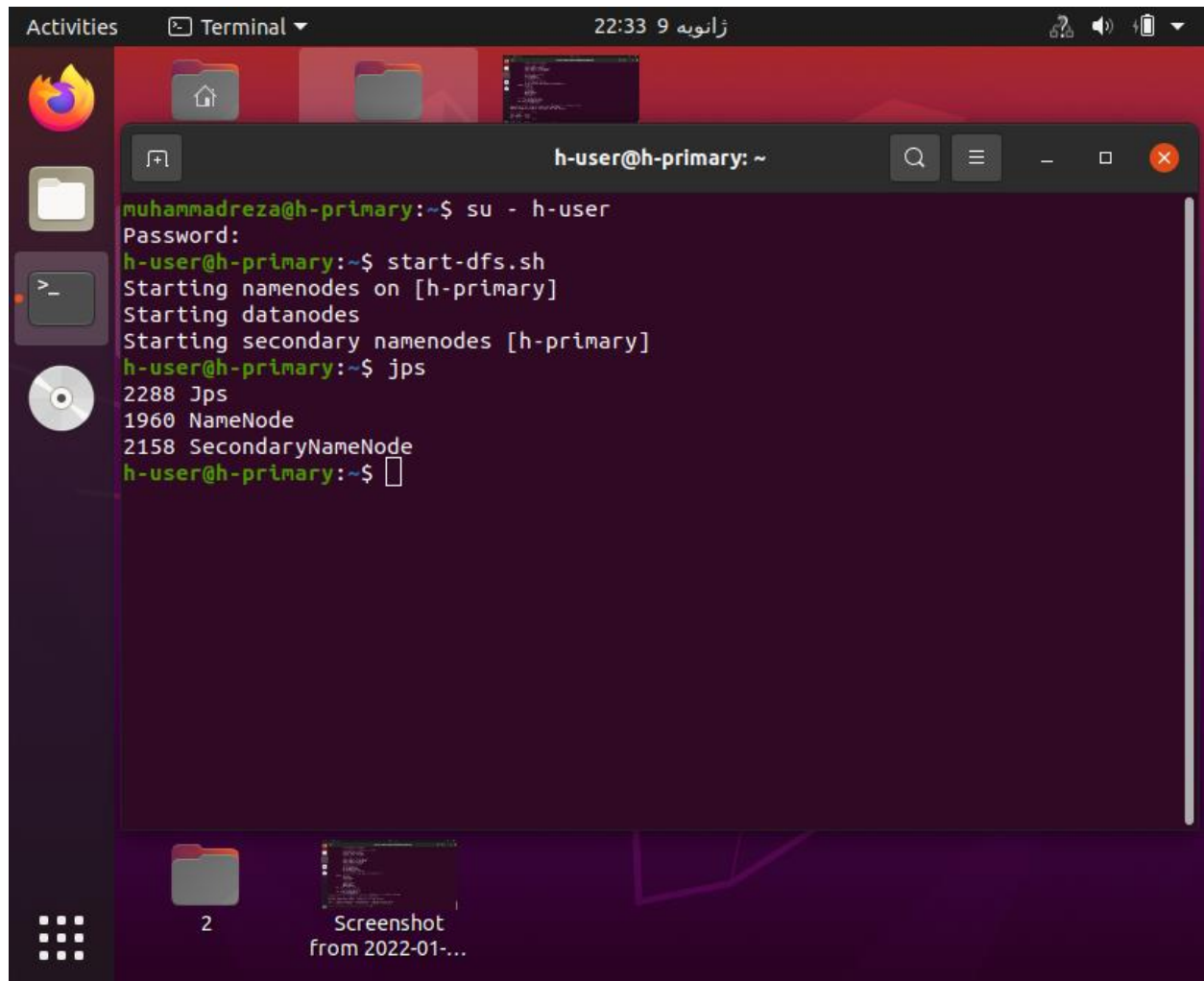


به نام خدا

محمدرضا قادری ۹۶۲۷۰۵۷

در ابتدا توسط فایلی که موجود بود سه تا ماشین مجازی برای انجام محاسبات ایجاد میکنیم با دستور زیر می‌توانیم ببینیم که هر دوتا workerها قابل اجرا هستند.



```
Activities Terminal 22:33 9 ژانویه
h-user@h-primary: ~
muhammadsa@h-primary:~$ su - h-user
Password:
h-user@h-primary:~$ start-dfs.sh
Starting namenodes on [h-primary]
Starting datanodes
Starting secondary namenodes [h-primary]
h-user@h-primary:~$ jps
2288 Jps
1960 NameNode
2158 SecondaryNameNode
h-user@h-primary:~$
```

```
Activities  Terminal  03:29 6 زانویه
h-user@h-secondary: ~
muhmadreza@h-secondary:~$ ip addr
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group defau
lt qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
2: enp0s3: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc fq_codel state UP g
roup default qlen 1000
    link/ether 08:00:27:ed:04:bb brd ff:ff:ff:ff:ff:ff
    inet 192.168.1.110/24 brd 192.168.1.255 scope global dynamic noprefixroute
    enp0s3
        valid_lft 258944sec preferred_lft 258944sec
    inet6 fe80::28e:3f56:40e0:26b0/64 scope link noprefixroute
        valid_lft forever preferred_lft forever
muhmadreza@h-secondary:~$ sudo nano /etc/hosts
[sudo] password for muhmadreza:
muhmadreza@h-secondary:~$ su - h-user
Password:
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

h-user@h-secondary:~$ jps
2321 Jps
2206 DataNode
h-user@h-secondary:~$
```

که می بینیم هم مستر می تواند ورکرها را بشناسد.

از طریق webGUI هم به صورت زیر می بینیم که هدوپ ایجاد شده و مشخصات و منابع قابل دسترس برای ورکرها توسط مستر هست مشخص شده است.

Activities Firefox Web Browser 22:35 9 زانويه

AUT WebMail x Namenode information x All Applications x +

192.168.1.114:9870/dfshealth.html#tab-overview

Heap Memory used 30.83 MB of 43.07 MB Heap Memory. Max Heap Memory is 479.5 MB.
Non Heap Memory used 50.08 MB of 51.21 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	38.13 GB
Configured Remote Capacity:	0 B
DFS Used:	99.37 MB (0.25%)
Non DFS Used:	19.35 GB
DFS Remaining:	16.71 GB (43.81%)
Block Pool Used:	99.37 MB (0.25%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.51% / 0.51% / 0.25%
Live Nodes	2 (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

Activities Firefox Web Browser 22:35 9 زانويه

AUT WebMail x Namenode information x All Applications x +

192.168.1.114:9870/dfshealth.html#tab-overview

Current transaction ID: 199

Journal Manager	State
FileJournalManager(root=/usr/local/hadoop/data/nameNode)	EditLogOutputStream(/usr/local/hadoop/data/nameNode/current/edits_inprogress_0000000000000000199)

