

How To Download Apache Spark

Step 1: Check whether Java is Installed.

Check whether JAVA is Installed or Not by executing this Command on cmd: **java --version**

```
PS C:\Users\manvi.gour> java --version
java 22.0.2 2024-07-16
Java(TM) SE Runtime Environment (build 22.0.2+9-70)
Java HotSpot(TM) 64-Bit Server VM (build 22.0.2+9-70, mixed mode, sharing)
PS C:\Users\manvi.gour> _
```

If not then Install it from <https://www.oracle.com/in/java/technologies/downloads/>

Step 2: Check whether Python is Installed.

Check whether Python is Installed or Not by executing this Command on cmd: **python --version**

```
PS C:\Users\manvi.gour> python --version
Python 3.12.4
PS C:\Users\manvi.gour> _
```

If not then Install it from <https://www.python.org/downloads/>

Step 3: Install Apache Spark.

Link: <https://spark.apache.org/downloads.html>

Step 3.1: Click on the Downlad Spark.

Download Apache Spark™

1. Choose a Spark release: 3.5.1 (Feb 23 2024) ▼

2. Choose a package type: Pre-built for Apache Hadoop 3.3 and later ▼

3. Download Spark: spark-3.5.1-bin-hadoop3.tgz

4. Verify this release using the 3.5.1 [signatures](#), [checksums](#) and [project release KEYS](#) by following these [procedures](#).

Note that Spark 3 is pre-built with Scala 2.12 in general and Spark 3.2+ provides additional pre-built distribution with Scala 2.13.

Step 3.2 : Click on the Downlad Link.



We suggest the following location for your download:

<https://d1cdn.apache.org/spark/spark-3.5.1/spark-3.5.1-bin-hadoop3.tgz>

Alternate download locations are suggested below.

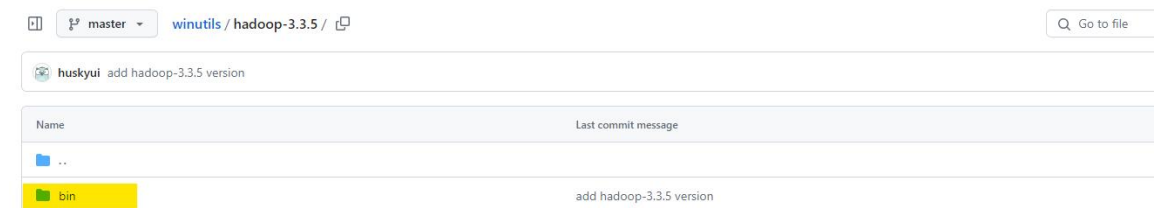
It is essential that you [verify the integrity](#) of the downloaded file using the PGP signature (`.asc` file) or a hash (`.md5` or `.sha*` file).

UTTD

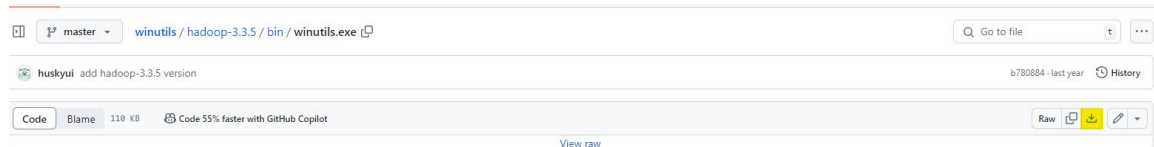
Step 4: Download Winutils.

Link to Download: <https://github.com/cdarlint/winutils/tree/master>





Step 4.1 : Choose Hadoop Version according to Apache Spark package type mentioned in Step 3.1.



Step 4.2: Goto Bin directory and Click Download.



Step 5: Extract Apache Spark Files and rename it as **spark-version**.

 winutils	31-07-2024 12:17	Application	110 KB
 spark-3.5.1-bin-hadoop3	31-07-2024 12:09	WinRAR archive	3,91,062 KB
 jdk-22_windows-x64_bin	31-07-2024 11:52	Windows Installer ...	1,67,008 KB
 spark-3.5.1	31-07-2024 12:10	File folder	

Step 6: Create Folder hadoop in C Drive.

