BigData Analytics

Dr. SOHAIL IMRAN

Hadoop on Docker

sohailimran2@yahoo.com

Ref: Web
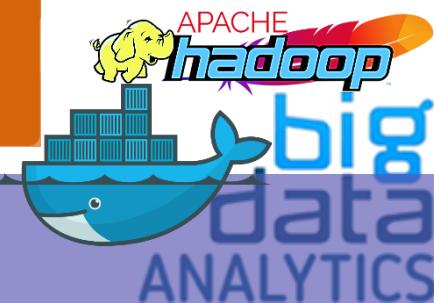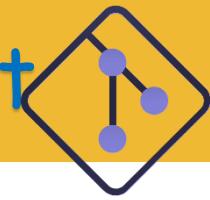
update

`sudo apt update`

```
sohailimran@DESKTOP-9VAOBA3:~$ sudo apt update
[sudo] password for sohailimran:
Hit:1 http://archive.ubuntu.com/ubuntu noble InRelease
Get:2 http://security.ubuntu.com/ubuntu noble-security InRelease [126 kB]
Get:3 http://archive.ubuntu.com/ubuntu noble-updates InRelease [126 kB]
Get:4 http://archive.ubuntu.com/ubuntu noble-backports InRelease [126 kB]
Get:5 http://archive.ubuntu.com/ubuntu noble-updates/main amd64 Packages [1585 kB]
Get:6 http://security.ubuntu.com/ubuntu noble-security/main amd64 Packages [1306 kB]
Get:7 http://archive.ubuntu.com/ubuntu noble-updates/main Translation-en [299 kB]
Get:8 http://archive.ubuntu.com/ubuntu noble-updates/main amd64 Components [175 kB]
Get:9 http://archive.ubuntu.com/ubuntu noble-updates/main amd64 c-n-f Metadata [15.7 kB]
Get:10 http://archive.ubuntu.com/ubuntu noble-updates/universe amd64 Packages [1498 kB]
Get:11 http://archive.ubuntu.com/ubuntu noble-updates/universe amd64 Components [378 kB]
Get:12 http://archive.ubuntu.com/ubuntu noble-updates/universe amd64 c-n-f Metadata [31.4 kB]
Get:13 http://archive.ubuntu.com/ubuntu noble-updates/restricted amd64 Packages [2244 kB]
Get:14 http://security.ubuntu.com/ubuntu noble-security/main Translation-en [215 kB]
Get:15 http://security.ubuntu.com/ubuntu noble-security/main amd64 Components [21.6 kB]
Get:16 http://security.ubuntu.com/ubuntu noble-security/main amd64 c-n-f Metadata [9336 B]
Get:17 http://security.ubuntu.com/ubuntu noble-security/universe amd64 Packages [906 kB]
Get:18 http://archive.ubuntu.com/ubuntu noble-updates/restricted Translation-en [509 kB]
Get:19 http://archive.ubuntu.com/ubuntu noble-updates/restricted amd64 Components [212 B]
Get:20 http://archive.ubuntu.com/ubuntu noble-updates/multiverse amd64 Components [940 B]
Get:21 http://archive.ubuntu.com/ubuntu noble-backports/main amd64 Components [7156 B]
Get:22 http://archive.ubuntu.com/ubuntu noble-backports/universe amd64 Components [11.0 kB]
Get:23 http://archive.ubuntu.com/ubuntu noble-backports/restricted amd64 Components [212 B]
Get:24 http://archive.ubuntu.com/ubuntu noble-backports/multiverse amd64 Components [212 B]
Get:25 http://security.ubuntu.com/ubuntu noble-security/universe amd64 Components [52.3 kB]
Get:26 http://security.ubuntu.com/ubuntu noble-security/universe amd64 c-n-f Metadata [19.4 kB]
Get:27 http://security.ubuntu.com/ubuntu noble-security/restricted amd64 Packages [2140 kB]
Get:28 http://security.ubuntu.com/ubuntu noble-security/restricted Translation-en [486 kB]
Get:29 http://security.ubuntu.com/ubuntu noble-security/restricted amd64 Components [208 B]
Get:30 http://security.ubuntu.com/ubuntu noble-security/multiverse amd64 Components [208 B]
Fetched 12.3 MB in 26s (478 kB/s)
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
64 packages can be upgraded. Run 'apt list --upgradable' to see them.
sohailimran@DESKTOP-9VAOBA3:~$
```

