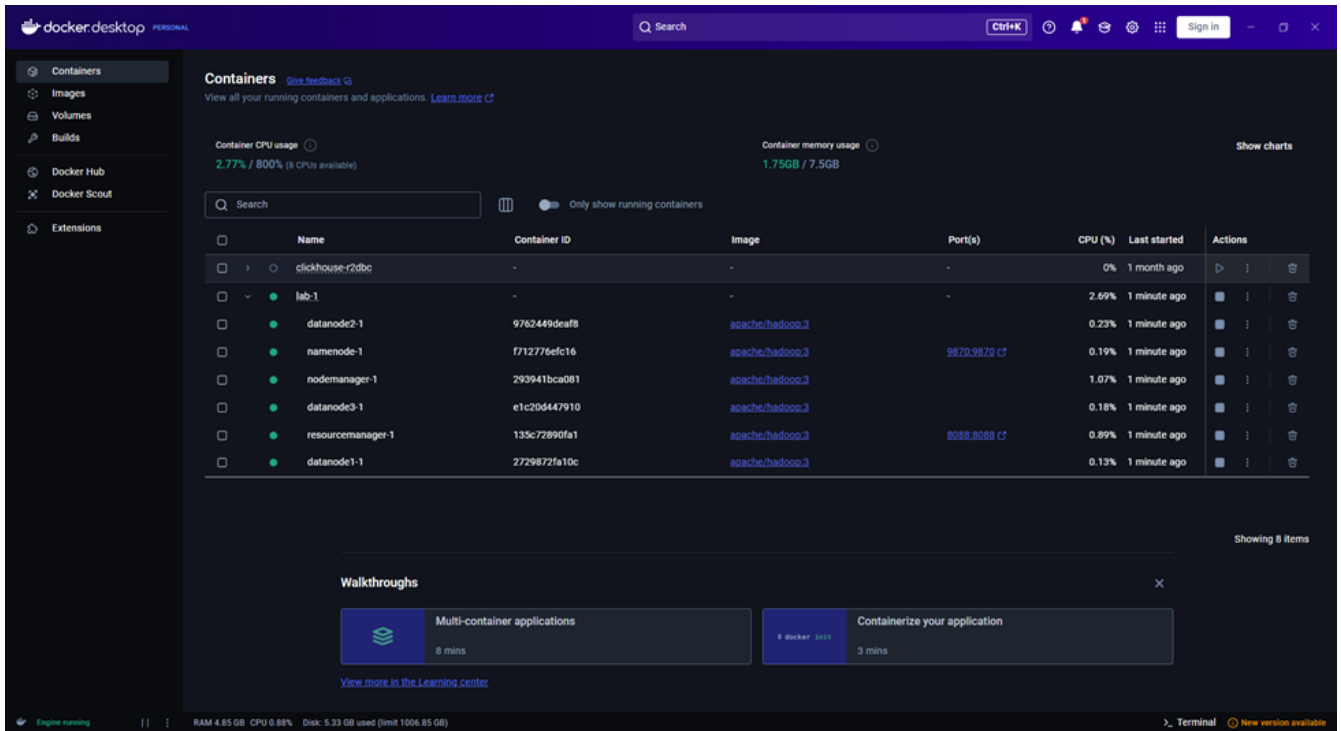
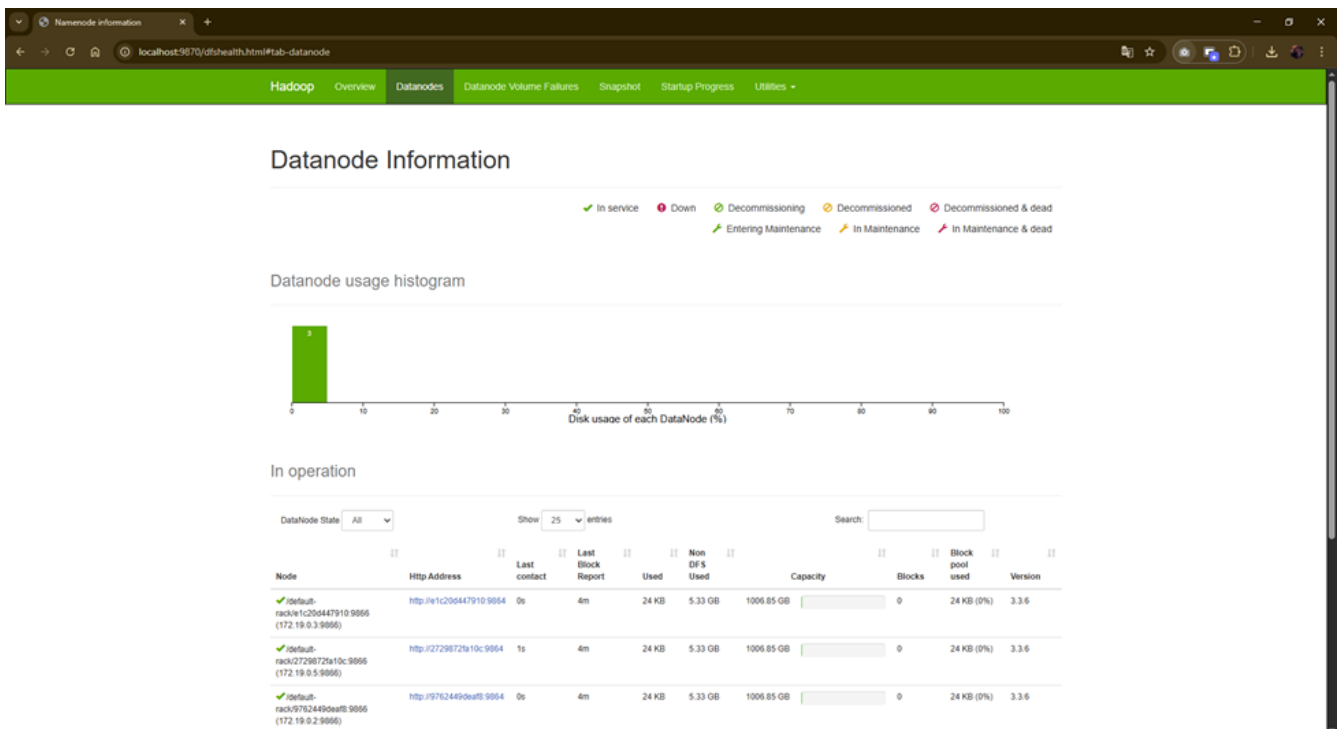


