

Douglas L Thain

List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/104485/publications.pdf>

Version: 2024-02-01

122
papers

3,131
citations

623734

14
h-index

330143

37
g-index

124
all docs

124
docs citations

124
times ranked

2793
citing authors

#	ARTICLE	IF	CITATIONS
1	Distributed computing in practice: the Condor experience. <i>Concurrency Computation Practice and Experience</i> , 2005, 17, 323-356.	2.2	1,345
2	Condor and the Grid. , 0, , 299-335.		221
3	Qthreads: An API for programming with millions of lightweight threads. <i>Parallel and Distributed Processing Symposium (IPDPS), Proceedings of the International Conference on</i> , 2008, , .	1.0	130
4	A Comparison and Critique of Eucalyptus, OpenNebula and Nimbus. , 2010, , .		117
5	Makeflow. , 2012, , .		107
6	All-pairs: An abstraction for data-intensive cloud computing. <i>Parallel and Distributed Processing Symposium (IPDPS), Proceedings of the International Conference on</i> , 2008, , .	1.0	75
7	Reproducibility in Scientific Computing. <i>ACM Computing Surveys</i> , 2019, 51, 1-36.	23.0	61
8	Integrating Containers into Workflows. , 2015, , .		58
9	All-Pairs: An Abstraction for Data-Intensive Computing on Campus Grids. <i>IEEE Transactions on Parallel and Distributed Systems</i> , 2010, 21, 33-46.	5.6	52
10	Chirp: a practical global filesystem for cluster and Grid computing. <i>Journal of Grid Computing</i> , 2009, 7, 51-72.	3.9	44
11	The quest for scalable support of data-intensive workloads in distributed systems. , 2009, , .		40
12	Scaling up Classifiers to Cloud Computers. , 2008, , .		36
13	Middleware support for many-task computing. <i>Cluster Computing</i> , 2010, 13, 291-314.	5.0	35
14	Harnessing parallelism in multicore clusters with the All-Pairs, Wavefront, and Makeflow abstractions. <i>Cluster Computing</i> , 2010, 13, 243-256.	5.0	35
15	Converting a High Performance Application to an Elastic Cloud Application. , 2011, , .		35
16	AWE-WQ: Fast-Forwarding Molecular Dynamics Using the Accelerated Weighted Ensemble. <i>Journal of Chemical Information and Modeling</i> , 2014, 54, 3033-3043.	5.4	29
17	Toward fine-grained online task characteristics estimation in scientific workflows. , 2013, , .		28
18	Wharf. , 2018, , .		26

#	ARTICLE	IF	CITATIONS
19	Combining Static and Dynamic Storage Management for Data Intensive Scientific Workflows. IEEE Transactions on Parallel and Distributed Systems, 2018, 29, 338-350.	5.6	21
20	DistIA: a cost-effective dynamic impact analysis for distributed programs. , 2016, , .		21
21	How to measure a large open-source distributed system. Concurrency Computation Practice and Experience, 2006, 18, 1989-2019.	2.2	20
22	Deploying High Throughput Scientific Workflows on Container Schedulers with Makeflow and Mesos. , 2017, , .		20
23	Practical Resource Monitoring for Robust High Throughput Computing. , 2015, , .		19
24	The Evolution of Global Scale Filesystems for Scientific Software Distribution. Computing in Science and Engineering, 2015, 17, 61-71.	1.2	19
25	Efficient access to many small files in a filesystem for grid computing. , 2007, , .		17
26	An invariant framework for conducting reproducible computational science. Journal of Computational Science, 2015, 9, 137-142.	2.9	17
27	GPU acceleration of Eulerianâ€“Lagrangian particle-laden turbulent flow simulations. International Journal of Multiphase Flow, 2018, 99, 437-445.	3.4	17
28	Troubleshooting thousands of jobs on production grids using data mining techniques. , 2008, , .		15
29	A Framework for Scalable Genome Assembly on Clusters, Clouds, and Grids. IEEE Transactions on Parallel and Distributed Systems, 2012, 23, 2189-2197.	5.6	14
30	DAGViz. , 2015, , .		14
31	A Job Sizing Strategy for High-Throughput Scientific Workflows. IEEE Transactions on Parallel and Distributed Systems, 2018, 29, 240-253.	5.6	14
32	Scheduling Grid workloads on multicore clusters to minimize energy and maximize performance. , 2009, , .		13
33	Experience with BXGrid: a data repository and computing grid for biometrics research. Cluster Computing, 2009, 12, 373-386.	5.0	13
34	Resource Management for Elastic Cloud Workflows. , 2012, , .		13
35	D <sc>ia</sc> P <sc>ro</sc>. ACM Transactions on Software Engineering and Methodology, 2016, 25, 1-50.	6.0	13
36	DataLab. , 2008, , .		12

