

Anarion Technologies – Apache Mahout

Note: Search product on Azure marketplace and click on "Get it now"

 Microsoft | Marketplace

Products ▾ 🔍 Search Marketplace

≡ All Products ▾ Categories ▾ Industries ▾ Partners ▾

Products > Apache Mahout on Ubuntu26



Apache Mahout on Ubuntu26

by Anarion Technologies
[Virtual Machines](#)

Starting at \$0.03 [Get it now](#) ☆ Save to my list

[Overview](#) [Plans + Pricing](#) [Ratings + reviews](#) [Details + support](#)

Production Ready VM on Ubuntu26.04 LTS + Free Support

Apache Mahout is a scalable machine learning library designed to build distributed data mining and analytics applications. Powered by Apache Spark, Mahout provides efficient implementations of machine learning algorithms for clustering, classification, collaborative filtering, recommendation systems, and similarity analysis. It enables organizations and developers to process large-scale datasets while leveraging distributed computing for high performance and scalability.

Apache Mahout offers a flexible mathematical environment and distributed linear algebra framework that simplifies the development of machine learning workflows. It supports advanced analytics tasks such as item similarity, row similarity, matrix factorization, and recommendation engines. Mahout integrates seamlessly with Apache Spark, making it well suited for big data analytics, enterprise data processing, research, and production machine learning workloads.

One of Apache Mahout's key strengths is its scalable distributed execution model, allowing machine learning algorithms

Click on Continue

Confirm your details to continue



Apache Mahout on Ubuntu26

By Anarion Technologies

Plan

Apache Mahout on Ubuntu26



Pricing

Starting at \$0.03/hour

Description

Apache Mahout on Ubuntu26

You're signed in as **Market Place Products** (mp@anariontech.com).

[Edit your details](#)

 By clicking **Get it now**, I give Microsoft permission to use or share my account information so that the provider can contact me regarding this product. I agree to the provider's [terms of use](#) and [privacy policy](#) and understand that the rights to use this product do not come from Microsoft, unless Microsoft is the provider. Use of Marketplace is governed by separate [terms](#) and [privacy](#).

Get it now

Click on Create

Home

Apache Mahout on Ubuntu26

Anarion Technologies



Apache Mahout on Ubuntu26

 Add to Favorites

Anarion Technologies | Virtual Machine

Subscription

Azure subscription 1

Plan

Apache Mahout on Ubuntu26

Create

Start with a pre-set configuration

Want to deploy programmatically? [Get started](#)

[Overview](#) [Plans + Pricing](#) [Usage Information + Support](#) [Ratings + Reviews](#)

Apache Mahout is a scalable machine learning library designed to build distributed data mining and analytics applications. Powered by Apache Spark, Mahout provides efficient implementations of machine learning algorithms for clustering, classification, collaborative filtering, recommendation systems, and similarity analysis. It enables organizations and developers to process large-scale datasets while leveraging distributed computing for high performance and scalability.

Apache Mahout offers a flexible mathematical environment and distributed linear algebra framework that simplifies the development of machine learning workflows. It supports advanced analytics tasks such as item similarity, row similarity, matrix factorization, and recommendation engines. Mahout integrates seamlessly with Apache Spark, making it well suited for big data analytics, enterprise data processing, research, and production machine learning workloads.

One of Apache Mahout's key strengths is its scalable distributed execution model, allowing machine learning algorithms to efficiently process massive datasets across multiple compute nodes. Its Spark-based architecture provides improved performance, fault tolerance, and the ability to handle complex analytical operations with minimal code changes. Mahout also includes command-line tools and APIs that simplify deployment and experimentation with distributed machine learning algorithms.

This virtual machine image comes pre-configured with Apache Mahout 14.1, Apache Spark 3.5.6, Apache Maven 3.8.8, OpenJDK 17, and Ubuntu 26.04 LTS, providing a secure and production-ready environment. The image is optimized for recommendation systems, predictive analytics, similarity computation, clustering, distributed machine learning,

Creating a virtual machine, enter or select appropriate values for zone, machine type, resource group and so on as per your choice.