# HDFS

1) Разверните и настройте кластер Hadoop. Кластер должен включать один главный узел (namenode) и более двух дочерних узлов (datanode) Для развертывания можно выбрать один из нескольких путей: 1) использовать в качестве узлов реальные машины, объединенные в локальную сеть. 2) использовать в качестве узлов виртуальные машины, объединенные в локальную сеть 3) размещать узлы в Docker-контейнерах 4) иное



2) Проверьте работоспособность кластера. Для этого выведите в консоль информацию о его конфигурации и состоянии, а также отобразите веб-страницу с аналогичной информацией



```
hdfs dfsadmin -report
```

```
Windows PowerShell
bash-4.2$ hdfs dfsadmin -report
Configured Capacity: 3243303530496 (2.95 TB)
Present Capacity: 3061156810752 (2.78 TB)
DFS Remaining: 3061156724736 (2.78 TB)
DFS Used: 86016 (84 KB)
DFS Used%: 0.00%
Replicated Blocks:
  Under replicated blocks: 0
  Blocks with corrupt replicas: 0
  Missing blocks: 0
  Missing blocks (with replication factor 1): 0
  Low redundancy blocks with highest priority to recover: 0
  Pending deletion blocks: 0
Erasure Coded Block Groups:
  Low redundancy block groups: 0
  Block groups with corrupt internal blocks: 0
  Missing block groups: 0
  Low redundancy blocks with highest priority to recover: 0
  Pending deletion blocks: 0

-----
Live datanodes (3):

Name: 172.19.0.2:9866 (lab-1-datanode2-1.lab-1_default)
Hostname: 9762449deaf8
Decommission Status : Normal
Configured Capacity: 1081101176832 (1006.85 GB)
DFS Used: 28672 (28 KB)
Non DFS Used: 5723217920 (5.33 GB)
DFS Remaining: 1020385574912 (950.31 GB)
```

