

Big data analytics on Apache Spark

International Journal of Data Science and Analytics
1, 145-164

DOI: [10.1007/s41060-016-0027-9](https://doi.org/10.1007/s41060-016-0027-9)

Citation Report

#	ARTICLE	IF	CITATIONS
1	Big Data Infrastructure: A Survey. Lecture Notes in Computer Science, 2017, , 249-258.	1.0	3
2	A comprehensive review from sequential association computing to Hadoop-MapReduce parallel computing in a retail scenario. Journal of Management Analytics, 2017, 4, 359-392.	1.6	22
3	Efficient Distributed Smith-Waterman Algorithm Based on Apache Spark. , 2017, , .		9
4	A RST-based stateful data analytics within spark. , 2017, , .		0
5	A Review of Infrastructures to Process Big Multimedia Data. International Journal of Computer Vision and Image Processing, 2017, 7, 54-64.	0.3	3
6	Docker vs. KVM: Apache spark application performance and ease of use. , 2017, , .		2
7	Spark Based Framework for Breast Cancer Analysis. SSRN Electronic Journal, 0, , .	0.4	1
8	Personalized Online Analytical Processing in Big Data Context Using User Profile and Search Context. International Journal of Strategic Information Technology and Applications, 2017, 8, 67-80.	0.6	1
9	Processing Big Data in Field of Marketing Models Using Apache Spark. Lecture Notes in Electrical Engineering, 2018, , 49-58.	0.3	1
10	All-at-once Decomposition of Coupled Billion-scale Tensors in Apache Spark. , 2018, , .		5
11	Power analysis of sorting algorithms on FPGA using OpenCL. , 2018, , .		2
12	An Approach for Optimizing the Performance for Apache Spark Applications. , 2018, , .		5
13	MapReduce Model for Random Forest Algorithm: Experimental Studies. Lecture Notes in Computer Science, 2018, , 184-194.	1.0	0
14	Optimal Task Scheduling for Distributed Cluster With Active Storage Devices and Accelerated Nodes. IEEE Access, 2018, 6, 48195-48209.	2.6	4
15	Efficiently Processing and Storing Library Linked Data using Apache Spark and Parquet. Information Technology and Libraries, 2018, 37, 29-49.	0.5	4
16	PF-Face: A Parallel Framework for Face Classification and Search from Massive Videos Based on Spark. , 2018, , .		5
17	Performance Comparison of Bitcoin Prediction in Big Data Environment. , 2018, , .		3
18	Big Data and Data Analytics in Aviation. , 2018, , 55-65.		4

#	ARTICLE	IF	CITATIONS
19	A Two-Stage Data Processing Algorithm to Generate Random Sample Partitions for Big Data Analysis. Lecture Notes in Computer Science, 2018, , 347-364.	1.0	13
20	An Efficient Traffic Monitoring Model Using a Stream Processing Platform Based on Smart Highways Events Generator. Advances in Intelligent Systems and Computing, 2019, , 35-44.	0.5	0
21	Exploring and cleaning big data with random sample data blocks. Journal of Big Data, 2019, 6, .	6.9	18
22	Scalable architecture for Big Data financial analytics: user-defined functions vs. SQL. Journal of Big Data, 2019, 6, .	6.9	12
23	A spark-based big data analysis framework for real-time sentiment prediction on streaming data. Software - Practice and Experience, 2019, 49, 1352-1364.	2.5	12
24	Implementing Big Data Lake for Heterogeneous Data Sources. , 2019, , .		27
25	Leveraging resource management for efficient performance of Apache Spark. Journal of Big Data, 2019, 6, .	6.9	20
26	jMetalPy: A Python framework for multi-objective optimization with metaheuristics. Swarm and Evolutionary Computation, 2019, 51, 100598.	4.5	143
27	An adaptive and real-time based architecture for financial data integration. Journal of Big Data, 2019, 6, .	6.9	16
28	Random Sample Partition: A Distributed Data Model for Big Data Analysis. IEEE Transactions on Industrial Informatics, 2019, 15, 5846-5854.	7.2	66
29	How Good is Query Optimizer in Spark?. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, 2019, , 595-609.	0.2	0
30	Integration of Apache Spark with Invasive Resource Manager. , 2019, , .		1
31	Distributed sentiment analysis of an agglutinative language via Spark by applying machine learning methods. , 2019, , .		4
32	Lifelong Machine Learning and root cause analysis for large-scale cancer patient data. Journal of Big Data, 2019, 6, .	6.9	6
33	Cluster-based Module to Manage Smart Grid Data for an Enhanced Situation Awareness: A Case Study. , 2019, , .		3
34	Apache Spark: A Big Data Processing Engine. , 2019, , .		21
35	Dynamic Container-based Resource Management Framework of Spark Ecosystem. , 2019, , .		11
36	The big data system, components, tools, and technologies: a survey. Knowledge and Information Systems, 2019, 60, 1165-1245.	2.1	78

