

نصب Hadoop و استفاده از آن

برای نصب Hadoop و انجام مثالهایی از MapReduce، نیاز داریم که حداقل دو VM داشته باشیم که یکی از آنها نقش نود master و بقیه، نقش نودهای slave را داشته باشند. در این گزارش، یک نود master و یک نود slave خواهیم داشت. برای ایجاد VM، از نرم افزار VirtualBox استفاده شده است و برای هر VM، سیستم عامل Ubuntu 18.04 نصب خواهد شد. لذا لازم است که فایل iso. مربوط به این نسخه از ubuntu را دانلود کرده باشیم که حجم آن حدود 2GB است.

نصب Virtual Box و ایجاد VM

با توجه به اینکه سیستم عامل host (لپ‌تاپ)، windows 10 است، نسخه VirtualBox 6.1.4 متناسب با آن با حجم 110 MB دانلود و نصب گردید. پس از نصب ابتدا مسیر پیش فرض محل ذخیره سازی فایل‌های VM از مسیر زیر، تغییر داده شد :

VirtualBox → File → preference → default machine folder

مسیر F:\VirtualBox-VMs برای ایجاد VMها و فایل‌های مربوطه تعریف شد.

سپس با انتخاب آیکن new، یک VM از نوع Linux با نسخه ubuntu(64bit) ایجاد شد و برای آن 3GB Memory و یک دیسک 20GB در نظر گرفته شد. سپس از مسیر زیر فایل iso. سیستم عامل ubuntu را به آن معرفی می‌کنیم :

Select vm → Setting → Storage → select cd icon → choose a disk file → ubuntu.iso

همچنین از مسیر زیر تعیین می‌کنیم که هنگام بالا آمدن VM، از CD بوت شود و سپس VM را start می‌کنیم تا نصب ubuntu آغاز شود :

Select vm → Setting → System → select CD as first boot order

Select vm → start

یک VM دیگر هم مشابه بالا به عنوان نود slave ایجاد می‌کنیم.

تنظیمات شبکه در Virtual Box

پس از نصب ubuntu، بصورت پیش فرض کارت شبکه بصورت NAT نصب می شود و این امکان را می دهد که از داخل VM مثلاً به اینترنت وصل شد و یا IP مربوط به host را ping کرد. VM بصورت پیش فرض آدرس 10.0.2.15 را می گیرد. در این حالت، از host نمی توان VM را ping کرد.

نیازی که برای راه اندازی Hadoop وجود دارد، اینست که دو تا VM بسازیم و این VM ها همدیگر را ping کنند و host هم آنها را ping کند. برای پیاده سازی این مساله، VM را stop کرده و ابتدا برای خود virtualBox یک کارت شبکه از نوع HostOnly می سازیم که البته پیش فرض یکی دارد و می توان از آن هم استفاده کرد.

VirtualBox : File → Host Network Manger

سپس برای هر VM به بخش setting رفته و قسمت network را انتخاب کرده و غیر از کارت شبکه موجود که NAT است، کارت شبکه دوم را هم از نوع HostOnly اضافه می کنیم و سپس VM را start می کنیم.

با دستور ip addr show در ترمینال لینوکس، میتوان ip این کارت جدید را دید. حال هم VM می تواند host را ping کند و هم host می تواند VM را ping کند. بهتر است ip استاتیک تنظیم کرد و فایل /etc/hosts را اصلاح کرد.

نصب Hadoop

مراحل آماده سازی نودهای master و slave بر اساس داکيومنت تهیه شده توسط آقای مولایی (TA) انجام شد که بصورت کلی شامل مراحل زیر می باشد :

❖ ویرایش فایل /etc/hosts

محتوای فایل برای نود master :

```
negar@negar-VirtualBox:~$ cat /etc/hosts
#negar-VirtualBox
192.168.56.101  hadoop-master
192.168.56.102  hadoop-slave-1
127.0.0.1      localhost
```

محتوای فایل برای نود slave :

```
negar@negar-VirtualBox:~$ cat /etc/hosts
#negar-VirtualBox
192.168.56.102  hadoop-slave-1
192.168.56.101  hadoop-master
127.0.0.1      localhost
```

نکته : یکی از مشکلاتی که بعد از نصب و اجرای hadoop برای اینجانب پیش آمد، این بود که روی پورت 9000 از 127.0.0.1 در واقع listen می‌کرد نه روی 192.168.56.101 که با جابجا کردن سطرهای مربوطه در این فایل، مشکل حل شد، لذا بهتر است که برای هر نود، نام تعیین شده برای آن بالاتر از سطر localhost قرار گیرد.

