





INSTALLING HADOOP

Go to the official website of oracle and download JDK (Java Development Kit) version 8

ORACLE



 View Accounts

 Contact Sales

Products Industries Resources Customers Partners Developers Company

Java / Technologies / JavaSE /

Java SE Downloads

Java SE Subscriptions

Java SE 8 Archive Downloads (JDK 8u202 and earlier)

Go to the [Oracle Java Archive](#)

The JDK is a development environment for building applications using the Java programming language.

The JDK includes tools useful for developing and testing programs written in the Java programming language and running on the Java™ platform.

WARNING: These older versions of the JRE and JDK are provided to help developers debug issues in older systems. **They are not updated with the latest security patches and are not recommended for use in production.**

These Java SE 8 update releases are provided under the [Binary Code License](#) ("BCL").






Java SE 8u211 and later updates are available, under the [Java SE OTN License](#).

For production use Oracle recommends downloading the latest JDK and JRE versions and allowing auto-update.

Only developers and Enterprise administrators should download these releases.

Downloading these releases requires an oracle.com account. If you don't have an oracle.com account you can use the links on the top of this page to learn more about it and register for one for free.

For current Java releases, please consult the [Oracle Software Download page](#).

Solaris SPARC 64-bit	88.1 MB	 jdk-8u202-solaris-sparcv9.tar.gz
Solaris x64 (SVR4 package)	124.37 MB	 jdk-8u202-solaris-x64.tar.Z
Solaris x64	85.38 MB	 jdk-8u202-solaris-x64.tar.gz
Windows x86	201.64 MB	 jdk-8u202-windows-i586.exe
Windows x64	211.58 MB	 jdk-8u202-windows-x64.exe

Java SE Runtime Environment 8u202

This software is licensed under the [Oracle Binary Code License Agreement for Java SE Platform Products](#)

Product / File Description	File Size	Download
----------------------------	-----------	----------

Solaris SPARC 64-bit88.1 MBjdk-8u202-solaris-sparcv9.tar.gz

Solaris x64 (SVR4)

Solaris x64

Windows x86

Windows x64

You must accept the [Oracle Binary Code License Agreement for the Java SE Platform Products](#) to download this software.

☒ I reviewed and accept the Oracle Binary Code License Agreement for the Java SE Platform Products

Required

Download jdk-8u202-windows-x64.exe

Java SE Runtime Environment 8u202

This software is licensed under the [Oracle Binary Code License Agreement for Java SE Platform Products](#)

Product / File Description	File Size	Download
----------------------------	-----------	----------

Login to your oracle account in order to complete the download process

Sign in to Oracle

Username or email

Next

[Forgot username?](#)

Don't have an Oracle Account?


Create Account

[© Oracle](#) | [Terms of Use](#) | [Privacy Policy](#)

Install JDK 8 using the downloaded file

 jdk-8u202-windows-x64	28-07-2024 22:30	Application	2,16,653 KB
---	------------------	-------------	-------------

Now, go to the official website of Apache Hadoop and download Hadoop (Preferred version: 3.3.6)

 Apache Hadoop

[Download](#) [Documentation](#) [Community](#) [Development](#) [Help](#)

Apache Software Foundation

Download

Hadoop is released as source code tarballs with corresponding binary tarballs for convenience. The downloads are distributed via mirror sites and should be checked for tampering using GPG or SHA-512.

