

[Hadoop][Lab] 04. HDFS

학습 목표

가. 리눅스 환경에서 HDFS에 대해 이해해 본다.

나. Hadoop 시스템에 기본 명령어를 실습을 통해 알아본다.

목차

[Hadoop][Lab] 04. HDFS

4-1 HDFS 상의 파일, 디렉토리 확인(ls)

4-2 HDFS 상의 디렉터리 생성(mkdir)

4-3 HDFS 상의 파일 생성(touchz)

4-4 로컬에서 HDFS상으로 파일 복사(copyFromLocal)

4-5 HDFS 상의 파일 내용 확인(cat)

4-6 HDFS 상의 파일 내용 복사(cp)

4-7 HDFS 상의 파일 내용 옮기기(mv)

4-8 HDFS 상의 파일 업로드하기(put)

4-1 hadoop 파일, 디렉토리 확인(ls)

Description : HDFS Command to display the list of Files and Directories in HDFS.
(HDFS의 파일 및 디렉터리 리스트를 보여주기 위한 HDFS 명령어이다.)

hdfs dfs -ls /

[Result]

```
[training@localhost ~]$ hdfs dfs -ls /
```

Found 3 items

```
drwxrwxrwt - hdfs supergroup      0 2013-09-10 10:35 /tmp
drwxrwxrwx - hue supergroup       0 2013-09-10 10:37 /user
drwxr-xr-x - hdfs supergroup      0 2013-09-05 20:08 /var
```

4-2 HDFS 상의 디렉터리 생성(mkdir)

Description : HDFS command to create the directory in HDFS

Usage : **hdfs dfs -mkdir /directory_name**

명령 내용 : 디렉터리 생성 후, 확인

hdfs dfs -mkdir /new_dir

hdfs dfs -ls /

[Result]

```
[training@localhost ~]$ hdfs dfs -ls /
```

Found 4 items

drwxr-xr-x	- training	supergroup	0	2019-02-10 10:09	/new_dir
drwxrwxrwt	- hdfs	supergroup	0	2013-09-10 10:35	/tmp
drwxrwxrwx	- hue	supergroup	0	2013-09-10 10:37	/user
drwxr-xr-x	- hdfs	supergroup	0	2013-09-05 20:08	/var

4-3 HDFS 상의 파일 생성(touchz)

Description : HDFS command to create a file in HDFS with file size 0 bytes.

Usage : **hdfs dfs -touchz /directory_name/filename**

명령 내용 : 파일 생성 후, 확인

hdfs dfs -mkdir /new_dir/sample

hdfs dfs -ls /new_dir

[Result]

```
[training@localhost ~]$ hdfs dfs -ls /new_dir
```

Found 1 items

drwxr-xr-x	- training	supergroup	0	2019-02-10 10:13	/new_dir/sample
------------	------------	------------	---	------------------	-----------------

4-4 로컬에서 HDFS상으로 파일 복사(copyFromLocal)

Description : HDFS Command to copy the file from a Local file system to HDFS.

Usage : **hdfs dfs -copyFromLocal <Local 원본> <hdfs 복사장소>**

사전 텍스트 파일을 하나 만들자.

```
[training@localhost ~]$ cat > text.txt
```

```
I am a boy
```

```
Hello!
```

```
You are great!^Z
```

```
[1]+ Stopped cat > text.txt
```

```
[training@localhost ~]$
```

명령 : 위의 리눅스 시스템 상의 파일을 HDFS의 상의 디렉터리 내에 복사한다.

```
hdfs dfs -copyFromLocal /home/training/text.txt /new_dir
```

```
hdfs dfs -ls /new_dir
```

[Result]

```
[training@localhost ~]$ hdfs dfs -ls /new_dir
```

```
Found 2 items
```

```
drwxr-xr-x - training supergroup 0 2019-02-10 10:13 /new_dir/sample
```

```
-rw-r--r-- 1 training supergroup 18 2019-02-10 10:26 /new_dir/text.txt
```

4-5 HDFS 상의 파일 내용 확인(cat)

Description : HDFS Command that reads a file on HDFS and prints the content of that file to the standard output.

(HDFS상의 파일을 읽은 후, 해당 파일의 내용을 표준 출력으로 출력하는 명령이다.)

Usage : **hdfs dfs -cat /directory_name/filename**

명령 내용 : 파일 생성 후, 확인

```
hdfs dfs -cat /new_dir/text.txt
```

[Result]

```
[training@localhost ~]$ hdfs dfs -cat /new_dir/text.txt
```

I am a boy

Hello!

