

HIVE INSTALLATION :

Docker Installation :

Install [Docker Desktop](#) . Drag into applications. Open the docker

Create a repository with folders,files shown below

Name	Date Modified	Size	Kind
> dags	15-Sep-2022 at 11:39 PM	--	Folder
docker-compose.yaml	Yesterday at 6:27 PM	10 KB	YAML
> logs	Yesterday at 3:55 PM	--	Folder
> plugins	14-Sep-2022 at 9:41 AM	--	Folder

In terminal navigate to the above repository and type command
edit docker-compose.yaml with containers required

```
...
services:
  frontend:
    image: awesome/webapp
    ports:
      - "443:8043"
    networks:
      - front-tier
      - back-tier
    configs:
      - httpd-config
    secrets:
      - server-certificate

  backend:
    image: awesome/database
    volumes:
      - db-data:/etc/data
    networks:
      - back-tier

volumes:
  db-data:
    driver: flocker
    driver_opts:
      size: "10GiB"
```

```

configs:
  httpd-config:
    external: true

secrets:
  server-certificate:
    external: true

networks:
  # The presence of these objects is sufficient to define them
  front-tier: {}
  back-tier: {}

```

...

Starting Docker and Pulling Cloudera image:

(Type below command in terminal)

<code>docker-compose up -d</code>	— Starts the container
<code>docker pull cloudera/quickstart:latest</code>	— Pulls Cloudera image where hive is inbuilt
<code>docker images</code>	— Check if cloudera image is successfully pulled

Starting Hadoop environment :

```

docker run --hostname=quickstart.cloudera --privileged=true -t -i -v [LocalPathAbs]:[ImagePath]
[IMAGEID_First 6 Characters] /usr/bin/docker-quickstart

```

P.S : Before running above command make sure you empty containers

```
MockProject -- @quickstart:/ -- com.docker.cli - docker run --hostname=quickstart.cloudera --privileged=true -t -i -p 8080:50070 -p 8081:50075 -p 8020:8020 -p 9000:...
rohithya@Sigmoids-MacBook-Air-4 MockProject % docker run --hostname=quickstart.cloudera --privileged=true -t -i -p 8080:50070 -p 8081:50075 -p 8020:8020 -p 9000:9000 --volumes-from=MockProject/Amazon-Mock-Project/datasets:/Storage 4239cd /usr/bin/docker-quickstart
WARNING: The requested image's platform (linux/amd64) does not match the detected host platform (linux/arm64/v8) and no specific platform was requested
Starting mysqld: [ OK ]

if [ "$1" == "start" ]; then
  if [ "${EC2}" == 'true' ]; then
    FIRST_BOOT_FLAG=/var/lib/cloudera-quickstart/.ec2-key-installed
    if [ ! -f "${FIRST_BOOT_FLAG}" ]; then
      METADATA_API=http://169.254.169.254/latest/meta-data
      KEY_URL=${METADATA_API}/public-keys/0/openssh-key
      SSH_DIR=/home/cloudera/.ssh
      mkdir -p ${SSH_DIR}
      chown cloudera:cloudera ${SSH_DIR}
      curl ${KEY_URL} >> ${SSH_DIR}/authorized_keys
      touch ${FIRST_BOOT_FLAG}
    fi
  fi
  if [ "${DOCKER}" != 'true' ]; then
    if [ -f /sys/kernel/mm/redhat_transparent_hugepage/defrag ]; then
      echo never > /sys/kernel/mm/redhat_transparent_hugepage/defrag
    fi
    cloudera-quickstart-ip
    HOSTNAME=quickstart.cloudera
    hostname ${HOSTNAME}
    sed -i -e "s/HOSTNAME=.*HOSTNAME=${HOSTNAME}/" /etc/sysconfig/network
  fi
  (
    cd /var/lib/cloudera-quickstart/tutorial;
    nohup python -m SimpleHTTPServer 80 &
  )
  # TODO: check for expired CM license and update config.js accordingly
fi
+ '[' start == start ']'
+ '[' '' == true ']'
+ '[' true != true ']'
+ cd /var/lib/cloudera-quickstart/tutorial
+ nohup python -m SimpleHTTPServer 80
nohup: appending output to 'nohup.out'
JMX enabled by default
```

(This starts the hadoop environment. Browser Namenode → localhost:8080)

localhost:8080/dfshealth.html#tab-overview

Inbox (8) - saker...udemyInbox (4,034) - sa...darwin boxgreyHRDE Freshers Induc...Sake Rohithya - L...tddLab - Google DriveMongoDB Tutorial...DAT board - Agile...

