Resource Summary Report

Generated by RRID on Apr 11, 2025

Apache Spark

RRID:SCR_016557
Type: Tool

Proper Citation

Apache Spark (RRID:SCR_016557)

Resource Information

URL: http://spark.apache.org/

Proper Citation: Apache Spark (RRID:SCR_016557)

Description: Software tool as a unified analytics engine for large scale data processing. An open source distributed general purpose cluster computing framework.

Resource Type: software resource, software application, data processing software, data analytics software

Defining Citation: DOI:10.1145/2934664

Keywords: unified, analytics, engine, large, scale, data, processing

Funding:

Availability: Free, Available for download, Freely available

Resource Name: Apache Spark

Resource ID: SCR_016557

License: Apache License, Version 2.0.

Record Creation Time: 20220129T080331+0000

Record Last Update: 20250411T055916+0000

Ratings and Alerts

No rating or validation information has been found for Apache Spark.

No alerts have been found for Apache Spark.

Data and Source Information

Source: SciCrunch Registry

Usage and Citation Metrics

We found 28 mentions in open access literature.

Listed below are recent publications. The full list is available at RRID.

Alonso R, et al. (2024) Improving digital twin experience through big data, IoT and social analysis: An architecture and a case study. Heliyon, 10(2), e24741.

Yönyül B, et al. (2024) MonARCh: an actor based architecture for dynamic linked data monitoring. PeerJ. Computer science, 10, e2133.

Hagan NKA, et al. (2024) SparkDWM: a scalable design of a Data Washing Machine using Apache Spark. Frontiers in big data, 7, 1446071.

Özgüven YM, et al. (2023) Distributed messaging and light streaming system for combating pandemics: A case study on spatial analysis of COVID-19 Geo-tagged Twitter dataset. Journal of ambient intelligence and humanized computing, 14(2), 773.

Perova Z, et al. (2023) PDCM Finder: an open global research platform for patient-derived cancer models. Nucleic acids research, 51(D1), D1360.

Laurie S, et al. (2022) The RD-Connect Genome-Phenome Analysis Platform: Accelerating diagnosis, research, and gene discovery for rare diseases. Human mutation, 43(6), 717.

Perez-Riverol Y, et al. (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic acids research, 50(D1), D543.

Schubotz M, et al. (2022) Caching and Reproducibility: Making Data Science Experiments Faster and FAIRer. Frontiers in research metrics and analytics, 7, 861944.

Munoz-Arcentales A, et al. (2021) Enabling Context-Aware Data Analytics in Smart Environments: An Open Source Reference Implementation. Sensors (Basel, Switzerland), 21(21).

Maarala AI, et al. (2021) Distributed hybrid-indexing of compressed pan-genomes for scalable and fast sequence alignment. PloS one, 16(8), e0255260.

Smith J, et al. (2021) Scalable analysis of multi-modal biomedical data. GigaScience, 10(9).

Chai X, et al. (2021) Analysis of spatiotemporal mobility of shared-bike usage during COVID-19 pandemic in Beijing. Transactions in GIS : TG, 25(6), 2866.

Marcon Y, et al. (2021) Orchestrating privacy-protected big data analyses of data from different resources with R and DataSHIELD. PLoS computational biology, 17(3), e1008880.

Rao P, et al. (2021) Gender Bias in the News: A Scalable Topic Modelling and Visualization Framework. Frontiers in artificial intelligence, 4, 664737.

Cortis K, et al. (2021) Over a decade of social opinion mining: a systematic review. Artificial intelligence review, 54(7), 4873.

Kratochvíl M, et al. (2020) GigaSOM.jl: High-performance clustering and visualization of huge cytometry datasets. GigaScience, 9(11).

Bobbili DR, et al. (2020) Excess of singleton loss-of-function variants in Parkinson's disease contributes to genetic risk. Journal of medical genetics, 57(9), 617.

Recanatini M, et al. (2020) Drug Research Meets Network Science: Where Are We? Journal of medicinal chemistry, 63(16), 8653.

Mih N, et al. (2020) Adaptations of Escherichia coli strains to oxidative stress are reflected in properties of their structural proteomes. BMC bioinformatics, 21(1), 162.

Lin AY, et al. (2020) CTO: a Community-Based Clinical Trial Ontology and its Applications in PubChemRDF and SCAIView. CEUR workshop proceedings, 2807.