1. Install a Java Development Kit (JDK) (remember Hadoop installation) from http://www.oracle.com/technetwork/java/javase/downloads/index.html . **You must install the JDK into a path with no spaces**, for example c:\jdk. Be sure to change the default location for the installation! **DO NOT INSTALL JAVA 9 – INSTALL JAVA 8**. Spark is not compatible with Java 9.

2. Download a **pre-built** version of Apache Spark from https://spark.apache.org/downloads.html

3. If necessary, download and install WinRAR so you can extract the .tgz file you downloaded. http://www.rarlab.com/download.htm

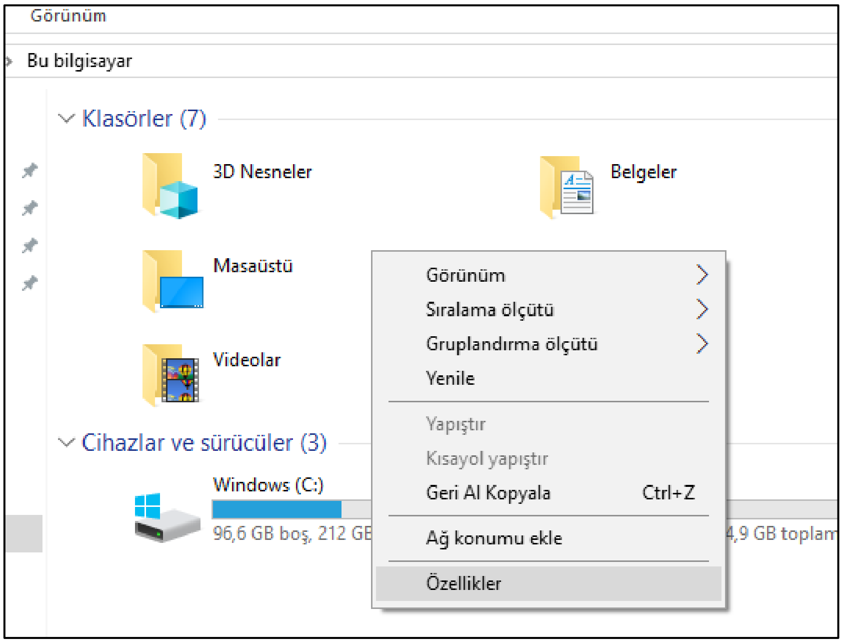
4. Extract the Spark archive, and copy its **contents** into **C:\spark** after creating that directory. You should end up with directories like c:\spark\bin, c:\spark\conf, etc.

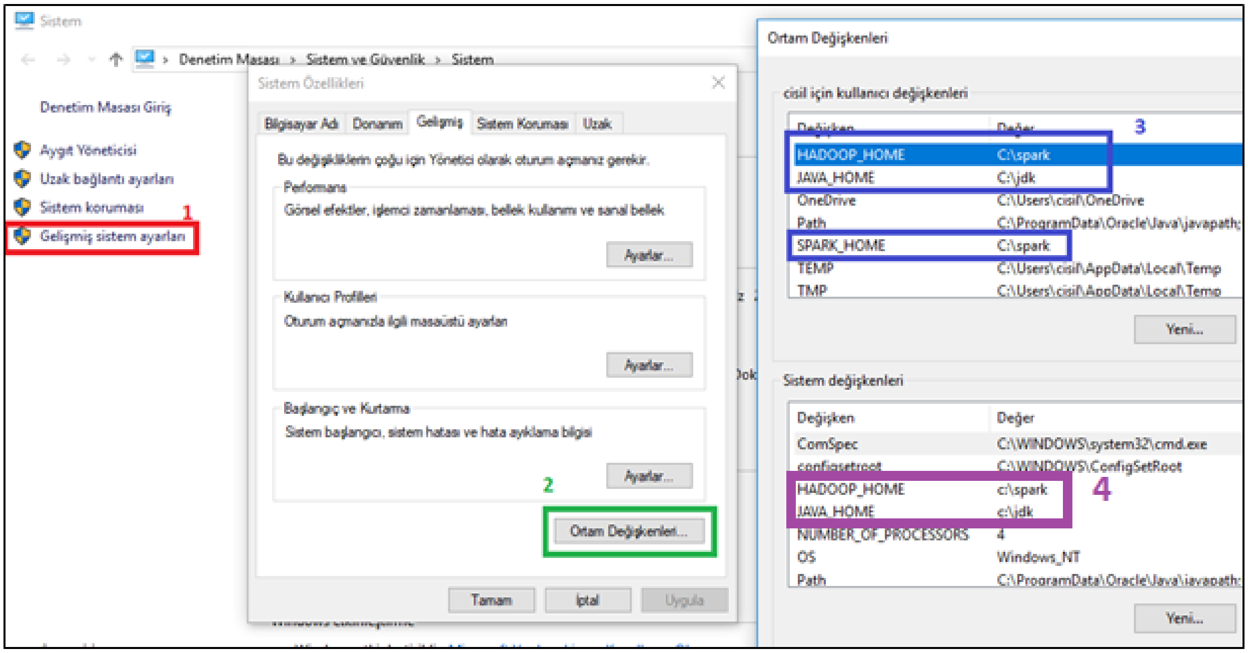
5. Download winutils.exe from our google drive directory (tutorial/installation_files) and move it into a **C:\spark\bin** folder that you've created. (note, this is a 64-bit application. If you are on a 32-bit version of Windows, you'll need to search for a 32-bit build of winutils.exe for Hadoop.)

