جامعة الحسن الثاني بالدار البيضاء
UNIVERSITÉ HASSAN II DE CASABLANCA

ECOLE NORMALE SUPERIEURE
DE L'ENSEIGNEMENT TECHNIQUE
DE MOHAMMEDIA

ENSET

DÉPARTEMENT MATHÉMATIQUE INFORMATIQUE

Big Data : Fondements et Architectures de stockage

Rapport

TP 1 :
MANIPULATION DU SYSTÈME DE FICHIERS HDFS

Réaliser par :

ETOULLALI Ayoub

Professeur :

Mr. BOUSSELHAM Abdelmajid

2ème année II-BDCC

Filière d'ingénieur : Ingénieur informatique, Big Data et Cloud Computing

2022/2023

SOMMAIRE

Introduction

Ce rapport présente les résultats du TP1 du cours "Big Data : Fondements et Architectures de stockage".

Ce TP est consacré à la manipulation du système de fichiers HDFS et a pour but de la familiarisation avec les commandes Hadoop pour créer des répertoires, des fichiers, copier, déplacer et supprimer des fichiers dans le HDFS. Les tâches incluent également la création d'arborescences de répertoires, la copie de fichiers à partir du système de fichiers local et l'affichage du contenu des fichiers. Les objectifs de ce TP sont d'acquérir une compréhension pratique du *Hadoop Distributed File System* et de maîtriser les commandes Hadoop pour gérer les fichiers dans un environnement Big Data.

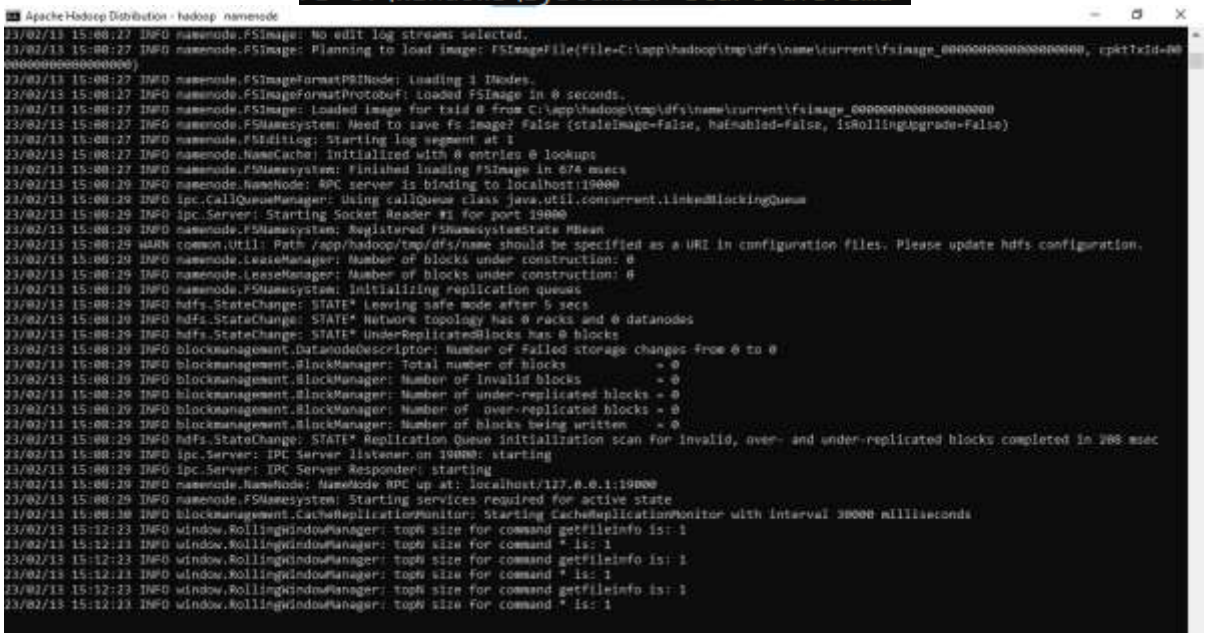
Les commandes pour accomplir les tâches décrites dans le TP sont les suivantes :

