

Apache Hadoop Installation Guide (Windows)

Installing Apache Hadoop

1. Go to the Apache Hadoop official website. Here is the link: <https://hadoop.apache.org/>
2. Click on the "**Downloads**" tab in the top menu.
3. Scroll down to the "**Stable Releases**" section and click on the "**Download**" link for the latest stable release of Apache Hadoop for Windows.
4. Save the Hadoop binary distribution ZIP file to your computer.
5. Extract the contents of the ZIP file to a directory of your choice. This will be your Hadoop installation directory.
6. Open a Command Prompt window with administrator privileges.
7. Navigate to the directory where you extracted the Hadoop ZIP file using the "cd" command. For example, if you extracted the file to a directory called "**hadoop**" in your C: drive, you can navigate to that directory with the following command: `cd C:\hadoop`.
8. Inside the Hadoop installation directory, go to the "**etc**" subdirectory and locate the "**hadoop-env.cmd**" file. Make a copy of this file and rename it to "**hadoop-env.cmd.bak**" for backup purposes.
9. Open the "**hadoop-env.cmd**" file with a text editor.
10. Find the line that sets the "JAVA_HOME" environment variable and set its value to the path of your Java JDK installation directory. For example: `set JAVA_HOME=C:\Program Files\Java\jdk1.8.0_281`.
11. Save and close the "hadoop-env.cmd" file.
12. In the Command Prompt, navigate back to the Hadoop installation directory, if needed.
13. Run the "**bin\winutils.exe**" script to create necessary directories for Hadoop to work. For example: `bin\winutils.exe mkdir -p /tmp/hadoop-${user.name}/dfs/data`.
14. Inside the Hadoop installation directory, locate the "**etc\hadoop**" subdirectory and open the "**core-site.xml**" file with a text editor.
15. Add the following configuration to the "**core-site.xml**" file, replacing the value of "**fs.defaultFS**" with the desired Hadoop file system URI:

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

16. Save and close the "**core-site.xml**" file.
17. Inside the "**etc\hadoop**" subdirectory, open the "**hdfs-site.xml**" file with a text editor.
18. Add the following configuration to the "hdfs-site.xml" file:

```
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
```

19. Save and close the "hdfs-site.xml" file.
20. In the Command Prompt, navigate to the "**sbin**" subdirectory inside the Hadoop installation directory. For example: `cd sbin`.
21. Run the "**start-dfs.cmd**" script to start the Hadoop Distributed File System (HDFS).
22. Run the "**start-yarn.cmd**" script to start the YARN resource manager.

23. Apache Hadoop is now installed on your Windows machine. You can start using it by running Hadoop applications or executing Hadoop commands in the Command Prompt. Note that you may need to configure additional settings, such as setting up SSH keys for Hadoop daemons, depending on your specific setup. Refer to the Apache Hadoop documentation for more information.

Note:

If all steps have been followed, Apache Hadoop should run fine on your machine. If, however, troubleshooting and support is required, please contact that Data Science Helpdesk to resolve¹.

Click [HERE](#) to schedule a meeting with the Helpdesk!

¹ This document was last updated on April 24, 2023

Apache Hadoop Installation Guide (MacOS)

Installing Apache Hadoop

1. Go to the Apache Hadoop official website. Here is the link: <https://hadoop.apache.org/>
2. Click on the "**Downloads**" tab in the top menu.
3. Scroll down to the "**Stable Releases**" section and click on the "**Download**" link for the latest stable release of Apache Hadoop for macOS.
4. Save the Hadoop binary distribution TAR.GZ file to your computer.
5. Open a Terminal window.
6. Navigate to the directory where you downloaded the Hadoop TAR.GZ file using the "cd" command.
7. Extract the contents of the TAR.GZ file to a directory of your choice. This will be your Hadoop installation directory. You can use the following command: `tar -zxvf <filename>.tar.gz`. For example: `tar -zxvf hadoop-3.3.1.tar.gz`.
8. After extraction, navigate to the Hadoop installation directory using the "cd" command. For example: `cd hadoop-3.3.1`.
9. Inside the Hadoop installation directory, go to the "**etc**" subdirectory and locate the "**hadoop-env.sh**" file. Make a copy of this file and rename it to "**hadoop-env.sh.bak**" for backup purposes.
10. Open the "**hadoop-env.sh**" file with a text editor.
11. Find the line that sets the "JAVA_HOME" environment variable and set its value to the path of your Java JDK installation directory. For example: `export JAVA_HOME=/Library/Java/JavaVirtualMachines/jdk1.8.0_281.jdk/Contents/Home`.
12. Save and close the "hadoop-env.sh" file.
13. In the Terminal, navigate back to the Hadoop installation directory, if needed.
14. Run the following command to create necessary directories for Hadoop to work: `./bin/hadoop namenode -format`.
15. Inside the "**etc/hadoop**" subdirectory, open the "**core-site.xml**" file with a text editor.
16. Add the following configuration to the "core-site.xml" file, replacing the value of "**fs.defaultFS**" with the desired Hadoop file system URI:

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

17. Save and close the "core-site.xml" file.
18. Inside the "etc/hadoop" subdirectory, open the "hdfs-site.xml" file with a text editor.
19. Add the following configuration to the "hdfs-site.xml" file:

```
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
```

20. Save and close the "hdfs-site.xml" file.
21. In the Terminal, navigate to the "sbin" subdirectory inside the Hadoop installation directory. For example: `cd sbin`.
22. Run the following command to start the Hadoop Distributed File System (HDFS): `./start-dfs.sh`.

23. Run the following command to start the YARN resource manager: `./start-yarn.sh`.
24. Apache Hadoop is now installed on your macOS machine. You can start using it by running Hadoop applications or executing Hadoop commands in the Terminal. Note that you may need to configure additional settings, such as setting up SSH keys for Hadoop daemons, depending on your specific setup. Refer to the Apache Hadoop documentation for more information.

Note:

If all steps have been followed, Apache Hadoop should run fine on your machine. If, however, troubleshooting and support is required, please contact that Data Science Helpdesk to resolve².

Click [HERE](#) to schedule a meeting with the Helpdesk!

² This document was last updated on April 24, 2023