

하둡을 웹 UI에 띄워보기

About

하둡을 깔고 초기에는 VM의 검정색 화면에서 `hdfs dfs -ls /` 등의 명령어를 치며 결과를 확인한다. 하지만 아무래도 터미널 기반이다보니 UI에 비해서 가독력이 떨어지는 것은 사실이다. 이러한 단점을 보완하고자 하둡의 UI가 나왔다고 할 수 있는데, 오늘은 하둡을 터미널이 아니라 UI에서 다뤄보는 것에 대해서 알아본다.

Prerequisite

하둡을 이미 VM에 설치 했다고 가정한다. 여러 대의 클러스터를 구축하기 위해 여러 개의 VM을 뺐다면 더 좋고 한대만 구동해도 상관 없다.

Let's get started!

하둡을 설치하고 난 후, 하둡 디렉토리는 완전히 깨끗한 상태이다. 이는 root를 확인함으로써 알 수 있다. 각자의 hdfs 경로를 찾아서 아래 명령어를 실행해보면 아래와 같다.

```
$ hdfs dfs -ls /
```

`/` 는 root directory을 뜻한다.

아래 명령어를 실행하면 오류가 뜨는데, 이는 `-ls` 뒤에 아무것도 지정해주지 않았기 때문이다. 아무것도 지정해주지 않으면 자동으로 home directory로 인식을 한다. 그런데 컴퓨터를 설치하고 home directory가 자동으로 만들어지는 것과는 달리 하둡은 이를 수동으로 만들어야한다.

```
$ hdfs dfs -ls
```

따라서 아래의 명령어를 통해 home directory을 만들어주자.

```
$ hdfs dfs -mkdir -p /user/hadoop
```

이제 하둡 UI를 켜본다.

먼저 하둡을 설치할 때 환경설정에서 설정한 http를 통해서 들어갈 것이다. `hdfs-site.xml` 에서 아래와 비슷한 부분을 찾아보자.

```
<property>
  <name>dfs.namenode.http-address</name>
  <value>hadoop-master-01:50070</value>
</property>
```

이 설정은 네임노드의 http 주소를 `hadoop-master-01:50070`으로 한다는 것이다. 즉, 인터넷 url에 `http://hadoop-master-01:50070` 을 치면 하둡의 네임노드를 UI로 볼 수 있다. 또는 GCP의 외부 ip를 `hadoop-master-01` 자리에 대신 넣어도 된다.

그러면 아래와 같이 깔끔한 UI를 확인할 수 있다.

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

Overview 'hadoop-master-01:9000' (active)

Started:	Thu Oct 15 10:23:31 +0900 2020
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Wed Sep 11 00:56:00 +0900 2019 by rohithsharmaks from branch-3.2.1
Cluster ID:	CID-eebc842-f20d-40e2-925a-4231a901e9cf
Block Pool ID:	BP-550805060-10.128.0.6-1602724990395

Summary

Security is off.
Safemode is off.

15 files and directories, 25 blocks (25 replicated blocks, 0 erasure coded block groups) = 40 total filesystem object(s).

Heap Memory used 67.91 MB of 283 MB Heap Memory. Max Heap Memory is 1.59 GB.

Non Heap Memory used 60.69 MB of 62.15 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	399.17 GB
Configured Remote Capacity:	0 B
DFS Used:	8.31 GB (2.08%)
Non DFS Used:	16.23 GB
DFS Remaining:	374.64 GB (93.85%)
Block Pool Used:	8.31 GB (2.08%)
DataNodes usages% (Min/Median/Max/stdDev):	1.76% / 1.99% / 2.77% / 0.41%
Live Nodes	4 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Thu Oct 15 10:23:31 +0900 2020

대충 보면, 노드의 개수, 죽은 노드의 개수, 총 용량 등을 한번에 정리해준다.

라이브 노드의 링크를 따라 들어가면 아래와 같이 각 노드별 사용량을 확인할 수 있다.



또한 Utilities -> Browse the file system은 -ls 대신 UI로 파일 위치를 확인할 수 있게 도와준다.

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

Browse Directory

/ Go! 📁 ⬆ 📄

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	drwxr-xr-x	hadoop	supergroup	0 B	Oct 15 10:48	0	0 B	kikang	🗑
<input type="checkbox"/>	drwxr-xr-x	hadoop	supergroup	0 B	Oct 15 10:46	0	0 B	test	🗑
<input type="checkbox"/>	drwxr-xr-x	hadoop	supergroup	0 B	Oct 15 10:47	0	0 B	user	🗑

Showing 1 to 3 of 3 entries

Previous 1 Next

나는 현재 root에 여러 파일을 만들어두었다. UI에서 만들어도 터미널에 즉각 연동이 되고, 그 반대로 해도 마찬가지이다.

UI를 이용해서 더 편리하게 하둡을 사용할 수 있을 것 같다.