Version	Release date	Source download	Binary download	Release notes
3.4.0	2024 Mar 17	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
3.3.6	2023 Jun 23	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
2.10.2	2022 May 31	source (checksum signature)	binary (checksum signature)	Announcement

To verify Apache Hadoop® releases using GPG:

- Download the release `hadoop-X.Y.Z-src.tar.gz` from a [mirror site](#).
- Download the signature file `hadoop-X.Y.Z-src.tar.gz.asc` from [Apache](#).
- Download the [Hadoop KEYS](#) file.
- `gpg --import KEYS`
- `gpg --verify hadoop-X.Y.Z-src.tar.gz.asc`

To perform a quick check using SHA-512:

- Download the release `hadoop-X.Y.Z-src.tar.gz` from a [mirror site](#).
- Download the checksum `hadoop-X.Y.Z-src.tar.gz.sha512` or `hadoop-X.Y.Z-src.tar.gz.mds` from [Apache](#).
- `shasum -a 512 hadoop-X.Y.Z-src.tar.gz`

After downloading, you will get .tar file. Use 7Zip or WinRAR to extract the files.

 hadoop-3.3.6.tar

28-07-2024 22:52

Compressed Archive ...

7,12,996 KB

Set environment variables for both JDK and Hadoop

Environment Variables

User variables for prath

Variable	Value
HADOOP_HOME	C:\hadoop-3.3.6
JAVA_HOME	C:\Program Files\Java\jdk1.8.0_202
OneDrive	C:\Users\prath\OneDrive
Path	C:\Program Files\MySQL\MySQL Shell 8.0\bin\;C:\Users\prath\A...
TEMP	C:\Users\prath\AppData\Local\Temp
TMP	C:\Users\prath\AppData\Local\Temp
UIPATH_LANGUAGE	en

New... Edit... Delete

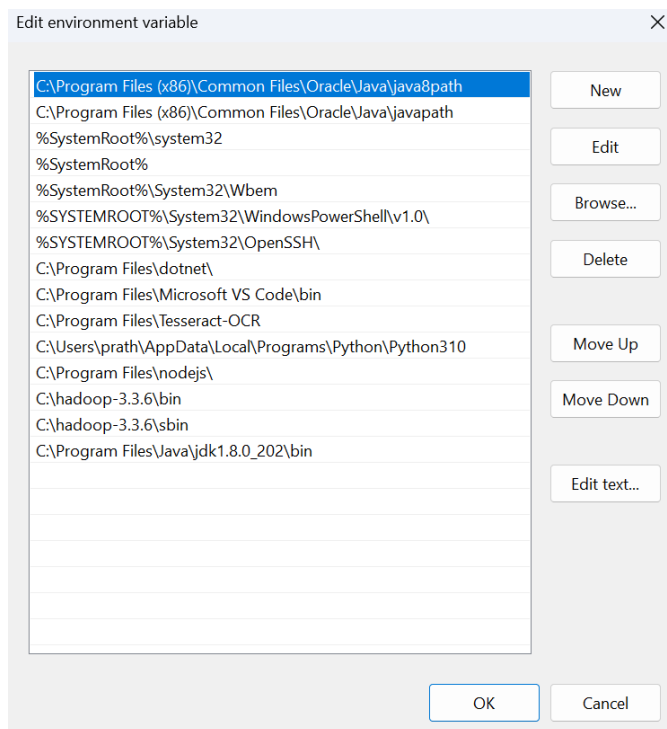
System variables

Variable	Value
_PSLockDownPolicy	0
ComSpec	C:\WINDOWS\system32\cmd.exe
DriverData	C:\Windows\System32\Drivers\DriverData
IGCCSVC_DB	AQAAANCMnd8BFdERjHoAwE/Cl+sBAAANpjf4fsnREyyg0soj3K...
NUMBER_OF_PROCESSORS	12
OS	Windows_NT
Path	C:\Program Files (x86)\Common Files\Oracle\Java\java8path;C:\P...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC

New... Edit... Delete

OK Cancel

Also, set the path in System variables under the variable “Path”.



