

# Praktikum Teknologi Perekayasaan Data

*Tanggal Praktikum : Monday, March 13, 2023*

## 1. Command Praktikum

- su - hdfs

```
[root@sandbox-hdp ~]# su - hdfs
Last login: Mon Mar 13 06:58:26 UTC 2023
```

- hdfs dfsadmin -report

```
[hdfs@sandbox-hdp ~]$ hdfs dfsadmin -report
Configured Capacity: 112248641024 (104.54 GB)
Present Capacity: 82478639697 (76.81 GB)
DFS Remaining: 78600169472 (73.20 GB)
DFS Used: 3878470225 (3.61 GB)
DFS Used%: 4.70%
Replicated Blocks:
  Under replicated blocks: 204
  Blocks with corrupt replicas: 0
  Missing blocks: 0
  Missing blocks (with replication factor 1): 0
  Low redundancy blocks with highest priority to recover: 204
  Pending deletion blocks: 0
Erasure Coded Block Groups:
  Low redundancy block groups: 0
  Block groups with corrupt internal blocks: 0
  Missing block groups: 0
  Low redundancy blocks with highest priority to recover: 0
  Pending deletion blocks: 0

-----
Live datanodes (1):

Name: 172.18.0.2:50010 (sandbox-hdp.hortonworks.com)
Hostname: sandbox-hdp.hortonworks.com
Decommission Status : Normal
Configured Capacity: 112248641024 (104.54 GB)
DFS Used: 3878470225 (3.61 GB)
Non DFS Used: 23623320495 (22.00 GB)
DFS Remaining: 78600169472 (73.20 GB)
DFS Used%: 3.46%
DFS Remaining%: 70.02%
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: 4
Last contact: Mon Mar 13 08:19:38 UTC 2023
Last Block Report: Mon Mar 13 06:40:34 UTC 2023
Num of Blocks: 1834

[hdfs@sandbox-hdp ~]$
```

- -safemode

```
[hdfs@sandbox-hdp ~]$ hdfs dfsadmin -safemode
Usage: hdfs dfsadmin [-safemode enter | leave | get | wait | forceExit]
[hdfs@sandbox-hdp ~]$ hdfs dfsadmin -safemode enter
Safe mode is ON
[hdfs@sandbox-hdp ~]$ hdfs dfsadmin -safemode leave
Safe mode is OFF
[hdfs@sandbox-hdp ~]$ █
```

- -finalizeUpgrade

```
[hdfs@sandbox-hdp ~]$ hdfs dfsadmin -finalizeUpgrade
Finalize upgrade successful
[hdfs@sandbox-hdp ~]$ █
```

- -refreshNodes

```
[hdfs@sandbox-hdp ~]$ hdfs dfsadmin -refreshNodes
Refresh nodes successful
[hdfs@sandbox-hdp ~]$
```

