

به نام خدا

تمرین سوم درس رایانش ابری - بهار ۱۴۰۲

مریم کرمانشاهانی - ۹۷۲۳۰۷۳

۱. **NameNode**: کانتینر **NameNode** مسئول مدیریت و کنترل کلیه فایل‌ها و داده‌های موجود در سیستم **Hadoop** است. آن مشخص می‌کند که فایل‌ها و داده‌ها در دیسک‌هایی که در **DataNode**‌ها ذخیره می‌شوند، چگونه سازماندهی شوند **NameNode**. همچنین اطلاعاتی درباره مکان فایل‌ها و داده‌ها را نگهداری می‌کند.

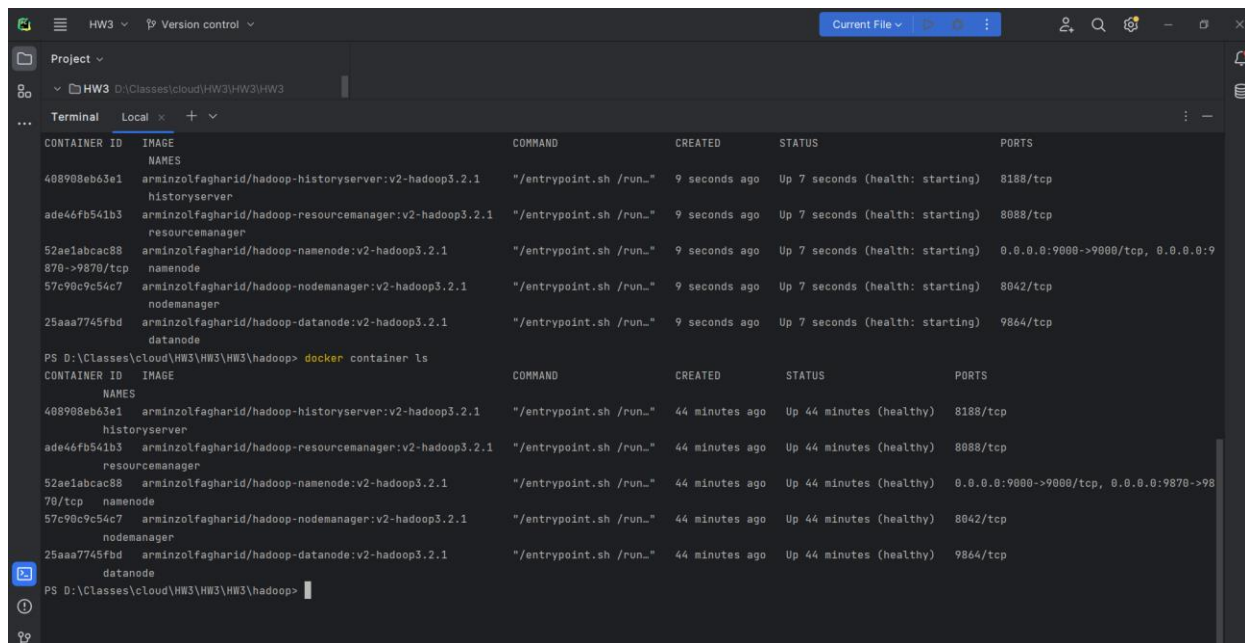
۲. **DataNode**: کانتینر **DataNode** نماینده فیزیکی هر دیسک و حافظه موجود در سیستم **Hadoop** است. این کانتینرها وظیفه ذخیره و مدیریت بخشی از داده‌ها را برعهده دارند. آنها داده‌ها را در دیسک‌های خود ذخیره می‌کنند و به **NameNode** اطلاع می‌دهند که داده‌های آنها در کدام **DataNode** واقع شده است.

۳. **NodeManager**: کانتینر **NodeManager** مسئول مدیریت و کنترل منابع محاسباتی در یک کانتینر **DataNode** است. این کانتینر بررسی می‌کند که منابع محاسباتی (مانند پردازنده و حافظه) کانتینر **DataNode** برای اجرای وظایف **MapReduce** مورد نیاز هستند یا خیر، و منابع مورد نیاز را به آن اختصاص می‌دهد.

