

## TP: Hadoop HDFS

1. Tout d'abord, vous devez cloner le dépôt Git suivant dans le répertoire de votre choix sur votre système. Je préfère le cloner dans le home pour la démonstration et l'accès facile.

```
"git clone https://github.com/big-data-europe/docker-hadoop"
```

2. puis entrez dans le référentiel du projet à l'aide de la commande change directory :

```
cd docker-hadoop
```

3. tapez : **ls -l**

```
(base) paras@IUDX:~/docker-hadoop$ ls -l
total 52
drwxrwxr-x 2 paras paras 4096 Mar 17 14:44 base
drwxrwxr-x 2 paras paras 4096 Mar 17 14:44 datanode
-rw-rw-r-- 1 paras paras 2522 Mar 17 14:44 docker-compose-v3.yml
-rw-rw-r-- 1 paras paras 1559 Mar 17 14:44 docker-compose.yml
-rw-rw-r-- 1 paras paras 2507 Mar 17 14:44 hadoop.env
drwxrwxr-x 2 paras paras 4096 Mar 17 14:44 historyserver
-rw-rw-r-- 1 paras paras 1437 Mar 17 14:44 Makefile
drwxrwxr-x 2 paras paras 4096 Mar 17 14:44 namenode
drwxrwxr-x 2 paras paras 4096 Mar 17 14:44 nginx
drwxrwxr-x 2 paras paras 4096 Mar 17 14:44 nodemanager
-rw-rw-r-- 1 paras paras 2171 Mar 17 14:44 README.md
drwxrwxr-x 2 paras paras 4096 Mar 17 14:44 resourcemanager
drwxrwxr-x 2 paras paras 4096 Mar 17 14:44 submit
```

Output of the command "ls -l"

## 4. Obtenez les images Docker et créez des conteneurs

```
docker-compose up -d
```

La commande ci-dessus téléchargera toutes les images Docker nécessaires à partir du hub Docker pour configurer les conteneurs HDFS. Le téléchargement des images peut prendre un peu de temps en fonction de votre vitesse Internet.

Maintenant, pour voir vos conteneurs Docker en cours d'exécution, utilisez la commande pour répertorier tous les conteneurs actifs.

```
docker container ls
or
docker ps
```

```

[base] name@9870:~/hadoop$ sudo docker container ls
CONTAINER ID   IMAGE                                COMMAND                  CREATED        STATUS        PORTS                               NAMES
c78b111f4e5   bde2826/hadoop-namenode:2.0.0-hadoop3.2.1-javaf  "/entrypoint.sh /run..." About a minute ago Up About a minute (healthy)      8042/tcp      namenode
73524c2a7c5   bde2826/hadoop-nodemanaget:2.0.0-hadoop3.2.1-javaf  "/entrypoint.sh /run..." About a minute ago Up About a minute (healthy)      8042/tcp      nodemanaget
63524ff4e5d   bde2826/hadoop-datadnode:2.0.0-hadoop3.2.1-javaf  "/entrypoint.sh /run..." About a minute ago Up About a minute (healthy)      8042/tcp      datadnode
3f521c434c9   bde2826/hadoop-historyserver:2.0.0-hadoop3.2.1-javaf  "/entrypoint.sh /run..." About a minute ago Up About a minute (healthy)      8188/tcp      historyserver
b79acc14f6a   bde2826/hadoop-resourcemanaget:2.0.0-hadoop3.2.1-javaf  "/entrypoint.sh /run..." About a minute ago Up 24 seconds (health: starting) 8042/tcp      resourcemanaget

```

Output of the command "sudo docker container ls"

Et l'état actuel peut également être vérifié à l'aide de la page Web <http://localhost:9870> :

The screenshot shows a web browser displaying the Hadoop Overview page. The page has a green header with tabs for Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The main content area is titled 'Overview 'namenode:9000' (active)'. Below this title is a table with the following information:


Started:	Sun Oct 16 14:45:58 +0200 2022
Version:	3.2.1, rb3cbb467e22ea829c3808f4b7b01d07e0tf3842
Compiled:	Tue Sep 10 17:56:00 +0200 2019 by rohithsharmaks from branch-3.2.1
Cluster ID:	CID-8c1e91b0-a928-47f4-ae22-0b1057523504
Block Pool ID:	BP-225166066-172.18.0.6-1665924357565

Below the table is a 'Summary' section. It states: 'Security is off.', 'Safemode is off.', and '19 files and directories, 8 blocks (8 replicated blocks, 0 erasure coded block groups) = 27 total filesystem object(s)'. At the bottom, it shows memory usage: 'Heap Memory used: 133.27 MB of 228.5 MB. Max Heap Memory is 852 MB.' A notification bubble in the bottom right corner says '1 nouvelle notification'.


# Premier pas avec Hadoop

Toutes les commandes interagissant avec le système Hadoop commencent par **hadoop fs** (aussi **hdfs dfs**). Ensuite, les options rajoutées sont très largement inspirées des commandes Unix standard.

1- Créer un répertoire dans HDFS, appelé input (Tapez l'une de ces commandes)



```
$ hadoop fs -mkdir -p input
```



```
$ hdfs dfs -mkdir -p input
```

## Ajouter des fichiers textes (.txt) dans le namenode

```
docker cp C:\Users\amina\OneDrive\Bureau\TP\words.txt  
namenode:/input/
```

Travail à faire (en utilisant les commandes HDFS):

1. Vérifier si le fichier words.txt est bien placé dans votre répertoire locale,
2. Placer le fichier words.txt dans le répertoire que vous avez créé
3. afficher le contenu de répertoire /input
4. télécharger le contenu de répertoire /input
5. lire les dernières lignes de fichier words.txt