

Apache Mahout 0.10.0 Release Notes

The Apache Mahout PMC is pleased to announce the release of Mahout 0.10.0. Mahout's goal is to create an environment for quickly creating machine learning applications that scale and run on the highest performance parallel computation engines available. Mahout comprises an interactive environment and library that supports generalized scalable linear algebra and includes many modern machine learning algorithms. This release has some major changes from 0.9, including the new Apache Spark back-end (with H2O in progress), a new matrix math DSL, streamlined content and bug fixes.

The Mahout Math environment we call “**Samsara**” for its symbol of universal renewal. It reflects a fundamental rethinking of how scalable machine learning algorithms are built and customized. Mahout-Samsara is here to help people create their own math while providing some off-the-shelf algorithm implementations. At its base are general linear algebra and statistical operations along with the data structures to support them. It's written in Scala with Mahout-specific extensions, and runs most fully on Spark.

To get started with Apache Mahout 0.10.0, download the release artifacts and signatures from <http://www.apache.org/dist/mahout/0.10.0/>

Many thanks to the contributors and committers who were part of this release. Please see below for the Release Highlights.

RELEASE HIGHLIGHTS

Mahout-Samsara has implementations for these generalized concepts:

- Linear algebra operations, multiply, transpose, slice, row and column iterators
- Distributed BLAS optimizer
- R-Like operators; for example $A.t \%* \% A$, which performs an optimized ‘thin’ $A'A$
- Packaged as extensions to Scala
- Includes a Scala REPL based interactive shell that runs on Spark
- Integrates with compatible libraries like MLLib

Mahout has historically been about highly scalable algorithms, and though we continue to support many of the past Hadoop MapReduce implementations (now with full Hadoop 2 support), Mahout also comes with the some new Mahout-Samsara based implementations:

- Distributed and in-core Stochastic Singular Value Decomposition (SSVD)
- Distributed Principal Component Analysis (PCA)
- Distributed and in-core QR Reduction (QR)
- Distributed Alternating Least Squares (ALS)
- Collaborative Filtering: Item and Row Similarity based on cooccurrence and supporting multimodal user actions
- Naive Bayes Classification

RELATION TO MACHINE LEARNING LIBS

Since Mahout is positioned as an environment it also allows seamless use of libraries like Mlib. If you need scalable linear algebra, think Mahout, if you need a specific algorithm check any compatible library as well.

STATS

A total of 205 separate JIRA issues are addressed in this release [2]. with 65 bugfixes.

GETTING STARTED

Download the release artifacts and signatures at <https://mahout.apache.org/general/downloads.html> The examples directory contains several working examples of the core functionality available in Mahout. These can be run via scripts in the examples/bin directory. Most examples do not need a Hadoop cluster in order to run.

FUTURE PLANS

0.10.1

As the project moves towards a 0.10.1 release, we are working on the following:

- Implement an end-to-end pipeline for an item similarity recommender workflow on top of H2O.
- Implement a more robust text processing pipeline
- Incorporate more statistical operations
- Support Spark DataFrames

Post 0.10.1

We already see the need for work in these areas:

- Mahout algebra performance improvements and bug fixes
- Streaming data
- Visualization
- Fuller H2O support
- Apache Flink support
- In-core matrix performance optimization

CONTRIBUTING

If you are interested in contributing, please see our [How to Contribute](#)[3] page or contact us via email at dev@mahout.apache.org.

CREDITS

As with any release, we wish to thank all of the users and contributors to Mahout. Please see the CHANGELOG [1] and JIRA Release Notes [2] for individual credits, as there are too many to list here.

[1] <https://github.com/apache/mahout/blob/master/CHANGELOG>

[2]

[https://issues.apache.org/jira/browse/MAHOUT-1678?jql=project%20%3D%20MAHOUT%20AND%20status%20in%20\(Resolved%2C%20Closed\)%20AND%20fixVersion%20%3D%2010.0](https://issues.apache.org/jira/browse/MAHOUT-1678?jql=project%20%3D%20MAHOUT%20AND%20status%20in%20(Resolved%2C%20Closed)%20AND%20fixVersion%20%3D%2010.0)

[3] <http://mahout.apache.org/developers/how-to-contribute.html>