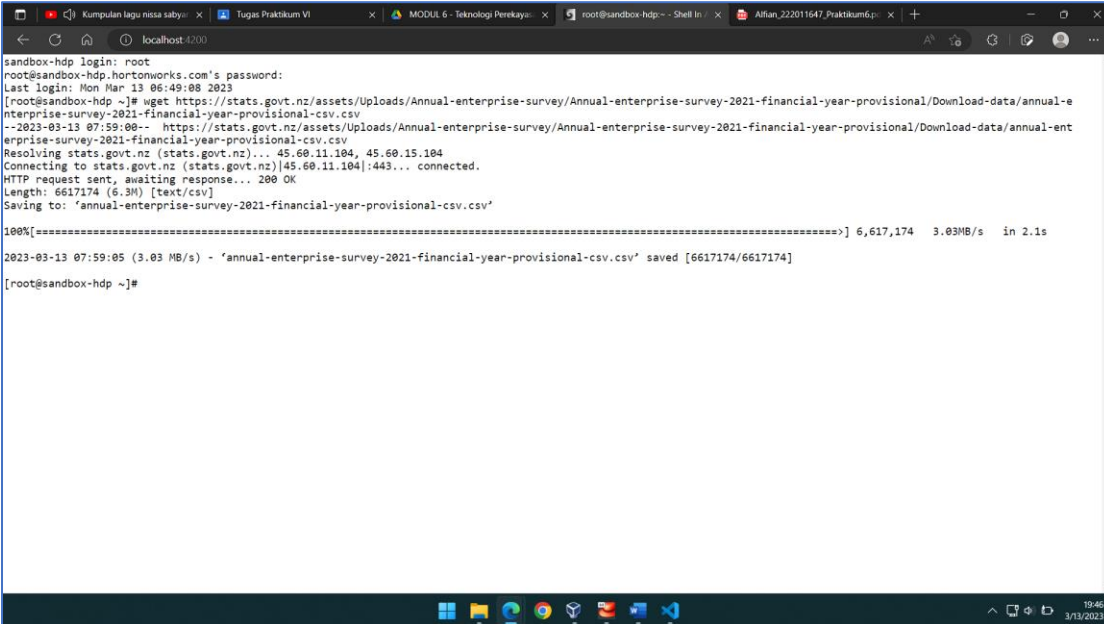
- -printTopology

```
[hdfs@sandbox-hdp ~]$ hdfs dfsadmin -printTopology
Rack: /default-rack
172.18.0.2:50010 (sandbox-hdp.hortonworks.com)
```

- envvars

```
[hdfs@sandbox-hdp ~]$ hdfs envvars
JAVA_HOME='/usr/lib/jvm/java'
HADOOP_HDFS_HOME='/usr/hdp/3.0.1.0-187/hadoop-hdfs'
HDFS_DIR='./'
HDFS_LIB_JARS_DIR='lib'
HADOOP_CONF_DIR='/usr/hdp/3.0.1.0-187/hadoop/conf'
HADOOP_TOOLS_HOME='/usr/hdp/3.0.1.0-187/hadoop'
HADOOP_TOOLS_DIR='share/hadoop/tools'
HADOOP_TOOLS_LIB_JARS_DIR='share/hadoop/tools/lib'
[hdfs@sandbox-hdp ~]$
```

2. Melakukan load data ke hadoop dengan mekanisme yang telah dijelaskan sebelumnya.



```
sandbox-hdp login: root
root@sandbox-hdp.hortonworks.com's password:
Last login: Mon Mar 13 06:49:00 2023
