

# Compte-Rendu TP1 : Installation et Configuration de Hadoop via WSL

DOUA GALAI

Génie Logiciel et Systèmes d'Information

19 Novembre 2025

## Introduction

Ce TP a pour objectif d'installer et configurer un environnement Hadoop fonctionnel sous WSL (Ubuntu), permettant d'exécuter des jobs MapReduce localement.

## 1 Environnement et Installation

### 1.1 Vérification de l'environnement WSL

```
Caused by: java.io.IOException: Input path  
9000/input  
    at org.apache.hadoop.mapreduce.lib.  
dedListStatus(FileInputFormat.java:315)  
    ... 26 more  
doua@GalaiDoua:~$ pwd  
/home/doua  
doua@GalaiDoua:~$ whoami  
doua  
doua@GalaiDoua:~$ lsb_release -a  
No LSB modules are available.  
Distributor ID: Ubuntu  
Description:    Ubuntu 24.04.3 LTS  
Release:        24.04  
Codename:       noble  
doua@GalaiDoua:~$
```

FIGURE 1 – Environnement Ubuntu 24.04.3 LTS sous WSL

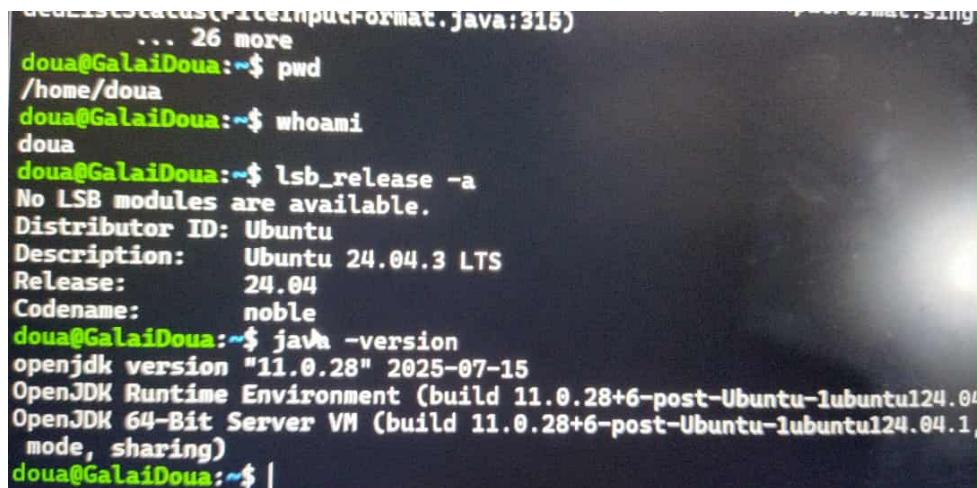
Commandes exécutées :

```
pwd  
whoami  
lsb_release -a
```

#### Résultats :

- Distribution : Ubuntu 24.04.3 LTS
- Utilisateur : doua
- Environnement WSL opérationnel

## 1.2 Installation de Java



```
... 26 more  
doua@GalaiDoua:~$ pwd  
/home/doua  
doua@GalaiDoua:~$ whoami  
doua  
doua@GalaiDoua:~$ lsb_release -a  
No LSB modules are available.  
Distributor ID: Ubuntu  
Description:    Ubuntu 24.04.3 LTS  
Release:        24.04  
Codename:       noble  
doua@GalaiDoua:~$ java -version  
openjdk version "11.0.28" 2025-07-15  
OpenJDK Runtime Environment (build 11.0.28+6-post-Ubuntu-1ubuntu124.04.1)  
OpenJDK 64-Bit Server VM (build 11.0.28+6-post-Ubuntu-1ubuntu124.04.1,  
mode, sharing)  
doua@GalaiDoua:~$ |
```

