

최신기술프로젝트15주차 과제

HDFS Assignment

컴퓨터공학과

2017103948 고다경

1. Load data on HDFS

```
→ hadoop-2.7.6 ./bin/hadoop fs -put input2.txt input2
→ hadoop-2.7.6 ./bin/hadoop fs -ls input
-rw-r--r--    1 hadoop supergroup        18 2021-04-09 13:23 input
→ hadoop-2.7.6 ./bin/hadoop fs -ls input2
-rw-r--r--    3 hadoop supergroup         6 2021-04-09 20:30 input2
```

⇒ input2.txt란 파일을 업로드한 후, 파일 목록 확인을 하였습니다.

2. Replication Configuration : 1 -> 3

3. Block-size Configuration: 128MB -> 64MB

```

<configuration>
  <property>
    <name>dfs.replication</name>
    <value>3</value>
  </property>
  <property>
    <name>dfs.blocksize</name>
    <value>67108864</value>
  </property>
</configuration>

```

⇒ `hdfs-site.xml` 파일에서 Replication을 3으로, Block-size를 67108864로 수정하였습니다.

```
[→ hadoop-2.7.6 ./bin/hdfs dfs -stat "%b %F %g %n %o %r %u %y %Y" input2
6 regular file supergroup input2 67108864 3 hadoop 2021-04-09 11:30:09 1617967809727
```

⇒ <path>에 있는 파일/디렉토리에 대한 통계를 print합니다.

⇒ %b: bytes로 된 파일 사이즈

⇒ %F: 유형

⇒ %g: 소유자의 그룹 이름

⇒ %n: 이름

⇒ %o: 블록 크기

현재 블록 크기는 정해진 대로 6710884인 것을 볼 수 있습니다.

⇒ %r: replication

현재 replication이 3개 인 것을 볼 수 있습니다.

⇒ %u: owner의 USERNAME

⇒ %y: 수정한 날짜(UTC 날짜로 표기)

⇒ %Y: 수정한 날짜(1970년 1월 1일 UTC 이후의 milliseconds 표시)

```
[→ hadoop-2.7.6 ./bin/hdfs dfsadmin -report
Configured Capacity: 94997024768 (88.47 GB)
Present Capacity: 87898128384 (81.86 GB)
DFS Remaining: 87898075136 (81.86 GB)
DFS Used: 53248 (52 KB)
DFS Used%: 0.00%
Under replicated blocks: 1
Blocks with corrupt replicas: 0
Missing blocks: 0
Missing blocks (with replication factor 1): 0

-----
Live datanodes (1):

Name: 127.0.0.1:50010 (localhost)
Hostname: class.novalocal
Decommission Status : Normal
Configured Capacity: 94997024768 (88.47 GB)
DFS Used: 53248 (52 KB)
Non DFS Used: 3135684608 (2.92 GB)
DFS Remaining: 87898075136 (81.86 GB)
DFS Used%: 0.00%
DFS Remaining%: 92.53%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 1
Last contact: Fri Apr 09 22:26:58 KST 2021
```

⇒ 현재 HDFS의 용량과 전체 노드의 용량 상태를 확인했습니다.

```

+ hadoop-2.7.6 ./bin/hadoop fsck / -files -blocks
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

Connecting to namenode via http://localhost:50070/fsck?ugi=hadoop&files=1&blocks=1&path=%2F
FSCK started by hadoop (auth:SIMPLE) from /127.0.0.1 for path / at Fri Apr 09 20:32:09 KST 2021
/ <dir>
/user <dir>
/user/hadoop <dir>
/user/hadoop/input 18 bytes, 1 block(s): OK
0. BP-1419620828-127.0.1.1-1617942046591:blk_1073741825_1001 len=18 repl=1

/user/hadoop/input2 6 bytes, 1 block(s): Under replicated BP-1419620828-127.0.1.1-1617942046591:blk_1073741826_1002.
Target Replicas is 3 but found 1 replica(s).
0. BP-1419620828-127.0.1.1-1617942046591:blk_1073741826_1002 len=6 repl=1

/user/hadoop-2.7.6 <dir>
Status: HEALTHY
Total size: 24 B
Total dirs: 4
Total files: 2
Total symlinks: 0
Total blocks (validated): 2 (avg. block size 12 B)
Minimally replicated blocks: 2 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 1 (50.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 1
Average block replication: 1.0
Corrupt blocks: 0
Missing replicas: 2 (50.0 %)
Number of data-nodes: 1
Number of racks: 1
FSCK ended at Fri Apr 09 20:32:09 KST 2021 in 2 milliseconds

The filesystem under path '/' is HEALTHY
+ hadoop-2.7.6

```

⇒ Replication이 1일 때, Default replication factor가 1인 것을 볼 수 있습니다.

```

+ hadoop-2.7.6 ./bin/hadoop fsck / -files -blocks
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

Connecting to namenode via http://localhost:50070/fsck?ugi=hadoop&files=1&blocks=1&path=%2F
FSCK started by hadoop (auth:SIMPLE) from /127.0.0.1 for path / at Fri Apr 09 22:27:09 KST 2021
/ <dir>
/user <dir>
/user/hadoop <dir>
/user/hadoop/input 18 bytes, 1 block(s): OK
0. BP-1419620828-127.0.1.1-1617942046591:blk_1073741825_1001 len=18 repl=1

/user/hadoop/input2 6 bytes, 1 block(s): Under replicated BP-1419620828-127.0.1.1-1617942046591:blk_1073741826_1002. Target Replicas is 3 but found 1 replica(s).
0. BP-1419620828-127.0.1.1-1617942046591:blk_1073741826_1002 len=6 repl=1

/user/hadoop-2.7.6 <dir>
Status: HEALTHY
Total size: 24 B
Total dirs: 4
Total files: 2
Total symlinks: 0
Total blocks (validated): 2 (avg. block size 12 B)
Minimally replicated blocks: 2 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 1 (50.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 3
Average block replication: 1.0
Corrupt blocks: 0
Missing replicas: 2 (50.0 %)
Number of data-nodes: 1
Number of racks: 1
FSCK ended at Fri Apr 09 22:27:09 KST 2021 in 18 milliseconds

The filesystem under path '/' is HEALTHY

```

⇒ Replication을 3으로 수정한 후, Default replication factor가 3으로 변경된 것을 확인했습니다.