All

Best match

WSL
System

APACHE
hadoop

big
data
ANALYTICS
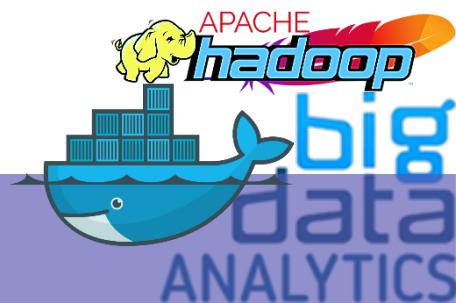
Dr.SOHAIL IMRAN

install git


```
curl --version
sudo apt install git
```

```
sohailimran@DESKTOP-9VAO  ×    +  ∨                                    —  □  ×

sohailimran@DESKTOP-9VAOBA3:~$ curl --version
curl 8.5.0 (x86_64-pc-linux-gnu) libcurl/8.5.0 OpenSSL/3.0.13 zlib/1.3 brotli/1.1.0 zstd/1.5.5 libidn2/2
.3.7 libpsl/0.21.2 (+libidn2/2.3.7) libssh/0.10.6/openssl/zlib nghttp2/1.59.0 librtmp/2.3 OpenLDAP/2.6.7
Release-Date: 2023-12-06, security patched: 8.5.0-2ubuntu10.6
Protocols: dict file ftp ftps gopher gophers http https imap imaps ldap ldaps mqtt pop3 pop3s rtmp rtsp
scp sftp smb smbs smtp smtps telnet tftp
Features: alt-svc AsynchDNS brotli GSS-API HSTS HTTP2 HTTPS-proxy IDN IPv6 Kerberos Largefile libz NTLM
PSL SPNEGO SSL threadsafe TLS-SRP UnixSockets zstd
sohailimran@DESKTOP-9VAOBA3:~$ sudo apt install git
[sudo] password for sohailimran:
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
git is already the newest version (1:2.43.0-1ubuntu7.3).
git set to manually installed.
0 upgraded, 0 newly installed, 0 to remove and 64 not upgraded.
sohailimran@DESKTOP-9VAOBA3:~$
```

Dr.SOHAIL IMRAN

2

git clone

`git clone https://github.com/big-data-europe/docker-hadoop`
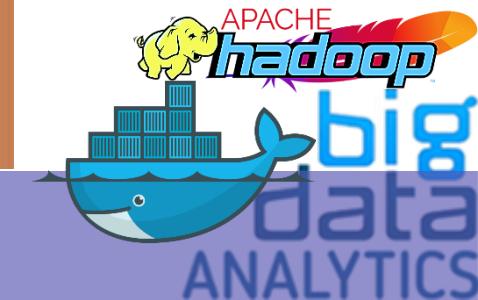
```
sohailimran@DESKTOP-9VAOBA3:~$ git clone https://github.com/big-data-europe/docker-hadoop
Cloning into 'docker-hadoop'...
remote: Enumerating objects: 539, done.
remote: Counting objects: 100% (189/189), done.
remote: Compressing objects: 100% (23/23), done.
remote: Total 539 (delta 169), reused 166 (delta 166), pack-reused 350 (from 1)
Receiving objects: 100% (539/539), 108.00 KiB | 654.00 KiB/s, done.
Resolving deltas: 100% (251/251), done.
sohailimran@DESKTOP-9VAOBA3:~$ ls
Dockerfile   app2.py   docker-hadoop
```

`cd docker-hadoop`

```
sohailimran@DESKTOP-9VAOBA3:~$ cd docker-hadoop
sohailimran@DESKTOP-9VAOBA3:~/docker-hadoop$ ls
Makefile    datanode                 hadoop.env      nginx            submit
README.md   docker-compose-v3.yml    historyserver   nodemanager
base        docker-compose.yml       namenode        resourcemanager
```

