

## Experiment - 9

### Lab Manual

**Aim : "Leveraging machine learning using Mahout like tools."**

Machine learning is trending for multiple reasons including vast usage of the World Wide Web and services using smart devices. BigData and machine learning are connected by requirement that for machine learning underlying data source can itself be Big Data, which in turn distributed by nature.

Mahout is part of hadoop ecosystem to create scalable performant machine learning applications. Mahout is a distributed linear algebra framework for data scientists and others to quickly implement their own algorithms as quoted by the site.

As of 0.13, Mahout achieves following algorithms:

- Distributed Linear Algebra
- Preprocessors
- Regression
- Clustering
- Recommenders

**Optionally**, you may download hadoop source and utilize or contribute on your personal computer.

Mahout source uses apache maven as Software Project Management and comprehension tool.

```
root@aharnish: /opt/mahout
File Edit View Search Terminal Help
root@aharnish:/opt/mahout# ls
bin          conf          doap_Mahout.rdf  experimental  pom.xml
buildtools   core          engine           LICENSE.txt   README.md
community    distribution  examples         NOTICE.txt   website
root@aharnish:/opt/mahout#
```

Mahout will build the following packages:

Apache Mahout

Mahout Core

Mahout Engine

Mahout HDFS Support

Mahout Community

Mahout Spark CLI Drivers

## Apache Maven

pom.xml file is the configuration file located in the project folder. It contains information about the dependencies, etc. It is a smart tool which does further dependency checks based on entries into pom.xml and fetches live from the web. Make sure to have connected to net if you are expecting live fetch.

The appropriate version number of module can be identified quickly from mvn repository site.  
i.e. for mahout-core package

Mahout java library can be used within a project after adding dependencies into pom.xml

```
<dependency>
    <groupId>org.apache.mahout</groupId>
```

```

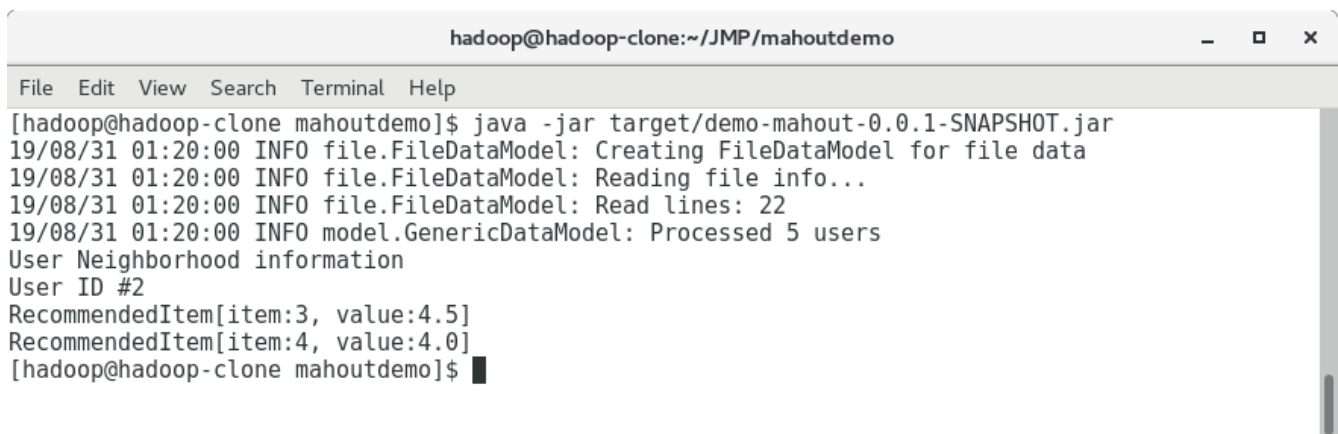
        <artifactId>mahout-core</artifactId>
        <version>0.9</version>
    </dependency>

    <dependency>
        <groupId>org.apache.mahout</groupId>
        <artifactId>mahout-math</artifactId>
        <version>0.13.0</version>
    </dependency>

    <dependency>
        <groupId>org.apache.mahout</groupId>
        <artifactId>mahout-integration</artifactId>
        <version>0.13.0</version>
    </dependency>

