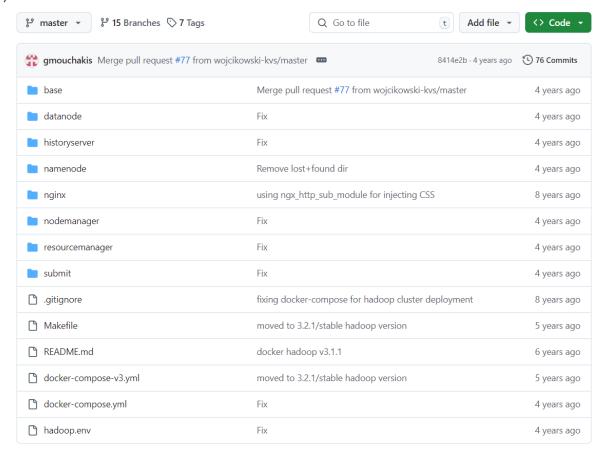
# Big Data Tp1

# Partie 1: Installation et Configuration

### 1)cloner:



C:\Users\21656\Desktop\ING 2\big-data\tp\tp1>git clone https://github.com/big-data-europe/docker-hadoop.git

### 2)docker compose up:

C:\Users\21656\Desktop\ING 2\big-data\tp\tp1\docker-hadoop>docker ps							
CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS	PORTS		
(E (- 2E20#	hd-2020/h-d d-td2 0 0 h-d2 2 1 d	NAMES	About a minute and	Ha Abant a mianta (baaltha)	00011/4		
ccbf5fa27284	bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8	"/entrypoint.sh /run" datanode	About a minute ago	Up About a minute (healthy)	9864/tcp		
9dc6b6f1ff23	bde2020/hadoop-historyserver:2.0.0-hadoop3.2.1-java8	"/entrypoint.sh /run"	About a minute ago	Up About a minute (healthy)	8188/tcp		
01560075	112020/11	historyserver			0000/1		
c81560075edd	bde2020/hadoop-resourcemanager:2.0.0-hadoop3.2.1-java8	"/entrypoint.sh /run" resourcemanager	About a minute ago	Up About a minute (healthy)	8088/tcp		
c6aaedb5c95e	bde2020/hadoop-nodemanager:2.0.0-hadoop3.2.1-java8	"/entrypoint.sh /run"	About a minute ago	Up About a minute (healthy)	8042/tcp		
		nodemanager					
42889edbd740	bde2020/hadoop-namenode:2.0.0-hadoop3.2.1-java8	"/entrypoint.sh /run"	About a minute ago	Up About a minute (healthy)	0.0.0.0:		
9000->9000/tcp, :::9000->9000/tcp, 0.0.0.0:9870->9870/tcp, :::9870->9870/tcp namenode							

#### Partie 2: Commandes de HDFS

Ouvrir le bash dans le conteneur namenode :

```
 C:\Users\21656\Desktop\ING\ 2\big-data\tp\tp1\docker-hadoop>docker\ exec\ -it\ namenode\ bash\ root@42889edbd740:/\#
```

Créer un fichier nommé bonjour contenant Bonjour Hadoop et HDFS dans le SE linux du namenode :

```
root@42889edbd740:/# echo "Bonjour Hadoop et HDFS">bonjour.txt
```

Interaction avec Hadoop:

```
root@42889edbd740:/# hdfs dfs -ls
ls: `.': No such file or directory
```