❖ نصب openjdk روی هر دو نود

```
$ sudo apt-get update
```

```
$ sudo apt-get install openjdk-8-jdk
```

```
negar@negar-VirtualBox:~$ java -version
openjdk version "1.8.0_252"
OpenJDK Runtime Environment (build 1.8.0_252-8u252-b09-1~18.04-b09)
OpenJDK 64-Bit Server VM (build 25.252-b09, mixed mode)
```

❖ نصب openssh روی هر دو نود

دستور سوم؛ نسبت به سند نصب آقای مولایی، اضافه شده است.

```
$ sudo apt-get install openssh-server
```

```
$ ssh-keygen -t rsa
```

```
$ cat /home/hadoop/.ssh/id_rsa.pub >> .ssh/authorized_keys
```

دستور زیر روی نود master اجرا می‌شود:

```
$ ssh-copy-id negar@hadoop-slave-1
```

دستور زیر روی نود slave اجرا می‌شود:

```
$ ssh-copy-id negar@hadoop-master
```

پس از نصب موفقیت‌آمیز ssh، باید بتوان از روی نود master با صدور دستور زیر بدون نیاز به وارد کردن Password به نود slave متصل شد :

```
negar@negar-VirtualBox:~$ ssh hadoop-slave-1
Welcome to Ubuntu 18.04.4 LTS (GNU/Linux 5.3.0-51-generic x86_64)
```

```
negar@negar-VirtualBox:~$ exit
logout
Connection to hadoop-slave-1 closed.
```

و بالعکس از روی نود slave نیز باید بتوان به همین صورت به نود master متصل شد :

```
negar@negar-VirtualBox:~$ ssh hadoop-master
Welcome to Ubuntu 18.04.4 LTS (GNU/Linux 5.3.0-51-generic x86_64)
```

```
negar@negar-VirtualBox:~$ exit
logout
Connection to hadoop-master closed.
```

نکته : به دلیل اینکه بعداً با استفاده از نرم افزار Bitvise ssh client برای انتقال فایلها از host به این نودها استفاده می شود و این نرم افزار از پروتکل ssh استفاده کرده و معمولاً از متد password برای اتصال به نودها استفاده می کند، بهتر است که فایل `/etc/ssh/sshd_config` را ویرایش کرده و `PasswordAuthentication yes` را در آن داشته باشیم. پس از آن لازم است که سرویس ssh را restart کنیم.

```
negar@negar-VirtualBox:~$ cat /etc/ssh/sshd_config | grep PasswordAuthentication
PasswordAuthentication yes
# PasswordAuthentication. Depending on your PAM configuration,
# PAM authentication, then enable this but set PasswordAuthentication
negar@negar-VirtualBox:~$
negar@negar-VirtualBox:~$ service ssh restart
==== AUTHENTICATING FOR org.freedesktop.systemd1.manage-units ====
Authentication is required to restart 'ssh.service'.
Authenticating as: negar,,, (negar)
Password:
==== AUTHENTICATION COMPLETE ====
negar@negar-VirtualBox:~$
```

❖ دانلود hadoop و انتقال آن به نودهای master و slave

نسخه hadoop 3.2.1 از سایت <https://hadoop.apache.org/releases.html> دانلود شده و از طریق نرم افزار Bitvise ssh client به نودهای master و slave منتقل شده و با دستور زیر extract شد :

```
$ sudo tar -xzf hadoop-3.2.1.tar.gz
```

سپس برای دادن مجوز `read,write,exec` به `all` دستور زیر اجرا شد، در این دستور -R اضافه شده، تا دسترسی برای تمام زیرشاخه ها هم داده شود.

```
$ sudo chmod -R a+rwX ~/hadoop-3.2.1/
```

❖ تنظیم PATH و JAVA_HOME روی هر دو نود

```
negar@negar-VirtualBox:~$ cat /etc/environment
PATH="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/home/negar/hadoop-3.2.1/bin:/home/negar/hadoop-3.2.1/sbin"
JAVA_HOME="/usr/lib/jvm/java-8-openjdk-amd64/jre"
```

