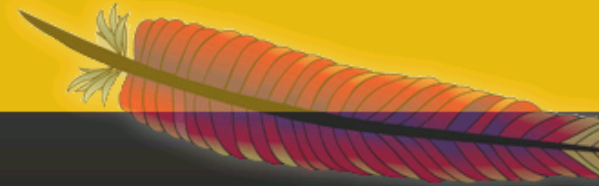


Mahout Primer

Apache 2.0 licensed



Apache Mahout is distributed under a commercially friendly Apache Software license

What is Apache Mahout?

The Apache Mahout™ machine learning library's goal is to build scalable machine learning libraries.

With scalable we mean:

Scalable to reasonably large data sets. Our core algorithms for clustering, classification and batch based collaborative filtering are implemented on top of Apache Hadoop using the map/reduce paradigm. However we do not restrict contributions to Hadoop based implementations: Contributions that run on a single node or on a non-Hadoop cluster are welcome as well. The core libraries are highly optimized to allow for good performance also for non-distributed algorithms

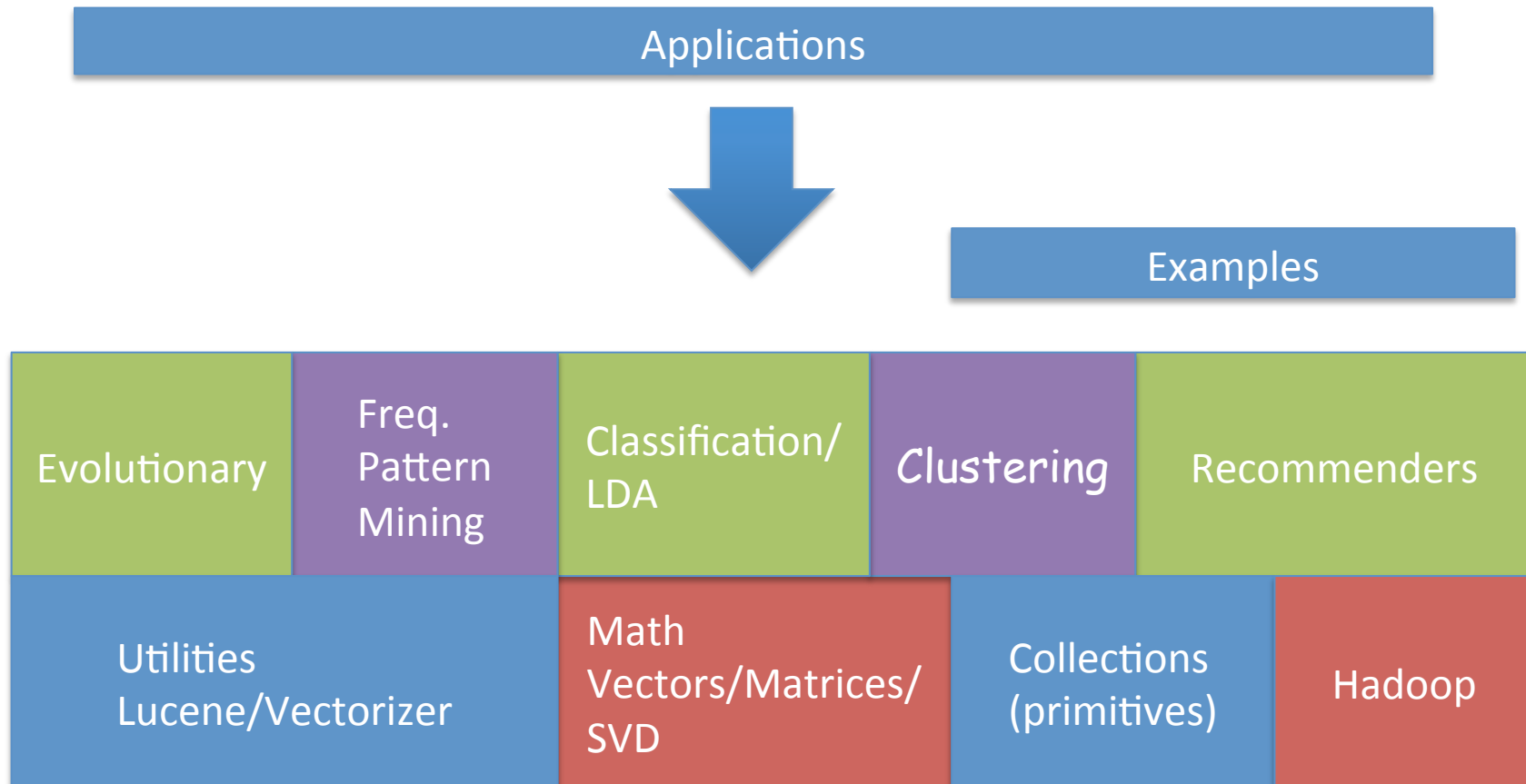
Scalable to support your business case. Mahout is distributed under a commercially friendly Apache Software license.

Scalable community. The goal of Mahout is to build a vibrant, responsive, diverse community to facilitate discussions not only on the project itself but also on potential use cases. Come to the mailing lists to find out more.

Mahout currently has

- Collaborative Filtering
- User and Item based recommenders
- K-Means, Fuzzy K-Means clustering
- Mean Shift clustering
- Dirichlet process clustering
- Latent Dirichlet Allocation
- Singular value decomposition
- Parallel Frequent Pattern mining
- Complementary Naive Bayes classifier
- Random forest decision tree based classifier
- High performance java collections (previously colt collections)
- A vibrant community
- and many more cool stuff to come by this summer thanks to Google summer of code

Mahout



Focus

- Scalable Machine Learning Platform
- Both to multi core, and on clusters
- Support for multiple data sources
- Utilities and data structures for performant features extraction
- Evaluation tools