Create a virtual machine

[Basics](#) [Disks](#) [Networking](#) [Management](#) [Advanced](#) [Tags](#) [Review + create](#)

Create a virtual machine that runs Linux or Windows. Select an image from Azure marketplace or use your own customized image. Complete the Basics tab then Review + create to provision a virtual machine with default parameters or review each tab for full customization. [Learn more](#)

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription *

Azure subscription 1

Resource group *

Demo

[Create new](#)

Instance details

Virtual machine name *

Demo

Region *

(US) East US

Review + create

< Previous

Next : Disks >

Create a virtual machine ...

Basics Disks Networking Management Advanced Tags Review + create

Azure VMs have one operating system disk and a temporary disk for short-term storage. You can attach additional data disks. The size of the VM determines the type of storage you can use and the number of data disks allowed. [Learn more](#)

Disk options

OS disk type * ⓘ Premium SSD (locally-redundant storage) ▼

Encryption type * (Default) Encryption at-rest with a platform-managed key ▼

Enable Ultra Disk compatibility ⓘ ☐
Ultra disk is available only for Availability Zones in eastus.

Data disks

You can add and configure additional data disks for your virtual machine or attach existing disks. This VM also comes with a temporary disk.

i Adding unmanaged data disks is currently not supported at the time of VM creation. You can add them after the VM is

Review + create

< Previous

Next : Networking >

Create a virtual machine ...

Basics Disks Networking Management Advanced Tags Review + create

Define network connectivity for your virtual machine by configuring network interface card (NIC) settings. You can control ports, inbound and outbound connectivity with security group rules, or place behind an existing load balancing solution. [Learn more](#)

Network interface

When creating a virtual machine, a network interface will be created for you.

Virtual network * ⓘ (new) Demo-vnet ▼
[Create new](#)

Subnet * ⓘ (new) default (10.1.0.0/24) ▼

Public IP ⓘ (new) Demo-ip ▼
[Create new](#)

NIC network security group ⓘ
☐ None
☒ Basic
☐ Advanced

Review + create

< Previous

Next : Management >

Create a virtual machine ...

Basics Disks Networking Management Advanced Tags Review + create

Configure monitoring and management options for your VM.

Azure Security Center

Azure Security Center provides unified security management and advanced threat protection across hybrid cloud workloads. [Learn more](#)

✓ Your subscription is protected by Azure Security Center basic plan.

Monitoring

Boot diagnostics ⓘ ☐ On
☒ Off

Enable OS guest diagnostics ⓘ ☐

Identity

[Review + create](#)

[< Previous](#)

[Next : Advanced >](#)

Create a virtual machine ...

Basics Disks Networking Management Advanced Tags Review + create

Tags are name/value pairs that enable you to categorize resources and view consolidated billing by applying the same tag to multiple resources and resource groups. [Learn more about tags](#)

Note that if you create tags and then change resource settings on other tabs, your tags will be automatically updated.

Name ⓘ	Value ⓘ	Resource
		12 selected v

[Review + create](#)

[< Previous](#)

[Next : Review + create >](#)

[Home](#) >

Create a virtual machine

✓ Validation passed

Basics Disks Networking Management Monitoring Advanced Tags **Review + create**

i Cost given below is an estimate and not the final price. For all your pricing needs, please use the pricing calculator. [↗](#)

Price

1 X Standard B1s
by Microsoft
[Terms of use](#) | [Privacy policy](#)

Subscription credits apply ⓘ

0.8652 INR/hr
[Pricing for other VM sizes](#)

TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the [Azure Marketplace Terms](#) for additional details.

