

Installation de Hadoop sous Ubuntu

Étape 1 : Installer Java

```
sudo apt update
sudo apt install openjdk-11-jdk -y
java -version
```

Étape 2 : Télécharger et extraire Hadoop

Accédez à <https://hadoop.apache.org/releases.html>, puis exécutez :

```
cd /usr/local
sudo wget https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar
  ↪ .gz
sudo tar -xvzf hadoop-3.3.6.tar.gz
sudo mv hadoop-3.3.6 hadoop
```

Étape 3 : Configurer les variables d'environnement

Modifier le fichier `.bashrc` :

```
nano ~/.bashrc
```

Ajouter à la fin :

```
export HADOOP_HOME=/usr/local/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
```

Puis exécutez :

```
source ~/.bashrc
```

Étape 4 : Configurer les fichiers Hadoop

1. hadoop-env.sh

```
nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh
```

Ajouter/modifier :

```
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
```

2. core-site.xml

```
nano $HADOOP_HOME/etc/hadoop/core-site.xml
```

À ajouter dans <configuration> :

```
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://localhost:9000</value>
</property>
```

3. hdfs-site.xml

```
nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml
```

Ajouter :

```
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
<property>
  <name>dfs.namenode.name.dir</name>
  <value>file:///usr/local/hadoopdata/hdfs/namenode</value>
</property>
<property>
  <name>dfs.datanode.data.dir</name>
  <value>file:///usr/local/hadoopdata/hdfs/datanode</value>
</property>
```

Créer les répertoires :

```
sudo mkdir -p /usr/local/hadoopdata/hdfs/namenode
sudo mkdir -p /usr/local/hadoopdata/hdfs/datanode
sudo chown -R $USER:$USER /usr/local/hadoopdata
```

4. mapred-site.xml

```
cp $HADOOP_HOME/etc/hadoop/mapred-site.xml.template $HADOOP_HOME/etc/hadoop/mapred
  ↪ -site.xml
nano $HADOOP_HOME/etc/hadoop/mapred-site.xml
```

Ajouter :

```
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
```

5. yarn-site.xml

```
nano $HADOOP_HOME/etc/hadoop/yarn-site.xml
```

Ajouter :

```
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
```

Étape 5 : Configuration SSH (optionnelle pour local)

```
sudo apt install openssh-server -y
ssh-keygen -t rsa
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
chmod 0600 ~/.ssh/authorized_keys
ssh localhost
```

Étape 6 : Formater HDFS

```
hdfs namenode -format
```

```
hdfs datanode -format
```

Étape 7 : Démarrer les services Hadoop

```
start-dfs.sh
start-yarn.sh
```

Vérifier les processus :

```
jps
```

Vous devriez voir :

- NameNode
- DataNode
- ResourceManager
- NodeManager

Étape 8 : Interfaces Web

Composant	URL
NameNode UI	http://localhost:9870
ResourceManager	http://localhost:8088

Étape 9 : Tester HDFS

```
hdfs dfs -mkdir /input
hdfs dfs -put $HADOOP_HOME/etc/hadoop/*.xml /input
hdfs dfs -ls /input
```