Dr. SOHAIL IMRAN

APACHE
hadoop

big
data
ANALYTICS

docker-compose.yml

YAML

```
sohailimran@DESKTOP-9VAOBA3:~/docker-hadoop$ ls
Makefile     base       docker-compose-v3.yml    hadoop.env      namenode    nodemanager       submit
README.md    datanode   docker-compose.yml       historyserver   nginx       resourcemanager
```
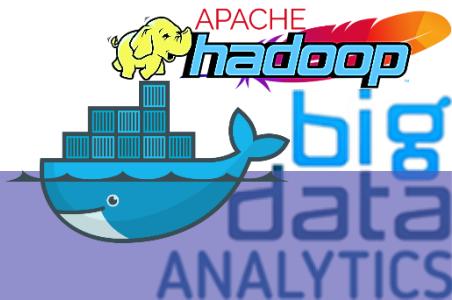
sudo nano docker-compose.yml

`sohailimran@DESKTOP-9VAOBA3:~/docker-hadoop$ sudo nano docker-compose.yml`

```
GNU nano 7.2                           docker-compose.yml *
version: "3"
services:
  namenode:
    image: bde2020/hadoop-namenode:2.0.0-hadoop3.2.1-java8
    container_name: namenode
    restart: always
    ports:
      - 9870:9870
      - 9000:9000
      - 9864:9864
    volumes:
      - hadoop_namenode:/hadoop/dfs/name
    environment:
      - CLUSTER_NAME=docker-hadoop
    env_file:
      - ./hadoop.env
```

```
datanode:
  image: bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8
  container_name: datanode
  ports:
    - 9865:9864
  restart: always
  volumes:
    - hadoop_datanode:/hadoop/dfs/data
  environment:
    SERVICE_PRECONDITION: "namenode:9870"
  env_file:
    - ./hadoop.env
```

1. Ctrl + O
2. Enter
3. Ctrl + X

APACHE
hadoop

big data ANALYTICS

Dr. SOHAIL IMRAN

4

compose

docker compose version

```
sohailimran@SI:~/docker-hadoop$ docker compose version
Docker Compose version v2.34.0
```
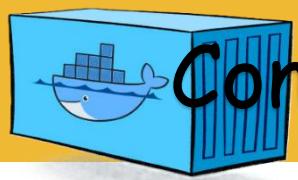
docker compose up -d

```
sohailimran@DESKTOP-9VAO    ×    +    ∨

sohailimran@DESKTOP-9VAOBA3:~/docker-hadoop$ docker compose up -d
WARN[0000] /home/sohailimran/docker-hadoop/docker-compose.yml: the attribute 'version' is
 obsolete, it will be ignored, please remove it to avoid potential confusion
[+] Running 23/28
 ✔datanode Pulled                                                       466.9s
 ✔namenode Pulled                                                       466.6s
 ✔resourcemanager Pulled                                                466.6s
 ✔nodemanager1 Pulled                                                   466.6s
 ✔historyserver Pulled                                                  466.6s
[+] Running 9/9
 ✔Network docker-hadoop_default              Cr...                        0.1s
 ✔Volume docker-hadoop_hadoop_historyserver  Created                      0.0s
 ✔Volume docker-hadoop_hadoop_namenode       Created                      0.0s
 ✔Volume docker-hadoop_hadoop_datanode       Created                      0.0s
 ✔Container datanode                          Started                      1.2s
 ✔Container historyserver                     Started                      1.2s
 ✔Container nodemanager                       Started                      1.3s
 ✔Container resourcemanager                   Starte...                    1.2s
 ✔Container namenode                          Started                      1.3s
sohailimran@DESKTOP-9VAOBA3:~/docker-hadoop$
```

