

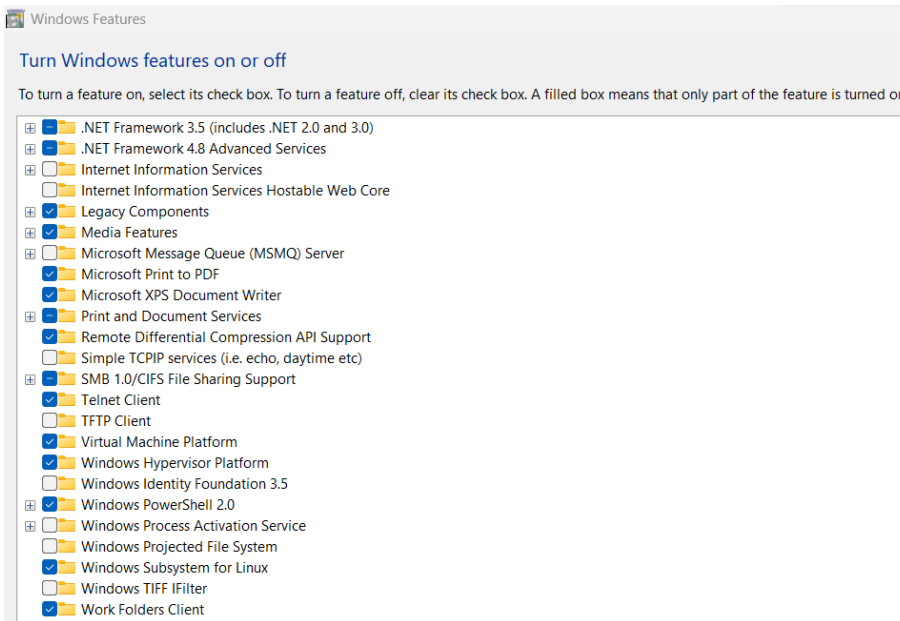
Q>Create a Hadoop multi-node cluster using either Virtual Machines or Docker.

Permissions Required for windows to install Docker.

Start→ Control Panel→Programs→Programs and Features→Turn on Windows features on or off→

Enable the Following:

- windows subsystem for Linux
- Windows Hypervisor Platform



Restart the System.

Docker Installation Link 64bit: <https://docs.docker.com/desktop/install/windows-install/>

Download the Dataset from Classroom:

<https://drive.google.com/file/d/1Ja6fD9ico9emdX8YNBioqOK1fD4L8g97/view?usp=sharing>

Extract the ZIP File

Open the folder –go to specific files –in file manger path ----remove path and write Powershell

Downloads > docker-hadoop > docker-hadoop-master

Name	Date modified	Type	Size
Today			
.gitignore	7/29/2023 11:08 AM	GITIGNORE File	1 KB
docker-compose.yml	7/29/2023 11:08 AM	YML File	2 KB
docker-compose-v3.yml	7/29/2023 11:08 AM	YML File	3 KB
hadoop.env	7/29/2023 11:08 AM	ENV File	3 KB
Makefile	7/29/2023 11:08 AM	File	2 KB
README.md	7/29/2023 11:08 AM	MD File	3 KB
historyserver	7/29/2023 11:08 AM	File folder	
namenode	7/29/2023 11:08 AM	File folder	
nginx	7/29/2023 11:08 AM	File folder	
nodemanager	7/29/2023 11:08 AM	File folder	
resourcemanager	7/29/2023 11:08 AM	File folder	
submit	7/29/2023 11:08 AM	File folder	
base	7/29/2023 11:08 AM	File folder	
datanode	7/29/2023 11:08 AM	File folder	

Docker-compose.yml contains 5 Data nodes:

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

volumes:

- hadoop_namenode:/hadoop/dfs/name

environment:

- CLUSTER_NAME=test

env_file:

- ./hadoop.env

datanode01:

image: bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8

container_name: datanode01

restart: always

volumes:

- hadoop_datanode01:/hadoop/dfs/data01

environment:

SERVICE_PRECONDITION: "namenode:9870"

env_file:

- ./hadoop.env

datanode02:

image: bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8

container_name: datanode02

restart: always

volumes:

- hadoop_datanode02:/hadoop/dfs/data02

environment:

SERVICE_PRECONDITION: "namenode:9870"

env_file:

- ./hadoop.env

datanode03:

image: bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8

container_name: datanode03

restart: always

volumes:

- hadoop_datanode03:/hadoop/dfs/data03

environment:

SERVICE_PRECONDITION: "namenode:9870"

env_file:

- ./hadoop.env

datanode04:

image: bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8

container_name: datanode04

restart: always

volumes:

- hadoop_datanode04:/hadoop/dfs/data04

environment:

SERVICE_PRECONDITION: "namenode:9870"

env_file:

- ./hadoop.env

datanode05:

image: bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8

container_name: datanode05

restart: always

volumes:

- hadoop_datanode05:/hadoop/dfs/data05

environment:

SERVICE_PRECONDITION: "namenode:9870"

env_file:

- ./hadoop.env

resourcemanager:

image: bde2020/hadoop-resourcemanager:2.0.0-hadoop3.2.1-java8

container_name: resourcemanager

restart: always

environment:

SERVICE_PRECONDITION: "namenode:9000 namenode:9870 datanode:9864"

env_file:

- ./hadoop.env

nodemanager1:

image: bde2020/hadoop-nodemanager:2.0.0-hadoop3.2.1-java8

container_name: nodemanager

restart: always

environment:

SERVICE_PRECONDITION: "namenode:9000 namenode:9870 datanode:9864 resourcemanager:8088"

env_file:

- ./hadoop.env

historyserver:

image: bde2020/hadoop-historyserver:2.0.0-hadoop3.2.1-java8

container_name: historyserver

restart: always

environment:

SERVICE_PRECONDITION: "namenode:9000 namenode:9870 datanode:9864 resourcemanager:8088"

volumes:

- hadoop_historyserver:/hadoop/yarn/timeline

env_file:

- ./hadoop.env

volumes:

hadoop_namenode:

hadoop_datanode01:

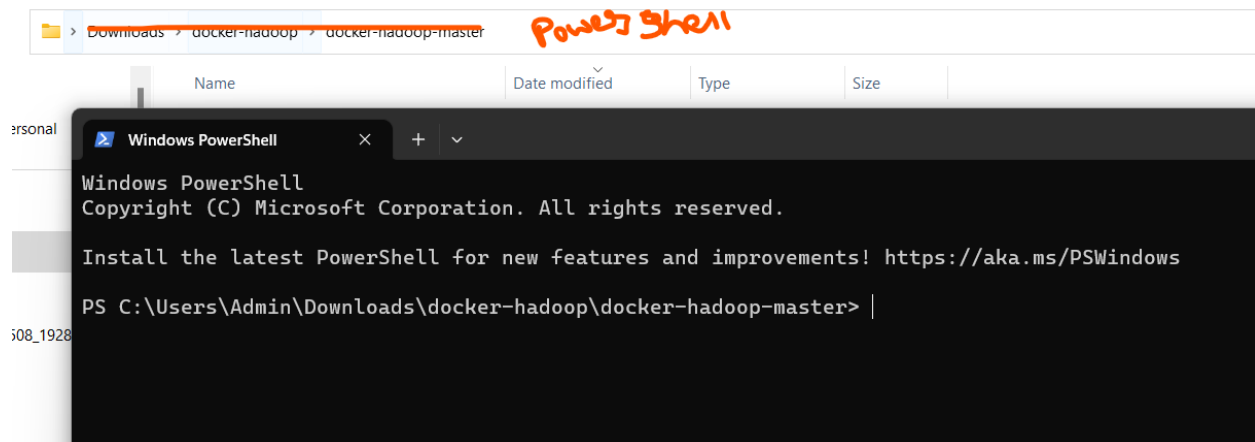
hadoop_datanode02:

hadoop_datanode03:

hadoop_datanode04:

hadoop_datanode05:

hadoop_historyserver:



In Command Prompt type:

- Docker compose up -d
- ls

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\Admin\Downloads\docker-hadoop\docker-hadoop-master> docker compose up -d
[+] Running 0/0
  - Volume "docker-hadoop-master_hadoop_datanode02" Creating                                0.0s
[+] Running 14/19:50:03+05:30 level=warning msg="Found orphan containers ([datanode]) for this project. If you remove Volume "docker-hadoop-master_hadoop_datanode02" Created                                0.0s
  ✓ Volume "docker-hadoop-master_hadoop_datanode05" Created                                0.0s
  ✓ Volume "docker-hadoop-master_hadoop_datanode04" Created                                0.0s
  ✓ Volume "docker-hadoop-master_hadoop_datanode03" Created                                0.0s
  ✓ Volume "docker-hadoop-master_hadoop_datanode01" Created                                0.0s
  ✓ Container datanode01 Started                                                            4.2s
  ✓ Container datanode02 Started                                                            4.2s
  ✓ Container datanode03 Started                                                            4.2s
  ✓ Container datanode05 Started                                                            4.2s
  ✓ Container datanode04 Started                                                            4.2s
  ✓ Container resourceanalyzer Running                                                       0.0s
  ✓ Container namenode Running                                                             0.0s
  ✓ Container nodemanager Running                                                           0.0s
  ✓ Container historyserver Running                                                         0.0s
PS C:\Users\Admin\Downloads\docker-hadoop\docker-hadoop-master> ls

Directory: C:\Users\Admin\Downloads\docker-hadoop\docker-hadoop-master

Mode                LastWriteTime         Length Name
----                -
d-----          7/29/2023 11:08 AM             base
d-----          7/29/2023 11:08 AM          datanode
d-----          7/29/2023 11:08 AM    historyserver
d-----          7/29/2023 11:08 AM      namenode
d-----          7/29/2023 11:08 AM         nginx
d-----          7/29/2023 11:08 AM    nodemanager
d-----          7/29/2023 11:08 AM  resourceanalyzer
d-----          7/29/2023 11:08 AM        submit
-a-----          7/29/2023 11:08 AM              6 .gitignore
```

- docker-compose ps
- ls

```
PS C:\Users\Admin\Downloads\docker-hadoop\docker-hadoop-master> docker-compose ps
NAME                PORTS                IMAGE                COMMAND                SERVICE                CREATED                STATUS
datanode            9864/tcp             bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8  "/entrypoint.sh /run..." datanode                2 weeks ago           Up 9 hours (he
datanode01         9864/tcp             bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8  "/entrypoint.sh /run..." datanode01             9 minutes ago         Up 8 minutes (
datanode02         9864/tcp             bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8  "/entrypoint.sh /run..." datanode02             9 minutes ago         Up 8 minutes (
datanode03         9864/tcp             bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8  "/entrypoint.sh /run..." datanode03             9 minutes ago         Up 8 minutes (
datanode04         9864/tcp             bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8  "/entrypoint.sh /run..." datanode04             9 minutes ago         Up 8 minutes (
datanode05         9864/tcp             bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8  "/entrypoint.sh /run..." datanode05             9 minutes ago         Up 8 minutes (
historyserver      8188/tcp             bde2020/hadoop-historyserver:2.0.0-hadoop3.2.1-java8  "/entrypoint.sh /run..." historyserver           2 weeks ago           Up 9 hours (he
namenode           0.0.0.0:9000->9000/tcp, 0.0.0.0:9870->9870/tcp bde2020/hadoop-namenode:2.0.0-hadoop3.2.1-java8  "/entrypoint.sh /run..." namenode                2 weeks ago           Up 9 hours (he
nodemanager        8042/tcp             bde2020/hadoop-nodemanager:2.0.0-hadoop3.2.1-java8  "/entrypoint.sh /run..." nodemanager1           2 weeks ago           Up 9 hours (he
resourceanalyzer   8088/tcp             bde2020/hadoop-resourceanalyzer:2.0.0-hadoop3.2.1-java8  "/entrypoint.sh /run..." resourceanalyzer        2 weeks ago           Up 9 hours (he
PS C:\Users\Admin\Downloads\docker-hadoop\docker-hadoop-master> ls

Directory: C:\Users\Admin\Downloads\docker-hadoop\docker-hadoop-master

Mode                LastWriteTime         Length Name
----                -
d-----          7/29/2023 11:08 AM             base
d-----          7/29/2023 11:08 AM          datanode
```

- docker network list
- docker network inspect docker-hadoop-master_default