```

The demo application as published on [TutorialsPoint](#) website can be downloaded and tested after project structure and mvn commands.



```

hadoop@hadoop-clone: ~/JMP/mahoutdemo
File Edit View Search Terminal Help
[hadoop@hadoop-clone mahoutdemo]$ java -jar target/demo-mahout-0.0.1-SNAPSHOT.jar
19/08/31 01:20:00 INFO file.FileDataModel: Creating FileDataModel for file data
19/08/31 01:20:00 INFO file.FileDataModel: Reading file info...
19/08/31 01:20:00 INFO file.FileDataModel: Read lines: 22
19/08/31 01:20:00 INFO model.GenericDataModel: Processed 5 users
User Neighborhood information
User ID #2
RecommendedItem[item:3, value:4.5]
RecommendedItem[item:4, value:4.0]
[hadoop@hadoop-clone mahoutdemo]$

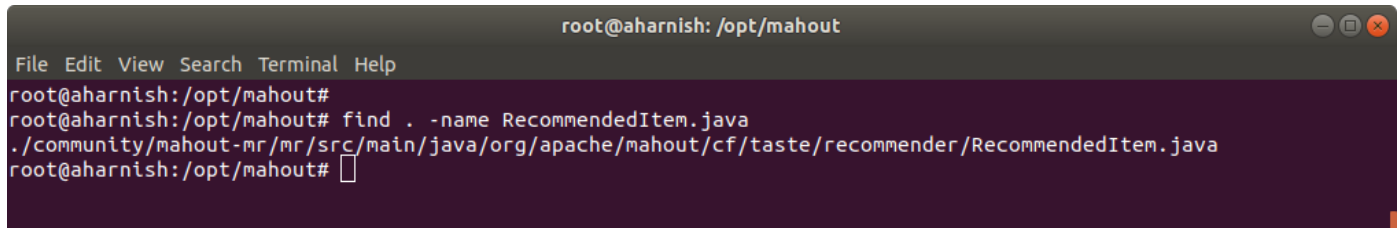
```

May explore web based documentation and java source code from mahout downloads:

<http://mahout.apache.org/docs/0.13.0/api/docs/>

<http://mahout.apache.org/docs/0.13.0/api/docs/mahout-mr/org/apache/mahout/cf/taste/recommender/RecommendedItem.html>

<http://mahout.apache.org/docs/0.13.0/api/docs/mahout-mr/>

A terminal window titled 'root@aharnish: /opt/mahout' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal shows the following commands and output:

```
root@aharnish:/opt/mahout#  
root@aharnish:/opt/mahout# find . -name RecommendedItem.java  
./community/mahout-mr/mr/src/main/java/org/apache/mahout/cf/taste/recommender/RecommendedItem.java  
root@aharnish:/opt/mahout#
```

## References :

1. <https://mahout.apache.org/>
2. <https://mahout.apache.org/users/basics/algorithms.html>
3. <https://mahout.apache.org/general/downloads>
4. <https://maven.apache.org/>
5. <https://mvnrepository.com/artifact/org.apache.mahout/mahout-core>
6. [https://www.tutorialspoint.com/mahout/mahout\\_environment.htm](https://www.tutorialspoint.com/mahout/mahout_environment.htm)
7. <https://mahout.apache.org/docs/0.13.0/api/docs/>
8. <http://mahout.apache.org/developers/buildingmahout>

## Exercise:

1. Run recommendation of a sample dataset using Mahout Library.

### **Setup Commands references:**

apt install git

apt install maven

git clone <https://github.com/apache/mahout.git> mahout

cd /opt/mahout; find . -name RecommendedItem.java