

# Environment Setup

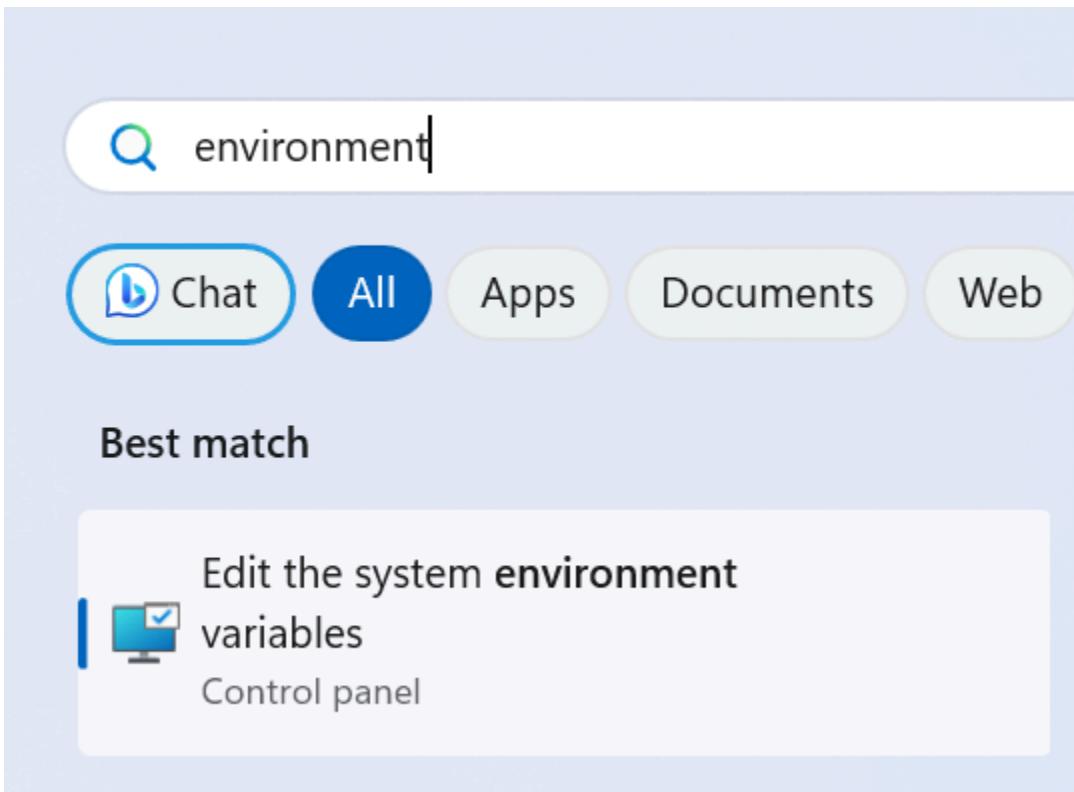
## Java Development Environment Setup

### JDK Installation

Download  [Download JDK 11](#)

