**Apache Spark 3.5 Installation on Windows**

Steps to install Apache Spark 3.5 Installation on Windows – In this article, I will explain step-by-step how to do Apache Spark 3.5 Installation on Windows OS 10 or 11.

## Install **Java 21** or Later

To install Apache Spark 3.5 on Windows 10 or 11, you would need to install Java/JDK 8, 11, 17, or the latest version

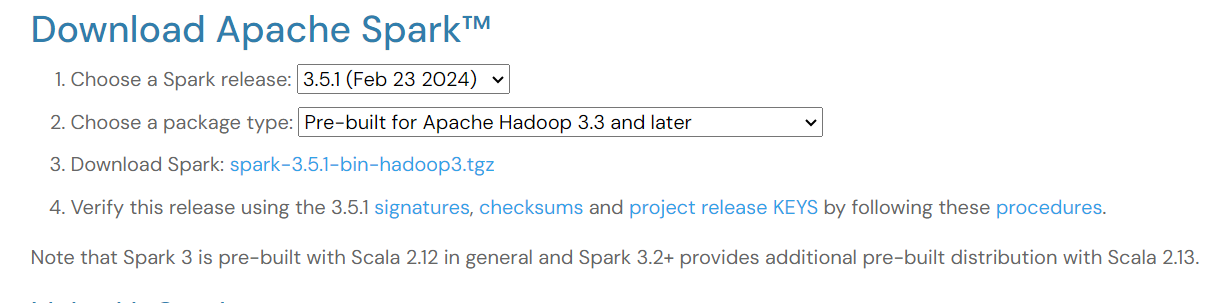
Download Java:- <https://download.oracle.com/java/21/latest/jdk-21_windows-x64_bin.exe> ([sha256](https://download.oracle.com/java/21/latest/jdk-21_windows-x64_bin.exe.sha256))

## **Apache Spark** Installation on Windows

Apache Spark comes in compressed tar/zip files; hence, installation on Windows is not much of a deal as you need to download and untar the file. Download Apache Spark by accessing the Spark Download page and selecting the link from “Download Spark (point 3 from below screenshot)”.

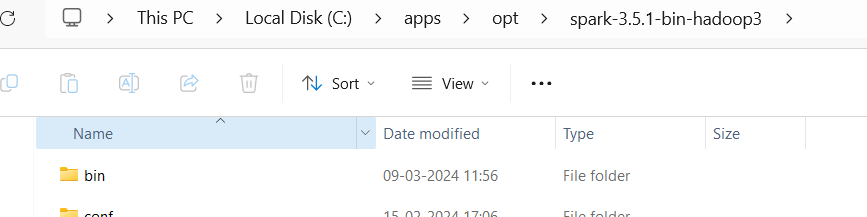
If you want to use a different version of Spark & Hadoop, select the one you want from the drop-down; the link on point 3 changes to the selected version and provides you with an updated link to download.

Download Spark: - <https://spark.apache.org/downloads.html>



After download, untar the binary using [7zip](https://www.7-zip.org/download.html) or any zip utility to extract the zip file and copy the extracted directory spark-3.5.1-bin-hadoop3 to c:\apps\opt\spark-3.5.1-bin-hadoop3

C:\apps\opt\spark-3.5.1-bin-hadoop3



## Spark Environment Variables

Post Java and Apache Spark installation on Windows, set JAVA\_HOME, SPARK\_HOME, HADOOP\_HOME, and PATH environment variables. If you know how to set the environment variable on Windows, add the following.

**JAVA\_HOME = C:\Program Files\Java\jdk-21**

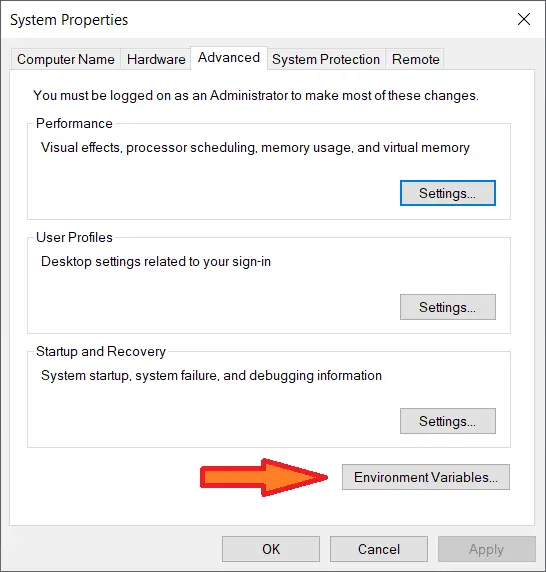
**SPARK\_HOME = C:\apps\opt\spark-3.5.1-bin-hadoop3**

**HADOOP\_HOME = C:\apps\opt\spark-3.5.1-bin-hadoop3**

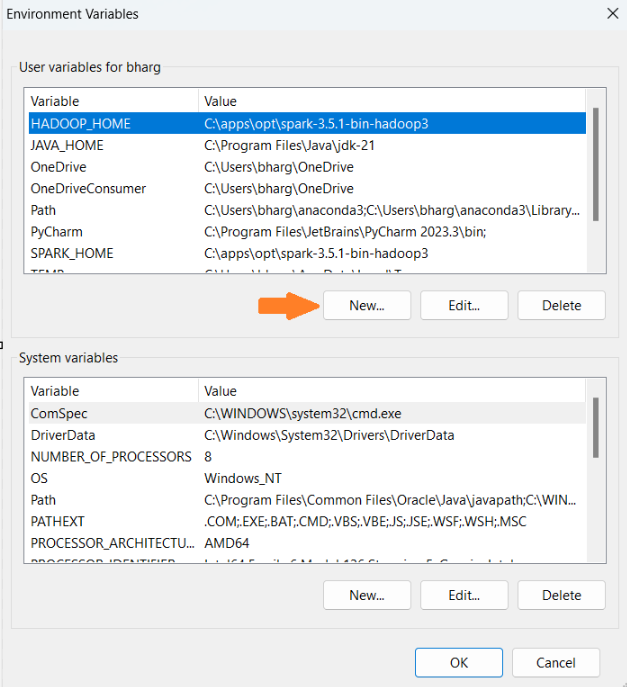
**PATH=%SPARK\_HOME%\bin;%JAVA\_HOME%\bin;%HADOOP\_HOME%\bin;**

Follow the steps below if you are unaware of how to add or edit environment variables on Windows.

