**Assumption**: make sure you have the Java 8 JDK installed

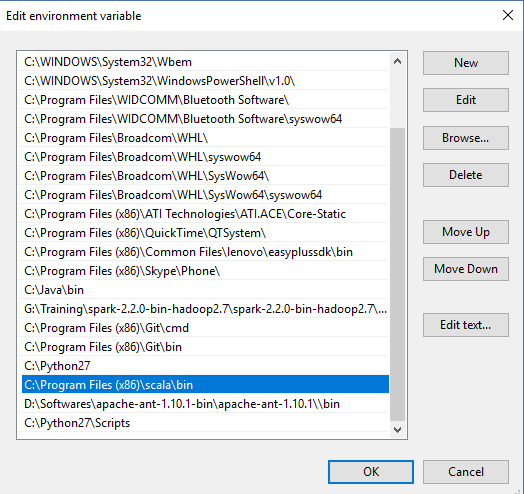
**Install Scala/Spark on Local**

Step1: Download latest version of scala from **https://scala-lang.org/download/** then as per your OS install the version.

Step2: For windows, msi installer will be doing all things Also for Linux/Ubuntu, rpm will be doing the same.

Step3: For windows follow

* go to my computer/PC and left click on that.
* select properties --> Advance System Settings -> Advance tab --> Environment variables --> System variables --> path
* Edit the path and click on new.
* After this, provide path of scala till bin. for ex (C:\Program Files (x86)\scala\bin)



* Say ok multiple times until all windows are closed.

Step 4: Go to command prompt (run) and type "scala".

Step 5: Scala prompt will open, installation **successful**.

Step 6: After Scala, let’s start with Spark. Download Spark from below;

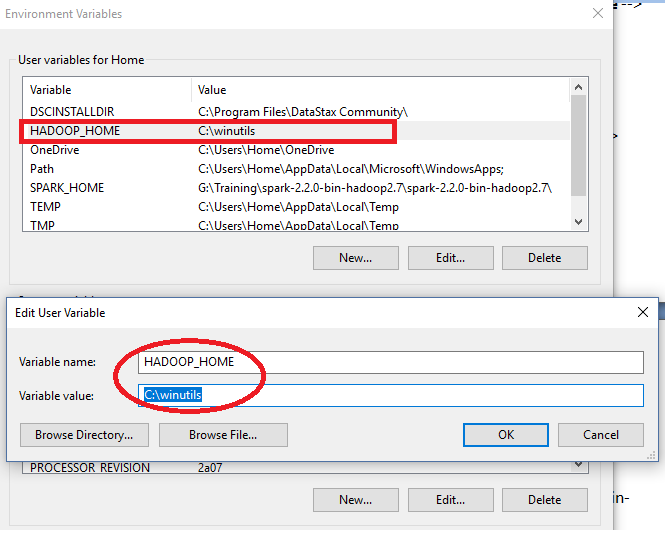
https://spark.apache.org/downloads.html

Step 7: Extract the folder downloaded.