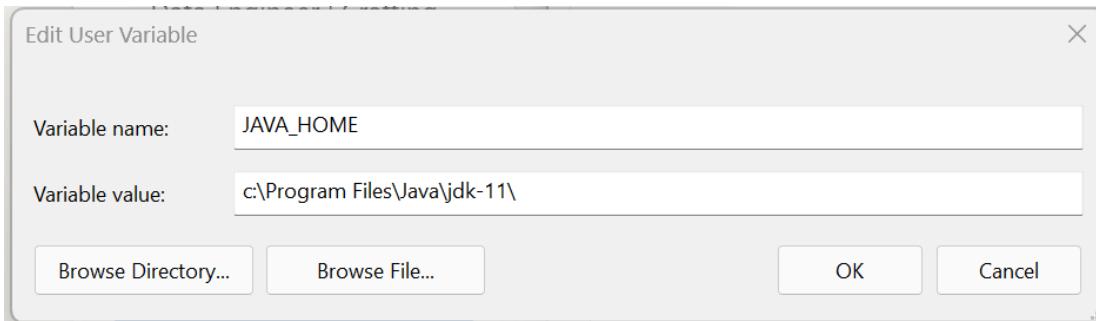
### Configure Environment Variables

1. Press the Windows key and search for **Environment** and click on **Edit the system environment variables**.



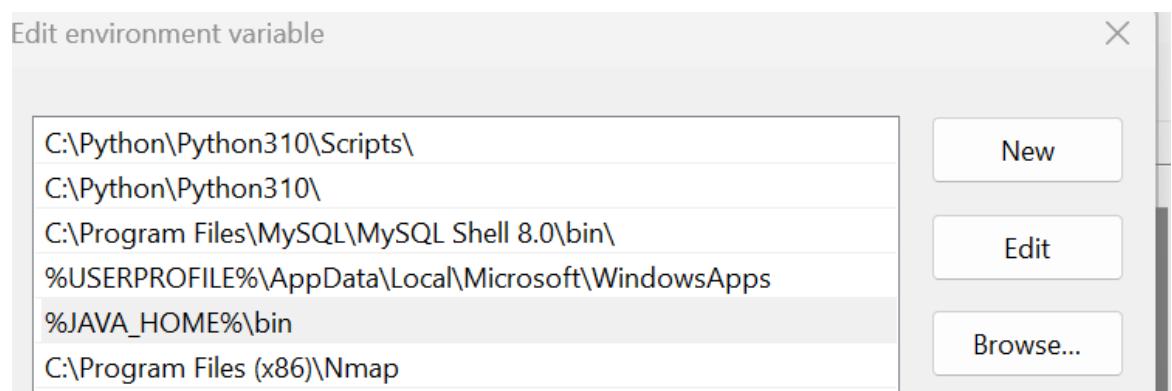
Click the Advanced tab, → Click the Environment Variables button

2. Under User variables click **New** and add the following



3. Under User variables search for Path variable and click Edit

Add the following



Click OK

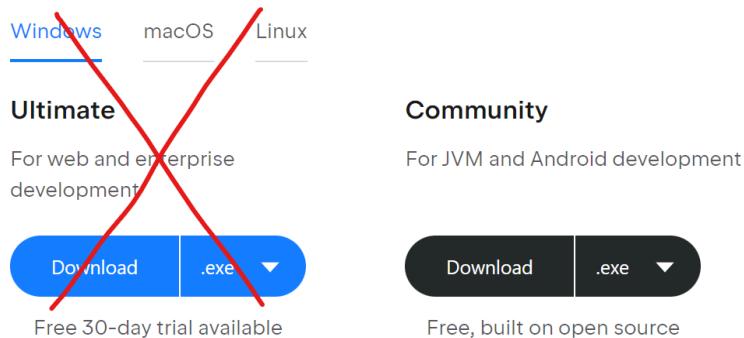
4. Check on cmd, see below:

```
C:\Users\npnpp>java -version
java version "11.0.18" 2023-01-17 LTS
Java(TM) SE Runtime Environment 18.9 (build 11.0.18+9-LTS-195)
Java HotSpot(TM) 64-Bit Server VM 18.9 (build 11.0.18+9-LTS-195, mixed mode)
```

## IntelliJ IDEA IDE Installation (**Skip for Python Developers**)

**Step 01 :** Download IntelliJ IDEA Community Edition (**Skip if already downloaded**)

## Download IntelliJ IDEA



## Python Development Environment Setup

### Python Installation

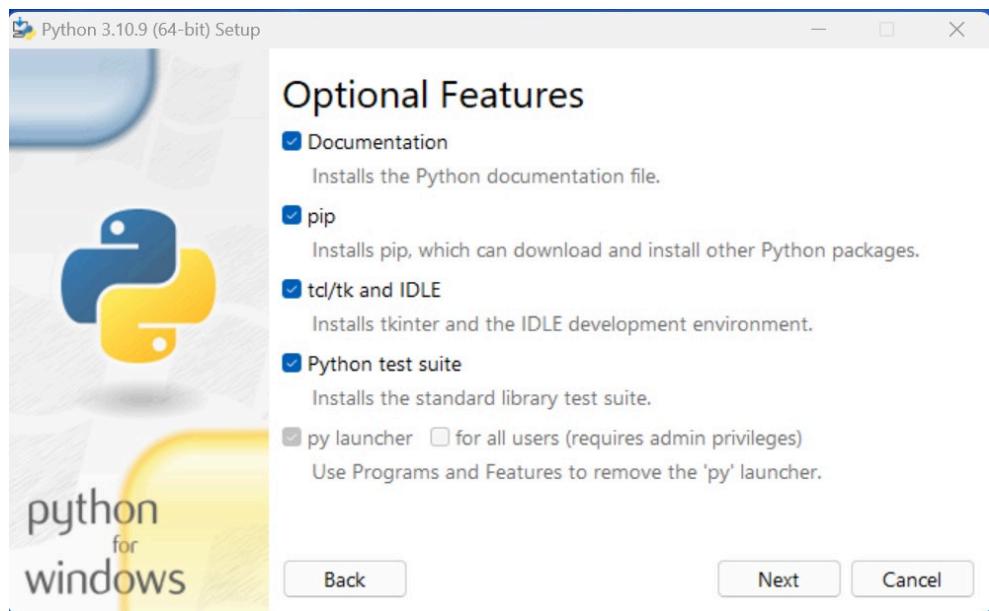
**Step 01** : Download Python from the link given below

