Signal Timing from Noisy Streaming Data. , 2017, , .		1
83	Data quality challenges with missing values and mixed types in joint sequence analysis. , 2017, , .		6
84	Apply Block Index Technique to Scientific Data Analysis and I/O Systems. , 2017, , .		9
85	Similarity Join over Array Data. , 2016, , .		15
86	Performance Analysis, Design Considerations, and Applications of Extreme-Scale In Situ Infrastructures. , 2016, , .		51
87	In Situ Storage Layout Optimization for AMR Spatio-temporal Read Accesses. , 2016, , .		2
88	Exploring memory hierarchy and network topology for runtime AMR data sharing across scientific applications. , 2016, , .		2
89	AMRZone: A Runtime AMR Data Sharing Framework for Scientific Applications. , 2016, , .		1
90	Data Elevator: Low-Contention Data Movement in Hierarchical Storage System. , 2016, , .		40

#	Article	IF	Citations
91	Towards Real-Time Detection and Tracking of Spatio-Temporal Features: Blob-Filaments in Fusion Plasma. IEEE Transactions on Big Data, 2016, 2, 262-275.	4.4	13
92	Performance Analysis Tool for HPC and Big Data Applications on Scientific Clusters. , 2016, , 139-161.		2
93	Indexing Blocks to Reduce Space and Time Requirements for Searching Large Data Files. , 2016, , .		7
94	Machine learning based job status prediction in scientific clusters. , 2016, , .		10
95	Novel Data Reduction Based on Statistical Similarity. , 2016, , .		13
96	Usage Pattern-Driven Dynamic Data Layout Reorganization. , 2016, , .		11
97	SDS-Sort. , 2016, , .		7
98	Visualization and Analysis for Near-Real-Time Decision Making in Distributed Workflows. , 2016, , .		6
99	Solving the Optimal Trading Trajectory Problem Using a Quantum Annealer. IEEE Journal on Selected Topics in Signal Processing, 2016, 10, 1053-1060.	7.3	123
100	A science data gateway for environmental management. Concurrency Computation Practice and Experience, 2016, 28, 1994-2004.	1.4	8
101	MPO: A System to Document and Analyze Distributed Heterogeneous Workflows. Lecture Notes in Computer Science, 2016, , 166-170.	1.0	1
102	Solving the optimal trading trajectory problem using a quantum annealer. , 2015, , .		8
103	Extracting Baseline Electricity Usage Using Gradient Tree Boosting. , 2015, , .		5
104	PATHA: Performance Analysis Tool for HPC Applications. , 2015, , .		4
105	Heavy-tailed distribution of parallel I/O system response time. , 2015, , .		3
106	In-memory Query System for Scientific Dataseis. , 2015, , .		6
107	Security for the scientific data services framework. , 2015, , .		0
108	Spatially clustered join on heterogeneous scientific data sets. , 2015, , .		6

#	ARTICLE	IF	CITATIONS
109	Parallel In Situ Detection of Connected Components in Adaptive Mesh Refinement Data. , 2015, , .		5
110	Simplifying index file structure to improve I/O performance of parallel indexing. , 2014, , .		4
111	Model-Driven Data Layout Selection for Improving Read Performance. , 2014, , .		8
112	Parallel query evaluation as a Scientific Data Service. , 2014, , .		11
113	Parallel data analysis directly on scientific file formats. , 2014, , .		70
114	Hello ADIOS: the challenges and lessons of developing leadership class I/O frameworks. Concurrency Computation Practice and Experience, 2014, 26, 1453-1473.	1.4	170
115	Exploring Irregular Time Series through Non-Uniform Fast Fourier Transform. , 2014, , .		2
116	Implementing the Palomar Transient Factory Real-Time Detection Pipeline in GLADE: Results and Observations. Lecture Notes in Computer Science, 2014, , 53-66.	1.0	6
117	Accelerating gene context analysis using bitmaps. , 2013, , .		11
118	Expediting scientific data analysis with reorganization of data. , 2013, , .		16
119	Why high performance visual data analytics is both relevant and difficult. Proceedings of SPIE, 2013, , .	0.8	1
120	SDS. , 2013, , .		21
121	Fast Change Point Detection for electricity market analysis. , 2013, , .		2
122	A big data approach to analyzing market volatility. Algorithmic Finance, 2013, 2, 241-267.	0.3	19
123	Optimizing fastquery performance on lustre file system. , 2013, , .		5
124	Parallel I/O, analysis, and visualization of a trillion particle simulation. , 2012, , .		47
125	Efficient Attribute-Based Data Access in Astronomy Analysis. , 2012, , .		3
126	Federal Market Information Technology in the Postâ€"FlashCrash Era: <i>Roles for Supercomputing</i>. Journal of Trading, 2012, 7, 9-25.	0.2	16