۴. **ResourceManager**: کانتینر **ResourceManager** مسئول مدیریت و تنظیم وظایف **MapReduce** در سیستم **Hadoop** است. این کانتینر منابع محاسباتی را به نودهای **MapReduce** تخصیص می‌دهد و وظایف مورد نظر را در سرورهای منابع محاسباتی قرار می‌دهد.

۵. **HistoryManager**: کانتینر **HistoryManager** مسئول ثبت و نگهداری وظایف اجرا شده در سیستم **Hadoop** است. این کانتینر وظیفه ذخیره و نمایش تاریخچه اجرای وظایف **MapReduce** را برعهده دارد تا بتوان بررسی کرد که وظایف به درستی اجرا شده اند یا خیر.

با استفاده از این کانتینرها، سیستم **Hadoop** به صورت توزیع‌شده و قابل مقیاس‌پذیری عمل می‌کند و به مدیریت و پردازش حجم بزرگی از داده‌ها قادر است.



در عکس پایین بهتر معلوم میشود:

```
D:\Classes\cloud\VM3\VM3\hadoop>docker ps
CONTAINER ID   IMAGE                                     COMMAND                  CREATED        STATUS        PORTS                               NAMES
da53d380f7c7   arminzolfagharid/hadoop-nodemanager:v2-hadoop3.2.1   "/entrypoint.sh /run_"   About an hour ago   Up About an hour (healthy)   8042/tcp   nodemanager
a9ff32f9f529   arminzolfagharid/hadoop-datanode:v2-hadoop3.2.1     "/entrypoint.sh /run_"   About an hour ago   Up About an hour (healthy)   9864/tcp   datanode
2f828d90bab4   arminzolfagharid/hadoop-namenode:v2-hadoop3.2.1     "/entrypoint.sh /run_"   About an hour ago   Up About an hour (healthy)   0.0.0.0:9000->9000/tcp, 0.0.0.0:9870->9870/tcp   namenode
ceb6b8bfce5a   arminzolfagharid/hadoop-resourcemanager:v2-hadoop3.2.1 "/entrypoint.sh /run_"   About an hour ago   Up About an hour (healthy)   8088/tcp   resourcemanager
21807896c453   arminzolfagharid/hadoop-historyserver:v2-hadoop3.2.1 "/entrypoint.sh /run_"   About an hour ago   Up About an hour (healthy)   8188/tcp   historyserver

D:\Classes\cloud\VM3\VM3\hadoop>docker exec -it nodemanager bash
root@da53d380f7c7:/# jps
1862 Jps
361 NodeManager
root@da53d380f7c7:/# exit
exit

D:\Classes\cloud\VM3\VM3\hadoop>docker exec -it datanode bash
root@a9ff32f9f529:/# jps
1905 Jps
363 DataNode
root@a9ff32f9f529:/# exit
exit

D:\Classes\cloud\VM3\VM3\hadoop>docker exec -it namenode bash
root@2f828d90bab4:/# jps
421 NameNode
1915 Jps
root@2f828d90bab4:/# exit
exit

D:\Classes\cloud\VM3\VM3\hadoop>docker exec -it resourcemanager bash
root@ceb6b8bfce5a:/# jps
358 ResourceManager
2056 Jps
root@ceb6b8bfce5a:/# exit
exit

D:\Classes\cloud\VM3\VM3\hadoop>docker exec -it historymanager bash
Error: No such container: historymanager
```

```

361 NodeManager
root@da53d380f7c7:/# exit
exit

D:\Classes\cloud\HW3\HW3\hadoop>docker exec -it datanode bash
root@a9ff32f9f529:/# jps
1905 Jps
363 DataNode
root@a9ff32f9f529:/# exit
exit

D:\Classes\cloud\HW3\HW3\hadoop>docker exec -it namenode bash
root@2f828d90bab4:/# jps
421 NameNode
1915 Jps
root@2f828d90bab4:/# exit
exit

D:\Classes\cloud\HW3\HW3\hadoop>docker exec -it resourcemanager bash
root@ce6b8bffce5a:/# jps
358 ResourceManager
2056 Jps
root@ce6b8bffce5a:/# exit
exit

D:\Classes\cloud\HW3\HW3\hadoop>docker exec -it historymanager bash
Error: No such container: historymanager

D:\Classes\cloud\HW3\HW3\hadoop>docker exec -it historyserver bash
root@21807896c453:/# jps
368 ApplicationHistoryServer
1851 Jps
root@21807896c453:/#

