

Hadoop HDFS 練習

0. 啟動 Hadoop

```
source /opt/hadoop/etc/hadoop/hadoop-env.sh
```

```
start-all.sh
```

1. **mkdir**，利用 mkdir 創資料夾

創二個 HDFS 目錄(hadoop1 和 hadoop2)

```
hadoop fs -mkdir /hadoop1 /hadoop2
```

```
bao@master:~$ hadoop fs -mkdir /hadoop1 /hadoop2
```

2. **ls**，利用 ls 查看 hdfs 有哪些資料夾

查看剛剛創建的 2 個資料夾

```
hadoop fs -ls /
```

```
hadoop dfs -ls /
```

```
bao@master:~$ hadoop fs -ls /
Found 2 items
drwxr-xr-x    - bao supergroup          0 2020-02-26 18:26 /hadoop1
drwxr-xr-x    - bao supergroup          0 2020-02-26 18:26 /hadoop2
```

利用 <http://master:9870> 查看剛剛創建的資料夾

(點選 Utilities - Browse the filesystem)

Browse Directory

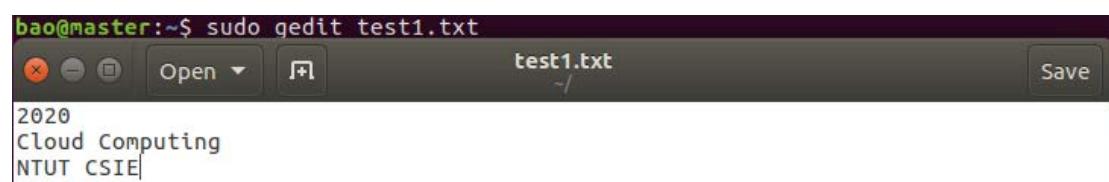
| Browse Directory | | | | | | | | |
|-----------------------------|------------|---------|------------|------------|-------|--------------|---------------|-------------|
| | | | | | | | | |
| | | | | | | | | |
| / | | | | | | | | |
| Show | 25 | entries | Permission | Owner | Group | Size | Last Modified | Replication |
| | | | | | | | | |
| <input type="checkbox"/> | drwxr-xr-x | - | bao | supergroup | 0 B | Feb 26 18:26 | 0 | 0 B |
| <input type="checkbox"/> | drwxr-xr-x | - | bao | supergroup | 0 B | Feb 26 18:26 | 0 | 0 B |
| Showing 1 to 2 of 2 entries | | | | | | | | |
| | | | | | | Previous | 1 | Next |

Hadoop, 2019.

3. **put**，利用 put 把本機檔案放到 hdfs 上

創一個 txt 檔

```
sudo gedit test1.txt
```



透過 put 指令把檔案放到 hdfs 上

```
hadoop fs -put test1.txt /hadoop1
```

```
bao@master:~$ hadoop fs -put test1.txt /hadoop1
2020-02-26 17:00:31,805 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
```

利用 ls 指令確認

```
hadoop fs -ls /hadoop1
```

```
bao@master:~$ hadoop fs -ls /hadoop1
Found 1 items
-rw-r--r--  2 bao supergroup          31 2020-02-26 17:00 /hadoop1/test1.txt
```

4. **get**，利用 get 把 hdfs 上的檔案下載到本機

```
hadoop fs -get /hadoop1/test1.txt fromHDFS.txt
```

```
bao@master:~$ hadoop fs -get /hadoop1/test1.txt fromHDFS.txt
2020-02-26 17:05:39,065 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
bao@master:~$ ls
core      Documents  examples.desktop   hdfs      Pictures    Templates    tmp
Desktop   Downloads  fromHDFS.txt       Music     Public      test1.txt   Videos
```

5. **cat**，透過 cat 指令確認檔案是否一樣

```
cat fromHDFS.txt
```

```
bao@master:~$ cat fromHDFS.txt
2020
Cloud Computing
NTUT CSIE
```

6. **rm**，利用 rm 把 hdfs 上的檔案刪除

```
hadoop fs -rm /hadoop1/test1.txt
```

```
bao@master:~$ hadoop fs -rm /hadoop1/test1.txt
Deleted /hadoop1/test1.txt
```

7. **rm -r**，利用 rm -r 刪掉整個資料夾

```
hadoop fs -rm -r /hadoop1
```

```
bao@master:~$ hadoop fs -rm -r /hadoop1
Deleted /hadoop1
```