#	Article	IF	Citations
127	TECA: A Parallel Toolkit for Extreme Climate Analysis. Procedia Computer Science, 2012, 9, 866-876.	1.2	37
128	A Communication-Avoiding Thick-Restart Lanczos Method on a Distributed-Memory System. Lecture Notes in Computer Science, 2012, , 345-354.	1.0	3
129	Minimizing Index Size by Reordering Rows and Columns. Lecture Notes in Computer Science, 2012, , 467-484.	1.0	6
130	Parallel in situ indexing for data-intensive computing. , 2011, , .		34
131	BMF: Bitmapped Mass Fingerprinting for Fast Protein Identification. , 2011, , .		2
132	FastQuery: A Parallel Indexing System for Scientific Data. , 2011, , .		39
133	Finding regions of interest on toroidal meshes. Computational Science & Discovery, 2011, 4, 015003.	1.5	8
134	Parallel index and query for large scale data analysis. , 2011, , .		58
135	Federal market information technology in the post flash crash era. , 2011, , .		10
136	Detecting atmospheric rivers in large climate datasets. , 2011, , .		15
137	Scientific data services. , 2011, , .		7
138	Massive-Scale RDF Processing Using Compressed Bitmap Indexes. Lecture Notes in Computer Science, 2011, , 470-479.	1.0	6
139	FastQuery: A General Indexing and Querying System for Scientific Data. Lecture Notes in Computer Science, 2011, , 573-574.	1.0	10
140	Coupling visualization and data analysis for knowledge discovery from multi-dimensional scientific data. Procedia Computer Science, 2010, 1, 1757-1764.	1.2	8
141	Multi-level bitmap indexes for flash memory storage. , 2010, , .		9
142	Adaptive Projection Subspace Dimension for the Thick-Restart Lanczos Method. ACM Transactions on Mathematical Software, 2010, 37, 1-18.	1.6	30
143	Analyses of multi-level and multi-component compressed bitmap indexes. ACM Transactions on Database Systems, 2010, 35, 1-52.	1.5	52
144	Finding Tropical Cyclones on a Cloud Computing Cluster: Using Parallel Virtualization for Large-Scale Climate Simulation Analysis. , 2010, , .		4

#	Article	IF	Citations
145	FastBit: interactively searching massive data. Journal of Physics: Conference Series, 2009, 180, 012053.	0.3	92
146	Efficient joins with compressed bitmap indexes. , 2009, , .		4
147	Automatic beam path analysis of laser wakefield particle acceleration data. Computational Science & Discovery, 2009, 2, 015005.	1.5	9
148	Optimizing two-pass connected-component labeling algorithms. Pattern Analysis and Applications, 2009, 12, 117-135.	3.1	280
149	Fast connected-component labeling. Pattern Recognition, 2009, 42, 1977-1987.	5.1	344
150	Efficiently Extracting Operational Profiles from Execution Logs Using Suffix Arrays. , 2009, , .		45
151	Occam's razor and petascale visual data analysis. Journal of Physics: Conference Series, 2009, 180, 012084.	0.3	2
152	Finding Regions of Interest in Large Scientific Datasets. Lecture Notes in Computer Science, 2009, , 130-147.	1.0	4
153	Data Parallel Bin-Based Indexing for Answering Queries on Multi-core Architectures. Lecture Notes in Computer Science, 2009, , 110-129.	1.0	5
154	Accelerating Queries on Very Large Datasets. Chapman & Hall/CRC Computational Science, 2009, , .	0.5	5
155	High performance multivariate visual data exploration for extremely large data. , 2008, , .		18
156	Adaptive Bitmap Indexes for Space-Constrained Systems. , 2008, , .		5
157	Efficient Operational Profiling of Systems Using Suffix Arrays on Execution Logs. , 2008, , .		1
158	Breaking the Curse of Cardinality on Bitmap Indexes. Lecture Notes in Computer Science, 2008, , 348-365.	1.0	43
159	Bitmap Index Design Choices and Their Performance Implications. , 2007, , .		19
160	Enabling Real-Time Querying of Live and Historical Stream Data. International Conference on Scientific and Statistical Database Management: [proceedings] International Conference on Scientific and Statistical Database Management, 2007, , .	0.0	48
161	Bitmap Indices for Data Warehouses. , 2007, , 157-178.		33
162	Optimizing bitmap indices with efficient compression. ACM Transactions on Database Systems, 2006, 31, 1-38.	1.5	264

