



UNIVERSITE GASTON BERGER

L'excellence au service du développement

INSTITUT  POLYTECHNIQUE
DE SAINT-LOUIS

Evaluation : **TP1 HDFS Prise en main**

Michel Nassalang

1. Prise en main de HDFS

Pour démarrer hdfs on a la commande : start-all.sh

Pour voir les commandes possibles avec hdfs : hdfs dfs

Pour voir ce que fait une commande avec hdfs : hdfs dfs -usage COMMANDE

2. Création des répertoires et copies de fichiers

- Pour la création de dossier dans hdfs, on a la commande : **hdfs dfs -mkdir path**

Pour le dossier nassalang créé, on a : **hdfs dfs -mkdir /user/vagrant/nassalang**

Pour le dossier retails : **hdfs dfs -mkdir /user/vagrant/nassalang/retails**

Pour le dossier orders : **hdfs dfs -mkdir /user/vagrant/nassalang/retails/orders**

- Pour la copie du fichier dans hdfs, on a : **hdfs dfs -copyFromLocal localdst dst**

Pour la copie du fichier /home/vagrant/shareFolder/data/retail_db/orders/part-00000 dans le répertoire hdfs créé précédemment on a comme commande :

hdfs dfs -copyFromLocal /home/vagrant/shareFolder/data/retail_db/orders/part-00000 /user/vagrant/nassalang/retails/orders

```
[vagrant@10 ~]$ hdfs dfs -ls /user/vagrant
Found 4 items
-rw-r--r-- 1 vagrant supergroup          42 2021-11-20 13:15 /user/vagrant/README.md
drwxr-xr-x - vagrant supergroup           0 2021-11-27 10:39 /user/vagrant/datasets
-rwxrwxrwx 1 vagrant supergroup       351323 2021-11-20 13:16 /user/vagrant/derby.log
drwxr-xr-x - vagrant supergroup           0 2021-11-27 10:43 /user/vagrant/master1
[vagrant@10 ~]$ hdfs dfs -mkdir /user/vagrant/nassalang
[vagrant@10 ~]$ hdfs dfs -ls /user/vagrant
Found 5 items
-rw-r--r-- 1 vagrant supergroup          42 2021-11-20 13:15 /user/vagrant/README.md
drwxr-xr-x - vagrant supergroup           0 2021-11-27 10:39 /user/vagrant/datasets
-rwxrwxrwx 1 vagrant supergroup       351323 2021-11-20 13:16 /user/vagrant/derby.log
drwxr-xr-x - vagrant supergroup           0 2021-11-27 10:43 /user/vagrant/master1
drwxr-xr-x - vagrant supergroup           0 2023-05-21 15:40 /user/vagrant/nassalang
[vagrant@10 ~]$ hdfs dfs -mkdir /user/vagrant/nassalang/retails
[vagrant@10 ~]$ hdfs dfs -mkdir /user/vagrant/nassalang/retails/orders
[vagrant@10 ~]$ hdfs dfs -copyFromLocal /home/vagrant/shareFolder/data/retail_db/orders/part-00000 /user/vagrant/nassalang/retails/orders
2023-05-21 15:44:16,125 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = 
[vagrant@10 ~]$ hdfs dfs -ls /user/vagrant/nassalang/retails/orders
Found 1 items
-rw-r--r-- 1 vagrant supergroup       3068827 2023-05-21 15:44 /user/vagrant/nassalang/retails/orders/part-00000
[vagrant@10 ~]$
```

