**SPACK AND SCALA INSTALLATION**

To install Scala in the system, the system must have JAVA in it.

**Install Java 8**

Apache Spark requires Java 8. You can check to see if Java is installed using the command prompt.

Open the command line by clicking **Start** > type *cmd* > click **Command Prompt**.

Type the following command in the command prompt:

java -version

If you don’t have Java installed:

1. Open a browser window, and navigate to <https://java.com/en/download/>.

2. Click the **Java Download** button and save the file to a location of your choice.

3. Once the download finishes double-click the file to install Java.

**Download Apache Spark**

1. Open a browser and navigate to <https://spark.apache.org/downloads.html>.

2. Under the *Download Apache Spark*heading, there are two drop-down menus. Use the current non-preview version.

* In our case, in ***Choose a Spark release***drop-down menu select **2.4.5 (Feb 05 2020)**.
* In the second drop-down ***Choose a package type*,** leave the selection **Pre-built for Apache Hadoop 2.7**.

3. Click the ***spark-2.4.5-bin-hadoop2.7.tgz***link.

4. A page with a list of mirrors loads where you can see different servers to download from. Pick any from the list and save the file to your Downloads folder.

### Verify Spark Software File

1. Verify the integrity of your download by checking the **checksum** of the file. This ensures you are working with unaltered, uncorrupted software.

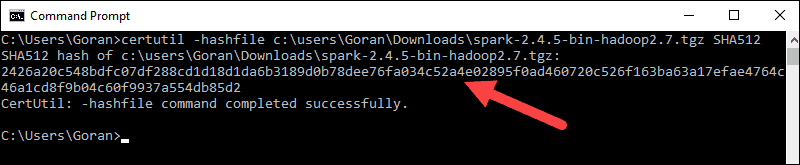
2. Navigate back to the Spark Download page and open the **Checksum** link, preferably in a new tab.

3. Next, open a command line and enter the following command:

certutil -hashfile c:\users\username\Downloads\spark-2.4.5-bin-hadoop2.7.tgz SHA512

4. Change the username to your username. The system displays a long alphanumeric code, along with the message **Certutil: -hashfile completed successfully**.

5. Compare the code to the one you opened in a new browser tab. If they match, your download file is uncorrupted.



### Install Apache Spark

Installing Apache Spark involves **extracting the downloaded file** to the desired location.

1. Create a new folder named Spark in the root of your C: drive. From a command line, enter the following:

cd \

mkdir Spark

2. In Explorer, locate the Spark file you downloaded.

3. Right-click the file and extract it to C:\Spark using the tool you have on your system (e.g., 7-Zip).

4. Now, your C:\Spark folder has a new folder spark-2.4.5-bin-hadoop2.7 with the necessary files inside.

**Add winutils.exe File**

Download the **winutils.exe** file for the underlying Hadoop version for the Spark installation you downloaded.

1. Navigate to this URL <https://github.com/cdarlint/winutils> and inside the **bin** folder, locate **winutils.exe**, and click it.



2. Find the **Download**button on the right side to download the file.

3. Now, create new folders ***Hadoop***and **bin** on C: using Windows Explorer or the Command Prompt.

4. Copy the winutils.exe file from the Downloads folder to **C:\hadoop\bin**.

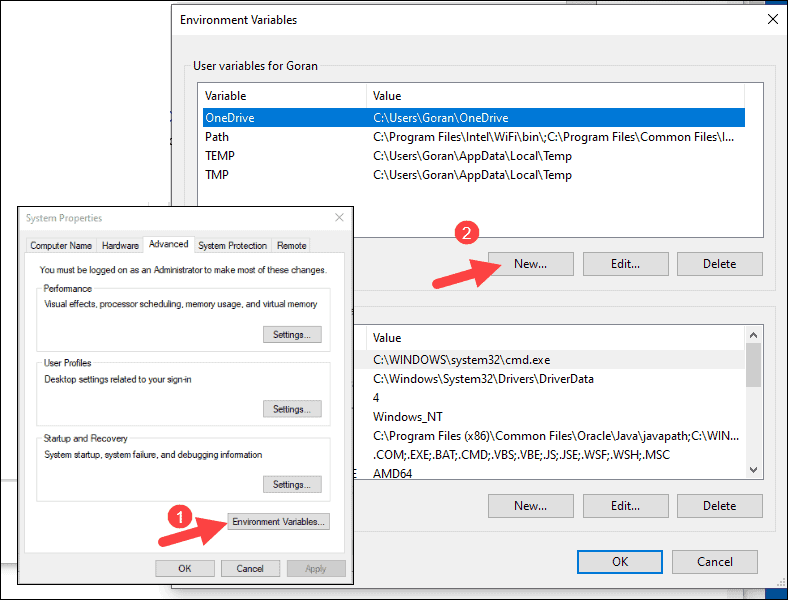
**Configure Environment Variables**

Configuring [environment variables in Windows](https://phoenixnap.com/kb/windows-set-environment-variable) adds the Spark and Hadoop locations to your system PATH. It allows you to run the Spark shell directly from a command prompt window.