#65e#IQ	22-07-2024 10:10	File folder
cdnpV0HEiX0O	22-07-2024 10:10	File folder
hadoop	31-07-2024 12:18	File folder
PerfLogs	22-07-2024 10:10	File folder
Program Files	31-07-2024 11:54	File folder

Step 6.1: Create bin Folder inside hadoop folder.

This PC > Windows (C:) > hadoop >				
Name	Date modified	Type	Size	
bin	31-07-2024 12:18	File folder		

Step 6.3: Paste the Winutils file inside the bin folder.

This PC > Windows (C:) > hadoop > bin				
Name	Date modified	Type	Size	
winutils	31-07-2024 12:17	Application	110 KB	

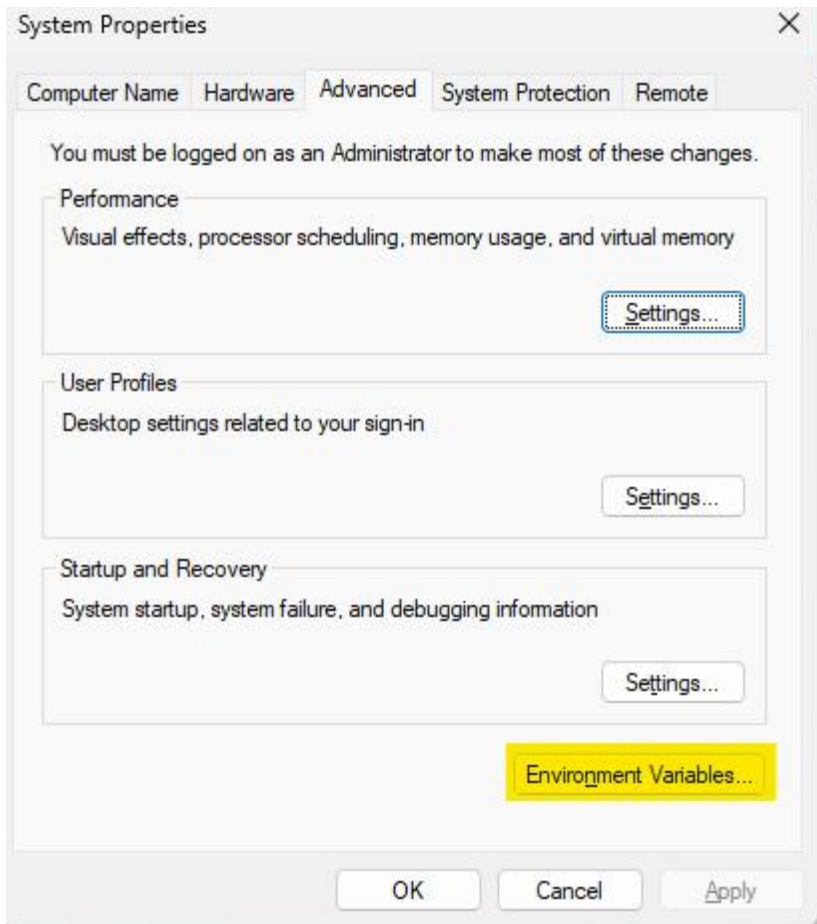
Step 7: Create Folder spark in C Drive.

This PC > Windows (C:) >				
Name	Date modified	Type	Size	
#65e#IQ	22-07-2024 10:10	File folder		
cdnpV0HEiX0O	22-07-2024 10:10	File folder		
hadoop	31-07-2024 12:18	File folder		
PerfLogs	22-07-2024 10:10	File folder		
Program Files	31-07-2024 11:54	File folder		
Program Files (x86)	22-07-2024 10:10	File folder		
spark	31-07-2024 12:20	File folder		