[< Previous](#)

[Next >](#)

Create

After Process of Create Virtual Machine. You have got an Option Go to Resource Group

Click Go to Resource Group

Copy the Public IP Address

SSH into Terminal and Run these Commands:

```
$ sudo su
```

```
$ cd ../../
```

```
$ sudo apt update && sudo apt upgrade -y
```

```

testing@testing:~$ sudo su
root@testing:/home/testing# cd ../../
root@testing:/# sudo apt update && sudo apt upgrade -y
Hit:1 http://azure.archive.ubuntu.com/ubuntu resolute InRelease
Get:2 http://azure.archive.ubuntu.com/ubuntu resolute-updates InRelease [137
kB]
Hit:3 http://azure.archive.ubuntu.com/ubuntu resolute-backports InRelease
Get:4 http://azure.archive.ubuntu.com/ubuntu resolute-security InRelease [13
7 kB]
Get:5 https://aquasecurity.github.io/trivy-repo/deb resolute InRelease [3238
B]
Get:6 http://azure.archive.ubuntu.com/ubuntu resolute-updates/main amd64v3 P
ackages [282 kB]
Get:7 http://azure.archive.ubuntu.com/ubuntu resolute-updates/main Translati
on-en [74.0 kB]
Get:8 http://azure.archive.ubuntu.com/ubuntu resolute-updates/main amd64 Com
ponents [49.4 kB]
Get:9 http://azure.archive.ubuntu.com/ubuntu resolute-updates/main amd64 c-n
-f Metadata [5244 B]
Get:10 http://azure.archive.ubuntu.com/ubuntu resolute-updates/universe amd6
4v3 Packages [159 kB]
Get:11 http://azure.archive.ubuntu.com/ubuntu resolute-updates/universe Tran
slation-en [50.1 kB]
Get:12 http://azure.archive.ubuntu.com/ubuntu resolute-updates/universe amd6

```

Verify Java

\$ java -version

```

root@testing:/# java -version
openjdk version "17.0.19" 2026-04-21
OpenJDK Runtime Environment (build 17.0.19+10-1-26.04.2-Ubuntu)
OpenJDK 64-Bit Server VM (build 17.0.19+10-1-26.04.2-Ubuntu, mixed mode, sha
ring)

```

Verify Apache Spark:

\$ spark-submit --version

```
root@testing:/# spark-submit --version
Welcome to
```

[illegible]

```
Using Scala version 2.12.18, OpenJDK 64-Bit Server VM, 17.0.19
Branch HEAD
Compiled by user runner on 2025-05-23T06:34:26Z
Revision 303c18c74664f161b9b969ac343784c088b47593
Url https://github.com/apache/spark
Type --help for more information.
root@testing:/#
```

Verify Mahout Libraries:

```
$ ls -lh /opt/mahout/lib
```

```
root@testing:/# ls -lh /opt/mahout/lib
total 24M
-rw-r--r-- 1 root root 2.2M Jun 29 12:50 mahout-core_2.12-14.1.jar
-rw-r--r-- 1 root root 24K Jun 29 12:50 mahout-hdfs_2.12-14.1.jar
-rw-r--r-- 1 root root 59K Jun 29 12:51 mahout-spark-cli-drivers_2.12-14.1.jar
-rw-r--r-- 1 root root 21M Jun 29 12:51 mahout-spark_2.12-14.1-dependency-reduced.jar
-rw-r--r-- 1 root root 243K Jun 29 12:51 mahout-spark_2.12-14.1.jar
-rw-r--r-- 1 root root 77K Jun 29 13:02 scopt_2.12-3.7.1.jar
root@testing:/#
```

```
$ cd /opt/mahout
```

```
$ ./bin/mahout spark-rowsimilarity --help
```

```

root@testing:/# cd /opt/mahout
root@testing:/opt/mahout# ./bin/mahout spark-rowsimilarity --help
Adding lib/ to CLASSPATH

spark-rowsimilarity Mahout
Usage: spark-rowsimilarity [options]

Input, output options
  -i, --input <value>      Input path, may be a filename, directory name, or comma delimited list of HDFS supported URIs (required)
  -o, --output <value>     Path for output directory, any HDFS supported URI (required)

Algorithm control options:
  -mo, --maxObservations <value>
                             Max number of observations to consider per row (optional). Default: 500
  -m, --maxSimilaritiesPerRow <value>
                             Limit the number of similarities per item to this number (optional). Default: 100

Note: Only the Log Likelihood Ratio (LLR) is supported as a similarity measure.

Input and Output text file schema options (same for both):
  -rd, --rowKeyDelim <value>
                             Separates the rowID key from the vector values list (optional). Default: "\t"