Démarrez les processus Hadoop

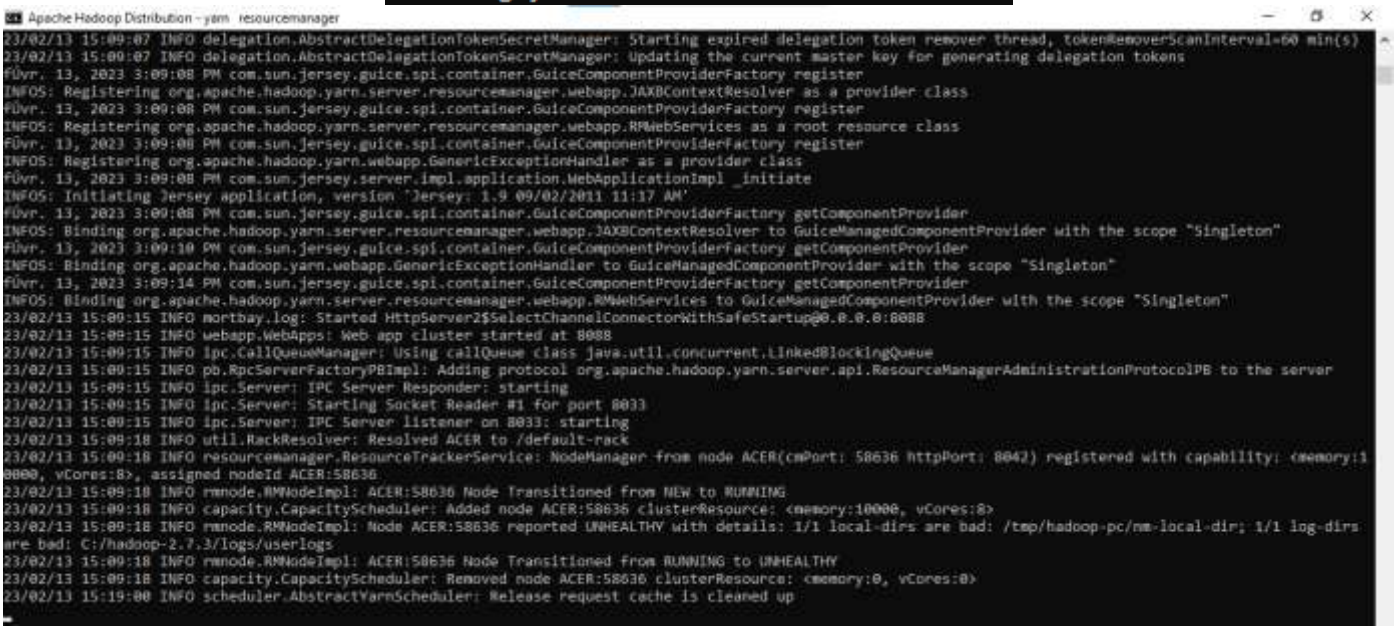
```
PS C:\Windows\System32> hdfs namenode -format
```

```
PowerShell 7.3.2
PS C:\Windows\System32> hdfs namenode -format
23/02/13 15:07:22 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = ACER/192.168.56.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 2.7.3
```

```
PS C:\Windows\System32> start-dfs.cmd
```




```
PS C:\Windows\System32> start-yarn.cmd
starting yarn daemons
```