```

Overview 'namenode:9000' (active)

Started:	Mon Jun 05 11:59:54 +0330 2023
Version:	3.2.1, rb3cbbb467e22ea829b38084b7b01d07e0bf3842
Compiled:	Tue Sep 10 20:26:00 +0430 2019 by rohithsharmaks from branch-3.2.1
Cluster ID:	CID-3a75f0c5-9e3f-43ff-976f-37a62cd91a5a
Block Pool ID:	BP-625210781-172.20.0.6-1685953791718

Summary

Security is off.

Safemode is off.

12 files and directories, 5 blocks (5 replicated blocks, 0 erasure coded block groups) = 17 total filesystem object(s).

Heap Memory used 134.76 MB of 341.5 MB Heap Memory. Max Heap Memory is 1.71 GB.

Non Heap Memory used 53.64 MB of 54.94 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	250.92 GB
Configured Remote Capacity:	0 B

NameNode Journal Status

Current transaction ID: 37

Journal Manager	State
FileJournalManager(root=/hadoop/dfs/name)	EditLogFileOutputStream(/hadoop/dfs/name/current/edits_inprogress_00000000000000000001)

NameNode Storage

Storage Directory	Type	State
/hadoop/dfs/name	IMAGE_AND_EDITS	Active

NameNode Storage

DFS Storage Types

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes in Service
DISK	250.92 GB	76 KB (0%)	228.63 GB (91.12%)	76 KB	1

Browse Directory

Show 25 entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Jun 10 20:32	0	0 B	root	

Showing 1 to 1 of 1 entries

Hadoop, 2019.

بعد از بالا آوردن کانتینرها ورد ترمینال namenode شدم که این کار از طریق خود اپلیکیشن داکر هم امکان پذیر بود بعد از آن دایرکتوری هایی برای قراردادن فایل هایم ساختم

```
# docker container ls
/bin/sh: 1: docker: not found
# ls
KEYS boot entrypoint.sh hadoop home lib64 mnt proc run/sbin sys usr
bin dev etc hadoop-data lib media opt root run.sh srv tmp var
# mkdir cloud/step2
mkdir: cannot create directory 'cloud/step2': No such file or directory
# mkdir cloud
# cd cl
/bin/sh: 5: cd: can't cd to cl
# cd cloud
# ls
# mkdir step2
# cd ..
# ls
KEYS boot dev etc hadoop-data lib media opt root run.sh srv tmp var
bin cloud entrypoint.sh hadoop home lib64 mnt proc run/sbin sys usr
# cd cloud
# ls
step2
# mkdir step1
# mkdir step3
#
```

با دستورات زیر هم فایل های پایتون و مجموعه داده ام را به کانتینر منتقل کردم.

```
(venv) PS D:\Classes\cloud\HW3\python-code> cd .\step1\  
(venv) PS D:\Classes\cloud\HW3\python-code\step1> cd ..  
(venv) PS D:\Classes\cloud\HW3\python-code> cd step2  
(venv) PS D:\Classes\cloud\HW3\python-code\step2> docker cp mapper.py namenode:/cloud/step2  
Error: No such container:path: namenode:/cloud  
(venv) PS D:\Classes\cloud\HW3\python-code\step2> docker cp mapper.py namenode:/cloud/step2  
(venv) PS D:\Classes\cloud\HW3\python-code\step2> docker cp reducer.py namenode:/cloud/step2  
(venv) PS D:\Classes\cloud\HW3\python-code\step2> docker cp dataset.csv namenode:/cloud  
(venv) PS D:\Classes\cloud\HW3\python-code\step2>
```

بعد هم باید فایل را به هدوپ منتقل کنم.