6. Open the the **c:\spark\conf** folder, and make sure "File Name Extensions" is checked in the "view" tab of Windows Explorer. Rename the log4j.properties. template file to log4j.properties. Edit this file (using Wordpad or something similar) and change the error level from INFO to ERROR for log4j.rootCategory

7. Right-click your Windows menu, select Control Panel, System and Security, and then System. Click on "Advanced System Settings" and then the "Environment Variables" button.

8. Mapping should be done as follow:
Please be sure that you did for mapping both user and system.





Bu değişikliklerin çoğu için Yönetici olarak oturum açmanız gerekir.

Performans

Görsel efektler, işlemci zamanlaması, bellek kullanımı ve sanal bellek

Kullanıcı Profilleri

Oturum açmanızla ilgili masaüstü ayarları

Başlangıç ve Kurtarma

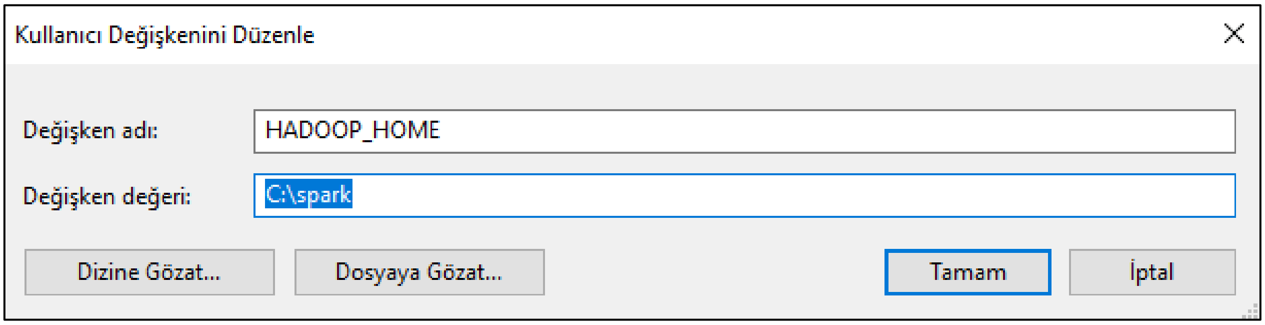
Sistem başlangıcı, sistem hatası ve hata ayıklama bilgisi

cisil için kullanıcı değişkenleri

Değişken	Değer
HADOOP_HOME	C:\spark
JAVA_HOME	C:\jdk
OneDrive	C:\Users\cisil\OneDrive
Path	C:\ProgramData\Oracle\Java\javapath;
SPARK_HOME	C:\spark
TEMP	C:\Users\cisil\AppData\Local\Temp
TMP	C:\Users\cisil\AppData\Local\Temp