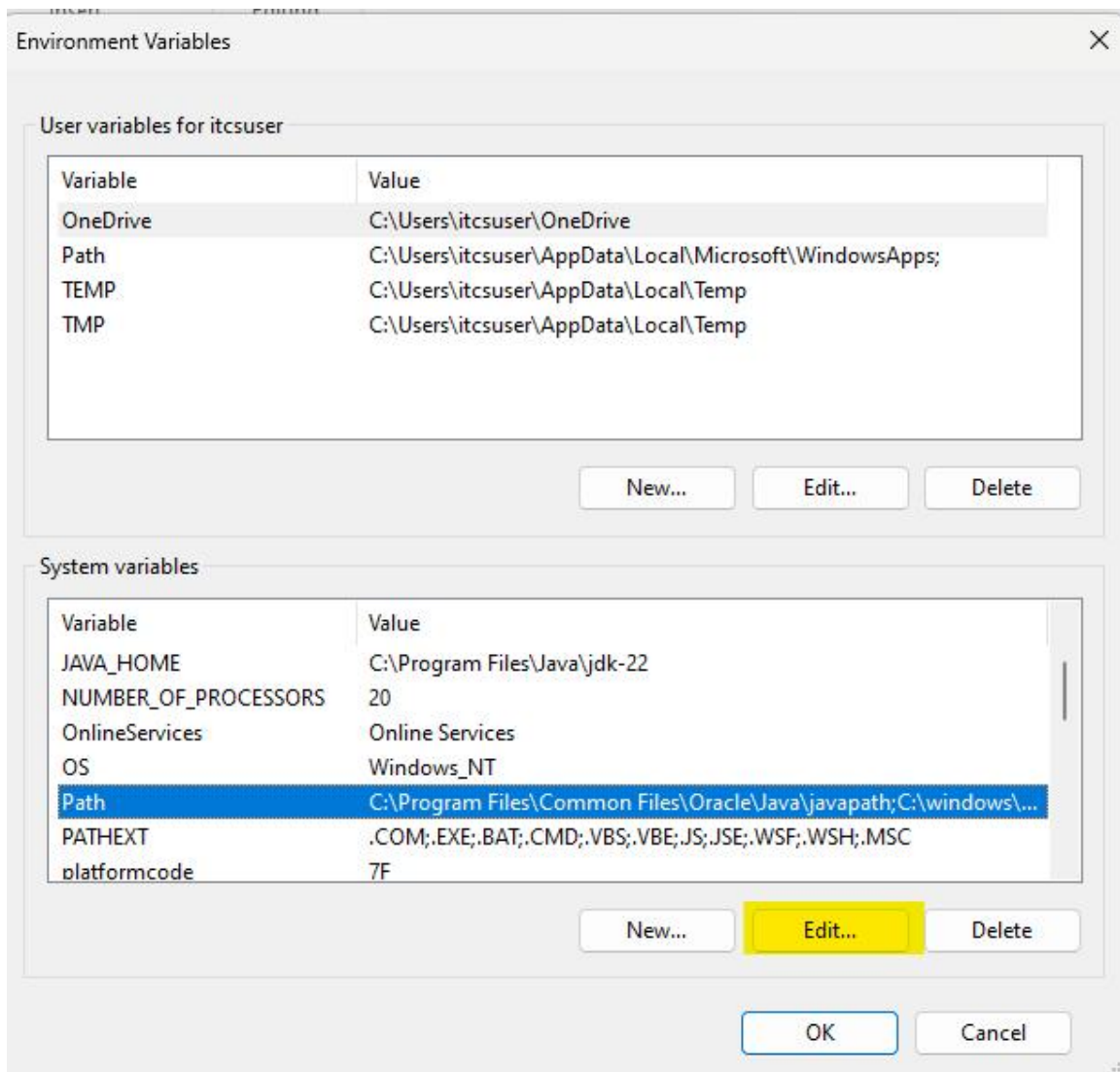
Step 7.1: Copy the Contents of Extracted Apache Spark File and Paste it in the spark Folder created.

