

MapReduce

1. Сначала надо настроить веб-морду Yarna:

В папке **docker-hadoop-master** лежит код контейнера. Там есть файл **docker-compose.yml** которым мы запускаем

Внутри файла **docker-compose.yml** найти **resource manager** и прописать порты:

ports:

- 8000:**8088** (8088 – не менять. Это порт 8088 на контейнере будет транслировать в указанный мной порт)

Убил все контейнеры:

docker kill \$(docker ps -q)

просмотрел, что их нет:

docker ps

```
max@max:~$ docker kill $(docker ps -q)
86dba0941d71
a7cc4668a8a6
08e6c8323525
2de332698ef7
54c3e6dc3092
max@max:~$ docker ps
CONTAINER ID   IMAGE          COMMAND                  CREATED        STATUS        PORTS          NAMES
max@max:~$
```

По новой поднимаю контейнеры:

Перешел в папку, где лежит образ контейнера:

cd Downloads/docker-hadoop-master

и там запустил

docker-compose up -d

```
max@max:~/Downloads$ cd docker-hadoop-master
max@max:~/Downloads/docker-hadoop-master$ docker-compose up -d
Starting historyserver    ... done
Starting nodemanager      ... done
Starting namenode         ... done
Starting datanode         ... done
Recreating resource manager ... done
max@max:~/Downloads/docker-hadoop-master$
```

Запускаю namenode

docker exec -it namenode bash

```
max@max:~/Downloads/docker-hadoop-master$ docker exec -it namenode bash
root@08e6c8323525:/#
```

2. Скачал файл .jar по ссылке <https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-mapreduce-examples/2.7.1/hadoop-mapreduce-examples-2.7.1.jar>

Закинул **hadoop-mapreduce-examples-2.7.1.jar** в контейнер:, запустил namenode

```
max@max:~/Downloads/docker-hadoop-master$ sudo docker cp "/home/max/Downloads/hadoop-mapreduce-examples-2.7.1.jar" namenode:/
max@max:~/Downloads/docker-hadoop-master$ docker exec -it namenode bash
root@08e6c8323525:/#
```

`docker cp "/home/max/Downloads/hadoop-mapreduce-examples-2.7.1.jar" namenode:/`

- Из файла [hadoop-mapreduce-examples-2.7.1.jar](#) запускаю на yarn процесс подсчета числа π (програмка внутри сборника [hadoop-mapreduce-examples-2.7.1.jar](#) лежит)

`yarn jar hadoop-mapreduce-examples-2.7.1.jar pi 1 10`

```
root@08e6c8323525:/# yarn jar hadoop-mapreduce-examples-2.7.1.jar pi 1 10
Number of Maps = 1
Samples per Map = 10
2021-12-09 11:27:48,099 INFO sasl.SaslDataTransferClient: SASL encryption trust ch
rusted = false
Wrote input for Map #0
Starting Job
2021-12-09 11:27:48 363 INFO client.RMProxy: Connecting to ResourceManager at reso
```

- В Браузере запустил

<http://localhost:8000/cluster/apps> и наблюдаю веб-мроду yarn

The screenshot shows the Hadoop web interface at `localhost:8000/cluster/apps`. The page title is "All Applications". On the left, there is a sidebar with a "Cluster" menu and a "Tools" button. The main content area displays various metrics and a table of applications.

Cluster Metrics											
Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCore	Unhealthy Nodes	Rebo	
1	0	1	0	0 B	16 GB						

Cluster Nodes Metrics											
Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebo						
1	0	0	0	0	0						

Scheduler 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>								

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Runn
application_1639052267798_0001	root	QuasiMonteCarlo	MAPREDUCE	default	0	Thu Dec 9 15:18:20 +0300 2021	Thu Dec 9 15:18:21 +0300 2021	Thu Dec 9 15:18:36 +0300 2021	FINISHED	SUCCEEDED	N/A
application_1639049586486_0001	root	QuasiMonteCarlo	MAPREDUCE	default	0	Thu Dec 9 14:33:10 +0300 2021	Thu Dec 9 14:33:19 +0300 2021	Thu Dec 9 14:33:33 +0300 2021	FINISHED	SUCCEEDED	N/A
application_1639048248022_0001	root	QuasiMonteCarlo	MAPREDUCE	default	0	Thu Dec 9 14:27:49 +0300 2021	Thu Dec 9 14:27:51 +0300 2021	Thu Dec 9 14:28:04 +0300 2021	FINISHED	SUCCEEDED	N/A

Showing 1 to 3 of 3 entries

This screenshot shows the same Hadoop web interface but with more detailed metrics and a larger table of applications.

Cluster Metrics													
Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCore					
5	0	1	4	2	10 GB	16 GB	0 B	2					

Cluster Nodes Metrics													
Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebo								
1	0	0	0	0	0								

Scheduler 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>										

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCo	Allocated Memory MB	Reserved CPU VCo	Reserved Memory MB
application_1639052267798_0005	root	QuasiMonteCarlo	MAPREDUCE	default	0	Thu Dec 9 17:01:36 +0300 2021	Thu Dec 9 17:01:36 +0300 2021	N/A	RUNNING	UNDEFINED	2	2	10240	0	0
application_1639052267798_0004	root	QuasiMonteCarlo	MAPREDUCE	default	0	Thu Dec 9 17:01:06 +0300 2021	Thu Dec 9 17:01:06 +0300 2021	Thu Dec 9 17:01:20 +0300 2021	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A