#	ARTICLE	IF	CITATIONS
37	An Asymptotic Ensemble Learning Framework for Big Data Analysis. IEEE Access, 2019, 7, 3675-3693.	2.6	17
38	Implementation of an Alternating Least Square Model Based Collaborative Filtering Movie Recommendation System on Hadoop and Spark Platforms. Lecture Notes on Data Engineering and Communications Technologies, 2019, , 237-249.	0.5	2
39	A dynamic, interpretable, and robust hybrid data analytics system for train movements in large-scale railway networks. International Journal of Data Science and Analytics, 2020, 9, 95-111.	2.4	15
40	Large-scale predictive modeling and analytics through regression queries in data management systems. International Journal of Data Science and Analytics, 2020, 9, 17-55.	2.4	2
41	Parallel particle swarm optimization classification algorithm variant implemented with Apache Spark. Concurrency Computation Practice and Experience, 2020, 32, e5451.	1.4	6
42	A R2N2 Approach For Cardiac Behavior Forecast on Non-Trending Big HealthCare Data. , 2020, , .		0
43	An Adaptable Big Data Value Chain Framework for End-to-End Big Data Monetization. Big Data and Cognitive Computing, 2020, 4, 34.	2.9	17
44	Chi-BD-DRF: Design of Scalable Fuzzy Classifiers for Big Data via A Dynamic Rule Filtering Approach. , 2020, , .		5
45	Storing, combining and analysing turkey experimental data in the Big Data era. Animal, 2020, 14, 2397-2403.	1.3	3
46	Comparative Study of Classification Algorithms to Analyze and Predict a Twitter Sentiment in Apache Spark. IOP Conference Series: Materials Science and Engineering, 2020, 928, 032045.	0.3	4
47	RCBO â€œ A Big Data Classification Based On an Efficient RCBO Optimization Technique and Apache Spark. , 2020, , .		1
48	An empirical study of alternating least squares collaborative filtering recommendation for MovieLens on Apache Hadoop and Spark. International Journal of Grid and Utility Computing, 2020, 11, 674.	0.1	3
49	AxBy: Approximate Computation Bypass for Data-Intensive Applications. , 2020, , .		3
50	A Novel Method for Classifying Function of Spatial Regions Based on Two Sets of Characteristics Indicated by Trajectories. International Journal of Data Warehousing and Mining, 2020, 16, 1-19.	0.4	2
51	Automatic parallelization of representative-based clustering algorithms for multicore cluster systems. International Journal of Data Science and Analytics, 2020, 10, 135-159.	2.4	3
52	A Tile-Based Framework with a Spatial-Aware Feature for Easy Access and Efficient Analysis of Marine Remote Sensing Data. Remote Sensing, 2020, 12, 1932.	1.8	2
53	A survey of data partitioning and sampling methods to support big data analysis. Big Data Mining and Analytics, 2020, 3, 85-101.	7.5	152
54	Big data analytics for retail industry using MapReduce-Apriori framework. Journal of Management Analytics, 2020, 7, 424-442.	1.6	31