```

Run a Spark Row Similarity job:

**`./bin/mahout spark-rowsimilarity \ -i <input_file> \ -o
<output_directory> \ -ma local`**

Replace: <input_file> with your input dataset.

<output_directory> with the directory where you want to save the output.

```
root@testing:/opt/mahout# ./bin/mahout spark-rowsimilarity \
-i /tmp/mahout-demo/input.txt \
-o /tmp/mahout-demo/output \
-ma local
Adding lib/ to CLASSPATH
Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties
26/07/10 12:54:32 WARN SparkConf: The configuration key 'spark.kryoserializer.buffer.mb' has been deprecated as of Spark 1.4 and may be removed in the future. Please use spark.kryoserializer.buffer instead. The default value for spark.kryoserializer.buffer.mb was previously specified as '0.064'. Fractional values are no longer accepted. To specify the equivalent now, one may use '64k'.
Exception in thread "main" java.lang.IllegalArgumentException: requirement failed: MAHOUT_HOME is required to spawn mahout-based spark jobs
    at scala.Predef$.require(Predef.scala:281)
    at org.apache.mahout.sparkbindings.package$.getMahoutHome(package.scala:214)
    at org.apache.mahout.sparkbindings.package$.findMahoutContextJars(package.scala:222)
    at org.apache.mahout.sparkbindings.package$.mahoutSparkContext(package.scala:77)
    at org.apache.mahout.drivers.MahoutSparkDriver.start(MahoutSparkDriver.scala:87)
    at org.apache.mahout.drivers.RowSimilarityDriver$.start(RowSimilarityDriver.scala:109)
    at org.apache.mahout.drivers.RowSimilarityDriver$.process(RowSimilarityDriver.scala:134)
    at org.apache.mahout.drivers.RowSimilarityDriver$.anonfun$main$1(RowSimilarityDriver.scala:103)
    at org.apache.mahout.drivers.RowSimilarityDriver$.anonfun$main$1$adapted(RowSimilarityDriver.scala:101)
    at scala.Option.map(Option.scala:230)
    at org.apache.mahout.drivers.RowSimilarityDriver$.main(RowSimilarityDriver.scala:101)
    at org.apache.mahout.drivers.RowSimilarityDriver.main(RowSimilarityDriver.scala)
root@testing:/opt/mahout# mkdir -p /tmp/mahout-demo
root@testing:/opt/mahout# cat > /tmp/mahout-demo/input.txt <<EOF
user1 item1:1 item2:1 item3:1
user2 item1:1 item2:1
user3 item2:1 item3:1
user4 item1:1 item3:1
EOF
root@testing:/opt/mahout# rm -rf /tmp/mahout-demo/output
root@testing:/opt/mahout# export MAHOUT_HOME=/opt/mahout
root@testing:/opt/mahout# ./bin/mahout spark-rowsimilarity \
-i /tmp/mahout-demo/input.txt \
-o /tmp/mahout-demo/output \
-ma local
Adding lib/ to CLASSPATH
Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties
26/07/10 12:55:28 WARN SparkConf: The configuration key 'spark.kryoserializer.buffer.mb' has been deprecated as of Spark 1.4 and may be removed in the future. Please use spark.kryoserializer.buffer instead. The default value for spark.kryoserializer.buffer.mb was previously specified as '0.064'. Fractional values are no longer accepted. To specify the equivalent now, one may use '64k'.
```

Thanks!!!