Now, check the versions of both Hadoop and JDK in command prompt.

```
C:\Users\prath>hadoop version
Hadoop 3.3.6
Source code repository https://github.com/apache/hadoop.git -r 1be78238728da9266a4f88195058f08fd012bf9c
Compiled by ubuntu on 2023-06-18T08:22Z
Compiled on platform linux-x86_64
Compiled with protoc 3.7.1
From source with checksum 5652179ad55f76cb287d9c633bb53bbd
This command was run using /C:/hadoop-3.3.6/share/hadoop/common/hadoop-common-3.3.6.jar
```

```
C:\Users\prath>java -version
java version "1.8.0_421"
Java(TM) SE Runtime Environment (build 1.8.0_421-b09)
Java HotSpot(TM) Client VM (build 25.421-b09, mixed mode, sharing)
```

Move to the directory where you had extracted Hadoop.

Go to `hadoop-3.3.6/etc/hadoop` and include these lines of code in the following files.

core-site.xml

```
<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>fs.defaultFS</name>
<value>hdfs://localhost:9000</value>
</property>
</configuration>
```

Create a folder “data” in the main directory of Hadoop (hadoop-3.3.6/). Within that, create two subfolders named “namenode” and “datanode”.

hdfs-site.xml

```
<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<name>dfs.namenode.name.dir</name>
<value>C:\hadoop-3.3.6\data\namenode</value>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>C:\hadoop-3.3.6\data\datanode</value>
</property>
</configuration>
```

mapred-site.xml

```
<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

yarn-site.xml

```
<configuration>
  <!-- Configure NodeManager auxiliary services -->
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>

  <!-- Specify the class for the MapReduce shuffle service -->
  <property>
    <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.shuffleHandler</value>
  </property>
</configuration>
```

Go to hadoop-env.cmd file and set the path to your java directory

```
@rem The java implementation to use. Required.
set JAVA_HOME=C:\Progra~1\Java\jdk1.8.0_202
```

Open command prompt and type the command:

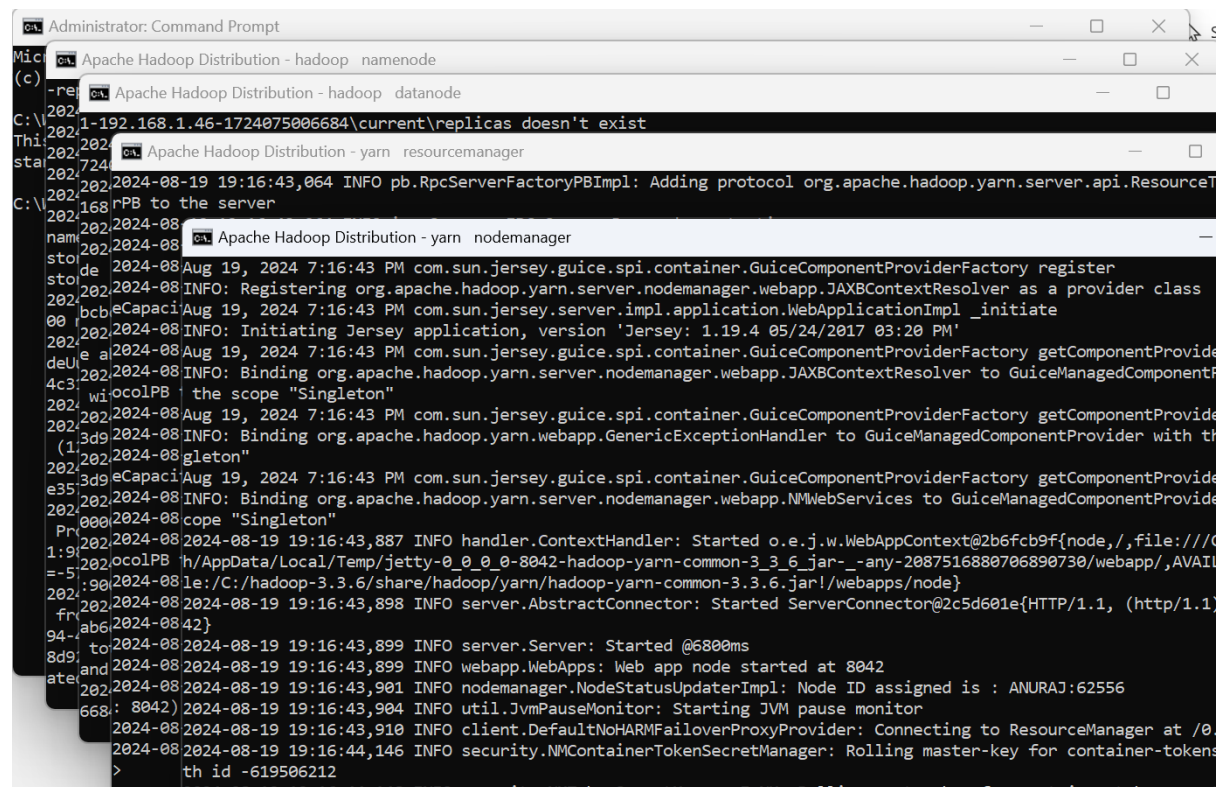
hdfs namenode -format