#練習:

創一個檔案為 test1.txt 並把它放到 hdfs 上的/hadoop2 資料夾

```
bao@master:~$ hadoop fs -put test1.txt /hadoop2
2020-02-26 18:30:24,382 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
bao@master:~$ hadoop fs -ls /hadoop2
Found 1 items
-rw-r--r--  2 bao supergroup          31 2020-02-26 18:30 /hadoop2/test1.txt
```

8. **cat**，利用 cat 查看 hdfs 上的檔案內容

```
hadoop fs -cat /hadoop2/test1.txt
```

```
bao@master:~$ hadoop fs -cat /hadoop2/test1.txt
2020-02-26 17:29:01,586 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
2020
Cloud Computing
NTUT CSIE
```

9. **ls -R**，利用 ls -R 查看所有在 hdfs 上的檔案和資料夾

```
bao@master:~$ hadoop fs -ls -R /  
drwxr-xr-x  - bao supergroup          0 2020-02-26 17:24 /hadoop2  
-rw-r--r--  2 bao supergroup         31 2020-02-26 17:24 /hadoop2/test1.txt
```

10. **chgrp**，利用 chgrp 改變文件所屬的群組

```
hadoop fs -chgrp root /hadoop2/test1.txt
```

```
bao@master:~$ hadoop fs -ls /hadoop2/test1.txt  
-rw-r--r--  2 bao supergroup         31 2020-02-26 17:24 /hadoop2/test1.txt  
bao@master:~$ hadoop fs -cngrp root /hadoop2/test1.txt  
bao@master:~$ hadoop fs -ls /hadoop2/test1.txt  
-rw-r--r--  2 bao root             31 2020-02-26 17:24 /hadoop2/test1.txt
```

11. **chmod**，利用 chmod 改變文件的權限

```
hadoop fs -chmod 700 /hadoop2/test1.txt
```

```
bao@master:~$ hadoop fs -ls /hadoop2/test1.txt  
-rw-r--r--  2 bao root             31 2020-02-26 17:24 /hadoop2/test1.txt  
bao@master:~$ hadoop fs -chmod 700 /hadoop2/test1.txt  
bao@master:~$ hadoop fs -ls /hadoop2/test1.txt  
-rwx-----  2 bao root             31 2020-02-26 17:24 /hadoop2/test1.txt
```

#練習：

創建一個資料夾/hadoop3 到 hdfs，利用 ls 指令查看目前 hdfs 上全部檔案和資料夾

```
bao@master:~$ hadoop fs -mkdir /hadoop3  
bao@master:~$ hadoop fs -ls -R /  
drwxr-xr-x  - bao supergroup          0 2020-02-26 17:24 /hadoop2  
-rwx-----  2 bao root             31 2020-02-26 17:24 /hadoop2/test1.txt  
drwxr-xr-x  - bao supergroup          0 2020-02-26 17:59 /hadoop3
```

12. **cp**，將文件從 HDFS 原本路徑複製到 HDFS 目標路徑

```
hadoop fs -cp /hadoop2/test1.txt /hadoop3
```

```
bao@master:~$ hadoop fs -ls -R /  
drwxr-xr-x  - bao supergroup          0 2020-02-26 18:30 /hadoop2  
-rwx-----  2 bao root             31 2020-02-26 18:30 /hadoop2/test1.txt  
drwxr-xr-x  - bao supergroup          0 2020-02-27 11:36 /hadoop3  
bao@master:~$ hadoop fs -cp /hadoop2/test1.txt /hadoop3  
2020-02-27 11:36:54,223 INFO sasl.SaslDataTransferClient: SASL encryption trust  
check: localHostTrusted = false, remoteHostTrusted = false  
2020-02-27 11:36:54,362 INFO sasl.SaslDataTransferClient: SASL encryption trust  
check: localHostTrusted = false, remoteHostTrusted = false  
bao@master:~$ hadoop fs -ls -R /  
drwxr-xr-x  - bao supergroup          0 2020-02-26 18:30 /hadoop2  
-rwx-----  2 bao root             31 2020-02-26 18:30 /hadoop2/test1.txt  
drwxr-xr-x  - bao supergroup          0 2020-02-27 11:36 /hadoop3  
-rw-r--r--  2 bao supergroup         31 2020-02-27 11:36 /hadoop3/test1.txt
```