This PC > Windows (C:) > spark				
Name	Date modified	Type	Size	
bin	31-07-2024 12:19	File folder		
conf	31-07-2024 12:19	File folder		
data	31-07-2024 12:19	File folder		
examples	31-07-2024 12:19	File folder		
jars	31-07-2024 12:19	File folder		
kubernetes	31-07-2024 12:19	File folder		
licenses	31-07-2024 12:19	File folder		
python	31-07-2024 12:19	File folder		
R	31-07-2024 12:19	File folder		
sbin	31-07-2024 12:19	File folder		
yarn	31-07-2024 12:19	File folder		
LICENSE	15-02-2024 17:06	File	23 KB	
NOTICE	15-02-2024 17:06	File	57 KB	
README	15-02-2024 17:06	Markdown Source...	5 KB	
RELEASE	15-02-2024 17:06	File	1 KB	

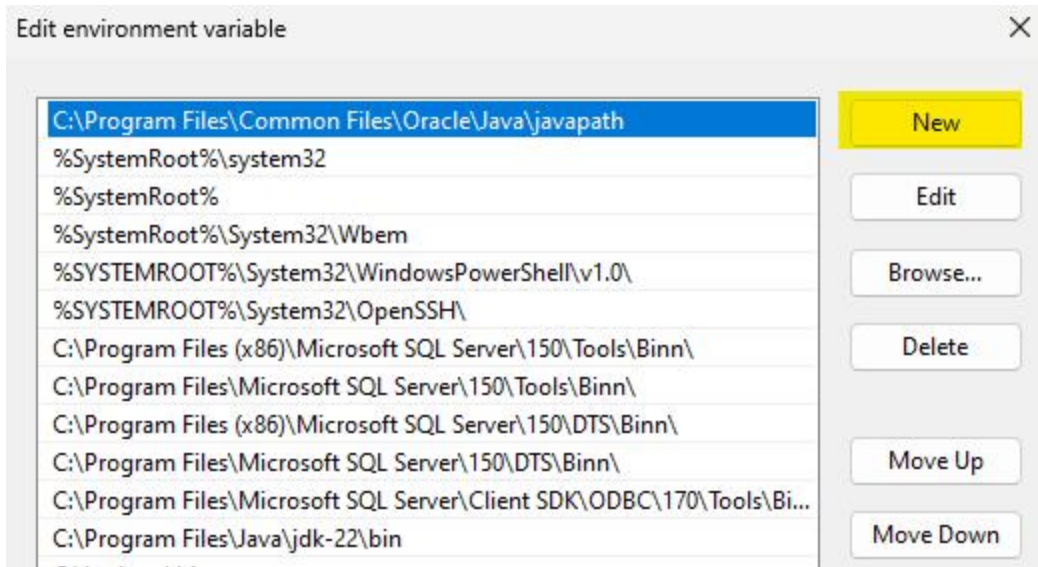
Step 8: Open System Properties to set Environment Variables.



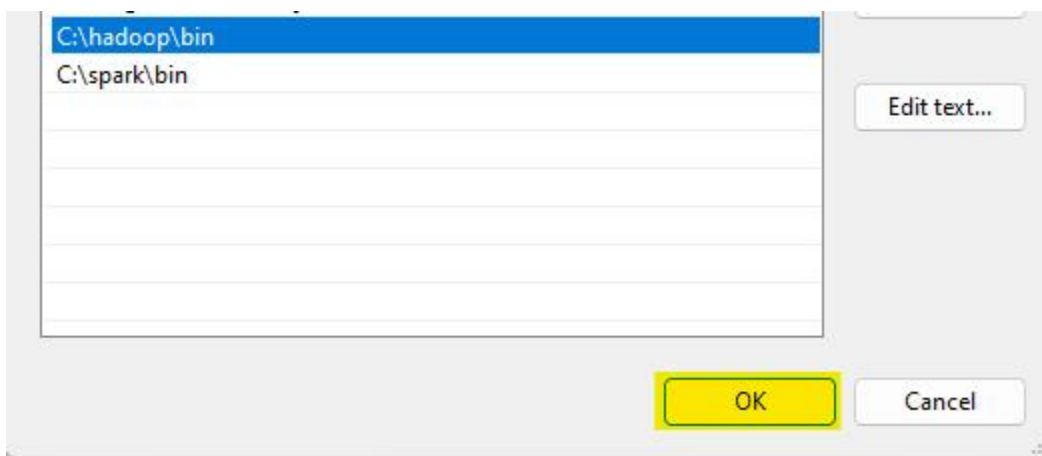
Step 8.1: Click on Path and then select Edit.