3. Visualisation des données par hdfs

Pour afficher les lignes d'un fichier, on utilise les commandes head et tail :

hdfs dfs -head file ; hdfs dfs -tail file

Pour le fichier part-00000: on a : **hdfs dfs -head /user/vagrant/nassalang/retails/orders/part-00000**; **hdfs dfs -tail /user/vagrant/nassalang/retails/orders/part-00000**

```
rw-r--r-- 1 vagrant supergroup 3068827 2023-05-21 15:44 /user/vagrant/nassalang/retails/orders/part-00000
[vagrant@10 ~]$ hdfs dfs -head /user/vagrant/nassalang/retails/orders/part-00000; hdfs dfs -tail /user/vagrant/nassalang/retails/orders/part-00000
2023-05-21 16:33:29,509 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
1,2013-07-25 00:00:00.0,11599,CLOSED
2,2013-07-25 00:00:00.0,256,PENDING_PAYMENT 7. Créer dans ce répertoire HDFS un nouveau répertoire nommé caaaa
3,2013-07-25 00:00:00.0,12111,COMPLETE
4,2013-07-25 00:00:00.0,8827,CLOSED 8. Copier le fichier local
5,2013-07-25 00:00:00.0,11318,COMPLETE
6,2013-07-25 00:00:00.0,7130,COMPLETE
7,2013-07-25 00:00:00.0,4530,COMPLETE
8,2013-07-25 00:00:00.0,2911,PROCESSING
9,2013-07-25 00:00:00.0,5657,PENDING_PAYMENT
10,2013-07-25 00:00:00.0,5648,PENDING_PAYMENT 9. Afficher les 5 premières lignes et les 5 dernières lignes du fichier copier dans HDFS
11,2013-07-25 00:00:00.0,918,PAYMENT_REVIEW
12,2013-07-25 00:00:00.0,1837,CLOSED 10. Lister les dossiers présents dans /user/hive/warehouse/test.db/
13,2013-07-25 00:00:00.0,9149,PENDING_PAYMENT
14,2013-07-25 00:00:00.0,9842,PROCESSING
15,2013-07-25 00:00:00.0,2568,COMPLETE
16,2013-07-25 00:00:00.0,7276,PENDING_PAYMENT 11. Créer dans votre répertoire HDFS un dossier nommé employee dans
17,2013-07-25 00:00:00.0,2667,COMPLETE /user/vagrant/ (userhome) /
18,2013-07-25 00:00:00.0,1205,CLOSED
19,2013-07-25 00:00:00.0,9488,PENDING_PAYMENT
20,2013-07-25 00:00:00.0,9198,PROCESSING
21,2013-07-25 00:00:00.0,2711,PENDING
22,2013-07-25 00:00:00.0,333,COMPLETE
23,2013-07-25 00:00:00.0,4367,PENDING_PAYMENT
24,2013-07-25 00:00:00.0,11441,CLOSED
25,2013-07-25 00:00:00.0,2023-05-21 16:33:35,269 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted =
9,PENDING
58861,2014-06-13 00:00:00.0,3031,PENDING_PAYMENT
58862,2014-06-15 00:00:00.0,7326,PROCESSING
58863,2014-06-16 00:00:00.0,3361,CLOSED
58864,2014-06-18 00:00:00.0,9634,ON_HOLD
58865,2014-06-19 00:00:00.0,4567,SUSPECTED_FRAUD 12. Copier le contenu du répertoire /user/hive/warehouse/test.db/employee
58866,2014-06-20 00:00:00.0,3890,PENDING_PAYMENT dans le répertoire /user/vagrant/ (userhome) /employee en le renommant
58867,2014-06-23 00:00:00.0,869,CANCELED avec 000000_1_employee.db
58868,2014-06-24 00:00:00.0,10184,PENDING
58869,2014-06-25 00:00:00.0,7456,PROCESSING
58870,2014-06-26 00:00:00.0,3343,COMPLETE 13. Copier le dossier /user/vagrant/ (userhome) /employee se trouvant dans HDFS
```

Pour lister les fichiers présents dans un dossier on utilise : **hdfs dfs -ls path**

Pour le repertoire /user/hive/warehouse/test.db/, on a : **hdfs dfs -ls /user/hive/warehouse/test.db/**

```
68883,2014-07-23 00:00:00.0,5533,COMPLETE
[vagrant@10 ~]$ hdfs dfs -ls /user/hive/warehouse/test.db/
Found 2 items
drwxr-xr-x - vagrant supergroup 0 2021-03-24 15:04 /user/hive/warehouse/test.db/crimes
drwxr-xr-x - vagrant supergroup 0 2021-03-24 16:10 /user/hive/warehouse/test.db/employee
[vagrant@10 ~]$ |
```