[root@sandbox-hdp ~]# wget https://stats.govt.nz/assets/Uploads/Annual-enterprise-survey/Annual-enterprise-survey-2021-financial-year-provisional/Download-data/annual-enterprise-survey-2021-financial-year-provisional-csv.csv
--2023-03-13 07:59:00-- https://stats.govt.nz/assets/Uploads/Annual-enterprise-survey/Annual-enterprise-survey-2021-financial-year-provisional/Download-data/annual-enterprise-survey-2021-financial-year-provisional-csv.csv
Resolving stats.govt.nz (stats.govt.nz)... 45.60.11.104, 45.60.15.104
Connecting to stats.govt.nz (stats.govt.nz)|45.60.11.104|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 6617174 (6.3M) [text/csv]
Saving to: 'annual-enterprise-survey-2021-financial-year-provisional-csv.csv'

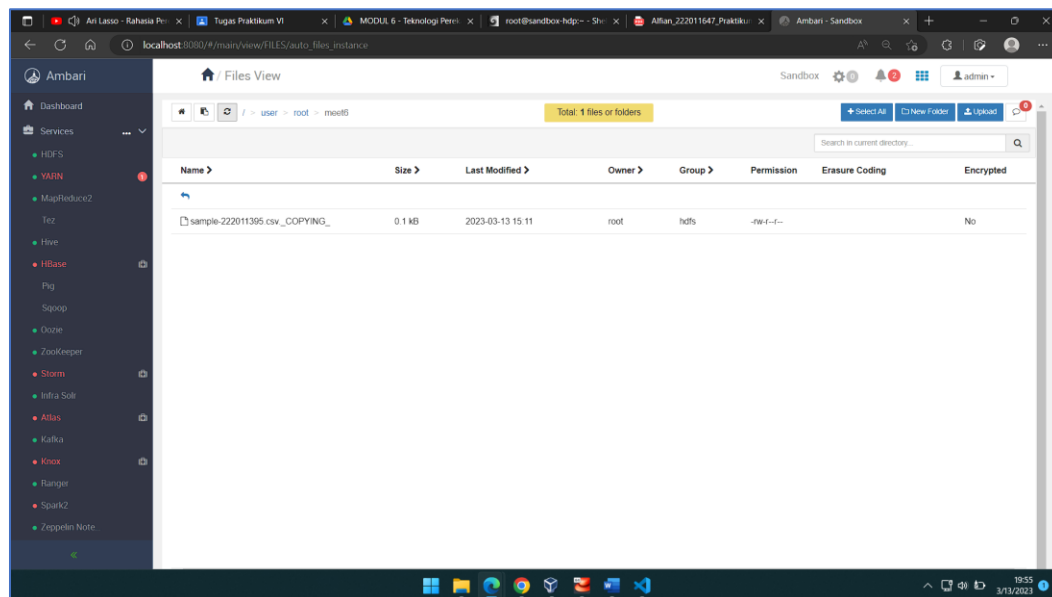
100%[=====] 6,617,174 3.03MB/s in 2.1s

2023-03-13 07:59:05 (3.03 MB/s) - 'annual-enterprise-survey-2021-financial-year-provisional-csv.csv' saved [6617174/6617174]

[root@sandbox-hdp ~]#
```