#	ARTICLE	IF	CITATIONS
55	Bayesian neural networks for flight trajectory prediction and safety assessment. Decision Support Systems, 2020, 131, 113246.	3.5	87
56	On the design of a framework integrating an optimization engine with streaming technologies. Future Generation Computer Systems, 2020, 107, 538-550.	4.9	10
57	Deadline-Aware Cost Optimization for Spark. IEEE Transactions on Big Data, 2021, 7, 115-127.	4.4	10
58	DI-Mondrian: Distributed improved Mondrian for satisfaction of the L-diversity privacy model using Apache Spark. Information Sciences, 2021, 546, 1-24.	4.0	22
59	Big data based hybrid machine learning model for improving performance of medical Internet of Things data in healthcare systems. , 2021, , 47-62.		27
60	Power-Law Distributed Graph Generation With MapReduce. IEEE Access, 2021, 9, 94405-94415.	2.6	2
61	Analysis of Big Data Storage Tools for Data Lakes based on Apache Hadoop Platform. International Journal of Advanced Computer Science and Applications, 2021, 12, .	0.5	2
62	Big data workflow platforms for science. AIP Conference Proceedings, 2021, , .	0.3	1
63	Big-Parallel-ETL: New ETL for Multidimensional NoSQL Graph Oriented Data. Journal of Physics: Conference Series, 2021, 1743, 012037.	0.3	3
64	The Power of a Collective: Team of Agents Solving Instances of the Flow Shop and Job Shop Problems. Lecture Notes in Computer Science, 2021, , 406-419.	1.0	1
65	Just-in-Time Sentiment Analysis for Streamed Data in Greek. Lecture Notes in Computer Science, 2021, , 249-263.	1.0	0
66	Choosing a Data Storage Format in the Apache Hadoop System Based on Experimental Evaluation Using Apache Spark. Symmetry, 2021, 13, 195.	1.1	11
67	Trade-Off Analysis Between Parallelism and Accuracy of SLIC on Apache Spark. , 2021, , .		2
69	Realtime analysis and visualization of data for instant decisions: A futuristic requirement of the digital world. Materials Today: Proceedings, 2021, , .	0.9	2
70	Efficient Performance Prediction for Apache Spark. Journal of Parallel and Distributed Computing, 2021, 149, 40-51.	2.7	21
71	Smart Sensing with Edge Computing in Precision Agriculture for Soil Assessment and Heavy Metal Monitoring: A Review. Agriculture (Switzerland), 2021, 11, 475.	1.4	29
72	Performance Comparison between S3, HDFS and RDS storage technologies for real-time big-data applications. , 2021, , .		0
73	Towards Energy Efficiency in Data Centers: An Industrial Experience Based on Reuse and Layout Changes. Applied Sciences (Switzerland), 2021, 11, 4719.	1.3	2