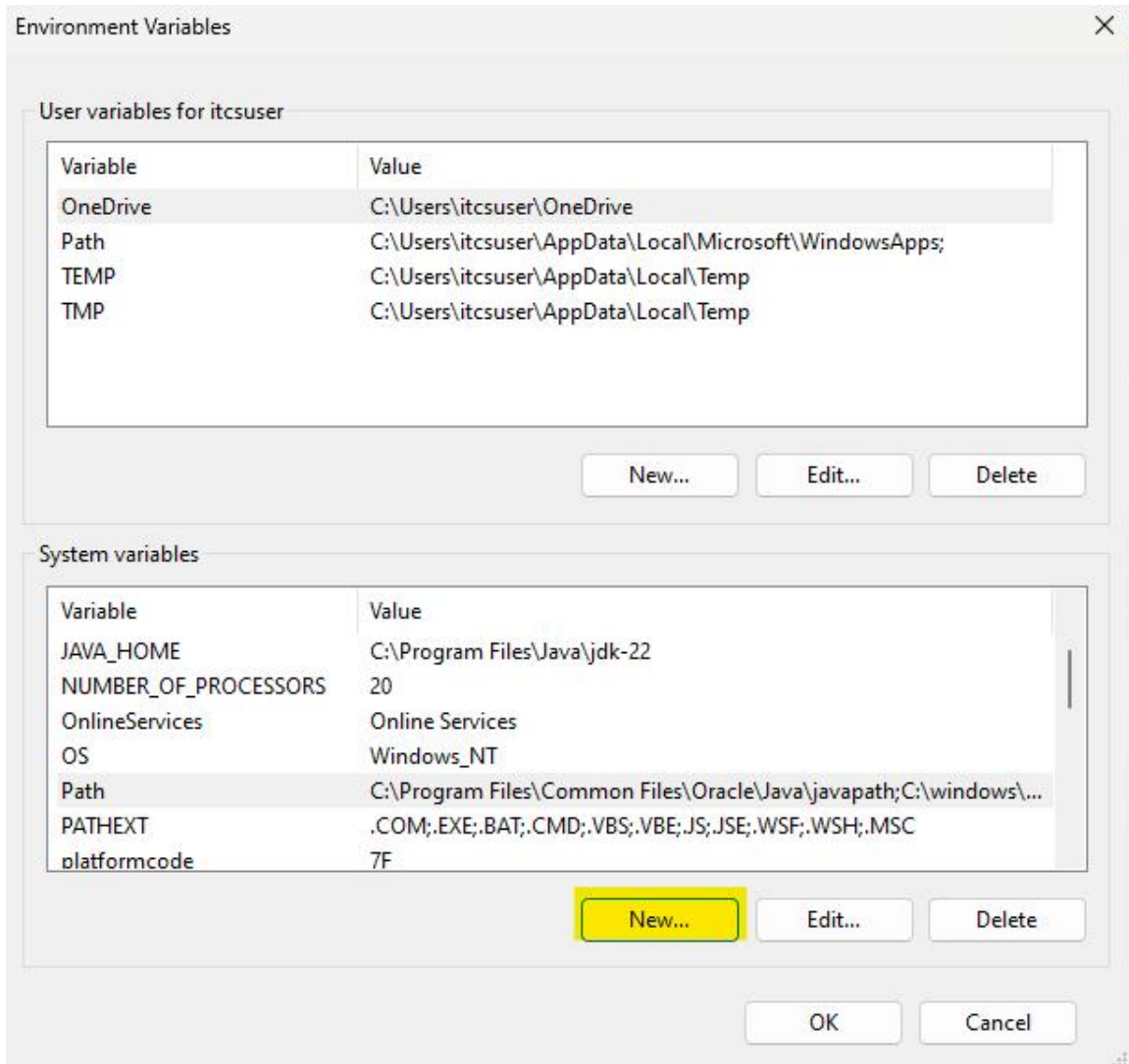
Step 8.2: Select New Option.