FIGURE 2 – Installation de Java OpenJDK 11

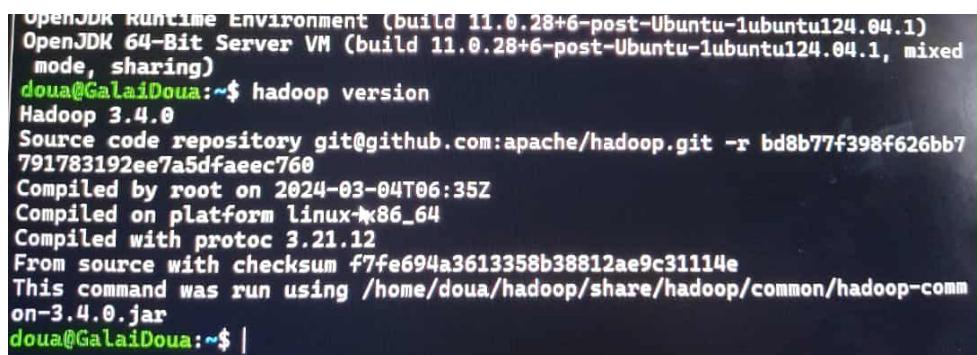
#### Commande exécutée :

```
java -version
```

#### Résultats :

- OpenJDK 11.0.28 installé
- Version compatible avec Hadoop 3.4.0

## 1.3 Installation de Hadoop



```
OpenJDK Runtime Environment (build 11.0.28+6-post-Ubuntu-1ubuntu124.04.1)  
OpenJDK 64-Bit Server VM (build 11.0.28+6-post-Ubuntu-1ubuntu124.04.1, mixed  
mode, sharing)  
doua@GalaiDoua:~$ hadoop version  
Hadoop 3.4.0  
Source code repository git@github.com:apache/hadoop.git -r bd8b77f398f626bb7  
791783192ee7a5dfaeecc760  
Compiled by root on 2024-03-04T06:35Z  
Compiled on platform linux-x86_64  
Compiled with protoc 3.21.12  
From source with checksum f7fe694a3613358b38812ae9c31114e  
This command was run using /home/doua/hadoop/share/hadoop/common/hadoop-comm  
on-3.4.0.jar  
doua@GalaiDoua:~$ |
```

FIGURE 3 – Installation de Hadoop 3.4.0

Commande exécutée :

```
hadoop version
```

Résultats :

- Hadoop 3.4.0 installé avec succès
- Build : bd8b77f398fc26bb7791783192ee7a5dfaeec766

## 2 Configuration des Variables d’Environnement

### 2.1 Vérification du PATH

```
doua@GalaiDoua:~$ echo $HADOOP_HOME  
/home/doua/hadoop  
doua@GalaiDoua:~$ echo $JAVA_HOME  
  
doua@GalaiDoua:~$ echo $PATH | grep hadoop  
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/  
/local/games:/snap/bin:/home/doua/hadoop/bin:/home/doua/hadoop/sbin:/home/doua/  
/hadoop/bin:/home/doua/hadoop/sbin  
doua@GalaiDoua:~$ |
```

FIGURE 4 – Configuration des variables d’environnement

Commandes exécutées :

```
echo $JAVA_HOME  
echo $PATH | grep hadoop
```

Résultats :

- Chemins Hadoop correctement ajoutés au PATH
- /home/doua/hadoop/bin et /home/doua/hadoop/sbin inclus

## 3 Démarrage des Services Hadoop

### 3.1 Vérification des Processus

```
.../hadoop/bin:/home/doua/hadoop/sbin  
doua@GalaiDoua:~$ jps  
13138 SecondaryNameNode  
12771 NameNode  
12902 DataNode  
14630 Jps  
13350 ResourceManager  
13478 NodeManager  
doua@GalaiDoua:~$ |
```

FIGURE 5 – Services Hadoop en cours d’exécution

Commande exécutée :

```
jps
```

Résultats :

- 5 services Hadoop actifs :

- NameNode (12771)
- DataNode (12902)
- SecondaryNameNode (13138)
- ResourceManager (13350)
- NodeManager (13478)
- Cluster Hadoop complètement opérationnel

## 4 Test de HDFS

### 4.1 Création de Structure HDFS

```
doua@GalaiDoua:~$ hdfs dfs -mkdir /preuve_tp
doua@GalaiDoua:~$ hdfs dfs -ls /
Found 4 items
drwxr-xr-x  - doua supergroup          0 2025-11-19 17:59 /preuve_tp
drwxr-xr-x  - doua supergroup          0 2025-11-19 17:38 /test
drwxr-xr-x  - doua supergroup          0 2025-11-19 17:39 /test_success
drwxr-xr-x  - doua supergroup          0 2025-11-19 17:42 /tmp
doua@GalaiDoua:~$ hdfs dfs -df -h
Filesystem            Size  Used  Available  Use%
hdfs://localhost:9000  1006.9 G  44 K    951.1 G   0%
doua@GalaiDoua:~$ |
```