#	ARTICLE	IF	CITATIONS
74	Prediction of Heart Stroke using A Novel Framework â€“ PySpark. International Journal of Preventive Medicine and Health, 2021, 1, 1-4.	1.5	1
75	AVUBDI: A Versatile Usable Big Data Infrastructure and Its Monitoring Approaches for Process Industry. Frontiers in Chemical Engineering, 2021, 3, .	1.3	2
76	Credit Card Fraud Detection using Apache Spark Analysis. , 2021, , .		2
78	Big data classification using deep learning and apache spark architecture. Neural Computing and Applications, 2021, 33, 15253.	3.2	1
79	Rider Chaotic Biography Optimization-driven Deep Stacked Auto-encoder for Big Data Classification Using Spark Architecture. International Journal of Web Services Research, 2021, 18, 42-62.	0.5	6
80	DHkmeans-â„“diversity: distributed hierarchical K-means for satisfaction of the â„“-diversity privacy model using Apache Spark. Journal of Supercomputing, 2022, 78, 2616-2650.	2.4	5
81	Network Intrusion Detection in the Wild - the Orange use case in the SIMARGL project. , 2021, , .		2
82	Recent Developments in Parallel and Distributed Computing for Remotely Sensed Big Data Processing. Proceedings of the IEEE, 2021, 109, 1282-1305.	16.4	45
83	A parallel and accurate method for large-scale image segmentation on a cloud environment. Journal of Supercomputing, 2022, 78, 4330-4357.	2.4	2
84	Big data analytics for default prediction using graph theory. Expert Systems With Applications, 2021, 176, 114840.	4.4	30
86	Cloud-agnostic architectures for machine learning based on Apache Spark. Advances in Engineering Software, 2021, 159, 103029.	1.8	6
87	Experimental Characteristics Study of Data Storage Formats for Data Marts Development within Data Lakes. Applied Sciences (Switzerland), 2021, 11, 8651.	1.3	1
88	A Computational Intelligence Approach to Predict Energy Demand Using Random Forest in a Cloudera Cluster. Applied Sciences (Switzerland), 2021, 11, 8635.	1.3	6
89	Detection of COVID-19 in Chest X-ray Images: A Big Data Enabled Deep Learning Approach. International Journal of Environmental Research and Public Health, 2021, 18, 10147.	1.2	36
90	Real-Time DDoS Attack Detection System Using Big Data Approach. Sustainability, 2021, 13, 10743.	1.6	86
91	WS-PDC: Persistent Distributed Channel-Based Web Services Applied on IFRS Data Processing and Loading. Lecture Notes in Networks and Systems, 2022, , 847-855.	0.5	0
92	Cricket Match Analytics Using the Big Data Approach. Electronics (Switzerland), 2021, 10, 2350.	1.8	25
93	Tuning configuration of apache spark on public clouds by combining multi-objective optimization and performance prediction model. Journal of Systems and Software, 2021, 180, 111028.	3.3	10

#	ARTICLE	IF	CITATIONS
94	PLDLS: A novel parallel label diffusion and label Selection-based community detection algorithm based on Spark in social networks. Expert Systems With Applications, 2021, 183, 115377.	4.4	19
95	JP-DAP: An Intelligent Data Analytics Platform for Metro Rail Transport Systems. IEEE Transactions on Intelligent Transportation Systems, 2022, 23, 9146-9156.	4.7	8
96	A Multi-Agent-Based System for Intrusion Detection. Smart Innovation, Systems and Technologies, 2021, , 177-191.	0.5	1
97	A Big Data Approach to Black Friday Sales. Intelligent Automation and Soft Computing, 2021, 27, 785-797.	1.6	57
98	Distributed and Parallel Ensemble Classification for Big Data Based on Kullback-Leibler Random Sample Partition. Lecture Notes in Computer Science, 2020, , 448-464.	1.0	8
99	A Near Metal Platform for Intensive Big Data Processing Using A Novel Approach. , 2020, , .		1
100	Predictors of outpatientsâ€™ no-show: big data analytics using apache spark. Journal of Big Data, 2020, 7, .	6.9	11
101	EMOGI. Proceedings of the VLDB Endowment, 2020, 14, 114-127.	2.1	19
102	A Capability Approach for Designing Business Intelligence and Analytics Architectures. , 2020, , .		2
103	A Fine-Grained Stateful Data Analytics Method Based on Resilient State Table. International Journal of Software Science and Computational Intelligence, 2018, 10, 66-79.	1.8	1
104	Big data clustering techniques based on Spark: a literature review. PeerJ Computer Science, 2020, 6, e321.	2.7	19
105	Spark framework for transcriptomic trimming algorithm reduces cost of reading multiple input files. , 2017, , .		1
106	Big Data Classification using the Deep Learning Enabled Spark Architecture: A Survey. SSRN Electronic Journal, 0, , .	0.4	1
107	Evaluating the Boundaries of Big Data Environments for Machine Learning. Lecture Notes in Computer Science, 2019, , 253-264.	1.0	0
108	Assessing the Dependability of Apache Spark System: Streaming Analytics on Large-Scale Ocean Data. Communications in Computer and Information Science, 2019, , 131-144.	0.4	0
109	Big Data Analysis Using Apache Spark MLlib and Hadoop HDFS with Scala and Java. Kurdistan Journal of Applied Research, 2019, 4, 7-14.	0.4	10
110	SparkFuzz. , 2020, , .		6
111	Industry 4.0: Roadmap for Applying Technologies in Shipbuilding and Manufacturing Sectors. , 2020, , .		0

