



- Télécharger Spark

The Google logo, featuring the word "Google" in its characteristic multi-colored font.

apache spark

Recherche Google

J'ai de la chance



- Télécharger Spark

A screenshot of a Google search for 'apache spark'. The search bar shows 'apache spark' and the results are filtered to 'All'. The first result is an advertisement from Talend.com titled 'Start with Apache Spark - Apache Spark - talend.com'. Below it is the official Apache Spark result with the title 'Apache Spark™ - Lightning-Fast Cluster Computing' and the URL 'https://spark.apache.org/'. The snippet describes Spark as a fast and general engine for big data processing. At the bottom, there is a link to the 'Documentation | Apache Spark' page.

Google

apache spark

All News Images Videos Books More Settings Tools

About 18,600,000 results (0.63 seconds)

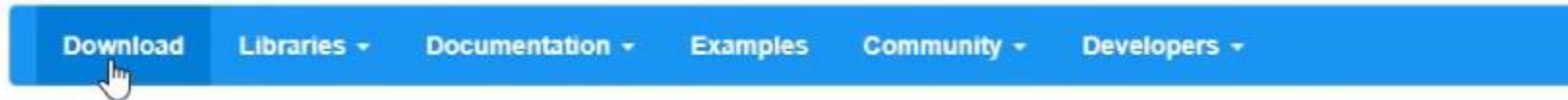
Start with Apache Spark - Apache Spark - talend.com
info talend.com/Spark
Accelerate Real-Time Analytics with Spark. Watch Free Talend Webinar!
Multi-Cloud Integration · Integrate Data Instantly · 900+ Data Connectors · Enable Data Agility
Types: Data Integration, Big Data, Data Preparation, Data Quality, Master Data Management, ESB
Master Data Management Talend ETL
White Papers Big Data Integration

Apache Spark™ - Lightning-Fast Cluster Computing
<https://spark.apache.org/>
Apache Spark is a fast and general engine for big data processing, with built-in modules for streaming, SQL, machine learning and graph processing.
Apache Spark 2.2.0 · Spark Streaming · Spark SQL · Contributing to Spark

Documentation | Apache Spark
<https://spark.apache.org/documentation.html>
Apache Spark Documentation: Setup instructions, programming guides, and other documentation are available for each stable version of Spark below.



- Télécharger Spark



Download Apache Spark™

1. Choose a Spark release:
2. Choose a package type:
3. Download Spark: [spark-2.2.0-bin-hadoop2.7.tgz](#)
4. Verify this release using the [2.2.0 signatures and checksums](#) and [project release KEYS](#).

Note: Starting version 2.0, Spark is built with Scala 2.11 by default. Scala 2.10 users should download the Spark source package and



