

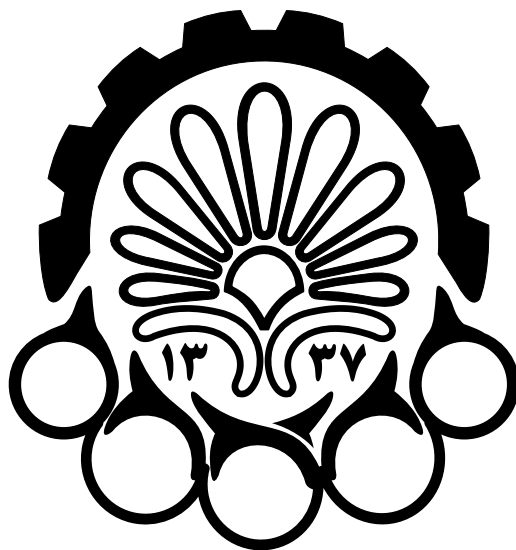
تمرین سوم درس مبانی رایانش ابری

بهار ۱۴۰۲

استاد درس : دکتر جوادی

مهلا شریفی

شماره دانشجویی : ۹۸۳۱۰۳۵



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

گام اول: راه اندازی خوشه Hadoop

• نمایش کانتینرهای ایجاد شده

CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS	PORTS	NAMES
2423a89da1ac	arminzolfagharid/hadoop-nodemanager:v2-hadoop3.2.1	"/entrypoint.sh /run..."	12 minutes ago	Up 12 minutes (unhealthy)	8042/tcp	nodemanager
35b904b65f78	arminzolfagharid/hadoop-historyserver:v2-hadoop3.2.1	"/entrypoint.sh /run..."	12 minutes ago	Up 12 minutes (unhealthy)	8188/tcp	historyserver
80e437990723	arminzolfagharid/hadoop-resourcemanager:v2-hadoop3.2.1	"/entrypoint.sh /run..."	12 minutes ago	Up 12 minutes (unhealthy)	8088/tcp	resourcemanager
33bb6931900a	arminzolfagharid/hadoop-namenode:v2-hadoop3.2.1	"/entrypoint.sh /run..."	12 minutes ago	Up 24 seconds (health: starting)	0.0.0.0:9000->9000/tcp, 0.0.0.0:9870->9870/tcp	namenode
8c769c985307	arminzolfagharid/hadoop-datanode:v2-hadoop3.2.1	"/entrypoint.sh /run..."	12 minutes ago	Up 12 minutes (unhealthy)	9864/tcp	datanode

COMMAND	CREATED	STATUS	PORTS
"/entrypoint.sh /run..."	12 minutes ago	Up 12 minutes (unhealthy)	8042/tcp
"/entrypoint.sh /run..."	12 minutes ago	Up 12 minutes (unhealthy)	8188/tcp
1 "/entrypoint.sh /run..."	12 minutes ago	Up 12 minutes (unhealthy)	8088/tcp
"/entrypoint.sh /run..."	12 minutes ago	Up 24 seconds (health: starting)	0.0.0.0:9000->9000/tcp, 0.0.0.0:9870->9870/tcp
"/entrypoint.sh /run..."	12 minutes ago	Up 12 minutes (unhealthy)	9864/tcp

• توضیح وظیفه هرکدام از کانتینرها در Hadoop

NameNode - 1

Namenode مسئول مدیریت فایل سیستم توزیع شده‌ی هدوپ یا همان hdfs است.

و یکسری متادیتا راجع به فایل‌ها من جمله ساختار دایرکتوری و مجوز را نگهداری می‌کند.

به طور کلی به هماهنگی بازیابی و ذخیره اطلاعات می‌پردازد.

HistoryServer- 1

به طور کلی در هدوپ یکسری جاب اجرا می‌گردد.

سابقه و گزارش‌های جاب‌های تکمیل شده توسط history server ذخیره می‌گردد.

با استفاده از ذخیره سازی این اطلاعات امکان تحلیل و debug و monitoring محیا می‌گردد.

ResourceManager- 3

مدیریت منابع را به نحوی کارآمد انجام می‌دهد.

اطلاعات منابع در دست این کانتینر است و زمان بندی و اختصاص منابع به اپلیکیشن های مختلفی که قرار است با استفاده از منابع کلاستر اجرا شوند، توسط این بخش انجام می پذیرد.

NodeManager- 4

بر روی هر ماشین جداگانه در کلاستر اجرا می شود و مسئول مدیریت منابع در سطح گره است .
استفاده از منابع را در گره ردیابی می کند و به resource manager گزارش می دهد.
و نیز کانتینرهایی که وظیفه اجرای وظایف یا برنامه های خاص بر عهده دارند را راه اندازی و نظارت می کند

DataNode -5

مسئول ذخیره و بازیابی داده ها در هدوپ است.
درخواست های خواندن و نوشتن را مدیریت می کند. و پس از دریافت اطلاعات بلوک ها را از namenode، به ایجاد بلوک ها می پردازد.