#	ARTICLE	IF	CITATIONS
37	Weaver. , 2010, , .		12
38	Taming complex bioinformatics workflows with weaver, makeflow, and starch. , 2010, , .		12
39	Facilitating the Reproducibility of Scientific Workflows with Execution Environment Specifications. <i>Procedia Computer Science</i> , 2017, 108, 705-714.	2.0	12
40	Towards Data Intensive Many-Task Computing. <i>Advances in Systems Analysis, Software Engineering, and High Performance Computing Book Series</i> , 0, , 28-73.	0.5	12
41	Grid Deployment of Legacy Bioinformatics Applications with Transparent Data Access. , 2006, , .		11
42	Expert-Citizen Engineering: "Crowdsourcing" Skilled Citizens. , 2011, , .		11
43	Folding proteins at 500 ns/hour with Work Queue. , 2012, 2012, 1-8.		11
44	Scaling up genome annotation using MAKER and work queue. <i>International Journal of Bioinformatics Research and Applications</i> , 2014, 10, 447.	0.2	11
45	Lessons Learned from Crowdsourcing Complex Engineering Tasks. <i>PLoS ONE</i> , 2015, 10, e0134978.	2.5	11
46	Designing Self-Tuning Split-Map-Merge Applications for High Cost-Efficiency in the Cloud. <i>IEEE Transactions on Cloud Computing</i> , 2017, 5, 303-316.	4.4	10
47	Efficient Integration of Containers into Scientific Workflows. , 2018, , .		10
48	Transparent access to Grid resources for user software. <i>Concurrency Computation Practice and Experience</i> , 2006, 18, 787-801.	2.2	9
49	Lessons Learned Building TeamTrak: An Urban/Outdoor Mobile Testbed. , 2007, , .		9
50	Highly scalable genome assembly on campus grids. , 2009, , .		9
51	Making work queue cluster-friendly for data intensive scientific applications. , 2013, , .		8
52	Umbrella. , 2015, , .		8
53	Harnessing parallelism in multicore clusters with the all-pairs and wavefront abstractions. , 2009, , .		8
54	Biocompute 2.0: an improved collaborative workspace for data intensive bio-science. <i>Concurrency Computation Practice and Experience</i> , 2011, 23, 2305-2314.	2.2	7

#	ARTICLE	IF	CITATIONS
55	Reduction of Workflow Resource Consumption Using a Density-based Clustering Model. , 2018, , .		7
56	A Lightweight Model for Right-Sizing Master-Worker Applications. , 2018, , .		7
57	Challenges in Executing Data Intensive Biometric Workloads on a Desktop Grid. , 2007, , .		6
58	Attaching Cloud Storage to a Campus Grid Using Parrot, Chirp, and Hadoop. , 2010, , .		6
59	PRUNE: A preserving run environment for reproducible scientific computing. , 2016, , .		6
60	Access control for a replica management database. , 2006, , .		5
61	Data mining on the grid for the grid. Parallel and Distributed Processing Symposium (IPDPS), Proceedings of the International Conference on, 2008, , .	1.0	5
62	Visualizing massively multithreaded applications with ThreadScope. Concurrency Computation Practice and Experience, 2010, 22, 45-67.	2.2	5
63	ROARS. , 2010, , .		5
64	Scripting distributed scientific workflows using Weaver. Concurrency Computation Practice and Experience, 2012, 24, 1685-1707.	2.2	5
65	Adapting bioinformatics applications for heterogeneous systems: a case study. Concurrency Computation Practice and Experience, 2014, 26, 866-877.	2.2	5
66	Adapting Collaborative Software Development Techniques to Structural Engineering. Computing in Science and Engineering, 2015, 17, 27-34.	1.2	5
67	SHADHO: Massively Scalable Hardware-Aware Distributed Hyperparameter Optimization. , 2018, , .		5
68	Cacheable Decentralized Groups for Grid Resource Access Control. , 2006, , .		4
69	Case Studies in Designing Elastic Applications. , 2013, , .		4
70	Accelerating Comparative Genomics Workflows in a Distributed Environment with Optimized Data Partitioning. , 2014, , .		4
71	Scaling Data Intensive Physics Applications to 10k Cores on Non-dedicated Clusters with Lobster. , 2015, , .		4
72	Scaling Up Bioinformatics Workflows with Dynamic Job Expansion: A Case Study Using Galaxy and Makeflow. , 2015, , .		4