- Télécharger Spark



Choisir la version à télécharger, puis cliquer sur le lien



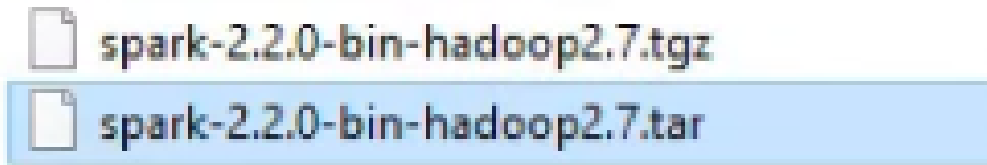
- Télécharger Spark



cliquer sur le lien miroir pour commencer le téléchargement



- Une fois le téléchargement est terminé

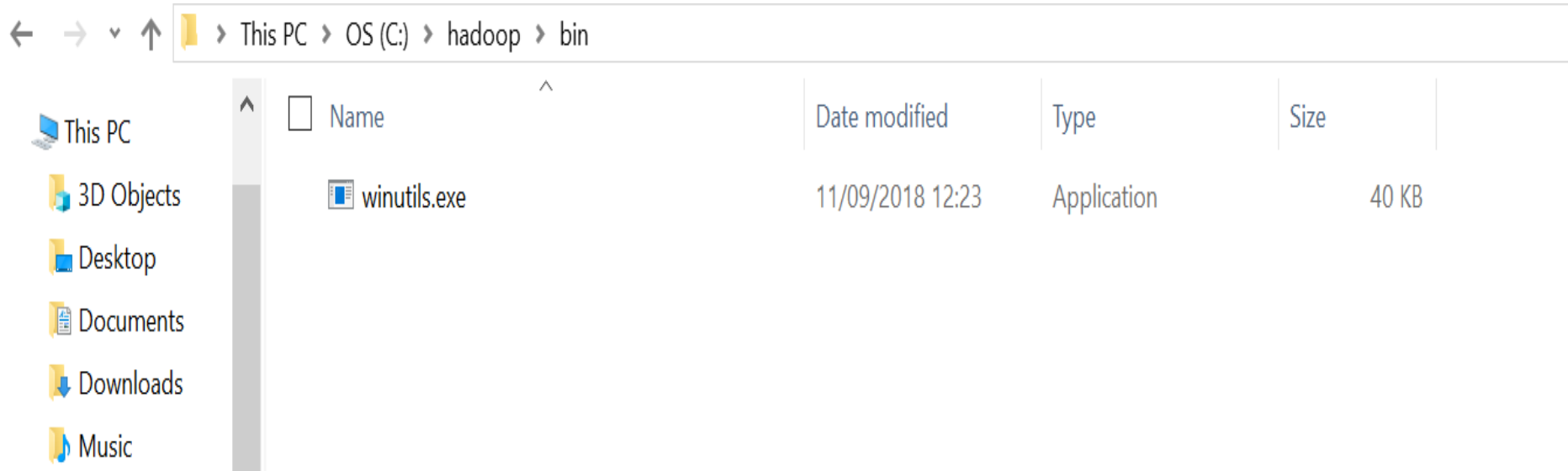


- Extraire le fichier vers un fichier .tar ou rar
 - Créer un dossier appelé « Apache-spark » sur votre partition principale (c:)
 - Extraire le fichier vers ce dossier (i.e., Apache-spark)

Installation de Spark sur windows

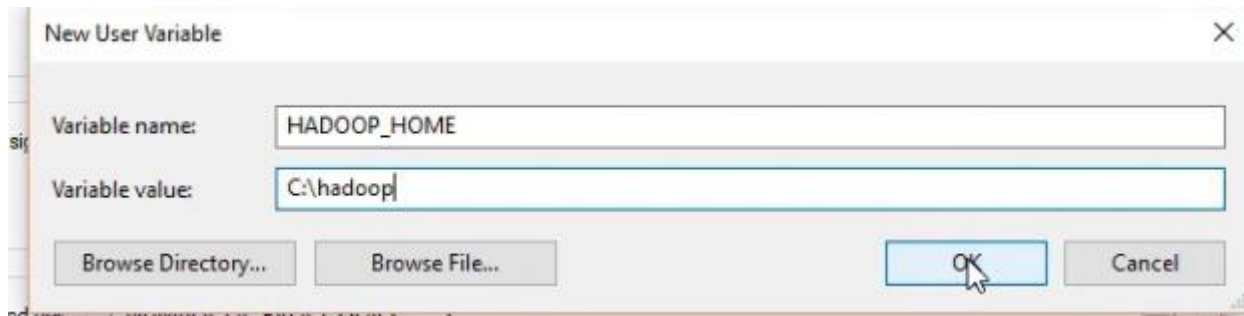


- Télécharger winutils.exe (la seule différence entre les autres SE à cause des fichiers système NTFS de windows)
- créer un dossier dans la partition principale nommé « hadoop »
- Créer dedans un autre dossier nommé « bin »
- Copier winutils.exe dans ce dossier

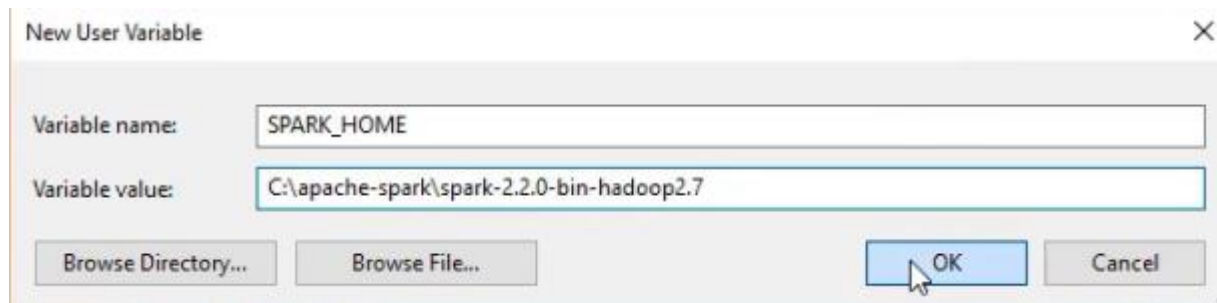




- Créer une variable d'environnement appelé « HADOOP_HOME » ayant comme valeur c:\hadoop (le dossier que vous venez de créer)