#	ARTICLE	IF	CITATIONS
112	Multi-Timescale Load Forecast of Large Power Customers Based on Online Data Recovery and Time Series Neural Networks. Journal of Circuits, Systems and Computers, 2022, 31, .	1.0	2
113	Propositional Aspect between Apache Spark and Hadoop Map-Reduce for Stock Market Data. , 2020, , .		1
114	A Study of Big Data Analytics using Apache Spark with Python and Scala. , 2020, , .		7
115	Large scale data mining for banking credit risk prediction. , 2020, , .		0
116	A Sentiment Analysis Service Platform for Streamed Multilingual Tweets. , 2020, , .		0
117	A Review of Infrastructures to Process Big Multimedia Data. , 2020, , 1-12.		0
118	Generalizing Streaming Pipeline Design for Big Data. Advances in Intelligent Systems and Computing, 2020, , 149-160.	0.5	2
119	Classification of Big Data Using Spark Framework. Lecture Notes in Electrical Engineering, 2021, , 847-854.	0.3	2
120	Evaluation Platform for DDM Algorithms With the Usage of Non-Uniform Data Distribution Strategies. International Journal of Information Technologies and Systems Approach, 2021, 15, 1-23.	0.8	6
121	A Critical Analysis of Apache Hadoop and Spark for Big Data Processing. , 2021, , .		6
122	A Comparative Analysis of Big Data Frameworks: An Adoption Perspective. Applied Sciences (Switzerland), 2021, 11, 11033.	1.3	8
123	Recent implications towards sustainable and energy efficient AI and big data implementations in cloud-fog systems: A newsworthy inquiry. Journal of King Saud University - Computer and Information Sciences, 2021, , .	2.7	1
124	An efficient approach for detecting anomalous events in real-time weather datasets. Concurrency Computation Practice and Experience, 2022, 34, e6707.	1.4	6
125	Efficient Execution of Dynamic Programming Algorithms on Apache Spark. , 2020, , .		2
126	Big data Predictive Analytics for Apache Spark using Machine Learning. , 2020, , .		0
127	A Survey On Cloud-Based Distributed Computing System Frameworks. , 2020, , .		2
128	Modelling and simulating vessel emissions in real time based on terrestrial AIS data. , 2020, , .		3
129	The Subnetwork Investigation of Scale-Free Networks Based on the Self-Similarity. SSRN Electronic Journal, 0, , .	0.4	0