 [Download python-3.10.9](#)

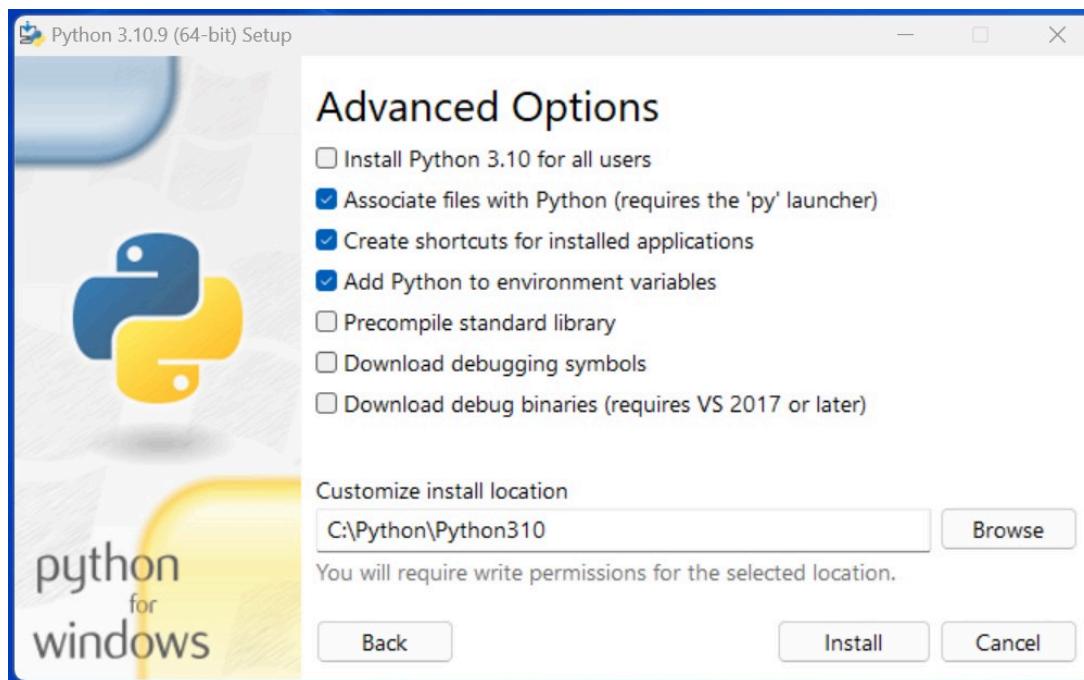
**Step 02** : Click python-3.10.9.exe which you have downloaded and click on **Add Python 3.10.9 to PATH** as shown in the screenshot and click on **Customize installation** then click Next



**Step 03** : Click Next

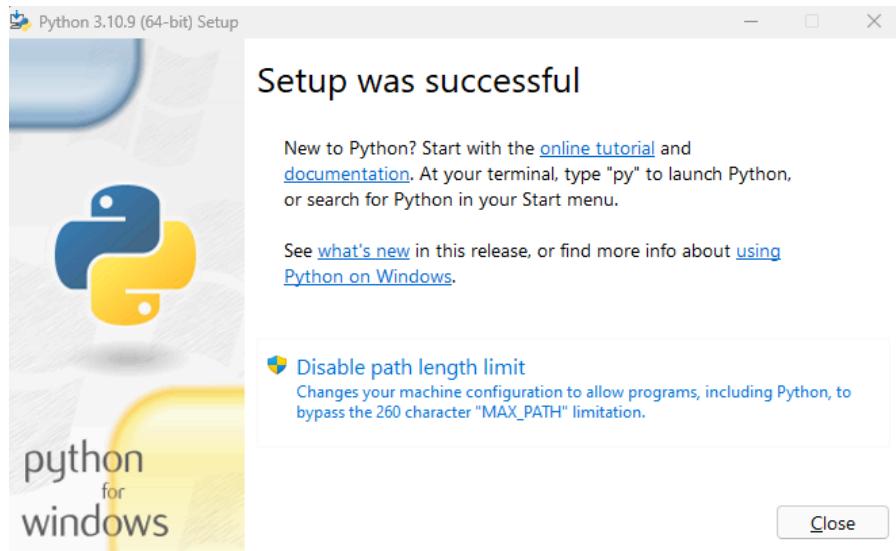


**Step 04 :** Configure Customize install location as C:\Python\Python310 shown and the Ensure Add Python to environment variables is checked then click Next