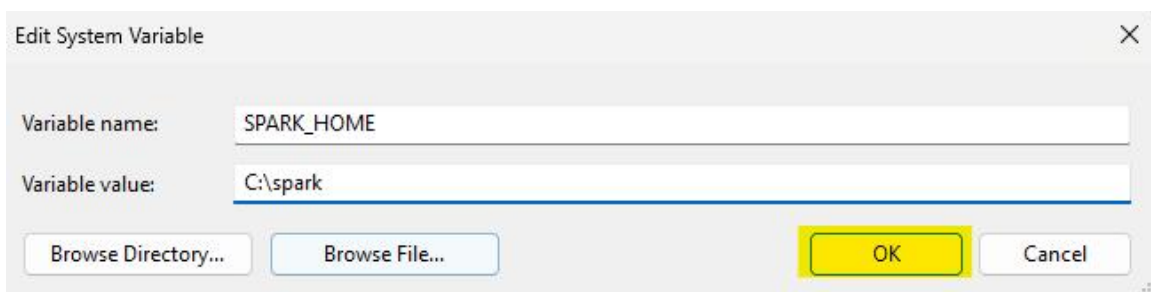
Step 8.3: Add bin directory to both the Hadoop and Spark Folder and Click on OK.



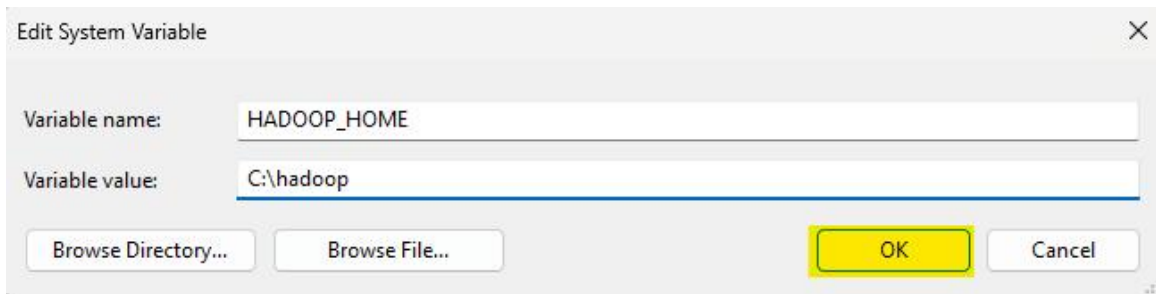
Step 9: Select New option in System Variables



Step 9.1: Add Directory to Spark Folder as Variable value and Variable name as SPARK_HOME and Click on OK.

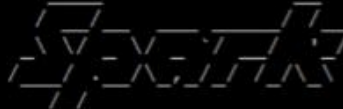


Step 9.2: Add Directory to hadoop Folder as Variable value and Variable name as HADOOP_HOME and Click on OK.



Step 10: Open Command Prompt and type **spark-shell** if the setup was successful it would display the following output.

```
log4j:WARN Please initialize the log4j system properly.  
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.  
Using Spark's repl log4j profile: org/apache/spark/log4j-defaults-repl.properties  
To adjust logging level use sc.setLogLevel("INFO")  
Welcome to
```



The Scala logo consists of two interlocking triangles forming a larger triangle, composed of smaller geometric shapes.

version 1.6.1

```
Using Scala version 2.10.5 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_91)  
Type in expressions to have them evaluated.  
Type :help for more information.
```

Step 10.1: Copy and Paste this URL to your Browser.

```

Spark context Web UI available at http://IMSMUMDT001396.WRITERCORPORATION.COM:4040
Spark context available as 'sc' (master = local[*], app id = local-1722417108066).
Spark session available as 'spark'.
Welcome to

```

You will get local host site then it means the Installation was Successful

Spark Jobs ^(?)

User: manvi.gour

Total Uptime: 52 s

Scheduling Mode: FIFO

[▶ Event Timeline](#)