13. **du**，顯示目錄中所有文件的大小

```
hadoop fs -du /hadoop2
```

```
bao@master:~$ hadoop fs -du /hadoop2  
31 62 /hadoop2/test1.txt
```

14. **du -s**，顯示該目錄/文件的總大小

```
bao@master:~$ hadoop fs -du -s /hadoop2  
31 62 /hadoop2
```

15. **expunge**，清空垃圾桶

```
bao@master:~$ hadoop fs -expunge
```

16. **getmerge**，將來源目錄 <src> 下所有文件集合到本機 <localdst> 檔案內

語法：hadoop fs -getmerge <src> <localdst>

在本機創 in1 資料夾，新增 2 檔案 input1 和 input2 存到 in1 資料夾，用 put 把 in1 資料夾丟到 hdfs 的 hadoop2

```
mkdir -p in1
```

```
echo "this is one;" > in1/input1
```

```
echo "this is two;" > in1/input2
```

```
hadoop fs -put in1 /hadoop2
```

```
bao@master:~$ mkdir -p in1
bao@master:~$ echo "this is one;" > in1/input1
bao@master:~$ echo "this is two;" > in1/input2
bao@master:~$ hadoop fs -put in1 /hadoop2
2020-02-27 11:56:14,239 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
2020-02-27 11:56:15,132 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
```

接著利用 getmerge 把 in1 裡面的檔案存成 merge.txt 下載到本機，查看結果：

```
hadoop fs -getmerge /hadoop2/in1 merge.txt
```

```
cat merge.txt
```

```
bao@master:~$ hadoop fs -getmerge /hadoop2/in1 merge.txt
2020-02-27 11:57:43,508 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
bao@master:~$ ls
Desktop examples.desktop in1 Pictures test1.txt
Documents fromHDFS.txt merge.txt Public tmp
Downloads hdfs Music Templates Videos
bao@master:~$ cat merge.txt
this is one;
this is two;
```

17. **moveFromLocal**，將 local 端的資料夾剪下移動到 HDFS 上

```
hadoop fs -moveFromLocal in1 /hadoop3
```

```
bao@master:~$ ls
Desktop examples.desktop in1 Pictures test1.txt
Documents fromHDFS.txt merge.txt Public tmp
Downloads hdfs Music Templates Videos
bao@master:~$ hadoop fs -moveFromLocal in1 /hadoop3
2020-02-27 12:01:35,103 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
2020-02-27 12:01:35,231 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
bao@master:~$ ls
Desktop examples.desktop merge.txt Public tmp
Documents fromHDFS.txt Music Templates Videos
Downloads hdfs Pictures test1.txt
bao@master:~$ hadoop fs -ls /hadoop3
Found 2 items
drwxr-xr-x - bao supergroup 0 2020-02-27 12:01 /hadoop3/in1
-rw-r--r-- 2 bao supergroup 31 2020-02-27 11:36 /hadoop3/test1.txt
```

18. **stat**，印出時間資訊

```
hadoop fs -stat /hadoop3
```

```
bao@master:~$ hadoop fs -stat /hadoop3
2020-02-27 04:01:35
```

19. `tail`，將文件的最後 1k 內容輸出

語法：`hadoop fs -tail [-f] 檔案 (-f 參數用來顯示如果檔案增大，則秀出被 append 上得內容)`

把`/opt/hadoop/LICENSE.txt`檔案上傳 hdfs 上 `hadoop3` 資料夾，利用 `tail` 查看文件最後 1K 內容

```
hadoop fs -put /opt/hadoop/LICENSE.txt /hadoop3
```

```
hadoop fs -tail /hadoop3/LICENSE.txt
```

```
bao@master:~$ hadoop fs -put /opt/hadoop/LICENSE.txt /hadoop3
2020-02-27 12:08:11,562 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
bao@master:~$ hadoop fs -tail /hadoop3/LICENSE.txt
2020-02-27 12:08:27,132 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
      this trademark restriction does not form part of this License.

      Creative Commons may be contacted at https://creativecommons.org/.

-----
For: hadoop-hdfs-project/hadoop-hdfs/src/main/java/org/apache/hadoop/hdfs
/server/datanode/checker/AbstractFuture.java and
hadoop-hdfs-project/hadoop-hdfs/src/main/java/org/apache/hadoop/hdfs
/server/datanode/checker/TimeoutFuture.java

Copyright (c) 2007 The Guava Authors

Licensed under the Apache License, Version 2.0 (the "License"); you may not
use this file except in compliance with the License. You may obtain a copy of
the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS, WITHOUT
WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the
License for the specific language governing permissions and limitations under
the License.
```