3) Изучите и выполните базовые операции над файлами и директориями в HDFS (используйте консольные команды):

3.1) Создайте текстовый файл и внесите в него произвольную текстовую информацию

### Создание файла

```
echo "Леонардо, Донателло, Микеланджело, Рафаэле" > my_hdfs_file.txt
```

```
cat my_hdfs_file.txt
```

```
Windows PowerShell
bash-4.2$ echo "Леонардо, Донателло, Микеланджело, Рафаэле" > my_hdfs_file.txt
bash-4.2$ cat my_hdfs_file.txt
Леонардо, Донателло, Микеланджело, Рафаэле
bash-4.2$
```

## Загрузка файла

```
hdfs dfs -mkdir -p /lab1
hdfs dfs -put my_hdfs_file.txt /lab1
hdfs dfs -ls /lab1
```

```
Windows PowerShell
bash-4.2$ hdfs dfs -mkdir -p /lab1
bash-4.2$ hdfs dfs -put my_hdfs_file.txt /lab1
bash-4.2$ hdfs dfs -ls /lab1
Found 1 items
-rw-r--r--  2 hadoop supergroup      79 2025-04-15 19:26 /lab1/my_hdfs_file.txt
bash-4.2$
```

## 3.2) Настройте фактор репликации для создаваемого файла

```
hdfs dfs -setrep -w 2 /lab1/my_hdfs_file.txt
```

```
hdfs fsck /lab1/my_hdfs_file.txt -files -blocks -locations
```

```
Windows PowerShell
bash-4.2$ hdfs fsck /lab1/my_hdfs_file.txt -files -blocks -locations
Connecting to namenode via http://namenode:9870/fsck?ugi=hadoop&files=1&blocks=1&locations=1&path=%2Flab1%2Fmy_hdfs_file.txt
FSCK started by hadoop (auth:SIMPLE) from /172.19.0.7 for path /lab1/my_hdfs_file.txt at Tue Apr 15 19:30:47 UTC 2025

[lab1/my_hdfs_file.txt 79 bytes, replicated: replication=2, 1 block(s): OK
0. BP-95005107-172.19.0.7-1744743520127:blk_1073741825-1001 len=79 Live_repl=2 [DatanodeInfoWithStorage[172.19.0.2:9866,DS-4efa29b0-dd4a-4630-94fa-0a767a7bdb54,DISK], DatanodeInfoWithStorage[172.19.0.3:9866,DS-c73e02f9-8bf9-4c20-9f27-9e7029890051,DISK]]

Status: HEALTHY
Number of data-nodes: 3
Number of racks: 1
Total dirs: 0
Total symlinks: 0

Replicated Blocks:
Total size: 79 B
Total files: 1
Total blocks (validated): 1 (avg. block size 79 B)
Minimally replicated blocks: 1 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 0 (0.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 2
Average block replication: 2.0
Missing blocks: 0
Corrupt blocks: 0
Missing replicas: 0 (0.0 %)
```