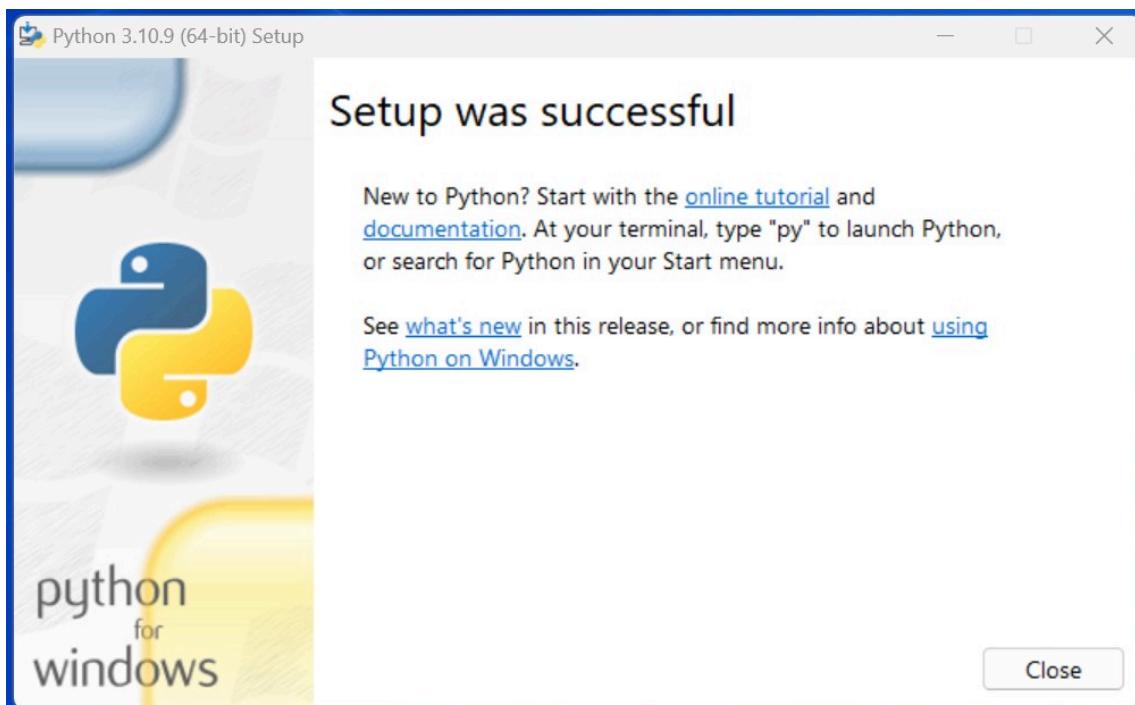


**Step 05**

Disable path length limit



**Step 06 :** Click Close



**Step 04 :** Check for correct installation

Open Command Prompt

```
C:\Users\npnnp>python
Python 3.10.9 (tags/v3.10.9:1dd9be6, Dec  6 2022, 20:01:21) [MSC v.1934 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
```

## PyCharm Installation

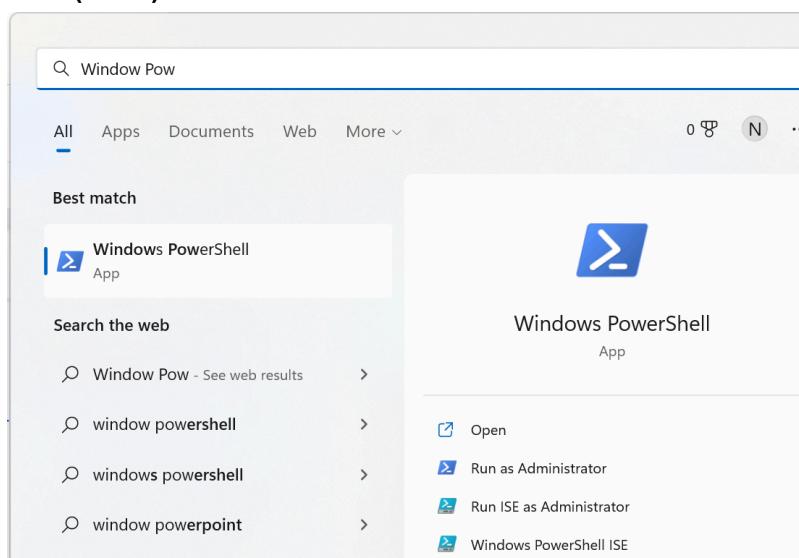
### Step 01 : Download PyCharm Community Edition

Navigate to the given link and scroll down to find PyCharm Community Edition

 [Download PyCharm Community Edition](#)



### Edit Windows Powershell (SKIP)



Open Powershell in **Administrator Mode** and execute the below command and give 'Y' wherever it prompts

```
$> Set-ExecutionPolicy -ExecutionPolicy RemoteSigned -Scope LocalMachine
```