❖ کانفیگ Hadoop Master و کپی تنظیمات به نود slave

فایل core-site.xml را از مسیر نصب hadoop بصورت زیر ویرایش کرده و بخشهای <configuration> را به آن اضافه می‌کنیم که در واقع پورت 9000 را روی نود master برای listen کردن به نودهای slave تعریف می‌کند:

```
negar@negar-VirtualBox:~$ cat hadoop-3.2.1/etc/hadoop/core-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://192.168.56.101:9000</value>
  </property>
</configuration>
```

به همین ترتیب، فایل hadoop-3.2.1/etc/hadoop/hdfs-site.xml را ویرایش کرده و بخش <configuration> را برای معرفی مسیر nameNode و dataNode بصورت زیر به آن اضافه می‌کنیم:

```
<configuration>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>/home/negar/hadoop-3.2.1/data/nameNode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>/home/negar/hadoop-3.2.1/data/dataNode</value>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
</configuration>
```

همانطور که از تصویر پیداست لازم است که قبل از آن مسیرهای `data/dataNode` و `data/nameNode` در مسیر نصب `hadoop` ایجاد شود و سپس در فایل `hdfs-site.xml` به آن ارجاع شود. همچنین با توجه به اینکه یک نود `slave` داریم، مقدار مشخصه `dfs.replication` را برابر با 1 مقدار می‌دهیم.

سپس این تنظیمات را با استفاده از دستور زیر به نود `slave` هم کپی می‌کنیم:

```
$ scp hadoop-3.2.1/etc/hadoop/* negar@hadoop-slave-1:/home/negar/hadoop-3.2.1/etc/hadoop/
```

چنانچه روی نداشتن دسترسی در مقصد، مشکلی وجود داشت، می‌توان ابتدا با دستور زیر `owner` این فلدر را به کاربری که برای `ssh` کانفیگ شده، تغییر داد و سپس از `scp` استفاده کرد.

```
$ sudo chown -R negar:negar Hadoop-3.2.1/
```

در مرحله بعد `hdfs` را بصورت زیر فرمت می‌کنیم :

```
$ source /etc/environment
```

```
$ hdfs namenode -format
```

حال میتوان `hdfs` را `start` کرد. خروجی اجرای `start-dfs.sh` و سپس `jps` در نود `master` بصورت زیر است :

```
negar@negar-VirtualBox:~$ start-dfs.sh
Starting namenodes on [hadoop-master]
Starting datanodes
Starting secondary namenodes [negar-VirtualBox]
negar@negar-VirtualBox:~$ jps
25889 Jps
25773 SecondaryNameNode
25486 NameNode
negar@negar-VirtualBox:~$
```

خروجی اجرای `jps` در نود `slave` بصورت زیر می‌باشد :

```
negar@negar-VirtualBox:~$ jps
20915 Jps
20840 DataNode
negar@negar-VirtualBox:~$
```

حال می‌توان با وارد کردن آدرس <http://hadoop-master:9870/> به کنسول وبی `hadoop` متصل شد. شکل زیر صفحه اول آن را نشان می‌دهد. (از `Host` با `ip` مربوط به `master` متصل شده ایم)

با توجه به اینکه یک نود `slave` داریم در بخش، `Live Nodes` عدد 1 نشان داده شده است.

Overview 'hadoop-master:9000' (active)

Started:	Fri May 29 09:50:00 +0430 2020
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Tue Sep 10 20:26:00 +0430 2019 by rohithsharmaks from branch-3.2.1
Cluster ID:	CID-b12f1e6-28f8-44ed-b6cb-55ae7448b94c
Block Pool ID:	BP-743812391-10.0.2.15-1588434409659

Summary

Security is off.
 Safemode is off.
 24 files and directories, 13 blocks (13 replicated blocks, 0 erasure coded block groups) = 37 total filesystem object(s).
 Heap Memory used 34.31 MB of 60.81 MB Heap Memory. Max Heap Memory is 728.88 MB.
 Non-Heap Memory used 47.51 MB of 48.94 MB Committed Non-Heap Memory. Max Non-Heap Memory is <unbounded>.