20. `test`，測試檔案，`-e` 檢查檔案是否存在(0=存在, 1=否)，`-z` 檢查文件是否為空(0=空, 1=不為空)，`-d` 檢查是否為目錄(0=存在, 1=否)，利用 `echo $?` 來看回傳值為 0 or 1

`-e`

```
hadoop -test -e /hadoop3/test1.txt
```

```
echo $?
```

```
bao@master:~$ hadoop fs -ls -R /
drwxr-xr-x  - bao supergroup          0 2020-02-27 11:56 /hadoop2
drwxr-xr-x  - bao supergroup          0 2020-02-27 11:56 /hadoop2/in1
-rw-r--r--  2 bao supergroup         14 2020-02-27 11:56 /hadoop2/in1/input1
-rw-r--r--  2 bao supergroup         14 2020-02-27 11:56 /hadoop2/in1/input2
-rwx-----  2 bao root              31 2020-02-26 18:30 /hadoop2/test1.txt
drwxr-xr-x  - bao supergroup          0 2020-02-27 12:08 /hadoop3
-rw-r--r--  2 bao supergroup    150569 2020-02-27 12:08 /hadoop3/LICENSE.txt
drwxr-xr-x  - bao supergroup          0 2020-02-27 12:01 /hadoop3/in1
-rw-r--r--  2 bao supergroup         14 2020-02-27 12:01 /hadoop3/in1/input1
-rw-r--r--  2 bao supergroup         14 2020-02-27 12:01 /hadoop3/in1/input2
-rw-r--r--  2 bao supergroup         31 2020-02-27 11:36 /hadoop3/test1.txt
bao@master:~$ hadoop fs -test -e /hadoop3/test1.txt
bao@master:~$ echo $?
0
bao@master:~$ hadoop fs -test -e /hadoop3/test2.txt
bao@master:~$ echo $?
1
```

-z

```
hadoop fs -test -z /hadoop3/
```

```
echo $?
```

```
bao@master:~$ hadoop fs -ls -R /
drwxr-xr-x    - bao supergroup          0 2020-02-27 11:56 /hadoop2
drwxr-xr-x    - bao supergroup          0 2020-02-27 11:56 /hadoop2/in1
-rw-r--r--    2 bao supergroup         14 2020-02-27 11:56 /hadoop2/in1/input1
-rw-r--r--    2 bao supergroup         14 2020-02-27 11:56 /hadoop2/in1/input2
-rwx-----    2 bao root             31 2020-02-26 18:30 /hadoop2/test1.txt
drwxr-xr-x    - bao supergroup          0 2020-02-27 12:08 /hadoop3
-rw-r--r--    2 bao supergroup      150569 2020-02-27 12:08 /hadoop3/LICENSE.txt
drwxr-xr-x    - bao supergroup          0 2020-02-27 12:01 /hadoop3/in1
-rw-r--r--    2 bao supergroup         14 2020-02-27 12:01 /hadoop3/in1/input1
-rw-r--r--    2 bao supergroup         14 2020-02-27 12:01 /hadoop3/in1/input2
-rw-r--r--    2 bao supergroup         31 2020-02-27 11:36 /hadoop3/test1.txt
bao@master:~$ hadoop fs -test -z /hadoop3/
bao@master:~$ echo $?
0
bao@master:~$ hadoop fs -test -z /hadoop3/test1.txt
bao@master:~$ echo $?
1
```

-d

```
hadoop fs -test -d /hadoop2 ; echo $?
```

```
bao@master:~$ hadoop fs -ls /
Found 2 items
drwxr-xr-x    - bao supergroup          0 2020-02-27 11:56 /hadoop2
drwxr-xr-x    - bao supergroup          0 2020-02-27 12:08 /hadoop3
bao@master:~$ hadoop fs -test -d /hadoop2 ; echo $?
0
bao@master:~$ hadoop fs -test -d /hadoop4 ; echo $?
1
```