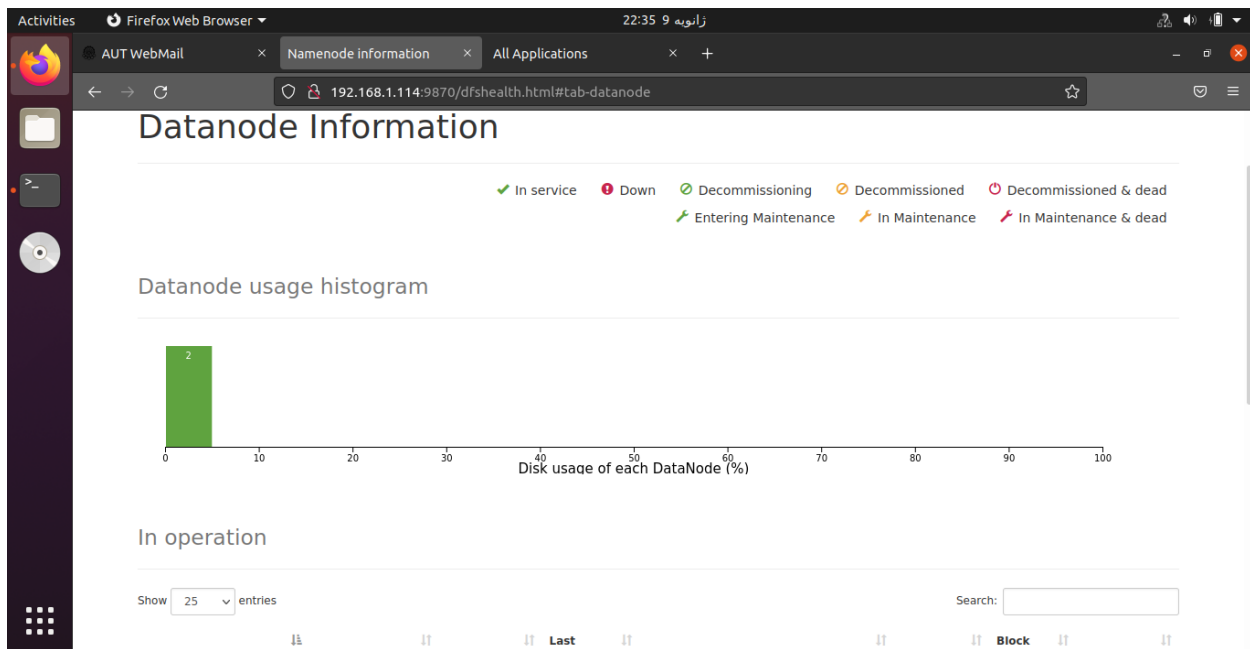
NameNode Storage

Storage Directory	Type	State
/usr/local/hadoop/data/nameNode	IMAGE_AND_EDITS	Active

DFS Storage Types

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes In Service
DISK	38.13 GB	99.37 MB (0.25%)	16.71 GB (43.81%)	99.37 MB	2

Hadoop, 2021.



Activities Firefox Web Browser 22:35 9 زانويه

AUT WebMail x Namenode information x All Applications x +

192.168.1.114:9870/dfshealth.html#tab-datanode

In operation

Show 25 entries Search:

Node	Http Address	Last contact	Last Block Report	Capacity	Blocks	Block pool used	Version
✔ h-secondary:9866 (192.168.1.115:9866)	http://h-secondary:9864	2s	2m	19.07 GB <div><div></div></div>	1	99.3 MB (0.51%)	3.2.2