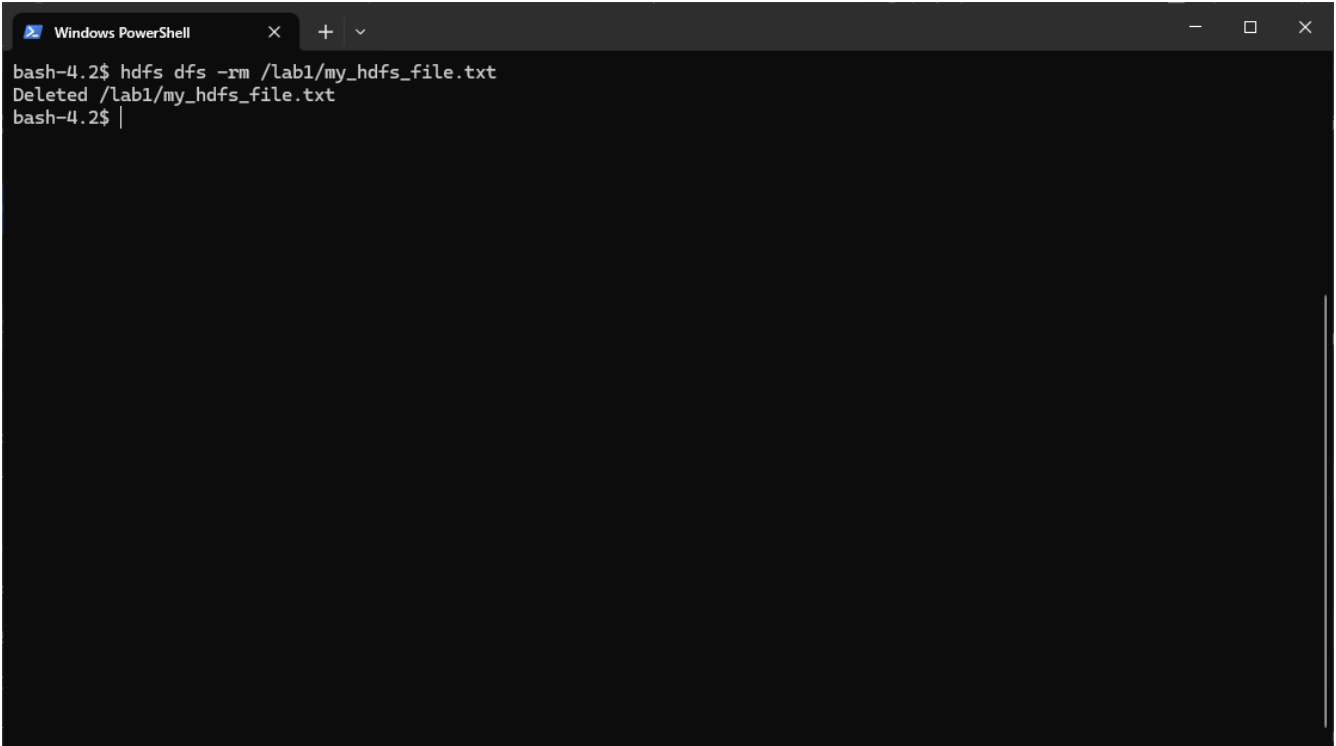
3.3) Отобразите изменения в каталогах локальной файловой системы (не HDFS), связанные с созданием нового файла

```
ls -a
```

```
Windows PowerShell
bash-4.2$ ls -a
.  LICENSE-binary  NOTICE-binary  README.txt  etc  lib  licenses-binary  sbin
.. LICENSE.txt    NOTICE.txt    bin         include  libexec  my_hdfs_file.txt  share
bash-4.2$
```