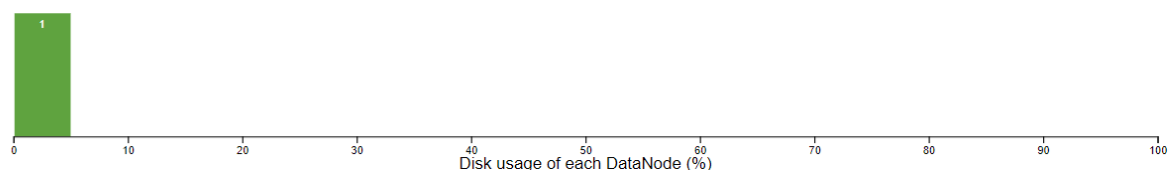
Configured Capacity:	19.56 GB
Configured Remote Capacity:	0 B
DFS Used:	493.84 MB (2.47%)
Non DFS Used:	7.17 GB
DFS Remaining:	10.89 GB (55.66%)
Block Pool Used:	493.84 MB (2.47%)
DataNodes usages% (Min/Median/Max/stdDev):	2.47% / 2.47% / 2.47% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0

با کلیک بر روی Live Nodes، اطلاعات و ip مربوط به Datanode نمایش داده می شود :

Datanode Information

✓ In service
❗ Down
🔄 Decommissioning
🚫 Decommissioned
🔴 Decommissioned & dead
🛠 Entering Maintenance
🔧 In Maintenance
🔴 In Maintenance & dead

Datanode usage histogram



In operation

Show entries

Search:

Node	Http Address	Last contact	Last Block Report	Capacity	Blocks	Block pool used	Version
✓ negar-VirtualBox:9866 (192.168.56.102:9866)	http://negar-VirtualBox:9866	0s	10m	19.56 GB <div><div></div></div>	13	493.84 MB (2.47%)	3.2.1

❖ تنظیمات YARN

فایل `.bashrc` را هم در نود `master` و هم در `slave` بصورت زیر ویرایش می کنیم :

```
export HADOOP_HOME="/home/negar/hadoop-3.2.1"
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_HDFS_HOME=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_YARN_HOME=$HADOOP_HOME
```

همچنین فایل `hadoop-3.2.1/etc/hadoop/yarn-site.xml` را ویرایش کرده و بخش زیر را به آن اضافه

می کنیم و در واقع نود `master` را به عنوان `resource manager` تعیین می کنیم :

```
<configuration>

<!-- Site specific YARN configuration properties -->
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>hadoop-master</value>
</property>
</configuration>
```

حال با اجرای دستورات زیر در نود `master` ابتدا `yarn` را `start` کرده و سپس لیست نودهای آن که در اینجا

یکی است (یک نود `slave` داریم) را مشاهده می کنیم :

```
negar@negar-VirtualBox:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
negar@negar-VirtualBox:~$ yarn node -list
2020-05-29 10:25:12,334 INFO client.RMProxy: Connecting to ResourceManager at hadoop-master/192.168.56.101:8032
Total Nodes:1
      Node-Id          Node-State Node-Http-Address      Number-of-Running-Containers
negar-VirtualBox:39685  RUNNING   negar-VirtualBox:8042
0
negar@negar-VirtualBox:~$
```

با دستور زیر نیز میتوان تنظیمات کلی `hadoop` را مشاهده کرد که تصویر خروجی آن در ادامه آمده است :

```
$ hdfs dfsadmin -report
```



```

negar@negar-VirtualBox:~$ hdfs dfsadmin -report
Configured Capacity: 21001486336 (19.56 GB)
Present Capacity: 12207972352 (11.37 GB)
DFS Remaining: 11690147840 (10.89 GB)
DFS Used: 517824512 (493.84 MB)
DFS Used%: 4.24%
Replicated Blocks:
    Under replicated blocks: 0
    Blocks with corrupt replicas: 0
    Missing blocks: 0
    Missing blocks (with replication factor 1): 0
    Low redundancy blocks with highest priority to recover: 0
    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: 192.168.56.102:9866 (hadoop-slave-1)
Hostname: negar-VirtualBox
Decommission Status : Normal
Configured Capacity: 21001486336 (19.56 GB)
DFS Used: 517824512 (493.84 MB)
Non DFS Used: 7703101440 (7.17 GB)
DFS Remaining: 11690147840 (10.89 GB)
DFS Used%: 2.47%
DFS Remaining%: 55.66%
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 May 29 10:36:36 IRDT 2020
Last Block Report: Fri May 29 09:50:09 IRDT 2020
Num of Blocks: 13

negar@negar-VirtualBox:~$ █