#	ARTICLE	IF	CITATIONS
130	Prediction of Heart Stroke using A Novel Framework â€“ PySpark. International Journal of Preventive Medicine and Health, 2021, 1, 1-4.	1.5	3
131	Real-Time Big Data Analytics Perspective on Applications, Frameworks and Challenges. , 2021, , .		4
132	Edge-Based Sampling for Complex Network with Self-Similar Structure. , 2021, , .		1
133	Near-Real Time Quality Prediction in a Plastic Injection Molding Process Using Apache Spark. , 2021, , .		2
135	Apache Spark and MLib-Based Intrusion Detection System or How the Big Data Technologies Can Secure the Data. Information (Switzerland), 2022, 13, 58.	1.7	18
136	An Efficient Spark-Based Hybrid Frequent Itemset Mining Algorithm for Big Data. Data, 2022, 7, 11.	1.2	7
137	Paradise: Real-Time, Generalized, and Distributed Provenance-Based Intrusion Detection. IEEE Transactions on Dependable and Secure Computing, 2023, 20, 1624-1640.	3.7	4
139	An Architecture for Service Integration to Fully Support Novel Personalized Smart Tourism Offerings. Sensors, 2022, 22, 1619.	2.1	8
140	Comparative analysis of Spark and Hadoop through Imputation of Data on Big Datasets. , 2021, , .		2
141	Jespipe: A Plugin-Based, Open MPI Framework for Adversarial Machine Learning Analysis. , 2021, , .		3
142	Big Data Analytics Tools: A Comparative Study. , 2021, , .		3
143	Apache Spark based analysis on word count application in Big Data. , 2022, , .		3
144	A Big Data Pipeline and Machine Learning for Uniform Semantic Representation of Data and Documents from IT Systems of Italian Ministry of Justice. International Journal of Grid and High Performance Computing, 2022, 14, 0-0.	0.7	12
145	Big Data Methods in Learning Analytics System by Using Dask Cluster Computer Framework. Lecture Notes on Data Engineering and Communications Technologies, 2022, , 314-323.	0.5	1
147	Collaborative Filtering-Based Music Recommendation in Spark Architecture. Mathematical Problems in Engineering, 2022, 2022, 1-8.	0.6	0
148	An efficient system using implicit feedback and lifelong learning approach to improve recommendation. Journal of Supercomputing, 0, , 1.	2.4	2
149	Runtime prediction of big data jobs: performance comparison of machine learning algorithms and analytical models. Journal of Big Data, 2022, 9, .	6.9	5
150	Predictions on diabetic patient datasets using big data analytics and machine learning techniques. , 2022, , 179-199.		0

#	ARTICLE	IF	CITATIONS
151	Big data analytics in healthcare. , 2022, , 3-11.		2
152	A Compendious Research on Big Data File Formats. , 2022, , .		0
153	Modified valence aware dictionary for sentiment reasoning classifier for detection and classification of Covid-19 related rumors from social media data streams. Concurrency Computation Practice and Experience, 2022, 34, .	1.4	1
154	Big Data Analysis for Healthcare Application using Minhash and Machine Learning in Apache Spark Framework. , 2022, , .		0
155	Spatiotemporal data partitioning for distributed random forest algorithm: Air quality prediction using imbalanced big spatiotemporal data on spark distributed framework. Environmental Technology and Innovation, 2022, 27, 102776.	3.0	8
156	Scalable Machine Learning Using PySpark. , 2022, , .		2
157	Real-Time Big Data Analytics for Data Stream Challenges: An Overview. European Journal of Education and Pedagogy, 2022, 2, 1-6.	0.2	2
158	A Machine-Learning Pipeline for Large-Scale Power-Quality Forecasting in the Mexican Distribution Grid. Applied Sciences (Switzerland), 2022, 12, 8423.	1.3	3
159	A Highly Discriminative Hybrid Feature Selection Algorithm for Cancer Diagnosis. Scientific World Journal, The, 2022, 2022, 1-15.	0.8	4
160	Parallel incremental efficient attribute reduction algorithm based on attribute tree. Information Sciences, 2022, 610, 1102-1121.	4.0	11
161	Random Partition Based Adaptive Distributed Kernelized SVM for Big Data. IEEE Access, 2022, 10, 95623-95637.	2.6	1
162	Distributed Anomalies Detection Using Isolation Forest and Spark. Communications in Computer and Information Science, 2022, , 700-712.	0.4	0
163	A taxonomy of tools and approaches for distributed genomic analyses. Informatics in Medicine Unlocked, 2022, 32, 101024.	1.9	0
164	An Approach to Business Workflow Software Architectures: A Case Study for Bank Account Transaction Type Prediction. Lecture Notes in Computer Science, 2022, , 709-723.	1.0	0
165	Street Smart in 5G: Vehicular Applications, Communication, and Computing. IEEE Access, 2022, 10, 105631-105656.	2.6	10
166	Optimization of Population Mobility Model of Artificial Society Based on Big Data Analytics. , 2022, , .		0
167	Forestry Big Data: A Review and Bibliometric Analysis. Forests, 2022, 13, 1549.	0.9	9
168	Impact on stock exchange due to Covid-19 using apache spark. AIP Conference Proceedings, 2022, , .	0.3	0