```
PS C:\Users\Admin\Downloads\docker-hadoop\docker-hadoop-master> docker network list
NETWORK ID          NAME                DRIVER              SCOPE
e5510d1ea236        bridge             bridge             local
64952dd1876f        docker-hadoop-master_default bridge             local
b6c6e4048f43        host              host              local
5b37308a827d        none              null              local
PS C:\Users\Admin\Downloads\docker-hadoop\docker-hadoop-master> docker network inspect docker-hadoop-master_default
[
  {
    "Name": "docker-hadoop-master_default",
    "Id": "64952dd1876f2ddab4867c008c6ac8cffb53d7ef5397075f4a8406ada71d2c68",
    "Created": "2023-07-15T06:49:32.058793556Z",
    "Scope": "local",
    "Driver": "bridge",
    "EnableIPv6": false,
    "IPAM": {
      "Driver": "default",
      "Options": null,
      "Config": [
        {
          "Subnet": "172.18.0.0/16",
          "Gateway": "172.18.0.1"
        }
      ]
    },
    "Internal": false,
    "Attachable": false,
    "Ingress": false,
    "ConfigFrom": {
      "Network": ""
    },
    "ConfigOnly": false,
    "Containers": {

```

- docker exec -it namenode /bin/bash
- ls

```
PS C:\Users\Admin\Downloads\docker-hadoop\docker-hadoop-master> docker exec -it namenode /bin/bash
root@a6d68f55202f:/# ls
KEYS bin boot dev entrypoint.sh etc hadoop hadoop-data home lib lib64 media mnt opt proc root run run.sh sbin srv sys tmp usr var
root@a6d68f55202f:/#
```

SERVICE_PRECONDITION: "namenode:9870"

Open the URL to check nodes are running or not:

<http://localhost:9870/dfshealth.html#tab-overview>

Hadoop	Overview	Datanodes	Datanode Volume Failures	Snapshot	Startup Progress	Utilities ▾
--------	----------	-----------	--------------------------	----------	------------------	-------------

Overview 'namenode:9000' (active)

Started:	Sat Jul 29 11:14:14 +0530 2023
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Tue Sep 10 21:26:00 +0530 2019 by rohitsharmaks from branch-3.2.1
Cluster ID:	CID-edec7259-8f3a-48c7-ba93-795bff67671a
Block Pool ID:	BP-1692714887-172.18.0.6-1689403781090

Summary

Security is off.

Safemode is off.

14 files and directories, 7 blocks (7 replicated blocks, 0 erasure coded block groups) = 21 total filesystem object(s).

Heap Memory used 50.04 MB of 288 MB Heap Memory. Max Heap Memory is 841 MB.

Non Heap Memory used 60.17 MB of 62.06 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	5.9 TB
----------------------	--------

Open Docker and Check

Containers

Images

Volumes

Dev Environments BETA

Docker Scout EARLY ACCESS

Learning Center

Extensions

Add Extensions

docker-hadoop-master

C:\Users\Admin\Downloads\docker-hadoop\docker-hadoop-master

Open

major services

bde2020/hadoop-histo

Running

resource manager

bde2020/hadoop-resou

Running

namenode

bde2020/hadoop-name

Running

9000:9000

Show all ports (2)

datanode02

bde2020/hadoop-datar

Running

datanode03

bde2020/hadoop-datar

Running

datanode05

bde2020/hadoop-datar

Running

datanode01

bde2020/hadoop-datar

Running

datanode04

bde2020/hadoop-datar

Running

1689403781090:blk_1073741833_1009 src: /172.18.0.11:32952 dest: /172.18.0.9:9866 of size 2

2023-07-29 19:50:39 datanode