21. **text**，將檔案（如壓縮檔, textrecordinputstream）輸出為純文字格式

語法 : `hadoop fs -text <src>`

創一個 test1.txt，把 test1.txt 壓縮成 test1.tar.gz，上傳到 hadoop3

```
tar zcvf test1.tar.gz test1.txt
```

```
hadoop fs -put test1.tar.gz /hadoop3
```

```
bao@master:~$ ls
Desktop  examples.desktop  merge.txt  Public      tmp
Documents fromHDFS.txt   Music     Templates  Videos
Downloads hdfs           Pictures   test1.txt
bao@master:~$ tar zcvf test1.tar.gz test1.txt
test1.txt
bao@master:~$ ls
Desktop  examples.desktop  merge.txt  Public      test1.txt
Documents fromHDFS.txt   Music     Templates  tmp
Downloads hdfs           Pictures   test1.tar.gz  Videos
bao@master:~$ hadoop fs -put test1.tar.gz /hadoop3
2020-02-27 12:27:42,005 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
```

查看結果

```
hadoop fs -text /hadoop3/test1.tar.gz
```

```
bao@master:~$ hadoop fs -ls /hadoop3
Found 4 items
-rw-r--r--  2 bao supergroup      150569 2020-02-27 12:08 /hadoop3/LICENSE.txt
drwxr-xr-x  - bao supergroup          0 2020-02-27 12:01 /hadoop3/in1
-rw-r--r--  2 bao supergroup      154 2020-02-27 12:27 /hadoop3/test1.tar.gz
-rw-r--r--  2 bao supergroup      31 2020-02-27 11:36 /hadoop3/test1.txt
bao@master:~$ hadoop fs -text /hadoop3/test1.tar.gz
2020-02-27 12:30:18,917 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
test1.txt000064400000000000000000000000003713625444125011363 0ustar rootroot2020
Cloud Computing
NTUT CSIE
```

22. **touchz**，建立一個空文件

```
hadoop fs -touchz /hadoop4
```

```
bao@master:~$ hadoop fs -touchz /hadoop4
bao@master:~$ hadoop fs -ls /
Found 3 items
drwxr-xr-x  - bao supergroup          0 2020-02-27 11:56 /hadoop2
drwxr-xr-x  - bao supergroup          0 2020-02-27 12:27 /hadoop3
-rw-r--r--  2 bao supergroup          0 2020-02-27 12:32 /hadoop4
bao@master:~$ hadoop fs -test -z /hadoop4 ; echo $?
0
```

#觀察 block 分佈情形

1. 產生日錄

```
hadoop fs -mkdir /tmp
```

```
bao@master:~$ hadoop fs -mkdir /tmp
bao@master:~$ hadoop fs -ls /
Found 4 items
drwxr-xr-x  - bao supergroup          0 2020-02-27 11:56 /hadoop2
drwxr-xr-x  - bao supergroup          0 2020-02-27 12:27 /hadoop3
-rw-r--r--  2 bao supergroup          0 2020-02-27 12:32 /hadoop4
drwxr-xr-x  - bao supergroup          0 2020-02-27 12:34 /tmp
```

2. 產生一個 100MB 的檔案

```
dd if=/dev/zero of=100mb.img bs=1M count=100
```

```
hadoop fs -put 100mb.img /tmp
```

```
bao@master:~$ dd if=/dev/zero of=100mb.img bs=1M count=100
100+0 records in
100+0 records out
104857600 bytes (105 MB, 100 MiB) copied, 0.0931984 s, 1.1 GB/s
bao@master:~$ ls
100mb.img  Downloads      hdfs      Pictures    test1.tar.gz  Videos
Desktop    examples.desktop merge.txt  Public      test1.txt
Documents  fromHDFS.txt   Music     Templates  tmp
bao@master:~$ hadoop fs -put 100mb.img /tmp
2020-02-27 12:38:29,845 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
bao@master:~$ hadoop fs -ls /tmp
Found 1 items
-rw-r--r--  2 bao supergroup  104857600 2020-02-27 12:38 /tmp/100mb.img
```