✔ h-tertiary:9866 (192.168.1.116:9866)	http://h-tertiary:9864	2s	2m	19.07 GB <div><div></div></div>	3	64 KB (0%)	3.2.2

Showing 1 to 2 of 2 entries

Previous 1 Next

Entering Maintenance

No nodes are entering maintenance.

Decommissioning

Activities Firefox Web Browser 22:37 9 ژانویه

AUT WebMail x Browsing HDFS x All Applications x +

192.168.1.114:8088/cluster

All Applications

er Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources	Reserved Resources
0	0	0	0	0	<memory:0, vCores:0>	<memory:0, vCores:0>	<memory:0, vCores:0>

er Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes
0	0	0	0	0

uler Metrics

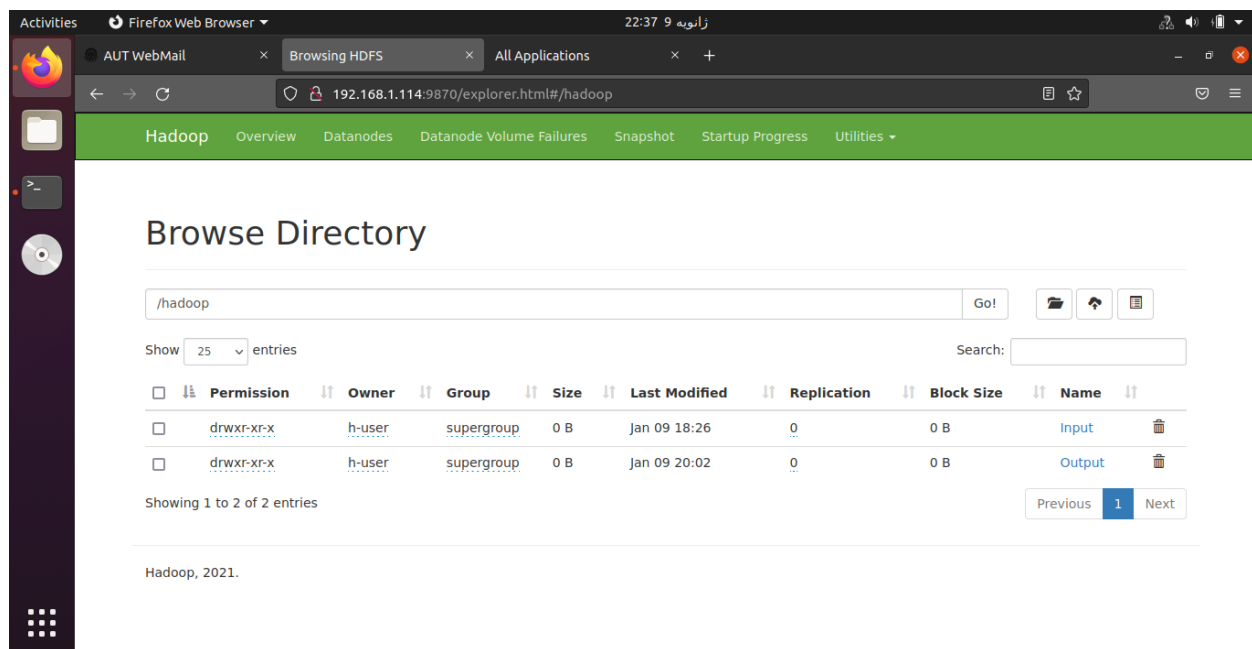
Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>