- Créer une variable d'environnement appelé « SPARK_HOME » ayant comme valeur « C:\Apache-spark\spark-2.3.1-bin-hadoop2.7 » (la distribution binaire de spark se trouvant dans le dossier nommé « C:\Apache-spark »)





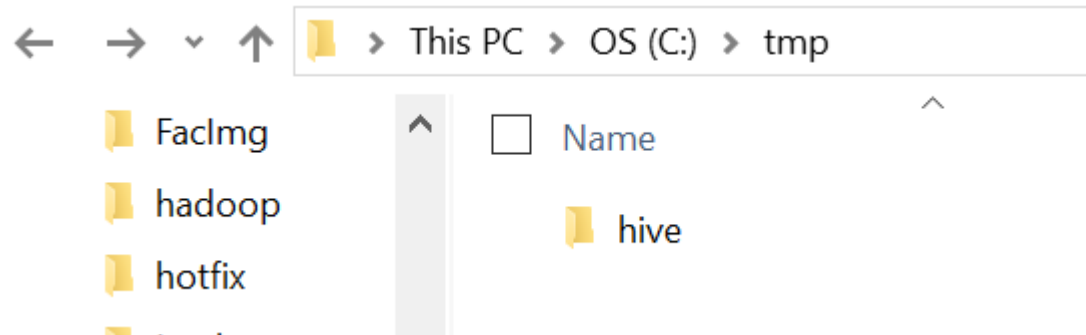
- Ajouter spark et hadoop à la variable PATH
- Cliquer sur « PATH » puis edit et new



- Quitter les fenêtres en cliquant sur « ok »



- Avant de démarrer spark, créer un dossier appelé « tmp » dans lequel créer un autre dossier appelé « hive »