root@42889edbd740:/# hdfs dfs -mkdir -p /

```
root@42889edbd740:/# hdfs dfs -ls /user
Found 1 items
drwxr-xr-x - root supergroup 0 2024-09-27 14:13 /user/root
```

```
root@42889edbd740:/# hdfs dfs -ls /
  Found 1 items
                                                                                                         0 2024-09-27 14:00 /rmstate
  drwxr-xr-x
                                     - root supergroup
  oot@42889edbd740:/# hdfs dfs -ls -R -h /
rwxr-xr-x - root supergroup
                                                   h /
0 2024-09-27 14:00 /rmstate
0 2024-09-27 14:00 /rmstate
0 2024-09-27 14:00 /rmstate/FSRMStateRoot
0 2024-09-27 14:00 /rmstate/FSRMStateRoot/AMRMTokenSecretManagerRoot
18 2024-09-27 14:00 /rmstate/FSRMStateRoot/AMRMTokenSecretManagerRoot/AMRMTokenSecretManagerN
2 2024-09-27 14:00 /rmstate/FSRMStateRoot/EDpochNode
0 2024-09-27 14:00 /rmstate/FSRMStateRoot/RMAppRoot
0 2024-09-27 14:00 /rmstate/FSRMStateRoot/RMDTSecretManagerRoot/DelegationKey_1
17 2024-09-27 14:00 /rmstate/FSRMStateRoot/RMDTSecretManagerRoot/DelegationKey_2
4 2024-09-27 14:00 /rmstate/FSRMStateRoot/RMSTSOTAGEROOT/RMSTSOTAGEROOT/PSTATEROOT/PSTATEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOTAGEROOT
drwxr-xr-x - root supergroup
drwxr-xr-x - root supergroup
drwxr-xr-x - root supergroup
drwxr-xr-x - root supergroup
-rw-r-r-- 3 root supergroup
drwxr-xr-x - root supergroup
drwxr-xr-x - root supergroup
drwxr-xr-x - root supergroup
-rw-r-r-- 3 root supergroup
-rw-r-r-- 3 root supergroup
-rw-r-r-- 3 root supergroup
drwxr-xr-x - root supergroup
  root@42889edbd740:/# hdfs dfs -mkdir -p /ahmed
  root@42889edbd740:/# hdfs dfs -ls /
  Found 3 items
                                                                                                         0 2024-09-27 14:22 /ahmed
  drwxr-xr-x - root supergroup
  drwxr-xr-x - root supergroup
                                                                                                        0 2024-09-27 14:00 /rmstate
                                                                                                       0 2024-09-27 14:13 /user
  drwxr-xr-x - root supergroup
 root@42889edbd740:/# hdfs dfs -mkdir data
root@42889edbd740:/# hdfs dfs -mkdir data/tp1
root@42889edbd740:/# hdfs dfs -mkdir data/tp2
 root@42889edbd740:/# hdfs dfs -mkdir data/tp1/ex1
 root@42889edbd740:/# echo "bonjour hadoop">data/tp1/ex1/file1.txt
 bash: data/tp1/ex1/file1.txt: No such file or directory
 root@42889edbd740:/# echo "bonjour hadoop" > data/tp1/ex1/file1.txt
 bash: data/tp1/ex1/file1.txt: No such file or directory
 root@42889edbd740:/# echo "bonjour hadoop" > /data/tp1/ex1/file1.txt
 bash: /data/tp1/ex1/file1.txt: No such file or directory
 root@42889edbd740:/# hdfs dfs -ls data
 Found 2 items
drwxr-xr-x - root supergroup
drwxr-xr-x - root supergroup
                                                                                                  0 2024-09-27 14:26 data/tp1
                                                                                                 0 2024-09-27 14:26 data/tp2
 root@42889edbd740:/# hdfs dfs -ls data/tp1
 Found 1 items
drwxr-xr-x - root supergroup
                                                                                                  0 2024-09-27 14:26 data/tp1/ex1
root@d32ae7953bba:/# cat bonjour.txt
Boniour Hadoop et HDFS
root@d32ae7953bba:/# hdfs dfs -put bonjour.txt
2024-09-29 06:09:57,692 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteH
ostTrusted = false
root@d32ae7953bba:/# hdfs dfs -ls -R
                                                                                    23 2024-09-29 06:09 bonjour.txt
 -rw-r--r-- 3 root supergroup
                                                                                  0 2024-09-27 14:26 data
drwxr-xr-x - root supergroup
                                                                                   0 2024-09-27 14:26 data/tp1
drwxr-xr-x - root supergroup
drwxr-xr-x - root supergroup
-rw-r--r- 3 root supergroup
drwxr-xr-x - root supergroup
                                                                                    0 2024-09-27 15:05 data/tp1/ex1
                                                                                    0 2024-09-27 15:05 data/tp1/ex1/file1.txt
0 2024-09-27 14:26 data/tp2
root@d32ae7953bba:/# hdfs dfs -cat bonjour.txt
2024-09-29 06:11:25,501 INFO sasl.SaslÓataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteH
ostTrusted = false
Boniour Hadoop et HDFS
son four nations of nurs
root@d32ae7953bba:/# hdfs dfs -cat bonjour.txt | more
2024-09-29 06:12:01,084 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteH
ostTrusted = false
Bonjour Hadoop et HDFS
root@d32ae7953bba:/# hdfs dfs -tail bonjour.txt
2024-09-29 06:12:50,313 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteH
ostTrusted = false 
Bonjour Hadoop et HDFS
```