## Apache Spark Development Environment Setup

### Scala Installation

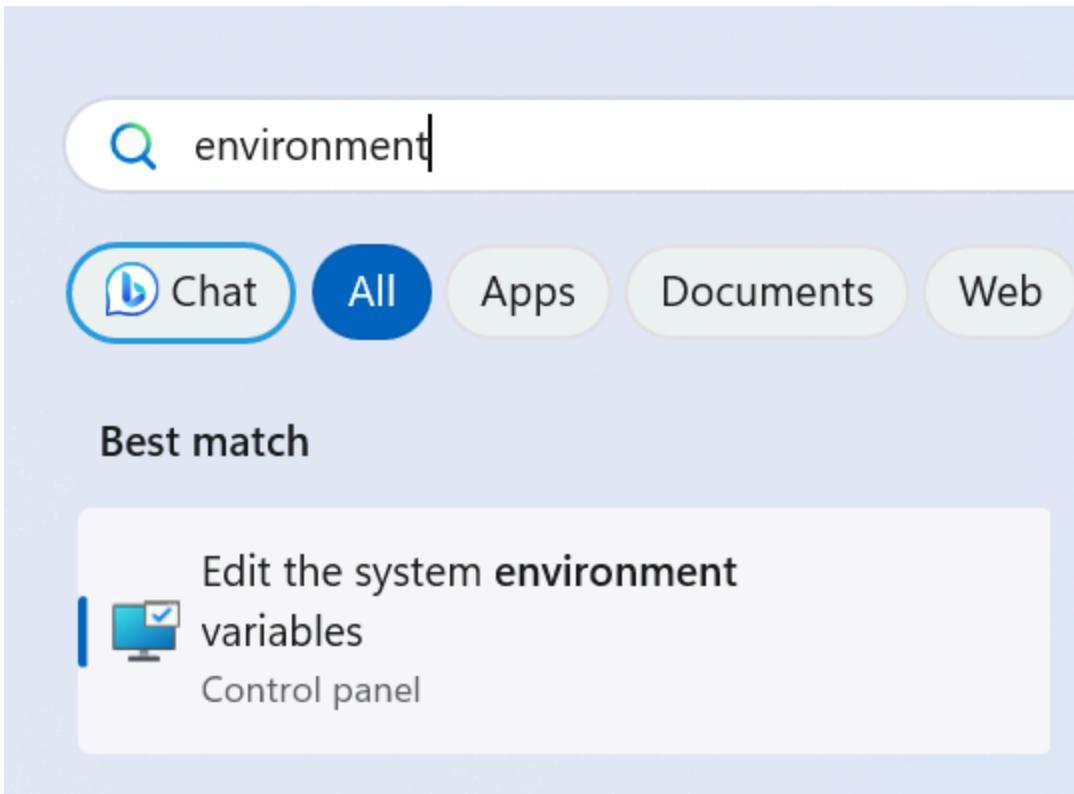
1. Download and install [Scala 2.12.8](#)

### Apache Spark Installation

1. Download [apache-spark.zip](#) to C:\data-engineering-env-setup\
2. Right Click on **apache-spark.zip** , Select **7-Zip** and select **Extract Here**