- Hadoop a besoin de ce dossier pour qu'il s'exécute en vitesse

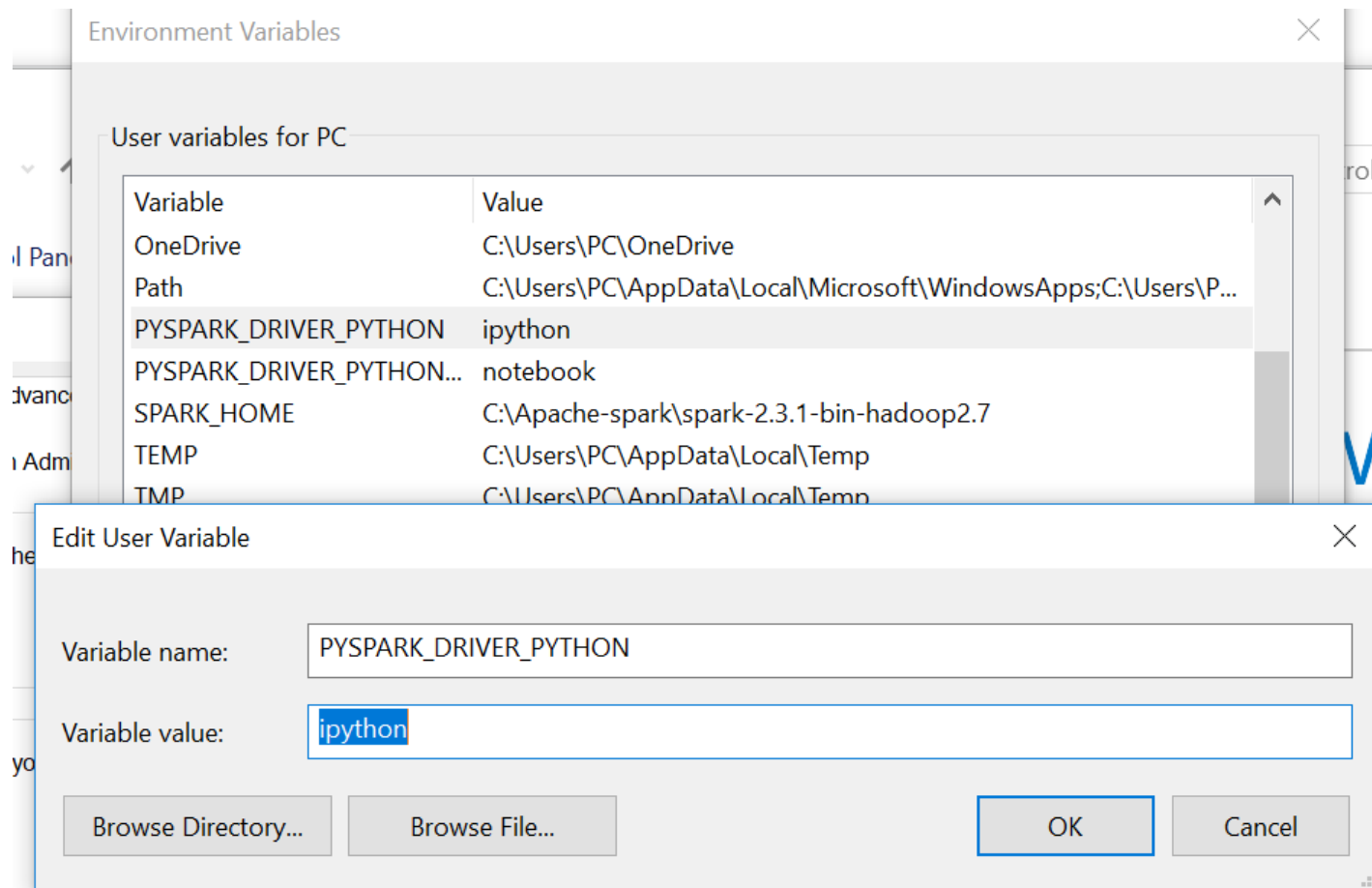


- Sur l'invité de commande taper cette commande

```
C:\> C:\WINDOWS\system32\cmd.exe  
Microsoft Windows [Version 10.0.17134.286]  
(c) 2018 Microsoft Corporation. All rights reserved.  
  
C:\Users\PC>winutils chmod 777 c:\tmp\hive  
  
C:\Users\PC>_
```

