

## گزارش تمرین سوم درس رایانش ابری

۱. صحت نقش‌های برعهده گرفته شده، با استفاده از دستور jps.

```
vagrant@hadoop-worker1: ~  
vagrant@hadoop-worker1:~$ sudo jps  
5441 NodeManager  
15986 Jps  
5499 DataNode  
vagrant@hadoop-worker1:~$
```

```
vagrant@hadoop-master: ~  
vagrant@hadoop-master:~$ sudo jps  
5994 NameNode  
17342 Jps  
5935 ResourceManager  
vagrant@hadoop-master:~$
```

```
vagrant@hadoop-worker2: ~  
vagrant@hadoop-worker2:~$ sudo jps  
15876 Jps  
5290 NodeManager  
5375 DataNode  
vagrant@hadoop-worker2:~$
```

## ۴. دسترسی پذیری WebGUI از کامپیوتر شخصی.

Activities Brave Web Browser 19:31 7 روزن

HDFS password: [sudo] password: HDFS password: Solved: su had vagrant-hadoop Microsoft Word Namenode Nodes of the cluster does hadoop Go client for H Adventures in C++ Untitled document

Google Calendar Peygir Divar Git ClickUp Divar Slite Divar Rasad Divar Sentry Design Patterns Home | Telepr... Architecture o... Telepresence... Protocol Buff... Scalability & S... ورود | مدیریت و...

### Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Overview 'hadoop-master:9000' (✓active)

Started:	Tue Jun 07 14:43:26 +0430 2022
Version:	3.3.2, r0bcb014209e219273cb6fd4152df7d713cbac61
Compiled:	Mon Feb 21 22:09:00 +0330 2022 by chao from branch-3.3.2
Cluster ID:	CID-1397e610-af4d-4387-9297-00e9f1290e84
Block Pool ID:	BP-1463864613-10.20.30.11-1654596797102

## Summary

Security is off.  
Safemode is off.

5 files and directories, 7 blocks (7 replicated blocks, 0 erasure coded block groups) = 12 total filesystem object(s).

Heap Memory used 55.53 MB of 79 MB Heap Memory. Max Heap Memory is 496 MB.

Non Heap Memory used 70.76 MB of 73.92 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	77.43 GB
Configured Remote Capacity:	0 B
	99.36 MB (0.13%)

10.20.30.11:9870/dfshealth.html#tab-datanode-volume-failures

Activities Brave Web Browser 19:31 7 روزن

HDFS password: [sudo] password: HDFS password: Solved: su had vagrant-hadoop Microsoft Word Namenode Nodes of the cluster does hadoop Go client for H Adventures in C++ Untitled document

Google Calendar Peygir Divar Git ClickUp Divar Slite Divar Rasad Divar Sentry Design Patterns Home | Telepr... Architecture o... Telepresence... Protocol Buff... Scalability & S... ورود | مدیریت و...

### hadoop Nodes of the cluster

Cluster

About  
Nodes  
Node Labels  
Applications  
NEW  
NEW\_SAVING  
SUBMITTED  
ACCEPTED  
RUNNING  
FINISHED  
FAILED  
KILLED  
Scheduler

Tools

#### Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources	Reserved Resources	Physical Mem Used %	Physical Mem Free %
0	0	0	0	0	<memory:0 B, vCores:0>	<memory:4 GB, vCores:16>	<memory:0 B, vCores:0>	54	0

#### Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
2	0	0	0	0	0	0

#### Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority	Scheduler Capacity
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:2048, vCores:4>	0	0

Show 20 entries

Node Labels	Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Containers	Allocation Tags	Mem Used	Mem Avail	Phys Mem Used %	VCores Used	VCores Avail	Phys VCores Used %
/default-rack	RUNNING	hadoop-worker2:36933	hadoop-worker2:8042	Tue Jun 07 13:40:59 +0000 2022	0	0 B	2 GB	53	0	8	0			
/default-rack	RUNNING	hadoop-worker1:42269	hadoop-worker1:8042	Tue Jun 07 13:40:18 +0000 2022	0	0 B	2 GB	54	0	8	0			