Dr. SOHAIL IMRAN

5

APACHE
hadoop

big
data
ANALYTICS
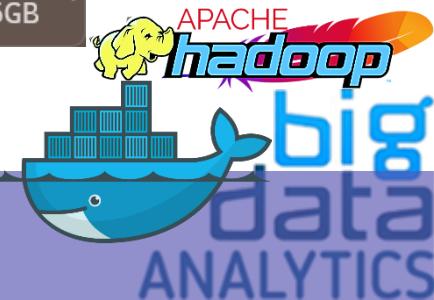
Containers & Images

docker ps

```
sohailimran@DESKTOP-9VAOBA3:~/docker-hadoop$ docker ps
CONTAINER ID   IMAGE                                              COMMAND                CREATED          STATUS
                 PORTS                                                                                     NAMES
c01aabcb58dc   bde2020/hadoop-historyserver:2.0.0-hadoop3.2.1-java8   "/entrypoint.sh /run…"   27 minutes ago   Up 27 min
utes (healthy)   8188/tcp                                                                                  historyserver
7a76d0a80800   bde2020/hadoop-nodemanager:2.0.0-hadoop3.2.1-java8    "/entrypoint.sh /run…"   27 minutes ago   Up 27 min
utes (healthy)   8042/tcp                                                                                  nodemanager
da398b19589c   bde2020/hadoop-namenode:2.0.0-hadoop3.2.1-java8      "/entrypoint.sh /run…"   27 minutes ago   Up 27 min
utes (healthy)    0.0.0.0:9000->9000/tcp, [::]:9000->9000/tcp, 0.0.0.0:9870->9870/tcp, [::]:9870->9870/tcp   namenode
2aac6d26ea24   bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8      "/entrypoint.sh /run…"   27 minutes ago   Up 27 min
utes (healthy)   9864/tcp                                                                                  datanode
5599790f93c2   bde2020/hadoop-resourcemanager:2.0.0-hadoop3.2.1-java8  "/entrypoint.sh /run…"   27 minutes ago   Up 27 min
utes (healthy)   8088/tcp                                                                                  resourcemanager
```

docker images

```
sohailimran@DESKTOP-9VAOBA3:~/docker-hadoop$ docker images
REPOSITORY                        TAG                       IMAGE ID       CREATED         SIZE
sohailimran/py-hello              py2                       9b05a16b76d2   7 days ago      183MB
mysql                             5.7                       4bc6bc963e6d   23 months ago   689MB
bde2020/hadoop-nodemanager        2.0.0-hadoop3.2.1-java8   b5515b537a30   5 years ago     2.05GB
bde2020/hadoop-resourcemanager    2.0.0-hadoop3.2.1-java8   91ca2afe8616   5 years ago     2.05GB
bde2020/hadoop-namenode           2.0.0-hadoop3.2.1-java8   51ad9293ec52   5 years ago     2.05GB
bde2020/hadoop-historyserver      2.0.0-hadoop3.2.1-java8   ab2971d5f8c3   5 years ago     2.05GB
bde2020/hadoop-datanode           2.0.0-hadoop3.2.1-java8   ddf6e9ad55af   5 years ago     2.05GB
```

Dr. SOHAIL IMRAN

6

APACHE
hadoop
big data
ANALYTICS

localhost:9870

localhost

localhost:9870/dfshealth.html#tab-overview

Hadoop

Overview | Datanodes | Datanode Volume Failures | Snapshot | Startup Progress | Utilities ▾

Overview 'namenode:9000' (active)
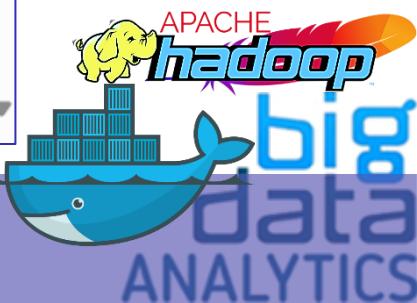
Started:	Thu Nov 27 00:10:30 +0500 2025
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Tue Sep 10 20:56:00 +0500 2019 by rohithsharmaks from branch-3.2.1
Cluster ID:	CID-f990dd03-1b7b-43b2-9690-44bfb37b5719
Block Pool ID:	BP-549882162-172.18.0.6-1764163124836
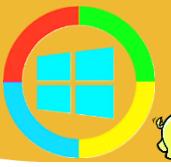
Summary

Security is off.

Safemode is off.