root@d32ae7953bba:/# hdfs dfs -rm bonjour.txt Deleted bonjour.txt

```
root@d32ae7953bba:/# hdfs dfs -copyFromLocal bonjour.txt
2024-09-29 06:14:26,807 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteH
2024-09-29 00:14.20,007 1m 5 3051
ostTrusted = false
root@d32ae7953bba:/# hdfs dfs -ls -R
                                        23 2024-09-29 06:14 bonjour.txt
0 2024-09-27 14:26 data
0 2024-09-27 14:26 data/tpl
-rw-r--r- 3 root supergroup
drwxr-xr-x - root supergroup
drwxr-xr-x - root supergroup
           - root supergroup
3 root supergroup
- root supergroup
                                         0 2024-09-27 15:05 data/tp1/ex1
0 2024-09-27 15:05 data/tp1/ex1/file1.txt
                                         0 2024-09-27 14:26 data/tp2
root@d32ae7953bba:/# hdfs dfs -chmod go+w bonjour.txt
root@d32ae7953bba:/# hdfs dfs -ls -R
 -rw-rw-rw- 3 root supergroup
                                                           23 2024-09-29 06:14 bonjour.txt

    root supergroup

                                                             0 2024-09-27 14:26 data
drwxr-xr-x
                                                             0 2024-09-27 14:26 data/tp1
                 - root supergroup
drwxr-xr-x
drwxr-xr-x - root supergroup
                                                             0 2024-09-27 15:05 data/tp1/ex1
                 3 root supergroup
                                                             0 2024-09-27 15:05 data/tp1/ex1/file1.txt
 -rw-r--r--
                                                             0 2024-09-27 14:26 data/tp2
drwxr-xr-x - root supergroup
root@d32ae7953bba:/# hdfs dfs -chmod go-r bonjour.txt
root@d32ae7953bba:/# hdfs dfs -ls -R
                  3 root supergroup
                                                            23 2024-09-29 06:14 bonjour.txt
 -rw--w--w-
 drwxr-xr-x

    root supergroup

                                                             0 2024-09-27 14:26 data
 drwxr-xr-x - root supergroup
                                                             0 2024-09-27 14:26 data/tp1
 drwxr-xr-x
                    - root supergroup
                                                             0 2024-09-27 15:05 data/tp1/ex1
 -rw-r--r--
                    3 root supergroup
                                                             0 2024-09-27 15:05 data/tp1/ex1/file1.txt
drwxr-xr-x
                    root supergroup
                                                             0 2024-09-27 14:26 data/tp2
root@d32ae7953bba:/# hdfs dfs -mv bonjour.txt dossier/bonjour.txt
root@d32ae7953bba:/# hdfs dfs -ls -R
drwxr-xr-x - root supergroup
                                                             0 2024-09-27 14:26 data
drwxr-xr-x - root supergroup
                                                             0 2024-09-27 14:26 data/tp1
drwxr-xr-x
                                                            0 2024-09-27 15:05 data/tp1/ex1
                   root supergroup
 -rw-r--r--
                  3 root supergroup
                                                             0 2024-09-27 15:05 data/tp1/ex1/file1.txt
drwxr-xr-x
                    root supergroup
                                                             0 2024-09-27 14:26 data/tp2
                                                             0 2024-09-29 06:18 dossier
drwxr-xr-x
                    root supergroup
                                                           23 2024-09-29 06:14 dossier/bonjour.txt
                  3 root supergroup
 -rw--w--w-
root@d32ae7953bba:/# hdfs dfs -get dossier/bonjour.txt bien.txt
2024-09-29 06:20:01,465 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteH
pstTrusted = false
root@d32ae7953bba:/# ls
KEYS
             bonjour.txt
                                                                  hadoop-data lib
                                                                                                              sbin
                                                                                                                     tmp
             bonjour1.txt dev
                                                   file2.txt
                                                                                  lib64 opt
bien.txt
                                                                 hdfs
                                                                                                              srv
                                                                                                   run
                                                                                                                     usr
bin
             bonjour2.txt entrypoint.sh hadoop
                                                                                  media proc run.sh sys
                                                                 home
                                                                                                                     var
root@d32ae7953bba:/# hd+s d+s -cp dossier/bonjour.txt dossier/bien.txt
2024-09-29 06:21:53,749 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteH
0stTrusted = false
03c1rl3ceu - Talse
2024-09-29 06:21:53,824 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteH
ostTrusted = false
root@d32ae7953bba:/# hdfs dfs -ls -R
                                        0 2024-09-27 14:26 data
0 2024-09-27 14:26 data/tp1
0 2024-09-27 15:05 data/tp1/ex1
           - root supergroup
- root supergroup
drwxr-xr-x
             - root supergroup
drwxr-xr-x
                                       0 2024-09-27 15:05 data/tp1/ex1/file1.txt
0 2024-09-27 15:05 data/tp1/ex1/file1.txt
0 2024-09-27 14:26 data/tp2
0 2024-09-29 06:21 dossier
23 2024-09-29 06:21 dossier/bien.txt
23 2024-09-29 06:14 dossier/bonjour.txt
            3 root supergroup
            - root supergroup
drwxr-xr-x
            - root supergroup
            3 root supergroup
 -rw-r--r--
              3 root supergroup
rnot@d32ae7953bba:/# hdfs dfs -cat dossier/bien.txt
2024-09-29 06:22:37,367 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteH
ostTrusted = false
Bonjour Hadoop et HDFS
```

```
root@d32ae7953bba:/# hdfs dfs -count -h dossier
1 2 46 dossier
```