Showing 1 to 2 of 2 entries

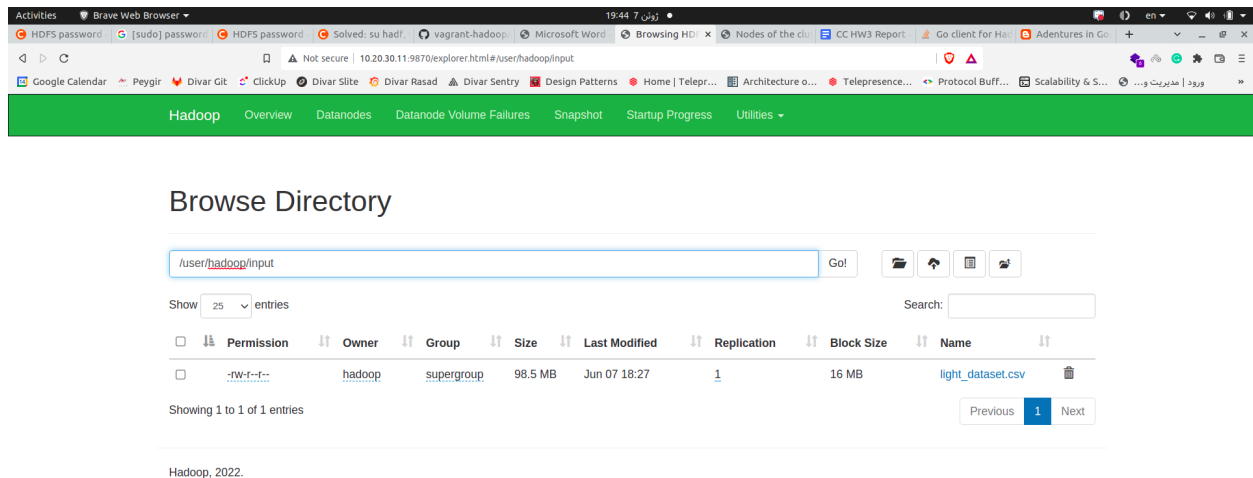
First Previous 1

۵. همان‌طور که در تصویر بالا مشاهده می‌شود، نودهای درخواستی ایجاد شده‌اند. سطر اول، مختص `hadoop-worker2` و سطر بعدی مختص `hadoop-worker1` می‌باشد. هر کدام ۲ گیگابایت مموری در اختیار داشته، و به ترتیب ۵۳ و ۵۴٪ از مموری در اختیار خود را مصرف کرده‌اند. هر دو در وضعیت اجرا (`Node Status=RUNNING`) هستند.

## گام دوم

۱. نتیجه‌ی ایجاد پوشه‌ی `/user/hadoop/`:  
این نتیجه به کمک دستور زیر ایجاد شده است.

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



۳. بارگزاری فایل دیتاست (light\_dataset) در مسیر مشخص شده.  
با کمک دستور زیر این بارگزاری انجام شد.

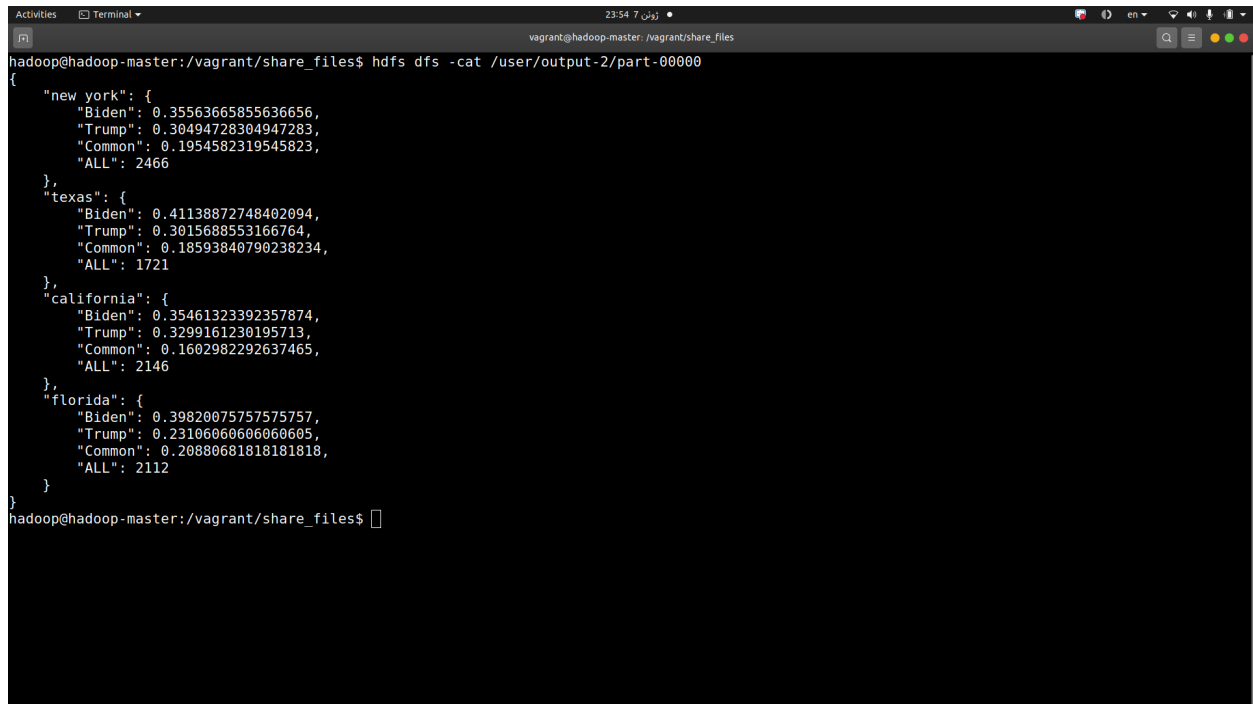
```
hdfs dfs -put datasets/light_dataset.csv /user/hadoop/input/
```

```
vagrant@hadoop-master: /vagrant/share_files
hadoop@hadoop-master:/vagrant/share_files$ hdfs dfs -ls /user/hadoop/input/
Found 1 items
-rw-r--r-- 1 hadoop supergroup 103280184 2022-06-07 13:57 /user/hadoop/input/light_dataset.csv
hadoop@hadoop-master:/vagrant/share_files$
```