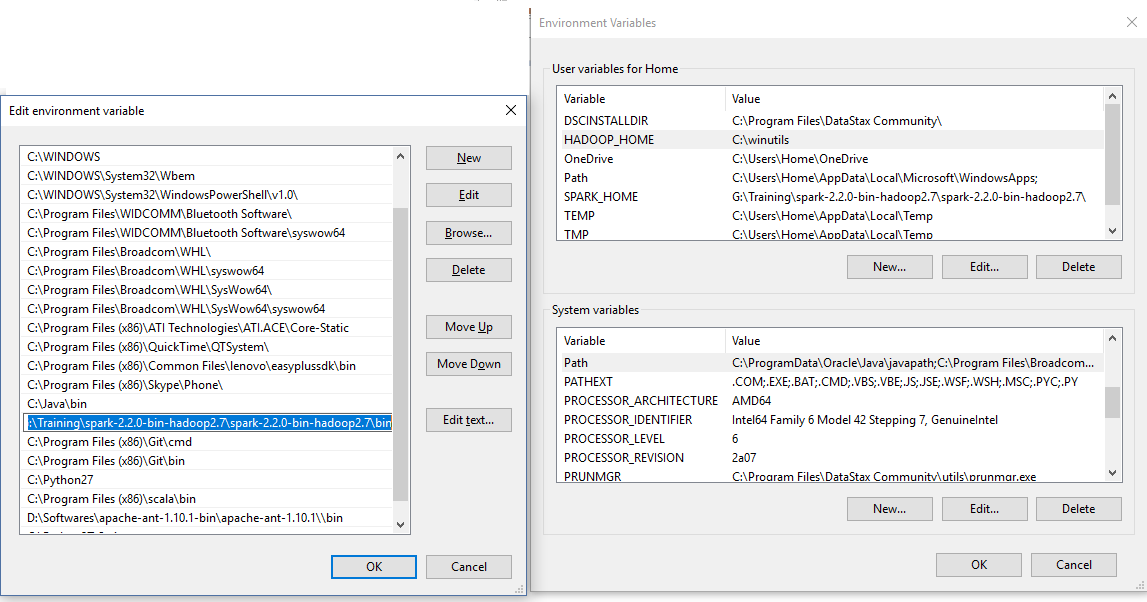
Step 8: Spark has dependency on winutils. So, copy and paste the **winutils** folder from **Google drive** --> Spark Installations and extract it locally.

Step 9: For windows follow

* Go to my computer/PC and left click on that.
* Select properties --> Advance System Settings -> Advance tab --> Environment variables --> System variables --> path
* Edit the path and click on new "user variables for Home" , add new variable as shown below :



* After this, provide path of spark till bin. for ex (C:\Program Files (x86)\spark\spark-2.2.0-bin-hadoop2.7\bin) as below :



* Say ok multiple times until all windows are closed.

Step 9: Open command prompt, and say “**Spark-shell**” to play with scala version of Spark or “**pyspark**” to play with Python.

SPARK FOR MAC

**Steps to install PySpark on Mac OS using Homebrew**

* Step 1 – Install Homebrew
* Step 2 – Install Java
* Step 3 – Install Scala (Optional)
* Step 4 – Install Python
* Step 5 – Install PySpark
* Step 6 – Start PySpark shell and Validate Installation

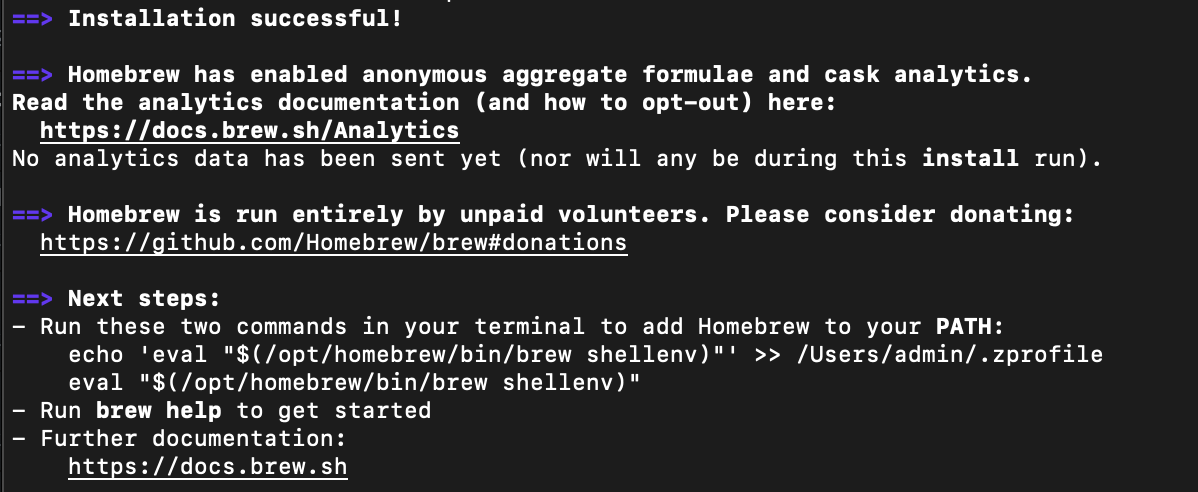
**1. Install PySpark on Mac using Homebrew**

Homebrew is a **Missing Package Manager for macOS (or Linux)** that is used to install third-party packages like Java, PySpark on Mac OS. In order to use this, first, you need to install it by using the below command.

