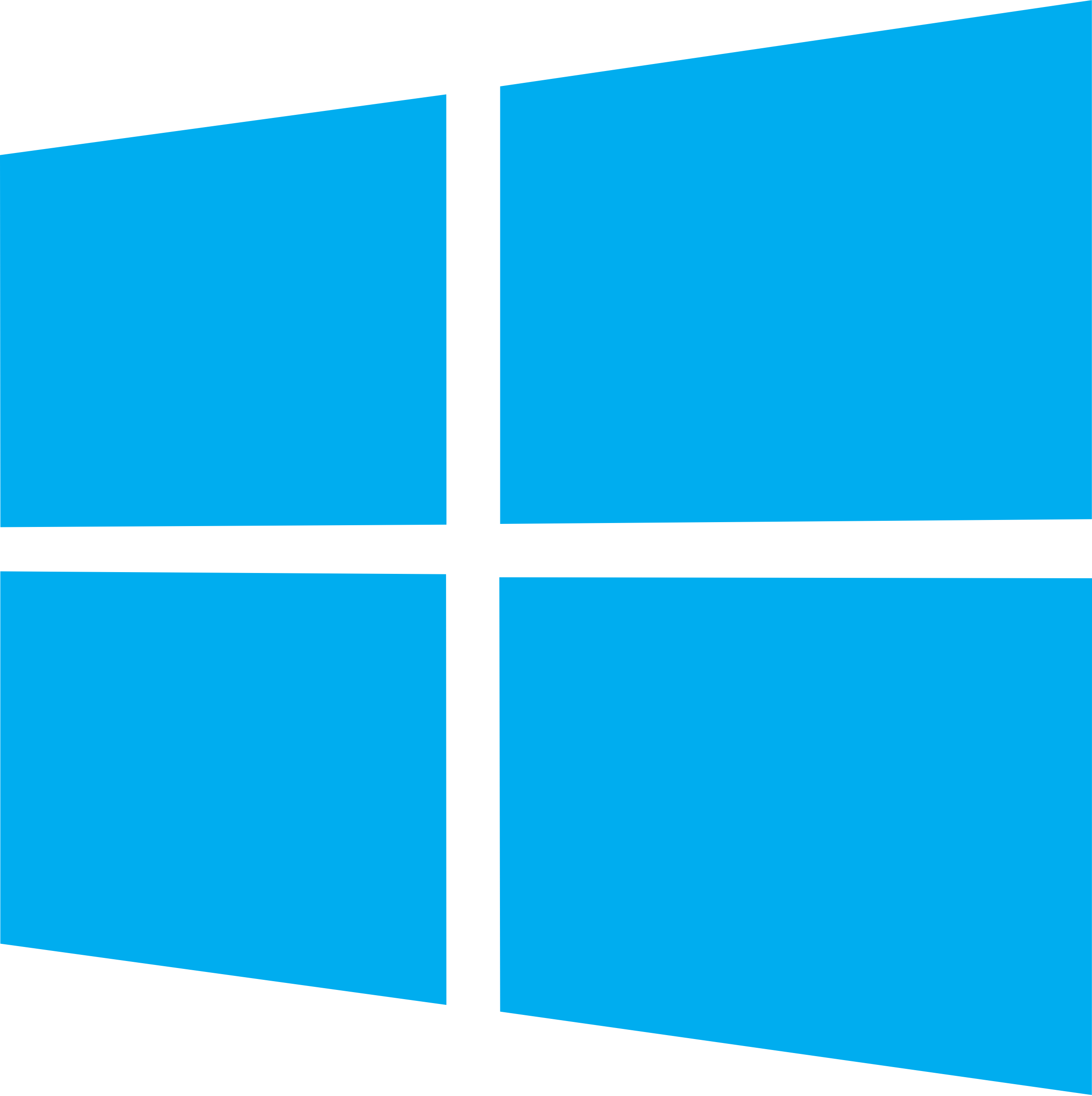
**on**

Prerequisite: Java

**1.Download winutils file**

**link** <https://github.com/steveloughran/winutils/blob/master/hadoop-2.7.1/bin/winutils.exe>

**2. Download spark**

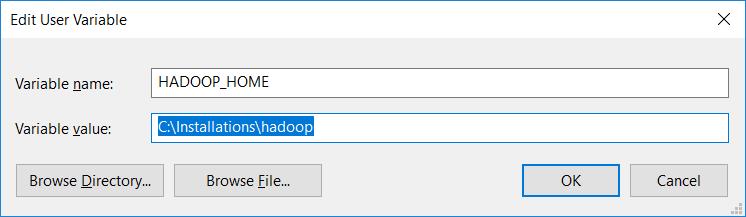
**link** <https://spark.apache.org/downloads.html>

**3. Create folder in D drive. Inside it create another folder** *Installations***.**

Place winutils in a directory D:\Installations\Hadoop\bin

**4. Unzip the spark in the D:\Installations directory**

Path of spark directory should be D:\Installations\spark-2.4.3

**5. Now set HADOOP\_HOME = C:\Installations\Hadoop environment variables.**

**6. Install Pyspark in python environment**

*pip install pyspark*

**7. Test the spark installation in python interpreter or jupyter notebook**

*from pyspark import SparkContext*

*sc = SparkContext()*

**8. Run spark from CMD spark-shell**

*cd /d D:\Installations\spark-2.4.3-bin-hadoop2.7\bin*

*spark-shell*

**References:**

1. <https://dzone.com/articles/working-on-apache-spark-on-windows>
2. <https://medium.com/@josemarcialportilla/getting-spark-python-and-jupyter-notebook-running-on-amazon-ec2-dec599e1c297>
3. <https://medium.com/big-data-engineering/how-to-install-apache-spark-2-x-in-your-pc-e2047246ffc3>