利用 hdfs fsck....觀察分布情形

hdfs fsck /tmp/100mb.img -files -blocks -locations

```
bao@master:~$ hdfs fsck /tmp/100mb.img -files -blocks -locations
Connecting to namenode via http://master:9870/fsck?ugi=bao&files=1&blocks=1&locations=1&path=%2Ftmp%2F100mb.img
FSCK started by bao (auth:SIMPLE) from /192.168.91.128 for path /tmp/100mb.img at Thu Feb 27 13:27:35 CST 2020
/tmp/100mb.img 104857600 bytes, replicated: replication=2, 1 block(s): Under replicated BP-134250155-192.168.91.128-1582540231094:blk_1073741835_1011. Target R replicas is 2 but found 1 live replica(s), 0 decommissioned replica(s), 0 decommissioning replica(s).
0. BP-134250155-192.168.91.128-1582540231094:blk_1073741835_1011 len=104857600 L live_repl=1 [DatanodeInfoWithStorage[192.168.91.128:9866,DS-9f88aa5b-f8e7-4e33-b50a-59b35c781f2e,DISK]]
```

Status: HEALTHY

| | |
|-----------------------|---|
| Number of data-nodes: | 1 |
| Number of racks: | 1 |
| Total dirs: | 0 |
| Total symlinks: | 0 |

Replicated Blocks:

| | |
|------------------------------|---------------------------------|
| Total size: | 104857600 B |
| Total files: | 1 |
| Total blocks (validated): | 1 (avg. block size 104857600 B) |
| Minimally replicated blocks: | 1 (100.0 %) |
| Over-replicated blocks: | 0 (0.0 %) |
| Under-replicated blocks: | 1 (100.0 %) |
| Mis-replicated blocks: | 0 (0.0 %) |
| Default replication factor: | 2 |
| Average block replication: | 1.0 |
| Missing blocks: | 0 |
| Corrupt blocks: | 0 |
| Missing replicas: | 1 (50.0 %) |

Erasure Coded Block Groups:

| | |
|--|-----|
| Total size: | 0 B |
| Total files: | 0 |
| Total block groups (validated): | 0 |
| Minimally erasure-coded block groups: | 0 |
| Over-erasure-coded block groups: | 0 |
| Under-erasure-coded block groups: | 0 |
| Unsatisfactory placement block groups: | 0 |
| Average block group size: | 0.0 |
| Missing block groups: | 0 |
| Corrupt block groups: | 0 |
| Missing internal blocks: | 0 |

```
FSCK ended at Thu Feb 27 13:27:35 CST 2020 in 38 milliseconds
```

```
The filesystem under path '/tmp/100mb.img' is HEALTHY
```

3. 修改副本個數再觀察一次，利用-setrep修改副本個數

```
hadoop fs -setrep 3 /tmp/100mb.img
```

```
bao@master:~$ hadoop fs -setrep 3 /tmp/100mb.img
Replication 3 set: /tmp/100mb.img

bao@master:~$ hdfs fsck /tmp/100mb.img -files -blocks -locations
Connecting to namenode via http://master:9870/fsck?ugi=bao&files=1&blocks=1&locations=1&path=%2Ftmp%2F100mb.img
FSCK started by bao (auth:SIMPLE) from /192.168.91.128 for path /tmp/100mb.img at Thu Feb 27 14:58:31 CST 2020
/tmp/100mb.img 104857600 bytes, replicated: replication=3, 1 block(s): Under replicated BP-134250155-192.168.91.128-1582540231094:blk_1073741835_1011. Target R replicas is 3 but found 1 live replica(s), 0 decommissioned replica(s), 0 decommissioning replica(s).
0. BP-134250155-192.168.91.128-1582540231094:blk_1073741835_1011 len=104857600 L
ive_repl=1 [DatanodeInfoWithStorage[192.168.91.128:9866,DS-9f88aa5b-f8e7-4e33-b
50a-59b35c781f2e,DISK]]


Status: HEALTHY
Number of data-nodes: 1
Number of racks: 1
Total dirs: 0
Total symlinks: 0

Replicated Blocks:
Total size: 104857600 B
Total files: 1
Total blocks (validated): 1 (avg. block size 104857600 B)
Minimally replicated blocks: 1 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 1 (100.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 2
Average block replication: 1.0
Missing blocks: 0
Corrupt blocks: 0
Missing replicas: 2 (66.666664 %)

Erasure Coded Block Groups:
Total size: 0 B
Total files: 0
Total block groups (validated): 0
Minimally erasure-coded block groups: 0
Over-erasure-coded block groups: 0
Under-erasure-coded block groups: 0
Unsatisfactory placement block groups: 0
Average block group size: 0.0
Missing block groups: 0
Corrupt block groups: 0
Missing internal blocks: 0
FSCK ended at Thu Feb 27 14:58:31 CST 2020 in 125 milliseconds

The filesystem under path '/tmp/100mb.img' is HEALTHY
```