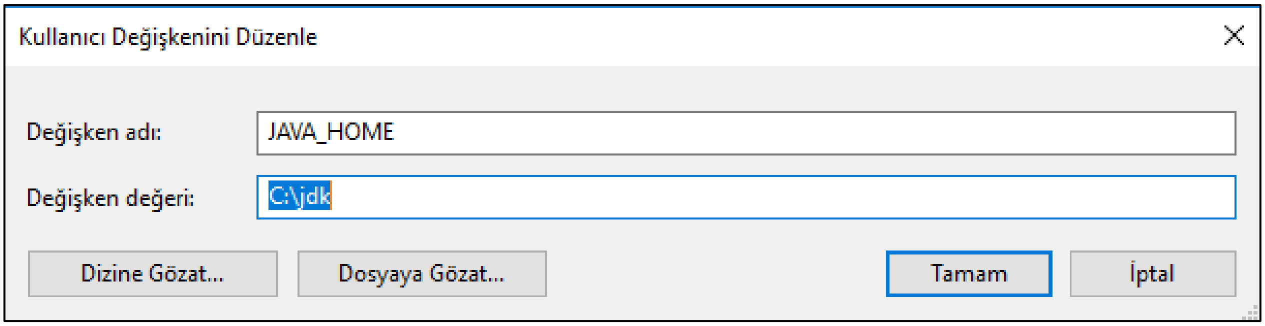
Sistem değişkenleri

Değişken	Değer
ComSpec	C:\WINDOWS\system32\cmd.exe
configsetroot	C:\WINDOWS\ConfigSetRoot
HADOOP_HOME	c:\spark
JAVA_HOME	c:\jdk
NUMBER_OF_PROCESSORS	4
OS	Windows_NT
Path	C:\ProgramData\Oracle\Java\javapath;