HadoopOverviewDatanodesDatanode Volume FailuresSnapshotStartup ProgressUtilities

Overview 'quickstart.cloudera:8020' (active)

Started:	Thu Nov 03 11:23:01 +0530 2022
Version:	2.6.0-cdh5.7.0, rc00978c67b0d3fe9f3b896b5030741bd40bf541a
Compiled:	Thu Mar 24 00:06:00 +0530 2016 by jenkins from Unknown
Cluster ID:	CID-11ef0663-e698-48f8-bbee-7b664322ae19
Block Pool ID:	BP-1120155954-10.0.0.1-1459909528739

Summary

Security is off.

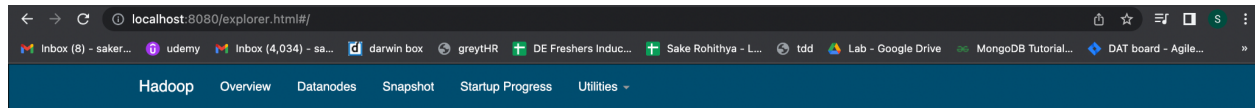
Safemode is off.

999 files and directories, 915 blocks = 1,914 total filesystem object(s).

Heap Memory used 111.59 MB of 204.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 41.81 MB of 42.44 MB Committed Non Heap Memory. Max Non Heap Memory is 130 MB.

Configured Capacity:	58.37 GB
DFS Used:	829.38 MB (1.39%)



Browse Directory

/								Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
drwxrwxrwx	hdfs	supergroup	0 B	Wed Apr 06 07:56:30 +0530 2016	0	0 B	benchmarks	
drwxr-xr-x	hbase	supergroup	0 B	Thu Nov 03 11:25:24 +0530 2022	0	0 B	hbase	
-rw-r--r--	root	supergroup	9.83 MB	Thu Nov 03 11:43:35 +0530 2022	1	128 MB	project	
drwxr-xr-x	root	supergroup	0 B	Thu Nov 03 11:48:08 +0530 2022	0	0 B	rohithya	
drwxrwxrwt	hdfs	supergroup	0 B	Thu Nov 03 11:25:50 +0530 2022	0	0 B	tmp	
drwxr-xr-x	hdfs	supergroup	0 B	Wed Apr 06 07:57:53 +0530 2016	0	0 B	user	
drwxr-xr-x	hdfs	supergroup	0 B	Wed Apr 06 07:57:47 +0530 2016	0	0 B	var	
Hadoop, 2014.								

Making directory in HDFS and copying data from local storage to HDFS dir

```
[root@quickstart ~]# hadoop fs -mkdir /rohithya
```

```
[root@quickstart ~]# hadoop fs -copyFromLocal Storage/data.csv /rohithya
```

Start the hive environment

```
[root@quickstart ~]# hive
```

Table creation and uploading data into table

```
hive> CREATE EXTERNAL TABLE ecomm_t (id int,order_status string,order_products_value double,order_freight_value double,order_items_qty double,customer_city string,customer_state string,customer_zip_code_prefix int,product_name_length int,product_description_length int,product_photos_qty int,review_score double,order_purchase_timestamp timestamp,order_aproved_at timestamp,order_delivered_customer_date timestamp) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/rohithya';
```

Querying the table to check if data is uploaded successfully or not :

```
MockProject -- @quickstart:/ -- com.docker.cli - docker run --hostname=quickstart.cloudera --privileged=true -t -i -p 8080:50070 -p 8081:50075 -p 8020:8020 -p 9000:...
```

```
hive> desc ecomm_t;
OK
id                int
order_status      string
order_products_value double
order_freight_value double
order_items_qty   double
customer_city     string
customer_state    string
customer_zip_code_prefix int
product_name_length int
product_description_length int
product_photos_qty int
review_score      double
order_purchase_timestamp timestamp
order_approved_at timestamp
order_delivered_customer_date timestamp
Time taken: 0.286 seconds, Fetched: 15 row(s)
hive> select * from ecomm_t limit 5;
OK
1    delivered    79.0    17.8    1.0    Luziania    GO    728    50    201    2    5.0    NULL    NULL    NULL
2    delivered    119.9    27.16    1.0    Joinville    SC    892    50    511    3    5.0    NULL    NULL    NULL
3    delivered    519.99    41.69    1.0    Serra ES    291    48    1156    2    1.0    NULL    NULL    NULL
4    delivered    29.5    17.92    1.0    RIO DE JANEIRO RJ    222    21    207    2    4.0    NULL    NULL    NULL
5    delivered    26.77    23.11    1.0    Sao Paulo    SP    40    41    451    1    5.0    NULL    NULL    NULL
Time taken: 0.192 seconds, Fetched: 5 row(s)
hive>
```

