

HDFS 클러스터 구동

hdfs namenode -format

cd sbin

start-dfs.cmd

start-yarn.cmd

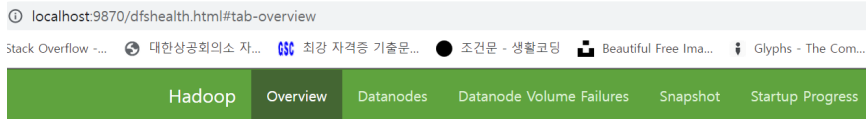
```

Apache Hadoop Distribution - hadoop namenode
2022-10-05 15:02:34,942 INFO blockmanagement.BlockReportLeaseManager: Register
red DN 0f645950-5bed-4441-8f45-cf6008cb2fb (127.0.0.1:9866).
2022-10-05 15:02:35,037 INFO blockmanagement.DatanodeDescriptor: Adding new s
torage ID 0x37a1a4eb7716e546: from storage 05-89abee89-e8e8-45cc-afdc-861080dad91a for DN 127.0.0.1:9866
2022-10-05 15:02:35,084 INFO BlockStateChange: BLOCK* processReport Oxa7c4667
29f8b4e62 with lease ID 0x37a1a4eb7716e546: Processing first storage report f
or 05-89abee89-e8e8-45cc-afdc-861080dad91a from datanode.DatanodeRegistration
(127.0.0.1:9866, datanodeUId=0f645950-5bed-4441-8f45-cf6008cb2fb, infoPort=
9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=57:cid=C1D-12210420-4d7
6-48c8-93d1-b8aca0ead016:nsid=1995317539:c=1664949682352)
2022-10-05 15:02:35,086 INFO BlockStateChange: BLOCK* processReport Oxa7c4667
29f8b4e62 with lease ID 0x37a1a4eb7716e546: from storage 05-89abee89-e8e8-45c
c-afdc-861080dad91a node.DatanodeRegistration(127.0.0.1:9866, datanodeUId=0f
645950-5bed-4441-8f45-cf6008cb2fb, infoPort=9864, infoSecurePort=0, ipcPort=
9867, storageInfo=lv=57:cid=C1D-12210420-4d76-48c8-93d1-b8aca0ead016:nsid=19
95317539:c=1664949682352), blocks: 0, hasStaleStorage: false, processing time
: 2 msecs, invalidatedBlocks: 0

Apache Hadoop Distribution - hadoop datanode
Verification scan starting in 6470054ms with interval of 21600000ms and throttle
Limit of ~1ms/s
2022-10-05 15:02:34,888 INFO datanode.DataNode: Block pool BP-846396654-192.168.
161.1-1664949682352 (Datanode Uuid 0f645950-5bed-4441-8f45-cf6008cb2fb) service
to localhost/127.0.0.1:9000 beginning handshake with NN
2022-10-05 15:02:34,892 INFO datanode.VolumeScanner: VolumeScanner(C:\hadoop-3.3
.2\dfs\data, DS-89abee89-e8e8-45cc-afdc-861080dad91a): no suitable block pools f
ound to scan. Waiting 181439984 ms.
2022-10-05 15:02:34,957 INFO datanode.DataNode: Block pool BP-846396654-192.168.
161.1-1664949682352 (Datanode Uuid 0f645950-5bed-4441-8f45-cf6008cb2fb) service
to localhost/127.0.0.1:9000 successfully registered with NN
2022-10-05 15:02:34,957 INFO datanode.DataNode: For namenode localhost/127.0.0.1
:9000 using BLOCKREPORT_INTERVAL of 21600000msecs CACHEREPORT_INTERVAL of 10000m
secs Initial delay: 0msecs; heartbeatInterval=3000
2022-10-05 15:02:35,113 INFO datanode.DataNode: Successfully sent block report 0
xa7c466729f8b4e62 with lease ID 0x37a1a4eb7716e546 to namenode: localhost/127.0.
0.1:9000, containing 1 storage report(s), of which we sent 1. The reports had 0
total blocks and used 1 RPC(s). This took 4 msecs to generate and 50 msecs for
RPC and NN processing. Got back one command: FinalizeCommand/5.
2022-10-05 15:02:35,114 INFO datanode.DataNode: Got finalize command for block p
ool BP-846396654-192.168.161.1-1664949682352