4. Création des repertoires et copies de contenu

Pour le dossier employee créé, on a : **hdfs dfs -mkdir /user/vagrant/nassalang/employee**

Pour la copie du contenu de dans ce repertoire créé précédemment , on a : **hdfs dfs -cp /user/hive/warehouse/test.db/employee/000000_1 /user/vagrant/nassalang/employee**

```
[vagrant@10 ~]$ hdfs dfs -ls /user/hive/warehouse/test.db/
Found 2 items
drwxr-xr-x - vagrant supergroup          0 2021-03-24 15:04 /user/hive/warehouse/test.db/crimes
drwxr-xr-x - vagrant supergroup          0 2021-03-24 16:10 /user/hive/warehouse/test.db/employee
[vagrant@10 ~]$ hdfs dfs -mkdir /user/vagrant/nassalang/employee
[vagrant@10 ~]$ hdfs dfs -cp /user/hive/warehouse/test.db/employee /user/vagrant/nassalang/employee
2023-05-21 17:05:48,536 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2023-05-21 17:05:49,061 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
[vagrant@10 ~]$ hdfs dfs -ls /user/vagrant/nassalang/employee
Found 1 items
drwxr-xr-x - vagrant supergroup          0 2023-05-21 17:05 /user/vagrant/nassalang/employee/employee
[vagrant@10 ~]$ hdfs dfs -ls /user/vagrant/nassalang/employee/employee
Found 1 items
-rw-r--r-- 1 vagrant supergroup          23 2023-05-21 17:05 /user/vagrant/nassalang/employee/employee/000000_1
[vagrant@10 ~]$ hdfs dfs -cp /user/vagrant/nassalang/employee/employee/000000_1 /user/vagrant/nassalang/employee
2023-05-21 17:07:35,382 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2023-05-21 17:07:35,695 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
[vagrant@10 ~]$ hdfs dfs -ls /user/vagrant/nassalang/employee
Found 2 items
-rw-r--r-- 1 vagrant supergroup          23 2023-05-21 17:07 /user/vagrant/nassalang/employee/000000_1
drwxr-xr-x - vagrant supergroup          0 2023-05-21 17:05 /user/vagrant/nassalang/employee/employee
[vagrant@10 ~]$ hdfs dfs -ls /user/vagrant/nassalang/employee/employee
```

Pour le renommage du fichier, on a : **hdfs dfs -mv**
/user/vagrant/nassalang/employee/000000_1
/user/vagrant/nassalang/employee/liste_employee.csv

```
[vagrant@10 ~]$ hdfs dfs -mv /user/vagrant/nassalang/employee/000000_1 /user/vagrant/nassalang/employee/liste_employee.csv
[vagrant@10 ~]$ hdfs dfs -ls /user/vagrant/nassalang/employee
Found 1 items
-rw-r--r-- 1 vagrant supergroup          23 2023-05-21 17:07 /user/vagrant/nassalang/employee/liste_employee.csv
[vagrant@10 ~]$ |
```

Pour la copie du dossier /user/vagrant/[username]/employee se trouvant dans HDFS vers un répertoire local de votre machine, on a : **hdfs dfs -copyToLocal**
/user/vagrant/nassalang/employee /home/vagrant/nassalang

```
[vagrant@10 ~]$ hdfs dfs -copyToLocal /user/vagrant/nassalang/employee /home/vagrant/nassalang
2023-05-21 17:38:18,197 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
[vagrant@10 ~]$ ls /home/vagrant/nassalang
employee  hbase-3.0.0-alpha-3-bin.tar.gz  retails
[vagrant@10 ~]$ ls /home/vagrant/nassalang/employee/
liste_employee.csv
[vagrant@10 ~]$ |
```