3.4) Удалите файл из HDFS

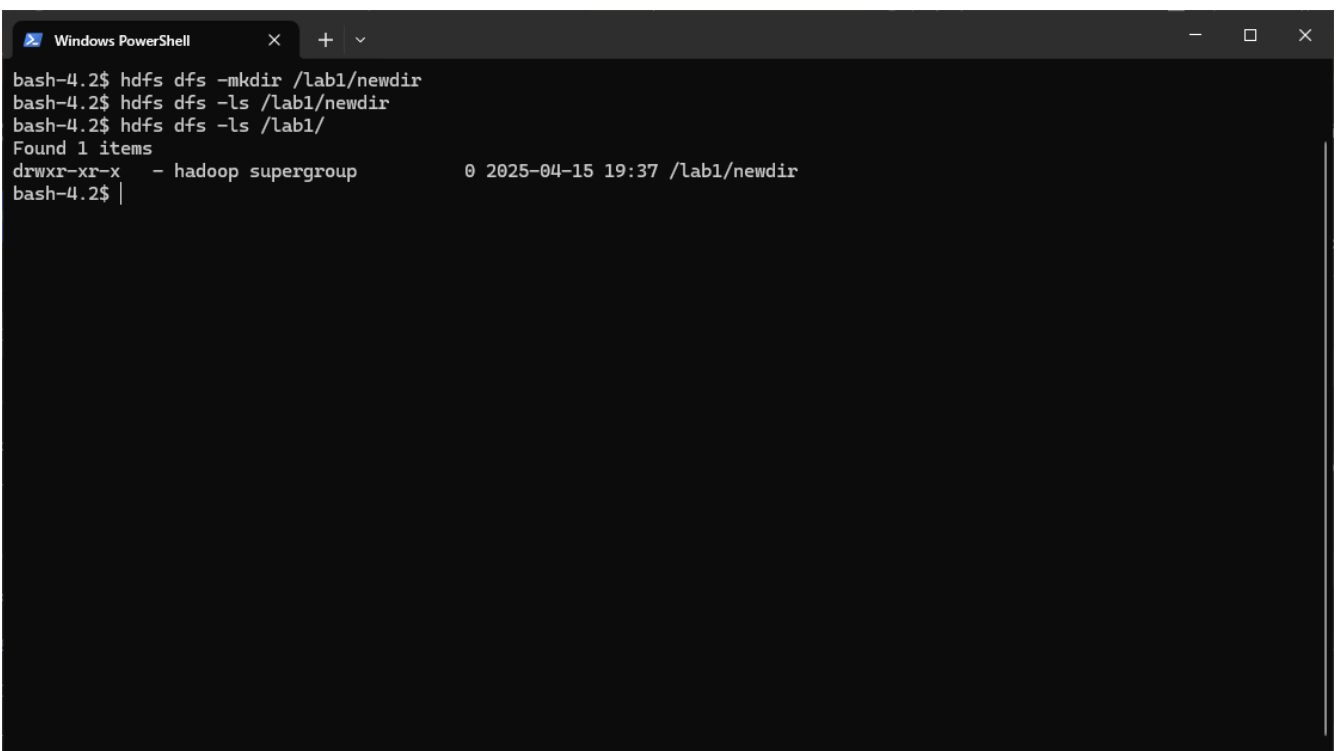
```
hdfs dfs -rm /lab1/my_hdfs_file.txt
```



```
Windows PowerShell
bash-4.2$ hdfs dfs -rm /lab1/my_hdfs_file.txt
Deleted /lab1/my_hdfs_file.txt
bash-4.2$
```