#	ARTICLE	IF	CITATIONS
169	Detecting Reconnaissance and Discovery Tactics from the MITRE ATT&CK Framework in Zeek Conn Logs Using Spark's Machine Learning in the Big Data Framework. <i>Sensors</i> , 2022, 22, 7999.	2.1	6
170	A Cluster-Based Partition Method of Remote Sensing Data for Efficient Distributed Image Processing. <i>Remote Sensing</i> , 2022, 14, 4964.	1.8	0
171	Plataforma de Monitoramento Inteligente da COVID-19. <i>Research, Society and Development</i> , 2022, 11, e580111436659.	0.0	2
172	Improved multi-class classification approach for imbalanced big data on spark. <i>Journal of Supercomputing</i> , 2023, 79, 6583-6611.	2.4	3
173	An Apache Spark Implementation for Text Document Clustering. , 2022, , .		0
174	A brief survey on big data: technologies, terminologies and data-intensive applications. <i>Journal of Big Data</i> , 2022, 9, .	6.9	8
175	PERBANDINGAN WAKTU EKSEKUSI PERAMALAN HARGA KOMODITAS PANGAN MENGGUNAKAN SPARKR DAN R STUDIO. <i>Jurnal Teknologi Informasi Jurnal Keilmuan Dan Aplikasi Bidang Teknik Informatika</i> , 2022, 16, 73-80.	0.1	0
176	A Scalable Analytical Framework for Complex Event Episode Mining With Various Domains Applications. <i>IEEE Access</i> , 2022, 10, 130672-130685.	2.6	0
177	Three-Way Decisions on Streaming Computing Platforms Supporting Decision-Making in Complex Large Real-World Environments. <i>IEEE Access</i> , 2022, 10, 122314-122328.	2.6	1
178	Enabling real time big data solutions for manufacturing at scale. <i>Journal of Big Data</i> , 2022, 9, .	6.9	6
179	Distributed-Memory FastFlow Building Blocks. <i>International Journal of Parallel Programming</i> , 2023, 51, 1-21.	1.1	5
180	Data Lake Architecture for Smart Fish Farming Data-Driven Strategy. <i>Applied System Innovation</i> , 2023, 6, 8.	2.7	1
181	Apache Spark-based scalable feature extraction approaches for protein sequence and their clustering performance analysis. <i>International Journal of Data Science and Analytics</i> , 0, , .	2.4	0
182	A comprehensive study and review of tuning the performance on database scalability in big data analytics. <i>Journal of Intelligent and Fuzzy Systems</i> , 2023, 44, 5231-5255.	0.8	4
183	A Machine Learning Approach for Predicting Execution Statistics of Spark Application. , 2022, , .		1
184	Towards Query Performance Improvement in Big Data Environment. , 2023, , 772-788.		0
185	Mining Skyline Patterns from Big Data Environments based on a Spark Framework. <i>Journal of Grid Computing</i> , 2023, 21, .	2.5	1
186	On the building of efficient self-adaptable health data science services by using dynamic patterns. <i>Future Generation Computer Systems</i> , 2023, 145, 478-495.	4.9	2