-

- با استفاده از دستور jps در هر کانتینر، صحت نقش آن کانتینر در Hadoop را بررسی کنید و اسکرین شات آن را بیاورید.

```
C:\WINDOWS\system32>docker exec -it nodemanager bash
root@2423a89da1ac:/# jps
584 NodeManager
811 Jps
```

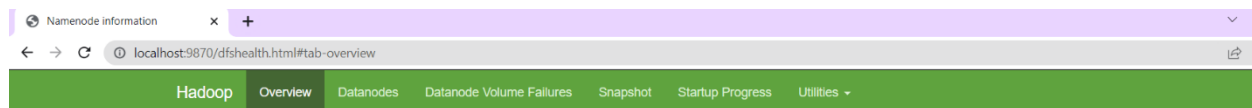
```
C:\WINDOWS\system32>docker exec -it historyserver bash
root@05b904b65f78:/# jps
592 ApplicationHistoryServer
772 Jps
```

```
C:\WINDOWS\system32>docker exec -it resourcemanager bash
root@10e437990723:/# jps
583 ResourceManager
957 Jps
```

```
C:\WINDOWS\system32>docker exec -it namenode bash
root@03bb6931900a:/# jps
421 NameNode
623 Jps
```

```
C:\WINDOWS\system32>docker exec -it datanode bash
root@8c769c985307:/# jps
824 Jps
587 DataNode
```

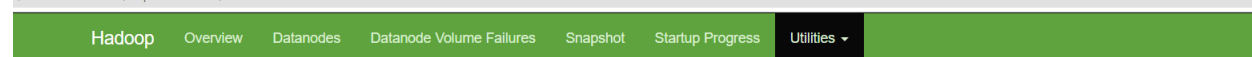
• نمایش WebUI برای NameNode (localhost:9870) و فایل سیستم آن



Overview 'namenode:9000' (active)

Started:	Fri Jun 09 16:51:46 +0330 2023
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Tue Sep 10 20:26:00 +0430 2019 by rohithsharmaks from branch-3.2.1
Cluster ID:	CID-5bf37fac-4a3f-4fdb-97e8-534358b40908
Block Pool ID:	BP-1894593815-172.20.0.6-1686316904584

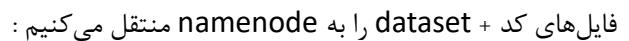
localhost:9870/explorer.html#



Browse Directory

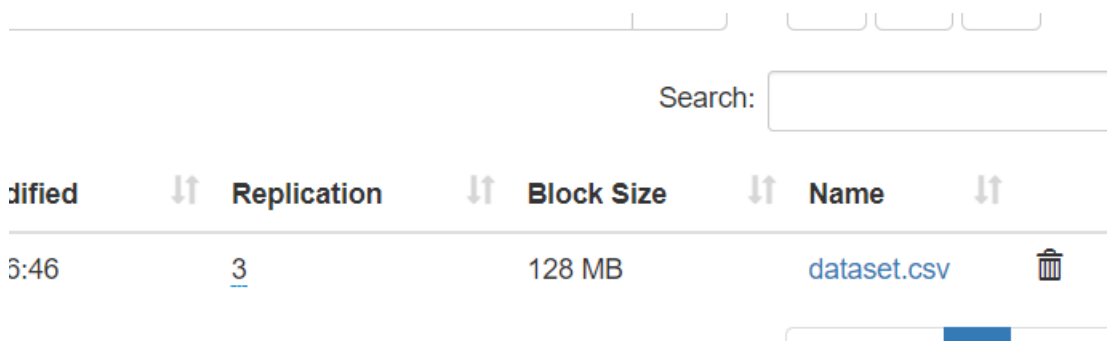
/								Go!			
Show	25	entries	Search:								
<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name			
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Jun 09 16:51	0	0 B	rmstate			
Showing 1 to 1 of 1 entries									Previous 1 Next		

Hadoop, 2019.



دیتاست را از namenode به hdfs منتقل کردیم. همانطور که در دستور کار ذکر شده است.