16 files and directories, 7 blocks (7 replicated blocks, 0 erasure coded block groups) = 23 total filesystem object(s).

dfs status

Heap Memory used 144.61 MB of 312 MB Heap Memory. Max Heap Memory is 1.7 GB.

Non Heap Memory used 53.65 MB of 55 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	1006.85 GB
Configured Remote Capacity:	0 B
DFS Used:	92 KB (0%)
Non DFS Used:	4.25 GB
DFS Remaining:	951.39 GB (94.49%)
Block Pool Used:	92 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (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	7
Number of Blocks Pending Deletion (including replicas)	0

Dr. SOHAIL IMRAN

```
docker inspect -f '{{ .Config.Hostname }}' namenode
```

```
sohailimran@DESKTOP-9VAOBA3:~$ docker inspect -f '{{ .Config.Hostname }}' namenode
e89ff71ebea5
```

```
docker inspect namenode --format='{{json .NetworkSettings.Networks}}'
```

```
sohailimran@DESKTOP-9VAOBA3:~$ docker inspect namenode --format='{{json .NetworkSettings.Networks}}'
{"docker-hadoop_default":{"IPAMConfig":null,"Links":null,"Aliases":["namenode","namenode"],"MacAddress":"6a:2b:f4:60:2a:fb
","DriverOpts":null,"GwPriority":0,"NetworkID":"21e7d0f80f5938110c3ffe1c8eae9c9bd5c5cdd794717c8f722271afe72dfb6e","Endpoin
tID":"f4399829155c8264773635bbe380080839ba3f1544bd16ecb44d8ae60a6cfcb4","Gateway":"172.18.0.1","IPAddress":"172.18.0.2","I
PPrefixLen":16,"IPv6Gateway":"","GlobalIPv6Address":"","GlobalIPv6PrefixLen":0,"DNSNames":["namenode","e89ff71ebea5"]}}
```
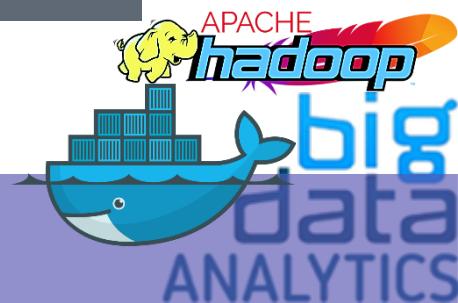
```
docker inspect -f '{{range .NetworkSettings.Networks}}{{.IPAddress}}{{end}}' namenode
```

```
sohailimran@DESKTOP-9VAOBA3:~$ docker inspect -f '{{range .NetworkSettings.Networks}}{{.IPAddress}}{{end}}' namenode
172.18.0.2
```

```
docker ps --format "table {{.Names}}\t{{.Ports}}"
```

```
sohailimran@DESKTOP-9VAOBA3:~/docker-hadoop$ docker ps --format "table {{.Names}}\t{{.Ports}}"
NAMES              PORTS
historyserver      8188/tcp
nodemanager        8042/tcp
resourcemanager    8088/tcp
namenode           0.0.0.0:9000->9000/tcp, [::]:9000->9000/tcp, 0.0.0.0:9864->9864/tcp, [::]:9864->9864/tcp,
 0.0.0.0:9870->9870/tcp, [::]:9870->9870/tcp
datanode           0.0.0.0:9865->9864/tcp, [::]:9865->9864/tcp
```

- namenode : ports 9000, 9864, 9870 exposed
- datanode : port 9865 (host) → 9864 (container) exposed
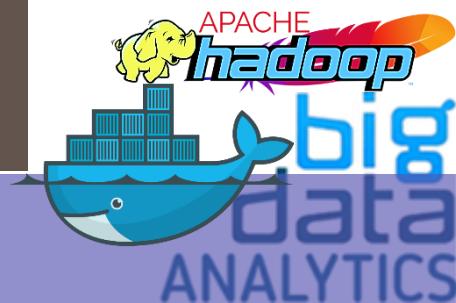
dfsadmin -report