#	ARTICLE	IF	CITATIONS
187	Sustainable Big Data Analytics Process Pipeline Using Apache Ecosystem. , 2022, , 1247-1259.		0
188	Big Data Processing: A review. , 0, , 23-30.		0
189	Mobile Big Data Analytics Using Deep Learning and Apache Spark. , 0, , 16-28.		2
190	Leveraging State-of-the-Art Engines for Large-Scale Data Analysis in High Energy Physics. Journal of Grid Computing, 2023, 21, .	2.5	2
191	Comparative Analysis of Techniques for Big-data Performance Testing. , 2022, , .		1
192	Friend Recommendation System Using Map-Reduce and Spark: A Comparison Study. , 2023, , .		1
193	Toward a smart health: big data analytics and IoT for real-time miscarriage prediction. Journal of Big Data, 2023, 10, .	6.9	2
194	Enhancing Digital Health Services with Big Data Analytics. Big Data and Cognitive Computing, 2023, 7, 64.	2.9	6
195	An intelligent traffic monitoring approach based on Hadoop ecosystem. , 2022, , .		2
196	A comparative study of big data use in Egyptian agriculture. Journal of Electrical Systems and Information Technology, 2023, 10, .	1.2	2
197	Deep Web Search Log Mining Using Spark for Efficient and Scalable Web Information Retrieval. Advances in Intelligent Systems and Computing, 2023, , 19-31.	0.5	0
198	Big Data Approach For IoT Botnet Traffic Detection Using Apache Spark Technology. , 2023, , .		1
199	An Intelligent Energy Management Information System with Machine Learning Algorithms in Oil and Gas Industry. Wireless Communications and Mobile Computing, 2023, 2023, 1-7.	0.8	0
203	SQL Query Optimization in Distributed NoSQL Databases for Cloud-Based Applications. Lecture Notes in Computer Science, 2023, , 21-41.	1.0	0
205	Framework for integrating healthcare big data using IoMT technology. , 2023, , 191-210.		0
208	Zero-Shot Cost Models for Parallel Stream Processing. , 2023, , .		0
212	More Sparking Soundex-Based Privacy-Preserving Record Linkage. Lecture Notes in Computer Science, 2023, , 73-93.	1.0	1
213	Towards Verified Scalable Parallel Computing with Coq and Spark. , 2023, , .		1

#	ARTICLE	IF	CITATIONS
215	Extending Composable Data Services into SmartNICs. , 2023, , .		1
216	SPOAHA: Spark Program Optimizer Based on Artificial Hummingbird Algorithm. Lecture Notes in Computer Science, 2023, , 317-331.	1.0	0
219	Building a Recommendation System for E-Commerce Using Machine Learning and Big Data Technologies. Lecture Notes in Networks and Systems, 2023, , 376-387.	0.5	0
220	Analysis of Optimization Strategies for Big Data Storage Management: A Study. , 2023, , .		0
222	A Novel Spark-Based Algorithm for Mining Frequent Utility Patterns. , 2023, , .		0
225	Association Rule Mining in Distributed Environment: A Survey. Lecture Notes in Networks and Systems, 2023, , 113-120.	0.5	0
226	Exploring the Power of Big Data for IoT: A Comprehensive Review. , 2023, , .		0
230	Technology Landscape for Making Climate and Environmental Science "Actionable", 2023, , 55-82.		0
232	A Comparative Study of Bigdata Tools: Hadoop Vs Spark Vs Storm. , 2023, , .		1
233	Accelerating I/O in Distributed Data Processing Systems with Apache Arrow CHFS. , 2023, , .		0
234	Temporal Variability Analysis of Domain Names using Graph Techniques on Big Data. , 2023, , .		0
235	Examining Amazon Customer Reviews using PySpark and AWS: A Data Lake Approach. , 2023, , .		0
236	Automated Analysis with Event Log Enrichment of the European Public Procurement Processes. Lecture Notes in Computer Science, 2023, , 178-188.	1.0	0
239	An Architectural Model for Integrating Big Data in Educational Information Systems. Lecture Notes in Networks and Systems, 2024, , 282-288.	0.5	0
240	Cloud Spark Cluster to Analyse English Prescription Big Data for NHS Intelligence. Lecture Notes in Networks and Systems, 2024, , 361-375.	0.5	0
242	FAI: A Fraudulent Account Identification System. Lecture Notes in Computer Science, 2024, , 253-257.	1.0	0