3. Melakukan pengecekan file di HDFS

- Proses copy file



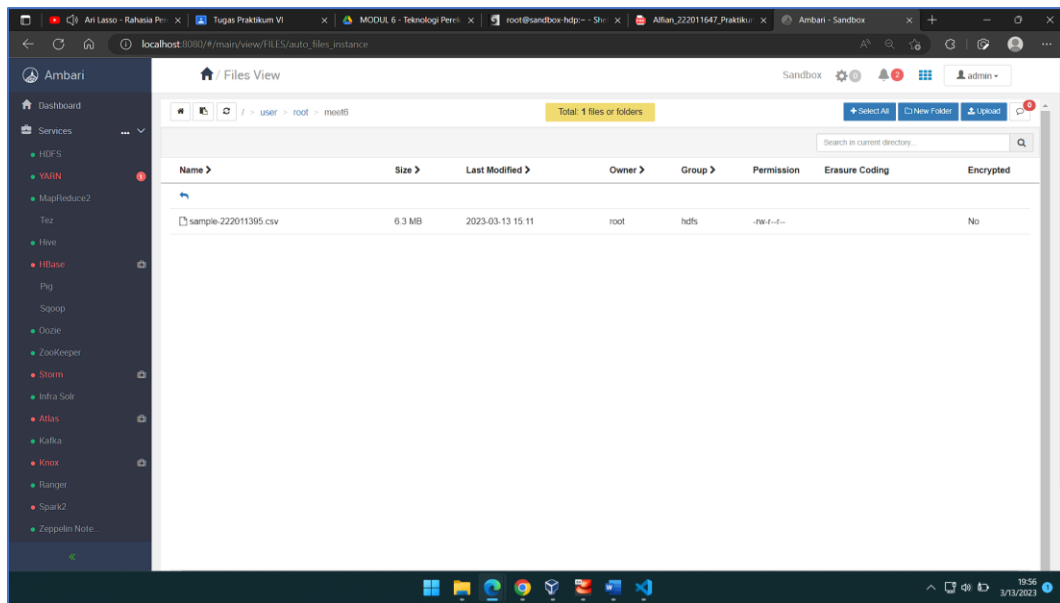
```
sandbox-hdp login: root
root@sandbox-hdp.hortonworks.com's password:
Last login: Mon Mar 13 06:49:08 2023
[root@sandbox-hdp ~]# wget https://stats.govt.nz/assets/Uploads/Annual-enterprise-survey/Annual-enterprise-survey-2021-financial-year-provisional/Download-data/annual-enterprise-survey-2021-financial-year-provisional-csv.csv
--2023-03-13 07:59:00-- https://stats.govt.nz/assets/Uploads/Annual-enterprise-survey/Annual-enterprise-survey-2021-financial-year-provisional/Download-data/annual-enterprise-survey-2021-financial-year-provisional-csv.csv
Resolving stats.govt.nz (stats.govt.nz)... 45.60.11.104, 45.60.15.104
Connecting to stats.govt.nz (stats.govt.nz)|45.60.11.104|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 6617174 (6.3M) [text/csv]
Saving to: 'annual-enterprise-survey-2021-financial-year-provisional-csv.csv'

100%[=====] 6,617,174  3.03MB/s  in 2.1s

2023-03-13 07:59:05 (3.03 MB/s) - 'annual-enterprise-survey-2021-financial-year-provisional-csv.csv' saved [6617174/6617174]

[root@sandbox-hdp ~]# ls
anaconda-ks.cfg  annual-enterprise-survey-2021-financial-year-provisional-csv.csv  geo  geolocation.csv
[root@sandbox-hdp ~]# ls -l
total 6988
-rw-r--r-- 1 root root 3302 May 31 2018 anaconda-ks.cfg
-rw-r--r-- 1 root root 6617174 Jun 29 2022 annual-enterprise-survey-2021-financial-year-provisional-csv.csv
drwx-xr-x 2 root root 4096 Feb 13 05:53 geo
-rw-r--r-- 1 root root 526677 Feb 13 05:53 geolocation.csv
[root@sandbox-hdp ~]# hadoop fs -ls
Found 3 items
drwx----- 0 root hdfs 0 2023-02-13 06:00 .Trash
-rw-r--r-- 1 root hdfs 0 2023-02-13 06:12 cobalagi
-rw-r--r-- 1 root hdfs 526677 2023-02-13 05:55 geolocation.csv
[root@sandbox-hdp ~]# hadoop fs -mkdir meet6
Found 4 items
drwx----- 0 root hdfs 0 2023-02-13 06:00 .Trash
-rw-r--r-- 1 root hdfs 0 2023-02-13 06:12 cobalagi
-rw-r--r-- 1 root hdfs 526677 2023-02-13 05:55 geolocation.csv
drwx-xr-x 0 2023-03-13 08:05 meet6
[root@sandbox-hdp ~]# mv annual-enterprise-survey-2021-financial-year-provisional-csv.csv sample-222011395.csv
[root@sandbox-hdp ~]# ls
anaconda-ks.cfg  geo
[root@sandbox-hdp ~]# hadoop fs -copyFromLocal sample-222011395.csv praktikum6/sample-222011395.csv
copyFromLocal: praktikum6/sample-222011395.csv: No such file or directory: hdfs://sandbox-hdp.hortonworks.com:8020/user/root/praktikum6/sample-222011395.csv
[root@sandbox-hdp ~]# hadoop fs -copyFromLocal sample-222011395.csv meet6/sample-222011395.csv
[root@sandbox-hdp ~]#
```