20 entries

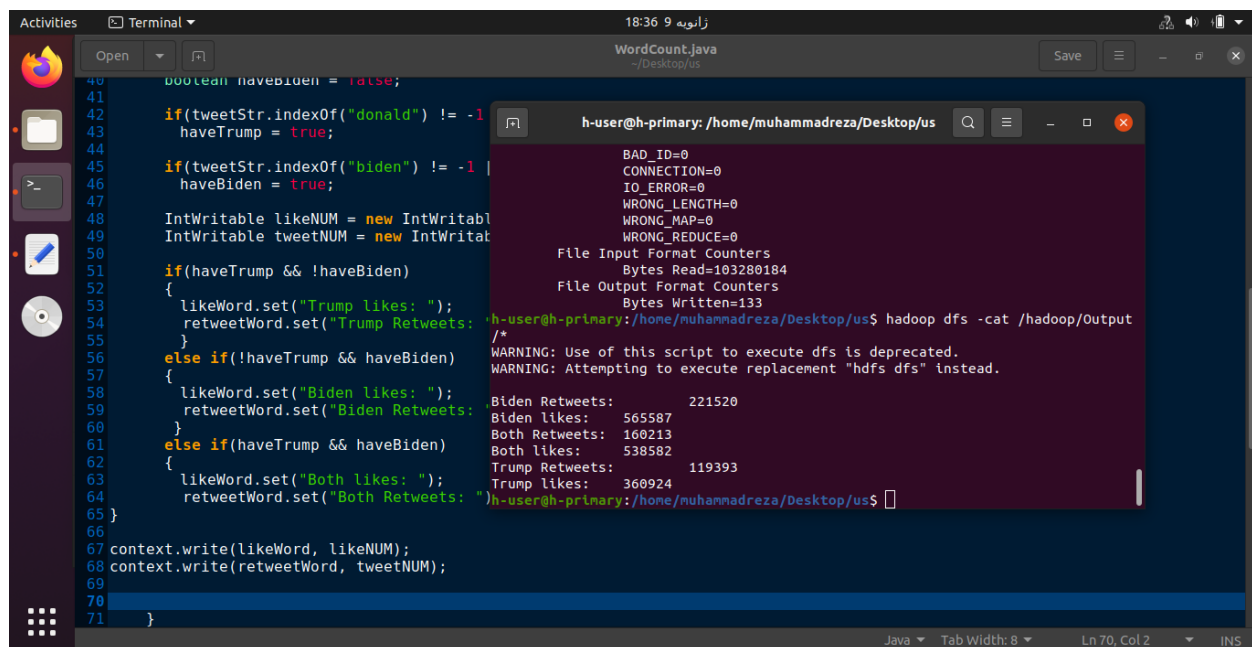
User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCoers	Allocated Memory MB	Allocated GPUs	Reserved CPU VCoers	Reserved Memory MB
No data available in table															