۴. اجرای کدها و دریافت نتیجه.

```
hadoop@hadoop-master:/vagrant/share_files$ hdfs dfs -cat /user/output/part-00000
{
  "Trump": {
    "likes": 370451,
    "retweets": 126000,
    "Web": 24885,
    "iPhone": 17710,
    "Android": 15395
  },
  "Biden": {
    "likes": 801346,
    "retweets": 267913,
    "Web": 24735,
    "iPhone": 21786,
    "Android": 14031
  },
  "Common": {
    "likes": 204750,
    "retweets": 78843,
    "Web": 17541,
    "iPhone": 9479,
    "Android": 7916
  }
}
hadoop@hadoop-master:/vagrant/share_files$
```

۵. اجرای کدها و دریافت نتیجه.

A terminal window titled 'Terminal' with a dark background. The prompt is 'hadoop@hadoop-master:~/vagrant/share\_files\$'. The command entered is 'hdfs dfs -cat /user/output-2/part-00000'. The output is a JSON object with four state-level entries: 'new york', 'texas', 'california', and 'florida'. Each entry contains counts for 'Biden', 'Trump', 'Common', and 'ALL'.

```
hadoop@hadoop-master:~/vagrant/share_files$ hdfs dfs -cat /user/output-2/part-00000
{
  "new york": {
    "Biden": 0.35563665855636656,
    "Trump": 0.30494728304947283,
    "Common": 0.1954582319545823,
    "ALL": 2466
  },
  "texas": {
    "Biden": 0.41138872748402094,
    "Trump": 0.3015688553166764,
    "Common": 0.18593840790238234,
    "ALL": 1721
  },
  "california": {
    "Biden": 0.35461323392357874,
    "Trump": 0.3209161230195713,
    "Common": 0.1602982292637465,
    "ALL": 2146
  },
  "florida": {
    "Biden": 0.39820075757575757,
    "Trump": 0.23106060606060605,
    "Common": 0.20880681818181818,
    "ALL": 2112
  }
}
```

hadoop@hadoop-master:~/vagrant/share\_files\$

۶. اجرای کدها و دریافت نتیجه.

```
hadoop@hadoop-master:/vagrant/share_files$ hdfs dfs -cat /user/output-3/part-00000
{
  "new_york": {
    "Biden": 0.3359094672922992,
    "Trump": 0.31162020425062104,
    "Common": 0.21032293679271322,
    "ALL": 3623
  },
  "california": {
    "Biden": 0.3626609442060086,
    "Trump": 0.32103004291845494,
    "Common": 0.16351931330472103,
    "ALL": 2330
  }
}
hadoop@hadoop-master:/vagrant/share_files$
```

این تفاوت، به دلیل موارد زیر می‌تواند باشد.

- حدودی بودن موقعیت‌های جغرافیایی
- عدم تناسب موقعیت جغرافیایی با استان. مثلاً بعضی سطرها موقعیت جغرافیایی مشخصی دارند، اما استانشان ذکر نشده بود، یا برعکس.