در صورت ایجاد مشکل در نیم نود میتوانید والیوم ها را پاک کنید تا مشکل حل شود:

```
Terminal Local (2) Local (3)  
(venv) PS D:\Classes\cloud\HW3\python-code> cd .\step1\  
(venv) PS D:\Classes\cloud\HW3\python-code\step1> cd ..  
(venv) PS D:\Classes\cloud\HW3\python-code> cd step2  
(venv) PS D:\Classes\cloud\HW3\python-code\step2> docker cp mapper.py namenode:/cloud/step2  
Error: No such container:path: namenode:/cloud  
(venv) PS D:\Classes\cloud\HW3\python-code\step2> docker cp mapper.py namenode:/cloud/step2  
(venv) PS D:\Classes\cloud\HW3\python-code\step2> docker cp reducer.py namenode:/cloud/step2  
(venv) PS D:\Classes\cloud\HW3\python-code\step2> docker cp dataset.csv namenode:/cloud  
(venv) PS D:\Classes\cloud\HW3\python-code\step2> docker volume rm hadoop_hadoop.datanode
```

سوال سوم

```
Bytes Written=124  
2023-06-10 17:17:05,829 INFO streaming.StreamJob: Output directory: /user/root/output/step1  
root@52ae1abcac88:/cloud/step1# hdfs dfs -cat /user/root/output/step1/part-00000  
2023-06-10 17:17:32,795 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false  
Both 204750.0 78843.0 67163 48975 37343  
Biden 801355.0 267920.0 67163 48975 37343  
Trump 370451.0 126000.0 67163 48975 37343  
root@52ae1abcac88:/cloud/step1#
```

سوال چهارم

```
get /cloud/results/step2/part-00000 : file exists  
root@52ae1abcac88:/cloud/results# hdfs dfs -cat /user/root/output/step2/part-00000  
2023-06-10 17:15:28,040 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false  
New York 0.2311455360782715 0.4235628041011007 0.3452914790206278 2453  
Texas 0.19752391671356218 0.4670793472144063 0.3353967360720315 1777  
California 0.19982547993019198 0.407940666317626526 0.39223385689354273 2292  
Florida 0.24298597194388777 0.4634268537074148 0.2935871743486974 1996  
root@52ae1abcac88:/cloud/results#
```

سوال پنجم

```
Docker Desktop Update to latest  
Search for local and remote images, containers, and more... Ctrl+K  
Containers  
Images  
Volumes  
Dev Environments  
Extensions  
Add Extensions  
namenode arminzolfagharid/hadoop-namenode:v2-hadoop3.2.1  
52ae1abcac88  
9000:9000 9870:9870  
STATUS  
Running (26 minutes ago)  
Logs Inspect Terminal Files Stats  
Open in external terminal  
root@52ae1abcac88:/cloud/results/step2# cd ..  
root@52ae1abcac88:/cloud/results# hdfs dfs -get /user/root/output/step3/part-00000 /cloud/results/step3/  
2023-06-10 17:12:54,289 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false  
root@52ae1abcac88:/cloud/results# hdfs dfs -cat /user/root/output/step3/part-00000  
2023-06-10 17:13:15,790 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false  
New York 0.25124515771997785 0.3929164360819037 0.3558384061981184 3614  
California 0.20056383407168749 0.41602899718082964 0.3834071687474829 2483  
root@52ae1abcac88:/cloud/results# hadoop jar /opt/hadoop-3.2.1/share/hadoop/tools/lib/hadoop-streaming-3.2.1.jar -file /cloud/step2/mapper.py -mapper "python3 mapper.py" -file /c
```

همانطور که میبینیم نتیجه مربوط به سوال پنجم عدد های بیتشری است و به طور کلی اعداد
بیشتری بررسی شده ؛ این یعنی رنج گسترده تری در برگرفته شده.