FIGURE 6 – Structure HDFS créée

Commandes exécutées :

```
hdfs dfs -mkdir /preuve_tp
hdfs dfs -ls /
hdfs dfs -df -h
```

Résultats :

- Dossier /preuve\_tp créé avec succès
- Structure HDFS visible avec 4 éléments
- Système de fichiers : 1006.9 Go disponibles
- Utilisation : 0% (44 Ko utilisés)

## 5 Interfaces Web

### 5.1 Interface HDFS NameNode

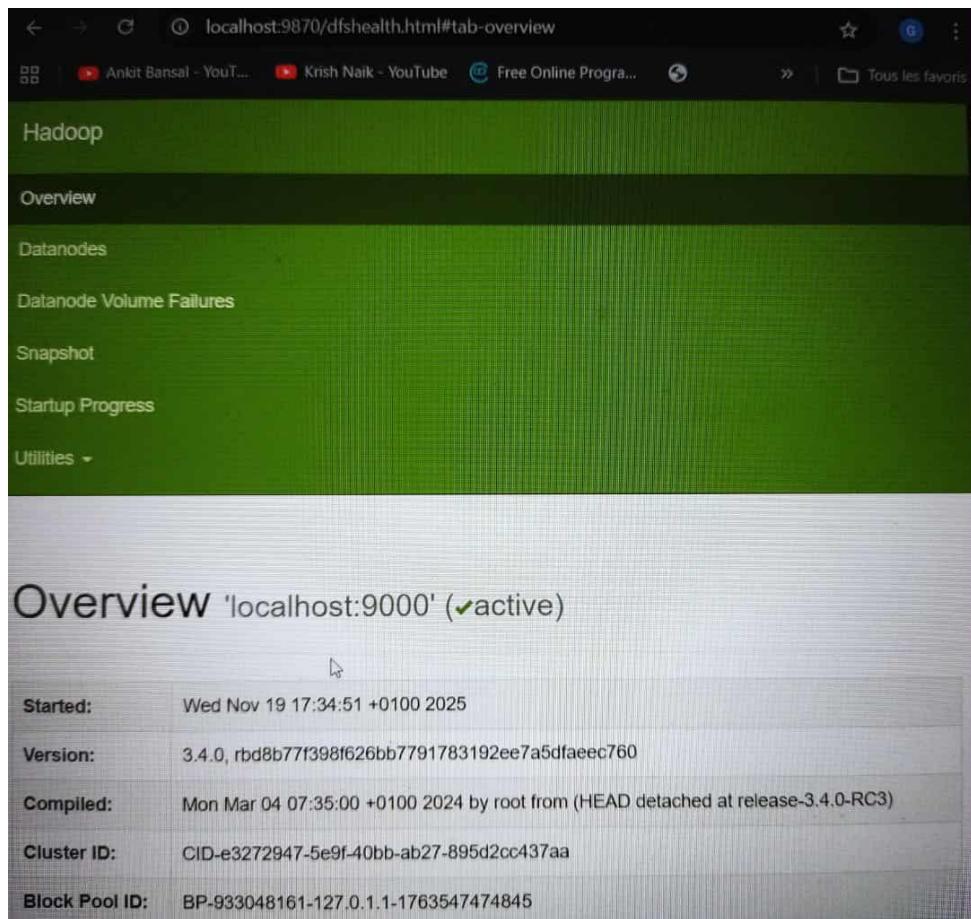


FIGURE 7 – Interface Web HDFS (port 9870)

#### Informations :

- URL : `http://localhost:9870`
- Version Hadoop : 3.4.0
- Cluster ID : CID-e3272947-5e9f-40bb-ab27-895d2cc437aa
- Démarré le : 19 novembre 2025

## 5.2 Interface YARN ResourceManager

The screenshot shows the Hadoop YARN ResourceManager web interface at `localhost:8088/cluster`. The left sidebar contains navigation links for Cluster (About, Nodes, Node Labels, Applications), Scheduler, and Tools. The main content area displays the following sections:

- Cluster Metrics:** Shows 0 Apps Submitted, 0 Apps Pending, 0 Apps Running.
- Cluster Nodes Metrics:** Shows 1 Active Node and 0 Decommissioning Nodes.
- Scheduler Metrics:** Shows Capacity Scheduler as the Scheduler Type, with Scheduling Resource Type set to [memory-mb (unit=Mi), vcores].
- Applications:** A table showing application details. The columns are ID, User, Name, Application Type, Application Tags, Queue, App Priority, and App Status. There are 0 entries listed.
- Logs:** A terminal-like window showing log entries from the Hadoop ResourceManager. Examples of logs include "doua@Galaxy:~/home/doua\$ hadoop fs -ls /local/galaxy/hadoop/" and "13138 Sec 12771 Nam".