1. Click **Start** and type *environment*.

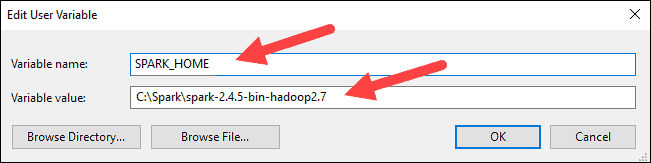
2. Select the result labeled ***Edit the system environment variables***.

3. A System Properties dialog box appears. In the lower-right corner, click **Environment Variables** and then click **New** in the next window.

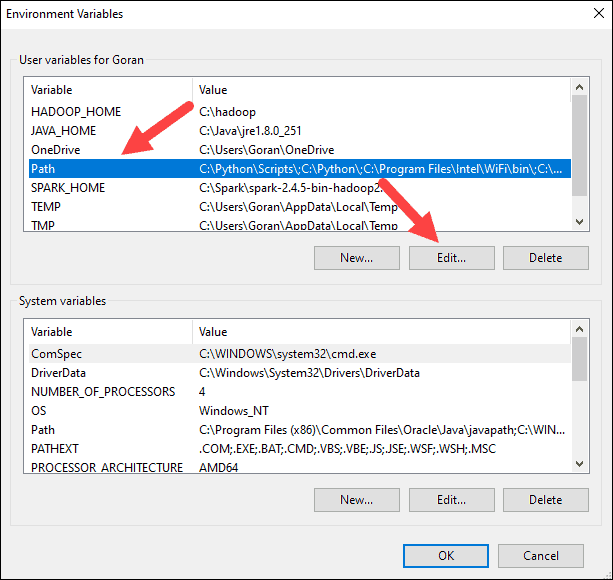


4. For *Variable Name* type ***SPARK\_HOME***.

5. For *Variable Value*type **C:\Spark\spark-2.4.5-bin-hadoop2.7**and click OK. If you changed the folder path, use that one instead.

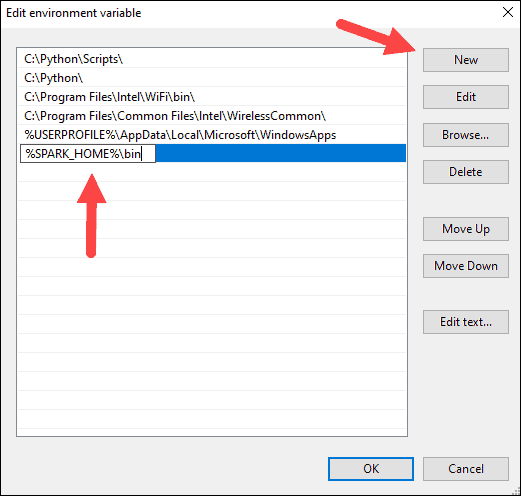


6. In the top box, click the **Path** entry, then click **Edit**. Be careful with editing the system path. Avoid deleting any entries already on the list.



7. You should see a box with entries on the left. On the right, click **New**.

8. The system highlights a new line. Enter the path to the Spark folder **C:\Spark\spark-2.4.5-bin-hadoop2.7\bin**. We recommend using **%SPARK\_HOME%\bin**to avoid possible issues with the path.



9. Repeat this process for Hadoop and Java.

* For Hadoop, the variable name is **HADOOP\_HOME** and for the value use the path of the folder you created earlier: **C:\hadoop.**Add **C:\hadoop\bin**to the **Path variable**field, but we recommend using **%HADOOP\_HOME%\bin**.
* For Java, the variable name is **JAVA\_HOME** and for the value use the path to your Java JDK directory (in our case it’s **C:\Program Files\Java\jdk1.8.0\_251**).

10. Click **OK** to close all open windows.

**Note:** Star by restarting the Command Prompt to apply changes. If that doesn't work, you will need to reboot the system.

**Step 8: Launch Spark**

1. Open a new command-prompt window using the right-click and **Run as administrator**:

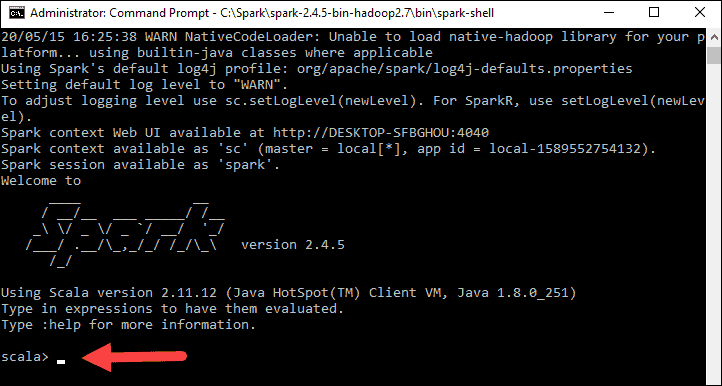
2. To start Spark, enter:

C:\Spark\spark-2.4.5-bin-hadoop2.7\bin\spark-shell

If you set the **environment path** correctly, you can type **spark-shell** to launch Spark.

3. The system should display several lines indicating the status of the application. You may get a Java pop-up. Select **Allow access** to continue.

Finally, the Spark logo appears, and the prompt displays the **Scala shell**.



4., Open a web browser and navigate to **http://localhost:4040/**.

5. You can replace **localhost**with the name of your system.

6. You should see an Apache Spark shell Web UI.