### Configure Environment Variables

1. Press the Windows key and search for **Environment** and click on **Edit the system environment variables.**

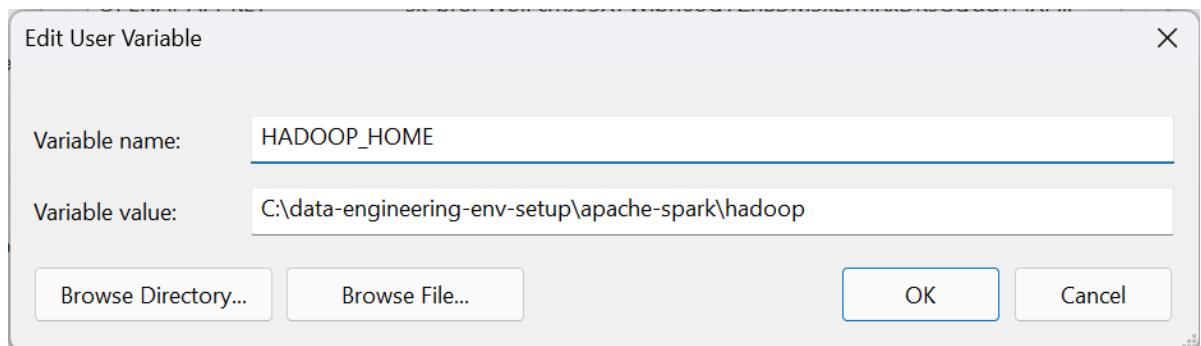


Click the Advanced tab, → Click the Environment Variables button

2. Add the following new User variables Click **New** and add below mentioned details

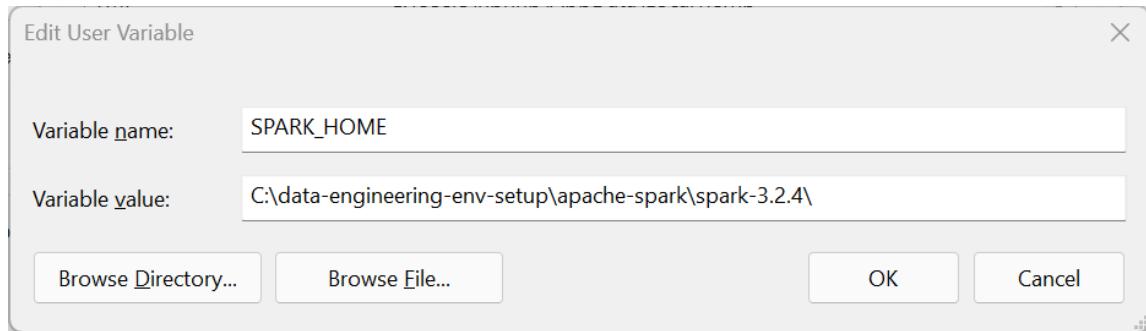
#### Configuring Hadoop

- Variable Name : HADOOP\_HOME
- Variable value: C:\data-engineering-env-setup\apache-spark\hadoop



#### Configuring Spark

- Variable Name : SPARK\_HOME
- Variable Value : C:\data-engineering-env-setup\apache-spark\spark-3.2.4



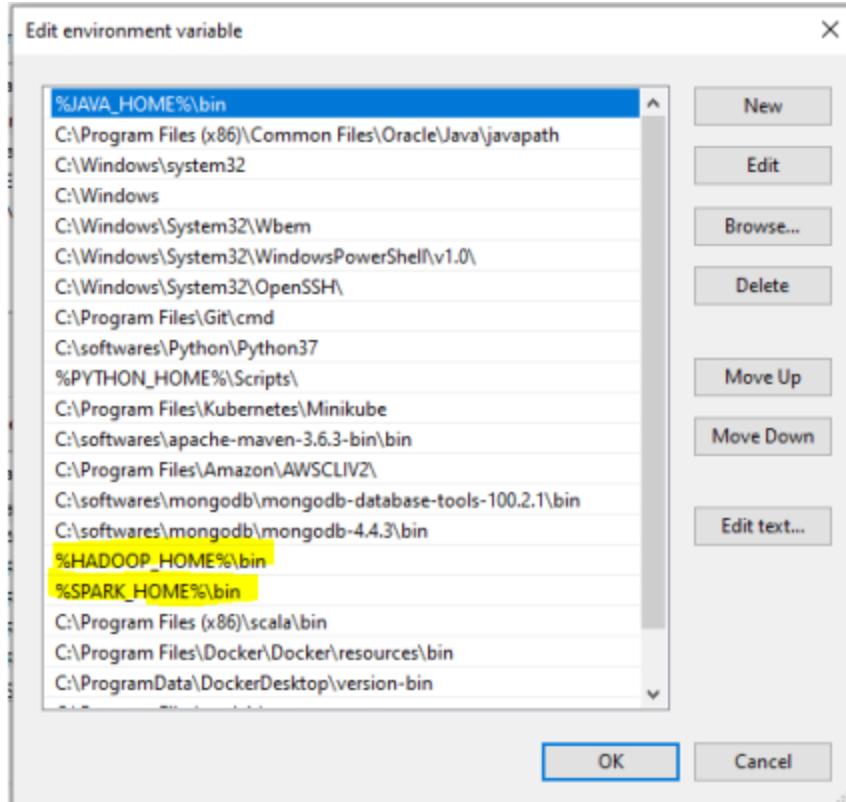
3. Now click on Path Variable under User variables

System variables	
Variable	Value
Path	%JAVA_HOME%\bin;C:\Program Files (x86)\Common Files\Oracle\J...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC
PROCESSOR_ARCHITECTURE	AMD64
PROCESSOR_IDENTIFIER	Intel64 Family 6 Model 42 Stepping 7, GenuineIntel
PROCESSOR_LEVEL	6
PROCESSOR_REVISION	2a07
PSModulePath	%ProgramFiles%\WindowsPowerShell\Modules;C:\Windows\svste...

4. Add the following paths to your PATH variable:

%HADOOP\_HOME%\bin

%SPARK\_HOME%\bin



Close the environment variable screen and the control panels.