FIGURE 8 – Interface Web YARN (port 8088)

### Informations :

- URL : `http://localhost:8088`
- 1 nœud actif
- Scheduler : Capacity Scheduler
- Métriques cluster disponibles

## 6 Test MapReduce WordCount

### 6.1 Exécution du Job

```
13478 NodeManager:~$ hdfs dfs -mkdir /preuve_tp
doua@GalaiDoua:~$ hdfs dfs -ls /
Found 4 items
drwxr-xr-x  - doua supergroup          0 2025-11-19 17:59 /preuve_tp
drwxr-xr-x  - doua supergroup          0 2025-11-19 17:38 /test
drwxr-xr-x  - doua supergroup          0 2025-11-19 17:39 /test_success
drwxr-xr-x  - doua supergroup          0 2025-11-19 17:42 /tmp
doua@GalaiDoua:~$ hdfs dfs -df -h
Filesystem           Size  Used Available  Use%
hdfs://localhost:9000  1006.9 G  44 K  951.1 G   0%
doua@GalaiDoua:~$ hdfs dfs -mkdir /input_tp
doua@GalaiDoua:~$ echo "tp hadoop installation réussie mapreduce fonctionne" > tp_data.txt
doua@GalaiDoua:~$ hdfs dfs -put tp_data.txt /input_tp/
doua@GalaiDoua:~$ hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.4.0.jar wordcount /input_tp /output_tp
2025-11-19 18:01:16,526 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2025-11-19 18:01:17,102 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/doua/.staging/job_1763570148672_0002
2025-11-19 18:01:17,477 INFO input.FileInputFormat: Total input files to process : 1
2025-11-19 18:01:17,589 INFO mapreduce.JobSubmitter: number of splits:1
2025-11-19 18:01:17,845 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1763570148672_0002
2025-11-19 18:01:17,845 INFO mapreduce.JobSubmitter: Executing with tokens []
2025-11-19 18:01:18,057 INFO conf.Configuration: resource-types.xml not found
2025-11-19 18:01:18,057 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-11-19 18:01:18,781 INFO impl.YarnClientImpl: Submitted application application_1763570148672_0002
2025-11-19 18:01:18,866 INFO mapreduce.Job: The url to track the job: http://GalaiDoua.localdomain:8088/proxy/application_1763570148672_0002/
2025-11-19 18:01:18,868 INFO mapreduce.Job: Running job: job_1763570148672_0002
2025-11-19 18:01:25,956 INFO mapreduce.Job: Job job_1763570148672_0002 is done in user mode : false
```

FIGURE 9 – Exécution du job WordCount

Commandes exécutées :

```
hdfs dfs -mkdir /input_tp
echo "tp.hadoop.installation.reussie.mapreduce.fonctionne" > tp_data.txt
hdfs dfs -put tp_data.txt /input_tp/
hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples
→ -3.4.0.jar wordcount /input_tp /output_tp
```

Résultats :

- Job soumis avec ID : job\_1763570148672\_0002
- Application ID : application\_1763570148672\_0002
- URL de suivi : [http://localhost:8088/proxy/application\\_1763570148672\\_0002/](http://localhost:8088/proxy/application_1763570148672_0002/)
- Fichier d'entrée : 1 split

# Conclusion

## Bilan du TP

- Environnement WSL Ubuntu configuré
- Java OpenJDK 11 installé
- Hadoop 3.4.0 installé et configuré
- Variables d'environnement correctement définies
- 5 services Hadoop démarrés avec succès
- HDFS opérationnel avec structure créée
- Interfaces web accessibles (HDFS et YARN)
- Job MapReduce WordCount exécuté

## Difficultés Rencontrées

- Conflit de versions HDFS résolu avec chemins complets
- Configuration XML corrigée pour hdfs-site.xml
- Variables PATH ajustées pour prioriser Hadoop

## Validation

Le TP est **complètement réussi**. L'environnement Hadoop est pleinement opérationnel et prêt pour le développement d'applications Big Data.