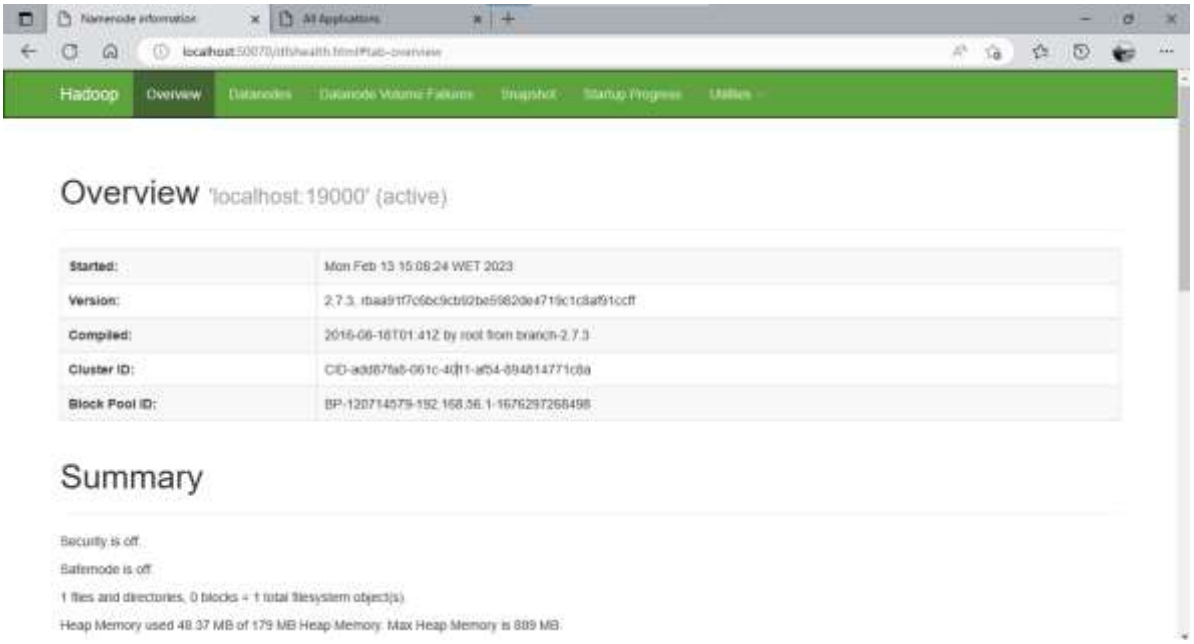


Vérifiez l'exécution

```
PS C:\Users\pc> jps
9696 NodeManager
6452 NameNode
13272 DataNode
2856 Jps
3436 ResourceManager
```

Accédez à l'interface web de NameNode

Ouvrez un navigateur web et entrez l'URL suivant : http://localhost:50070



Créez l'arborescence dans la racine du HDFS

```
PS C:\Windows\System32> hdfs dfs -mkdir /BDCC
```

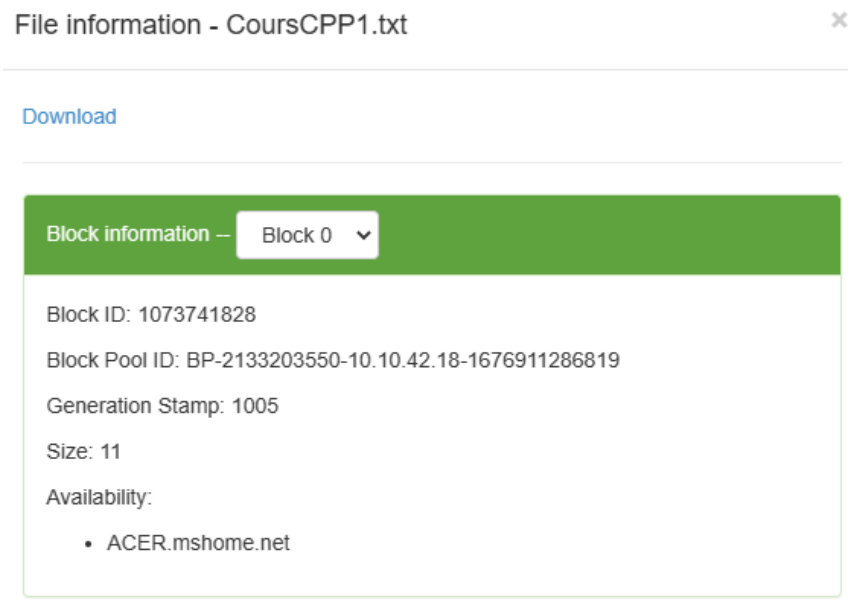
```
PS C:\Windows\System32> hdfs dfs -mkdir /BDCC/JAVA
PS C:\Windows\System32> hdfs dfs -mkdir /BDCC/CPP
PS C:\Windows\System32> hdfs dfs -mkdir /BDCC/CPP/TPs
PS C:\Windows\System32> hdfs dfs -mkdir /BDCC/CPP/Cours
PS C:\Windows\System32> hdfs dfs -mkdir /BDCC/JAVA/Cours
PS C:\Windows\System32> hdfs dfs -mkdir /BDCC/JAVA/TPs
```