### Test for correct Installation

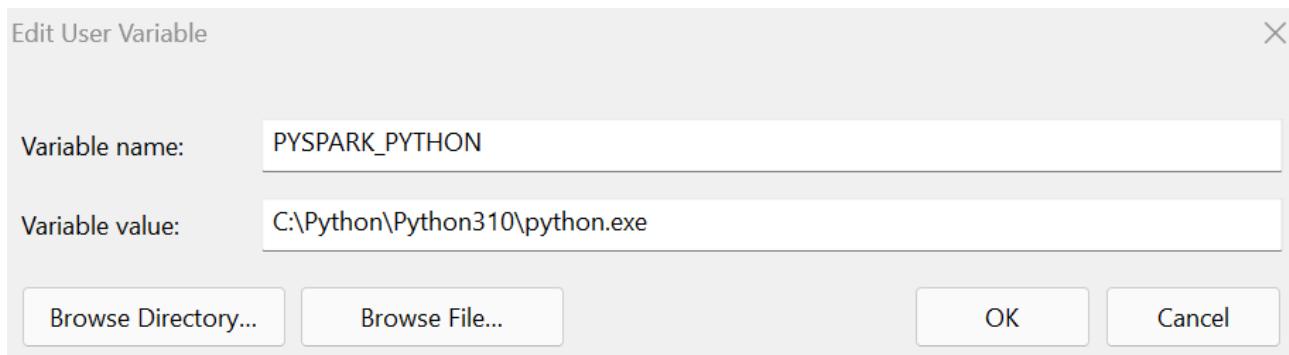
1. Close all the command prompt if opened
2. Open up windows command prompt
3. And type spark-shell

# PySpark Installation

1. Add the following new User variables Click New and add below mentioned details

## Configuring PySpark

- Variable Name : PYSPARK\_PYTHON
  - Variable Value : C:\Python\Python310\python.exe



- ## 2. Open Commands prompt

```
$> python.exe -m pip install --upgrade pip  
$> pip install pyspark==3.2.4
```

## Test for correct Installation

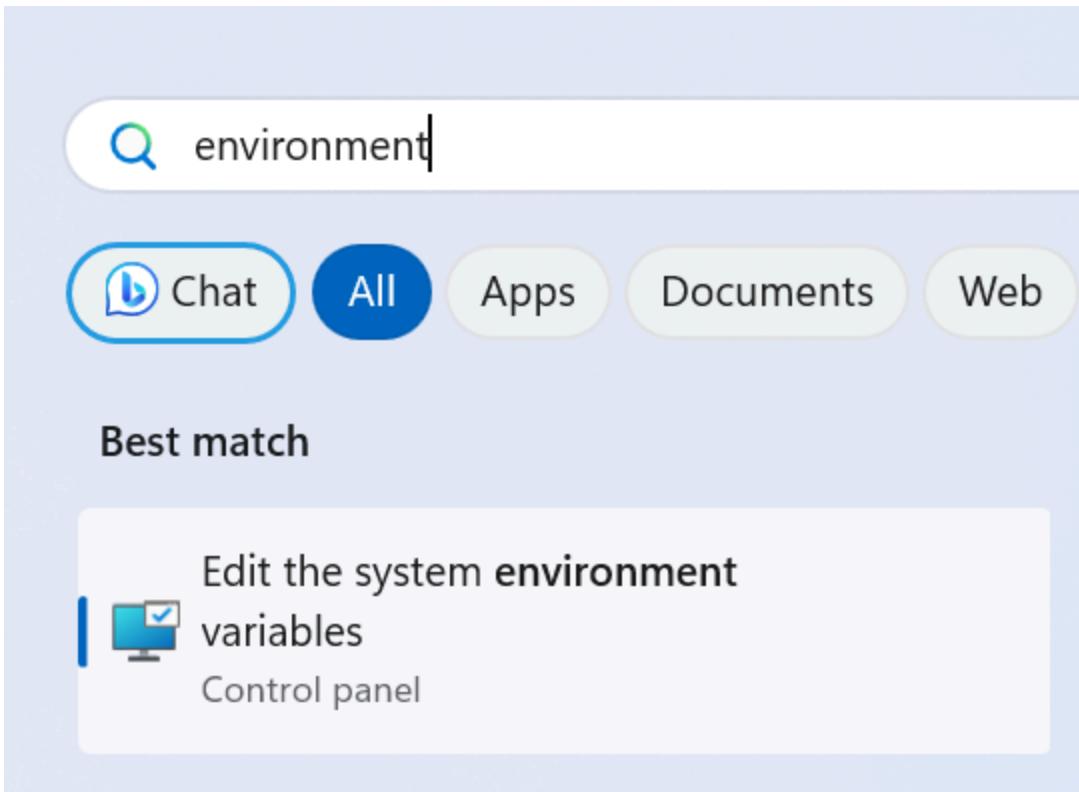
```
C:\Users\npnnp>pyspark  
Python 3.10.9 (tags/v3.10.9:1dd9be6, Dec 6 2022, 20:01:21) [MSC v.1934 64 bit (AMD64)] on win32  
Type "help", "copyright", "credits" or "license" for more information.  
WARNING: An illegal reflective access operation has occurred  
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/C:/big-data-masters-programme-spark/spark-3.2.4-bin-hadoop2.7/jars/spark-unsafe_2.12-3.2.4.jar) to constructor java.nio.DirectByteB  
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform  
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations  
WARNING: All illegal access operations will be denied in a future release  
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties  
Setting default log level to "WARN".  
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).  
23/07/16 09:16:25 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using  
asses where applicable  
Welcome to  
    /---/  
   _\ \_\_ \_\_ \_\_ / /  
  /__ / .__/\_\_/_/ /_\ \_\  version 3.2.4  
  /_  
  
Using Python version 3.10.9 (tags/v3.10.9:1dd9be6, Dec 6 2022 20:01:21)  
Spark context Web UI available at http://host.docker.internal:4040  
Spark context available as 'sc' (master = local[*], app id = local-1689479186368).  
SparkSession available as 'spark'.
```