Apache Hadoop Distribution - yarn resourcemanager
2022-10-05 15:04:46,810 INFO ipc.Server: Starting 05, 2022 3:04:47 오후 com.sun.jersey.server.impl.application.WebApplicationImpl _init
2022-10-05 15:04:46,814 INFO pb.RpcServerFac: Initializing Jersey application, version Jersey: 1.19 02/11/2015 03:25 AM
rPB to the server
2022-10-05 15:04:46,814 INFO ipc.Server: IPC 정보: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver to Guice
2022-10-05 15:04:46,815 INFO ipc.Server: IPC the scope "Singleton"
2022-10-05 15:04:46,823 INFO util.JvmPauseMonitor: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices to GuiceManagedComponent
2022-10-05 15:04:46,838 INFO ipc.CallQueueManager: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponent
eCapacity: 5000, scheduler: class org.apache.gleton
2022-10-05 15:04:46,855 INFO ipc.Server: Starting 05, 2022 3:04:47 오후 com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory
2022-10-05 15:04:46,863 INFO pb.RpcServerFac: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices to GuiceManagedComponent
pcollPB to the server
2022-10-05 15:04:46,864 INFO ipc.Server: IPC 2022-10-05 15:04:47,895 INFO handler.ContextHandler: Started o.e.j.w.WebAppContext@291373d
2022-10-05 15:04:46,864 INFO ipc.Server: IPC in/AppData/Local/Temp/jetty-0.0.0-8042-hadoop-yarn-common-3.3.2-jar--any-551939139471408
2022-10-05 15:04:46,989 INFO ipc.CallQueueManager: Binding org.apache.hadoop.yarn/hadoop-yarn-common-3.3.2-jar/webapps/node
eCapacity: 5000, scheduler: class org.apache
2022-10-05 15:04:46,992 INFO ipc.Server: Starting 05, 2022 3:04:47 오후 com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory
2022-10-05 15:04:46,996 INFO pb.RpcServerFac: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices to GuiceManagedComponent
pcollPB to the server
2022-10-05 15:04:46,996 INFO ipc.Server: IPC 2022-10-05 15:04:47,913 INFO server.Server: Started @16397ms
2022-10-05 15:04:46,997 INFO ipc.Server: IPC 2022-10-05 15:04:47,915 INFO webapp.WebApps: Web app node started at 8042
2022-10-05 15:04:46,997 INFO ipc.Server: IPC 2022-10-05 15:04:47,916 INFO nodemanager.NodeStatusUpdaterImpl: Node ID assigned is: DESK
2022-10-05 15:04:47,584 INFO webproxy.ProxyC2022-10-05 15:04:47,918 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2022-10-05 15:04:47,645 INFO recovery.FMStat2022-10-05 15:04:47,934 INFO client.DefaultNoHAFMFailoverProxyProvider: Connecting to Resou
2022-10-05 15:04:47,645 INFO resourceManager.th id 2054363633 INFO security.NMContainerTokenSecretManager: Rolling master-key fo
2022-10-05 15:04:48,308 INFO resourceManager 2022-10-05 15:04:48,333 INFO security.NMTokenSecretManagerInNM: Rolling master-key for con
httpPort: 8042) registered with capability: -1790017824
2022-10-05 15:04:48,311 INFO rmnode.RMNodeIm2022-10-05 15:04:48,334 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceMan
2022-10-05 15:04:48,340 INFO capacity.Capacit with total resource of <memory:8192, vCores:8>
vCores:8>

