

Resource Summary Report

Generated by ASWG on Aug 3, 2026

Apache Hadoop

RRID:SCR_011879

Type: Tool

Proper Citation

Apache Hadoop (RRID:SCR_011879)

Resource Information

URL: <http://hadoop.apache.org/>

Proper Citation: Apache Hadoop (RRID:SCR_011879)

Description: Software library providing a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures. The project includes these modules: * Hadoop Common: The common utilities that support the other Hadoop modules. * Hadoop Distributed File System (HDFS): A distributed file system that provides high-throughput access to application data. * Hadoop YARN: A framework for job scheduling and cluster resource management. * Hadoop MapReduce: A YARN-based system for parallel processing of large data sets.

Abbreviations: Hadoop

Resource Type: software toolkit, software resource

Keywords: computing

Availability: Open unspecified license

Resource Name: Apache Hadoop

Resource ID: SCR_011879

Alternate IDs: OMICS_01210

Record Creation Time: 20220129T080307+0000

Record Last Update: 20260803T040604+0000

Ratings and Alerts

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

No alerts have been found for Apache Hadoop.

Data and Source Information

Source: [SciCrunch Registry](#)

Usage and Citation Metrics

We found 57 mentions in open access literature.

Listed below are recent publications. The full list is available at [ASWG](#) .

Han X, et al. (2024) When we talk about Big Data, What do we really mean? Toward a more precise definition of Big Data. *Frontiers in big data*, 7, 1441869.

Prasanna S, et al. (2024) A scalable tool for analyzing genomic variants of humans using knowledge graphs and graph machine learning. *Frontiers in big data*, 7, 1466391.

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.

Milkman KL, et al. (2022) A 680,000-person megastudy of nudges to encourage vaccination in pharmacies. *Proceedings of the National Academy of Sciences of the United States of America*, 119(6).

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

Koppad S, et al. (2021) Cloud Computing Enabled Big Multi-Omics Data Analytics. *Bioinformatics and biology insights*, 15, 11779322211035921.

Tantry TP, et al. (2021) Self-learning software tools for data analysis in meta-analysis. *Korean journal of anesthesiology*, 74(5), 459.

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

Matalonga L, et al. (2021) Solving patients with rare diseases through programmatic reanalysis of genome-phenome data. *European journal of human genetics : EJHG*, 29(9), 1337.

Thulin A, et al. (2021) Discordance of PIK3CA and TP53 mutations between breast cancer brain metastases and matched primary tumors. *Scientific reports*, 11(1), 23548.

Maddocks S, et al. (2020) Burden of caring for children living with human immunodeficiency virus in a semi-rural South African community. *South African family*

practice : official journal of the South African Academy of Family Practice/Primary Care, 62(1), e1.

Brito LF, et al. (2020) Large-Scale Phenotyping of Livestock Welfare in Commercial Production Systems: A New Frontier in Animal Breeding. *Frontiers in genetics*, 11, 793.

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

Kapil G, et al. (2020) Attribute based honey encryption algorithm for securing big data: Hadoop distributed file system perspective. *PeerJ. Computer science*, 6, e259.

Prathapan S, et al. (2020) The rising complexity and burden of multimorbidity in a middle-income country. *PloS one*, 15(12), e0243614.

Koltes JE, et al. (2019) A Vision for Development and Utilization of High-Throughput Phenotyping and Big Data Analytics in Livestock. *Frontiers in genetics*, 10, 1197.

Barreto ML, et al. (2019) The Centre for Data and Knowledge Integration for Health (CIDACS): Linking Health and Social Data in Brazil. *International journal of population data science*, 4(2), 1140.

Tampuu A, et al. (2019) ViraMiner: Deep learning on raw DNA sequences for identifying viral genomes in human samples. *PloS one*, 14(9), e0222271.

Ciritoglu HE, et al. (2019) HaRD: a heterogeneity-aware replica deletion for HDFS. *Journal of big data*, 6(1), 94.

Rotenberg DJ, et al. (2018) The CAMH Neuroinformatics Platform: A Hospital-Focused Brain-CODE Implementation. *Frontiers in neuroinformatics*, 12, 77.