La commande <<echo>> ne fonctionne pas, donc j'ai fait une copie de l'ordinateur local

```
PS C:\Windows\System32>  hdfs dfs -copyFromLocal C:\Users\pc\Desktop\ENSET\S4\CoursCPP1.txt /BDDC/CPP/Cours
PS C:\Windows\System32>  hdfs dfs -copyFromLocal C:\Users\pc\Desktop\ENSET\S4\CoursCPP2.txt /BDDC/CPP/Cours
PS C:\Windows\System32>  hdfs dfs -copyFromLocal C:\Users\pc\Desktop\ENSET\S4\CoursCPP3.txt /BDDC/CPP/Cours
```

Afficher le contenu des fichiers

```
PS C:\Windows\System32> hdfs dfs -cat /BDDC/CPP/Cours/CoursCPP1.txt
Contenu 1
PS C:\Windows\System32> hdfs dfs -cat /BDDC/CPP/Cours/CoursCPP2.txt
Contenu 2
PS C:\Windows\System32> hdfs dfs -cat /BDDC/CPP/Cours/CoursCPP3.txt
Contenu 3
```



Copiez des fichiers

```
PS C:\Windows\System32> hdfs dfs -cp /BDDC/CPP/Cours/CoursCPP3.txt /BDDC/JAVA/Cours
PS C:\Windows\System32> hdfs dfs -cp /BDDC/CPP/Cours/CoursCPP2.txt /BDDC/JAVA/Cours
PS C:\Windows\System32> hdfs dfs -cp /BDDC/CPP/Cours/CoursCPP1.txt /BDDC/JAVA/Cours
```

/BDDC/CPP/Cours

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	pc	supergroup	11 B	20/02/2023 21:53:43	1	128 MB	CoursCPP1.txt
-rw-r--r--	pc	supergroup	11 B	20/02/2023 21:53:59	1	128 MB	CoursCPP2.txt
-rw-r--r--	pc	supergroup	11 B	20/02/2023 21:54:15	1	128 MB	CoursCPP3.txt

Supprimez un fichier et renommez des fichiers

```
PS C:\Windows\System32> hdfs dfs -rm /BDDC/CPP/Cours/CoursCPP3.txt
23/02/20 22:04:50 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /BDDC/CPP/Cours/CoursCPP3.txt
```