#	ARTICLE	IF	CITATIONS
163	Bitmap indices for fast end-user physics analysis in ROOT. Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2006, 559, 99-102.	0.7	8
164	Imaging and visual analysis---Detecting distributed scans using high-performance query-driven visualization. , 2006, , .		17
165	Accelerating Network Traffic Analytics Using Query-Driven Visualization. , 2006, , .		23
166	Detecting Distributed Scans Using High-Performance Query-Driven Visualization. , 2006, , .		20
167	Optimizing candidate check costs for bitmap indices. , 2005, , .		33
168	Optimizing connected component labeling algorithms. , 2005, , .		107
169	FastBit: an efficient indexing technology for accelerating data-intensive science. Journal of Physics: Conference Series, 2005, 16, 556-560.	0.3	55
170	Optimizing I/O Costs of Multi-dimensional Queries Using Bitmap Indices. Lecture Notes in Computer Science, 2005, , 220-229.	1.0	12
171	Evaluation Strategies for Bitmap Indices with Binning. Lecture Notes in Computer Science, 2004, , 120-129.	1.0	35
172	On the Performance of Bitmap Indices for High Cardinality Attributes. , 2004, , 24-35.		78
173	On the Performance of Bitmap Indices for High Cardinality Attributes1. , 2004, , 24-35.		29
174	A Block Orthogonalization Procedure with Constant Synchronization Requirements. SIAM Journal of Scientific Computing, 2002, 23, 2165-2182.	1.3	75
175	Strategies for processing ad hoc queries on large data warehouses. , 2002, , .		22
176	A performance comparison of bitmap indexes. , 2001, , .		40
177	Thick-Restart Lanczos Method for Large Symmetric Eigenvalue Problems. SIAM Journal on Matrix Analysis and Applications, 2000, 22, 602-616.	0.7	229
178	A parallel Lanczos method for symmetric generalized eigenvalue problems. Computing and Visualization in Science, 1999, 2, 37-46.	1.2	16
179	Thick-Restart Lanczos Method for Electronic Structure Calculations. Journal of Computational Physics, 1999, 154, 156-173.	1.9	48
180	Dynamic Thick Restarting of the Davidson, and the Implicitly Restarted Arnoldi Methods. SIAM Journal of Scientific Computing, 1998, 19, 227-245.	1.3	92

#	ARTICLE	IF	CITATIONS
181	Thick-restart Lanczos method for symmetric eigenvalue problems. Lecture Notes in Computer Science, 1998, , 43-55.	1.0	3
182	DQGMRES: a Direct Quasi-minimal Residual Algorithm Based on Incomplete Orthogonalization. Numerical Linear Algebra With Applications, 1996, 3, 329-343.	0.9	25
183	Molecular dynamics with quantum forces: Vibrational spectra of localized systems. Physical Review B, 1996, 53, 12071-12079.	1.1	32
184	Design of an iterative solution module for a parallel sparse matrix library (P_SPARSLIB). Applied Numerical Mathematics, 1995, 19, 343-357.	1.2	8
185	Abinitiomolecular-dynamics simulations of Si clusters using the higher-order finite-difference-pseudopotential method. Physical Review B, 1994, 50, 12234-12237.	1.1	84
186	Higher-order finite-difference pseudopotential method: An application to diatomic molecules. Physical Review B, 1994, 50, 11355-11364.	1.1	506
187	Statistical tests for deterministic effects in broad band time series. Physica D: Nonlinear Phenomena, 1993, 69, 172-188.	1.3	21
188	Compressing bitmap indexes for faster search operations. , 0, , .		50
189	Using bitmap index for interactive exploration of large datasets. , 0, , .		29
190	Network Traffic Analysis With Query Driven Visualization SC 2005 HPC Analytics Results. , 0, , .		9
191	Query-Driven Visualization of Large Data Sets. , 0, , .		22
192	HDF5-FastQuery: Accelerating Complex Queries on HDF Datasets using Fast Bitmap Indices. , 0, , .		34
193	Minimizing I/O Costs of Multi-Dimensional Queries with Bitmap Indices. , 0, , .		3
194	Query-Driven Visualization of Large Data Sets. , 0, , .		29
195	Federal Market Information Technology in the Post Flash Crash Era: Roles for Supercomputing. SSRN Electronic Journal, 0, , .	0.4	17
196	What Makes You Hold on to That Old Car? Joint Insights From Machine Learning and Multinomial Logit on Vehicle-Level Transaction Decisions. Frontiers in Future Transportation, 0, 3, .	1.3	2