root@d32ae7953bba:/# hdfs dfs -rm dossier/bonjour.txt Deleted dossier/bonjour.txt

```
root@d32ae7953bba:/# hdfs dfs -ls dossier
Found 1 items
-rw-r--r- 3 root supergroup 23 2024-09-29 06:21 dossier/bien.txt
root@d32ae7953bba:/# hdfs dfs -rm -r dossier
Deleted dossier
root@d32ae7953bba:/# hdfs dfs -ls
Found 1 items
drwxr-xr-x - root supergroup 0 2024-09-27 14:26 data
```

# Partie 3: Création d'une arborescence et téléchargement de fichier :

Création des répertoires TPs et data :

## Télechargement et copiage du fichier purchases.txt:

<b>base</b>	9/27/2024 2:32 PM	File folder	
atanode	9/27/2024 2:32 PM	File folder	
historyserver	9/27/2024 2:32 PM	File folder	
namenode	9/27/2024 2:32 PM	File folder	
nginx	9/27/2024 2:32 PM	File folder	
nodemanager	9/27/2024 2:32 PM	File folder	
resourcemanager	9/27/2024 2:32 PM	File folder	
submit	9/27/2024 2:32 PM	File folder	
gitignore	9/27/2024 2:32 PM	Git Ignore Source	1 KB
docker-compose	9/27/2024 2:32 PM	Yaml Source File	2 KB
docker-compose-v3	9/27/2024 2:32 PM	Yaml Source File	3 KB
hadoop.env	9/27/2024 2:32 PM	ENV File	3 KB
Makefile	9/27/2024 2:32 PM	File	2 KB
purchases	9/29/2024 7:57 AM	Text Document	206,361 KB
■ README	9/27/2024 2:32 PM	Markdown Source	3 KB

#### Copier le fichier du local vers namenode :

C:\Users\21656\Desktop\ING 2\big-data\tp\tp1\docker-hadoop>docker cp purchases.txt namenode:purshases.txt Successfully copied 211MB to namenode:purshases.txt

```
root@f4fbf8cfda46:/# ls
                                          home lib64 mnt proc
lib media opt pursh
KEYS boot entrypoint.sh
                            hadoop
                                                                              root run.sh srv tmp var
                             hadoop-data
                                                              purshases.txt
                                                                                     sbin
                                                                              run
                                                                                             svs
                                                                                                   usr
```

#### Téléchargement des fichiers en utilisant docker exec :

```
:\Users\21656\Desktop\ING 2\big-data\tp\tp1\docker-hadoop>docker exec -it namenode curl -o pg4300.txt https://www.guten
berg.org/cache/epub/4300/pg4300.txt

% Total % Received % Xferd Average Speed Time Time Time
Dload Upload Total Spent Left
100 1549k 100 1549k 0 0 373k 0 0:00:04 0:00:04 --:--:
                                                                                 Time Current
Left Speed
-:--:- 374k
100 1549k 100 1549k
 ::\Users\21656\Desktop\ING 2\big-data\tp\tp1\docker-hadoop>docker exec -it namenode curl -o pg135.txt https://www.gutenb
erg.org/cache/epub/135/pg135.txt
% Total % Received % Xferd Average Speed
                                                                                  Time Curre
Left Speed
---- 638k
                                                             Time
                                                                        Time
                                                                                          Current
                                     Dload Upload Total
0 517k 0 0:00:06
                                                                      Spent Left
0:00:06 --:--:
100 3290k 100 3290k
root@f4fbf8cfda46:/# ls
                                                                              pg4300.txt proc
KEYS boot entrypoint.sh
bin dev etc
                                                      home lib64 mnt pg135.txt
                                    hadoop
                                                                                                                  root run.sh srv tmp var
                                    hadoop-data lib
                                                             media opt
                                                                                              purshases.txt run
```