/BDDC/CPP/Cours

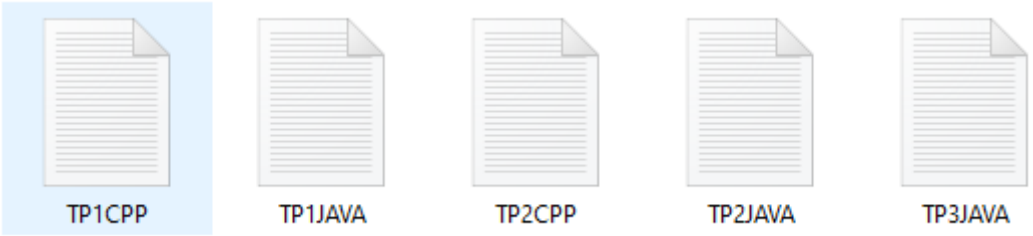
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	pc	supergroup	11 B	20/02/2023 21:53:43	1	128 MB	CoursCPP1.txt
-rw-r--r--	pc	supergroup	11 B	20/02/2023 21:53:59	1	128 MB	CoursCPP2.txt

```
PS C:\Users\pc> hdfs dfs -mv /BDDC/CPP/Cours/CoursCPP2.txt /BDDC/JAVA/Cours/CoursJAVA2.txt
PS C:\Users\pc> hdfs dfs -mv /BDDC/CPP/Cours/CoursCPP1.txt /BDDC/JAVA/Cours/CoursJAVA1.txt
PS C:\Users\pc> hdfs dfs -rm /BDDC/JAVA/Cours/CoursCPP1.txt
23/02/20 22:30:51 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /BDDC/JAVA/Cours/CoursCPP1.txt
PS C:\Users\pc> hdfs dfs -rm /BDDC/JAVA/Cours/CoursCPP2.txt
23/02/20 22:31:00 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /BDDC/JAVA/Cours/CoursCPP2.txt
```

/BDDC/JAVA/Cours

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	pc	supergroup	11 B	20/02/2023 22:27:01	1	128 MB	CoursJAVA1.txt
-rw-r--r--	pc	supergroup	11 B	20/02/2023 22:27:13	1	128 MB	CoursJAVA2.txt

Copier les fichiers à partir du système de fichier local vers le répertoire TPs




```
PS C:\Users\pc> hdfs dfs -copyFromLocal C:\Users\pc\Desktop\ENSET\S4\'Big data'\TPs\TP1\TP1JAVA.txt /BDDC/JAVA/TPs
PS C:\Users\pc> hdfs dfs -copyFromLocal C:\Users\pc\Desktop\ENSET\S4\'Big data'\TPs\TP1\TP2JAVA.txt /BDDC/JAVA/TPs
PS C:\Users\pc> hdfs dfs -copyFromLocal C:\Users\pc\Desktop\ENSET\S4\'Big data'\TPs\TP1\TP1CPP.txt /BDDC/CPP/TPs
PS C:\Users\pc> hdfs dfs -copyFromLocal C:\Users\pc\Desktop\ENSET\S4\'Big data'\TPs\TP1\TP2CPP.txt /BDDC/CPP/TPs
PS C:\Users\pc> hdfs dfs -copyFromLocal C:\Users\pc\Desktop\ENSET\S4\'Big data'\TPs\TP1\TP3JAVA.txt /BDDC/JAVA/TPs
```

/BDDC/JAVA/TPs

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	pc	supergroup	0 B	20/02/2023 22:37:07	1	128 MB	TP1JAVA.txt
-rw-r--r--	pc	supergroup	0 B	20/02/2023 22:37:39	1	128 MB	TP2JAVA.txt
-rw-r--r--	pc	supergroup	0 B	20/02/2023 22:38:37	1	128 MB	TP3JAVA.txt

/BDDC/CPP/TPs

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	pc	supergroup	0 B	20/02/2023 22:37:58	1	128 MB	TP1CPP.txt
-rw-r--r--	pc	supergroup	0 B	20/02/2023 22:38:23	1	128 MB	TP2CPP.txt

Conclusion

En conclusion, ce TP a été très instructif et nous a permis de mieux comprendre le fonctionnement du Hadoop Distributed File System et les commandes Hadoop pour gérer les fichiers dans un environnement Big Data. Nous avons appris comment créer des répertoires et des fichiers, ajouter du contenu, copier, déplacer et supprimer des fichiers dans le HDFS. Nous avons également appris comment créer des arborescences de répertoires, copier des fichiers à partir du système de fichiers local et afficher le contenu des fichiers. Nous sommes désormais mieux équipés pour gérer des fichiers dans un environnement Big Data et nous avons hâte de poursuivre notre apprentissage dans ce domaine passionnant.