### 3.5) Создайте новую директорию

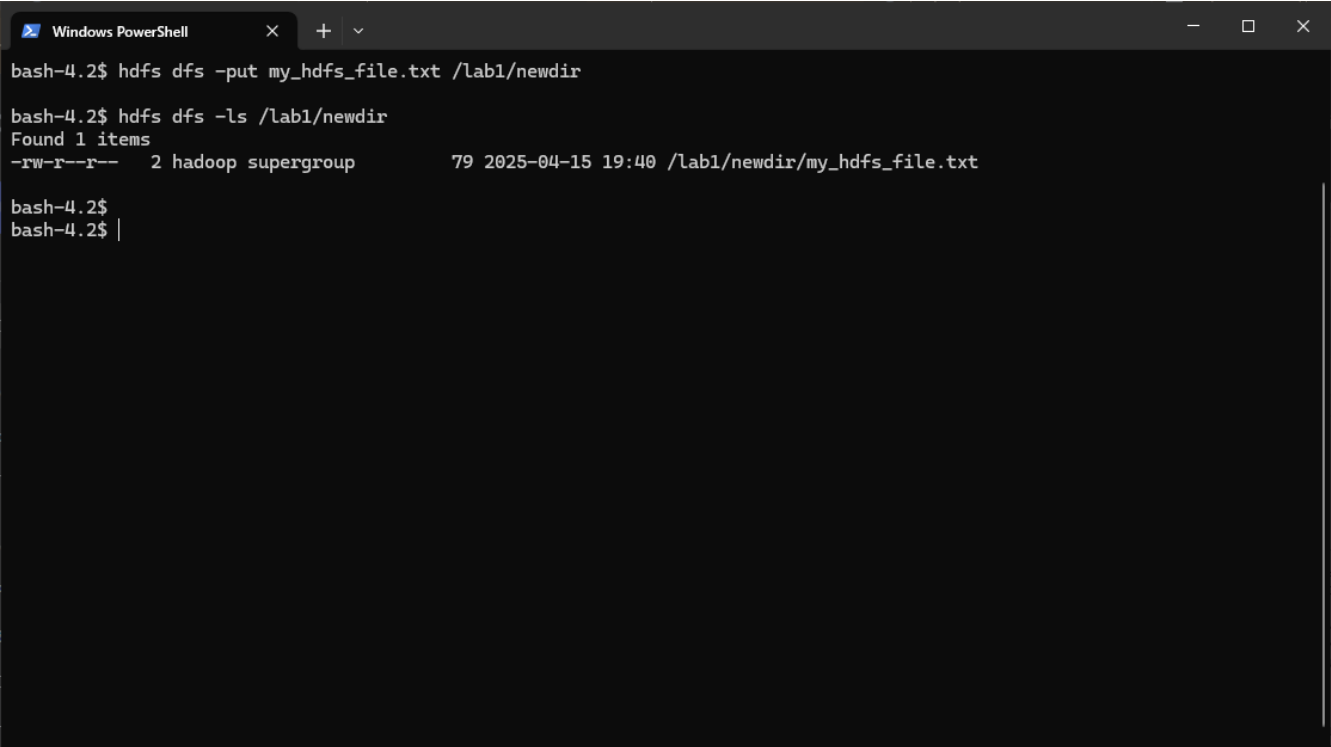
```
hdfs dfs -mkdir /lab1/newdir
hdfs dfs -ls /lab1
```



```
Windows PowerShell
bash-4.2$ hdfs dfs -mkdir /lab1/newdir
bash-4.2$ hdfs dfs -ls /lab1/newdir
bash-4.2$ hdfs dfs -ls /lab1/
Found 1 items
drwxr-xr-x  - hadoop supergroup          0 2025-04-15 19:37 /lab1/newdir
bash-4.2$
```

### 3.6) Переместите созданные ранее файлы в новую директорию

```
hdfs dfs -put my_hdfs_file.txt /lab1/newdir  
hdfs dfs -ls /lab1/newdir
```



The screenshot shows a Windows PowerShell terminal window with a dark background. The title bar at the top reads "Windows PowerShell" and includes standard window controls (minimize, maximize, close). The terminal content shows the following sequence of commands and output:

```
bash-4.2$ hdfs dfs -put my_hdfs_file.txt /lab1/newdir  
bash-4.2$ hdfs dfs -ls /lab1/newdir  
Found 1 items  
-rw-r--r--  2 hadoop supergroup      79 2025-04-15 19:40 /lab1/newdir/my_hdfs_file.txt  
bash-4.2$  
bash-4.2$ |
```