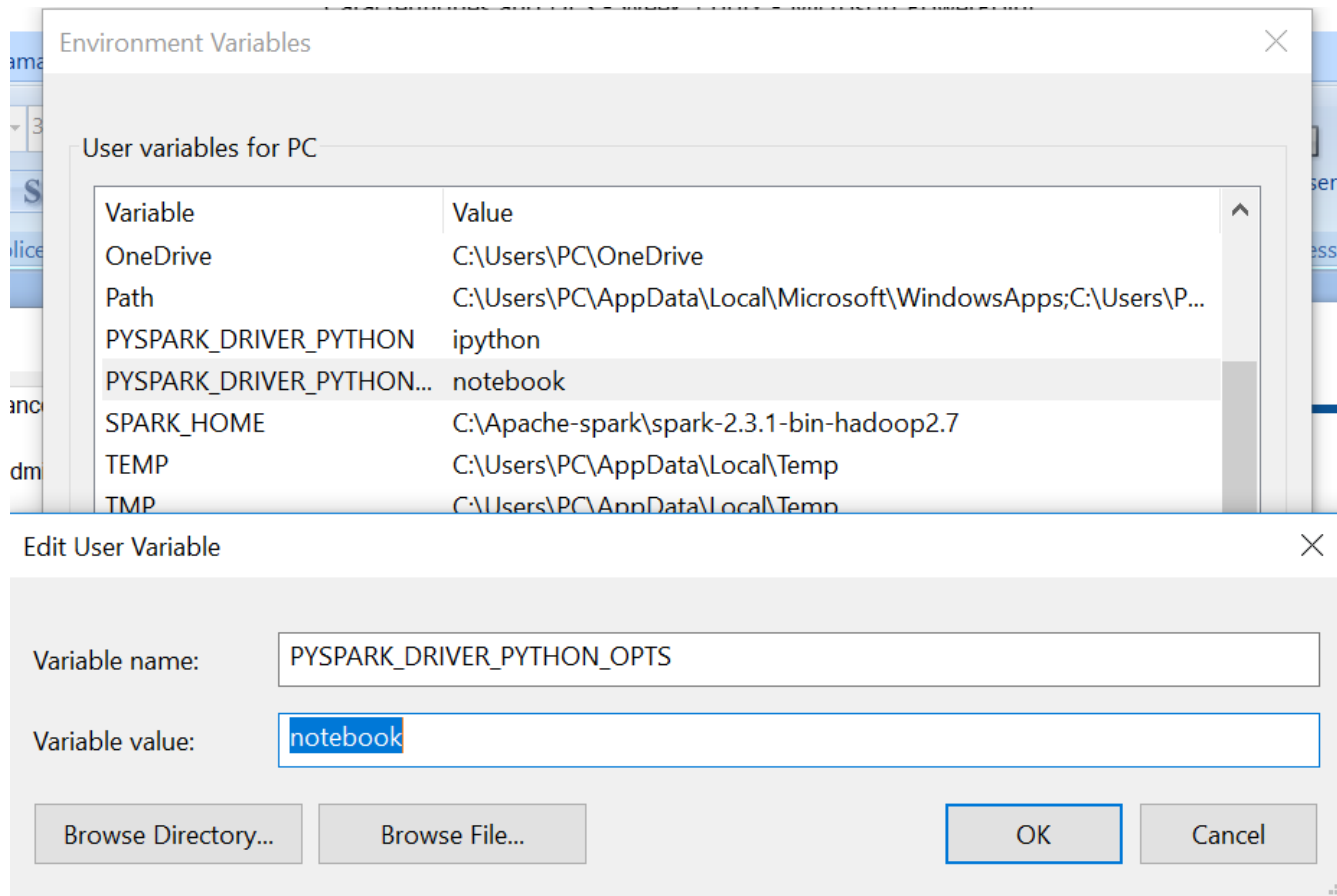


- Pour utiliser spark sur jupyter notbook et pas sur l'invité de commande, ajouter les deux variables d'environnement





- Pour utiliser spark sur jupyter notebook et pas sur l'invité de commande, ajouter les deux variables d'environnement





- Démarrer spark en tapant « pyspark » sur l'invité de commande anaconda

```
Anaconda Prompt - pyspark

(base) C:\Users\PC>spark
'spark' is not recognized as an internal or external command,
operable program or batch file.

(base) C:\Users\PC>pyspark
[I 23:06:49.123 NotebookApp] The port 8888 is already in use, trying another port.
[I 23:06:50.850 NotebookApp] JupyterLab beta preview extension loaded from C:\Users\PC\Anaconda3\lib\site-packages\jupyterlab
[I 23:06:50.850 NotebookApp] JupyterLab application directory is C:\Users\PC\Anaconda3\share\jupyter\lab
[I 23:06:51.352 NotebookApp] Serving notebooks from local directory: C:\Users\PC
[I 23:06:51.352 NotebookApp] 0 active kernels
[I 23:06:51.352 NotebookApp] The Jupyter Notebook is running at:
[I 23:06:51.367 NotebookApp] http://localhost:8889/?token=c32ec14427a311d25412a0e6ea869c1124d66c2929230a83
[I 23:06:51.367 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 23:06:51.367 NotebookApp]

Copy/paste this URL into your browser when you connect for the first time,
to login with a token:
http://localhost:8889/?token=c32ec14427a311d25412a0e6ea869c1124d66c2929230a83&token=c32ec14427a311d25412a0e6ea869c1124d66c2929230a83
[I 23:06:59.369 NotebookApp] Accepting one-time-token-authenticated connection from ::1
[I 23:14:21.600 NotebookApp] Kernel started: 9f1ac48a-c116-44b6-a7b8-454a531dbedf
[W 23:14:31.663 NotebookApp] Timeout waiting for kernel_info reply from 9f1ac48a-c116-44b6-a7b8-454a531dbedf
2018-10-10 23:15:15 WARN NativeCodeLoader:62 - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```