ing 0 to 0 of 0 entries

در قسمت دوم میبایستی ما در ابتدا یک پوشه برای پروژه خود روی هدوپ ایجاد کنیم که ورودی‌ها و خروجی‌ها را آنجا داشته باشیم بایستی در ابتدا یک پوشه برای پروژه ایجاد کنیم که در آن یک پوشه برای گذاشتن ورودی‌ها قرار می‌دهیم و سپس فایل ورودی مورد نظر رو در آنجا قرار می‌دهیم و در ادامه باید یک کد برای اجرای دستورات بزنیم سپس برای اینکه بتوانیم در هدوپ از آنها استفاده کنیم بایستی یک فایل قابل اجرا بسازیم چون با جاوا کار داریم این کار با jar انجام میشود که توسط کلاسهای ران گرفته از کد ساخته می‌شود. حالا برای قسمت اول که فقط قرار است لایک و ریتوییت‌های هر کدام را بشماریم فایل را می‌خوانیم و قسمت متن رو چک میکنیم و لایک و ریتوییت رو به صورت چندتایی میدیم.



برای دیتاست سبک داریم.



برای تعداد توییت ها در کشورهای متفاوت

```
Activities Terminal 19:58 9 زانویه
h-user@h-primary: /home/muhamadreza/Desktop/us

Spilled Records=204440
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=610
Total committed heap usage (bytes)=274866176

Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0

File Input Format Counters
Bytes Read=103280184
File Output Format Counters
Bytes Written=848

h-user@h-primary:/home/muhamadreza/Desktop/us$ hadoop dfs -cat /hadoop/Output/*
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

america 0.1894727008234462 0.36551935832588184 0.3118018390743237 74929
australia 0.1652930402930403 0.2445054945054945 0.4084249084249084 2184
canada 0.20989563200618477 0.2792810204870506 0.3850019327406262 5174
enirates 0.3069306930693069 0.21287128712871287 0.31683168316831684 202
engalnd 0.21666937978186554 0.2236692169949536 0.43154810353247597 6143
france 0.4670757611517583 0.1463299504366297 0.31980174651876325 4237
germany 0.26912100456621 0.1906392694063927 0.449486301369863 3504
iran 0.2807017543859649 0.23976608187134502 0.3684210526315789 171
italy 0.30056497175141245 0.21977401129943502 0.3768361581920904 1770
mexico 0.38602520045819017 0.17983963344788087 0.38831615120274915 873
netherlands 0.20743034055727555 0.31977001326846527 0.32109685979655017 2261
spain 0.3082901554404145 0.22409326424870465 0.3588082901554404 772

h-user@h-primary:/home/muhamadreza/Desktop/us$
```

و برای امریکا و فرانسه با مختصات به صورت زیر هست

```
Activities Terminal 20:03 9 زانویه
h-user@h-primary: /home/muhamadreza/Desktop/us

Map output records=102688
Map output bytes=1636026
Map output materialized bytes=1841408
Input split bytes=117
Combine input records=0
Combine output records=0
Reduce input groups=2
Reduce shuffle bytes=1841408
Reduce input records=102688
Reduce output records=2
Spilled Records=205376
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=663
Total committed heap usage (bytes)=274866176

Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0

File Input Format Counters
Bytes Read=103280184
File Output Format Counters
Bytes Written=143

h-user@h-primary:/home/muhamadreza/Desktop/us$ hadoop dfs -cat /hadoop/Output/*
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

america 0.1982425344283535 0.33982195473638016 0.33065847491275363 95706
france 0.3892867373245488 0.16299054712116873 0.3627900315095961 6982

h-user@h-primary:/home/muhamadreza/Desktop/us$
```

با توجه به اینکه خیلی از مختصات به صورت درست و دقیق نیست و خیلی از خانه های مختصات برای توییت خالی است و حتی ممکن است استفاده از **vpn** ها باعث ایجاد اختلافاتی در مختصات در زمان ایجاد توییت کند این عمر باعث می شود. (موارد در حالت استفاده از مختصات بیشتر است که منطقی هست) یا برای مثال **gpa** روشن بوده ولی به اشتباه و یا به صورت پیش فرض ست کرده بوده باشد به این کشورها.

