# Hadoop

## Installation windows

Installer le dernier JDK (1.8) :

<https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

Ajouter le jdk en variable d’environnement :

Variable : JAVA\_HOME

Value : C:\Java\jdk1.8.0\_211

Puis ajouter le jdk dans le path :

C:\Java\jdk1.8.0\_211\bin

Vérifier l’installation de java :

Sur le cmd : java -version

Télécharger Hadoop (le binary)

<https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.1.2/hadoop-3.1.2.tar.gz>

extraire le dossier .gz

Copier le dossier hadoop-3.1.2 à la racine de c

Ajouter le hadoop en variable d’environnement :

Variable : HADOOP\_HOME

Value : C:\hadoop-3.1.2\bin

Puis ajouter le jdk dans le path :

C:\hadoop-3.1.2\bin

C:\hadoop-3.1.2\sbin

Maintenant, on va dans C:\hadoop-3.1.2\etc\hadoop

On va modifier core-site.xml :

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>

</configuration>

On va modifier mapred-site.xml :

<configuration>

<property>

<name>mapreduce.framwork.name</name>

<value>yarn</value>

</property>

</configuration>

Ensuite, dans le dossier hadoop-3.1.2 :

Crée un dossier data, puis, on crée un dossier datanode et namenode dans data

Puis, dans etc, on va modifier le fichier hdfs-site.xml :

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>C:\hadoop-3.1.2\data\namenode</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>C:\hadoop-3.1.2\data\datanode</value>

</property>

</configuration>

Ensuite, on va modifier le fichier yarn-site.xml :

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

Ensuite, on va éditer le fichier hadoop-env.cmd :

set JAVA\_HOME=C:\Java\jdk1.8.0\_211

Maintenant, on part dans C:\hadoop-3.1.2\bin

Copier le contenu du bin de \Hadoop bin files\bin (fichier à télécharger) dans C:\hadoop-3.1.2\bin

Maintenant, dans le cmd, on vas exécuter la commande hdfs namenode -format (L’executer en mode admin si ça marche pas)

Ensuite, nous avons dans C:\hadoop-3.1.2\sbin de nouveaux script shell windows

On as les commandes :

start-all // vas tous démarrer

On va executer la commande start-all.cmd

# Spark

## Installation windows

Télécharger Spark (le binary)

<https://www.apache.org/dyn/closer.lua/spark/spark-2.4.3/spark-2.4.3-bin-hadoop2.7.tgz>

Dézipper le de tel sorte à avoir C:\spark-2.4.3

Ajouter le dossier en variable d’environnement :

Variable : SPARK\_HOME

Value : C:\spark-2.4.3

Puis ajouter le jdk dans le path :

C:\spark-2.4.3\bin

Maintenant on teste :

Sur le cm den mode admin, faire spark-shell