- Hasil Copy



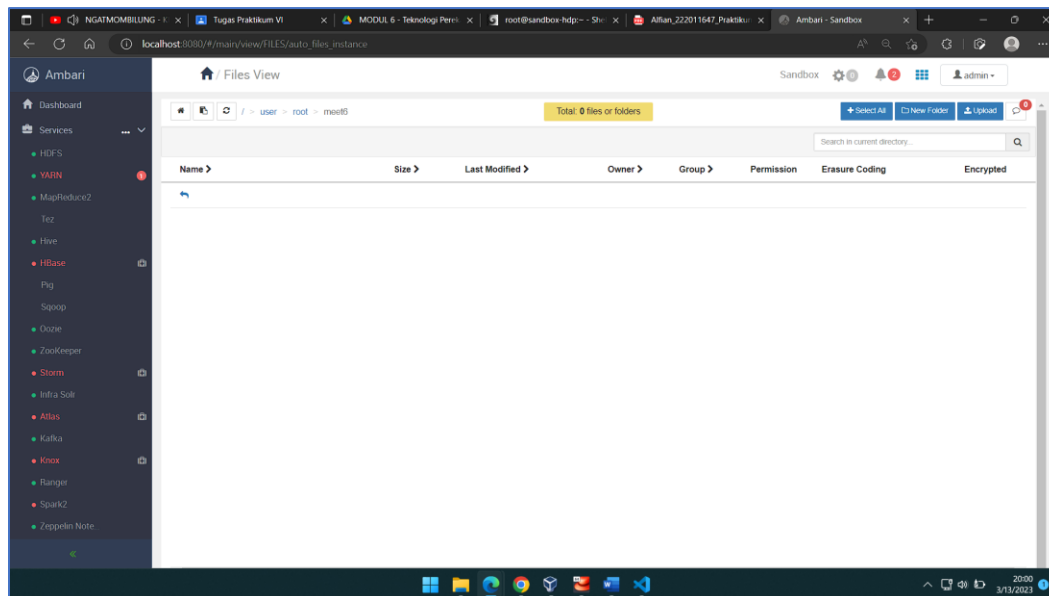
```
Connecting to stats.govt.nz (stats.govt.nz)[45.60.11.104]:443... connected.
HTTP request sent, awaiting response... 200 OK
length: 6617174 (6.3M) [text/csv]
Saving to: 'annual-enterprise-survey-2021-financial-year-provisional-csv.csv'

100%[=====] 6,617,174  3.0MB/s  in 2.1s

2023-03-13 07:59:05 (3.03 MB/s) - 'annual-enterprise-survey-2021-financial-year-provisional-csv.csv' saved [6617174/6617174]

[root@sandbox-hdp ~]# ls
anaconda-ks.cfg  annual-enterprise-survey-2021-financial-year-provisional-csv.csv  geo  geolocation.csv
[root@sandbox-hdp ~]# ls -l
total 6988
-rw-r--r-- 1 root root 3302 May 31 2018 anaconda-ks.cfg
-rw-r--r-- 1 root root 6617174 Jun 29 2022 annual-enterprise-survey-2021-financial-year-provisional-csv.csv
-rwxr-xr-x 2 root root 4096 Feb 13 05:53 geo
-rw-r--r-- 1 root root 526677 Feb 13 05:53 geolocation.csv
[root@sandbox-hdp ~]# hadoop fs -ls
Found 3 items
drwxr-xr-x 1 root hdfs 0 2023-02-13 06:00 .Trash
-rw-r--r-- 1 root hdfs 0 2023-02-13 06:12 cobalagi
-rw-r--r-- 1 root hdfs 526677 2023-02-13 05:55 geolocation.csv
[root@sandbox-hdp ~]# hadoop fs -mkdir meet6
[root@sandbox-hdp ~]# hadoop fs -ls
Found 4 items
drwxr-xr-x 1 root hdfs 0 2023-02-13 06:00 .Trash
-rw-r--r-- 1 root hdfs 0 2023-02-13 06:12 cobalagi
-rw-r--r-- 1 root hdfs 526677 2023-02-13 05:55 geolocation.csv
-rwxr-xr-x 1 root hdfs 0 2023-03-13 08:05 meet6
[root@sandbox-hdp ~]# mv annual-enterprise-survey-2021-financial-year-provisional-csv.csv sample-222011395.csv
[root@sandbox-hdp ~]# ls
anaconda-ks.cfg  geo
[root@sandbox-hdp ~]# hadoop fs -copyFromLocal sample-222011395.csv praktikum6/sample-222011395.csv
copyFromLocal: 'praktikum6/sample-222011395.csv': No such file or directory: 'hdfs://sandbox-hdp.hortonworks.com:8020/user/root/praktikum6/sample-222011395.csv'
[root@sandbox-hdp ~]# hadoop fs -copyFromLocal sample-222011395.csv meet6/sample-222011395.csv
[root@sandbox-hdp ~]# ls -l
total 6988
-rw-r--r-- 1 root root 3302 May 31 2018 anaconda-ks.cfg
-rwxr-xr-x 2 root root 4096 Feb 13 05:53 geo
-rw-r--r-- 1 root root 526677 Feb 13 05:53 geolocation.csv
-rw-r--r-- 1 root root 6617174 Jun 29 2022 sample-222011395.csv
[root@sandbox-hdp ~]# hadoop fs -ls meet6
Found 1 items
-rw-r--r-- 1 root hdfs 6617174 2023-03-13 08:11 meet6/sample-222011395.csv
[root@sandbox-hdp ~]#
```