## Kafka Installation

1. Download [confluent-7.6.1.zip](#) to C:\data-engineering-env-setup\
2. Right Click on apache-spark.zip , Select 7-Zip and select Extract Here

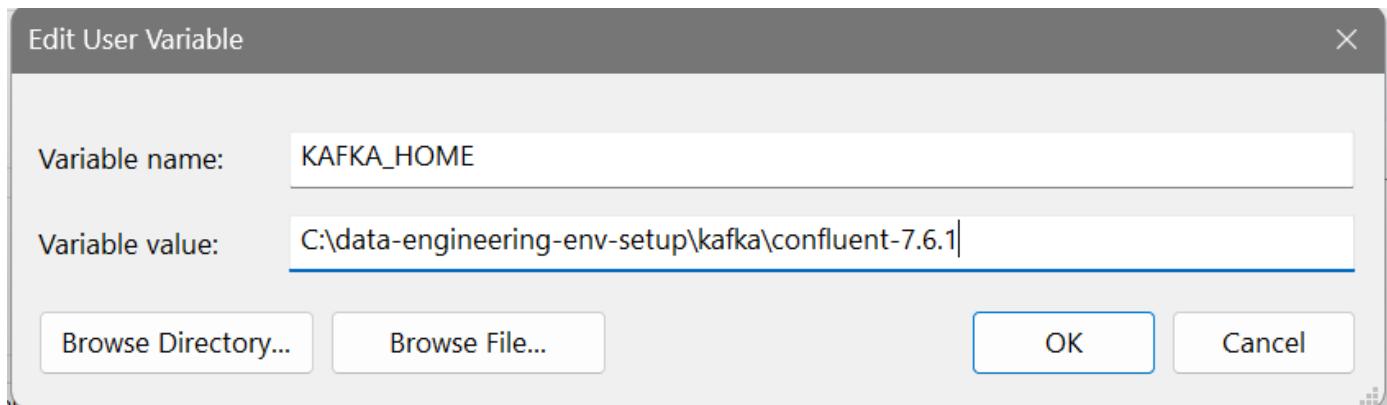
## Configure Environment variables

5. Press the Windows key and search for **Environment** and click on **Edit the system environment variables.**



Click the Advanced tab, → Click the Environment Variables button

6. Under User variables click **New** and add the following



7. Under User variables search for Path variable and click Edit

Add the following

**%KAFKA\_HOME%\bin\windows\**

## Edit environment variable

X

C:\Python\Python310\Scripts\  
C:\Python\Python310\  
%JAVA\_HOME%\bin  
%USERPROFILE%\AppData\Local\Microsoft\WindowsApps  
C:\Users\npnnnp\AppData\Local\Programs\Azure Data Studio\bin  
%CASSANDRA\_HOME%\bin  
%HADOOP\_HOME%\bin  
%SPARK\_HOME%\bin  
C:\Program Files (x86)\Nmap  
C:\Users\npnnnp\AppData\Local\Microsoft\WinGet\Packages\Da...  
**%KAFKA\_HOME%\bin\windows\**  
C:\Users\npnnnp\AppData\Local\Programs\Microsoft VS Code\bin  
C:\Users\npnnnp\AppData\Local\Google\Cloud SDK\google-clo...  
%FLINK\_HOME%\bin  
C:\Users\npnnnp\AppData\Roaming\npm  
C:\Users\npnnnp\AppData\Local\Microsoft\WinGet\Packages\As...

New

Edit

Browse...

Delete

Move Up

Move Down

Edit text...

OK

Cancel