4-6 HDFS 상의 파일 내용 복사(cp)

Description : HDFS Command to copy files from source to destination. This command allows multiple sources as well, in which case the destination must be a directory.

(HDFS 원본에서 대상으로 파일을 복사하는 명령입니다. 이 명령은 여러 소스도 허용합니다. 이 경우 대상은 **디렉토리** 여야합니다.)

Usage : **hdfs dfs -cp <src> <dest>**

명령 내용 : 파일 복사 후, 확인

```
hdfs dfs -cp /new_dir/text.txt /new_dir/test01.txt
```

```
hdfs dfs -ls /new_dir
```

[Result]

```
[training@localhost ~]$ hdfs dfs -ls /new_dir
```

Found 3 items

```
drwxr-xr-x - training supergroup      0 2019-02-10 10:13 /new_dir/sample
-rw-r--r--  1 training supergroup    18 2019-02-10 10:35 /new_dir/test01.txt
-rw-r--r--  1 training supergroup    18 2019-02-10 10:26 /new_dir/text.txt
```

(실습 1)

(가) cp 명령을 이용하여 text.txt 파일을 /new_dir2 디렉터리로 복사하기

(나) test01.txt 파일을 Local(리눅스 환경)에 복사하기(copyToLocal 명령 이용)

(다) test01.txt 파일을 지우기(rm)

(라) new_dir2 디렉터리 지우기

4-7 HDFS 상의 파일 내용 옮기기(mv)

Description : HDFS Command to move files from source to destination. This Command allows multiple sources as well, in which case the destination needs to be a directory.

(HDFS 원본에서 대상으로 파일을 이동하는 명령입니다. 이 명령은 여러 소스도 허용합니다. 이 경우 대상은 **디렉토리** 여야합니다.)

Usage : **hdfs dfs -mv <src> <dest>**

명령 내용 : 파일 복사 후, 확인

hdfs dfs -cp /new_dir/text.txt /new_dir/test02.txt

hdfs dfs -ls /new_dir

[Result]

[training@localhost ~]\$ **hdfs dfs -ls /new_dir**

Found 4 items

```
drwxr-xr-x  - training supergroup      0 2019-02-10 10:13 /new_dir/sample
-rw-r--r--  1 training supergroup    18 2019-02-10 10:35 /new_dir/test01.txt
-rw-r--r--  1 training supergroup    18 2019-02-10 10:43 /new_dir/test02.txt
-rw-r--r--  1 training supergroup    18 2019-02-10 10:26 /new_dir/text.txt
```

4-8 HDFS 상의 파일 업로드하기(put)

Description : HDFS Command to copy single source or multiple sources from local file system to the destination file system.

(HDFS 단일 소스 또는 여러 소스를 로컬 파일 시스템에서 대상 파일 시스템으로 복사하는 명령이다.)

Usage : **hdfs dfs -put <local src> <HDFS dest>**

명령 내용 : 파일 복사 후, 확인

hdfs dfs -mkdir /new_dir2

hdfs dfs -put /home/training/text.txt /new_dir2

hdfs dfs -ls /new_dir2

[Result]

```
[training@localhost ~]$ hdfs dfs -mkdir /new_dir2
[training@localhost ~]$ hdfs dfs -put /home/training/text.txt /new_dir2
[training@localhost ~]$ hdfs dfs -ls /new_dir2
Found 1 items
-rw-r--r-- 1 training supergroup      18 2019-02-10 10:51 /new_dir2/text.txt
```

* **get 명령은 반대로 hdfs 시스템으로부터 local file system으로 파일을 복사하는 명령이다.**

(실습 2)

(가) 어떤 HDFS 시스템상의 디렉터리에 파일과 디렉터리 그리고 바이트 수를 계산하고 싶다면 어떤 명령을 이용하여 확인이 가능할까?

(나) 위에서 확인한 /new_dir 의 파일들의 용량을 확인해보자.(du)

(실습3) 아래 명령의 어떤 기능을 하는지 찾아서 작성해 보자.

hdfs dfs -chown

hdfs dfs -chgrp

hdfs dfs -tail

hdfs dfs -chmod

hdfs dfs -appendToFile

[기타 관리 명령어]

서브 명령어	관련 관리작업
fsck	파일 시스템 상태 체크
balancer	HDFS 재균형
daemonlog	로그 레벨 동적 변경

dfsadmin

HDFS 상태 확인, HDFS 퇴거, DataNode 참가등