- Remove File



```
100%[=====] 6,617,174 3.03MB/s in 2.1s
2023-03-13 07:59:05 (3.03 MB/s) - 'annual-enterprise-survey-2021-financial-year-provisional-csv.csv' saved [6617174/6617174]

[root@sandbox-hdp ~]# ls
anaconda-ks.cfg annual-enterprise-survey-2021-financial-year-provisional-csv.csv geo geolocation.csv
[root@sandbox-hdp ~]# ls -l
total 6988
-rw-r--r-- 1 root root 3302 May 31 2018 anaconda-ks.cfg
-rw-r--r-- 1 root root 6617174 Jun 29 2022 annual-enterprise-survey-2021-financial-year-provisional-csv.csv
drwxr-xr-x 2 root root 4096 Feb 13 05:53 geo
-rw-r--r-- 1 root root 526677 Feb 13 05:53 geolocation.csv
[root@sandbox-hdp ~]# hadoop fs -ls
Found 3 items
drwx----- 1 root hdfs 0 2023-02-13 06:00 .Trash
-rw-r--r-- 1 root hdfs 0 2023-02-13 06:12 cobalagl
-rw-r--r-- 1 root hdfs 526677 2023-02-13 05:55 geolocation.csv
[root@sandbox-hdp ~]# hadoop fs -ls /meet6
Found 4 items
drwx----- 1 root hdfs 0 2023-02-13 06:00 .Trash
-rw-r--r-- 1 root hdfs 0 2023-02-13 06:12 cobalagl
-rw-r--r-- 1 root hdfs 526677 2023-02-13 05:55 geolocation.csv
drwxr-xr-x 1 root hdfs 0 2023-03-13 08:05 meet6
[root@sandbox-hdp ~]# mv annual-enterprise-survey-2021-financial-year-provisional-csv.csv sample-222011395.csv
[root@sandbox-hdp ~]# ls
anaconda-ks.cfg geo
[root@sandbox-hdp ~]# hadoop fs -copyfromlocal sample-222011395.csv praktikum/sample-222011395.csv
copyfromlocal: 'praktikum/sample-222011395.csv': No such file or directory: 'hdfs://sandbox-hdp.hortonworks.com:8020/user/root/praktikum/sample-222011395.csv'
[root@sandbox-hdp ~]# hadoop fs -copyfromlocal sample-222011395.csv meet6/sample-222011395.csv
[root@sandbox-hdp ~]# ls -l
total 6988
-rw-r--r-- 1 root root 3302 May 31 2018 anaconda-ks.cfg
drwxr-xr-x 2 root root 4096 Feb 13 05:53 geo
-rw-r--r-- 1 root root 526677 Feb 13 05:53 geolocation.csv
-rw-r--r-- 1 root root 6617174 Jun 29 2022 sample-222011395.csv
[root@sandbox-hdp ~]# hadoop fs -ls meet6
Found 1 items
-rw-r--r-- 1 root hdfs 6617174 2023-03-13 08:11 meet6/sample-222011395.csv
[root@sandbox-hdp ~]# hadoop fs -rm meet6/sample-222011395.csv
23/03/13 08:15:52 INFO fs.TrashPolicyDefault: Moved: 'hdfs://sandbox-hdp.hortonworks.com:8020/user/root/meet6/sample-222011395.csv' to trash at: hdfs://sandbox-hdp.hortonworks.com:8020/us
er/root/.Trash/Current/user/root/meet6/sample-222011395.csv
[root@sandbox-hdp ~]# hadoop fs -ls meet6
[root@sandbox-hdp ~]#
```