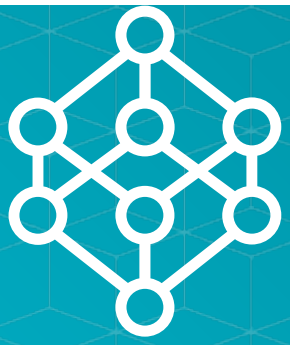

MAHOUT

THE APACHE SOFTWARE FOUNDATION

For Creating Scalable Performant Machine Learning Applications

Apache Mahout(TM) is a **distributed linear algebra framework** and **mathematically expressive Scala DSL** designed to let mathematicians, statisticians, and data scientists quickly *implement their own algorithms*. Apache Spark is the recommended out-of-the-box distributed back-end, *or can be extended to other distributed backends.*

- Mathematically Expressive Scala DSL
- Support for Multiple Distributed Backends (including Apache Spark)
- Modular Native Solvers for CPU/GPU/CUDA Acceleration

Apache Mahout User's Guide

Apache Mahout is a powerful, scalable, and versatile machine learning library designed for distributed data processing. It offers a comprehensive set of algorithms for various tasks, including classification, clustering, recommendation, and pattern mining. Built on top of the Apache Hadoop ecosystem, Mahout leverages MapReduce and Spark to enable data processing on large-scale datasets.

In this User's Guide, we provide an overview of Apache Mahout, its key features, and how to get started with using the library for your machine learning projects.

Key Features

- **Scalability**: Apache Mahout is designed to handle large-scale data processing by leveraging the power of Hadoop and Spark, making it an excellent choice for big data machine learning projects.
- **Versatility**: Mahout offers a wide range of machine learning algorithms, covering classification, clustering, recommendation, and more, ensuring that you have the right tools for your specific use case.
- **Extensibility**: The library is easily extensible, allowing you to add custom algorithms and processing steps to meet your unique requirements.
- **Integration**: Mahout seamlessly integrates with other components of the Hadoop ecosystem, such as HDFS and HBase, simplifying data storage and retrieval in your projects.