#### Mettre les fichiers su hdfs:

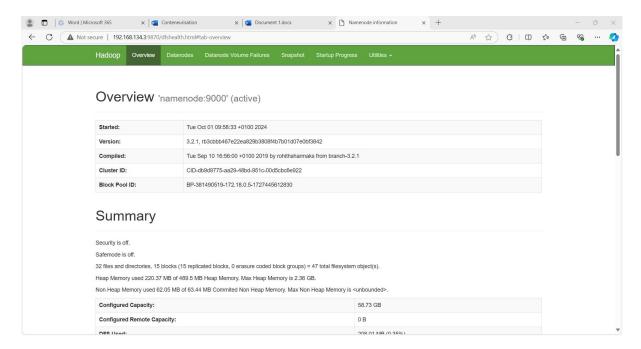
```
root@f4fbf8cfda46:/# hdfs dfs -put purshases.txt
2024-10-01 09:24:09,892 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteH
   ostTrusted = false 2024-10-01 09:24:10,500 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteH
  2024-10-01 09:24:10,000 INFO SAST:SASTDATATIANSFERCTION: SAST ENCRYPTION TRUST CHECK: TOTAL MOSTIFUSTED - FALSE, PERMOTER COSTIFUSTED - FALSE, PROMOTER COSTIFUSTED - FALSE COSTIFUSTED - 
   ostTrusted = false
root@f4fbf8cfda46:/# hdfs dfs -ls
root@f4fbf8ctuate
Found 5 items
drwxr-xr-x - root supergroup
drwxr-xr-x - root supergroup
-rw-r--r- 3 root supergroup
-rw-r--r- 3 root supergroup
-rw-r--r- 3 root supergroup
                                                                                                                                                                                         0 2024-10-01 09:00 TPs
0 2024-09-27 14:26 data
3369259 2024-10-01 09:24 pg135.txt
1586382 2024-10-01 09:24 pg4300.txt
211312924 2024-10-01 09:24 purshases.txt
```

#### Déplacement des fichiers dans le sous dossier data :

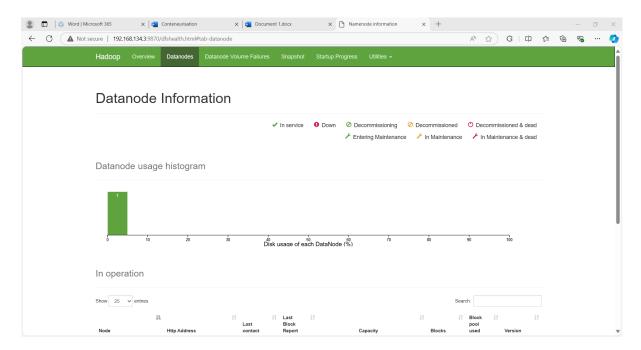
```
root@f4fbf8cfda46:/# hdfs dfs -mv purshases.txt TPs/data/purshases.txt
root@f4fbf8cfda46:/# hdfs dfs -mv pg4300.txt TPs/data/pg4300.txt
root@f4fbf8cfda46:/# hdfs dfs -mv pg135.txt TPs/data/pg135.txt
root@f4fbf8cfda46:/# hdfs dfs -ls TPs/data/
Found 3 items
-rw-r--r-- 3 root supergroup
                                     3369259 2024-10-01 09:24 TPs/data/pg135.txt
1586382 2024-10-01 09:24 TPs/data/pg4300.txt
              3 root supergroup
-rw-r--r--
              3 root supergroup 211312924 2024-10-01 09:24 TPs/data/purshases.txt
```

#### Etat du cluster:

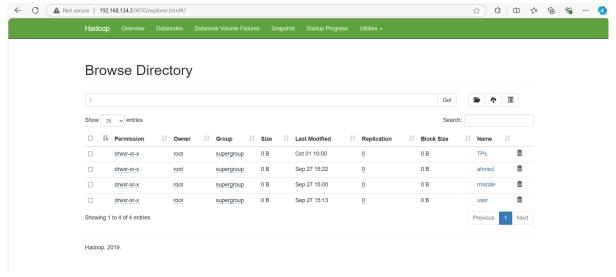
#### Overview:



## Datanodes:



# Utilities:



.