# Install Homebrew

/bin/bash -c "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/HEAD/install.sh)"

This prompts for the root password. You will need to type your root password to run this command. On a personal laptop, this is the same password you enter when you log into your Mac. If you don’t have root access, contact your system admin. You should see something like this below after the successful installation of homebrew.



Post-installation, you may need to run the below command to set the brew to your $PATH.

# Set brew to Path

echo 'eval "$(/opt/homebrew/bin/brew shellenv)"' >> /Users/admin/.zprofile

eval "$(/opt/homebrew/bin/brew shellenv)"

## 2. Install Java Version

PySpark uses Java underlying hence you need to have Java on your Mac. Since Java is a third party, you can install it using the Homebrew command brew. Since Oracle Java is not open source anymore, I am using the OpenJDK version 11. Run the below command in the terminal to install it.

# Install OpenJDK 11

brew install openjdk@11

**3. Install Scala**

Since Spark is written in Scala language it is obvious you would need Scala to run Spark programs however to run PySpark this is optional.

# Install Scala (optional)

brew install scala

**4. Install Python**

As you would know, PySpark is used to run Spark jobs in Python hence you also need Python to install on Mac OS. let’s install it by using Homebrew. If you already have latest then ignore this step.

# install Python

brew install python

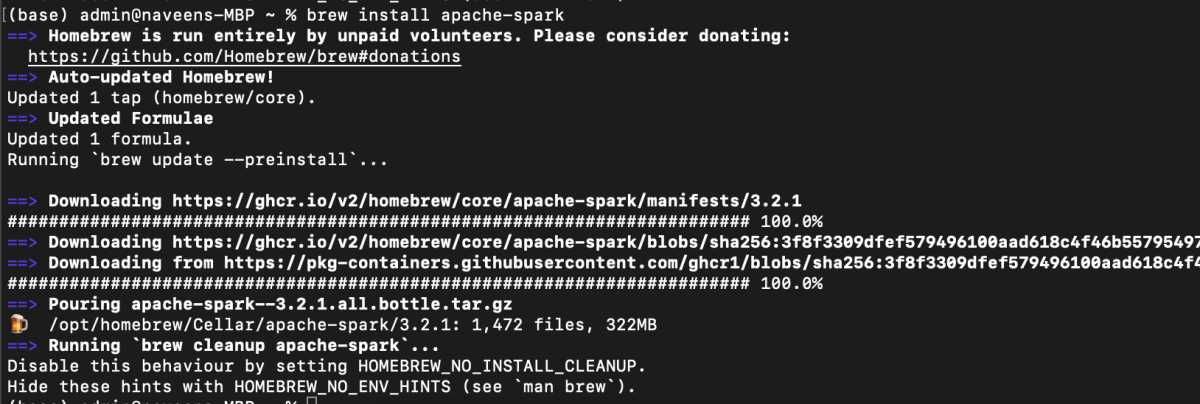
## 5. Install PySpark on Mac

PySpark is a Spark library written in Python to run Python applications using Apache Spark capabilities. Spark was basically written in Scala and later on due to its industry adaptation its API PySpark was released for Python using Py4J. Py4J is a Java library that is integrated within PySpark and allows python to dynamically interface with JVM objects, hence to run PySpark you also need Java to be installed along with Python, and Apache Spark. So to use PySpark, let’s install PySpark on Mac.

# Install Apache Spark

brew install apache-spark

This installs the latest version of Apache Spark which ideally includes PySpark.



After successful installation of Apache Spark run pyspark from the command line to launch PySpark shell.