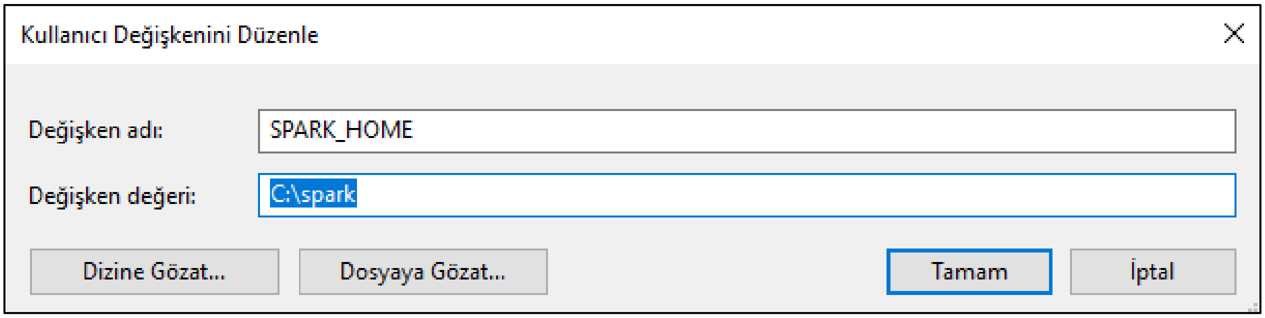
Değişken adı: HADOOP_HOME

Değişken değeri: C:\spark



Değişken adı: JAVA_HOME

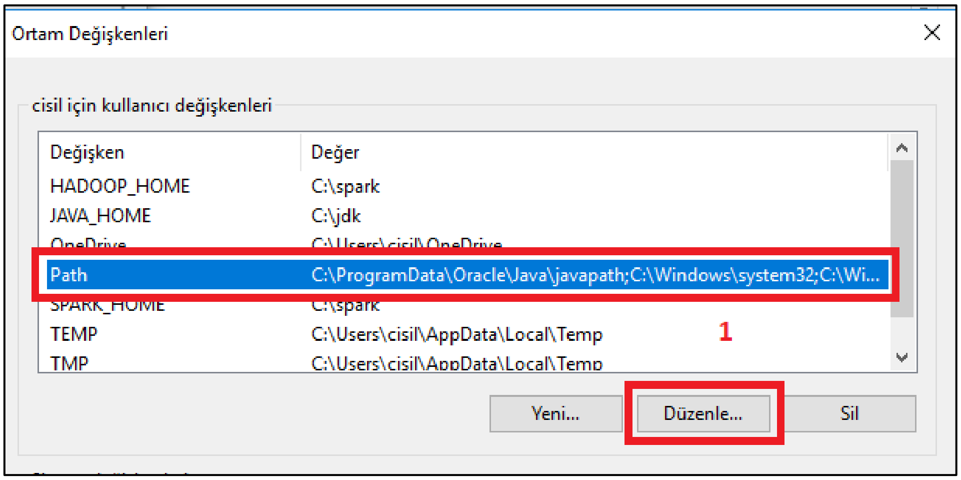
Değişken değeri: C:\jdk



Değişken adı: SPARK_HOME

Değişken değeri: C:\spark

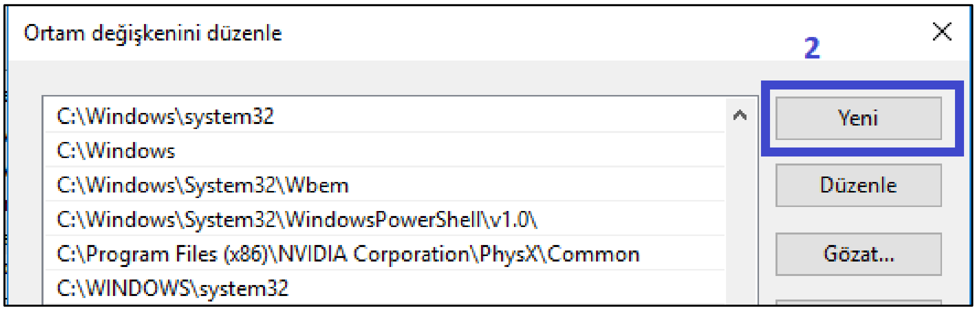
9. After mapping you should link to these maps to the path. Adding paths should be done as follow: It should be done both user and system.



Ortam Değişkenleri

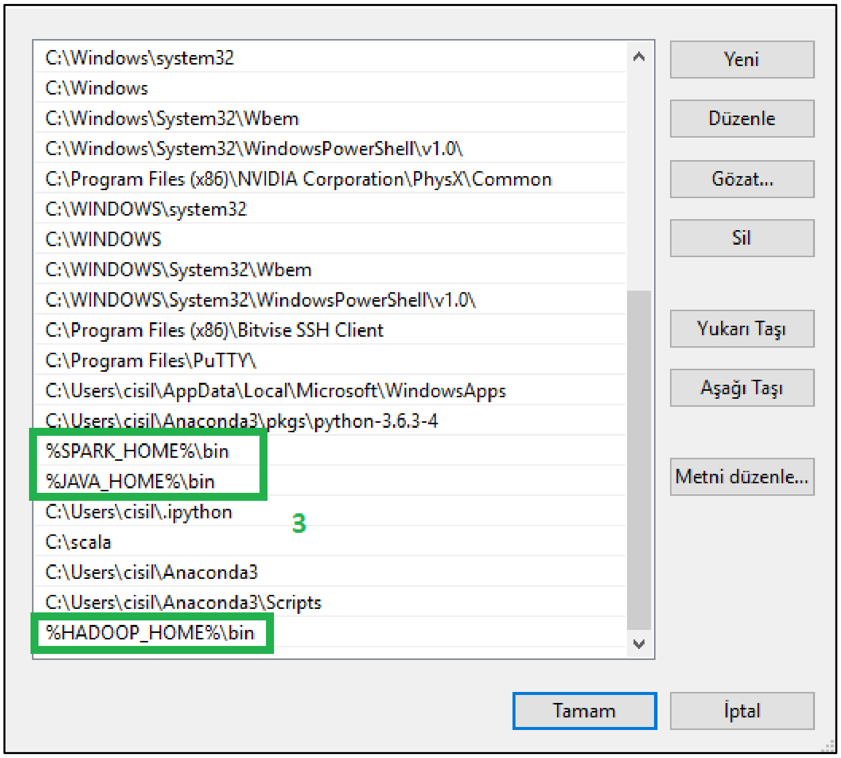
cisil için kullanıcı değişkenleri

Değişken	Değer
HADOOP_HOME	C:\spark
JAVA_HOME	C:\jdk
OneDrive	C:\Users\cisil\OneDrive
Path	C:\ProgramData\Oracle\Java\javapath;C:\Windows\system32;C:\Wi...
SPARK_HOME	C:\spark
TEMP	C:\Users\cisil\AppData\Local\Temp
TMP	C:\Users\cisil\AppData\Local\Temp