`docker exec -it namenode hdfs dfsadmin -report`

```
sohailimran@DESKTOP-9VAOBA3:~/docker-hadoop$ docker exec -it namenode hdfs dfsadmin -report
Configured Capacity: 1081101176832 (1006.85 GB)
Present Capacity: 1021521588224 (951.37 GB)
DFS Remaining: 1021519667200 (951.36 GB)
DFS Used: 1921024 (1.83 MB)
DFS Used%: 0.00%
Replicated Blocks:
        Under replicated blocks: 48
        Blocks with corrupt replicas: 0
        Missing blocks: 0
        Missing blocks (with replication factor 1): 0
        Low redundancy blocks with highest priority to recover: 48
        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 (1):

Name: 172.18.0.6:9866 (datanode.docker-hadoop_default)
Hostname: 0dd4b4a533d5
Decommission Status : Normal
Configured Capacity: 1081101176832 (1006.85 GB)
DFS Used: 1921024 (1.83 MB)
Non DFS Used: 4587233280 (4.27 GB)
DFS Remaining: 1021519667200 (951.36 GB)
DFS Used%: 0.00%
DFS Remaining%: 94.49%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 1
Last contact: Sat Nov 29 05:56:25 UTC 2025
Last Block Report: Sat Nov 29 04:16:58 UTC 2025
Num of Blocks: 48
```

Dr. SOHAIL IMRAN

11

APACHE
hadoop

big
data
ANALYTICS

cp to docker

```
docker exec namenode ls /
```
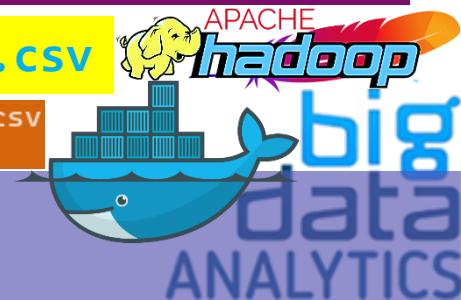
```
sohailimran@DESKTOP-9VAOBA3:~$ docker exec namenode ls /
KEYS
bin
boot
dev
entrypoint.sh
etc
hadoop
hadoop-data
home
lib
lib64
media
mnt
opt
proc
root
```

```
docker cp /mnt/d/DataSets/pdVsPySp.txt namenode:/home/pdVsPySp.txt
```

```
sohailimran@DESKTOP-9VAOBA3:~$ docker cp /mnt/d/DataSets/pdVsPySp.txt namenode:/home/pdVsPySp.txt
Successfully copied 1.54kB to namenode:/home/pdVsPySp.txt
```

```
docker cp /mnt/d/DataSets/salgrade.csv namenode:/home/salgrade.csv
```

```
sohailimran@DESKTOP-9VAOBA3:~$ docker cp /mnt/d/DataSets/salgrade.csv namenode:/home/salgrade.csv
Successfully copied 2.05kB to namenode:/home/salgrade.csv
```

docker to hdfs HDFS

docker exec -it namenode ls -lh /home/salgrade.csv

```
sohailimran@DESKTOP-9VAOBA3:~$ docker exec -it namenode ls -lh /home/salgrade.csv
-rwxrwxrwx 1 1000 1000 81 Jan 21  2024 /home/salgrade.csv
```

**docker exec -it namenode hdfs dfs -mkdir /user**
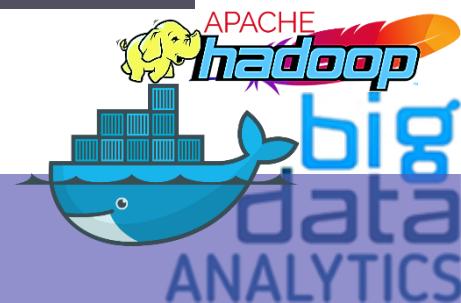
```
sohailimran@DESKTOP-9VAOBA3:~$ docker exec -it namenode hdfs dfs -mkdir /user
sohailimran@DESKTOP-9VAOBA3:~$ docker exec -it namenode hdfs dfs -ls /
Found 2 items
drwxr-xr-x   - root supergroup          0 2025-11-30 07:35 /rmstate
drwxr-xr-x   - root supergroup          0 2025-11-30 17:39 /user
```