```

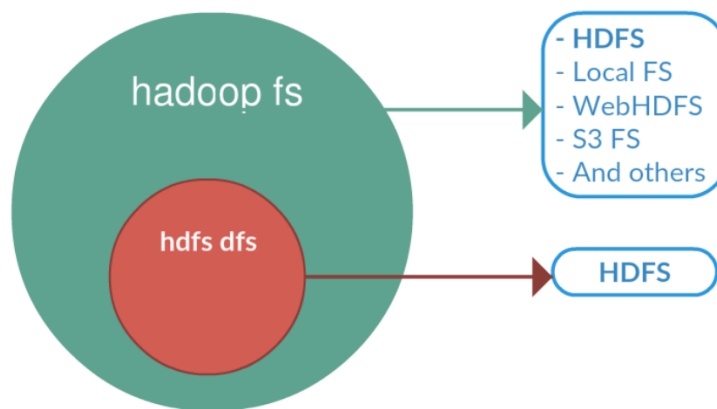
کپی و انتقال فایل به Hadoop

برای دیدن محتویات مسیر ریشه hdfs می توان از یکی از دستورات زیر استفاده کرد :

```
$ hadoop fs -ls /
```

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

نکته : دستور hadoop fs نسبت به hdfs dfs عمومی تر است و اجازه کار با انواع فایل سیستم ها را می دهد اما hdfs dfs تنها برای فایل سیستم HDFS قابل استفاده است. در اینجا که ما با HDFS کار می کنیم، این دو دستور مترادف هستند و در ادامه گزارش از hdfs dfs استفاده شده است. شکل زیر، این مساله را نشان می دهد.



فایل زیر را دانلود کرده و از طریق نرم افزار Bitvise به نود master در مسیر hw-data منتقل می کنیم:

<http://ocw.mit.edu/ans7870/6/6.006/s08/lecturenotes/files/t8.shakespeare.txt>

در ادامه روی hdfs یک دایرکتوری به نام homeworks ساخته و فایل بالا را از مسیر hw-data به آن منتقل می کنیم :

```
negar@negar-VirtualBox:~$ ll hw-data/
total 5340
drwxrwxr-x  2 negar negar   4096 May 29 12:18 ./
drwxr-xr-x 20 negar negar   4096 May 29 12:17 ../
-rw-rw-r--  1 negar negar 5458199 May 29 12:14 t8.shakespeare.txt
negar@negar-VirtualBox:~$ hdfs dfs -mkdir /homeworks
negar@negar-VirtualBox:~$ hdfs dfs -put hw-data/t8.shakespeare.txt /homeworks
2020-05-29 12:20:50,741 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
negar@negar-VirtualBox:~$ hdfs dfs -ls /
Found 2 items
drwxr-xr-x  - negar supergroup          0 2020-05-29 12:20 /homeworks
drwxr-xr-x  - negar supergroup          0 2020-05-14 10:36 /user
negar@negar-VirtualBox:~$ hdfs dfs -ls /homeworks
Found 1 items
-rw-r--r--  1 negar supergroup   5458199 2020-05-29 12:20 /homeworks/t8.shakespeare.txt
negar@negar-VirtualBox:~$
```

می‌توان بجای put از copyFromLocal استفاده کرد. هر دو برای کپی فایل از فایل سیستم local به hdfs استفاده می‌شوند. در شکل زیر کپی دیگری از همین فایل با copyFromLocal در hdfs ایجاد شده است.

```
negar@negar-VirtualBox:~$ hdfs dfs -copyFromLocal hw-data/t8.shakespeare.txt /homeworks/t8-copy.txt
2020-05-29 12:30:59,898 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTru
sted = false, remoteHostTrusted = false
negar@negar-VirtualBox:~$ hdfs dfs -ls /homeworks
Found 2 items
-rw-r--r-- 1 negar supergroup 5458199 2020-05-29 12:31 /homeworks/t8-copy.txt
-rw-r--r-- 1 negar supergroup 5458199 2020-05-29 12:20 /homeworks/t8.shakespeare.txt
negar@negar-VirtualBox:~$
```