Dataset با موفقیت به hdfs افزوده شده است :



اجرای موفقیت آمیز MapReduce اول به همراه نتیجه

```
WRONG_REDUCE=0
job output is in hdfs:///output/MapReduce1
Removing HDFS temp directory hdfs:///user/root/tmp/mrjob/MapReduce1.root.20230612.033408.947488...
Removing temp directory /tmp/MapReduce1.root.20230612.033408.947488...
```

اضافه شدن فایل خروجی به hdfs

↕↕		Block Size		↕↕		Name		↕↕			
		0 B				MapReduce1					
Group	↕↕	Size	↕↕	Last Modified	↕↕	Replication	↕↕	Block Size	↕↕	Name	↕↕
supergroup		0 B		Jun 12 07:04		3		128 MB		_SUCCESS	
supergroup		143 B		Jun 12 07:04		3		128 MB		part-00000	

```
root@03bb6931900a:~# hdfs dfs -cat hdfs:///output/MapReduce1/part-00000
2023-06-12 03:36:52,868 INFO sasl.SaslDataTransferClient: SASL encryption
"Both Candidate"      [10484.0, 2279.0, 219, 49, 34]
"Donald Trump"        [8593.0, 3450.0, 737, 355, 302]
"Joe Biden"           [27554.0, 14986.0, 1815, 1046, 750]
```

اجرای موفقیت آمیز MapReduce دوم به همراه نتیجه

```
WRONG_REDUCE=0
job output is in hdfs:///output/MapReduce2
Removing HDFS temp directory hdfs:///user/root/tmp/mrjob/MapReduce2.root.20230612.033740.446635...
Removing temp directory /tmp/MapReduce2.root.20230612.033740.446635...
```

↕↕		Replication		↕↕		Block Size		↕↕		Name		↕↕	
		0				0 B				MapReduce1			
		0				0 B				MapReduce2			

↕↕	Permission	↕↕	Owner	↕↕	Group	↕↕	Size	↕↕	Last Modified	↕↕	Replication	↕↕	Block Size	↕↕	Name	↕↕
	-rw-r--r--		root		supergroup		0 B		Jun 12 07:08		3		128 MB		_SUCCESS	
	-rw-r--r--		root		supergroup		336 B		Jun 12 07:08		3		128 MB		part-00000	

```

root@03bb6931900a:~# hdfs dfs -cat hdfs:///output/MapReduce2/part-00000
2023-06-12 03:40:09,346 INFO sasl.SaslDataTransferClient: SASL encryption trust check: local
"California" [0.002570694087403599, 0.016893132574366507, 0.010650018362100624, 2723]
"Florida" [0.0012568077084206116, 0.03142019271051529, 0.008797653958944282, 2387]
"New York" [0.0017445917655268667, 0.023028611304954642, 0.006629448709002094, 2866]
"Texas" [0.0005035246727089627, 0.026183282980866064, 0.005035246727089627, 1986]




```



اجرای موفقیت آمیز MapReduce سوم به همراه نتیجه

```

job output is in hdfs:///output/MapReduce3
Removing HDFS temp directory hdfs:///user/root/tmp/mrjob/MapReduce3.root.20230612.034054.034789...
Removing temp directory /tmp/MapReduce3.root.20230612.034054.034789...

```

	Replication	↕	Block Size	↕	Name	↕
	0		0 B		MapReduce1	
	0		0 B		MapReduce2	
	0		0 B		MapReduce3	

↑	Size	↕	Last Modified	↕	Replication	↕	Block Size	↕	Name	↕
	0 B		Jun 12 07:23		3		128 MB		_SUCCESS	
	171 B		Jun 12 07:23		3		128 MB		part-00000	

```

root@03bb6931900a:~# hdfs dfs -cat hdfs:///output/MapReduce3/part-00000
2023-06-12 03:54:17,000 INFO sasl.SaslDataTransferClient: SASL encryption trust check: local
false
"California" [0.002377717391304348, 0.018002717391304348, 0.010190217391304348, 2944]
"New York" [0.001188779362815027, 0.020446980504041846, 0.007370423204945316, 4206]

```

دلیل تفاوت در نتیجه فایل دوم و سوم :

ممکن است مختصات ها دقیق نباشند و برخی state های دیگر نیز مختصات مشابه داشته باشند. و یا قسمتی از یک state به دلیل مختصات نادقیق حذف شده باشد.

در صورتی که فرض بگیریم، دقیق نبودن مختصات باعث اختلاف بین نتایج نشده است، یک دلیل دیگر می تواند خالی بودن مقدار long و lat برای برخی از ردیف ها باشد. در حالی که ممکن است state را دارا باشند.

و برعکس این امر نیز امکان پذیر است یعنی state خالی باشد ولی longitude و latitude وجود داشته باشد.

• خروجی هریک از برنامه‌های MapReduce

- نمایش WebUI برای NameNode (localhost:9870) و فایل سیستم آن
(دیتاست و خروجی برنامه‌های MapReduce)

هر دو قسمت در بالا آورده شده‌اند .

علاوه بر کدها برخی دستوراتی که برای انجام تمرین استفاده شده است، نیز به پیوست ارسال شده است. (در فایل به نام desc.txt)