2023-07-29 14:20:39,179 INFO datanode.DataNode: DataNodeRegistration(172.18.0.4:9866, datanodeUuid=2f5ac87a-7d7c-4f45-9bab-3604ad7e55ca, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57; cid=CID-edec7259-8f3a-48c7-ba93-793bfff67671a;nsid=1378724001;c=1689403781090) Starting thread to transfer BP-1692714887-172.18.0.6-1689403781090:blk_1073741835_1011 to 172.18.0.8:9866 172.18.0.10:9866

2023-07-29 19:50:39 datanode

2023-07-29 14:20:39,182 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false

2023-07-29 19:50:39 datanode05

2023-07-29 14:20:39,185 INFO datanode.DataNode: Receiving BP-1692714887-172.18.0.6-1689403781090:blk_1073741835_1011 src: /172.18.0.4:58288 dest: /172.18.0.8:9866

2023-07-29 19:50:39 datanode

2023-07-29 14:20:39,188 INFO datanode.DataNode: DataTransfer, at 9c58deacea9a:9866 : Transmitted BP-1692714887-172.18.0.6-1689403781090:blk_1073741835_1011 (numBytes=17) to 28f050797dec/172.18.0.8:9866

2023-07-29 19:50:39 datanode05

2023-07-29 14:20:39,189 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false

2023-07-29 19:50:39 datanode05

2023-07-29 14:20:39,193 INFO datanode.DataNode: Received BP-1692714887-172.18.0.6-1689403781090:blk_1073741835_1011 src: /172.18.0.4:58288 dest: /172.18.0.8:9866 of size 17

2023-07-29 19:50:39 datanode03

2023-07-29 14:20:39,204 INFO datanode.DataNode: Receiving BP-1692714887-172.18.0.6-1689403781090:blk_1073741835_1011 src: /172.18.0.8:32952 dest: /172.18.0.10:9866

2023-07-29 19:50:39 datanode03

2023-07-29 14:20:39,206 INFO datanode.DataNode: Received BP-1692714887-172.18.0.6-1689403781090:blk_1073741835_1011 src: /172.18.0.8:32952 dest: /172.18.0.10:9866 of size 17

2023-07-29 19:54:03 nodemanager

2023-07-29 14:24:03,292 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0

2023-07-29 19:54:03 historyserver

2023-07-29 14:24:03,301 INFO timeline.LevelDbTimelineStore: Discarded 0 entities for timestamp 1690035843241 and earlier in 0.042 seconds

2023-07-29 19:54:29 nodemanager

2023-07-29 14:24:29,101 INFO tracker.NMLogAggregationStatusTracker: Rolling over the cached log aggregation status.

2023-07-29 19:59:03 historyserver

2023-07-29 14:29:03,447 INFO timeline.LevelDbTimelineStore: Discarded 0 entities for timestamp 1690036143422 and earlier in 0.017 seconds

2023-07-29 20:04:03 nodemanager

2023-07-29 14:34:03,671 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0

2023-07-29 20:04:03 historyserver

2023-07-29 14:34:03,739 INFO timeline.LevelDbTimelineStore: Discarded 0 entities for timestamp 1690036443724 and earlier in 0.01 seconds

2023-07-29 20:04:29 nodemanager

2023-07-29 14:34:29,397 INFO tracker.NMLogAggregationStatusTracker: Rolling over the cached log aggregation status.

Node	Http Address	Last contact	Last Block Report	Capacity	Blocks	Block pool used	Version
✓28f050797dec:9866 (172.18.0.8:9866)	http://28f050797dec:9864	0s	24m	1006.85 GB	2	52 KB (0%)	3.2.1
✓5bb9c9d1c2fe:9866 (172.18.0.10:9866)	http://5bb9c9d1c2fe:9864	0s	24m	1006.85 GB	2	52 KB (0%)	3.2.1
✓9c58deacea9a:9866 (172.18.0.4:9866)	http://9c58deacea9a:9864	0s	86m	1006.85 GB	7	92 KB (0%)	3.2.1
✓b6fd4b4c473d:9866 (172.18.0.9:9866)	http://b6fd4b4c473d:9864	0s	24m	1006.85 GB	3	60 KB (0%)	3.2.1
✓cf5eb79be626:9866 (172.18.0.11:9866)	http://cf5eb79be626:9864	0s	24m	1006.85 GB	4	68 KB (0%)	3.2.1
✓f7dcf39d08d5:9866 (172.18.0.7:9866)	http://f7dcf39d08d5:9864	0s	24m	1006.85 GB	3	60 KB (0%)	3.2.1

Showing 1 to 6 of 6 entries

Previous1Next