بطور برعکس اگر بخواهیم یک فایل را از hdfs به local منتقل کنیم از get یا copyToLocal استفاده می‌کنیم. در شکل زیر به هر دو روش، فایل کپی شده است:

```
negar@negar-VirtualBox:~$ hdfs dfs -get /homeworks/t8-copy.txt ~/hw-data/t8-local-1.txt
2020-05-29 12:38:58,717 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTru
sted = false, remoteHostTrusted = false
negar@negar-VirtualBox:~$ ll hw-data/
total 10672
drwxrwxr-x 2 negar negar 4096 May 29 12:38 ./
drwxr-xr-x 20 negar negar 4096 May 29 12:17 ../
-rw-r--r-- 1 negar negar 5458199 May 29 12:38 t8-local-1.txt
-rw-rw-r-- 1 negar negar 5458199 May 29 12:14 t8.shakespeare.txt
negar@negar-VirtualBox:~$ hdfs dfs -copyToLocal /homeworks/t8-copy.txt ~/hw-data/t8-local-2.txt
2020-05-29 12:39:39,307 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTru
sted = false, remoteHostTrusted = false
negar@negar-VirtualBox:~$ ll hw-data/
total 16004
drwxrwxr-x 2 negar negar 4096 May 29 12:39 ./
drwxr-xr-x 20 negar negar 4096 May 29 12:17 ../
-rw-r--r-- 1 negar negar 5458199 May 29 12:38 t8-local-1.txt
-rw-r--r-- 1 negar negar 5458199 May 29 12:39 t8-local-2.txt
-rw-rw-r-- 1 negar negar 5458199 May 29 12:14 t8.shakespeare.txt
negar@negar-VirtualBox:~$
```

برای ایجاد یک کپی از فایل در خود hdfs از cp و برای حذف یک فایل از hdfs از rm استفاده می‌شود. در شکل زیر ابتدا یک کپی با نام t8.txt ایجاد شده و سپس حذف شده است :

```
negar@negar-VirtualBox:~$ hdfs dfs -cp /homeworks/t8-copy.txt /homeworks/t8.txt
2020-05-29 12:45:57,803 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTru
sted = false, remoteHostTrusted = false
2020-05-29 12:45:57,951 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTru
sted = false, remoteHostTrusted = false
negar@negar-VirtualBox:~$ hdfs dfs -ls /homeworks
Found 3 items
-rw-r--r-- 1 negar supergroup 5458199 2020-05-29 12:31 /homeworks/t8-copy.txt
-rw-r--r-- 1 negar supergroup 5458199 2020-05-29 12:20 /homeworks/t8.shakespeare.txt
-rw-r--r-- 1 negar supergroup 5458199 2020-05-29 12:45 /homeworks/t8.txt
negar@negar-VirtualBox:~$ hdfs dfs -rm /homeworks/t8.txt
Deleted /homeworks/t8.txt
negar@negar-VirtualBox:~$ hdfs dfs -ls /homeworks
Found 2 items
-rw-r--r-- 1 negar supergroup 5458199 2020-05-29 12:31 /homeworks/t8-copy.txt
-rw-r--r-- 1 negar supergroup 5458199 2020-05-29 12:20 /homeworks/t8.shakespeare.txt
negar@negar-VirtualBox:~$
```




این موارد را از طریق اینترفیس web هم می‌توان مشاهده کرد که در شکل زیر نشان داده شده است که DFS Browser چگونه محتوای hdfs را نشان می‌دهد. با کلیک روی هر فایل، می‌توان دید که بلاکهای هر فایل و replicat های آن روی چه datanode هایی ذخیره شده‌اند.

Not secure | 192.168.56.101:9870/explorer.html#/homeworks



Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

Browse the file system
Logs
Log Level
Metrics
Configuration
Process Thread Dump

Browse Directory

/homeworks   

Show entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	negar	supergroup	5.21 MB	May 29 12:31	1	128 MB	t8-copy.txt	
<input type="checkbox"/>	-rw-r--r--	negar	supergroup	5.21 MB	May 29 12:20	1	128 MB	t8.shakespeare.txt	

Showing 1 to 2 of 2 entries

Hadoop, 2019.