docker exec -it namenode hdfs dfs -put /home/salgrade.csv /user/

```
sohailimran@DESKTOP-9VAOBA3:~$ docker exec -it namenode hdfs dfs -put /home/salgrade.csv /user/
2025-11-30 17:45:42,164 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted
= false, remoteHostTrusted = false
```
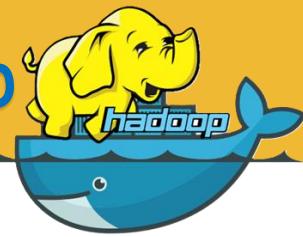
docker exec -it namenode hdfs dfs -ls /user/

```
sohailimran@DESKTOP-9VAOBA3:~$ docker exec -it namenode hdfs dfs -ls /user/
Found 1 items
-rw-r--r--   3 root supergroup         81 2025-11-30 17:45 /user/salgrade.csv
```

APACHE hadoop
big data
ANALYTICS

Dr.SOHAIL IMRAN

docker context use desktop-linux

```
PS C:\Users\DELL> docker context use desktop-linux
desktop-linux
Current context is now "desktop-linux"
```

docker exec namenode env | findstr HADOOP_HOME

```
PS C:\Users\DELL> docker exec namenode env | findstr HADOOP_HOME
HADOOP_HOME=/opt/hadoop-3.2.1
```
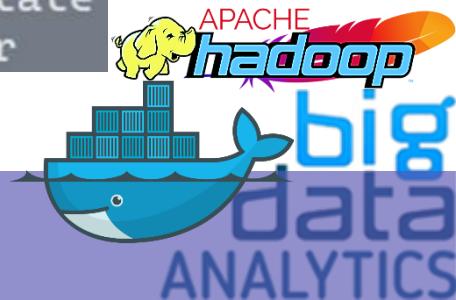
docker exec -it namenode bash

```
PS C:\Users\DELL> docker exec -it namenode bash
```

hdfs dfs -ls /

```
root@e89ff71ebea5:/# hdfs dfs -ls /
Found 2 items
drwxr-xr-x   - root supergroup          0 2025-11-26 13:18 /rmstate
drwxr-xr-x   - root supergroup          0 2025-11-26 13:25 /user
```

put to hdfs

```
hdfs dfs -mkdir /input
hdfs dfs -mkdir /output
```

```
root@e89ff71ebea5:/# hdfs dfs -mkdir /input
root@e89ff71ebea5:/# hdfs dfs -mkdir /output
```

```
hdfs dfs -ls /
```

```
root@e89ff71ebea5:/# hdfs dfs -ls /
Found 4 items
drwxr-xr-x    - root       supergroup          0 2025-11-27 13:09 /input
drwxr-xr-x    - root       supergroup          0 2025-11-27 13:09 /output
drwxr-xr-x    - root       supergroup          0 2025-11-26 13:18 /rmstate
drwxrwxrwx    - drillbits  supergroup          0 2025-11-26 20:47 /user
```
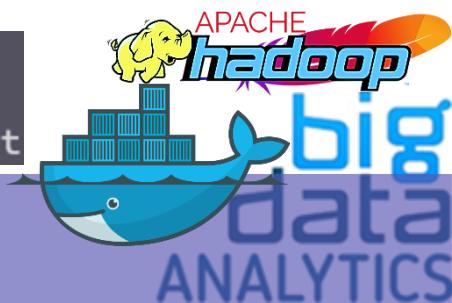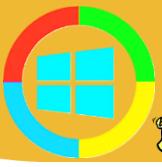
```
hdfs dfs -put /home/pdVsPySp.txt /input
```

```
root@e89ff71ebea5:/# hdfs dfs -put /home/pdVsPySp.txt /input
2025-11-28 20:31:14,530 INFO sasl.SaslDataTransferClient: SASL encryption trust
 check: localHostTrusted = false, remoteHostTrusted = false
```

```
hdfs dfs -ls /input
```

```
root@e89ff71ebea5:/# hdfs dfs -ls /input
Found 1 items
-rw-r--r--    3 root supergroup          0 2025-11-28 07:59 /input/pdVsPySp.txt
```