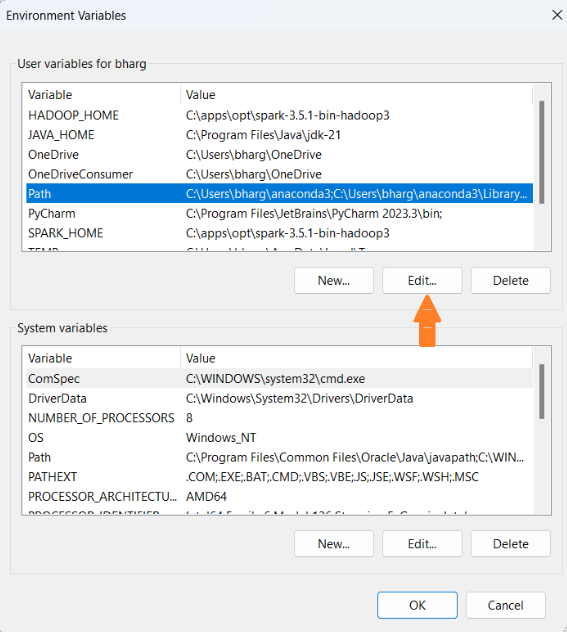
1. Open the System Environment Variables window and select **Environment Variables**.



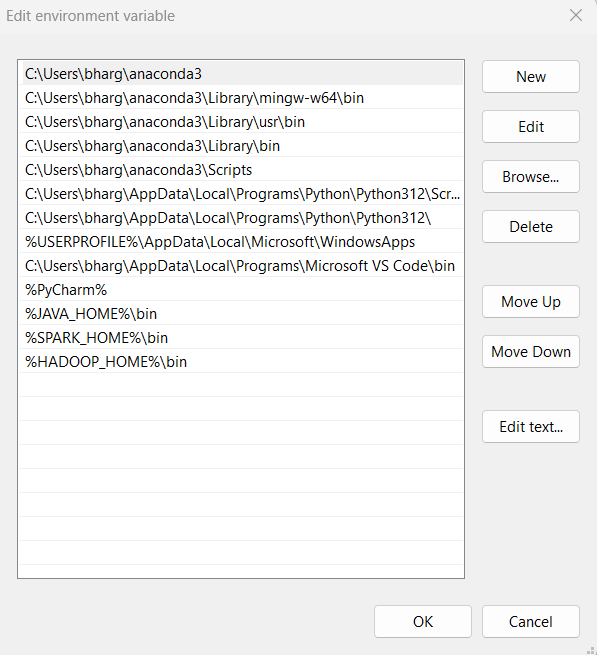
1. On the following Environment variable screen, add SPARK\_HOME, HADOOP\_HOME, JAVA\_HOME by selecting the **New** option.



1. This opens up the **New User Variables** window where you can enter the variable name and value. Add respective paths to these variables.
2. Now **Edit** the **PATH** variable.



1. Add Spark, Java, and Hadoop bin locations by selecting the **New** option.



## Spark with winutils.exe on Windows

Many beginners think Apache Spark needs a Hadoop cluster installed to run, but that’s not true; Spark can run on AWS using S3 and Azure using blob storage without Hadoop and HDFSe.t.c.

To run Apache Spark on Windows, you need winutils.exe as it uses POSIX like file access operations in Windows using Windows API.

winutils.exe enables Spark to use Windows-specific services, including running shell commands on a Windows environment.

Download  [winutils](https://github.com/bharggav/Py-Spark/blob/rel1/winutils.zip) for Hadoop 3.3 and copy it to %SPARK\_HOME%\bin folder. Winutils differ for each Hadoop version; hence, download the right version based on your Spark vs Hadoop distribution.

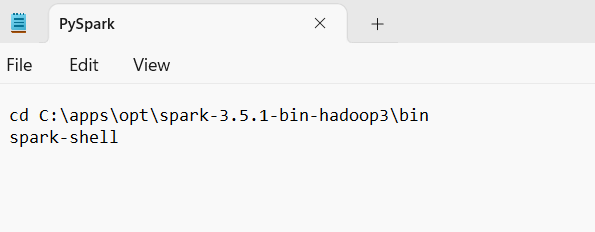
## Apache Spark shell

spark-shell is a CLI utility that comes with Apache Spark distribution, open command prompt, go to cd %SPARK\_HOME%/bin and type spark-shell command to run Apache Spark shell. You should see something like this below (ignore the error you see at the end). Sometimes, your Spark instance may take a minute or two to initialize to get to the below screen.

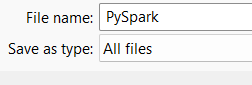
Or Create a CMD file on desktop.

**cd C:\apps\opt\spark-3.5.1-bin-hadoop3\bin**

**spark-shell**



Save as



File Name: - PySpark.cmd

Save as type: - All files



Double click the file.