```
C:\Users\prath>hdfs namenode -format
2024-08-19 19:13:25,285 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
```

```
2024-08-19 19:13:26,652 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2024-08-19 19:13:26,652 INFO util.GSet: VM type = 64-bit
2024-08-19 19:13:26,652 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
2024-08-19 19:13:26,652 INFO util.GSet: capacity = 2^15 = 32768 entries
2024-08-19 19:13:26,684 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1561018181-192.168.1.46-1724075006684
2024-08-19 19:13:26,746 INFO common.Storage: Storage directory C:\hadoop-3.3.6\data\namenode has been successfully formatted.
```

Open command prompt as administrator and type the command:

start-all.cmd



```
Administrator: Command Prompt
Apache Hadoop Distribution - hadoop namenode
2024-08-19 19:16:43,064 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceT
Apache Hadoop Distribution - hadoop datanode
2024-08-19 19:16:43,064 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceT
Apache Hadoop Distribution - yarn resourcemanager
2024-08-19 19:16:43,064 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceT
Apache Hadoop Distribution - yarn nodemanager
2024-08-19 19:16:43,064 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceT
2024-08-19 19:16:43,887 INFO handler.ContextHandler: Started o.e.j.w.WebAppContext@2b6fcb9f{node/,file:///C:/hadoop-3.3.6/share/hadoop/yarn/hadoop-yarn-common-3.3.6_jar-any-2087516880706890730/webapp/,AVAILABLE}
2024-08-19 19:16:43,899 INFO webapp.WebApps: Web app node started at 8042
2024-08-19 19:16:43,901 INFO nodemanager.NodeStatusUpdaterImpl: Node ID assigned is : ANURAJ:62556
2024-08-19 19:16:43,904 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2024-08-19 19:16:43,910 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8042
2024-08-19 19:16:44,146 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens
2024-08-19 19:16:44,148 INFO security.NMTokenSecretManagerToNM: Rolling master-key for container-tokens
```

Type jps and check if all the Hadoop services are running.

```
C:\Windows\System32>jps
14256 NodeManager
21760 Jps
14100 DataNode
20552 ResourceManager
22584 NameNode
```

Go to localhost:9870

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Overview 'localhost:9000' (✓active)

Started:	Mon Aug 19 19:16:39 +0530 2024
Version:	3.3.6, r1be78238728da9266a4f88195058f08fd012bf9c
Compiled:	Sun Jun 18 13:52:00 +0530 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-84c31ef0-3f5d-4043-8d92-f6f88dc15d51
Block Pool ID:	BP-1561018181-192.168.1.46-1724075006684

Summary

Security is off.


Safemode is off.

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

Heap Memory used 108.58 MB of 339.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 51.2 MB of 53.48 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Go to localhost:8088

 **hadoop**

AI

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW_SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Use
0	0	0	0	0	<memory:0 B, vC

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes
1	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation
Capacity Scheduler	[memory-mb (unit=M), vcores]	<memory:1024, vCores:1>

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FI
Showing 0 to 0 of 0 entries											