```



Overview 'localhost:9000' (✓active)

| | |
|----------------|--|
| Started: | Wed Oct 05 15:02:32 +0900 2022 |
| Version: | 3.3.2, r0bcb014209e219273cb6fd4152df7d713cbac61 |
| Compiled: | Tue Feb 22 03:39:00 +0900 2022 by chao from branch-3.3.2 |
| Cluster ID: | CID-f2210420-4d76-48c8-93d1-b8aca0ead016 |
| Block Pool ID: | BP-846396654-192.168.161.1-1664949682352 |

All Applications

Cluster Metrics

| Apps Submitted | Apps Pending | Apps Running | Apps Completed | Containers Running | Used Resources | Total Resources |
|----------------|--------------|--------------|----------------|--------------------|------------------------|-------------------------|
| 0 | 0 | 0 | 0 | 0 | <memory:0 B, vCores:0> | <memory:8 GB, vCores:8> |

Cluster Nodes Metrics

| Active Nodes | Decommissioning Nodes | Decommissioned Nodes | Lost Nodes |
|--------------|-----------------------|----------------------|------------|
| 1 | 0 | 0 | 0 |

Scheduler Metrics

| Scheduler Type | Scheduling Resource Type | Minimum Allocation | Maximum Allocation |
|--------------------|-------------------------------|-------------------------|-------------------------|
| Capacity Scheduler | [memory-mb (unit=Mi), vcores] | <memory:1024, vCores:1> | <memory:8192, vCores:4> |

Show 20 entries

| ID | User | Name | Application Type | Application Tags | Queue | Application Priority | StartTime | LaunchTime | FinishTime | State | FinalStatus | Running Containers | Allocated CPU VCores | Allocated Memory MB |
|----------------------------|------|------|------------------|------------------|-------|----------------------|-----------|------------|------------|-------|-------------|--------------------|----------------------|---------------------|
| No data available in table | | | | | | | | | | | | | | |

Showing 0 to 0 of 0 entries

CLI (Command Line Interface) 실습

HDFS CLI를 실행하는 방법(명령어)

1. `hadoop fs : local, hdfs` 등과 같은 파일 시스템을 가리킬 수 있는 일반 파일 시스템과 관련있다.
2. `hdfs dfs` : HDFS에만 해당하는 명령어이다.

** FS셸에는 HDFS와 직접 상호작용하여 다양한 명령과 Local FS, WebHDFS, S3 FS 등과 같은 `hadoop`이 지원하는 다른 파일 시스템이 포함되어 있다.

hadoop fs 실습시작

1. `hadoop fs`상에 디렉토리 생성

: `C:\hadoop-3.3.2>hadoop fs -mkdir /user // 유저생성`

HDFS 내에 사용자의 홈 디렉터리는 "/user/사용자이름"으로 생성

: `C:\hadoop-3.3.2>hadoop fs -mkdir /user/fastcampus`

: `C:\hadoop-3.3.2>hadoop fs -mkdir /user/fastcampus/input`

: `C:\hadoop-3.3.2>hadoop fs -ls /user/fastcampus // 정상적으로 생성된 것을 확인`

2. 로컬에 있는 파일이나 디렉토리를 `hdfs`에 업로드 하기

1. `hadoop fs -put LICENSE.txt /user/fastcampus/input`

2. `C:\hadoop-3.3.2>hadoop fs -ls /user/fastcampus/input // 업로드되었는지 확인`

```

C:\Hadoop-3.3.2>hadoop fs -mkdir /user/fastcampus/input
mkdir: `hdfs://localhost:9000/user/fastcampus': No such file or directory

C:\Hadoop-3.3.2>hadoop fs -ls
ls: `.`: No such file or directory

C:\Hadoop-3.3.2>hadoop fs -mkdir /user

C:\Hadoop-3.3.2>hadoop fs -mkdir /user/fastcampus

C:\Hadoop-3.3.2>hadoop fs -mkdir /user/fastcampus/input

C:\Hadoop-3.3.2>hadoop fs -ls /user/fastcampus
Found 1 items
drwxr-xr-x - Minjin supergroup          0 2022-10-03 17:55 /user/fastcampus/input

C:\Hadoop-3.3.2>hadoop fs -put LICENSE.txt /user/fastcampus/input

C:\Hadoop-3.3.2>hadoop fs -ls /user/fastcampus/input
Found 1 items
-rw-r--r-- 1 Minjin supergroup      15217 2022-10-03 18:02 /user/fastcampus/input/LICENSE.txt

```

3. hdfs상에 존재하는 데이터를 로컬파일시스템으로 옮겨오기

- `hadoop fs -copyToLocal /user/fastcampus/input/LICENSE.txt ./` // 현재 디렉토리에 가져오기

```

C:\H>hadoop fs -copyToLocal /user/fastcampus/input/LICENSE.txt ./

C:\H>ls
C 드라이브의 볼륨에는 이름이 없습니다.
볼륨 일련 번호: 767F-BA0B

C:\H 디렉터리

2022-10-02 오후 02:36 <DIR>      apache-maven-3.8.6
2022-10-02 오후 05:53 <DIR>      fastcampus
2022-10-03 오후 06:08 <DIR>      hadoop-3.3.2
2022-09-30 오후 03:15      638,660,563 hadoop-3.3.2.tar.gz
2022-10-02 오후 05:22 <DIR>      hdfs-example
2022-10-03 오후 06:21      15,217 LICENSE.txt
2022-09-30 오후 01:47 <DIR>      PerfLogs
2022-09-30 오후 06:07 <DIR>      Program Files
2022-10-03 오후 05:04 <DIR>      Program Files (x86)
2022-10-03 오후 04:40 <DIR>      Temp
2022-09-30 오후 04:03 <DIR>      tmp
2022-10-01 오후 05:31 <DIR>      Users
2022-10-03 오후 04:40 <DIR>      Windows

      2개 파일      638,675,780 바이트
      11개 디렉터리 234,757,455,872 바이트 남음