#字數統計(WordCount)

創一個檔案

```
echo "Cloud computing class" > input
```

```
bao@master:~$ echo "Cloud computing class" > input
bao@master:~$ ls
100mb.img  Downloads      hdfs      Music      Templates      tmp
Desktop    examples.desktop  input      Pictures   test1.tar.gz  Videos
Documents  fromHDFS.txt    merge.txt  Public    test1.txt
```

接著在 hdfs 創一個資料夾，把檔案丟上去

```
hadoop fs -mkdir /input
```

```
hadoop fs -put input /input
```

```
bao@master:~$ hadoop fs -mkdir /input
bao@master:~$ hadoop fs -put input /input
2020-02-27 14:11:49,848 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
```

將產生的結果存到 output

```
hadoop jar /opt/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-
3.2.1.jar wordcount /input output
```

執行完成畫面

```
Map-Reduce Framework
  Map input records=1
  Map output records=3
  Map output bytes=34
  Map output materialized bytes=46
  Input split bytes=95
  Combine input records=3
  Combine output records=3
  Reduce input groups=3
  Reduce shuffle bytes=46
  Reduce input records=3
  Reduce output records=3
  Spilled Records=6
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=389
  CPU time spent (ms)=2450
  Physical memory (bytes) snapshot=441409536
  Virtual memory (bytes) snapshot=5312307200
  Total committed heap usage (bytes)=292552704
  Peak Map Physical memory (bytes)=251396096
  Peak Map Virtual memory (bytes)=2653122560
  Peak Reduce Physical memory (bytes)=190013440
  Peak Reduce Virtual memory (bytes)=2659184640
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=22
File Output Format Counters
  Bytes Written=28
```

```
hadoop fs -ls -R output
```

發現所有的執行成果會記錄在 part-r-*，使用 cat 指令查看檔案內容

```
hadoop fs -cat output/*
```

```
bao@master:~$ hadoop fs -ls -R output
-rw-r--r--  2 bao supergroup          0 2020-02-27 14:17 output/_SUCCESS
-rw-r--r--  2 bao supergroup        28 2020-02-27 14:17 output/part-r-00000
bao@master:~$ hadoop fs -cat output/*
2020-02-27 14:23:56,900 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
Cloud      1
class      1
computing   1
```

範例二：用標準表示法過濾內容 grep

grep 命令是擷取文件裡特定字元，在 Hadoop example 中此指令擷取文件中有此指定文字的字串，作計數統計

利用剛創建的 input，從/input 目錄中所有檔案中找出符合 cl 後面接 a-z 字母一個以上字串

```
hadoop jar /opt/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-
3.2.1.jar grep /input output2 'cl[a-z.]+'
```

```
bao@master:~$ hadoop jar /opt/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-exa
mples-3.2.1.jar grep /input output2 'cl[a-z.]+'
```

執行畫面

```
Map-Reduce Framework
  Map input records=1
  Map output records=1
  Map output bytes=14
  Map output materialized bytes=22
  Input split bytes=126
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=22
  Reduce input records=1
  Reduce output records=1
  Spilled Records=2
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=72
  CPU time spent (ms)=3430
  Physical memory (bytes) snapshot=474251264
  Virtual memory (bytes) snapshot=5312106496
  Total committed heap usage (bytes)=374341632
  Peak Map Physical memory (bytes)=286478336
  Peak Map Virtual memory (bytes)=2652938240
  Peak Reduce Physical memory (bytes)=187772928
  Peak Reduce Virtual memory (bytes)=2659168256
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=108
File Output Format Counters
  Bytes Written=8
```

結果画面

```
bao@master:~$ hadoop fs -ls output2
Found 2 items
-rw-r--r--  2 bao supergroup          0 2020-02-27 14:51 output2/_SUCCESS
-rw-r--r--  2 bao supergroup          8 2020-02-27 14:51 output2/part-r-00000
bao@master:~$ hadoop fs -cat output2/*
2020-02-27 14:53:11,426 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
1      class
```