Dr. SOHAIL IMRAN
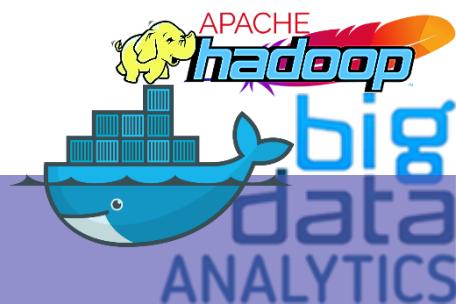
mapreduce

```
hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-
examples-3.2.1.jar wordcount /input /output/wordcount
```

```
root@e89ff71ebea5:/# hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.1.jar wordcount /input /output/wordcount
2025-11-28 20:37:18,644 INFO client.RMProxy: Connecting to ResourceManager at resourcemanager/172.18.0.2:8032
2025-11-28 20:37:18,971 INFO client.AHSProxy: Connecting to Application History server at historyserver/172.18.0.4:10200
2025-11-28 20:37:19,163 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_176435
9819985_0002
2025-11-28 20:37:19,269 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-11-28 20:37:19,384 INFO input.FileInputFormat: Total input files to process : 1
2025-11-28 20:37:19,430 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-11-28 20:37:19,945 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-11-28 20:37:19,957 INFO mapreduce.JobSubmitter: number of splits:1
2025-11-28 20:37:20,240 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2025-11-28 20:37:20,254 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1764359819985_0002
2025-11-28 20:37:20,254 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-11-28 20:37:20,418 INFO conf.Configuration: resource-types.xml not found
2025-11-28 20:37:20,419 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-11-28 20:37:20,764 INFO impl.YarnClientImpl: Submitted application application_1764359819985_0002
2025-11-28 20:37:20,822 INFO mapreduce.Job: The url to track the job: http://resourcemanager:8088/proxy/application_1764359819985_0002/
2025-11-28 20:37:20,822 INFO mapreduce.Job: Running job: job_1764359819985_0002
2025-11-28 20:37:26,927 INFO mapreduce.Job: Job job_1764359819985_0002 running in uber mode : false
2025-11-28 20:37:26,929 INFO mapreduce.Job:  map 0% reduce 0%
2025-11-28 20:37:31,991 INFO mapreduce.Job:  map 100% reduce 0%
2025-11-28 20:37:36,016 INFO mapreduce.Job:  map 100% reduce 100%
2025-11-28 20:37:36,025 INFO mapreduce.Job: Job job_1764359819985_0002 completed successfully
2025-11-28 20:37:36,084 INFO mapreduce.Job: Counters: 54
```

```
        Peak Reduce Virtual memory (bytes)=8453066752
    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=2665
    File Output Format Counters
        Bytes Written=1755
```

Dr.SOHAIL IMRAN

hdfs dfs -ls /output

```
root@e89ff71ebea5:/# hdfs dfs -ls /output
Found 1 items
drwxr-xr-x   - root supergroup          0 2025-11-28 10:32 /output/wordcount
```

hdfs dfs -get /output/wordcount /home/wordcount

```
root@e89ff71ebea5:/# hdfs dfs -get /output/wordcount /home/wordcount
2025-11-28 20:56:09,673 INFO sasl.SaslDataTransferClient: SASL encryption
 trust check: localHostTrusted = false, remoteHostTrusted = false
```

ls /home

```
sohailimran@DESKTOP-9VAOBA3:~$ ls /home
salgrade.csv   sohailimran
```

docker cp namenode:/home/wordcount /mnt/d/DataSets/wordcount

```
sohailimran@DESKTOP-9VAOBA3:~$ docker cp namenode:/home/wordcount /mnt/d/DataSets/wordcount
Successfully copied 2.56kB to /mnt/d/DataSets/wordcount
```

> Data (D:)

DataSets

wordcount

archive (24).zip

_SUCCESS

part-r-00000

part-r-00000

File   Edit   View   H1 ⌄

```
(HDFS),    1
(JVM),     1
(RDD)      1
(e.g.      1
Amazon  1
Apache  1
Complexity        1
```

Dr. SOHAIL IMRAN