```

4. hdfs상에 업로드되어있는 파일을 볼 수 있는 역할

- `hadoop fs -cat /user/fastcampus/input/LICENSE.txt`

```
C:\>hadoop fs -cat /user/fastcampus/input/LICENSE.txt

        Apache License
        Version 2.0, January 2004
        http://www.apache.org/licenses/

TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION

1. Definitions.

"License" shall mean the terms and conditions for use, reproduction,
and distribution as defined by Sections 1 through 9 of this document.

"Licensor" shall mean the copyright owner or entity authorized by
the copyright owner that is granting the License.

"Legal Entity" shall mean the union of the acting entity and all
other entities that control, are controlled by, or are under common
control with that entity. For the purposes of this definition,
"control" means (i) the power, direct or indirect, to cause the
direction or management of such entity, whether by contract or
otherwise, or (ii) ownership of fifty percent (50%) or more of the
outstanding shares, or (iii) beneficial ownership of such entity.

"You" (or "Your") shall mean an individual or Legal Entity
exercising permissions granted by this License.

"Source" form shall mean the preferred form for making modifications,
including but not limited to software source code, documentation
source, and configuration files.

"Object" form shall mean any form resulting from mechanical
transformation or translation of a Source form, including but
not limited to compiled object code, generated documentation,
and conversions to other media types.

"Work" shall mean the work of authorship, whether in Source or
Object form, made available under the License, as indicated by a
copyright notice that is included in or attached to the work
(an example is provided in the Appendix below).

"Derivative Works" shall mean any work, whether in Source or Object
form, that is based on (or derived from) the Work and for which the
editorial revisions, annotations, elaborations, or other modifications
represent, as a whole, an original work of authorship. For the purposes
of this License, Derivative Works shall not include works that remain
separable from, or merely link (or bind by name) to the interfaces of,
the Work and Derivative Works thereof.

"Contribution" shall mean any work of authorship, including
the original version of the Work and any modifications or additions
to that Work or Derivative Works thereof, that is intentionally
submitted to Licensor for inclusion in the Work by the copyright owner
or by an individual or Legal Entity authorized to submit on behalf of
the copyright owner. For the purposes of this definition, "submitted"
means any form of electronic, verbal, or written communication sent
to the Licensor or its representatives, including but not limited to
communication on electronic mailing lists, source code control systems,
and issue tracking systems that are managed by, or on behalf of, the
Licensor for the purpose of discussing and improving the Work, but
excluding communication that is conspicuously marked or otherwise
```

5. hdfs상에 있는 파일을 이동시키기

- `hdfs fs -mv /user/fastcampus/input/LICENSE.txt /user/fastcampus`

6. 파일 내용 뒷부분 보기

- `hadoop fs -tail /C:/LICENSE.txt`

7. 파일 내용 앞부분 보기

- `hadoop fs -head /C:/LICENSE.txt`

```
C:\>hadoop fs -head /C:/LICENSE.txt

        Apache License
        Version 2.0, January 2004
        http://www.apache.org/licenses/

TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION

1. Definitions.

"License" shall mean the terms and conditions for use, reproduction,
and distribution as defined by Sections 1 through 9 of this document.

"Licensor" shall mean the copyright owner or entity authorized by
the copyright owner that is granting the License.

"Legal Entity" shall mean the union of the acting entity and all
other entities that control, are controlled by, or are under common
control with that entity. For the purposes of this definition,
"control" means (i) the power, direct or indirect, to cause the
direction or management of such entity, whether by contract or
otherwise, or (ii) ownership of fifty percent (50%) or more of the
outstanding shares, or (iii) beneficial ownership of such entity.
```

명령어 확인 : `hadoop fs -help`

- 로컬에 있는 파일이나 디렉토리를 hdfs에 업로드 하는 역할
: `hadoop fs -help put` (= put을 어떻게 사용하는지 확인가능) : `hadoop fs -copyFromLocal`
- hdfs상에 존재하는 데이터를 로컬파일시스템으로 옮겨오는 역할
: `hadoop fs -get`
: `hadoop fs -copyToLocal`
- hdfs상에 업로드되어있는 파일을 볼 수 있는 역할
: `hadoop fs -cat`
- hdfs상에 있는 파일을 이동하는 역할
: `hdfs fs -mv`