#	ARTICLE	IF	CITATIONS
73	Confuga: Scalable Data Intensive Computing for POSIX Workflows. , 2015, , .		4
74	Taming metadata storms in parallel filesystems with metaFS. , 2017, , .		4
75	VC3. , 2018, , .		4
76	Automatic Dependency Management for Scientific Applications on Clusters. , 2018, , .		4
77	Operating System Support for Space Allocation in Grid Storage Systems. , 2006, , .		3
78	Biocompute. , 2010, , .		3
79	Adapting bioinformatics applications for heterogeneous systems. , 2011, , .		3
80	Fine-Grained Access Control in the Chirp Distributed File System. , 2012, , .		3
81	ROARS: a robust object archival system for data intensive scientific computing. Distributed and Parallel Databases, 2012, 30, 325-350.	1.6	3
82	Shifting the bioinformatics computing paradigm: A case study in parallelizing genome annotation using MAKER and Work Queue. , 2012, , .		3
83	Automated packaging of bioinformatics workflows for portability and durability using makeflow. , 2013, , .		3
84	Opportunistic High Energy Physics Computing in User Space with Parrot. , 2014, , .		3
85	Conducting reproducible research with Umbrella: Tracking, creating, and preserving execution environments. , 2016, , .		3
86	Early Experience Using Amazon Batch for Scientific Workflows. , 2018, , .		3
87	Autoscaling High-Throughput Workloads on Container Orchestrators. , 2020, , .		3
88	Harnessing HPC resources for CMS jobs using a Virtual Private Network. EPJ Web of Conferences, 2021, 251, 02032.	0.3	3
89	Abstractions for Cloud Computing with Condor. , 2010, , 153-171.		3
90	Data Intensive Computing with Clustered Chirp Servers. Advances in Systems Analysis, Software Engineering, and High Performance Computing Book Series, 0, , 140-154.	0.5	3

#	ARTICLE	IF	CITATIONS
91	On-Demand Transient Data Storage and Backup in Mobile Systems. , 2007, , .		2
92	Reflections on the virtues of modularity: a case study in linux security modules. Software - Practice and Experience, 2009, 39, 1235-1251.	3.6	2
93	Towards long term data quality in a large scale biometrics experiment. , 2010, , .		2
94	A system for management of Computational Fluid Dynamics simulations for civil engineering. , 2012, , .		2
95	Design of an active storage cluster file system for DAG workflows. , 2013, , .		2
96	Balancing Thread-Level and Task-Level Parallelism for Data-Intensive Workloads on Clusters and Clouds. , 2015, , .		2
97	Exploiting volatile opportunistic computing resources with Lobster. Journal of Physics: Conference Series, 2015, 664, 032035.	0.4	2
98	An analysis of reproducibility and non-determinism in HEP software and ROOT data. Journal of Physics: Conference Series, 2017, 898, 102007.	0.4	2
99	Log Discovery for Troubleshooting Open Distributed Systems with TLQ. , 2020, , .		2
100	Positioning Dynamic Storage Caches for Transient Data. , 2006, , .		1
101	Work in progress - integrating undergraduate research and education with the TeamTrak mobile computing system. Proceedings - Frontiers in Education Conference, FIE, 2007, , .	0.0	1
102	BXGrid: A Data Repository and Workflow Abstraction for Biometrics Research. , 2008, , .		1
103	Using Small Abstractions to Program Large Distributed Systems. , 2008, , .		1
104	DeltaDB: A Scalable Database Design for Time-Varying Schema-Free Data. , 2014, , .		1
105	Opportunistic Computing with Lobster: Lessons Learned from Scaling up to 25k Non-Dedicated Cores. Journal of Physics: Conference Series, 2017, 898, 052036.	0.4	1
106	An Algebra for Robust Workflow Transformations. , 2018, , .		1
107	Flexible Partitioning of Scientific Workflows Using the JX Workflow Language. , 2019, , .		1
108	Dynamic Sizing of Continuously Divisible Jobs for Heterogeneous Resources. , 2019, , .		1

#	ARTICLE	IF	CITATIONS
109	Not All Tasks Are Created Equal: Adaptive Resource Allocation for Heterogeneous Tasks in Dynamic Workflows. , 2021, , .		1
110	Emerging Frameworks for Advancing Scientific Workflows Research, Development, and Education. , 2021, , .		1
111	Dynamic Task Shaping for High Throughput Data Analysis Applications in High Energy Physics. , 2022, , .		1
112	Experience With A Literate Approach to Computer Science. , 2006, , .		0
113	GEMS: User Control for Cooperative Scientific Repositories. Computer Communications and Networks, 2011, , 57-87.	0.8	0
114	Expanding Tasks of Logical Workflows Into Independent Workflows for Improved Scalability. , 2014, , .		0
115	Balancing push and pull in Confuga, an active storage cluster file system for scientific workflows. Concurrency Computation Practice and Experience, 2017, 29, e3834.	2.2	0
116	Report on the first workshop on negative and null results in eScience. Concurrency Computation Practice and Experience, 2017, 29, e3908.	2.2	0
117	Scaling up a CMS tier-3 site with campus resources and a 100 Gb/s network connection: what could go wrong?. Journal of Physics: Conference Series, 2017, 898, 082041.	0.4	0
118	A First Look at the JX Workflow Language. , 2018, , .		0
119	MAKER as a Service: Moving HPC Applications to Jetstream Cloud. , 2018, , .		0
120	Deploying and extending CMS Tier 3s using VC3 and the OSG Hosted CE service. EPJ Web of Conferences, 2019, 214, 03035.	0.3	0
121	Coordination of Access to Large-Scale Datasets in Distributed Environments. Chapman & Hall/CRC Computational Science, 2009, , .	0.5	0
122	Accelerating Comparative Genomics Work ows in a Distributed Environment with Optimized Data Partitioning and Work ow Fusion. Scalable Computing, 2015, 16, .	1.0	0