Ortam değişkenini düzenle

```
C:\Windows\system32
C:\Windows
C:\Windows\System32\Wbem
C:\Windows\System32\WindowsPowerShell\v1.0\
C:\Program Files (x86)\NVIDIA Corporation\PhysX\Common
C:\WINDOWS\system32
```



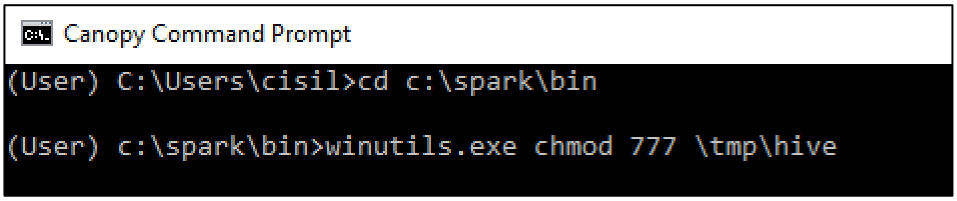
```
C:\Windows\system32
C:\Windows
C:\Windows\System32\Wbem
C:\Windows\System32\WindowsPowerShell\v1.0\
C:\Program Files (x86)\NVIDIA Corporation\PhysX\Common
C:\WINDOWS\system32
C:\WINDOWS
C:\WINDOWS\System32\Wbem
C:\WINDOWS\System32\WindowsPowerShell\v1.0\
C:\Program Files (x86)\Bitvise SSH Client
C:\Program Files\PuTTY\
C:\Users\cisil\AppData\Local\Microsoft\WindowsApps
C:\Users\cisil\Anaconda3\pkgs\python-3.6.3-4
%SPARK_HOME%\bin
%JAVA_HOME%\bin
C:\Users\cisil\.ipython
C:\scala
C:\Users\cisil\Anaconda3
C:\Users\cisil\Anaconda3\Scripts
%HADOOP_HOME%\bin
```

10. Close the environment variable screen and the control panels.

11. Install the latest **Enthought Canopy for Python 3.5** from https://store.enthought.com/downloads/#default
Don't install a Python 2.7 version!

12. If you still have problems, you can try the command below before pyspark command.

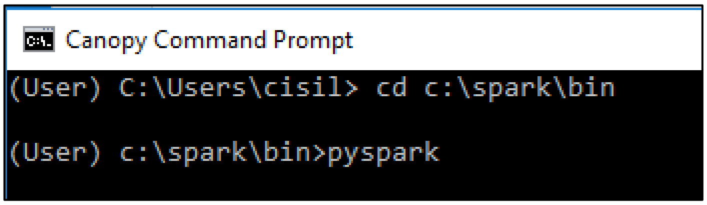
C:\spark\bin>winutils.exe chmod 777 \tmp\hive



```
(User) C:\Users\cisil>cd c:\spark\bin

(User) c:\spark\bin>winutils.exe chmod 777 \tmp\hive
```

13. To run pyspark you write cd c:\spark\bin then write pyspark as can be seen below:



```
(User) C:\Users\cisil> cd c:\spark\bin

(User) c:\spark\bin>pyspark
```

14. Needs to be checked whether is properly installed or not!

- Open up Canopy and select "Canopy Command Prompt" from the Tools menu.
- Enter **cd c:\spark\bin** and then **dir** to get a directory listing.
- Look for a text file we can play with, like README.md or CHANGES.txt
- Enter **pyspark**
- At this point you should have a >>> prompt. If not, double check the steps above.
- Enter **rdd = sc.textFile("README.md")** (or whatever text file you've found) Enter **rdd.count()**
- You should get a count of the number of lines in that file! Congratulations, you just ran your first Spark program!
- Enter **quit()** to exit the spark shell, and close the console window
- It seems the installation is fine ;)